REARRANGING ESTIMATORS OF THE VALUE-AT-RISK AND OTHER RISK MEASURES

VICTOR CHERNOZHUKOV† IVÁN FERNÁNDEZ-VAL§ ALFRED GALICHON‡

Abstract. Quantile estimation procedures such as Quantile Regression (QR) have become widely used in quantitative finance over the last years, in large part because they are a basic building tool for risk measures such as VaR estimation or TailVaR. A problem which has been documented with the QR procedure is that the estimated curves do not necessarily satisfy the natural property of monotonicity, which can lead the Value-at-Risk to fail to be increasing with respect to the confidence level as it should logically be. For instance, for a given portfolio, the estimated VaR at level 90% might be greater than the estimated VaR at level 95%. The present paper shows how to fix this problem by applying a procedure called increasing rearrangement to the (possibly non-monotone) quantiles estimates. We show that this procedure enjoys the most desirable approximation property that the rearranged quantile curve is “closer” to the true curve than the initial estimator. We also compute the asymptotic statistical properties of the rearrangement procedure in large sample.

1. Introduction and motivation

This paper shows how to address a common problem in quantile estimation, arising notably when trying to estimate a Value-at-Risk (VaR) level: the increasing feature of quantile functions is not robust to finite-sample error or misspecification in the estimator.

VaR and Quantile Estimation. Over a little more than a decade, the Value-at-Risk has arguably become the single mostly used tool in financial risk management. From its adoption by financial institutions in the early 1990’s for internal control purposes, until its recommendation by the Basel I and II Committees on Banking Supervision, and Solvency II for insurance, this tool has attracted a great number of research on its statistical properties and its impact on financial management.

Date: First version is August 12, 2007. The present version is of November 4, 2007.
Given a contingent loss $Y$ (we take the convention that a positive $Y$ denotes a net loss), we denote its distribution function by $F_Y$ and its quantile function $Q_Y(t)$, given by $Q_Y(t) = F_Y^{-1}(t)$. The Value-at-Risk associated with payoff $Y$ at level $\alpha$ is the minimal amount of capital $x = \text{VaR}_\alpha(Y)$ which fails to cover the losses of $Y$ less than a fraction $\alpha$ of the times. In statistical terms, it is the quantile function evaluated at $\alpha$, namely

$$\text{VaR}_\alpha(Y) = Q_Y(\alpha).$$

A number of other risks measures of less popular use can also be constructed using the quantile function. Indeed, [Artzner et al. (1999)] have shown that a desirable property for a risk measure is convexity, which VaR fails to fulfil. Among measures which have this property, the Expected Shortfall (sometimes known as TailVaR), which is the expected value of a loss in the event of a loss, is computed as

$$\text{ES}_\alpha(Y) = \frac{1}{1-\alpha} \int_0^1 (Q_Y(u) - Q_Y(\alpha)) du,$$

and more generally, it has been shown that a class of risk measures with the natural properties of regularity and coherence (see Appendix A for definitions and references) is given by

$$\rho(Y) = \int_0^1 \varphi(u) Q_Y(u) du,$$

where $\varphi$ is a nondecreasing function.

Reliable procedures for estimating quantile functions appear therefore as an essential component in risk measurement, to compute the VaR or derived risk measures. And while the VaR only requires the estimation of the quantile function at a single level, other risk measures require the estimation of the quantile function on a part of- or the whole interval $[0,1]$: for instance, the expected shortfall at level $\alpha$ requires the estimation of the $\text{VaR}$ number at every significance level beyond $\alpha$. A number of methodologies for performing such estimation have been proposed to do so. A first class of approaches use a factor model such as RiskMetrics to forecast the expectation and the standard deviation of the payoff. Under the assumption of normality, the VaR is then a linear expression in these parameters. The assumption of normality is a strong one however, and another class of procedures have been proposed, based on the estimation of historical conditional quantiles of the value of the
portfolio under consideration. Such estimation is typically done nonparametrically using the Quantile Regression: given a conditioning variable $X$, the model specification is

$$Q_{Y|X}(u|x) = x'\beta(u)$$

where $x \in \mathbb{R}^d$ is the conditioning variable, and $\beta : [0, 1] \to \mathbb{R}^d$ is a function to be estimated. For the use of Quantile Regression in a Value-at-Risk context, cf. for instance [Chernozhukov and Umantsev (2001)], and [Engle and Manganelli (2007)] in an autoregressive context. It has been shown by [Koenker and Bassett (1978)] that $\beta(t)$ can be estimated using the following linear optimization, based on the least asymmetric absolute deviation problem:

$$\hat{\beta}(t) = \arg \min_{\beta \in \mathbb{R}^d} \sum_{k=1}^n \rho_t(Y_k - X_k'\beta)$$

where $\rho_t$ is the loss function $\rho_t(u) = tu^+ + (1-t)u^-$. The estimator of the quantile curve is then

$$\hat{Q}_{Y|X}(t|x) = x'\hat{\beta}(t).$$

However, a problem that has been documented with quantile regression is that the map $t \mapsto x'\hat{\beta}(t)$ might not be increasing with respect with $t$, either due to sample finiteness or to misspecification of the Quantile Regression (see eg. [He (1997)]). In other words for a same risk $Y$, the estimated VaR at level 90% might be greater than the estimated VaR at level 95%. In a Value-at-Risk context, this problem and its potential side-effects on financial risk management has been documented eg. in [Gourieroux and Jasiak (2007)]. The problem studied here is not particular to quantile regression, but arises in a number of procedures when estimating either quantile or distribution curves.

**Restoring monotonicity.** The purpose of the present paper is to show how to restore the natural monotonicity property of the quantile curves using a tool known in Variational Analysis under the name of Increasing Rearrangement, and to study both the statistical asymptotic properties and the approximation properties of the operation. Consider the random variable $Y_x := \hat{Q}(U|x)$ where $U \sim U(0, 1)$. The distribution function of this variable is given by

$$\hat{F}(y|x) = \int_0^1 1\{\hat{Q}(u|x) \leq y\}du.$$
which after inversion, leads to a proper (i.e. nondecreasing) quantile function quantile function

\[
\hat{F}^{-1}(u|x) = \inf\{y : \hat{F}(y|x) \geq u\}.
\]

If the original curve is increasing in \(u\), then the rearranged quantile function \(\hat{F}^{-1}(u|x)\) coincides with the original curve \(\hat{Q}(u|x)\), but when this is not the case, both curves differ. Thus, starting with a possibly non-monotone original curve \(u \mapsto \hat{Q}(u|x)\), the rearrangement procedure described above produces a monotone quantile curve \(u \mapsto \hat{F}^{-1}(u|x)\).

The present paper establishes the empirical properties of the rearranged quantile curves \(u \mapsto \hat{F}^{-1}(u|x)\). We show first an approximation result: the rearranged curve \(\hat{F}^{-1}(u|x)\) is closer to the true conditional quantile curve \(Q_0(u|x)\) than the original curve. Formally, for each \(x\), we have that for all \(p \in [1, \infty]\)

\[
\left( \int_{\mathcal{U}} |Q_0(u|x) - \hat{F}^{-1}(u|x)|^p du \right)^{1/p} \leq \left( \int_{\mathcal{U}} |Q_0(u|x) - \hat{Q}(u|x)|^p du \right)^{1/p}.
\]

Importantly, the inequality is strict for \(p \in (1, \infty)\) whenever \(u \mapsto \hat{Q}(u|x)\) is decreasing on a subset of \(\mathcal{U} := (0, 1)\) of positive Lebesgue measure, while \(u \mapsto Q_0(u|x)\) is strictly increasing. This property is independent of the sample size, and thus continues to hold in the population, and also regardless of whether the linear quantile estimator \(x'\hat{\beta}(u)\) estimates \(Q_0(u|x)\) consistently or not, i.e., whether \(Q_0(u|x) = x'\beta(u)\) or \(Q_0(u|x) \neq x'\beta(u)\). In other words, the rearranged quantile curves have smaller estimation error than the original curves whenever the latter are not monotone. This property does not depend on the way the quantile model is estimated.

To introduce the rest of results, let us fix the value of the regressor \(X\) to \(x\) and suppose that the estimator satisfies a functional central limit theorem, that is, weakly converges to a Gaussian process \(G(u|x)\),

\[
\sqrt{n}(\hat{Q}(u|x) - Q(u|x)) \Rightarrow G(u|x),
\]

as a stochastic process indexed by \(u\) in the metric space of bounded functions \(\ell^\infty(0, 1)\), and where \(n\) is the sample size. This condition holds for the basic linear quantile regression and also holds for censored, instrumental, and other types of quantile regressions (cf. discussion in [Angrist et al. (2006)]).
The second main result of the paper is that

$$\sqrt{n}(\hat{F}^{-1}(u|x) - F^{-1}(u|x)) \Rightarrow G(u|x), \quad (1.4)$$

as a stochastic process indexed by $u$, in $\ell^\infty(0, 1)$; which, remarkably, coincides with the first order asymptotics (1.3) of the original curve. This result has a convenient practical implication: if the population curve is monotone, then the empirical non-monotone curve can be re-arranged to be monotonic without affecting its (first order) asymptotic properties. To derive the above results we find the functional Hadamard derivative of $F^{-1}(u|x)$ with respect to perturbations of the underlying curve $x'/\beta(u)$ in discontinuous directions, tangentially to the set of continuous functions, and then use the functional delta method. Establishing the Hadamard differentiability of the rearranged distribution and quantile curves in discontinuous directions is the key theoretical result which allows to relate the first order asymptotic behavior of the rearranged object to that of the original estimated object.

**Literature review.** [Koenker and Bassett (1978)] were the first to introduce Quantile Regression. For an overview on the topic, and an account of the relevant literature, we refer to [Koenker (2005)]. See the papers by [Chernozhukov and Umantsev (2001)] as well as [Gourieroux and Jasiak (2007)] and [Engle and Manganelli (2007)] for application of Quantile Regression to VaR estimation, and [Bassett et al. (2004)] for general convex risk measures. Many authors have investigated the sensitivity of the Value-at-Risk estimators with respect to various estimation parameters. [Gourieroux and Liu (2006)] have investigated the sensitivity of a family of convex risk measures with respect to a parameter controlling the convexity of the measure. [Gourieroux et al. (2000)] have investigated the sensitivity of the VaR with respect to the portfolio allocation.

There is a large literature on the Rearrangement operator in Variational Analysis, for which we refer to the literature review in [Chernozhukov et al. (2006a)]. For statistical purposes, let us mention the important papers of Dette, Neumeyer, and Pilz (2006), who previously applied rearrangements to kernel mean regressions and derived their pointwise asymptotic normality; Davydov and Zitikis (2005), which considered tests of monotonicity based on rearranged kernel regression; Fougeres (1997), which used the rearrangement for
density estimation; and Neumeyer (2007), which proved several consistency results. The idea of the rearrangement operator is also implicitly present in [Gourieroux and Liu (2006)].

Several authors have previously recognized and addressed the problem of non-monotonicity of the VaR estimator with respect to the confidence level, in links with properties of Quantile Regression. [He (1997)] first documented the problem and proposed to impose a location-scale regression model, which naturally satisfies monotonicity. In a context motivated by VaR estimation, [Gourieroux and Jasiak (2007)] proposed to project the empirical distribution to a parametric class of proper (increasing) quantiles using the minimization of a Kullback-Leibler divergence.

Organization of the paper. The rest of the paper is organized as follows. Section 2.1 gives useful basic results regarding the rearrangement operator. Section 2.2 gives the first order variation of this operator, more precisely its Hadamard derivative, which is suitable for statistical purposes. Section 2.3 provides empirical properties of the operator, both in finite sample and in large sample. Section 3 gives a numerical illustration of the method, and section 4 concludes the paper. Appendix A recalls some classical definitions and results on risk measures. The proof of all the mathematical statements are given in Appendix B. A formal justification for the main result is given in Appendix C.

2. Properties of the Monotone Rearrangement Procedure

In this section the treatment of the problem is somewhat more general than in the introduction; namely, we replace the linear functional form $x'\beta(u)$ by $Q(u, x)$. Define $Y_x = Q(U, x)$, where $U \sim \text{Uniform}(U)$ with $U = (0,1)$. Let $F(y|x) = \int_0^1 1\{Q(u, x) \leq y\}du$ be the distribution function of $Y_x$, and $F^{-1}(u|x)$ be the quantile function of $Y_x$.

2.1. Some Basic Properties. We start by developing some basic properties for the population counterparts of the rearranged distribution curve $F(y|x)$ and its inverse $F^{-1}(u|x)$.

Recall first some definitions. Let $g : \mathcal{U} \subset \mathbb{R} \rightarrow \mathbb{R}$ a continuously differentiable function. A point $u \in \mathcal{U}$ is called a regular point of $g$ if the derivative of $g$ at this point does not vanish, i.e., $g'(u) \neq 0$. A point $u$ which is not a regular point is called a critical point. A
value \( y \in g(\mathcal{U}) \) is called a regular value of \( g \) if \( g^{-1}(\{y\}) \) contains only regular points, i.e., if \( \forall u \in g^{-1}(\{y\}), g'(u) \neq 0 \). A value \( y \) which is not a regular value is called a critical value.

Denote by \( Y_x \) the support of \( Y_x \), and let \( Y^*_x \) be the subset of regular values of \( u \mapsto Q(u, x) \) in \( Y_x \). Denote \( \mathcal{Y} = \{(y, x) : y \in Y_x, x \in \mathcal{X}\} \), \( \mathcal{Y}^* = \{(y, x) : y \in Y^*_x, x \in \mathcal{X}\} \), and \( \mathcal{U} \mathcal{X} = \{(u, x) : (F^{-1}(u|x), x) \in \mathcal{Y}\} \). We assume throughout that \( Y_x \subset \mathcal{Y} \), which is compact subset of \( \mathbb{R} \), and that \( x \in \mathcal{X} \), a compact subset of \( \mathbb{R}^d \).

We make the following assumptions about \( Q(u, x) \):

(a) \( Q(u, x) : \mathcal{U} \times \mathcal{X} \to \mathbb{R} \) is a continuously differentiable function in both arguments,
(b) \( Q'(u, x) := \partial Q(u, x)/\partial u \) does not vanish almost everywhere on \( \mathcal{U} \), for each \( x \in \mathcal{X} \),
(c) the number of times \( \partial Q(u, x)/\partial u \) changes its sign is bounded, uniformly in \( x \in \mathcal{X} \).

In some applications, the curves of interest are not functions of \( x \), or we might be interested in a particular value \( x \). In this case, the set \( \mathcal{X} \) is taken to be a singleton \( \mathcal{X} = \{x\} \).

**Proposition 1** (Basic Properties of \( F(y|x) \) and \( F^{-1}(u|x) \)). Under assumptions (a) - (c), the functions \( F(y|x) \) and \( F^{-1}(u|x) \) satisfy the following properties:

1. The set of critical values, \( Y_x \setminus Y^*_x \), is finite, and \( \int_{Y_x \setminus Y^*_x} dF(y|x) = 0 \).

2. For any \( y \in Y^*_x \)

\[
F(y|x) = \int_{Q(u,x) \leq y} du = \sum_{k=1}^{K(y,x)} \text{sign}\{Q'(u_k(y, x), x)\} u_k(y, x) + 1\{Q'(u_{K(y,x)}, x) < 0\},
\]

where \( \{u_k(y, x), \text{ for } k = 1, ..., K(y,x) < \infty\} \) are the roots of \( Q(u, x) = y \) in increasing order.

3. For any \( y \in \mathcal{Y}^* \), the ordinary derivative \( f(y|x) = \partial F(y|x)/\partial y \) exists and takes the form

\[
f(y|x) = \sum_{k=1}^{K(y,x)} \frac{1}{|Q'(u_k(y, x), x)|},
\]

which is continuous at each \( y \in \mathcal{Y}^*_x \). \( F(y|x) \) is absolutely continuous in \( y \in \mathcal{Y} \) and is strictly increasing in \( y \in \mathcal{Y} \).
4. The quantile function $F^{-1}(u|x)$ partially coincides with $Q(u, x)$ on regions where the latter function is increasing; namely

$$F^{-1}(u|x) = Q(u, x),$$

provided that the equation $Q(u, x) = y$ has unique solution for $y = F^{-1}(u|x)$.

5. The quantile function $F^{-1}(u|x)$ is equivariant to location and scale transformations of $Q(u, x)$.

6. The quantile function $F^{-1}(u|x)$ has the ordinary continuous derivative

$$\frac{1}{f(F^{-1}(u|x)|x)},$$

when $F^{-1}(u|x) \in \mathcal{Y}^*_x$, and 0 when $F^{-1}(u|x) \in \mathcal{Y}_x \setminus \mathcal{Y}^*_x$. This function is also a Radon-Nikodym derivative with respect to the Lebesgue measure.

7. The map $(y, x) \mapsto F(y|x)$ is continuous on $\mathcal{X} \mathcal{Y}$ and the map $(u, x) \mapsto F^{-1}(u|x)$ is continuous on $\mathcal{X} \mathcal{U}$.

The following synthetic simple example illustrates some of these basic properties in a situation where the initial “approximated” population quantile curve is chosen to be highly non monotone. Consider the following hypothetical quantile function:

$$Q(u) = 5 \left( x + \frac{1}{\pi} \sin(2\pi u) \right).$$

(2.1)

The left panel of Figure 1 shows that this function is non monotone in $[0, 1]$. In particular, the slope of $Q(u)$ changes sign twice at $1/3$ and $2/3$. Both panels of Figure 1 illustrate results 1, 2, 4, and 7 of the proposition by plotting together the original and rearranged quantile and distribution curves. Here we can see that the rearranged functions are continuous and monotonically increasing, do not have mass points, and coincide with the original counterparts for points where the initial distribution is uniquely defined. For the rest of the points $y$, the rearranged distribution is a linear combination of the solutions to the equation $Q(u) = y$. The coefficients in this linear combination alternate between 1 and -1 depending on whether the original quantile curve is increasing or decreasing at the corresponding solution point.
Figure 1. Left: The pseudo-quantile function $Q(u)$ and the rearranged quantile function $F^{-1}(u)$. Right: The pseudo-distribution function $Q^{-1}(y)$ and the rearranged distribution function $F(y)$.

Figure 2. Left: The density (sparsity) function of the rearranged quantile function $F^{-1}(u)$. Right: The density function of the rearranged distribution function $F(y)$. 
Figures 2 illustrates the third and sixth results of the proposition by plotting the density functions of $F(y)$ and $F^{-1}(u)$, respectively. The density function for $F(y)$ in the left panel is continuous at the regular values of $Q(u)$, and grows asymptotically to infinity at the critical values. Similarly, the density function for $F^{-1}(u)$ in the right panel is continuous at the regular points of $Q(u)$, and takes the value zero at the critical points.

2.2. First variation. Next, we establish the main results of the paper on Hadamard differentiability of $F^{-1}(u|x)$ with respect to $Q(u, x)$, tangentially to the space of continuous functions on $U \times X$. This differentiability property is important for statistical applications, as it shall relate the first order approximation error of the rearranged quantile to that of the original estimator.

In the remainder of the paper we shall suppose that the function $u \mapsto Q(u, x)$ is a proper quantile, namely that it has $Q'(u, x) > 0$ for each $x \in X$. We shall denote by $\ell^\infty(U \times X)$ the set of bounded and measurable functions $h : U \times X \rightarrow \mathbb{R}$. $C(U \times X)$ denotes the set of continuous functions mapping $h : U \times X \rightarrow \mathbb{R}$, and $\ell^1(U \times X)$ denotes the set of measurable functions $h : U \times X \rightarrow \mathbb{R}$ such that $\int_U \int_X |h(u, x)| \, du \, dx < \infty$, where $du$ and $dx$ denote the integration with respect to the Lebesgue measure on $U$ and $X$, respectively.

**Proposition 2** (Hadamard Derivative of $F^{-1}(u|x)$ with respect to $Q(u, x)$). Define $F(y|x, h_t) := \int_0^1 1\{Q(u|x) + th_t(u|x) \leq y\} du$. Under assumptions (a) - (c), as $t \rightarrow 0$,

$$\tilde{D}_{h_t}(u, x, t) := \frac{F^{-1}(u|x, h_t) - F^{-1}(u|x)}{t} \rightarrow h$$

uniformly in $(u, x) \in U X$.

In practice, the object of our interest will often consist in integral expressions of $F^{-1}(u|x)$, as it is the case either if we are interested in an unconditional VaR (in which case we need to integrate $x$ over its forecasted distribution), or if we are interested in alternative risk measures such as Expected Shortfall (in which case we need to integrate over $y$). In fact, we have the following result, which follows from the above convergence over the entire domain:

**Corollary 1.** We have:

$$\int_U g(u|x, u') \tilde{D}_{h_t}(u|x, t) du \rightarrow \int_U g(u|x, u') h(u|x) du$$
uniformly in \((u', x) \in U\mathcal{X}\), for any measurable \(g\) such that \(\sup_{u', x} |g(u|x, u')| \in \ell^1(U)\) and such that \((x, u') \mapsto g(u|x, u')\) is continuous for a.e. \(u\).

2.3. Empirical Properties of \(\hat{F}^{-1}(u|x)\). We are now ready to state the main results for this section.

Proposition 3 (Improvement in Estimation Property Provided by Rearrangement). Suppose that \(\hat{Q}(\cdot | \cdot)\) is an estimator (not necessarily consistent) for some true quantile curve \(Q_0(\cdot | \cdot)\). Then, the rearranged curve \(\hat{F}^{-1}(u|x)\) is closer to the true curve than \(\hat{Q}(u|x)\) in the sense that, for each \(x \in \mathcal{X}\),

\[
\left( \int_{U} |Q_0(u|x) - \hat{F}^{-1}(u|x)|^p du \right)^{1/p} \leq \left( \int_{U} |Q_0(u|x) - \hat{Q}(u|x)|^p du \right)^{1/p}, \quad p \in [1, \infty],
\]

where the inequality is strict for \(p \in (1, \infty)\) whenever \(\hat{Q}(u|x)\) is decreasing on a subset of \(U\) of positive Lebesgue measure, while \(Q_0(u|x)\) is increasing on \(U\).

The above property is independent of the sample size and of the way the estimate of the curve is obtained, and thus continues to hold in the population.

Therefore, the rearranged estimated VaRs taken as function of the confidence level are closer to the true ones than the original estimators, in any reasonable functional norm. This means that a non-monotone estimator can never be closer to the true one than its rearranged version. The result is extremely general as this property holds true in finite sample and for any distribution.

We now turn to large sample results. The following proposition investigates the asymptotic distributions of the rearranged functions.

Proposition 4 (Limit Distribution of \((u, x) \mapsto \hat{F}^{-1}(u|x))\). Suppose that \(\hat{Q}(\cdot | \cdot)\) is an estimator for \(Q(\cdot | \cdot)\) that takes its values in the space of bounded measurable functions defined on \(U\mathcal{X}\), and that, in \(\ell^\infty(U\mathcal{X})\),

\[
\sqrt{n}(\hat{Q}(u|x) - Q(u|x)) \Rightarrow G(u|x),
\]
as a stochastic process indexed by \((u, x) \in \mathcal{UX}\), where \((u, x) \mapsto G(u|x)\) is a Gaussian process with continuous paths, \(n\) is the sample size. Assume that \(Q(u|x)\) satisfies the basic conditions (a) and (b) and is increasing.

Then in \(\ell^\infty(\mathcal{UX}_K)\), where \(K\) is any compact subset of \(\mathcal{YX}\), with \(\mathcal{UX}_K = \{(u, x) : (F^{-1}(u|x), x) \in K\}\),

\[
\sqrt{n}(\hat{F}^{-1}(u|x) - F^{-1}(u|x)) \Rightarrow G(u|x)
\]

as a stochastic process indexed by \((u, x) \in \mathcal{UX}_K\). Moreover the convergence is uniform over the entire \(\mathcal{YX}\) and \(\mathcal{UX}\).

The main condition is that the initial estimate satisfies a functional central limit theorem with a Gaussian limit process (that is not necessarily centered at zero). Thus the first order properties of the rearranged and initial quantile estimates coincide. Hence, all the inference tools that apply to the original quantile estimates also apply to the rearranged quantile estimates. In particular, if the bootstrap is valid for the original estimate, it is also valid for the rearranged estimate, by the functional delta method for the bootstrap.

Let us now consider the linear functionals of the rearranged quantile estimates. We deduce the following result:

**Corollary 2 (Empirical Properties of Integrated Curves).** Under the conditions of Proposition 6 and the same restrictions on the function \(g\) as in Proposition 4, the following results are true with the limits being continuous on the specified domains:

\[
\sqrt{n} \int_{\mathcal{U}} g(u|x, u')(\hat{F}^{-1}(u|x) - F^{-1}(u|x))du \Rightarrow \int_{\mathcal{U}} g(u|x, u')G(u|x)du,
\]

as stochastic process indexed by \((u', x) \in \mathcal{UX}\), in \(\ell^\infty(\mathcal{UX})\).

### 3. Illustration and Empirical Example

**INSERT HERE**

### 4. Conclusion

This paper shows how to deal with a problematic artefact with a popular VaR measurement method such as Quantile Regression, that the estimated level of the Value-at-Risk
associated with a given portfolio might not be increasing with respect to the confidence level. Not only does this violate a natural logical restriction; but it might also have side effects for management purposes, not to talk about negative impact on the communication on the tools towards the end user. Having started from a possibly non-monotone family of VaR estimators, the procedure proposed here produces a corrected family of estimators that not only satisfies the natural monotonicity requirement, but also has smaller estimation error than the original estimators. An asymptotic distribution theory is derived for the rearranged curves, and the relevance of the approach is illustrated with a simulation experiment.

† Massachusetts Institute of Technology, Department of Economics and Operations Research Center, University College London, CEMMAP. E-mail: vchern@mit.edu. Research support from the Castle Krob Chair, National Science Foundation, the Sloan Foundation, and CEMMAP is gratefully acknowledged.

‡ Boston University, Department of Economics. E-mail: ivanf@bu.edu.

§ Ecole Polytechnique, Département d’Economie. E-mail: alfred.galichon@polytechnique.edu.

APPENDIX A. COHERENT RISK MEASURES

Here we motivate the choice of a risk measure given above. In their seminal paper, [Artzner et al. (1999)] have introduced the following natural properties for a risk measure:

**Definition 1.** A functional \( \varrho : L_d \to \mathbb{R} \) is called a convex risk measure if it satisfies the following properties:

- Monotonicity: \( X \geq Y \Rightarrow \varrho(X) \leq \varrho(Y) \)
- Translation invariance: \( \varrho(X + m) = \varrho(X) + m \varrho(1) \)
- Convexity: \( \varrho(\lambda X + (1 - \lambda)Y) \leq \lambda \varrho(X) + (1 - \lambda)\varrho(Y) \) for all \( \lambda \in (0, 1) \).

**Definition 2.** A functional \( \varrho : L_1 \to \mathbb{R} \) is called a coherent risk measure if it convex and satisfies:

- Positive homogeneity: \( \varrho(\lambda X) = \lambda \varrho(X) \) for all \( \lambda \geq 0 \).
Definition 3. A functional $\varrho : L_1 \to \mathbb{R}$ is law-invariant if $\varrho(X) = \varrho(\tilde{X})$ whenever $L_X = L_{\tilde{X}}$.

In the case of univariate prospects, comonotonic additivity is used in addition to law invariance to define regular risk measures:

Definition 4. A functional $\varrho : L_1 \to \mathbb{R}$ is called a regular risk measure if it satisfies:

- Law invariance
- Comonotonic additivity: $\varrho(X + Y) = \varrho(X) + \varrho(Y)$ when $X, Y$ are comonotonic, i.e. weakly increasing transformation of each other.

[Kusuoka (2001)] has shown the following equivalence result:

Theorem 1. A regular risk measure $\varrho$ is coherent if and only if for some random variable $W$ with law $L_W = P$ with support in $\mathbb{R}$, and $U \sim U(0, 1)$ uniformly distributed on $(0, 1)$, we have

$$\varrho(X) := \varrho_P(X) = E[Q_W(U)Q_X(U)].$$

In particular, the TailVaR (expected shortfall), defined as

$$\text{TailVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 Q_X(u)du$$

is a convex risk measure. The VaR is, however, not a convex risk measure.

**Appendix B. Proofs of the results in the main text**

The proofs of the results given here are specialization of more general results given in [Chernozhukov et al. (2006a)]. They are restated here for the sake of completeness.

B.1. Proof of Proposition 1. First, note that the distribution of $Y_x$ has no atoms, i.e.,

$$\Pr[Y_x = y] = \Pr[Q(U|x) = y] = \Pr[U \in \{u \in \mathcal{U} : u \text{ is a root of } Q(u|x) = y\}] = 0,$$

since the number of roots of $Q(u|x) = y$ is finite under (a) - (b), and $U \sim \text{Uniform}(\mathcal{U})$. Next, by assumptions (a)-(b) the number of critical values of $Q(u|x)$ is finite, hence claim (1) follows.
Next, for any regular \( y \), we can write \( F(y|x) \) as
\[
\int_0^1 1\{Q(u|x) \leq y\} du = \sum_{k=0}^{K(y|x)-1} \int_{u_k(y|x)}^{u_{k+1}(y|x)} 1\{Q(u|x) \leq y\} du + \int_{u_{K(y|x)}(y|x)}^1 1\{Q(u|x) \leq y\} du,
\]
where \( u_0(y|x) := 0 \) and \( \{u_k(y|x), k = 1, \ldots, K(y|x) < \infty\} \) are the roots of \( Q(u|x) = y \) in increasing order. Note that the sign of \( Q'(u|x) \) alternates over consecutive \( u_k(y|x) \), determining whether \( 1\{Q(y|x) \leq y\} = 1 \) on the interval \([u_{k-1}(y|x), u_k(y|x)]\). Hence the first term in the previous expression simplifies to \( \sum_{k=0}^{K(y|x)-1} 1\{Q'(u_{k+1}(y|x)|x) \geq 0\}(u_{k+1}(y|x) - u_k(y|x)) \); while the last term simplifies to \( 1\{Q'(u_{K(y|x)}(y|x)|x) \leq 0\}(1 - u_{K(y|x)}(y|x)) \). An additional simplification yields the expression given in claim (2) of the proposition.

The proof of claim (3) follows by taking the derivative of expression in claim (2), noting that at any regular value \( y \) the number of solutions \( K(y|x) \) and \( \text{sign}(Q'(u_k(y|x)|x)) \) are locally constant; moreover,
\[
u_k(y|x) = \frac{\text{sign}(Q'(u_k(y|x)|x))}{|Q'(u_k(y|x)|x)|}.
\]
Combining these facts we get the expression for the derivative given in claim (3).

To show the absolute continuity of \( F(y|x) \) with \( f(y|x) \) being the Radon-Nykodym derivative, it suffices to show that for each \( y' \in \mathcal{Y}_x \), \( \int_{-\infty}^{y'} f(y|x) dy = \int_{-\infty}^{y'} dF(y|x) \), cf. Theorem 31.8 in Billingsley (1995). Let \( V_t^x \) be the union of closed balls of radius \( t \) centered on the critical points \( \mathcal{Y}_x \setminus \mathcal{Y}_x^\ast \), and define \( \mathcal{Y}_x^t = \mathcal{Y}_x \setminus V_t^x \). Then, \( \int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} f(y|x) dy = \int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} dF(y|x) \). Since the set of critical points \( \mathcal{Y}_x \setminus \mathcal{Y}_x^\ast \) is finite and has mass zero under \( F(y|x) \), \( \int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} dF(y|x) \) is finite as \( t \to 0 \). Therefore, \( \int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} f(y|x) dy \) is finite and has mass zero under \( F(y|x) \), \( \int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} f(y|x) dy \) is finite and has mass zero under \( F(y|x) \). Therefore, \( \int_{-\infty}^{y'} 1\{y \in \mathcal{Y}_x^t\} f(y|x) dy \) as \( t \to 0 \).

Claim (4) follows by noting that at the regions where \( s \to Q(s|x) \) is increasing and one-to-one, we have that \( F(y|x) = \int_{Q(s|x) \leq y} ds = \int_{s \leq Q^{-1}(y|x)} ds = \frac{1}{Q^{-1}(y|x)} \). Inverting the equation \( u = F(F^{-1}(u|x)|x) = \frac{1}{F^{-1}(u|x)|x} \) yields \( F^{-1}(u|x) = Q(u|x) \).

Claim (5). We have \( Y_x = Q(U|x) \) has quantile function \( F^{-1}(u|x) \). A quantile function is known to be equivariant to monotone increasing transformations, including location-scale transformations. Thus, this is true in particular for \( F^{-1}(u|x) \).

Claim (6) is immediate from claim (3).
Claim (7). The proof of continuity of \( F(y|x) \) is subsumed in the step 1 of the proof of Proposition 3 (see below). Therefore, for any sequence \( x_t \to x \) we have that \( F(y|x_t) \to F(y|x) \) uniformly in \( y \), and \( F(y|x) \) is continuous. Let \( u_t \to u \) and \( x_t \to x \). Since \( F(y|x) = u \) has a unique root \( y = F^{-1}(u|x) \), the root of \( F(y|x_t) = u_t \), i.e., \( y_t = F^{-1}(u_t|x_t) \), converges to \( y \) by a standard argument, see, e.g., van der Vaart and Wellner (1997). \( \square \)

**B.2. Proof of Proposition 2.** We first show the following result (which is a particular case of [Chernozhukov et al. (2006a)], Proposition 2 in the case where the quantile function is increasing):

**Proposition A.** [Hadamard Derivative of \( F(y|x) \) with respect to \( Q(u|x) \)] Define \( F(y|x, h_t) := \int_0^1 \{Q(u|x) + th_t(u|x) \leq y\} du \). Under assumptions (a)-(b), as \( t \to 0 \),

\[
D_{h_t}(y|x, t) := \frac{F(y|x, h_t) - F(y|x)}{t} \to D_h(y|x) := \frac{h(u_1(y|x)|x)}{Q'(u_1(y|x)|x)}.
\]  

The convergence holds uniformly in any compact subset of \( \mathcal{Y}^* := \{(y, x) : y \in \mathcal{Y}^*_x, x \in \mathcal{X}\} \), for every \( |h_t - h|_\infty \to 0 \), where \( h_t \in \ell^\infty(\mathcal{U}\mathcal{X}) \), and \( h \in C(\mathcal{U}\mathcal{X}) \).

**Proof of Proposition A.** Recall that \( K(y|x) = 1 \) as \( Q(u|x) \) is a proper quantile function. We have that for any \( \delta > 0 \), there exists \( \epsilon > 0 \) such that for \( u \in B_\epsilon(u_1(y|x)) \) and for small enough \( t \geq 0 \)

\[
1\{Q(u|x) + th_t(u|x) \leq y\} \leq 1\{Q(u|x) + t(h_1(y|x)|x) - \delta \leq y\},
\]

whereas for all \( u \not\in B_\epsilon(u_1(y|x)) \), as \( t \to 0 \),

\[
1\{Q(u|x) + th_t(u|x) \leq y\} = 1\{Q(u|x) \leq y\}.
\]

Therefore,

\[
\frac{\int_0^1 \{Q(u|x) + th_t(u|x) \leq y\} du - \int_0^1 \{Q(u|x) \leq y\} du}{t} \leq \int_{B_\epsilon(u_1(y|x))} \frac{1\{Q(u|x) + t(h_1(y|x)|x) - \delta \leq y\} - 1\{Q(u|x) \leq y\}}{t} du,
\]

which by the change of variable \( y' = Q(u|x) \) is equal to

\[
\frac{1}{t} \int_{\mathcal{Y}|y, y - t(h_1(y|x)|x) - \delta)} \frac{1}{|Q'(Q^{-1}(y'|x)|x)|} dy',
\]
where $J$ is the image of $B_t(u_1(y|x))$ under $u \mapsto Q(\cdot|x)$. The change of variable is possible because for $\epsilon$ small enough, $Q(\cdot|x)$ is one-to-one between $B_t(u_1(y|x))$ and $J$.

Fixing $\epsilon > 0$, for $t \to 0$, we have that $J \cap [y, y - t(h(u_1(y|x)|x) - \delta)] = [y, y - t(h(u_1(y|x)|x) - \delta)]$, and $|Q'(Q^{-1}(y'|x)|x)| \to |Q'(u_1(y|x)|x)|$ as $Q^{-1}(y'|x) \to u_1(y|x)$. Therefore, the right hand term in (B.2) is no greater than

$$\frac{-h(u_1(y|x)|x) + \delta}{|Q'(u_1(y|x)|x)|} + o(1).$$

Similarly $\frac{-h(u_1(y|x)|x) - \delta}{|Q'(u_1(y|x)|x)|} + o(1)$ bounds (B.2) from below. Since $\delta > 0$ can be made arbitrarily small, the result follows. $\Box$

We are now ready to prove Proposition 2.

**Proof of Proposition 2.**

For a fixed $x$ the result follows by Proposition A, by step 1 of the proof below, and by an application of the Hadamard differentiability of the quantile operator shown by Doss and Gill (1992). Step 2 establishes uniformity over $x \in X$.

**Step 1.** Let $K$ be a compact subset of $\mathcal{X}$. Let $(y_t, x_t)$ be a sequence in $K$, convergent to a point, say $(y, x)$. Then, for every such sequence, $\epsilon_t := t\|h_t\|_\infty + \|Q(\cdot|x_t) - Q(\cdot|x)\|_\infty + |y_t - y| \to 0$, and

$$|F(y_t|x_t, h_t) - F(y|x)| \leq \int_0^1 \left| 1\{Q(u|x_t) + t h_t(u|x) \leq y_t\} - 1\{Q(u|x) \leq y\} \right| du,$$

$$\leq \int_0^1 1\{|Q(u|x) - y| \leq \epsilon_t\} du \to 0,$$

where the last step follows from the absolute continuity of $y \mapsto F(y|x)$, the distribution function of $Q(U|x)$. By setting $h_t = 0$ the above argument also verifies that $F(y|x)$ is continuous in $(y, x)$. Lemma 1 implies uniform convergence of $F(y|x, h_t)$ to $F(y|x)$, which in turn implies by a standard argument the uniform convergence of quantiles $F^{-1}(u|x, h_t) \to F^{-1}(u|x)$, uniformly over $K^*$, where $K^*$ is any compact subset of $\mathcal{X}^*$.

**Step 2.** We have that uniformly over $K^*$,

$$\frac{F(F^{-1}(u|x, h_t)|x, h_t) - F(F^{-1}(u|x, h_t)|x)}{t} = D_h(F^{-1}(u|x, h_t)|x) + o(1),$$

$$= D_h(F^{-1}(u|x)|x) + o(1),$$
using Step 1, Proposition A, and the continuity properties of $D_h(y|x)$. Further, uniformly over $K^*$, by Taylor expansion and Proposition 1, as $t \to 0$,
\[
\frac{F(F^{-1}(u|x,h_t)|x) - F(F^{-1}(u|x)|x)}{t} = f(F^{-1}(u|x)|x) \frac{F^{-1}(u|x,h_t) - F^{-1}(u|x)}{t} + o(1),
\]
and (as will be shown below)
\[
\frac{F(F^{-1}(u|x,h_t)|x,h_t) - F(F^{-1}(u|x)|x)}{t} = o(1),
\]
as $t \to 0$. Observe that the left hand side of (B.6) equals that of (B.5) plus that of (B.4). The result then follows. □

B.3. Proof of Proposition 3. A direct proof of this result is given in Proposition 1 of [Chernozhukov et al. (2006b)]. The weak inequality immediately follows from the inequality due to Lorentz (1953): Let $Q$ and $G$ be two functions mapping $U$ to $K$, a bounded subset of $\mathbb{R}$. Let $Q^*$ and $G^*$ denote their corresponding increasing rearrangements. Then, we have
\[
\int_U L(Q^*(u), G^*(u)) du \leq \int_U L(Q(u), G(u)) du,
\]
for any submodular discrepancy function $L : \mathbb{R}^2 \to \mathbb{R}_+$. In our case, $G(u) = Q_0(u) = G^*(u) = Q_0^*(u)$ almost everywhere. Thus, the true function is its own rearrangement. Moreover, $L(v,w) = |w - v|^p$ is submodular for $p \in [1,\infty)$. For $p = \infty$, the inequality follows by taking limit as $p \to \infty$. The proof of the strict inequality is given in [Chernozhukov et al. (2006b)], Proposition 1. □

B.4. Proof of Proposition 4. This Proposition simply follows by the functional delta method (e.g., van der Vaart, 1998). Instead of restating what this method is, it takes less space to simply recall the proof in the current context.

To show the first part, consider the map $g_n(y,x|h) = \sqrt{n}(F(y|x,h/\sqrt{n}) - F(y|x))$. The sequence of maps satisfies $g_{n'}(y,x|h_{n'}) \to D_h(y|x)$ in $\ell^\infty(K)$ for every subsequence $h_{n'} \to h$ in $\ell^\infty(U\mathcal{X})$, where $h$ is continuous. It follows by the Extended Continuous Mapping Theorem that, in $\ell^\infty(K)$, $g_n(y,x|\sqrt{n}(\hat{Q}(u|x) - Q(u|x))) \Rightarrow D_G(y|x)$ as a stochastic process indexed by $(y,x)$, since $\sqrt{n}(\hat{Q}(u|x) - Q(u|x)) \Rightarrow G(u|x)$ in $\ell^\infty(U\mathcal{X})$.

Conclude similarly for the second part. □
Appendix C. An informal justification of the first variation formula

In this section we provide an informal justification for the main results in the paper. In this section, we will manipulate mathematical objects without any attempt at justification.

We start by some facts about the formal manipulation of the Dirac delta function $\delta$.

**Fact 1.** The Dirac function is the formal derivative of the step function:

$$\frac{d}{dz}1\{z \leq a\} = \delta(z - a).$$

Fact 1 states that the Dirac can be formally defined as the derivative of the Heaviside step function. Thus if the step function $1\{z \leq a\}$ is interpreted as the distribution function of the random variable which is the constant $a$, then the Dirac function is formally interpreted as the density of this random variable with respect to the Lebesgue measure (which does not exist in a classical sense of course as we fail to have uniform continuity).

**Fact 2.** For a smooth function $g$, one has

$$\int_{-\infty}^{+\infty} g(z) \delta(z - a) \, dz = g(a).$$

Fact 2 states that the convolution of the Dirac function centered at $a$ and any smooth function $g$ is the value of $g$ at point $a$. In other words, the function $x \rightarrow \delta(x - a)$ assigns an infinite weight to the point $a$, and zero weight to the other points. Again, this can be interpreted in a probabilistic setting: if $g$ is the density of the probability distribution of a random variable $X$, then using the previous interpretation of $\delta$ as the probability density function of the random variable $0$, the convolution product can be interpreted as the probability density of the variable $X + 0 = X$.

**Fact 3.** For a smooth increasing function $\phi$, one has

$$\delta(\phi(u) - a) = \frac{\delta(u - \phi^{-1}(a))}{|\phi'(\phi^{-1}(a))|}.$$

Fact 3 is obtained by making a change of variables in the integral in Fact 2. Probabilistically, this expression is related to the celebrated Borel’s paradox: for a random couple $(X, U)$ with an absolutely continuous distribution, the conditional distribution of $X$ given $U = u$ is in general not the same as the one of $X$ given $\phi(U) = \phi(u)$. □
Now let us use these facts to justify the result of Proposition 2 in the main text, namely the weak convergence

$$\sqrt{n} \left( \hat{F}_n^{-1} (u|x) - F^{-1} (u|x) \right) \Rightarrow x' G (u).$$

For that, we shall establish the convergence

$$\sqrt{n} \left( \hat{F}_n (y|x) - F (y|x) \right) \Rightarrow - \frac{x' G (u_0)}{x' \beta' (u_0)},$$

where $u_0$ is the solution of equation $y = x' \beta (u)$. Indeed, one has

$$\hat{F}_n (y|x) - F (y|x) = \int 1 \{ x' \beta (\tau) + x' G (\tau) / \sqrt{n} < y \} - 1 \{ x' \beta (\tau) < y \} d\tau,$$

so we can formally get the first variation $f (z + \varepsilon) \simeq f (z) + \varepsilon f' (z)$ to the step function $f (z) = 1 \{ z \leq 0 \}$. By Fact 1, we know $f' (z) = \delta (z)$, so we get $1 \{ z + \varepsilon \leq 0 \} \simeq 1 \{ z \leq 0 \} + \varepsilon \delta (z)$, which in the present context, gives

$$\hat{F}_n (y|x) - F (y|x) \simeq - \int_{-\infty}^{+\infty} \delta (y - x' \beta (\tau)) \frac{x' G (\tau)}{\sqrt{n}} d\tau,$$

and making use of Facts 3 and 2, one gets

$$\hat{F}_n (y|x) - F (y|x) \simeq - \frac{1}{\sqrt{n}} \frac{x' G (u_0)}{x' \beta' (u_0)}.$$

The result then follows by application of a functional delta method to the inversion of the distribution function: indeed, if $\hat{F} (y) - F (y) = \varepsilon (y)$, then at first order, one has

$$\hat{F}^{-1} (u) - F^{-1} (u) \simeq - F' (F^{-1} (u)) \varepsilon (F^{-1} (u)),$$

which, applied to the previous expression, formally leads to the desired first order approximation

$$\sqrt{n} \left( \hat{F}_n^{-1} (u|x) - F^{-1} (u|x) \right) \simeq x' G (u).$$

A rigorous proof of this formal argument (without appealing to the theory of generalized functions) is given in [Chernozhukov et al. (2006a)].
References


