Testing axial symmetry by means of directional regression quantiles[*]

Šárka Hudecová[†1] and Miroslav Šiman[2]

[1]*Department of Probability and Statistics, Faculty of Mathematics and Physics,
Charles University.
Sokolovská 83, CZ 18675 Prague 8, Czech Republic
e-mail:* hudecova@karlin.mff.cuni.cz

[2]*The Czech Academy of Sciences, Institute of Information Theory and Automation.
Pod Vodárenskou věží 4, CZ 18200 Prague 8, Czech Republic*

Abstract: The article describes how directional quantiles can be useful for testing the null hypothesis that a multivariate distribution is symmetric around a line in a given direction. It also generalizes the proposed tests to residual distributions in a linear regression setup, discusses their use for statistical inference regarding equality of distributions, equality of scale, or exchangeability, and illustrates the achievements with carefully designed pictures and examples.

Contents

1. Introduction

Symmetry not only makes the world beautiful and interesting, but also plays a key role in mathematical statistics where it simplifies and reduces statistical models via sufficiency. Statisticians already recognize several kinds of multivariate symmetry; see, e.g., [22] for a survey. Nevertheless, there seem to be no available results on testing axial symmetry of a given multivariate distribution beyond dimension two except for the very recent permutation test in [12]. A random vector $\boldsymbol{Y}$ with $\mathrm{E}\|\boldsymbol{Y}\| < \infty$ is here formally defined to be axially symmetric around an axis with direction $\boldsymbol{u}$ when $\mathcal{L}\{\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})\} = \mathcal{L}\{\mathbf{R}_{\boldsymbol{u}}(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))\}$ for the orthonormal matrix $\mathbf{R}_{\boldsymbol{u}} = 2\boldsymbol{u}\boldsymbol{u}^{\top} - \mathbf{I}$, satisfying $\mathbf{R}_{\boldsymbol{u}}\boldsymbol{u} = \boldsymbol{u}$ and $\mathbf{R}_{\boldsymbol{u}}\boldsymbol{v} = -\boldsymbol{v}$ for all vectors $\boldsymbol{v}$ orthogonal to $\boldsymbol{u}$.

To be more precise, a slightly similar test for the first eigenvector of a covariance or scatter matrix is often used in the principal component analysis; see, e.g., [6] and the references given there. In the bivariate case, a simple nonparametric test for testing symmetry around a given line has been proposed only in [20] and applied to testing exchangeability and symmetry around a coordinate axis. Some tests of the latter hypotheses have already been discussed in the non-parametric statistical literature; see, e.g., [10] and [18] for bivariate tests and [23] for a multivariate test of conditional symmetry. Related symmetries have also been studied for directional data [4].

Nevertheless, the problem of testing axial symmetry (about an arbitrary line) becomes really interesting only in spaces beyond dimension two where axial symmetry does not coincide with hyperplane symmetry. Furthermore, the assumption of a particular axis of symmetry is often too restrictive. It would often be more convenient to assume only its direction as there can never exist two or more such axes parallel with one another. The tests presented below thus nicely fill in the gap as they work for general distributions in any dimension and assume only the axial direction under the null hypothesis. They can also test conditional axial symmetry in a regression context with a few regressors and responses, which also distinguishes them from possible competitors. That is to say that a follow-up article [11] will describe various tests of symmetry around a general subspace based on a completely different principle, only in the mulivariate case (with no regressors), and with more stringent assumptions.

The presented tests originate from the directional quantile regression, introduced in [7] and further elucidated in [19], that appears very useful even for a single direction, namely for the direction of the assumed axis of symmetry. Then it more or less resembles the ordinary quantile regression of [15] and [13], but with stochastic regressors and applied to certain projections. The regression framework behind the tests makes their regression extensions very intuitive and straightforward.

Unfortunately, it turns out that the presence of response-dependent and stochastic regressors makes the traditional inference about the regression quantile process invalid except for some very special cases. Consequently, the tests presented here had to be derived anew with the response-dependent and stochastic regressors taken into consideration. Therefore, the results probably con-

tribute something original even to the theory of ordinary quantile regression with such type of regressors.

The tests are likely to become useful even in full generality because axial symmetry plays a key role in molecular symmetry (influencing chemical properties of the matter) and because it naturally occurs whenever mirrors or reflections are employed, e.g., in optics, acoustics, particle physics, astronomy, and crystallography. For example, the applications might result from the law of reflection, the axial rotation of heavenly bodies, the natural axial symmetry of simple living organisms, the axial symmetry of electrostatic potential of symmetric molecules, or from the radars rotating around an axis. The need for testing axial symmetry might also arise in the same situations when rotational symmetry is investigated for directional data; see [4] and the references therein. Furthermore, the conditional axial symmetry is also closely linked to the directional predictability of linear models with vector responses.

In any way, the test of (conditional) axial symmetry may be used to check other common statistical hypotheses such as (conditional) symmetry about a particular coordinate axis, (conditional) exchangeability, and equality of (conditional) distributions or their scales, all that possibly up to a suitable shift. The applications are further elaborated and illustrated in the text.

The rest of this article is organized as follows. Section 2 introduces the necessary minimum regarding the directional regression quantiles of interest, Section 3 provides some motivation for the tests and for their use in various situations, Section 4 mainly derives and discusses the tests for the null hypothesis of axial symmetry about a line with a given direction in the location case, Section 5 extends the tests to a linear regression context, and concluding Section 6 illustrates the achievements with a few representative examples and comparisons. The Appendix collects most of the technical remarks and the proofs of all the assertions.

2. Directional quantiles

Suppose that $\boldsymbol{Y} = (Y^{(1)}, \ldots, Y^{(m)})^\top \in \mathbb{R}^m$ and $\boldsymbol{X} = (1, X^{(2)}, \ldots, X^{(p)})^\top = (1, \boldsymbol{Z}^\top)^\top \in \mathbb{R}^p$ stand for a random vector of responses and for a random vector of regressors, respectively, and that the following Assumption 1 holds in the whole article.

Assumption 1. *The joint probability distribution $\mathcal{L}$ of $(\boldsymbol{Y}^\top, \boldsymbol{Z}^\top)^\top$ is absolutely continuous with finite expectation, cumulative distribution function F, and probability density function f that is continuous, bounded, and positive in the interior of a connected support.*

In general, directional multivariate quantiles extend univariate quantiles to multivariate spaces directionwise. In [7], the directional (regression) τ-quantile of the m-dimensional response $\boldsymbol{Y}$ corresponding to the p-dimensional covariate $\boldsymbol{X}$ is defined for any direction $\boldsymbol{u} \in \mathcal{S}^{m-1} := \{\boldsymbol{v} \in \mathbb{R}^m : \|\boldsymbol{v}\| = 1\}$ and for any quantile level $\tau \in (0, 1)$ as the unique hyperplane

$$\pi_{\tau\boldsymbol{u}} = \{(\boldsymbol{y}^\top, \boldsymbol{z}^\top) \in \mathbb{R}^m \times \mathbb{R}^{p-1} : \boldsymbol{b}_{\tau\boldsymbol{u}}^\top \boldsymbol{y} - \boldsymbol{a}_{\tau\boldsymbol{u}}^\top (1, \boldsymbol{z}^\top)^\top = 0\}$$

where $(\boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top, \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top)^\top \in \mathbb{R}^{m+p}$ solves the minimization problem

$$\min_{(\boldsymbol{b}^\top, \boldsymbol{a}^\top)^\top \in \mathbb{R}^{m+p},\ \boldsymbol{b}^\top \boldsymbol{u}=1} \mathrm{E}\{\rho_\tau(\boldsymbol{b}^\top \boldsymbol{Y} - \boldsymbol{a}^\top \boldsymbol{X})\} \tag{2.1}$$

where $\rho_\tau(x) = x(\tau - \mathrm{I}(x < 0)) = \max\{(\tau-1)x, \tau x\}$ is the quantile check function. Equivalently,

$$\pi_{\tau\boldsymbol{u}} = \{(\boldsymbol{y}^\top, \boldsymbol{z}^\top) \in \mathbb{R}^m \times \mathbb{R}^{p-1} : \boldsymbol{u}^\top \boldsymbol{y} - \boldsymbol{c}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{\Gamma}_{\boldsymbol{u}}^\top \boldsymbol{y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top (1, \boldsymbol{z}^\top)^\top = 0\}$$

where $\boldsymbol{\Gamma_u}$ is an $m \times (m-1)$ matrix complementing $\boldsymbol{u}$ to an orthonormal matrix and $(\boldsymbol{c}_{\tau\boldsymbol{u}}{}^\top, \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top)^\top \in \mathbb{R}^{m+p-1}$ solves the unconstrained (standard quantile regression) minimization problem

$$\min_{(\boldsymbol{c}^\top, \boldsymbol{a}^\top)^\top \in \mathbb{R}^{m+p-1}} \mathrm{E}\{\rho_\tau(\boldsymbol{u}^\top \boldsymbol{Y} - \boldsymbol{c}^\top \boldsymbol{\Gamma}_{\boldsymbol{u}}^\top \boldsymbol{Y} - \boldsymbol{a}^\top \boldsymbol{X})\} \tag{2.2}$$

with one scalar response $\boldsymbol{u}^\top \boldsymbol{Y}$ and $p + m - 1$ scalar regressors grouped in $(\boldsymbol{Y}^\top \boldsymbol{\Gamma_u}, \boldsymbol{X}^\top)^\top$. (It is basically a standard quantile regression problem when the basis of the response space is changed from the canonical one to $(\boldsymbol{u}|\boldsymbol{\Gamma_u})$.) Then $\boldsymbol{b}_{\tau\boldsymbol{u}} = \boldsymbol{u} - \boldsymbol{\Gamma_u}\boldsymbol{c}_{\tau\boldsymbol{u}}$, $\boldsymbol{c}_{\tau\boldsymbol{u}} = -\boldsymbol{\Gamma}_{\boldsymbol{u}}^\top \boldsymbol{b}_{\tau\boldsymbol{u}}$, and the choice of $\boldsymbol{\Gamma_u}$ does not impact $\pi_{\tau\boldsymbol{u}}$.

The (regression) $(\tau\boldsymbol{u})$-quantile hyperplane $\pi_{\tau\boldsymbol{u}}$ divides the space into the upper and lower $(\tau\boldsymbol{u})$-quantile (regression) halfspaces $H_{\tau\boldsymbol{u}}^+$ and $H_{\tau\boldsymbol{u}}^-$:

$$H_{\tau\boldsymbol{u}}^+ = \{(\boldsymbol{y}^\top, \boldsymbol{z}^\top)^\top \in \mathbb{R}^m \times \mathbb{R}^{p-1} : \boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top (1, \boldsymbol{z}^\top)^\top \geq 0\} \text{ and}$$

$$H_{\tau\boldsymbol{u}}^- = \{(\boldsymbol{y}^\top, \boldsymbol{z}^\top)^\top \in \mathbb{R}^m \times \mathbb{R}^{p-1} : \boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top (1, \boldsymbol{z}^\top)^\top < 0\}.$$

In the location (multivariate) case with $p = 1$, both $\boldsymbol{Z}$ and $\boldsymbol{z}$ simply disappear from all the definitions.

Think of the population necessary and sufficient conditions for $(\boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top, \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top)^\top$ to solve (2.1):

$$\boldsymbol{0} = \frac{1}{1-\tau}\mathrm{E}\{\mathrm{I}(\boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{Y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{X} \geq 0)\boldsymbol{X}\}$$
$$- \frac{1}{\tau}\mathrm{E}\{\mathrm{I}(\boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{Y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{X} < 0)\boldsymbol{X}\}, \tag{2.3}$$

$$\boldsymbol{D}_{\tau\boldsymbol{u}} = \frac{1}{1-\tau}\mathrm{E}\{\mathrm{I}(\boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{Y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{X} \geq 0)\boldsymbol{Y}\}$$
$$- \frac{1}{\tau}\mathrm{E}\{\mathrm{I}(\boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{Y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{X} < 0)\boldsymbol{Y}\}, \tag{2.4}$$

$$1 = \boldsymbol{b}^\top \boldsymbol{u}, \tag{2.5}$$

where $\boldsymbol{D}_{\tau\boldsymbol{u}} = \lambda_{\tau\boldsymbol{u}}\boldsymbol{u}/\{\tau(1-\tau)\}$ and $\lambda_{\tau\boldsymbol{u}}$ is the Lagrange multiplier associated with (2.5).

The probability interpretation of the directional regression quantiles follows from (2.3) thanks to the first unit coordinate of $\boldsymbol{X}$:

$$\tau = \mathrm{E}\{\mathrm{I}(\boldsymbol{b}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{Y} - \boldsymbol{a}_{\tau\boldsymbol{u}}{}^\top \boldsymbol{X} < 0)\}. \tag{2.6}$$

Consequently, $(\boldsymbol{D}_{\tau\boldsymbol{u}}^\top, \boldsymbol{0}^\top)^\top$ can be interpreted as the vector linking the mass centers $\boldsymbol{\mu}(H_{\tau\boldsymbol{u}}^-)$ and $\boldsymbol{\mu}(H_{\tau\boldsymbol{u}}^+)$ of $H_{\tau\boldsymbol{u}}^-$ and $H_{\tau\boldsymbol{u}}^+$ through the overall center of mass and pointing to $H_{\tau\boldsymbol{u}}^+$ if f is interpreted as the mass density.

Loosely speaking, the (regression) $(\tau\boldsymbol{u})$-quantile hyperplane $\pi_{\tau\boldsymbol{u}}$ cuts off the probability mass equal to τ and splits the space into two halfspaces with their own probability mass centers whose mutual position is determined by the direction $\boldsymbol{u}$ up to certain scalar multiplier.

3. Motivation

For simplicity, first assume the purely multivariate case when $\mathcal{L} = \mathcal{L}(\boldsymbol{Y})$ is axially symmetric around an axis with direction $\boldsymbol{u}$. Then the uniquely defined $\pi_{\tau\boldsymbol{u}}$, cutting off the right amount of probability mass, must be orthogonal to $\boldsymbol{u}$ for any $\tau \in (0,1)$ because $\boldsymbol{\mu}(H_{\tau\boldsymbol{u}}^+) - \boldsymbol{\mu}(H_{\tau\boldsymbol{u}}^-)$ is then parallel to $\boldsymbol{u}$ and the necessary and sufficient conditions (2.3) to (2.5) are then satisfied for $\boldsymbol{b}_{\tau\boldsymbol{u}} = \boldsymbol{u}$, i.e., for $\boldsymbol{c}_{\tau\boldsymbol{u}} = \boldsymbol{0}$. In other words, the axial symmetry around an axis with direction $\boldsymbol{u}$ implies $\boldsymbol{c}_{\tau\boldsymbol{u}} = \boldsymbol{0}$ for any $\tau \in (0,1)$, which makes its testing by means of $\boldsymbol{c}_{\tau\boldsymbol{u}}$ and its scalar functions very promising, especially for elliptical distributions:

Proposition 1. *Consider random vector $\boldsymbol{Y}$ with distribution $\mathcal{L}(\boldsymbol{Y})$ satisfying Assumption 1.*

(1) *If $\mathcal{L}(\boldsymbol{Y})$ is symmetric around an axis with direction $\boldsymbol{u}$, then $\boldsymbol{c}_{\tau\boldsymbol{u}} = \boldsymbol{0}$ for any $\tau \in (0,1)$.*
(2) *If $\mathcal{L}(\boldsymbol{Y})$ is elliptically symmetric and $\boldsymbol{c}_{\tau\boldsymbol{u}} = \boldsymbol{0}$ for any $\tau \in (0,1)$, then $\mathcal{L}(\boldsymbol{Y})$ is symmetric around an axis with direction $\boldsymbol{u}$.*
(3) *If $\mathcal{L}(\boldsymbol{Y})$ is elliptically symmetric but not symmetric around an axis with direction $\boldsymbol{u}$, then $\boldsymbol{c}_{\tau\boldsymbol{u}} \neq \boldsymbol{0}$ is a constant vector independent of $\tau \in (0,1)$.*

The impact of such axial symmetry tests far exceeds their primary purpose because they can also be employed

1. for testing the hypothesis of axial symmetry about a particular line in direction $\boldsymbol{u}$. That is to say that if the population distribution of $\boldsymbol{Y}$ is symmetric about an axis in direction $\boldsymbol{u}$, then the particular axis of symmetry is of the form $\mathrm{E}(\boldsymbol{Y}) + t\boldsymbol{u}$, $t \in \mathbb{R}$. Testing the particularity should thus involve a test of the mean vector after the test of the axial direction. After a suitable affine transformation (turning the particular line of interest to the last coordinate axis), one would only have to test that the first $(m-1)$ coordinates of the mean vector are zero.
2. for testing exchangeability after a suitable shift. If a multivariate distribution is exchangeable after a suitable shift, then it is symmetric around an axis with direction $\boldsymbol{u} = (1, 1 \ldots, 1)^\top / \sqrt{m}$. The converse is true only in $\mathbb{R}^2$.
3. for testing equality of independent univariate distributions up to their location. Their joint distribution would then be exchangeable after a suitable shift. Conversely, if the joint distribution of independent univariate distributions is exchangeable, then the univariate distributions are the same.
4. for testing that independent univariate distributions are equally scaled on condition that they can differ only in their location and scale parameters.
5. for testing equality of independent multivariate distributions (after a suitable shift) because then their univariate marginal distributions corre-

sponding to any particular dimension must be independent and one could test it with a compound test.

Of course, one can combine 1. with 2., 3., 4. and 5. if a particular knowledge about the location is available.

In the regression case, the null hypothesis $H_0^S(\boldsymbol{u})$ of axial symmetry around an axis with direction $\boldsymbol{u}$ pertains to the conditional distribution:

$$H_0^S(\boldsymbol{u}): \ \mathcal{L}\{\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}|\boldsymbol{X})|\boldsymbol{X}\} = \mathcal{L}\{\mathbf{R_u}(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}|\boldsymbol{X}))|\boldsymbol{X}\}, \ \mathbf{R_u} = 2\mathbf{uu}^\top - \mathbf{I},$$

almost surely. In the location case, it reduces to

$$H_0^S(\boldsymbol{u}): \ \mathcal{L}\{\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y})\} = \mathcal{L}\{\mathbf{R_u}(\boldsymbol{Y} - \mathrm{E}(\boldsymbol{Y}))\}, \ \mathbf{R_u} = 2\mathbf{uu}^\top - \mathbf{I}.$$

However, the following text focuses only on the null hypothesis

$$H_0(\boldsymbol{u}): \ \mathbf{c}_{\tau\boldsymbol{u}} = \mathbf{0} \text{ for all } \tau \in (0,1) \tag{3.1}$$

because $H_0^S(\boldsymbol{u})$ always implies $H_0(\boldsymbol{u})$ in the location case and because the two hypotheses are equivalent for elliptical distributions; see Proposition 1.

Furthermore, all that easily extends to certain linear regression location-scale models because Proposition 1 can easily be rephrased for conditional distributions $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x})$ and because the conditional gradient conditions turn into the unconditional ones if all the τ-quantiles of $\boldsymbol{u}'\boldsymbol{Y}$ given $\boldsymbol{X}$ are linear in $\boldsymbol{X}$, thanks to $\mathrm{E}(\cdot) = \mathrm{E}_{\boldsymbol{X}}\mathrm{E}_{\boldsymbol{Y}|\boldsymbol{X}}(\cdot|\boldsymbol{X})$ used in (2.2).

In particular, the theory can be applied even to the following linear regression common-scale model with parametric matrix $\mathbf{B}$, parametric vector $\boldsymbol{d} \neq \mathbf{0}$, and a centered absolutely continuous error term $\boldsymbol{\varepsilon} \in \mathbb{R}^m$, $\mathrm{E}\|\boldsymbol{\varepsilon}\| < \infty$, independent of absolutely continuous regressor vector $\boldsymbol{X}$, $\mathrm{E}\|\boldsymbol{X}\| < \infty$:

$$\boldsymbol{Y} = \mathbf{B}\boldsymbol{X} + (\boldsymbol{d}^\top\boldsymbol{X})\boldsymbol{\varepsilon}. \tag{3.2}$$

Then $H_0^S(\boldsymbol{u})$ implies $H_0(\boldsymbol{u})$ for any $\boldsymbol{u}$, $\|\boldsymbol{u}\| = 1$, and the reverse is true for elliptically distributed $\boldsymbol{\varepsilon}$. If $\boldsymbol{\varepsilon}$ is elliptically distributed, but not around an axis with direction $\boldsymbol{u}$, then $\boldsymbol{c}_{\tau\boldsymbol{u}} \neq \mathbf{0}$ is a constant vector independent of $\tau \in (0,1)$.

In fact, $H_0^S(\boldsymbol{u})$ implies $H_0(\boldsymbol{u})$ for a given directional vector $\boldsymbol{u}$ even in a more general model

$$\boldsymbol{Y} = \mathbf{B}\boldsymbol{X} + \boldsymbol{\Gamma_u}g_0(\boldsymbol{d}_0, \boldsymbol{X}) + \mathbf{Q}\mathbf{D}(\boldsymbol{X})\mathbf{Q}^\top\boldsymbol{\varepsilon} \tag{3.3}$$

where $\mathbf{Q} = (\boldsymbol{u}, \boldsymbol{\Gamma_u})$, $\mathbf{D}(\boldsymbol{X}) = \mathrm{diag}(g_1(\boldsymbol{d}_1, \boldsymbol{X}), \ldots, g_m(\boldsymbol{d}_m, \boldsymbol{X}))$ is an almost surely regular diagonal matrix, $g_1(\boldsymbol{d}_1, \boldsymbol{X}) = \boldsymbol{d}_1^\top\boldsymbol{X}$, $g_0, g_2, \ldots, g_m$ are almost surely non-zero scalar functions, and $\boldsymbol{d_0}, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_m$ are parametric vectors such that $\boldsymbol{d}_1 \neq \mathbf{0}$.

To sum up, $H_0^S(\boldsymbol{u})$ can be tested by means of $H_0(\boldsymbol{u})$ not only in the location case, but also in certain linear regression location-scale models such as (3.2) and (3.3) described above. Consequently, the rest of the article deals only with the problem of testing $H_0(\boldsymbol{u})$, which may be of independent interest. The presented theory for the regression case also covers models (3.2) and (3.3) thanks to Remark 4 after Proposition 3, stated in the Appendix.

Of course, the proposed asymptotic nonparametric tests are the most useful when they have no simple parametric competitors, e.g., when the specification of the underlying probability distribution is unknown or unrelated to the axial symmetry considered. In fact, they may sometimes be the only reasonable options available in the multidimensional and regression cases.

In what follows, sample variants will usually be denoted with $\widehat{}$ and the tests will be called after their test statistics.

4. Location case

The asymptotic representation and distribution of $\widehat{c}_{\tau u}$ in the i.i.d. case is known from [7] for any fixed $\tau \in (0,1)$. If the assumptions were different, then $H_0(\boldsymbol{u})$ of (3.1) could be tested, for example, by means of the rank score statistic of [16] known from the ordinary quantile regression; see also Section 3.7.3 of [13]. The same test statistic is also adopted here but its asymptotic distribution becomes complicated by the presence of stochastic and response-dependent regressors, and it does not follow directly from any known theory.

Consider $\tau \in (0,1)$, i.i.d. vectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$, $\mathbb{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^\top \in \mathbb{R}^{n \times m}$, $\overline{\mathbf{Y}}_n = n^{-1} \mathbb{Y}^\top \mathbf{1}_n$, the rank score vector $\widehat{\mathbf{a}}(\tau) \in \mathbb{R}^n$ defined as

$$\widehat{\mathbf{a}}(\tau) := \operatorname{argmax}\{\mathbf{u}^\top \mathbb{Y}^\top \mathbf{a} : \ n^{-1} \mathbf{1}_n^\top \mathbf{a} = (1 - \tau), \ \mathbf{a} \in [0,1]^n\},$$

and its centered modification $\widehat{\mathbf{b}}(\tau) \equiv \left(\widehat{b}_1(\tau), \ldots, \widehat{b}_n(\tau)\right)^\top := \widehat{\mathbf{a}}(\tau) - (1-\tau)\mathbf{1}_n$. The τ-indexed process

$$\widehat{\mathbf{S}}_n(\tau) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top (\mathbf{Y}_i - \overline{\mathbf{Y}}_n)\widehat{b}_i(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top \mathbf{Y}_i \widehat{b}_i(\tau) = \frac{1}{\sqrt{n}} \boldsymbol{\Gamma}_{\mathbf{u}}^\top \mathbb{Y}^\top \widehat{\mathbf{b}}(\tau) \tag{4.1}$$

is the cornerstone on which the tests will be built.

Proposition 2. *Assume $H_0(\boldsymbol{u})$ with Assumption 1, and $\mathrm{E}(\|\mathbf{Y}\|^{2+\delta}) < \infty$ for some $\delta > 0$.*

(1) Then, for any $\tau \in (0,1)$, $\widehat{\mathbf{S}}_n(\tau) = \widehat{\mathbf{S}}_n^0(\tau) + o_P(1)$, where

$$\widehat{\mathbf{S}}_n^0(\tau) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top \mathbf{Y}_i [\tau - \mathrm{I}\{\mathbf{u}^\top \mathbf{Y}_i < \widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\}] \tag{4.2}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top [\mathbf{Y}_i - \mathrm{E}(\mathbf{Y})][\tau - \mathrm{I}\{\mathbf{u}^\top \mathbf{Y}_i < F_{\boldsymbol{u}}^{-1}(\tau)\}] + o_P(1), \tag{4.3}$$

$F_{\boldsymbol{u}}^{-1}(\tau)$ is the τ-quantile of $\mathbf{u}^\top \mathbf{Y}$ and $\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)$ is its sample counterpart.

(2) If $0 < \tau_1 < \tau_2 < \cdots < \tau_k < 1$, $k \in \mathbb{N}$, then the vector $\widehat{\mathbf{S}}_n(\tau_1, \ldots, \tau_k) = \left(\widehat{\mathbf{S}}_n(\tau_1)^\top, \ldots, \widehat{\mathbf{S}}_n(\tau_k)^\top\right)^\top$ is asymptotically multivariate normal with mean zero and block covariance matrix $\boldsymbol{\Sigma}(\tau_1, \ldots, \tau_k) = \left(\boldsymbol{\Sigma}_{ij}(\tau_i, \tau_j)\right)_{i,j=1}^k$ where

$$\boldsymbol{\Sigma}_{ij}(\tau_i, \tau_j) = \tau_i \tau_j \boldsymbol{\Gamma}_{\mathbf{u}}^\top \mathrm{Var}\,(\mathbf{Y})\boldsymbol{\Gamma}_{\mathbf{u}} - \tau_i \mathbf{C}_{\tau_j} - \tau_j \mathbf{C}_{\tau_i} + \mathbf{C}_{\min\{\tau_i, \tau_j\}}$$

and $\mathbf{C}_t = \mathrm{E}[\mathbf{\Gamma_u^\top}\{\mathbf{Y} - \mathrm{E}(\mathbf{Y})\}\{\mathbf{Y} - \mathrm{E}(\mathbf{Y})\}^\top \mathbf{\Gamma_u}\mathrm{I}\{\mathbf{u}^\top\mathbf{Y} < F_{\boldsymbol{u}}^{-1}(t)\}]$. *Consequently, if* $\widehat{\mathbf{\Sigma}}(\tau_1,\dots,\tau_k)$ *is a consistent estimator of* $\mathbf{\Sigma}(\tau_1,\dots,\tau_k)$, *then*

$$T_D := \widehat{\mathbf{S}}_n^\top(\tau_1,\dots,\tau_k)\{\widehat{\mathbf{\Sigma}}(\tau_1,\dots,\tau_k)\}^{-1}\widehat{\mathbf{S}}_n(\tau_1,\dots,\tau_k) \xrightarrow{D} \chi^2_{k(m-1)}. \quad (4.4)$$

(3) *If moreover* $\mathrm{E}[\mathbf{\Gamma_u^\top}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Gamma_u}\mathrm{I}\{\mathbf{u}^\top\mathbf{Y} < F_{\boldsymbol{u}}^{-1}(\tau)\}] = \tau\mathrm{E}(\mathbf{\Gamma_u^\top}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Gamma_u})$ *for any* $\tau \in (0,1)$, *then* $\widehat{\mathbf{S}}_n(\tau) \xrightarrow{D} \{\mathbf{\Gamma_u^\top}\mathrm{Var}\,(\mathbf{Y})\mathbf{\Gamma_u}\}^{1/2}\mathbf{B}_{m-1}(\tau)$ *and*

$$T_C := \sup_{\tau\in[\varepsilon,1-\varepsilon]} \widehat{T}_n(\tau) \xrightarrow{D} \sup_{\tau\in[\varepsilon,1-\varepsilon]} Q^2_{m-1}(\tau) \qquad (4.5)$$

for any $\varepsilon \in (0,0.5)$, *where* $\mathbf{B}_{m-1}(\tau)$ *is the* $(m-1)$-*dimensional Brownian bridge on* $[0,1]$,

$$\widehat{T}_n(\tau) := \frac{1}{\tau(1-\tau)}\widehat{\mathbf{S}}_n^\top(\tau)(\mathbf{\Gamma_u^\top}\mathbf{S_Y}\mathbf{\Gamma_u})^{-1}\widehat{\mathbf{S}}_n(\tau), \ \tau\in(0,1), \qquad (4.6)$$

$\mathbf{S_Y}$ *is the sample covariance matrix of* $\mathbf{Y}$ *computed from* $\mathbf{Y}_1,\dots,\mathbf{Y}_n$, *and* $Q_{m-1}(\tau) = \|\mathbf{B}_{m-1}(\tau)\|\{\tau(1-\tau)\}^{-1/2}$ *is the Bessel process of order* $m-1$.

Simplification of Proposition 2 in certain cases, weakening its assumptions, and good invariance properties of $\widehat{T}_n(\tau)$ of (4.6) are discussed in Remarks 1 to 3 in the Appendix. In particular, $\widehat{T}_n(\tau)$, $\tau \in (0,1)$, is invariant with respect to the choice of $\mathbf{\Gamma_u}$, to shifts, to rotations (if $(\boldsymbol{u},\mathbf{\Gamma_u})$ is rotated accordingly) and to scale transformations preserving axial symmetry. The same invariance then holds even for T_C of (4.5).

As $H_0(\boldsymbol{u})$ implies a weaker null hypothesis $\widetilde{H_0}(\boldsymbol{u}) : \mathbf{c}_{\tau_i\boldsymbol{u}} = \mathbf{0}$ for $0 < \tau_1 < \cdots < \tau_k < 1$, testing $H_0(\boldsymbol{u})$ can also be based on Proposition 2(2) and may be beneficial at least in the family of elliptical distributions in view of Proposition 1. This is also confirmed empirically in Section 6.

To sum up, $H_0(\boldsymbol{u})$ can be tested in the location case by means of the test statistics T_D and T_C of (4.4) and (4.5) and their asymptotic distributions under $H_0(\boldsymbol{u})$. Of course, test statistics asymptotically equivalent to T_C or T_D under $H_0(\boldsymbol{u})$ can be used as well, and the corresponding assumptions must be fulfilled.

5. Regression case

Assume i.i.d. sample $(\boldsymbol{Y}_i^\top, \boldsymbol{X}_i^\top)^\top$, $i = 1,\dots,n$, and write $(\widehat{\mathbf{c}}_{\tau\boldsymbol{u}}^\top, \widehat{\mathbf{a}}_{\tau\boldsymbol{u}}^\top)^\top$ for the sample analogue to $(\mathbf{c}_{\tau\boldsymbol{u}}^\top, \mathbf{a}_{\tau\boldsymbol{u}}^\top)^\top$, $\mathbb{Y}$ for $(\boldsymbol{Y}_1,\dots\boldsymbol{Y}_n)^\top$ and $\mathbb{X}$ for $(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n)^\top$.

Consider $H_0(\boldsymbol{u})$ of (3.1), define centered regression rank score vector $\widehat{\mathbf{b}}(\tau) = \left(\widehat{\mathbf{b}}_1(\tau),\dots,\widehat{\mathbf{b}}_n(\tau)\right)^\top$ as

$$\widehat{\mathbf{b}}(\tau) := \mathrm{argmax}\{\mathbf{u}^\top\mathbb{Y}^\top\mathbf{a} : \mathbb{X}^\top\mathbf{a} = (1-\tau)\mathbb{X}^\top\mathbf{1}_n, \ \mathbf{a} \in [0,1]^n\} - (1-\tau)\mathbf{1}_n, \ (5.1)$$

and focus again on the τ-indexed process

$$\widehat{\mathbf{S}}_n(\tau) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{\Gamma_u^\top}\mathbf{Y}_i\widehat{\mathbf{b}}_i(\tau) = \frac{1}{\sqrt{n}}\mathbf{\Gamma_u^\top}\mathbb{Y}^\top\widehat{\mathbf{b}}(\tau) = \frac{1}{\sqrt{n}}\mathbf{\Gamma_u^\top}\mathbb{Y}^\top(\mathbf{I}_n - \mathbf{M}_\mathbb{X})\widehat{\mathbf{b}}(\tau),$$

where $\mathbf{M}_{\mathbb{X}}$ stands for the projection matrix $\mathbf{M}_{\mathbb{X}} = \mathbb{X}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top$.

Proposition 3. *Assume the null hypothesis $H_0(\boldsymbol{u})$ with Assumption 1 and both $\mathrm{E}(\|\mathbf{Z}\|^{2+\delta}) < \infty$ and $\mathrm{E}(\|\mathbf{Y}\|^{2+\delta}) < \infty$ for some $\delta > 0$.*

(1) *Then, for any $\tau \in (0,1)$, $\widehat{\mathbf{S}}_n(\tau) = \widehat{\mathbf{S}}_n^0(\tau) + o_P(1)$ for*

$$\widehat{\mathbf{S}}_n^0(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\{\tau - \mathrm{I}(\mathbf{u}^\top\mathbf{Y}_i < \mathbf{a}_{\tau\boldsymbol{u}}^\top\mathbf{X}_i)\}\{\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}_i - \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{H}(\tau)\mathbf{G}(\tau)^{-1}\mathbf{X}_i\},$$

$\mathbf{G}(\tau) = \mathrm{E}\{f_{\mathbf{u}^\top\mathbf{Y}|\mathbf{X}}(\mathbf{a}_{\tau\boldsymbol{u}}^\top\mathbf{X})\mathbf{X}\mathbf{X}^\top\}$, $\mathbf{H}(\tau) = \mathrm{E}\{f_{\mathbf{u}^\top\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}}(\mathbf{a}_{\tau\boldsymbol{u}}^\top\mathbf{X})\mathbf{Y}\mathbf{X}^\top\}$, *$f_{\mathbf{u}^\top\mathbf{Y}|\mathbf{X}}$ being the density of $\mathbf{u}^\top\mathbf{Y}$ given $\mathbf{X}$, and $f_{\mathbf{u}^\top\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}}$ being the density of $\mathbf{u}^\top\mathbf{Y}$ given $(\mathbf{X}^\top, \mathbf{Y}^\top\boldsymbol{\Gamma}_{\mathbf{u}})^\top$.*

(2) *If $0 < \tau_1 < \tau_2 < \cdots < \tau_k < 1$, $k \in \mathbb{N}$, then vector $\widehat{\mathbf{S}}_n(\tau_1,\ldots,\tau_k) = \left(\widehat{\mathbf{S}}_n(\tau_1)^\top, \ldots, \widehat{\mathbf{S}}_n(\tau_k)^\top\right)^\top$ is asymptotically multivariate normal with zero mean and block covariance matrix $\boldsymbol{\Sigma}(\tau_1,\ldots,\tau_k) = \left(\boldsymbol{\Sigma}_{ij}(\tau_i,\tau_j)\right)_{i,j=1}^k$, where*

$$\begin{aligned}\boldsymbol{\Sigma}_{ij}(\tau_i,\tau_j) =& \mathrm{E}\{\tau_i - \mathrm{I}(\mathbf{u}^\top\mathbf{Y} < \mathbf{a}_{\tau_i\boldsymbol{u}}^\top\mathbf{X})\}\{\tau_j - \mathrm{I}(\mathbf{u}^\top\mathbf{Y} < \mathbf{a}_{\tau_j\boldsymbol{u}}^\top\mathbf{X})\}\\&\times\{\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y} - \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{H}(\tau_i)\mathbf{G}(\tau_i)^{-1}\mathbf{X}\}\{\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y} - \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{H}(\tau_j)\mathbf{G}(\tau_j)^{-1}\mathbf{X}\}^\top.\end{aligned}$$

Consequently, if $\widehat{\boldsymbol{\Sigma}}(\tau_1,\ldots,\tau_k)$ is a consistent estimator of $\boldsymbol{\Sigma}(\tau_1,\ldots,\tau_k)$, then

$$T_D := \widehat{\mathbf{S}}_n^\top(\tau_1,\ldots,\tau_k)\{\widehat{\boldsymbol{\Sigma}}(\tau_1,\ldots,\tau_k)\}^{-1}\widehat{\mathbf{S}}_n(\tau_1,\ldots,\tau_k) \xrightarrow{D} \chi^2_{k(m-1)}. \quad (5.2)$$

(3) *Furthermore, assume that $\mathbf{u}^\top\mathbf{Y} = \boldsymbol{\gamma}^\top\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y} + \boldsymbol{\alpha}^\top\mathbf{X} + \phi(\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y})\varepsilon$ where $\phi(\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y})$ is a real function linear in $\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}$, $\boldsymbol{\gamma} \in \mathbb{R}^{m-1}$ and $\boldsymbol{\alpha} \in \mathbb{R}^p$ are vector parameters, and ε is independent with $(\mathbf{Y}^\top\boldsymbol{\Gamma}_{\mathbf{u}}, \mathbf{X}^\top)^\top$. Then it holds for any $\tau_i, \tau_j \in (0,1)$ that $\boldsymbol{\Sigma}_{ij}(\tau_i,\tau_j) = (\min\{\tau_i,\tau_j\} - \tau_i\tau_j)\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{W}\boldsymbol{\Gamma}_{\mathbf{u}}$, where*

$$\mathbf{W} = \mathrm{E}[\mathbf{Y} - \mathrm{E}(\mathbf{Y}\mathbf{X}^\top)\{\mathrm{E}(\mathbf{X}\mathbf{X}^\top)\}^{-1}\mathbf{X}][\mathbf{Y} - \mathrm{E}(\mathbf{Y}\mathbf{X}^\top)\{\mathrm{E}(\mathbf{X}\mathbf{X}^\top)\}^{-1}\mathbf{X}]^\top.$$

In addition, $\widehat{\mathbf{S}}_n(\tau) \xrightarrow{D} (\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{W}\boldsymbol{\Gamma}_{\mathbf{u}})^{1/2}\mathbf{B}_{m-1}(\tau)$ and

$$T_C := \sup_{\tau\in[\varepsilon,1-\varepsilon]}\widehat{T}_n(\tau) \xrightarrow{D} \sup_{\tau\in[\varepsilon,1-\varepsilon]}Q^2_{m-1}(\tau) \quad (5.3)$$

for any $\varepsilon \in (0,0.5)$, where $\mathbf{B}_{m-1}(\tau)$ is the $(m-1)$-dimensional Brownian bridge on $[0,1]$, $Q_{m-1}(\tau)$ is the Bessel process of order $m-1$,

$$\widehat{T}_n(\tau) := \frac{1}{\tau(1-\tau)}\widehat{\mathbf{S}}_n^\top(\tau)(\boldsymbol{\Gamma}_{\mathbf{u}}^\top\widehat{\mathbf{W}}\boldsymbol{\Gamma}_{\mathbf{u}})^{-1}\widehat{\mathbf{S}}_n(\tau), \; \tau \in (0,1), \quad (5.4)$$

and $\widehat{\mathbf{W}}$ is a consistent estimator of $\mathbf{W}$.

Note that a natural consistent estimator of $\mathbf{W}$ is

$$\widehat{\mathbf{W}}_0 = n^{-1}\{\mathbb{Y} - \mathbb{X}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{Y}\}^\top\{\mathbb{Y} - \mathbb{X}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\mathbb{Y}\}$$

and that the condition of Proposition 3(3) is satisfied if $(\mathbf{Y}^\top, \mathbf{Z}^\top)^\top$ comes from a multivariate normal distribution.

Remarks 4 and 5 in the Appendix further show that the assumption of Proposition 3(3) can be weakened and that $\widehat{T}_n(\tau)$ of (5.4) inherits all the good invariance properties from the location case and adds to them certain regression invariance, if a reasonable estimator of $\widehat{\mathbf{W}}$ such as $\widehat{\mathbf{W}}_0$ is employed.

Testing $H_0(\boldsymbol{u})$ can be based on Proposition 3(2) because $H_0(\boldsymbol{u})$ implies $\widetilde{H_0}(\boldsymbol{u}) : \mathbf{c}_{\tau_i \boldsymbol{u}} = \mathbf{0}$ for $0 < \tau_1 < \cdots < \tau_k < 1$ even in the regression case. It may be beneficial especially if $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{X})$ is elliptical.

To sum up, if the assumptions are satisfied, then $H_0(\boldsymbol{u})$ can be tested in the regression case by means of the test statistics T_D and T_C of (5.2) and (5.3) using their asymptotic distributions under $H_0(\boldsymbol{u})$ stated ibidem. In principle, one could then use any statistics asymptotically equivalent to T_D or T_C under $H_0(\boldsymbol{u})$.

6. Demonstrative examples

This section illustrates the new testing possibilities with a few carefully designed and representative examples involving the test T_C based on the whole quantile process (4.5,5.3) and its χ^2 modification T_D based on a few particular quantiles (4.4,5.2), both location and regression case, both elliptical and non-elliptical distributions (with light and heavy tails), both large ($n = 5\,000$) and not too large ($n = 100$) data samples, responses with dimension $m = 2, \ldots, 20$, and regressors with dimension $p = 1, \ldots, 20$.

The results have been obtained by means of the packages quantreg [14] and ks [2] for R [24] where the latter package was employed only for the computation of conditional densities. The reported p-values regarding the process-based test T_C have been calculated thanks to the algorithm for computing the tail probabilities of restricted suprema of the squared standardized tied-down Bessel process of any order [3].

For simplicity, the centered regression rank scores have been produced by the function *ranks* of quantreg with score equal to tau. They slightly differ from those of (5.1) for observations with zero residuals due to the weighting performed by the function but it does not affect the asymptotic behavior of the tests.

6.1. Simulated data

Fig. 1 shows the process $\{\widehat{\boldsymbol{c}}_{\tau \boldsymbol{u}}\}_\tau$ obtained from a random sample of size $n = 5\,000$ from bivariate normal distribution $\mathcal{N}(0,1) \times \mathcal{N}(0,4)$ for two axial directions where only one of them corresponds to an axis of symmetry. The left picture strongly speaks for axial symmetry while the other vehemently denies it. The pictures are in harmony with Proposition 1 and confirm that the tests may be sensitive even to small departures from the null hypothesis of axial symmetry.

Figs. 2 to 4 use the process-based test T_C with the sample variance-covariance estimator and approximate the supremum over $\tau \in [0.05, 0.95]$ with the maxi-
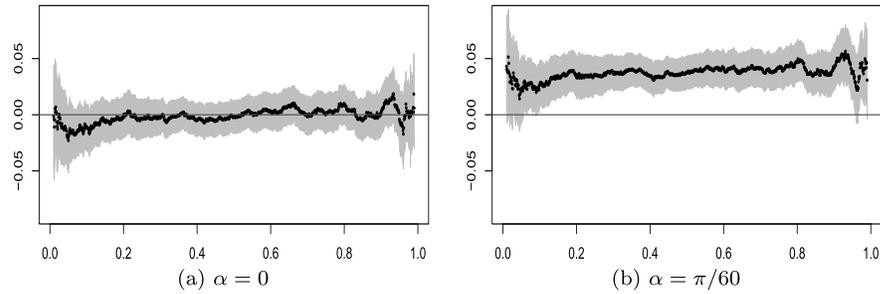
FIG 1. **Quantile regression process** $\{\widehat{\boldsymbol{c}}_{\tau\boldsymbol{u}}\}_{\tau}$. *The (scalar) coefficient estimates $\widehat{\boldsymbol{c}}_{\tau\boldsymbol{u}}$, $\tau \in [0.01, 0.99]$, together with their 90% (pointwise) confidence band (vertical axis), and their dependence on the quantile level τ (horizontal axis) for samples drawn from $\mathcal{N}(0,1) \times \mathcal{N}(0,4)$ and $\boldsymbol{u} = (\cos(\alpha), \sin(\alpha))^{\top}$ where $\alpha = 0$ (left) or $\alpha = \pi/60$ (right).*
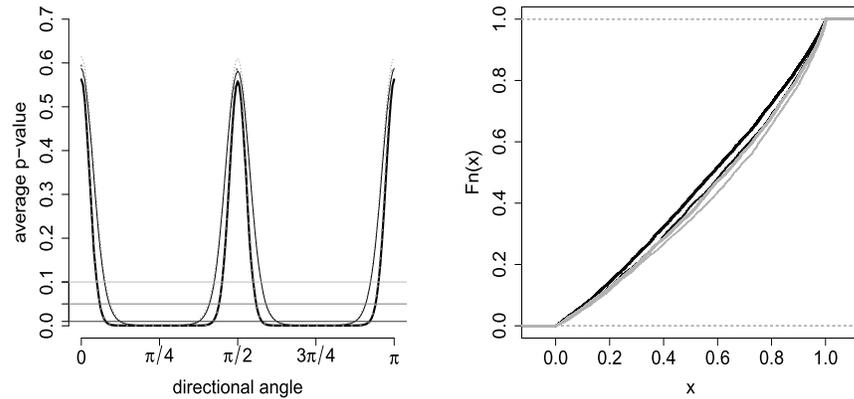
FIG 2. **Testing axial symmetry.** *The average of sample p-values from the process-based tests T_C of $H_0(\boldsymbol{u})$, $\boldsymbol{u} = (\cos(\alpha), \sin(\alpha), 0)^{\top}$ for $\alpha \in [0, \pi]$ (left), and their empirical distribution functions for $\alpha = 0$ (right). The plots correspond to samples of $n = 100$ (thin) and $n = 200$ (thick) independent observations drawn from the true model $(Y_1, Y_2, Y_3)^{\top} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^{\top} \sim \mathcal{N}(0,1) \times \mathcal{N}(0,4) \times \mathcal{N}(0,9)$ that have been modeled only with the intercept (solid black) or with the intercept and another regressor $Z \sim \mathcal{N}(0,1)$ independent of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)^{\top}$ (dotted or gray).*

mum over a dense equispaced grid on the interval $[0.05, 0.95]$, i.e., with a conservative (lower) estimate.

Fig. 2 presents average p-values obtained from 1 000 samples of size $n = 100$ or 200 of trivariate normally distributed responses from $\mathcal{N}(0,1) \times \mathcal{N}(0,4) \times \mathcal{N}(0,9)$ with or without a single stochastic independent regressor following the standard normal distribution. The individual p-values have been computed from each sample for several directions in the form $\boldsymbol{u} = (\cos(\alpha), \sin(\alpha), 0)^{\top}$ for $\alpha \in [0, \pi]$, and their empirical distributions for the axial directional angle $\alpha = 0$ (corresponding to the null hypothesis) are also reported.
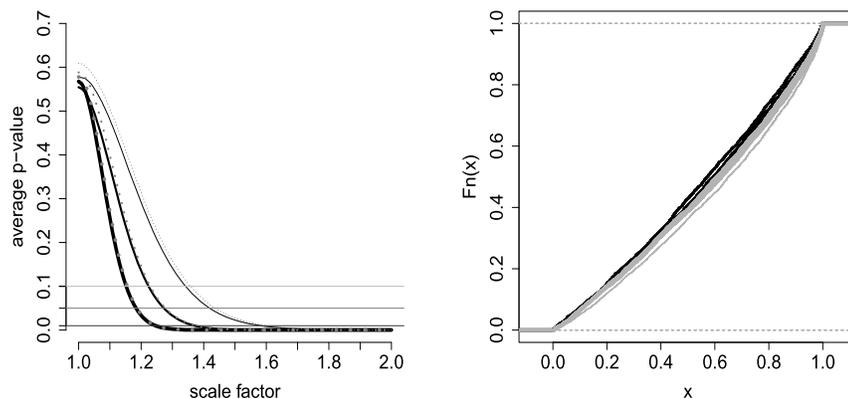
FIG 3. **Testing equality of scale.** *The average of sample p-values from the process-based tests T_C for various scale factors $R \in [1,2]$ (left) and their empirical distribution functions under the null hypothesis of $R = 1$ (right). The plots have been obtained for samples containing $n = 100$ (thin), $n = 200$ (normal), and $n = 400$ (thick) observations drawn from the true model $(Y_1, Y_2)^\top = (\varepsilon_1, R\varepsilon_2)^\top$ with independent $\varepsilon_1 \sim \mathcal{N}(0,1)$ and $\varepsilon_2 \sim \mathcal{N}(0,1)$ that have been modeled only with the intercept (solid black) or with the intercept and another regressor $Z \sim \mathcal{U}(-1,1)$ independent of ε_1 and ε_2 (dotted or gray).*

Next Fig. 3 similarly displays average p-values obtained for direction $\boldsymbol{u} = (\cos(\pi/4), \sin(\pi/4))^\top$ from 1 000 bivariate random samples of size $n = 100, 200$, or 400. Each sample $\boldsymbol{Y}_i = (Y_i^{(1)}, Y_i^{(2)})^\top$, $i = 1, \ldots, n$, comes from the bivariate standard normal distribution and results in several individual p-values, each computed from the responses $(Y_i^{(1)}, RY_i^{(2)})^\top$, $i = 1, \ldots, n$, for a scaling factor $R \in [1,2]$ and possibly also for one independent scalar regressor from the uniform distribution on $[-1, 1]$. The null hypothesis of equally scaled marginal distributions corresponds to the value $R = 1$ and to the reported empirical distribution functions of the sample p-values.

The observed average p-values produced by the tests T_C are slightly higher than 0.55 under the null hypothesis in both cases, although they are always based on 1 000 replications. This discrepancy diminishes with growing sample sizes and occurs mainly due to the data samples too small for the asymptotic approximation to hold perfectly, although the use of discrete approximation and (upper) p-value estimates may also play a marginal role. Apart from the small issue with the seemingly conservative size, the tests T_C behave as expected, which means in line with Propositions 1–3.

Fig. 4 presents empirical power for comparison with average p-values. It shows the empirical power of the test T_C obtained from 1 000 random samples of size $n = 5\,000$ coming from a twenty-dimensional distribution with independent marginals for twenty-dimensional directions $\boldsymbol{u} = (\cos(\alpha), \sin(\alpha), 0, \ldots, 0)^\top$ for $\alpha \in [-\pi/60, \pi/60]$. Two centered distributions were used: (heavy-tailed) $t_3 \times 2t_3 \times \cdots \times 20t_3$ and (light-tailed) $L(1) \times 2L(1) \times \cdots \times 20L(1)$ where L denotes the Laplace distribution. The plots employ the critical values for testing levels
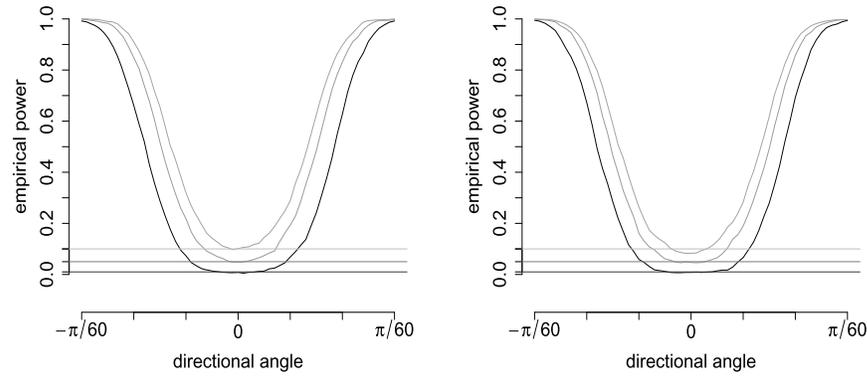
FIG 4. **Test power.** *Empirical powers from the process-based test T_C of axial symmetry around a line in direction $\boldsymbol{u} = (\cos(\alpha), \sin(\alpha), 0, \ldots, 0)^\top \in \mathbb{R}^{20}$ for $\alpha \in [-\pi/60, \pi/60]$. The plots correspond to samples of size $n = 5\,000$ of independent centered observations $(Y_1, Y_2, \ldots, Y_{20})^\top$ with independent marginals, namely $Y_i \sim i \cdot t_3$ (left) and $Y_i \sim i \cdot L(1)$ (right), $i = 1, \ldots, n$. The critical values for testing levels 0.01 (black), 0.05 (dark gray), and 0.10 (light gray) come from [3].*

$0.01, 0.05$, and 0.10 obtained from the tables published in [3] (that produce similar results as those of [1]).

The remaining figures illustrate the χ^2-tests T_D, based on the sample variance-covariance matrix estimators. They use $(m+p-1)$-dimensional data samples of size n obtained from the multivariate uniform distribution on $[-2, 2]^{m+p-1}$ ($\mathcal{U}$), multivariate standard normal distribution ($\mathcal{N}$), and/or multivariate standard t distribution with df degrees of freedom (t_{df}). The first $p - 1$ dimensions are taken for the stochastic regressor vectors $\boldsymbol{Z}_i$, the next m dimensions are considered as the error vectors $\boldsymbol{\varepsilon}_i$, and the response vectors then follow the model $\boldsymbol{Y}_i = \boldsymbol{1}_m + \boldsymbol{1}_m \boldsymbol{1}_{p-1}^\top \boldsymbol{Z}_i + \boldsymbol{\varepsilon}_i + \mathrm{I}[i > n/2]\Delta\boldsymbol{1}_m$, $i = 1, \ldots, n$, where $\Delta = 0$ for each null hypothesis and $\Delta > 0$ for each alternative considered. The tests themselves are based on n_τ quantile levels τ equidistantly distributed in the interval $[\varepsilon_\tau, 1 - \varepsilon_\tau]$ including its end points (and $\tau = 0.5$ for $n_\tau = 1$). They check the null hypothesis $H_0(\boldsymbol{u})$, $\boldsymbol{u} = (1, \boldsymbol{0}^\top)^\top$, of axial symmetry of the (conditional) response distribution around an axis parallel with the first coordinate axis in the response space.

The figures illustrate the dependence of the tests T_D on ε_τ (Fig. 5), n_τ (Fig. 6), m (Fig. 7), p (Fig. 8), n (Fig. 9), Δ (Fig. 10), or on the data distribution (Fig. 11). Solid lines generally correspond to the test behavior under the null hypothesis, dashed lines generally correspond to the test behavior under the shift alternative with $\Delta = 0.5$. The dimension of responses (m), the dimension of regressor vectors $\boldsymbol{X}_i = (1, \boldsymbol{Z}_i^\top)^\top$ including the first unit coordinate (p), and the underlying distributions $\mathcal{L}$ of $(\boldsymbol{Z}_i^\top, \boldsymbol{\varepsilon}_i^\top)^\top$, $i = 1, \ldots, n$, are indicated below each picture when it makes sense; usually $\mathcal{L} \in \{\mathcal{N}\}$ or $\mathcal{L} \in \{\mathcal{U}, \mathcal{N}, t_7\}$. Always $n = 1\,000$ except for Fig. 9 that shows how the tests depend on the number of observations n. If $m = 4$ and $p = 2$ or vice versa, then the regression
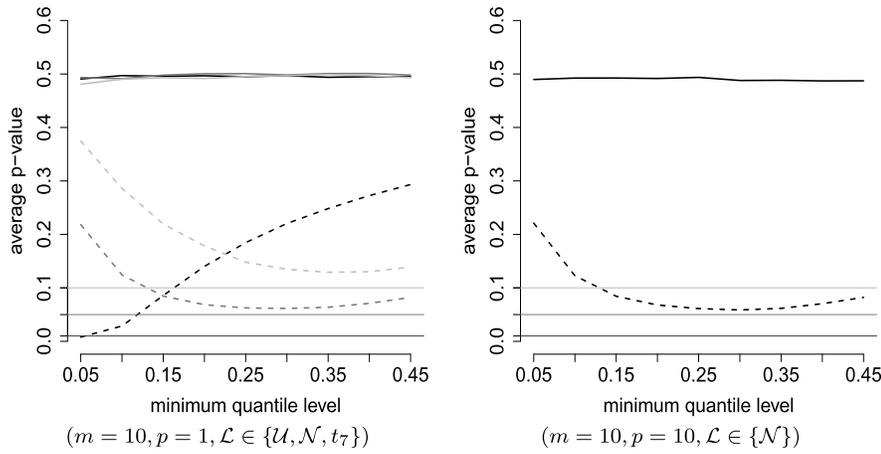
FIG 5. **Test dependence on the minimum quantile level (ε_τ) considered.** *The average sample p-values of the χ^2-tests T_D of axial symmetry (along the first coordinate axis) in dependence on the minimum quantile level $\varepsilon_\tau \in \{0.05, 0.10, \ldots, 0.45\}$ for $\Delta = 0$ (the null hypothesis, solid line) and $\Delta = 0.5$ (an alternative, dashed line). The color lightens with the order of the distribution in the list; the distribution mentioned first is thus always associated with black.*
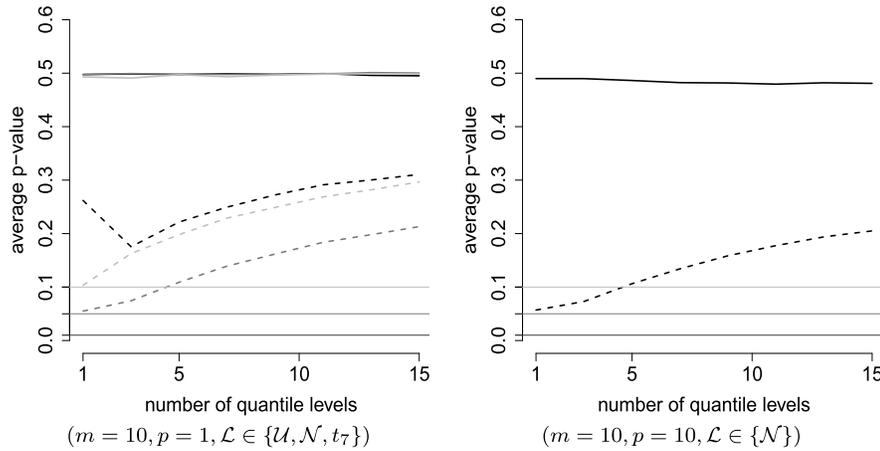

FIG 6. **Test dependence on the number of quantile levels (n_τ).** *The average sample p-values of the χ^2-tests T_D of axial symmetry (along the first coordinate axis) in dependence on the number $n_\tau \in \{1, 3, \ldots, 15\}$ of considered quantile levels, equidistantly spaced between $\tau = 0.2$ and $\tau = 0.8$, for $\Delta = 0$ (the null hypothesis, solid line) and $\Delta = 0.5$ (the alternative, dashed line). The color lightens with the order of the distribution in the list; the distribution mentioned first is thus always associated with black.*

variant of the tests uses the computationally demanding estimation of conditional densities with the normal kernel and the bandwidth optimal for normal

densities (for the sake of quick computation). The reported average p-values are based on $N = 1\,000$ replications in such cases and on $N = 10\,000$ replications in the others.

Fig. 5 uses the tests T_D for two values of τ, $\tau = \varepsilon_\tau$ and $\tau = 1 - \varepsilon_\tau$, and shows their dependence on ε_τ, $\varepsilon_\tau = 0.05, 0.10, \ldots, 0.45$. Accordingly, it seems quite prudent to choose $\tau = 0.2$ and $\tau = 0.8$ for the two values of τ as a reasonable compromise, at least for the alternatives considered here. This is why this couple of quantile levels is used in the other settings whenever two values of τ are employed, i.e., in Figs. 7 and 8.

Fig. 6 investigates the dependence of the tests T_D on the number of quantile levels considered. It uses $\tau = 0.5$ for $n_\tau = 1$, and n_τ equidistant values of τ from the interval $[0.2, 0.8]$ including its end points for $n_\tau = 3, 5, \ldots, 15$. The results are probably highly dependent on the choice of alternatives, but they nevertheless indicate that n_τ should not be chosen pointlessly too high; see also Proposition 1. Therefore, the remaining pictures employ only $n_\tau = 1$ and $\tau = 0.5$ (Figs. 9, 10, and 11) or $n_\tau = 2$ and $\tau = 0.2$ and 0.8 (Figs. 7 and 8).
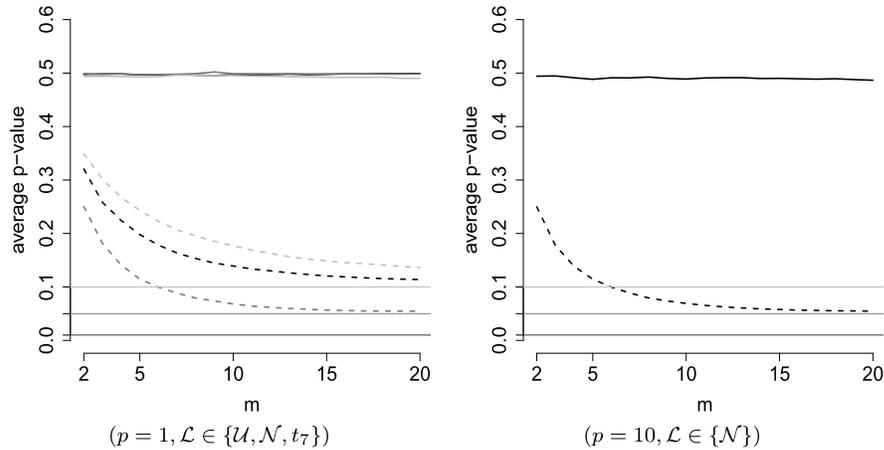

$(p = 1, \mathcal{L} \in \{\mathcal{U}, \mathcal{N}, t_7\})$ $\qquad$ $(p = 10, \mathcal{L} \in \{\mathcal{N}\})$

FIG 7. **Test dependence on the dimension of responses (m).** *The average sample p-values of the χ^2-tests T_D of axial symmetry (along the first coordinate axis) in dependence on the dimension of responses $m \in \{2, \ldots, 20\}$ for $\Delta = 0$ (the null hypothesis, solid line) and $\Delta = 0.5$ (the alternative, dashed line). The color lightens with the order of the distribution in the list; the distribution mentioned first is thus always associated with black.*

Fig. 7 displays the dependence of the tests T_D on the dimension of responses $m = 2, \ldots, 20$, and Fig. 8 shows their behavior for different dimensions of the regressor $p = 1, \ldots, 20$. The power of the test grows with m under the considered alternative, while it seems to be unaffected by changes in p.

Fig. 9 shows how the tests T_D depend on the number of observations $n = 100, \ldots, 1\,000$, while Fig. 10 illustrates the dependence of the test on the shift Δ in the data generating model. The observed dependences are as expected. Finally, Fig. 11 illustrates the test performance for the multivariate t_{df} distri-
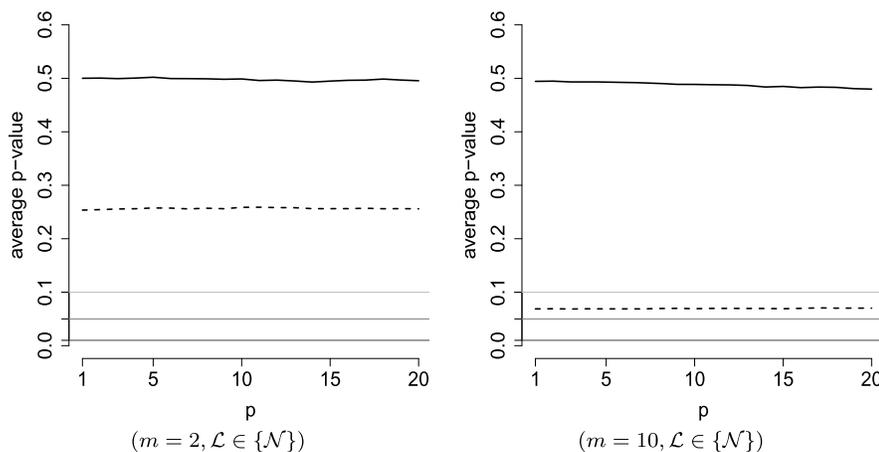
Fig 8. **Test dependence on the dimension of regressors (p).** *The average sample p-values of the χ^2-tests of axial symmetry (along the first coordinate axis) in dependence on the dimension of regressors (including the intercept) $p \in \{1, \ldots, 20\}$ for $\Delta = 0$ (the null hypothesis, solid line) and $\Delta = 0.5$ (the alternative, dashed line).*
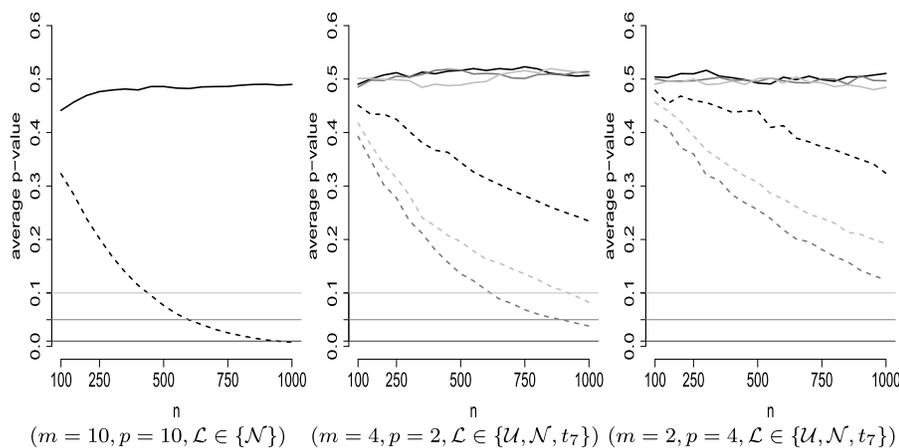

Fig 9. **Test dependence on the number of observations (n).** *The average sample p-values of the χ^2-tests of axial symmetry (along the first coordinate axis) in dependence on the number of observations $n \in \{100, 150, \ldots, 1\,000\}$ for $\Delta = 0$ (the null hypothesis, solid line) and $\Delta = 0.5$ (the alternative, dashed line). The color lightens with the order of the distribution in the list; the distribution mentioned first is thus always associated with black.*

bution and the average p-value is plotted as a function of degrees of freedom $df = 5, \ldots, 20$.

To sum up the simulation results, the χ^2-tests T_D behave as expected and in harmony with the theory. They seem to be correctly sized at least for the dimension of responses as high as $m = 10$ and for the dimension of regressors
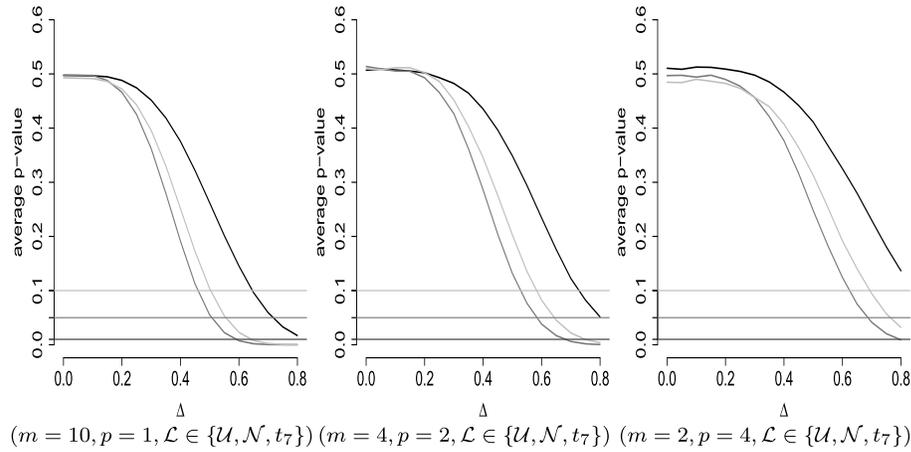
$(m = 10, p = 1, \mathcal{L} \in \{\mathcal{U}, \mathcal{N}, t_7\})$ $(m = 4, p = 2, \mathcal{L} \in \{\mathcal{U}, \mathcal{N}, t_7\})$ $(m = 2, p = 4, \mathcal{L} \in \{\mathcal{U}, \mathcal{N}, t_7\})$

FIG 10. ***Test dependence on the shift (Δ) in the alternatives.*** *The average sample p-values of the χ^2-tests T_D of axial symmetry (along the first coordinate axis) in dependence on the shift parameter $\Delta \in \{0, 0.05, 0.1, \ldots, 0.8\}$ in the data generating model. The color lightens with the order of the distribution in the list; the distribution mentioned first is thus always associated with black.*

as high as $p = 10$ when the conditional density estimation is not needed, $n \geq 20(m + p)$, and $n_\tau \leq 5$ or so, which covers the most typical situations. In general, it can also be recommended not to choose n_τ pointlessly too high and ε_τ pointlessly too small. One should be more careful, though, if the density estimation is required, i.e., in the non-normal regression case. Nevertheless, the tests may then behave reasonably even for $m + p - 1 = 4$ if n is large enough such as $n = 1\,000$, at least in the cases considered here.

6.2. Test comparison

The general asymptotic tests presented here have no close competitors, perhaps except for the follow-up tests for elliptical distributions; see [11] for the tests and the comparison. Consequently, the comparison with other available tests is possible only in very special cases.

The only known test working with general axes of symmetry is that of [20]. Therefore, it is used here as a benchmark with the two recommended values of the auxiliary binning parameter k (namely $k = 2$ and $k = 3$). However, the test is only bivariate and works with fixed axes instead of axial directions. It is applied to the problem of testing symmetry of a bivariate normal distribution (with zero mean, unit marginal variances and correlation ϱ) around the x-axis and compared there with the process-based test T_C presented here for testing level $\alpha = 0.05$ in terms of empirical size ($\varrho = 0$) and power ($\varrho > 0$) obtained by means of 10 000 simulated independent random samples with $n = 60, 75, 100$ or 150 observations. The benchmark empirical powers and sizes were copied from
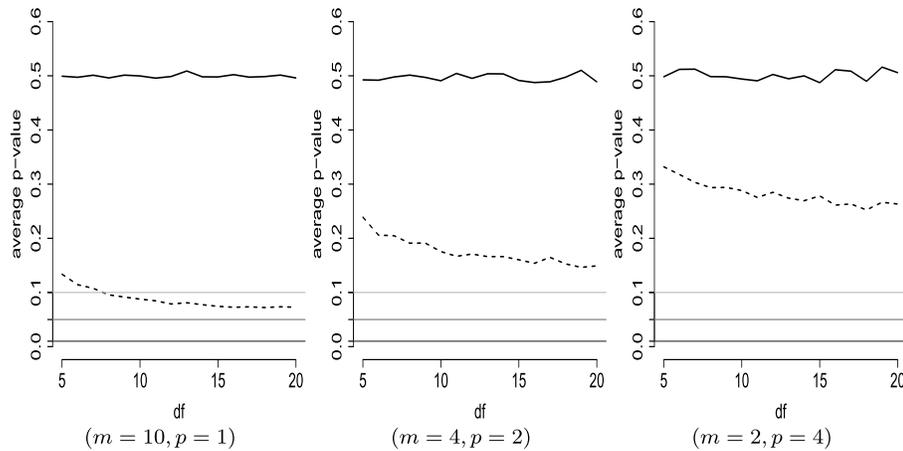
FIG 11. **Test dependence on the degrees of freedom (df) of the t distribution used.** *The average sample p-values of the χ^2-tests T_D of axial symmetry (along the first coordinate axis) in dependence on the number of degrees of freedom $df \in \{5, 6, \ldots, 20\}$ of the multivariate t distribution underlying the data for $\Delta = 0$ (the null hypothesis, solid line) and $\Delta = 0.5$ (the alternative, dashed line).*

the original Table 3 of [20] for maximum reliability. See Table 1 for the results.

The test T_C appears conservative for the small sample sizes considered, which is in line with Figs. 2 and 3. In spite of the fact, it still beats the benchmark considerably in terms of empirical test power.

6.3. Real data

This section applies the process-based test T_C to three real data sets for the sake of illustration.

First, consider the famous Fisher's Iris (flower) data set as included in R. It consists of 50 samples from each of the three Iris species considered. Each sample contains measurements regarding the length and the width of the sepals and petals (in centimeters). Assume the null hypothesis that the probability distribution of petal length is the same for all the three species up to a location shift. According to Section 3, it can be tested by means of the presented axial symmetry tests, direction $\boldsymbol{u} = (1, 1, 1)^\top / \sqrt{3}$ and the combined sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{50}$ where $\boldsymbol{X}_i = (X_i^{(1)}, X_i^{(2)}, X_i^{(3)})^\top$ and $X_i^{(j)}$ is the ith observation (of petal length) from the jth sample. If the process-based test T_C is applied, then the null hypothesis is rejected with p-value about 0.0008. Similar results can be obtained even for the petal width.

The axial symmetry of certain economic or financial data distributions may have some meaningful economic interpretations; see, e.g., [21] for an application to exchange rates. For example, consider 1146 log-returns of ten Forex 1M exchange rates (AUD/CHF, AUD/JPY, EUR/CAD, EUR/CHF, GBP/CAD,

TABLE 1

The table relates to the problem of testing symmetry of bivariate normal distribution (with zero mean, unit marginal variances and correlation ϱ) around the x-axis. It uses the test of [20] with the two recommended values of the auxiliary binning parameter k (namely $k = 2$ and $k = 3$) as a benchmark, denoted as $B(k = 2)$ and $B(k = 3)$. The benchmark is compared with the process-based test T_C in terms of empirical power (for $\varrho > 0$) or size (for $\rho = 0$) for testing level $\alpha = 0.05$, based on $10\,000$ simulations of independent samples with $n = 60, 75, 100$ or 150 observations. The benchmark empirical sizes and powers were only copied from the original Table 3 of [20] to minimize the chance of an error. The critical values for T_C were obtained from [1].

Test comparison for bivariate normal distribution and axial symmetry

ϱ	$B(k=2)$	$B(k=3)$	T_C	$B(k=2)$	$B(k=3)$	T_C
		$n = 60$			$n = 75$	
0.0	0.0544	0.0506	0.0238	–	0.0457	0.0297
0.1	0.0726	0.0660	0.0538	–	0.0660	0.0692
0.2	0.1346	0.1223	0.1630	–	0.1398	0.2151
0.3	0.2464	0.2501	0.3825	–	0.2900	0.5018
0.4	0.4336	0.4451	0.6737	–	0.5305	0.7965
0.5	0.6478	0.6640	0.8975	–	0.7786	0.9636
0.6	0.8405	0.8682	0.9877	–	0.9382	0.9982
0.7	0.9549	0.9737	0.9997	–	0.9928	0.9998
		$n = 100$			$n = 150$	
0.0	0.0498	–	0.0324	0.0548	0.0515	0.0298
0.1	0.0737	–	0.0926	0.1033	0.0963	0.1327
0.2	0.1773	–	0.3106	0.2736	0.2878	0.4740
0.3	0.3705	–	0.6618	0.5686	0.6038	0.8620
0.4	0.6255	–	0.9230	0.8331	0.8721	0.9897
0.5	0.8500	–	0.9937	0.9728	0.9833	1.0000
0.6	0.9707	–	1.0000	0.9977	0.9995	1.0000
0.7	0.9985	–	1.0000	1.0000	1.0000	1.0000

GBP/USD, NZD/CHF, USD/CHF, USD/NOK, XAU/AUD) from 13/11/2014 19:11 to 14/11/2014 14:16, combined to one ten-dimensional sample in the same order. Then the null hypothesis assuming both serial independence and exchangeability (up to a shift) would be rejected (for $\boldsymbol{u} = (1, 1, \ldots, 1)^\top / \sqrt{10}$) by T_C with the p-value much less than 0.01. The same null hypothesis applied only to the bivariate sample corresponding to the first two exchange rates would be rejected (for $\boldsymbol{u} = (1, 1)^\top / \sqrt{2}$) with p-value less than 0.0005. The same conclusion would be obtained for the bivariate sample of GBP/CAD and GBP/USD. The null hypothesis assuming both serial independence and symmetry around the last (and possibly shifted) coordinate axis ($\boldsymbol{u} = (0, \ldots, 1)^\top$) would also be rejected in the ten-variate sample as well as in the two bivariate ones (all p-values less than 0.0005).

If 626 (virtually serially uncorrelated) log-returns of four daily exchange rates (AUD/CZK, CAD/CZK, EUR/CZK, USD/CZK) from 2/5/2017 to 30/10/2019

were considered as a four-variate i.i.d. sample, then the null hypothesis of exchangeability would still be rejected with p-value less than 0.001. Similarly clear rejection would be obtained also for the null hypothesis assuming axial symmetry around the last coordinate axis or for the bivariate sample consisting only of EUR/CZK and USD/CZK.

As the third possible application, consider the Australian athletes data set `ais` contained in the R package DAAG [17]. Its subsets are used in both [12] and [8] for testing certain symmetry hypotheses including axial symmetries. In particular, the latter article tested (and rejected) the spherical symmetry of the joint distribution of the logarithms of the red blood cell count, white blood cell count and hemoglobin concentration, where all the three characteristics were obtained for 202 athletes. There is no wonder the spherical symmetry was rejected because the process-based test T_C rejects the axial symmetry for (coordinate) axial directions $\boldsymbol{u} = (0,0,1)^\top$ and $\boldsymbol{u} = (1,0,0)^\top$, always with p-value less than 0.0005. The results are in line with those regarding the spherical symmetry of all the three bivariate marginals, reported in [8].

Appendix

This section contains the proofs of Propositions 1 to 3 as well as the technical remarks commenting on them. The first three pertain to Proposition 2 while the remaining two comment on Proposition 3.

Proof of Proposition 1. Obviously, (1) follows from the text preceding the proposition. Claims (2) and (3) are proved here.

Suppose that $\boldsymbol{Y}$ is elliptically distributed with median $\boldsymbol{\mu} = \boldsymbol{0}$ and scatter matrix $\boldsymbol{\Sigma}$ with $\det(\boldsymbol{\Sigma}) = 1$ without any loss of generality. Fix $\tau \in (0,1)$ and define symmetric matrix $\mathbf{V} := \boldsymbol{\Sigma}^{-1/2}$.

Transformation $\boldsymbol{W} := \mathbf{V}\boldsymbol{Y}$ leads to the spherically distributed random vector $\boldsymbol{W}$ and to the transformed gradient conditions (2.3) and (2.4) that turn into

$$\tau = \mathrm{E}_{\boldsymbol{W}}\big\{\mathrm{I}(\boldsymbol{b_{\tau u}}^\top \mathbf{V}^{-1}\boldsymbol{W} - a_{\tau u} < 0)\big\} \text{ and}$$

$$\mathbf{V}\boldsymbol{D_{\tau u}} = \frac{1}{1-\tau}\mathrm{E}_{\boldsymbol{W}}\big\{\mathrm{I}(\boldsymbol{b_{\tau u}}^\top \mathbf{V}^{-1}\boldsymbol{W} - a_{\tau u} \geq 0)\boldsymbol{W}\big\}$$
$$- \frac{1}{\tau}\mathrm{E}_{\boldsymbol{W}}\big\{\mathrm{I}(\boldsymbol{b_{\tau u}}^\top \mathbf{V}^{-1}\boldsymbol{W} - a_{\tau u} < 0)\boldsymbol{W}\big\},$$

which are the gradient conditions for the directional τ-quantile of $\boldsymbol{W}$ in the unit direction $\boldsymbol{v} = \mathbf{V}\boldsymbol{u}/\|\mathbf{V}\boldsymbol{u}\|$, with $\boldsymbol{b_{\tau v}}(\boldsymbol{W}) = \|\mathbf{V}\boldsymbol{u}\|\mathbf{V}^{-1^\top}\boldsymbol{b_{\tau u}}(\boldsymbol{Y})$. As the spherical distribution of $\boldsymbol{W}$ is symmetric around axes in all directions ($\boldsymbol{v}$ not excluded), necessarily $\boldsymbol{b_{\tau v}}(\boldsymbol{W}) = \boldsymbol{v}$ and $\boldsymbol{c_{\tau u}}(\boldsymbol{Y}) = -\boldsymbol{\Gamma_u}^\top \boldsymbol{b_{\tau u}}(\boldsymbol{Y}) = -\boldsymbol{\Gamma_u}^\top \mathbf{V}^2\boldsymbol{u}/\|\mathbf{V}\boldsymbol{u}\|^2 = -\boldsymbol{\Gamma_u}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{u}/(\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{u})$, irrespective of $\tau \in (0,1)$. Consequently, $\boldsymbol{c_{\tau u}} = \boldsymbol{0}$ if and only if $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{\Sigma}$, i.e., the directional vector of an axis of symmetry of $\mathcal{L}(\boldsymbol{Y})$. $\square$

Proof of Proposition 2. The arguments mimic those of [5] and [16]. The definition of $\widehat{\mathbf{a}}(\tau)$ and its properties including $n^{-1}\mathbf{1}_n^\top\widehat{\mathbf{a}}(\tau) = (1 - \tau)$ imply

$$
\begin{aligned}
\widehat{\mathbf{S}}_n(\tau) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top(\mathbf{Y}_i - \overline{\mathbf{Y}}_n)\{\widehat{a}_i(\tau) - (1 - \tau)\} \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}_i\{\widehat{a}_i(\tau) - (1 - \tau)\} \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}_i[\tau - \mathrm{I}\{\mathbf{u}^\top\mathbf{Y}_i < \widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\}] \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}_i\mathrm{I}\{\mathbf{u}^\top\mathbf{Y}_i = \widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\}\{\widehat{a}_i(\tau) - 1\}
\end{aligned}
$$

where $\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)$ is the sample τ-quantile of $\mathbf{u}^\top\mathbf{Y}$. As for the second term, say II,

$$
\sup_{\tau\in(0,1)}\|II\| \leq \frac{m}{\sqrt{n}}\max_{i=1,\dots,n}\|\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}_i\| \to 0
$$

in probability because of the moment assumptions, and because its sum contains no more than m nonzero summands almost surely. Consequently,

$$
\widehat{\mathbf{S}}_n(\tau) = \widehat{\mathbf{G}}_n\{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\} + o_P(1) \text{ where } \widehat{\mathbf{G}}_n(a) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}_i\{\tau - \mathrm{I}(\mathbf{u}^\top\mathbf{Y}_i < a)\}.
$$

Define further $\mathbf{G}_n(a) = \mathrm{E}\{\widehat{\mathbf{G}}_n(a)\}$. The null hypothesis $H_0(\boldsymbol{u})$ and gradient conditions (2.3) and (2.4) lead to $\mathrm{E}\{\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y}\mathrm{I}(\mathbf{u}^\top\mathbf{Y} < a)\} = F_{\mathbf{u}}(a)\mathrm{E}(\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{Y})$ and

$$
\mathbf{G}_n(a) = \sqrt{n}\boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathrm{E}[\mathbf{Y}\{\tau - F_{\mathbf{u}}(a)\}],
$$

where $F_{\mathbf{u}}$ is the distribution function of $\mathbf{u}^\top\mathbf{Y}$. Then

$$
\widehat{\mathbf{G}}_n\{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\} = \widehat{\mathbf{G}}_n\{F_{\boldsymbol{u}}^{-1}(\tau)\} + \left[\mathbf{G}_n\{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\} - \mathbf{G}_n\{F_{\boldsymbol{u}}^{-1}(\tau)\}\right] + \mathbf{R}_n,
$$

where $\mathbf{R}_n = [\widehat{\mathbf{G}}_n\{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\} - \mathbf{G}_n\{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\}] - [\widehat{\mathbf{G}}_n\{F_{\boldsymbol{u}}^{-1}(\tau)\} - \mathbf{G}_n\{F_{\boldsymbol{u}}^{-1}(\tau)\}]$ $= -\mathbb{G}_n\{g_{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)} - g_{F_{\boldsymbol{u}}^{-1}(\tau)}\}$, where the empirical process $\mathbb{G}_n$ is defined as $\mathbb{G}_n g = \frac{1}{\sqrt{n}}\{\sum_{i=1}^n [g(\mathbf{Y}_i) - \mathrm{E}g(\mathbf{Y})]\}$ and $g_a(\mathbf{y}) = \boldsymbol{\Gamma}_{\mathbf{u}}^\top\mathbf{y}\mathrm{I}(\mathbf{u}^\top\mathbf{y} < a)$. Since $\mathrm{E}\|\mathbf{Y}\|^{2+\delta} < \infty$, one can show similarly as in Example 19.6. of [25] that $\mathcal{F} = \{g_a(\mathbf{y}) : a \in \mathbb{R}\}$ is a Donsker class. Furthermore, it can be directly verified by means of the Hölder inequality that if $I_i(\alpha) = \mathrm{E}\{g_{\alpha,i}(\mathbf{Y}) - g_{F_{\boldsymbol{u}}^{-1}(\tau),i}(\mathbf{Y})\}^2$, then $I_i(\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)) \to 0$ in probability as $n \to \infty$ for each component $i = 1,\dots,m-1$. It then follows from Lemma 19.24 of [25] that $\mathbf{R}_n = o_P(1)$. The Taylor theorem for $\mathbf{G}_n$ and the Bahadur representation of sample quantiles of [26] result in

$$
\widehat{\mathbf{G}}_n\{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau)\} = \widehat{\mathbf{G}}_n\{F_{\boldsymbol{u}}^{-1}(\tau)\} + \frac{\partial\mathbf{G}_n(a)}{\partial a}\{F_{\boldsymbol{u}}^{-1}(\tau)\}\{\widehat{F}_{\boldsymbol{u},n}^{-1}(\tau) - F_{\boldsymbol{u}}^{-1}(\tau)\} + o_P(1)
$$

$$= \widehat{\mathbf{G}}_n\{F_{\mathbf{u}}^{-1}(\tau)\} - f_{\mathbf{u}}\{F_{\mathbf{u}}^{-1}(\tau)\}\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathrm{E}(\mathbf{Y})\sqrt{n}\{\widehat{F}_{\mathbf{u},n}^{-1}(\tau) - F_{\mathbf{u}}^{-1}(\tau)\} + o_P(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\{\mathbf{Y}_i - \mathrm{E}(\mathbf{Y})\}[\tau - \mathrm{I}\{\mathbf{u}^{\top}\mathbf{Y}_i < F_{\mathbf{u}}^{-1}(\tau)\}] + o_P(1),$$

which proves (1) and implies (2) thanks to the Lindeberg-Lévy central limit theorem, because

$$\widehat{\mathbf{S}}_n(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\xi}_i(\tau) + o_P(1)$$

where $\boldsymbol{\xi}_i(\tau) = \boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\{\mathbf{Y}_i - \mathrm{E}(\mathbf{Y})\}[\tau - \mathrm{I}\{\mathbf{u}^{\top}\mathbf{Y}_i < F_{\mathbf{u}}^{-1}(\tau)\}]$, $i = 1, \ldots, n$, are independent and identically distributed random vectors.

Assume $\varepsilon \in (0, 0.5)$. A careful inspection of the previous steps reveals that all the $o_P(1)$ terms considered above converge to $\mathbf{0}$ uniformly for $\tau \in [\varepsilon, 1 - \varepsilon]$, and thus $\sup_{\tau \in [\varepsilon, 1-\varepsilon]}\|\widehat{\mathbf{S}}_n(\tau) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\xi}_i(\tau)\| = o_P(1)$. The assumption of (3) leads to

$$\mathrm{Cov}\left(\boldsymbol{\xi}_i(s), \boldsymbol{\xi}_i(t)\right) = (\min\{s, t\} - st)\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\{\mathrm{Var}\left(\mathbf{Y}\right)\}\boldsymbol{\Gamma}_{\mathbf{u}}.$$

Consequently, the theory of Donsker classes, [25, Chapter 19], applied to the process $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\xi}_i(\tau)$, implies that the process $\widehat{\mathbf{S}}_n$ indexed by $\tau \in [\varepsilon, 1-\varepsilon]$ converges in distribution to the process $\{\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\{\mathrm{Var}\left(\mathbf{Y}\right)\}\boldsymbol{\Gamma}_{\mathbf{u}}\}^{1/2}\mathbf{B}_{m-1}$, where $\mathbf{B}_{m-1}$ is the $(m-1)$-dimensional Brownian bridge. $\square$

Remark 1. The assumption in Proposition 2(3) is satisfied whenever the conditional quantile function of $\mathbf{u}^{\top}\mathbf{Y}$ given $\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}$ is linear in $\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}$ because then $\mathbf{u}^{\top}\mathbf{Y}$ and $\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}$ are independent under the null hypothesis $H_0(\boldsymbol{u})$. For example, this happens if $\mathbf{u}^{\top}\mathbf{Y}$ given $\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}$ is normally distributed with the mean linear in $\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}$ and constant variance. This occurs, e.g., if $\mathbf{Y}$ is multivariate normal.

Remark 2. The proof of Proposition 2 suggests that the τ-indexed process $\widehat{\mathbf{S}}_n$ still converges to a Gaussian process $\mathbf{U}$ even when the assumption in Proposition 2 (3) is violated. Therefore, the test statistic T_C of (4.5) could be used for testing $H_0(\boldsymbol{u})$ even in the general situation if the critical values were computed from the more general process $\mathbf{U}$, although it would be very computationally demanding.

Remark 3. The test statistic $\widehat{T}_n(\tau)$ of (4.6) does not depend on the choice of the auxiliary matrix $\boldsymbol{\Gamma}_{\mathbf{u}}$. It is also invariant with respect to all the transformations that preserve the axial symmetry, namely to certain shift, rotation and linear scale transformations described below. Note that

$$\widehat{T}_n(\tau) = \frac{1}{n\tau(1-\tau)}\widehat{\mathbf{b}}(\tau)^{\top}\mathbb{Y}\boldsymbol{\Gamma}_{\mathbf{u}}(\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{S}_{\mathbf{Y}}\boldsymbol{\Gamma}_{\mathbf{u}})^{-1}\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbb{Y}^{\top}\widehat{\mathbf{b}}(\tau)$$

thanks to (4.1).

Independence on the choice of $\boldsymbol{\Gamma}_{\mathbf{u}}$. Recall that $\boldsymbol{\Gamma}_{\mathbf{u}}$ is a complement of $\mathbf{u}$ to an orthonormal matrix. Consequently, if $\boldsymbol{\Delta}_{\mathbf{u}}$ is a different matrix such

that $(\mathbf{u}|\boldsymbol{\Delta_u})$ is orthonormal, then there must exist an $(m-1)\times(m-1)$ regular matrix $\mathbf{A}$ such that $\boldsymbol{\Gamma_u} = \boldsymbol{\Delta_u}\mathbf{A}$. Hence,

$$\boldsymbol{\Gamma_u}(\boldsymbol{\Gamma_u^\top}\mathbf{S_Y}\boldsymbol{\Gamma_u})^{-1}\boldsymbol{\Gamma_u^\top} = \boldsymbol{\Delta_u}\mathbf{A}(\mathbf{A}^\top\boldsymbol{\Delta_u^\top}\mathbf{S_Y}\boldsymbol{\Delta_u}\mathbf{A})^{-1}\mathbf{A}^\top\boldsymbol{\Delta_u^\top}$$
$$= \boldsymbol{\Delta_u}(\boldsymbol{\Delta_u^\top}\mathbf{S_Y}\boldsymbol{\Delta_u})^{-1}\boldsymbol{\Delta_u^\top}$$

and $\widehat{T}_n(\tau)$ does not depend on the particular choice of $\boldsymbol{\Gamma_u}$, because $\widehat{\mathbf{b}}(\tau)$ does not depend on it either.

Shift invariance. Assume shifted observations $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \mathbf{s}$, $\mathbf{s} \in \mathbb{R}^m$, grouped into the matrix $\widetilde{\mathbb{Y}} = (\widetilde{\mathbf{Y}}_1, \ldots, \widetilde{\mathbf{Y}}_n)^\top$. The rank score vector does not depend on the shift, because $\mathbf{u}^\top\widetilde{\mathbb{Y}}^\top\mathbf{a} = \mathbf{u}^\top\mathbb{Y}^\top\mathbf{a} - \mathbf{u}^\top\mathbf{s}\mathbf{1}^\top\mathbf{a} = \mathbf{u}^\top\mathbb{Y}^\top\mathbf{a} - n(1-\tau)\mathbf{u}^\top\mathbf{s}$, and also the original and shifted observations are the same after centering with their mean in (4.1). Therefore, $\widehat{T}_n(\tau)$ remains the same even for the shifted observations.

Rotation invariance. Assume $\widetilde{\mathbf{u}} = \mathbf{A}\mathbf{u}$, $\boldsymbol{\Gamma_{\widetilde{u}}} = \mathbf{A}\boldsymbol{\Gamma_u}$, and $\widetilde{\mathbf{Y}}_i = \mathbf{A}\mathbf{Y}_i$ for an orthonormal matrix $\mathbf{A}_{m\times m}$, $\mathbf{A}^\top\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^\top$. Then $\widehat{T}_n(\tau)$ is the same as if it were computed from rotated observations $\mathbf{A}\mathbf{Y}_i$ for testing the axial symmetry around a line in direction $\widetilde{\mathbf{u}}$. Indeed, the rank scores would be the same thanks to $\widetilde{\mathbb{Y}} = \mathbb{Y}\mathbf{A}^\top$ and $\widetilde{\mathbf{u}}^\top\widetilde{\mathbb{Y}}^\top\mathbf{a} = \mathbf{u}^\top\mathbb{Y}^\top\mathbf{a}$, $\mathbf{S_{AY}} = \mathbf{A}\mathbf{S_Y}\mathbf{A}^\top$, and

$$\mathbb{Y}\mathbf{A}^\top\mathbf{A}\boldsymbol{\Gamma_u}(\boldsymbol{\Gamma_u^\top}\mathbf{A}^\top\mathbf{S_{AY}}\mathbf{A}\boldsymbol{\Gamma_u})^{-1}\boldsymbol{\Gamma_u^\top}\mathbf{A}^\top\mathbf{A}\mathbb{Y}^\top = \mathbb{Y}\boldsymbol{\Gamma_u}(\boldsymbol{\Gamma_u^\top}\mathbf{S_Y}\boldsymbol{\Gamma_u})^{-1}\boldsymbol{\Gamma_u^\top}\mathbb{Y}^\top.$$

Scale invariance. Define orthonormal matrix $\mathbf{Q} = (\mathbf{u}|\boldsymbol{\Gamma_u})$ and assume $\widetilde{\mathbf{Y}}_i = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top\mathbf{Y}_i$ where $\mathbf{D} = \mathrm{diag}(d_{11}, \ldots, d_{mm})$ is a (real) regular diagonal matrix with $d_{11} > 0$ and $\mathbf{D}_{-1} := \mathrm{diag}(d_{22}, \ldots, d_{mm})$. Then $\widehat{T}_n(\tau)$ also does not change for the transformed observations. Indeed, the rank scores remain the same because $\widetilde{\mathbb{Y}} = \mathbb{Y}\mathbf{Q}\mathbf{D}\mathbf{Q}^\top$ and $\mathbf{u}^\top\widetilde{\mathbb{Y}}^\top\mathbf{a} = d_{11}\mathbf{u}^\top\mathbb{Y}^\top\mathbf{a}$. In addition, $\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\boldsymbol{\Gamma_u} = \boldsymbol{\Gamma_u}\mathbf{D}_{-1}$, $\mathbf{S_{\widetilde{Y}}} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top\mathbf{S_Y}\mathbf{Q}\mathbf{D}\mathbf{Q}^\top$, and

$$\widetilde{\mathbb{Y}}\boldsymbol{\Gamma_u}(\boldsymbol{\Gamma_u^\top}\mathbf{S_{\widetilde{Y}}}\boldsymbol{\Gamma_u})^{-1}\boldsymbol{\Gamma_u^\top}\widetilde{\mathbb{Y}}^\top$$
$$= \mathbb{Y}\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\boldsymbol{\Gamma_u}(\boldsymbol{\Gamma_u^\top}\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\mathbf{S_Y}\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\boldsymbol{\Gamma_u})^{-1}\boldsymbol{\Gamma_u^\top}\mathbf{Q}\mathbf{D}\mathbf{Q}^\top\mathbb{Y}^\top$$
$$= \mathbb{Y}\boldsymbol{\Gamma_u}\mathbf{D}_{-1}(\mathbf{D}_{-1}\boldsymbol{\Gamma_u^\top}\mathbf{S_Y}\boldsymbol{\Gamma_u}\mathbf{D}_{-1})^{-1}\mathbf{D}_{-1}\boldsymbol{\Gamma_u^\top}\mathbb{Y}^\top = \mathbb{Y}\boldsymbol{\Gamma_u}(\boldsymbol{\Gamma_u^\top}\mathbf{S_Y}\boldsymbol{\Gamma_u})^{-1}\boldsymbol{\Gamma_u^\top}\mathbb{Y}^\top.$$

Proof of Proposition 3. The proof proceeds along the same lines as the proof of Proposition 2. It turns out that

$$\widehat{\mathbf{S}}_n(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Gamma_u^\top}\mathbf{Y}_i\{\tau - \mathrm{I}(\mathbf{u}^\top\mathbf{Y}_i < \widehat{\mathbf{a}}_{\tau\boldsymbol{u}}^\top\mathbf{X}_i)\} + o_P(1) = \widehat{\mathbf{G}}_n(\widehat{\mathbf{a}}_{\tau\boldsymbol{u}}) + o_P(1)$$

where

$$\widehat{\mathbf{G}}_n(\mathbf{a}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\Gamma_u^\top}\mathbf{Y}_i\{\tau - \mathrm{I}(\mathbf{u}^\top\mathbf{Y}_i < \mathbf{a}^\top\mathbf{X}_i)\}.$$

Define $\mathbf{G}_n(\mathbf{a}) := \mathrm{E}\{\widehat{\mathbf{G}}_n(\mathbf{a})\} = \sqrt{n}\left[\tau\boldsymbol{\Gamma_u^\top}\mathrm{E}(\mathbf{Y}) - \boldsymbol{\Gamma_u^\top}\mathrm{E}\{F_{\mathbf{u}^\top\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma_u^\top}\mathbf{Y}}(\mathbf{a}^\top\mathbf{X})\mathbf{Y}\}\right]$ where $F_{\mathbf{u}^\top\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma_u^\top}\mathbf{Y}}$ is the distribution function of $\mathbf{u}^\top\mathbf{Y}$ given $(\mathbf{X}^\top, \mathbf{Y}^\top\boldsymbol{\Gamma_u})^\top$.

Then
$$\frac{\partial \mathbf{G}_n}{\partial \mathbf{a}} = -\sqrt{n}\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathrm{E}\{f_{\mathbf{u}^{\top}\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}}(\mathbf{a}^{\top}\mathbf{X})\mathbf{Y}\mathbf{X}^{\top}\},$$

$\partial \mathbf{G}_n/\partial \mathbf{a}(\mathbf{a}_{\tau\boldsymbol{u}}) = \sqrt{n}\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{H}(\tau)$, and arguments analogous to those used in the proof of Proposition 2 lead to $\mathbf{R}_n := \{\widehat{\mathbf{G}}_n(\widehat{\mathbf{a}}_{\tau\boldsymbol{u}}) - \mathbf{G}_n(\widehat{\mathbf{a}}_{\tau\boldsymbol{u}})\} - \{\widehat{\mathbf{G}}_n(\mathbf{a}_{\tau\boldsymbol{u}}) - \mathbf{G}_n(\mathbf{a}_{\tau\boldsymbol{u}})\} = o_P(1)$. The Bahadur representation for $\widehat{\mathbf{a}}_{\tau\boldsymbol{u}}$ under the null hypothesis $H_0(\boldsymbol{u})$ can be derived, e.g., by means of the general approach described in [9]:

$$\sqrt{n}(\widehat{\mathbf{a}}_{\tau\boldsymbol{u}} - \mathbf{a}_{\tau\boldsymbol{u}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{G}(\tau)^{-1}\{\tau - \mathrm{I}(\mathbf{u}^{\top}\mathbf{Y}_i < \mathbf{a}_{\tau\boldsymbol{u}}^{\top}\mathbf{X}_i)\}\mathbf{X}_i.$$

Then

$$\widehat{\mathbf{G}}_n(\widehat{\mathbf{a}}_{\tau\boldsymbol{u}}) = \widehat{\mathbf{G}}_n(\mathbf{a}_{\tau\boldsymbol{u}}) - \boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{H}(\tau)\sqrt{n}(\widehat{\mathbf{a}}_{\tau\boldsymbol{u}} - \mathbf{a}_{\tau\boldsymbol{u}}) + o_P(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\eta}_i(\tau) + o_P(1),$$

where $\boldsymbol{\eta}_i(\tau) = \{\tau - \mathrm{I}(\mathbf{u}^{\top}\mathbf{Y}_i < \mathbf{a}_{\tau\boldsymbol{u}}^{\top}\mathbf{X}_i)\}\{\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}_i - \boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{H}(\tau)\mathbf{G}(\tau)^{-1}\mathbf{X}_i\}$ and random vectors $\boldsymbol{\eta}_1(\tau),\ldots,\boldsymbol{\eta}_n(\tau)$ are independent and identically distributed. The gradient conditions (2.3) and (2.4) imply $\mathrm{E}\{\boldsymbol{\eta}_1(\tau_i)\} = \mathbf{0}$ and $\mathrm{Cov}\{\boldsymbol{\eta}_1(\tau_i), \boldsymbol{\eta}_1(\tau_j)\} = \boldsymbol{\Sigma}_{ij}(\tau_i, \tau_j)$, $i, j = 1, \ldots, k$. In addition, $\sup_{\tau\in[\varepsilon,1-\varepsilon]}\|\widehat{\mathbf{S}}_n(\tau) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\eta}_i(\tau)\| = o_P(1)$ for any $\varepsilon \in (0, 0.5)$, similarly as in the location case.

If $H_0(\boldsymbol{u})$ holds with the additional assumption of (3), then $\boldsymbol{\gamma} = \mathbf{0}$ and $\phi(\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y})$ is a constant ϕ which does not depend on $\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}$. Therefore, with the obvious meaning of symbols F_ε, F_ε^{-1}, and f_ε, one has $F_{\mathbf{u}^{\top}\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}}(w) = F_\varepsilon\{(w - \boldsymbol{\alpha}^{\top}\mathbf{X})\phi^{-1}\}$, and $F_{\mathbf{u}^{\top}\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}}^{-1} = \mathbf{a}_{\tau\boldsymbol{u}}^{\top}\mathbf{X} = F_\varepsilon^{-1}(\tau)\phi + \boldsymbol{\alpha}^{\top}\mathbf{X}$, and consequently $f_{\mathbf{u}^{\top}\mathbf{Y}|\mathbf{X}}(\mathbf{a}_{\tau\boldsymbol{u}}^{\top}\mathbf{X}) = f_{\mathbf{u}^{\top}\mathbf{Y}|\mathbf{X},\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}}(\mathbf{a}_{\tau\boldsymbol{u}}^{\top}\mathbf{X}) = f_\varepsilon\{(\mathbf{a}_{\tau\boldsymbol{u}}^{\top}\mathbf{X} - \boldsymbol{\alpha}^{\top}\mathbf{X})\phi^{-1}\}/\phi = f_\varepsilon\{F_\varepsilon^{-1}(\tau)\}/\phi$. Then the matrix $\mathbf{H}(\tau)\mathbf{G}(\tau)^{-1} = \mathrm{E}(\mathbf{Y}\mathbf{X}^{\top})\{\mathrm{E}(\mathbf{X}\mathbf{X}^{\top})\}^{-1}$ is independent of τ, and the rest follows similarly as in the proof of Proposition 2. $\square$

Remark 4. One can also assume in Proposition 3(3) that $\mathbf{u}^{\top}\mathbf{Y} = \boldsymbol{\gamma}^{\top}\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y} + \boldsymbol{\alpha}^{\top}\mathbf{X} + \phi(\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}, \mathbf{X})\varepsilon$ where $\phi(\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}, \mathbf{X})$ is a real function linear in $(\mathbf{Y}^{\top}\boldsymbol{\Gamma}_{\mathbf{u}}, \mathbf{X}^{\top})^{\top}$. This assumption is satisfied, e.g., by the regression models (3.2) and (3.3). Then, under the null hypothesis $H_0(\boldsymbol{u})$, $\phi(\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}, \mathbf{X}) =: \phi(\mathbf{X})$ does not depend on $\boldsymbol{\Gamma}_{\mathbf{u}}^{\top}\mathbf{Y}$ and the statement holds with

$$\mathbf{W} = \mathrm{E}\big[\mathbf{Y} - \mathrm{E}\{\phi(\mathbf{X})^{-1}\mathbf{Y}\mathbf{X}^{\top}\}[\mathrm{E}\{\phi(\mathbf{X})^{-1}\mathbf{X}\mathbf{X}^{\top}\}]^{-1}\mathbf{X}\big]$$
$$\times \big[\mathbf{Y} - \mathrm{E}\{\phi(\mathbf{X})^{-1}\mathbf{Y}\mathbf{X}^{\top}\}[\mathrm{E}\{\phi(\mathbf{X})^{-1}\mathbf{X}\mathbf{X}^{\top}\}]^{-1}\mathbf{X}\big]^{\top}.$$

Remark 5. Statistic $\widehat{T}_n(\tau)$ of (5.4) can be proved invariant with respect to the choice of $\boldsymbol{\Gamma}_{\mathbf{u}}$ in the same way as for $p = 1$. If $\widehat{\mathbf{W}}(\mathbb{Y}, \mathbb{X})$ is a reasonable scatter estimator satisfying $\widehat{\mathbf{W}}(\mathbb{Y}\mathbf{A}^{\top} + \mathbf{1}_n\boldsymbol{s}^{\top}, \mathbb{X}) = \mathbf{A}\widehat{\mathbf{W}}(\mathbb{Y}, \mathbb{X})\mathbf{A}^{\top}$ for any vector $\boldsymbol{s} \in \mathbb{R}^m$ and any regular matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$, then $\widehat{T}_n(\tau)$ also inherits the shift, rotation, and scale invariance from the location case. If $\widehat{\mathbf{W}}(\mathbb{Y} + \mathbb{X}\mathbf{A}, \mathbb{X}) = \widehat{\mathbf{W}}(\mathbb{Y}, \mathbb{X})$ for any matrix $\mathbf{A} \in \mathbb{R}^{p\times m}$, then $\widehat{T}_n(\tau)$ is invariant with respect to the transformations $\boldsymbol{Y} \mapsto \boldsymbol{Y} + \mathbf{A}^{\top}\boldsymbol{X}$, and if $\widehat{\mathbf{W}}(\mathbb{Y}, \mathbb{X}\mathbf{A}) = \widehat{\mathbf{W}}(\mathbb{Y}, \mathbb{X})$ for any regular matrix $\mathbf{A} \in \mathbb{R}^{p\times p}$,

then $\widehat{T}_n(\tau)$ is invariant with respect to the transformations $\boldsymbol{X} \mapsto \mathbf{A}^\top \boldsymbol{X}$. It is because the centered rank scores do not change with such transformations. Note also that the sample variance-covariance matrix estimator $\widehat{\mathbf{W}}_0(\mathbb{Y}, \mathbb{X})$ exhibits all the equivariance and invariance properties mentioned in this remark.

Acknowledgments

References

[1] ANDREWS, D. W. K. (2003). Tests for parameter instability and structural change with unknown change point: a corrigendum. *Econometrica* **71** 395–397. MR1956864

[2] DUONG, T. (2019). `ks`: Kernel Smoothing R package version 1.11.6.

[3] ESTRELLA, A. (2003). Critical values and p values of Bessel process distributions: computation and application to structural break tests. *Econometric Theory* **19** 1128–1143. MR2015979

[4] GARCIA-PORTUGUÉS, E., PAINDAVEINE, D. and VERDEBOUT, T. (2020). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc.* **115** 1873–1887. MR4189764

[5] GUTENBRUNNER, C. and JUREČKOVÁ, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20** 305–330. MR1150346

[6] HALLIN, M., PAINDAVEINE, D. and VERDEBOUT, T. (2010). Optimal rank-based testing for principal components. *Ann. Statist.* **38** 3245–3299. MR2766852

[7] HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670

[8] HENZE, N., HLÁVKA, Z. and MEINTANIS, S. G. (2014). Testing for spherical symmetry via the empirical characteristic function. *Statistics* **48** 1282–1296. MR3269735

[9] HJORT, N. L. and POLLARD, D. (2011). Asymptotics for minimisers of convex processes. arXiv:1107.3806.

[10] HOLLANDER, M. (1971). A nonparametric test for bivariate symmetry. *Biometrika* **58** 203–212. MR0275596

[11] HUDECOVÁ, Š. and ŠIMAN, M. (2020). Testing symmetry around a subspace. *Stat. Papers*. In press. https://doi.org/10.1007/s00362-020-01201-4.

[12] KALINA, J. (2019). Common multivariate estimators of location and scatter capture the symmetry of the underlying distribution. *Comm. Statist. Simulation Comput.* In press. https://doi.org/10.1080/03610918.2019.1615624.

[13] KOENKER, R. (2005). *Quantile regression*, 3 ed. Cambridge University Press, New York. MR2268657

[14] KOENKER, R. (2015). `quantreg`: Quantile Regression R package.

[15] KOENKER, R. and BASSETT, G. J. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

[16] KOENKER, R. and MACHADO, J. A. F. (1999). Regression quantiles. *J. Amer. Statist. Assoc.* **94** 1296–1310. MR2324109

[17] MAINDONALD, J. H. and BRAUN, W. J. (2019). `DAAG`: Data Analysis and Graphics Data and Functions R package version 1.22.1.

[18] MODARRES, R. (2008). Tests of bivariate exchangeability. *Int. Stat. Rev.* **76** 203–213.

[19] PAINDAVEINE, D. and ŠIMAN, M. (2011). On directional multiple-output quantile regression. *J. Multivariate Anal.* **102** 193–392. MR2739109

[20] RAO, K. S. M. and RAGHUNATH, M. (2012). A simple nonparametric test for bivariate symmetry about a line. *J. Statist. Plann. Inference* **142** 430–444. MR2843047

[21] ROUSSEEUW, P. J. and STRUYF, A. (2002). *A Depth Test for Symmetry* In HUBER-CAROL, C., BALAKRISHNAN, N., NIKULIN, M. S., MESBAH, M. (EDS.) *Goodness-of-Fit Tests and Model Validity* 401–412. Birkhäuser Boston, Boston, MA. MR1901851

[22] SERFLING, R. J. (2006). *Multivariate Symmetry and Asymmetry* In KOTZ S., BALAKRISHNAN N., READ C.B., VIDAKOVIC B. (EDS) *Encyclopedia of Statistical Sciences* 5338–5345. J. Wiley & Sons,New York.

[23] SU, L. (2006). A simple test for multivariate conditional symmetry. *Econom. Lett.* **93** 374–378. MR2281660

[24] R CORE TEAM (2019). `R`: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

[25] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press. MR1652247

[26] WU, W. B. (2005). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33** 1934–1963. MR2166566