Elliptical Multiple-Output Quantile Regression and Convex Optimization

Marc Hallin
SBS-EM, ECARES, Université libre de Bruxelles

Miroslav Siman
Czech Academy of Sciences

November 2015

ECARES working paper 2015-47

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Elliptical Multiple-Output Quantile Regression and Convex Optimization

Marc Hallin[a,*], Miroslav Šiman[b]

[a] *ECARES, Université libre de Bruxelles CP114/4, B-1050 Brussels, Belgium*
[b] *The Institute of Information Theory and Automation of the Czech Academy of Sciences, Pod Vodárenskou věží 4, CZ-182 08 Prague 8, Czech Republic*

Abstract

This article extends linear quantile regression to an elliptical multiple-output regression setup. The definition of the proposed concept leads to a convex optimization problem. Its elementary properties, and the consistency of its sample counterpart, are investigated. An empirical application is provided.

Keywords: quantile regression, elliptical quantile, multivariate quantile, multiple-output regression

2000 MSC: 62H12, 62J99, 62G05

1. Introduction

Due to their close relation to location and scatter, and their central role in the geometry of Gaussian and elliptical distributions, ellipsoids and the related Mahalanobis distances are quite logical tools for the statistical analysis of multivariate data. Quite naturally, thus, ellipsoids have been considered in the definition of multivariate quantiles and related concepts.

*Corresponding author.
Email address: mhallin@ulb.ac.be (Marc Hallin)

A definition of elliptical multivariate quantiles has been proposed by Hlubinka and Šiman (2013), which leads to a convex optimization problem, hence to a unique solution. That concept essentially deals with location, although its weighted version, based on covariate-driven weights, allows, in the presence of covariates, for a *local constant regression* extension. In the location case (when no covariates are available), Hlubinka and Šiman (2015) consider a more general nonlinear definition, leading to non-convex optimization. The uniqueness of the resulting quantile, therefore, is problematic.

This paper, inspired by Koenker and Bassett (1978), presents a linear multiple-output quantile regression extension of Hlubinka and Šiman (2013) and shows that it leads to a convex optimization problem with a uniquely defined solution for all multivariate continuous distributions with finite second-order moments and connected support, including those with multimodal densities that often arise in the context of mixtures (see, e.g., Došlá (2009)).

Section 2 presents the new concept, Sections 3 and 4 investigate its main properties in the population case and in the sample case, and Section 5 briefly illustrates it with a real data application.

2. Definition

Let $\tau \in (0, 1)$ and consider an m-dimensional response vector $\boldsymbol{Y}$ associated with a $(p+1)$-dimensional vector of regressors $(1, \boldsymbol{Z}')'$. Throughout, it is assumed that the joint distribution of $(\boldsymbol{Y}', \boldsymbol{Z}')'$ is absolutely continuous, with connected support and finite second-order moments.

In the location case (when $p = 0$), Hlubinka and Šiman (2013) define the multivariate (location) elliptical τ-quantile as the ellipsoid

$$\varepsilon_\tau^{\mathrm{loc}} = \varepsilon_\tau^{\mathrm{loc}}(\boldsymbol{Y}) := \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y}'\mathbb{A}_\tau\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{b}_\tau - c_\tau = 0\},$$

where $\mathbb{A}_\tau \in \mathbb{R}^{m\times m}$, $\boldsymbol{b}_\tau \in \mathbb{R}^{m\times 1}$, and $c_\tau > 0$ minimize, subject to $\mathbb{A}$ being symmetric and positive semidefinite with determinant one ($\mathbb{A}$ is thus a *shape matrix* in the sense of Paindaveine (2008)), the objective function

$$\Psi_\tau^{\mathrm{loc}}(\mathbb{A}, \boldsymbol{b}, c) := \mathrm{E}\,\rho_\tau(\boldsymbol{Y}'\mathbb{A}\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{b} - c)$$

with the usual check function $\rho_\tau(x) := x(\tau - \mathrm{I}(x < 0)) = \max\{(\tau-1)x, \tau x\}$. The positive semidefiniteness of $\mathbb{A}$ and the condition on its determinant ensure that $\varepsilon_\tau^{\mathrm{loc}}$ is indeed an ellipsoid, centered at $\boldsymbol{s}_\tau := -\mathbb{A}_\tau^{-1}\boldsymbol{b}_\tau/2$, with equation $(\boldsymbol{y} - \boldsymbol{s}_\tau)'\mathbb{A}_\tau(\boldsymbol{y} - \boldsymbol{s}_\tau) = \kappa_\tau$, where $\kappa_\tau := c_\tau + \boldsymbol{b}_\tau'\mathbb{A}_\tau^{-1}\boldsymbol{b}_\tau/4$. The condition $\det(\mathbb{A}) = 1$ can be viewed as an identification constraint: for any $K > 0$, the triples $(\mathbb{A}, \boldsymbol{b}, c)$ and $(K\mathbb{A}, K\boldsymbol{b}, Kc)$ indeed define the same ellipsoid.

The same definition can be reformulated as a convex optimization problem by relaxing the constraint $\det(\mathbb{A}) = 1$ into $(\det(\mathbb{A}))^{1/m} \geq 1$: the function $\mathbb{A} \mapsto (\det(\mathbb{A}))^{1/m}$, unlike $\mathbb{A} \mapsto \det(\mathbb{A})$, is concave on the cone of symmetric positive semidefinite matrices (see, e.g., Šilhavý (2015)), and the fact that $\Psi_\tau^{\mathrm{loc}}(K\mathbb{A}, K\boldsymbol{b}, Kc) = K\Psi_\tau^{\mathrm{loc}}(\mathbb{A}, \boldsymbol{b}, c)$ for any $K > 0$ implies that the optimal $\mathbb{A}_\tau$ is such that $(\det(\mathbb{A}_\tau))^{1/m} = \det(\mathbb{A}_\tau) = 1$ (see Section 2 of Hlubinka and Šiman (2013), where alternative identification constraints are also discussed).

In the presence of covariates (that is, when $p \geq 1$), the traditional homoscedastic multiple-output linear regression model suggests, for an elliptical multiple-output regression τ-quantile, a simple equation of the form

$$(\boldsymbol{y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{z})'\mathbb{A}_\tau(\boldsymbol{y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{z}) - \gamma = 0$$

with some $\mathbb{A} \in \mathbb{R}^{m\times m}$, $\boldsymbol{\beta} \in \mathbb{R}^{m\times 1}$, $\mathbb{B} \in \mathbb{R}^{m\times p}$, and $\gamma > 0$. The trouble is that the corresponding objective function

$$\mathrm{E}\,\rho_\tau\big((\boldsymbol{Y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{Z})'\mathbb{A}(\boldsymbol{Y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{Z}) - \gamma\big)$$

is not convex in $\boldsymbol{\beta}$ and $\mathbb{B}$, so that its minimization with respect to $\mathbb{A}$, $\boldsymbol{\beta}$, $\mathbb{B}$, and γ is not a *convex* optimization problem. And the same could be said even if γ were an affine linear function of $\boldsymbol{z}$.

In order to restore convexity, consider instead the more general definition

$$\varepsilon_\tau^{\mathrm{reg}} := \{(\boldsymbol{y}', \boldsymbol{z}')' \in \mathbb{R}^{m+p} \colon (\boldsymbol{y} - \boldsymbol{\beta}_\tau - \mathbb{B}_\tau\boldsymbol{z})'\mathbb{A}_\tau(\boldsymbol{y} - \boldsymbol{\beta}_\tau - \mathbb{B}_\tau\boldsymbol{z}) - (\gamma_\tau + \boldsymbol{c}_\tau'\boldsymbol{z} + \boldsymbol{z}'\mathbb{C}\boldsymbol{z}) = 0\} \tag{1}$$

of an elliptical regression quantile $\varepsilon_\tau^{\mathrm{reg}} = \varepsilon_\tau^{\mathrm{reg}}(\boldsymbol{Y}, \boldsymbol{Z})$, where a quadratic form of covariate-driven scale is allowed, and $\mathbb{A}_\tau, \boldsymbol{\beta}_\tau, \mathbb{B}_\tau, \gamma_\tau, \boldsymbol{c}_\tau$, and $\mathbb{C}_\tau$ jointly minimize

$$\Psi_\tau^{\mathrm{reg}} := \mathrm{E}\,\rho_\tau\big((\boldsymbol{y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{z})'\mathbb{A}(\boldsymbol{y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{z}) - (\gamma + \boldsymbol{c}'\boldsymbol{z} + \boldsymbol{z}'\mathbb{C}\boldsymbol{z})\big)$$

under the constraint that $\mathbb{C} \in \mathbb{R}^{p\times p}$ is symmetric and $\mathbb{A} \in \mathbb{R}^{m\times m}$ is symmetric positive semidefinite with $\det(\mathbb{A}) = 1$. This minimization, however, still does not take the form of a convex optimization problem.

Let therefore $\mathbb{M} := (\mathbb{M}^1, \ldots, \mathbb{M}^6)$, with $\mathbb{M}^1 := \mathbb{A} \in \mathbb{R}^{m\times m}$ symmetric positive semidefinite, $\mathbb{M}^2 := \mathbb{B}'\mathbb{A}\mathbb{B} - \mathbb{C} \in \mathbb{R}^{p\times p}$ symmetric, $\mathbb{M}^3 := -2\mathbb{B}'\mathbb{A} \in \mathbb{R}^{p\times m}$, $\mathbb{M}^4 := -2\boldsymbol{\beta}'\mathbb{A} \in \mathbb{R}^{1\times m}$, $\mathbb{M}^5 := 2\boldsymbol{\beta}'\mathbb{A}\mathbb{B} - \boldsymbol{c}' \in \mathbb{R}^{1\times p}$, and $\mathbb{M}^6 := \boldsymbol{\beta}'\mathbb{A}\boldsymbol{\beta} - \gamma \in \mathbb{R}$. The correspondence between $\mathbb{M}$ and $(\mathbb{A}, \boldsymbol{\beta}, \mathbb{B}, \gamma, \boldsymbol{c}, \mathbb{C})$ is one-to-one, with $\mathbb{A} = \mathbb{M}^1$, $\boldsymbol{\beta} = -\frac{1}{2}{\mathbb{M}^1}^{-1}\mathbb{M}^{4\prime}$, $\mathbb{B} = -\frac{1}{2}{\mathbb{M}^1}^{-1}\mathbb{M}^{3\prime}$, $\gamma = \frac{1}{4}\mathbb{M}^4{\mathbb{M}^1}^{-1}\mathbb{M}^{4\prime} - \mathbb{M}^6$, $\boldsymbol{c} = \frac{1}{2}\mathbb{M}^3{\mathbb{M}^1}^{-1}\mathbb{M}^{4\prime} - \mathbb{M}^{5\prime}$, and $\mathbb{C} = \frac{1}{4}\mathbb{M}^3{\mathbb{M}^1}^{-1}\mathbb{M}^{3\prime} - \mathbb{M}^2$: $\mathbb{M}$ thus provides a reparametrization of the problem.

In this new parametrization, the elliptical regression quantile $\varepsilon_\tau^{\rm reg}$ can be expressed as

$$\varepsilon_\tau^{\rm reg} = \{(\boldsymbol{y}', \boldsymbol{z}')' \in \mathbb{R}^{m+p} : r(\boldsymbol{y}, \boldsymbol{z}, \mathbb{M}_\tau) = 0\}$$

where

$$\begin{aligned} r(\boldsymbol{y}, \boldsymbol{z}, \mathbb{M}) &:= \boldsymbol{y}'\mathbb{M}^1\boldsymbol{y} + \boldsymbol{z}'\mathbb{M}^2\boldsymbol{z} + \boldsymbol{z}'\mathbb{M}^3\boldsymbol{y} + \mathbb{M}^4\boldsymbol{y} + \mathbb{M}^5\boldsymbol{z} + \mathbb{M}^6 \\ &= (\boldsymbol{y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{z})'\mathbb{A}(\boldsymbol{y} - \boldsymbol{\beta} - \mathbb{B}\boldsymbol{z}) - (\gamma + \boldsymbol{c}'\boldsymbol{z} + \boldsymbol{z}'\mathbb{C}\boldsymbol{z}), \end{aligned}$$

and $\mathbb{M}_\tau := (\mathbb{M}_\tau^1, \dots, \mathbb{M}_\tau^6)$ jointly minimize

$$\Psi_\tau^{\rm reg} = \Psi_\tau^{\rm reg}(\mathbb{M}) := \Psi_\tau^{\rm reg}(\mathbb{M}^1, \dots, \mathbb{M}^6) = \mathrm{E}\,\rho_\tau\big(r(\boldsymbol{Y}, \boldsymbol{Z}, \mathbb{M})\big),$$

subject to $(\det(\mathbb{M}^1))^{1/m} \geq 1$. As in the location case, positive homogeneity of $\Psi_\tau^{\rm reg}(\mathbb{M}^1, \dots, \mathbb{M}^6)$ implies $\det(\mathbb{M}_\tau) = 1$. The considerable advantage of the parametrization in terms of $\mathbb{M}$ is that it leads to a *convex* optimization problem, hence to a *unique* minimum under the assumptions made.

In principle, one might place further convex constraints on the parameters $\mathbb{M}^1$, …, $\mathbb{M}^6$ in order to simplify the model, such as

$$\begin{aligned} \mathbb{M}^3 = 0 &\iff \mathbb{B} = 0, \\ \mathbb{M}^2 = 0 \text{ and } \mathbb{M}^3 = 0 &\iff \mathbb{B} = 0 \text{ and } \mathbb{C} = 0, \\ \mathbb{M}^3 = 0 \text{ and } \mathbb{M}^5 = 0 &\iff \mathbb{B} = 0 \text{ and } \boldsymbol{c} = 0, \\ \mathbb{M}^2 = 0,\ \mathbb{M}^3 = 0, \text{ and } \mathbb{M}^5 = 0 &\iff \mathbb{B} = 0,\ \mathbb{C} = 0, \text{ and } \boldsymbol{c} = 0; \end{aligned}$$

the resulting optimization problems still are convex, hence also lead to uniquely defined elliptical regression quantiles. In particular, the last set of constraints corresponds to the location elliptical quantiles of Hlubinka and Šiman (2013) which, therefore, are included as a special case. Other natural constraints such as $\mathbb{C} = 0$ and $\boldsymbol{c} = 0$, however, cannot be expressed by means of convex constraints on $\mathbb{M}^1, \ldots, \mathbb{M}^6$. More general yet natural parametric forms of heteroskedasticity, also involving covariate-driven shape matrices, unfortunately, seem impossible within the convex optimization framework.

Finally, it is worth pointing out that $\mathbb{M}^1_\tau, \ldots, \mathbb{M}^6_\tau$, $\mathbb{A}_\tau$, $\boldsymbol{\beta}_\tau$, $\mathbb{B}_\tau$, γ_τ, $\boldsymbol{c}_\tau$, $\mathbb{C}_\tau$, and $\Psi^{\mathrm{reg}}_\tau(\mathbb{M}_\tau)$, but also the Lagrange multipliers associated with possible additional constraints, are potentially useful for statistical inference, especially when considered as τ-indexed processes.

3. Main properties: population case

As in Hlubinka and Šiman (2013), the (Karush-)Kuhn-Tucker necessary and sufficient conditions characterizing the elliptical regression τ-quantile translate to

$$1 = \det(\mathbb{M}^1_\tau), \tag{2}$$

$$0 = \mathrm{P}(r < 0) - \tau, \tag{3}$$

$$0 = \frac{1}{1-\tau}\,\mathrm{E}[\boldsymbol{Y}\,\mathrm{I}_{[r\geq 0]}] - \frac{1}{\tau}\,\mathrm{E}\big[\boldsymbol{Y}\,\mathrm{I}_{[r<0]}\big], \tag{4}$$

$$0 = \frac{1}{1-\tau}\,\mathrm{E}[\boldsymbol{Z}\,\mathrm{I}_{[r\geq 0]}] - \frac{1}{\tau}\,\mathrm{E}\big[\boldsymbol{Z}\,\mathrm{I}_{[r<0]}\big], \tag{5}$$

$$0 = \frac{1}{1-\tau}\,\mathrm{E}[\boldsymbol{Z}\boldsymbol{Y}'\,\mathrm{I}_{[r\geq 0]}] - \frac{1}{\tau}\,\mathrm{E}\big[\boldsymbol{Z}\boldsymbol{Y}'\,\mathrm{I}_{[r<0]}\big], \tag{6}$$

$$0 = \frac{1}{1-\tau}\,\mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}'\,\mathrm{I}_{[r\geq 0]}] - \frac{1}{\tau}\,\mathrm{E}\big[\boldsymbol{Z}\boldsymbol{Z}'\,\mathrm{I}_{[r<0]}\big], \tag{7}$$

and

$$\frac{L_\tau}{m\tau(1-\tau)}\,\det(\mathbb{M}^1_\tau)^{1/m}\mathbb{M}^{1\,-1}_\tau = \frac{1}{1-\tau}\,\mathrm{E}[\boldsymbol{Y}\boldsymbol{Y}'\,\mathrm{I}_{[r\geq 0]}] - \frac{1}{\tau}\,\mathrm{E}\big[\boldsymbol{Y}\boldsymbol{Y}'\,\mathrm{I}_{[r<0]}\big], \tag{8}$$

where $r = r(\boldsymbol{Y}, \boldsymbol{Z}, \mathbb{M}_\tau)$ and L_τ is the Lagrange multiplier associated with the determinant-based constraint $(\det(\mathbb{M}^1))^{1/m} \geq 1$ (recall that $\mathbb{M}^1$ is assumed symmetric positive semidefinite and $\mathbb{M}^2$ symmetric). Proceeding along the same line as in Hlubinka and Šiman (2013), one easily obtains that $L_\tau > 0$ (which is why (2) states $\det(\mathbb{M}_\tau^1) = 1$) and $L_\tau = \Psi_\tau^{\mathrm{reg}}(\mathbb{M}_\tau)$. Therefore, the Lagrange multiplier L_τ does not only measure the impact of the determinant-based constraint, but also equals the minimal value achieved by the objective function.

Conditions (2)–(8) are easy to interpret: (2) only scales the problem; (3) provides $\varepsilon_\tau^{\mathrm{reg}}$ with a clear probability interpretation, namely, that its probability content is τ; (4) and (5) further imply that

$$\mathrm{E}[(\boldsymbol{Y}', \boldsymbol{Z}')'|r \geq 0] = \mathrm{E}[(\boldsymbol{Y}', \boldsymbol{Z}')'|r < 0],$$

so that the probability mass centers of the interior of $\varepsilon_\tau^{\mathrm{reg}}$ and the exterior of $\varepsilon_\tau^{\mathrm{reg}}$ coincide; conditions (6)–(8) yield

$$\begin{pmatrix} L_\tau \frac{1}{m\tau(1-\tau)} {\mathbb{M}_\tau^1}^{-1} & 0 \\ 0 & 0 \end{pmatrix} = \mathrm{var}\big((\boldsymbol{Y}', \boldsymbol{Z}')'|r \geq 0\big) - \mathrm{var}\big((\boldsymbol{Y}', \boldsymbol{Z}')'|r < 0\big),$$

which relates $(\mathbb{M}_\tau^1)^{-1}$ to the difference between the “inner” and “outer” (conditional) variances. Due to an unfortunate typo, the same formula for the location case is repeatedly stated without the $\tau(1-\tau)$ factor in Hlubinka and Šiman (2013), namely in Part [4] of Theorem 2 and in the text preceding it.

It is easy to see that the elliptical regression quantiles $\varepsilon_\tau^{\mathrm{reg}}$ are both regression-equivariant and fully affine-equivariant: if $\boldsymbol{f} \in \mathbb{R}^{m\times 1}$, $\mathbb{F} \in \mathbb{R}^{m\times m}$, $\mathbb{G} \in \mathbb{R}^{m\times p}$, $\mathbb{H} \in \mathbb{R}^{p\times p}$, $d = \det(\mathbb{F})$, and $\varepsilon_\tau^{\mathrm{reg}}(\boldsymbol{Y}, \boldsymbol{Z})$ of (1) leads to quantile coefficients $\mathbb{A}_\tau$, $\boldsymbol{\beta}_\tau$, $\mathbb{B}_\tau$, γ_τ, $\boldsymbol{c}_\tau$, and $\mathbb{C}_\tau$, then $\varepsilon_\tau^{\mathrm{reg}}(\boldsymbol{Y} + \boldsymbol{f} + \mathbb{G}\boldsymbol{Z}, \boldsymbol{Z})$ leads to $\mathbb{A}_\tau$, $\boldsymbol{\beta}_\tau + \boldsymbol{f}$, $\mathbb{B}_\tau + \mathbb{G}$, γ_τ, $\boldsymbol{c}_\tau$, and $\mathbb{C}_\tau$, and $\varepsilon_\tau^{\mathrm{reg}}(\boldsymbol{f} + \mathbb{F}\boldsymbol{Y}, \mathbb{H}\boldsymbol{Z})$ leads to $d^2(\mathbb{F}^{-1})'\mathbb{A}_\tau\mathbb{F}^{-1}$, $\boldsymbol{\beta}_\tau + \boldsymbol{f}$, $\mathbb{B}_\tau\mathbb{H}^{-1}$, $d^2\gamma_\tau$, $d^2(\mathbb{H}^{-1})'\boldsymbol{c}_\tau$, and $d^2(\mathbb{H}^{-1})'\mathbb{C}_\tau\mathbb{H}^{-1}$.

In certain cases, the quantile cuts $\varepsilon_\tau^{\mathrm{reg}}(\boldsymbol{Y}, \boldsymbol{z}_0)$, $\boldsymbol{z}_0 \in \mathbb{R}^p$, coincide with the conditional quantiles or at least preserve the center, axes, and hyperplanes of symmetry of the conditional distribution. The details will be provided and proved in a more general context elsewhere.

4. Main properties: sample case

In the sample case with n observations $(\boldsymbol{Y}_i', \boldsymbol{Z}_i')'$, $i = 1, \ldots, n$, empirical versions $\varepsilon_{\tau;n}^{\mathrm{reg}}$ of the elliptical regression quantiles $\varepsilon_\tau^{\mathrm{reg}}$ can be defined by considering expectations with respect to empirical distributions. It makes sense, however, to consider here a slightly more general weighted setup with a positive weight w_i associated with the ith observation, $i = 1, \ldots, n$. Those weights can be useful for implementing bootstrap or for handling ties. The weighted optimization problem may then be rewritten as

$$\min_{\mathbb{M}^1,\, \ldots,\, \mathbb{M}^6,\, \boldsymbol{r}^+,\, \boldsymbol{r}^-} \Psi_{\tau;n}^{\mathrm{reg}}(\mathbb{M}) := \sum_{i=1}^n \tau w_i r_i^+ + \sum_{i=1}^n (1-\tau) w_i r_i^-$$

subject to the (differentiable) feasibility constraints

$$-\det(\mathbb{M}^1)^{1/m} + 1 \leq 0, \quad -r_i^+ \leq 0 \ \text{ and } \ -r_i^- \leq 0, \quad i = 1, \ldots, n, \tag{9}$$

$$r(\boldsymbol{Y}_i, \boldsymbol{Z}_i, \mathbb{M}) - r_i^+ + r_i^- = 0, \quad i = 1, \ldots, n, \tag{10}$$

$$\mathbb{M}^1 \text{ is a symmetric positive semidefinite matrix,} \tag{11}$$

$$\mathbb{M}^2 \text{ is a symmetric matrix,} \tag{12}$$

where r_i^+ and r_i^- are the positive and negative parts of the residual $r_i = r_i^+ - r_i^- := r(\boldsymbol{Y}_i, \boldsymbol{Z}_i, \mathbb{M})$, $i = 1, \ldots, n$.

As in Hlubinka and Šiman (2013), one can invoke the theory of convex optimization as exposed in Boyd and Vandenberghe (2004), check the refined Slater's constraint qualification, and apply the (Karush-)Kuhn-Tucker conditions. The matrices $\mathbb{M}^1_{\tau;n}, \ldots, \mathbb{M}^6_{\tau;n}$ thus solve the sample elliptical τ-quantile optimization problem if and only if there exist $r_i^+ \geq 0$ and $r_i^- \geq 0$, $i = 1, \ldots, n$, and dual variables $L \geq 0$, $\lambda_i^+ \geq 0$, $\lambda_i^- \geq 0$, and ν_i, $i = 1, \ldots, n$, such that

$$\text{the constraints (9)–(12) are satisfied for } \mathbb{M}^1_{\tau;n}, \ldots, \mathbb{M}^6_{\tau;n}, \tag{13}$$

$$L(-\det(\mathbb{M}^1_{\tau;n})^{1/m} + 1) = 0, \tag{14}$$

$$\lambda_i^+ r_i^+ = 0 \;\text{ and }\; \lambda_i^- r_i^- = 0, \quad i = 1, \ldots, n, \tag{15}$$

$$w_i\tau - \lambda_i^+ - \nu_i = 0 \;\text{ and }\; w_i(1-\tau) - \lambda_i^- + \nu_i = 0, \quad i = 1, \ldots, n, \tag{16}$$

$$\sum_{i=1}^n \nu_i = 0, \quad \sum_{i=1}^n \nu_i \boldsymbol{Z}_i = 0, \quad \text{and} \quad \sum_{i=1}^n \nu_i \boldsymbol{Y}_i = 0, \tag{17}$$

$$\sum_{i=1}^n \nu_i \boldsymbol{Z}_i \boldsymbol{Y}_i' = 0, \quad \sum_{i=1}^n \nu_i \boldsymbol{Z}_i \boldsymbol{Z}_i' = 0, \text{ and} \tag{18}$$

$$\sum_{i=1}^n \nu_i \boldsymbol{Y}_i \boldsymbol{Y}_i' = \frac{L}{m} \det(\mathbb{M}^1_{\tau;n})^{1/m} (\mathbb{M}^1_{\tau;n})^{-1}. \tag{19}$$

This implies $\lambda_i^+ = 0$ and $\nu_i = w_i\tau$ for $r_i^+ > 0$, $\lambda_i^- = 0$ and $\nu_i = w_i(\tau - 1)$ for $r_i^- > 0$, and $w_i(\tau - 1) \leq \nu_i \leq w_i\tau$ for $r_i = 0$. Furthermore,

$$\sum_{i=1}^n w_i \, \mathrm{I}_{[r_i < 0]} \leq n\tau \leq \sum_{i=1}^n w_i \, \mathrm{I}_{[r_i \leq 0]} \,.$$

Up to the small deviations caused by the data points with zero residuals, the necessary and sufficient conditions (13)–(19) roughly can be interpreted as the sample counterparts of the population conditions (2)–(8).

The strong duality theorem for convex optimization implies that, for the optimal solution $\mathbb{M}_{\tau;n} := (\mathbb{M}^1_{\tau;n}, \ldots, \mathbb{M}^6_{\tau;n})$,

$$\begin{aligned}
\Psi^{\mathrm{reg}}_{\tau;n}(\mathbb{M}_{\tau;n}) =& \tau \sum w_i r_i^+ + (1-\tau) \sum w_i r_i^- + \sum \lambda_i(-r_i^+) + \sum \lambda_i(-r_i^-) \\
& + L(-\det(\mathbb{M}^1_{\tau;n})^{1/m} + 1) + \sum \nu_i (r(\boldsymbol{Y}_i, \boldsymbol{Z}_i, \mathbb{M}_{\tau;n}) - r_i^+ + r_i^-) \\
=& \sum r_i^+(w_i\tau - \lambda_i - \nu_i) + \sum r_i^-(w_i(1-\tau) - \lambda_i + \nu_i) + 0 \\
& + \mathbb{M}^6_{\tau;n}(\sum \nu_i) + \mathbb{M}^5_{\tau;n}(\sum \nu_i \boldsymbol{Z}_i) + \mathbb{M}^4_{\tau;n}(\sum \nu_i \boldsymbol{Y}_i) \\
& + \mathrm{tr}(\mathbb{M}^3_{\tau;n} \sum \nu_i \boldsymbol{Y}_i \boldsymbol{Z}_i') + \mathrm{tr}(\mathbb{M}^2_{\tau;n} \sum \nu_i \boldsymbol{Z}_i \boldsymbol{Z}_i') + \mathrm{tr}(\mathbb{M}^1_{\tau;n} \sum \nu_i \boldsymbol{Y}_i \boldsymbol{Y}_i') \\
=& L \det(\mathbb{M}^1_{\tau;n})^{1/m} \, \mathrm{tr}(\mathbb{M}^1_{\tau;n} {\mathbb{M}^1_{\tau;n}}^{-1})/m = L \det(\mathbb{M}^1_{\tau;n})^{1/m},
\end{aligned}$$

where all sums run from $i = 1$ to n. If $\Psi^{\mathrm{reg}}_{\tau;n} = \Psi^{\mathrm{reg}}_{\tau;n}(\mathbb{M}^1_{\tau;n}, \ldots, \mathbb{M}^6_{\tau;n}) > 0$, then $L > 0$, $\det(\mathbb{M}^1_{\tau;n}) = 1$, and $\Psi^{\mathrm{reg}}_{\tau;n} = L$. If $\Psi^{\mathrm{reg}}_{\tau;n} = 0$, then necessarily $L = 0$. In both cases,

$$\Psi^{\mathrm{reg}}_{\tau;n}(\mathbb{M}^1_{\tau;n}, \ldots, \mathbb{M}^6_{\tau;n}) = L,$$

and the optimal value of the objective function again equals that of the Lagrange multiplier associated with the determinant-based constraint.

All statements so far in this section are valid without any assumption at all. There are typically $p(p+1)/2 + pm + p + m + 1$ zero residuals for all but a finite number of τ values if n is sufficiently large, except for some very special data configurations that can be ruled out almost surely under the absolute continuity assumption on the underlying population distribution. Consequently, the number of distinct sample elliptical regression τ-quantiles, $\tau \in (0,1)$, is finite and, for low n and large p, relatively small.

If $w_i := w(\boldsymbol{Y}_i, \boldsymbol{Z}_i)$, where w is a square-integrable density positive on the same domain as the population density of $(\boldsymbol{Y}', \boldsymbol{Z}')'$, then Theorem 5.14 of van der Vaart (1998) guarantees basic convergence, for $n \to \infty$, of the (weighted) sample elliptical regression quantile coefficient vector

$$\boldsymbol{m}_{\tau;n} := (\mathrm{vec}(\mathbb{M}^1_{\tau;n})', \mathrm{vec}(\mathbb{M}^2_{\tau;n})', \mathrm{vec}(\mathbb{M}^3_{\tau;n})', \mathbb{M}^4_{\tau;n}, \mathbb{M}^5_{\tau;n}, \mathbb{M}^6_{\tau;n})'$$

to its (uniquely defined) population counterpart

$$\boldsymbol{m}_{\tau} := (\mathrm{vec}(\mathbb{M}^1_{\tau})', \mathrm{vec}(\mathbb{M}^2_{\tau})', \mathrm{vec}(\mathbb{M}^3_{\tau})', \mathbb{M}^4_{\tau}, \mathbb{M}^5_{\tau}, \mathbb{M}^6_{\tau})',$$

in the sense that

$$\mathrm{P}\big(\{\|\boldsymbol{m}_{\tau;n} - \boldsymbol{m}_{\tau}\| > \varepsilon\} \text{ and } \{\boldsymbol{m}_{\tau;n} \in \mathcal{K}\}\big) \longrightarrow_{n\to\infty} 0$$

for any $\varepsilon > 0$ and any compact set $\mathcal{K}$ of the right dimension. The location version of this result for unit weights in Theorem 3 of Hlubinka and Šiman (2013) is stated incorrectly with the $\notin$ symbol instead of $\in$.

The optimization (semidefinite programming) behind the sample weighted elliptical regression quantiles can be done, e.g., with the CVX toolbox (Grant and Boyd, 2008, 2009) for MATLAB (The MathWorks, Inc., 2013), that can handle relatively large and multi-dimensional datasets.

5. A real-data example

The theory shows that elliptical regression quantiles are particularly suitable for large datasets without outliers. In this section, they are computed for body girth measurements data (Heinz et al., 2003) that are often used for illustrating various statistical methods, despite the fact that they do not constitute a random sample from any well-defined population.

In this example, $n = 260$ observations of calf maximum girth Y_1 (cm) and thigh maximum girth Y_2 (cm) of physically active women are modeled with the aid of a single regressor Z representing either their body mass index (BMI) or their age. Figure 1 displays the sample version of $\varepsilon_\tau^{\mathrm{reg}}((\boldsymbol{Y}'_1, \boldsymbol{Y}'_2)', \boldsymbol{z}_0)$ for $\tau \approx 0.032$, 0.560, and 0.933[1] at some empirical quantiles (of orders 0.1, 0.3, 0.5, 0.7, and 0.9) of the regressor Z. The figure clearly reveals different but meaningful trends and heteroskedasticity patterns for different quantile levels. Interested readers may compare these results with those obtained for the same data by the competing methods of Hallin et al. (2010, 2015).

Acknowledgements

The research of Miroslav Šiman was supported by the Czech Science Foundation project GA14-07234S. Marc Hallin acknowledges the support of the IAP research network grant P7/06 of the Belgian government (Belgian Science Policy), a Crédit aux Chercheurs of the Fonds National de la Recherche Scientifique, and the Discovery grant DP150100210 of the Australian Research Council. Both authors thank Davy Paindaveine and an anonymous referee for insightful comments.

References

Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge University Press, Cambridge. http://www.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf.

Došlá, Š., 2009. Conditions for bimodality and multimodality of a mixture of two unimodal densities. Kybernetika 45, 279–292.

[1] In the population case, those τ-quantiles would match the location halfspace depth contours of a bivariate normal distribution at levels 0.40, 0.10, and 0.01, respectively.

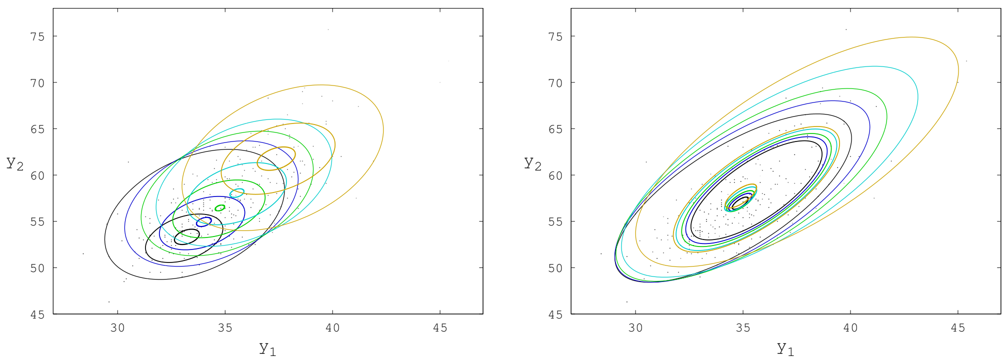

Figure 1: **Application to the body girth measurement data.** The figure illustrates the dependence of calf maximum girth Y_1 (in cm) and thigh maximum girth Y_2 (in cm) on a single regressor Z, which is either the body mass index (left) or age (right), by means of the empirical parametric elliptical regression quantiles $\varepsilon^{\text{reg}}_{\tau;n}((Y_1, Y_2)', z_0)$ obtained from $n = 260$ observations for $\tau \approx 0.032, 0.560,$ and 0.933 and for z_0 equal to the empirical pth quantile of Z, $p = 0.1$ (black), 0.3 (blue), 0.5 (green), 0.7 (cyan), and 0.9 (yellow). The colors are visible only in the online version of the article.

Grant, M., Boyd, S., 2008. Graph implementations for nonsmooth convex programs, in: Blondel, V., Boyd, S., Kimura, H. (Eds.), Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences, Springer, pp. 95–110. http://stanford.edu/∼boyd/graph_dcp.html

Grant, M., Boyd, S., 2009. CVX: Matlab software for disciplined convex programming (web page and software). http://stanford.edu/∼boyd/cvx

Hallin, M., Paindaveine, D., Šiman, M., 2010. Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. Annals of Statistics 38, 635–669.

Hallin, M., Lu, Z., Paindaveine, D., Šiman, M., 2015. Local bilinear multiple-output quantile regression. Bernoulli 21, 1435–1466.

Heinz, G., Peterson, L.J., Johnson, R.W., Kerjk, C.J., 2003. Exploring relationships in body dimensions. Journal of Statistics Education 11. Available at http://www.amstat.org/publications/jse/v11n2/datasets.heinz.html.

Hlubinka, D., Šiman, M., 2013. On elliptical quantiles in the quantile regression setup. Journal of Multivariate Analysis 116, 163–171.

Hlubinka, D., Šiman, M., 2015. On generalized elliptical quantiles in the nonlinear quantile regression setup. TEST 24, 249–264.

Koenker, R., Bassett, G.J., 1978. Regression quantiles. Econometrica 46, 33–50.

The MathWorks, Inc., 2013. MATLAB. Natick, Massachusetts, United States.

Paindaveine, D., 2008. A canonical definition of shape. Statistics & Probability Letters 78, 2240–2247.

Šilhavý, M., 2008. The Convexity of $\mathrm{C} \mapsto \mathrm{h}(\det \mathrm{C})$. Technische Mechanik 35, 60–61.

van der Vaart, A.W., 1998. Asymptotic Statistics, Cambridge University Press, Cambridge.