Journal of Multivariate Analysis 100 (2009) 622–635

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Distribution-free tests for polynomial regression based on simplicial depth

Robin Wellmann [a,*], Peter Harmand [b], Christine H. Müller [a]

[a] *University of Kassel, Germany*

[b] *University of Oldenburg, Germany*

ARTICLE INFO

Article history:
Received 31 July 2007
Available online 2 July 2008

AMS subject classifications:
primary 62G05
62G10
secondary 62J05
62J12
62G20

Keywords:
Distribution-free tests
Simplicial depth
Regression depth
Polynomial regression
Degenerated U-statistic
Spectral decomposition
Outlier robustness

ABSTRACT

A general approach for developing distribution-free tests for general linear models based on simplicial depth is presented. In most relevant cases, the test statistic is a degenerated U-statistic so that the spectral decomposition of the conditional expectation of the kernel function is needed to derive the asymptotic distribution. A general formula for this conditional expectation is derived. Then it is shown how this general formula can be specified for polynomial regression. Based on the specified form, the spectral decomposition and thus the asymptotic distribution is derived for polynomial regression of arbitrary degree. The power of the new test is compared via simulation with other tests. An application on cubic regression demonstrates the applicability of the new tests and in particular their outlier robustness.

1. Introduction

Simplicial depth for multivariate location was introduced by Liu [7,8]. It is based on the half-space depth of Tukey [16]. Both depth notions lead to a generalization of the median for multivariate data which is equivariant with respect to affine transformations. Moreover the concept of depth is useful to generalize ranks.

The simplicial depth has the advantage that it is a U-statistic so that in principle the asymptotic distribution is known. However, it is not easy to derive the asymptotic distribution. Arcones et al. [1] derived the asymptotic normality of the maximum simplicial depth estimator of Liu [7,8] via the convergence of the whole U-process. The convergence of the U-process was also shown by Dumbgen [5]. However the asymptotic normal distribution has a covariance matrix which depends on the underlying distribution. Hence this result cannot be used to derive distribution-free tests. Liu [9] and Liu and Singh [10] proposed distribution-free multivariate rank tests which generalize the Wilcoxon's rank sum test for two samples. While the asymptotic normality is derived for several depth notions for distributions on $\mathbb{R}^1$, it is shown only for the Mahalanobis depth for distributions on $\mathbb{R}^k$, $k > 1$. Hence it is unclear how to generalize the approach of Liu and Singh to other situations.

Several other depth concepts were introduced since the work of Tukey [16]. See for example the book of Mosler [13] and the references therein. Multivariate depth concepts were transferred to regression by Rousseeuw and Hubert [15], to logistic

* Corresponding address: Department of Mathematics, University of Kassel, D-34109 Kassel, Germany.
E-mail address: wellmann@mathematik.uni-kassel.de (R. Wellmann).

0047-259X/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2008.06.009

regression by Christmann and Rousseeuw [3] and to the Michaelis–Menten model by Van Aelst et al. [17]. The depth concept for regression is based on the notion of nonfit introduced by Rousseeuw and Hubert [15]. Thereby, a regression parameter θ is called a nonfit, if there is another parameter θ' which provides for all observations z_n smaller squared residuals $r(z_n, \theta')^2$. The depth of a regression parameter θ is then given by the minimum number of observations which must be removed so that θ becomes a nonfit.

More general concepts of depth were introduced and discussed by Zuo and Serfling [22,23] and Mizera [11]. Mizera [11] in particular generalized the regression depth of Rousseeuw and Hubert [15] by basing the nonfit on general quality functions instead of squared residuals. Using these quality functions, he introduced the "global depth", the "tangent depth" and the "local depth" and gave a sufficient condition for their equality. This approach makes it possible to define the depth of a parameter value with respect to given observations in various statistical models via general quality functions. Appropriate quality functions are in particular likelihood functions as studied by Mizera and Müller [12] for the location-scale model and by Müller [14] for generalized linear models.

As for multivariate location, there exist only few results concerning tests based on regression depth and its generalizations. Bai and He [2] derived the asymptotic distribution of the maximum depth estimator for regression so that tests could be based on this. However, this asymptotic distribution is given implicitly so that it is not convenient for inference. Van Aelst et al. [17] derived an exact test based on the regression depth of Rousseeuw and Hubert [15] but did it only for linear regression. Müller [14] derived tests by using the simplicial regression depth which generalizes Liu's [7,8] simplicial depth to regression.

In general, the simplicial depth d_S for a q dimensional parameter θ which is based on a given depth notion d is defined by

$$d_S(\theta, z) = \binom{N}{q+1}^{-1} \sum_{1 \le n_1 < n_2 < \cdots < n_{q+1} \le N} \mathbb{I}\{d(\theta, (z_{n_1}, \ldots, z_{n_{q+1}})) > 0\},$$

where $z = (z_1, \ldots, z_N)$ is the sample and $\mathbb{I}$ denotes the indicator function. This is a U-statistic with symmetrical kernel function $\psi_\theta \in \mathbb{L}_2(\bigotimes_{n=1}^{q+1} P_\theta^{Z_1})$, defined as

$$\psi_\theta(z_{n_1}, \ldots, z_{n_{q+1}}) := \mathbb{I}\{d(\theta, (z_{n_1}, \ldots, z_{n_{q+1}})) > 0\}.$$

That is, ψ_θ equals one if and only if the depth $d(\theta, (z_{n_1}, \ldots, z_{n_{q+1}}))$ of θ in $z_{n_1}, \ldots, z_{n_{q+1}}$ is greater than zero.

Müller [14] proposed basing the test statistic directly on the general simplicial depth d_S. For testing a hypothesis of the form $H_0 : \theta \in \Theta_0$, where Θ_0 is an arbitrary subset of the parameter space, the test statistic is defined as $T(z_1, \ldots, z_N) := \sup_{\theta \in \Theta_0} T_\theta(z_1, \ldots, z_N)$. Thereby $T_\theta(z_1, \ldots, z_N)$ is defined as

$$T_\theta(z_1, \ldots, z_N) := \frac{\sqrt{N}(d_S(\theta, (z_1, \ldots, z_N)) - \mu_\theta)}{(q+1)\sigma_\theta},$$

if $d_S(\theta, z)$ is not a degenerated U-statistic, i.e. $\psi_\theta^1(z_1) := E(\psi_\theta(Z_1, \ldots, Z_{q+1})|Z_1 = z_1)$ depends on z_1, and

$$T_\theta(z_1, \ldots, z_N) := N(d_S(\theta, (z_1, \ldots, z_N)) - \mu_\theta), \tag{1}$$

if $d_S(\theta, z)$ is a degenerated U-statistic. In several cases the asymptotic distribution of $T_\theta(Z_1, \ldots, Z_N)$ does not depend on θ, so that the null hypothesis H_0 is rejected if $T(z_1, \ldots, z_N)$ is less than the α-quantile of this asymptotic distribution. Thereby the quantities μ_θ and σ_θ are defined as $\mu_\theta = E(\psi_\theta(Z_1, \ldots, Z_{q+1}))$ and $\sigma_\theta^2 = \mathrm{Var}(\psi_\theta^1(Z_1))$.

Unfortunately, the simplicial depth d_S is a degenerated U-statistic in the most interesting case, where the true regression function is in the center of the data, which means that the median of the residuals is zero. Whereas nondegenerated U-statistics are asymptotically normal distributed, simple asymptotic results are not possible for degenerated U-statistics. In the degenerated case, the asymptotic distributions can be derived by using the second component of the Hoeffding decomposition. We have namely the following result (see e.g. [6], p. 79, 80, 90, [21], p. 650). If the reduced normalized kernel function

$$\psi_\theta^2(z_1, z_2) := E(\psi_\theta(Z_1, \ldots, Z_{q+1}) - \mu_\theta | Z_1 = z_1, Z_2 = z_2) \tag{2}$$

is $\mathbb{L}_2$-integrable, it has a spectral decomposition of the form

$$\psi_\theta^2(z_1, z_2) = \sum_{l=1}^{\infty} \lambda_l \varphi_l(z_1) \varphi_l(z_2), \tag{3}$$

where the functions φ_l are $\mathbb{L}_2$-integrable, normalized, and orthogonal. Then the asymptotic distribution of the simplicial depth is given by

$$N(d_S(\theta, (Z_1, \ldots, Z_N)) - \mu_\theta) \xrightarrow{\mathcal{L}} \binom{q+1}{2} \sum_{l=1}^{\infty} \lambda_l (U_l^2 - 1), \tag{4}$$

where $U_l \sim \mathcal{N}(0, 1)$ and $U_1, U_2, \ldots$ are independent. In the general case, it could happen that the eigenvalues λ_l depend on the underlying parameter θ. However, Müller [14] could show that this is not the case for polynomial regression in generalized linear models so that the asymptotic distribution does not depend on the regression parameter.

However, Müller [14] was only able to find the spectral decompositions for linear and quadratic regression in generalized linear models. These spectral decompositions were found by solving differential equations. In this paper we derive the spectral decomposition (3) for polynomial regression of arbitrary degree by a completely new approach.

For this approach, we use in Section 2 quality functions for defining the depth d used in the simplicial depth d_S. The simplicial depth is based on the tangent depth for these quality functions. However, this simplicial depth attains rather high values in subspaces of the parameter space, since it does not provide convex depth contours as all simplicial depth notions do not. This is in particular a disadvantage in testing if the aim is to reject the null hypothesis. To avoid this disadvantage, we introduce in Section 2 a harmonized depth and use it as the kernel function of the simplicial depth. This approach leads also to a method to calculate the maximum simplicial depth under the null hypothesis. While in [14] only null hypotheses could be rejected for which the null hypothesis is a point or a line within the parameter space, we are now able to treat hypotheses about arbitrary subspaces and polyhedrals, as Wellmann et al. [18] showed.

In Section 3, we derive a general formula for the conditional expectation (2) for the simplicial depth for generalized linear models introduced in Section 2 and we show that the asymptotic distribution can be obtained by calculating the spectral decomposition of a function $\mathcal{K}$, which only depends on the probability law of the vector product of regressor variables. This means in particular that the asymptotic distribution of the test statistic (1) does not depend on the unknown regression parameter. The function $\mathcal{K}$ is applied to the harmonized form of the simplicial regression depth but the proofs hold also for the unmodified form.

The general formula for $\mathcal{K}$ is specified for polynomial regression of arbitrary degree in Section 4. Based on the specified formula, the spectral decomposition is derived. The spectral decomposition is found by a Fourier series representation of a related function of $\mathbb{L}_2[-1, 1]$ which is used to derive the required representation of $\mathcal{K}$. We think that this approach can be used to find the spectral decomposition of other simplicial depth functions. In particular, Wellmann and Müller [19] derived the asymptotic distribution of the simplicial regression depth for different models of multiple regression.

In Section 5 the power of the simplicial depth test is compared with other tests via simulation. It turns out that the power of the simplicial depth test is better than the power of the tangent depth test of Van Aelst et al. [17], but there exist robust tests for linear regression which have a slightly better power than the simplicial depth test. An advantage of the simplicial depth test is that general forms of hypotheses can be tested, as shown in [18]. The null hypothesis could be an arbitrary polyhedron within the parameter space. This is also demonstrated in Section 6 where an application on tests in a cubic regression model is given. This example also shows that the new tests possess a higher outlier robustness than competing tests. All proofs are given in Section 7.

2. Simplicial depth for generalized linear models

We assume that the random vectors $Z_1, \ldots, Z_N$ are independent and identically distributed throughout the paper. The random vectors $Z_n = (U_n, X_n)$ have values in $\mathcal{Z} \subset \mathbb{R}^{1+q}$, where U_n is a real-valued random variable and the regressor X_n is a random vector. For a given transformation $h : \mathbb{R} \to \mathbb{R}$ we write $Y_n = h(U_n)$. The random error E_n is given by

$$E_n = Y_n - X_n^T\theta,$$

where $\theta \in \Theta = \mathbb{R}^q$. Random variables are denoted by capital letters and realizations by small letters. The value $s_n(\theta) := \operatorname{sign}_\theta(z_n) := \operatorname{sign}(y_n - x_n^T\theta)$ is the sign of the residual of the nth (transformed) observation. The family $\mathcal{P} = \{P_\theta^{(Z_1,\ldots,Z_N)} : \theta \in \Theta\}$ of probability measures with $\Theta = \mathbb{R}^q$ may be unknown, but for the purpose of deriving tests, we will assume that the following assumptions hold:

$$\begin{aligned}
&\bullet\quad P_\theta(S_1(\theta) = 1|X_1) \equiv \frac{1}{2} \text{ a.s.,}\\
&\bullet\quad P_\theta(S_1(\theta) = 0|X_1) \equiv 0 \text{ a.s.,}\quad \text{and}\\
&\bullet\quad P_\theta(X_1, \ldots, X_q \text{ are linearly dependent}) = 0.
\end{aligned}\tag{5}$$

Thus, the random errors only have to satisfy $P_\theta(E_n > 0|X_n) \equiv P_\theta(E_n < 0|X_n) \equiv \frac{1}{2}$. In particular, the errors may be heteroscedastic. Our results are not affected by the concrete form of the error distribution. The last two conditions of (5) are usually satisfied for continuous distributions. The first condition of (5) holds in particular with $h(U_n) = U_n$ for general linear models with symmetric error. For generalized linear models it can be achieved by using an appropriate transformation h. See for example [14] for the exponential distribution, where h should be $h(U_n) = \log(\frac{U_n}{\log(2)})$.

For this model, the following quality functions can be used:

Definition 1 (*Quality Functions for Generalized Linear Models*). Take φ to be a function with continuous derivatives, which has its maximum and its sole critical point in 0. Then the function

$$\mathcal{G}_{z_n} : \Theta \to \mathbb{R} \quad \text{with } \mathcal{G}_{z_n}(\theta) := \varphi(y_n - x_n^T\theta)$$

is said to be a quality function for generalized linear models.

Although quality functions are needed to define the tangent depth or the global depth of Mizera [11], the resulting depth functions do not depend on the choice of φ, so that we may restrict ourselves to the simplest case $\varphi(x) = -x^2$.

The general form is only needed to cover the likelihood case: Often, one would like to choose the likelihood functions $f_\theta^{Z_n}(z_n)$ to be the quality functions. However, they should be used only, if $P_\theta(E_n > 0|X_n) = \frac{1}{2}$ is satisfied, because otherwise the true regression function is not in the center of the data and thus the estimator is biased. A forthcoming paper will deal with the problems which appear if $P_\theta(E_n > 0|X_n) \equiv \frac{1}{2}$ is not satisfied.

Definition 2 (*Tangent Depth*). According to Mizera [11], we define the tangent depth of $\theta \in \Theta$ with respect to given observations $z_1, \ldots, z_N \in \mathcal{Z}$ to be

$$d_T(\theta, z) = \min_{u \neq 0} \#\{n : u^T \nabla \mathcal{G}_{z_n}(\theta) \geq 0\},$$

where $\mathcal{G}_{z_1}, \ldots, \mathcal{G}_{z_N}$ are quality functions for generalized linear models, $z := (z_1, \ldots, z_N)$ and $\nabla \mathcal{G}_{z_n}(\theta)$ denotes the vector of partial derivatives of $\mathcal{G}_{z_n}$ in θ.

As shown in [11], this depth notion counts the number of observations that needs to be removed such that there is a "better" parameter for all remaining observations. It is easy to see, that the tangent depth does not depend on the choice of φ. Furthermore, for all $\theta \in \Theta$ and for given observations $z_1, \ldots, z_N \in \mathcal{Z}$ we have:

$$d_T(\theta, z) = \min_{u \neq 0} \#\{n : s_n(\theta) u^T x_n \geq 0\}. \tag{6}$$

As in [15] it can be shown, that the parameter space $\Theta = \mathbb{R}^q$ is divided up into domains with constant depth by the hyperplanes

$$H_n = \{\theta \in \mathbb{R}^q : s_n(\theta) = 0\}, \quad n = 1, \ldots, N.$$

For given observations let $\mathrm{Dom}(z)$ be the set of all those domains. We define $\bar{d}_T(G, z) := d_T(\theta, z)$ for $G \in \mathrm{Dom}(z)$ and $\theta \in G$.

We will define the simplicial depth to be a U-statistic. If we would take the tangent depth to be the kernel function of the U-statistic, then the simplicial regression depth attains rather high values in subspaces of the parameter space, namely in $\mathrm{Border}(z) := \cup_{n=1}^N H_n$. This is in particular a disadvantage if the aim is to reject the null hypothesis. To avoid this disadvantage, we introduce a harmonized depth.

Definition 3 (*Harmonized Depth*). The harmonized depth of $\theta \in \Theta$ with respect to the observations $z_1, \ldots, z_N \in \mathcal{Z}$ is defined to be

$$\psi_\theta(z) = \min_{G \in \mathrm{Dom}(z), \theta \in \bar{G}} \bar{d}_T(G, z),$$

where $\bar{G}$ is the closure of G.

Definition 4 (*Simplicial Depth*). The simplicial depth is given by

$$d_S(\theta, z) = \binom{N}{q+1}^{-1} \sum_{1 \leq n_1 < n_2 < \cdots < n_{q+1} \leq N} \psi_\theta(z_{n_1}, \ldots, z_{n_{q+1}}).$$

This depth, which transfers the simplicial depth of Liu to regression models, is also called a simplicial depth because it counts the fraction of simplices that are bounded by $q+1$ hyperplanes from $H_1, \ldots, H_N$ and contains θ as an interior point. Algorithms for calculating the simplicial depth are based on this view as well and are given in [18]. The proposed tests are based on the asymptotic distribution of this depth notion.

3. The asymptotic distribution of the simplicial depth

The definition of tangent depth shows, that the depth of a parameter is the half-space depth of 0 with respect to the gradients of the quality functions. Thereby, the half-space depth of 0 with respect to given vectors $r_1, \ldots, r_N \in \mathbb{R}^q$ is defined as

$$d_H(0, r) := \min_{u \neq 0} \#\{n : u^T r_n \geq 0\},$$

where $r = (r_1, \ldots, r_{q+1})$ (see [16]). The next lemma is needed to derive the conditional expectations of the kernel function, which depends only on $q+1$ observations, but it can also be used to calculate the simplicial depth of a given parameter.

Lemma 1. *Let $r_1, \ldots, r_{q+1} \in \mathbb{R}^q$ be in general position. Then $d_H(0, r) \in \{0, 1\}$ and the following statements are equivalent:*

(i) $d_H(0, r) = 0$,
(ii) $r_1 \notin \mathbb{R}_{\leq 0} r_2 + \cdots + \mathbb{R}_{\leq 0} r_{q+1}$.

Proofs are given in Section 7. The next Proposition shows, that

$$\psi_\theta^1(z_1) := E(\psi_\theta(Z_1, \dots, Z_{q+1})|Z_1 = z_1)$$

does not depend on z_1, so that the simplicial depth is a degenerated U-statistic and has asymptotically the distribution of an infinite linear combination of χ^2-distributed random variables (see e.g. [6], p. 79, 80, 90, [21], p. 650). This distribution depends only on the conditional expectation

$$\psi_\theta^2(z_1, z_2) := E(\psi_\theta(Z_1, \dots, Z_{q+1})|Z_1 = z_1, Z_2 = z_2) - E(\psi_\theta(Z_1, \dots, Z_{q+1})).$$

Proposition 1. *Suppose that the assumptions in (5) are satisfied. Let $\theta \in \Theta$ and let $z_1, z_2 \in \mathcal{Z}$, such that x_1, x_2 are linearly independent and $s_1(\theta), s_2(\theta) \in \{-1, 1\}$. Then*

$$\psi_\theta^1(z_1) = \frac{1}{2^q}$$

and

$$\psi_\theta^2(z_1, z_2) = \frac{s_1(\theta)s_2(\theta)}{2^{q-1}} \left(P_\theta(x_1^T W x_2^T W < 0) - \frac{1}{2} \right),$$

where $W := X_3 \times \dots \times X_{q+1}$ is the vector product of $X_3, \dots, X_{q+1}$.

With this proposition, we obtain a main result: We get the asymptotic distribution of the simplicial depth for generalized linear models which satisfy assumptions (5) by calculating the spectral decomposition of the kernel $\mathcal{K}$, defined by

$$\mathcal{K}(x_1, x_2) := P_\theta(x_1^T W x_2^T W < 0) - \frac{1}{2}, \quad \text{for } x_1, x_2 \in \mathbb{R}^q. \tag{7}$$

Note that $x_i^\top W = \det(x_i, X_3, \dots, X_{q+1})$ for $i = 1, 2$. The spectral decomposition is a representation

$$\mathcal{K}(x_1, x_2) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(x_1)\varphi_j(x_2) \quad \text{in } \mathbb{L}_2\left(P^{X_1} \otimes P^{X_1}\right),$$

where $(\varphi_j)_{j=1}^\infty$ is an orthonormal system (ONS) in $\mathbb{L}_2\left(P^{X_1}\right)$ and $\lambda_1, \lambda_2, \dots \in \mathbb{R}$. The functions $(\varphi_j)_{j=1}^\infty$ are eigenfunctions and the values $\lambda_1, \lambda_2, \dots$ are the corresponding eigenvalues of the related integral operator $T_{\mathcal{K}}$, defined by

$$T_{\mathcal{K}} : \mathbb{L}_2(P^{X_1}) \to \mathbb{L}_2(P^{X_1}) \quad \text{with } T_{\mathcal{K}} f(s) = \int \mathcal{K}(s, t) f(t) \mathrm{d}P^{X_1}(t).$$

Now we show, how the asymptotic distribution can be obtained from the spectral decomposition of $\mathcal{K}$. The system $(\psi_j)_{j=1}^\infty$, defined by $\psi_j(z) := \mathrm{sign}_\theta(z)\varphi_j(v(z))$ for $z \in \mathcal{Z}$ is an ONS in $\mathbb{L}_2(P^{Z_1})$ and for ψ_θ^2 we have the representation

$$\begin{aligned}
\psi_\theta^2(z_1, z_2) &= \frac{\mathrm{sign}_\theta(z_1)\mathrm{sign}_\theta(z_2)}{2^{q-1}} \mathcal{K}(v(z_1), v(z_2)) \\
&= \sum_{j=1}^{\infty} \frac{1}{2^{q-1}} \lambda_j \mathrm{sign}_\theta(z_1)\varphi_j(v(z_1))\mathrm{sign}_\theta(z_2)\varphi_j(v(z_2)) \\
&= \sum_{j=1}^{\infty} \frac{1}{2^{q-1}} \lambda_j \psi_j(z_1)\psi_j(z_2).
\end{aligned}$$

Hence, it follows by (4), that

$$N\left(d_S(\theta, (Z_1, \dots, Z_N)) - \frac{1}{2^q}\right) \xrightarrow{\mathcal{L}} \sum_{l=1}^{\infty} \frac{(q+1)!}{(q-1)!2^q} \lambda_l \left(U_l^2 - 1\right), \tag{8}$$

where $U_1, U_2, \dots$ are i.i.d. random variables with $U_1 \sim \mathcal{N}(0, 1)$. Furthermore, this derivation shows, that the asymptotic distribution does not depend on the underlying parameter θ, if the distribution of W does not depend on it. However, in general it depends on the underlying distribution of the regressors X_n. In the next section, this general result is applied to polynomial regression.

4. Polynomial regression

In the model for polynomial regression of degree $r = q - 1$ we can write $Z_n = (U_n, x(T_n))$ with a real-valued random variable T_n, where $Y_n = h(U_n)$ is the dependent variable and $X_n := x(T_n) := (1, T_n, \ldots, T_n^r)^T$ is the regressor. The unknown parameter is $\theta = (\theta_1, \ldots, \theta_q)^\top \in \mathbb{R}^q$, so that

$$Y_n = \theta_1 + \theta_2 T_n + \cdots + \theta_q T_n^{q-1} + E_n.$$

Suppose that the assumptions in (5) are satisfied. Because of the independence of $T_1, \ldots, T_N$, the third assumption in (5) is equivalent to $P_\theta(T_1 = t) = 0$ for all $t \in \mathbb{R}$.

In this section, we derive the asymptotic distribution of the simplicial depth by calculating the spectral decomposition of the kernel $\mathcal{K}$, given in (7). While Müller [14] derived it only for $r = 1$ and $r = 2$ in another way, we have now the asymptotic distribution for polynomial regression of arbitrary degree. At first, we give a simple representation of the kernel $\mathcal{K}$, which is obtained from (7) via the formula for Vandermonde determinants (see the Section 7).

Proposition 2. *For all $t_1, t_2 \in \mathbb{R}$ we have*

$$\mathcal{K}(x(t_1), x(t_2)) = -2^{r-1}\left(\frac{1}{2} - |F^{T_1}(t_1) - F^{T_1}(t_2)|\right)^r,$$

where F^{T_1} is the distribution function of T_1.

Müller derived the same formula for the reduced normalized kernel function ψ_θ^2 (see Proposition 2 in [14]). Our proof is based not on ψ_θ^2, but on $\mathcal{K}$. This makes the proof much shorter. It remains to derive the spectral decomposition of $\mathcal{K}$, which we obtain in the next proposition via a Fourier series representation of $(\frac{1}{2} - |z|)^r$ in $\mathbb{L}_2[-1, 1]$.

Proposition 3. *The spectral decomposition of $\left(\frac{1}{2} - |s - t|\right)^r$ in $\mathbb{L}_2[0, 1]^2$ is given by*

$$\left(\frac{1}{2} - |s - t|\right)^r = \gamma_0^{(r)} \cdot 1 + \sum_{l=1}^{\infty} \gamma_l^{(r)} \cdot 2 \cdot [\cos(k\pi s)\cos(k\pi t) + \sin(k\pi s)\sin(k\pi t)]$$

where for r odd

$$\gamma_l^{(r)} = \begin{cases} 0, & \text{if } l \text{ is even}, \\ -\displaystyle\sum_{\substack{k\in\{1,\ldots,r\} \\ k \text{ odd}}} \frac{r!}{2^{r-k-1}(r-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}}, & \text{if } l \text{ is odd}, \end{cases}$$

and for r even

$$\gamma_l^{(r)} = \begin{cases} \dfrac{1}{(r+1)2^r}, & \text{if } l = 0, \\ -\displaystyle\sum_{\substack{k\in\{1,\ldots,r\} \\ k \text{ odd}}} \frac{r!}{2^{r-k-1}(r-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}}, & \text{if } l \text{ is even and } l > 0, \\ 0, & \text{if } l \text{ is odd}. \end{cases}$$

Let $(\psi_j)_{j\in J}$ be the ONS in $\mathbb{L}_2[0, 1]$, given in the proof of Proposition 3, such that $(\gamma_j^{(r)})_{j\in J}$ are the eigenvalues, related to $K(s, t) = (\frac{1}{2} - |s - t|)^r$. Then the system $(\varphi_j)_{j\in J}$, defined by $\varphi_j := \psi_j \circ F^{T_1} \circ x^{-1}$ is an ONS in $\mathbb{L}(P^{X_1})$ and we have the representation

$$\begin{aligned}\mathcal{K}(x_1, x_2) &= \mathcal{K}(x(x^{-1}(x_1)), x(x^{-1}(x_2))) \\ &= -2^{r-1}\left(\frac{1}{2} - |F^{T_1}(x^{-1}(x_1)) - F^{T_1}(x^{-1}(x_2))|\right)^r \\ &= \sum_{j\in J}(-2^{r-1}\gamma_j^{(r)})\varphi_j(x_1)\varphi_j(x_2).\end{aligned}$$

Hence, the next theorem holds:

Theorem 1. *If $P(Y_n - x(T_n)^\top\theta \geq 0|T_n) = \frac{1}{2}$ and T_n has a continuous distribution, then the asymptotic distribution of the simplicial likelihood depth $d_S(\theta, (Z_1, \ldots, Z_N))$ for polynomial regression is given by*

$$N\left(d_S(\theta, (Z_1, \ldots, Z_N)) - \frac{1}{2^{r+1}}\right) \xrightarrow{\mathcal{L}} \sum_{l=0}^{\infty} \lambda_{2l+1}(V_l^2 + W_l^2 - 2)$$

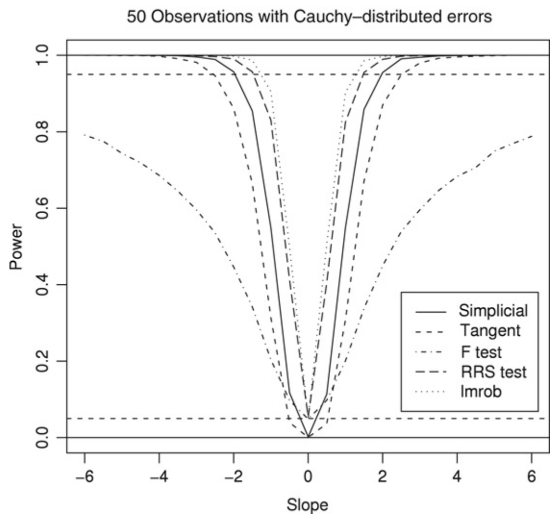

Fig. 1. Cauchy distributed errors.

for r even and

$$N\left(d_S(\theta,(Z_1,\ldots,Z_N))-\frac{1}{2^{r+1}}\right)\xrightarrow{\mathcal{L}}\lambda_0(U^2-1)+\sum_{l=1}^{\infty}\lambda_{2l}(V_l^2+W_l^2-2)$$

for r odd, where $U,\ V_0, W_0, V_1, W_1, \ldots$ are independent random variables with standard normal distribution and

$$\lambda_0 = -\frac{r+2}{2^{r+2}},$$
$$\lambda_l = \sum_{\substack{k\in\{1,\ldots,r\}\\ k\ odd}} \frac{(r+2)!}{2^{r-k+1}(r-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}} \quad \text{for } l \in \mathbb{N}.$$

The calculation of the test statistic and the critical values for any hypothesis of the form $H_0 : \theta \in \Theta_0$ where Θ_0 is a subspace of the parameter space or a polyhedron is described in [18]. There, a table of the critical values is also given. Although this test is an asymptotic test, it controls the alpha level also for the finite case. This is shown by the simulation study of the next section. In this section also the power is compared with other outlier robust tests.

5. Power comparison

In a simulation study, we compared the power of different tests for linear regression. We compared the simplicial depth test with the tangent depth test according to Van Aelst et al. [17] and Daniels [4], the F test, the regression rank-score test (RRS test, function `rrs.test()` from R package `quantreg`) and a test which is based on MM regression estimators (function `lmrob()` from R package `robustbase`). The functions `rrs.test()` and `lmrob()` are used with the default parameter values. All tests are performed to the level $\alpha = 0.05$. The power curves are obtained by simulation with 5000 repetitions.

In the first example, we tested the null hypothesis that the true regression line is horizontal. The regressors are realizations of independent normal distributed random variables with mean 0 and standard deviation 1. The errors are independent Cauchy distributed with location parameter 0 and scale parameter 1. We choose Cauchy distributed errors in order to simulate outliers. Fig. 1 shows the power for different slopes. The simplicial depth test has a better power than the tangent depth test and a slightly worse power than the RRS test and the lmrob test. The F test nearly keeps the level, but has poor power for such observations. It can also be seen that the true levels of the simplicial depth test and the tangent depth test are smaller than 0.05. This is because for both tests, the depth is maximized over all parameters from the null hypothesis. The null hypothesis is rejected in nearly all cases with the simplicial depth test, if the true slope is larger than 2.

In the second example, we also tested the null hypothesis that the true regression line is horizontal, but we assumed heteroscedastic errors. The regressors are independent normal distributed with mean 0 and standard deviation 2. The errors are independent Cauchy distributed with location parameter 0, but the scale parameter depends on the regressor x_n. Given x_n, the scale parameter of the error is equal to $2 + \exp(x_n)$, so that the variation increases from left to right. Fig. 2 shows, that also in this example the simplicial depth test has a better power than the tangent depth test and controls the level. Moreover, the RRS test has poorer power for small slopes and does not keep the level since the minimum is shifted to the left. The lmrob test had the best power in all comparisons, but did not converge for some data sets and might also have a problem with the level.

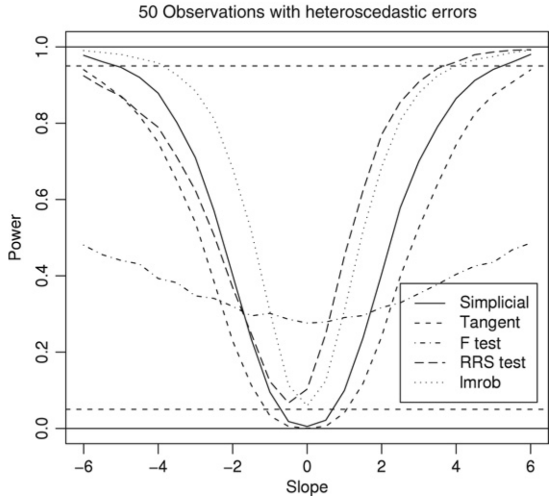

Fig. 2. Heteroscedastic errors.

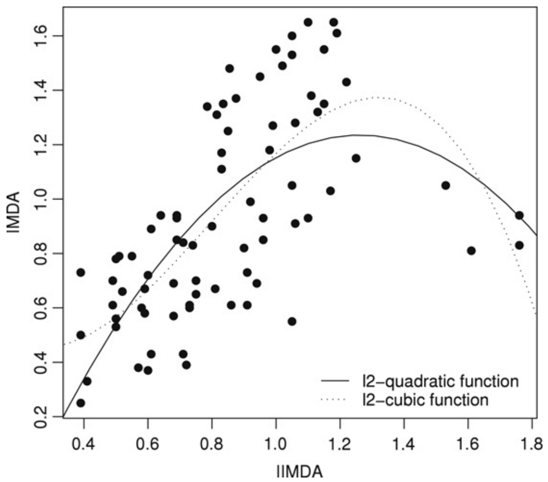

Fig. 3. Least squares quadratic and cubic function.

6. Application: Test about quadratic function against cubic function

The concentration of malondialdehyde (MDA) for 78 women twice after childbirth (IMDA and IIMDA) at two time points was measured, to find a relation between the levels of IMDA and IIMDA. MDA is a metabolite of lipid peroxides detectable in plasma. It was measured as an indicator of lipid peroxidation and oxidation stress of women postpartum (after childbirth). The data came from the Clinic of Gynaecology, Faculty Hospital with Policlinic, Bratislava-Ružinov (Slovakia).

We choose IMDA as the dependent variable. The data shown in Fig. 3 suggest an almost linear relationship with 4 outliers on the right-hand side. Thus an outlier robust test should not reject the quadratic model against the cubic model and it should not reject the linear model against the quadratic model. But it should reject the null hypothesis that the true regression line is constant in the model for linear regression.

We tested these hypotheses with the simplicial depth test, the RRS test, the lmrob test and the F test, although normality of the residuals with respect to the ordinary least squares estimation was rejected (p-value < 0.001). These tests are also used in the simulation study of Section 5. The results are given in Table 1, where '0' denotes a decision for H_0, '1' denotes a decision for H_1, and '—' indicates that this hypothesis cannot be tested.

It turned out that only the simplicial depth test provides the expected results. The results of the test linear against quadratic may be not surprising since the deepest quadratic function looks rather linear (see Fig. 4), whereas the least squares estimate is attracted by the outliers (see Fig. 3).

In Table 1 it is also demonstrated that the simplicial depth test can be used to test hypotheses which cannot be tested with the other tests. We used the fast algorithm given in [18] for the determination of $d_S(\theta, z)$, so that not all $\binom{N}{q+1}$ combinations

Table 1
Decisions of different tests

Model	Hypothesis	Simplicial depth test	RRS test	lmrob test	F test
Cubic	$\theta_4 = 0$	0	1	1	1
Quadratic	$\theta_3 = 0$	0	1	1	1
Linear	$\theta_2 = 0$	1	1	1	1
Quadratic	$\lvert\theta_3\rvert > 10$	1	–	–	–
Linear	$\theta_2 \notin [0, 3]$	1	–	–	–

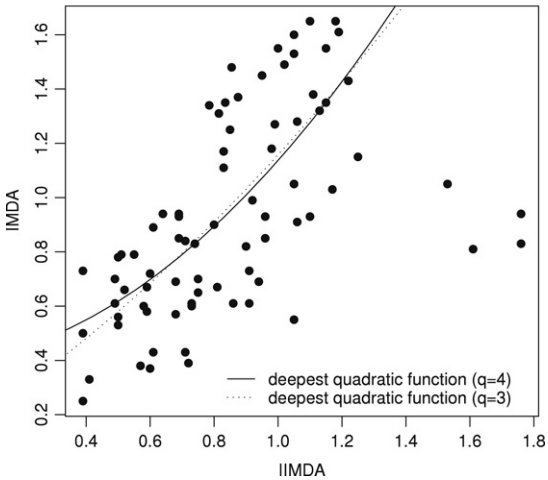

Fig. 4. Deepest quadratic functions for $q = 3$ and $q = 4$.

need to be calculated. Nevertheless, the method is computer intensive if the dimension of the null hypothesis is large. In Wellmann et al. it is also shown how the test statistic can be calculated if Θ_0 is an arbitrary polyhedron.

7. Proofs

For more details of the proofs see also [20].

Proof of Lemma 1. Since $r_1, \ldots, r_q$ are linearly independent, they belong to a hyperplane H with $0 \notin H$.
There are a $\gamma < 0$ and a $u \in \mathbb{R}^q$, such that $H = \{v \in \mathbb{R}^q : v^T u = \gamma\}$.
Since $r_1, \ldots, r_q$ do not belong to the half-space $\{v \in \mathbb{R}^q : v^T u \geq 0\}$, we have $d_H(0, r) \leq 1$.
It remains to show the equivalence.

(ii) $\Rightarrow$ (i): For any $j = 1, \ldots, q+1$ let H_j be the hyperplane that contains the points $(r_i)_{i \in \{1,\ldots,q+1\} \setminus \{j\}}$.

Step 1: There is a $j \in \{1, \ldots, q+1\}$ such that 0 and r_j are on different sides of H_j.

Proof. Since $r_2, \ldots, r_{q+1}$ is a basis of $\mathbb{R}^q$, there exists $\gamma_2, \ldots, \gamma_{q+1} \in \mathbb{R}$ such that $r_1 = \gamma_2 r_2 + \cdots + \gamma_{q+1} r_{q+1}$. Since $r_1 \notin \mathbb{R}_{\leq 0} r_2 + \cdots + \mathbb{R}_{\leq 0} r_{q+1}$ we may assume that $\gamma_2 > 0$. We prove that r_1 and 0 are on different sides of H_1, if r_2 and 0 are on the same side of H_2. Hence, we have:

(a) $r_1 = \gamma_2 r_2 + \cdots + \gamma_{q+1} r_{q+1}$ with $\gamma_2 > 0$.
(b) There are $\alpha > 0$ and $\beta_3, \ldots, \beta_{q+1} \in \mathbb{R}$ such that
$r_2 = r_1 - \alpha r_1 + \sum_{j=3}^{q+1} \beta_j (r_j - r_1) \in \mathbb{R}_{<0} r_1 + H_2$.

From these equations we obtain two different representations of r_2:

$$r_2 = \left(1 - \alpha - \sum_{j=3}^{q+1} \beta_j\right) r_1 + \beta_3 r_3 + \cdots + \beta_{q+1} r_{q+1},$$

$$r_2 = \frac{1}{\gamma_2} r_1 + \frac{-\gamma_3}{\gamma_2} r_3 + \cdots + \frac{-\gamma_{q+1}}{\gamma_2} r_{q+1}.$$

Comparing the coefficients leads to

$$0 < \frac{1}{\gamma_2} = 1 - \alpha - \sum_{j=3}^{q+1} \beta_j \quad \text{and} \quad \frac{-\gamma_k}{\gamma_2} = \beta_k \quad \text{for } k = 3, \ldots, q+1.$$

It follows that

$$\begin{aligned}\gamma_2+\cdots+\gamma_{q+1} &= \gamma_2-\gamma_2\beta_j-\cdots-\gamma_2\beta_{q+1}\\ &= \gamma_2\left(1-\sum_{j=3}^{q+1}\beta_j\right)\\ &= \frac{1-\sum_{j=3}^{q+1}\beta_j}{1-\alpha-\sum_{j=3}^{q+1}\beta_j} > 1.\end{aligned}$$

With (a) we have

$$r_1 = (\gamma_2+\gamma_3+\cdots+\gamma_{q+1})r_2+\gamma_3(r_3-r_2)\cdots+\gamma_{q+1}(r_{q+1}-r_2).$$

Thus there is a $\lambda\in(0,1)$ with: $\lambda r_1 \in r_2+\sum_{j=3}^{q+1}\mathbb{R}(r_j-r_2)=H_1$.
Hence, r_1 and 0 are on different sides of H_1. This finishes the proof of Step 1.

Step 2: Main proof. The vectors r_j and 0 are on different sides of this affine hyperplane H_j. Let $v\in H_j$. All vectors $r_1,\ldots,r_{q+1}$ are in the open half-space $\mathbb{R}_{>0}v+(H_j-v)$. The half-space $\mathbb{R}^q\setminus(\mathbb{R}_{>0}v+(H_j-v))$ does not contain the vectors $r_1,\ldots,r_{q+1}$. Thus, $d_H(0,r)=0$.

(i) $\Rightarrow$ (ii)

The vectors $r_1,\ldots,r_{q+1}$ belong to an open half-space H with $0\in\partial H$

$$\begin{aligned}&\Rightarrow -r_2,\ldots,-r_{q+1}\in H' := \mathbb{R}^q\setminus H\\ &\Rightarrow \mathbb{R}_{\le 0}r_2,\ldots,\mathbb{R}_{\le 0}r_{q+1}\subset H'\\ &\Rightarrow \mathbb{R}_{\le 0}r_2+\cdots+\mathbb{R}_{\le 0}r_{q+1}\subset H'.\end{aligned}$$

Because of $r_1\notin H'$ it follows that $r_1\notin\mathbb{R}_{\le 0}r_2+\cdots+\mathbb{R}_{\le 0}r_{q+1}$. □

Proof of Proposition 1. Let $z_1,\ldots,z_m\in\mathcal{Z}$, such that $x_1,\ldots,x_m$ are linearly independent and $s_n := s_n(\theta)\neq 0$ for $n=1,\ldots,m$, where $m\le q$. Let

$$\begin{aligned}&\tilde{X} := (X_{m+1},\ldots,X_{q+1}),\\ &Z' := (z_1,\ldots,z_m,Z_{m+1},\ldots,Z_{q+1}),\\ &\mathcal{X} := v(\mathcal{Z}),\\ &\boldsymbol{X}_{gp} := \{(x_{m+1},\ldots,x_{q+1})\in\mathcal{X}^{q+1-m} : \text{each subset of } q \text{ vectors from } x_1,\ldots,x_{q+1} \text{is lin. indep.}\},\\ &\boldsymbol{X}_{ex} := \{(x_{m+1},\ldots,x_{q+1})\in\boldsymbol{X}_{gp} : \exists s_{m+1},\ldots,s_{q+1}\in\{-1,1\} : d_H(0,(s_1x_1,\ldots,s_{q+1}x_{q+1}))=1\}.\end{aligned}$$

We have to calculate $E(\psi_\theta(Z_1,\ldots,Z_{q+1})|Z_1=z_1,\ldots,Z_m=z_m)=E(\psi_\theta(Z'))$.
Since $P(\tilde{X}\in\boldsymbol{X}_{gp})=1$ and $P(\psi_\theta(Z')\in\{0,1\})=1$, we have

$$E(\psi_\theta(Z'))=P(\psi_\theta(Z')=1 \text{ and } \tilde{X}\in\boldsymbol{X}_{gp}).$$

Because of $\{\psi_\theta(Z')=1\}\cap\{\tilde{X}\in\boldsymbol{X}_{gp}\}\subset\{\tilde{X}\in\boldsymbol{X}_{ex}\}$,
it follows that

$$E(\psi_\theta(Z'))=P(\psi_\theta(Z')=1 \text{ and } \tilde{X}\in\boldsymbol{X}_{ex}).$$

For $r_2,\ldots,r_{q+1}\in\mathbb{R}^q$ let $\sigma_{(r_2,\ldots,r_{q+1})} : \{2,\ldots,q+1\}\to\{-1,1\}$, such that

$$s_1x_1\in\mathbb{R}_{\le 0}\sigma_{(r_2,\ldots,r_{q+1})}(2)r_2+\cdots+\mathbb{R}_{\le 0}\sigma_{(r_2,\ldots,r_{q+1})}(q+1)r_{q+1},$$

if each subset of q vectors from $s_1x_1,r_2,\ldots,r_{q+1}$ is linearly independent.
Since $x_2,\ldots,x_m$ are fixed, we can write $\sigma_{\tilde{x}} := \sigma_{(x_2,\ldots,x_{q+1})}$ for $\tilde{x}=(x_{m+1},\ldots,x_{q+1})\in\boldsymbol{X}_{gp}$.

Now, we prove that

$$P(\psi_\theta(Z')=1 \text{ and } \tilde{X}\in\boldsymbol{X}_{ex})=P(\forall n=m+1,\ldots,q+1 : \mathrm{sign}_\theta(Z_n)=\sigma_{\tilde{X}}(n),\tilde{X}\in\boldsymbol{X}_{ex}\}.$$

Therefore let $z_{m+1},\ldots,z_{q+1}\in\mathcal{Z}$ with $\tilde{x} := (x_{m+1},\ldots,x_{q+1})\in\boldsymbol{X}_{ex}$.
Since $\tilde{x}\in\boldsymbol{X}_{ex}$ there are $s_{m+1},\ldots,s_{q+1}\in\{-1,1\}$ with $d_H(0,(s_1x_1,\ldots,s_{q+1}x_{q+1}))=1$.
With Lemma 1 it follows that $s_1x_1\in\mathbb{R}_{\le 0}s_2x_2+\cdots+\mathbb{R}_{\le 0}s_{q+1}x_{q+1}$.
Hence, the definition of $\sigma_{\tilde{x}}$ implies that

$$s_n=\sigma_{\tilde{x}}(n)\quad\text{for } n=2,\ldots,m. \tag{9}$$

Furthermore, we have

$$\begin{aligned}
&\psi_\theta(z) = 1\\
&\Leftrightarrow \theta \notin \cap_{n=1}^{q+1} H_n \quad \text{and} \quad d_T(\theta, z) = 1\\
&\Leftrightarrow \mathrm{sign}_\theta(z_n) \neq 0 \quad \text{for } n = 1, \ldots, q+1 \quad \text{and} \quad d_T(\theta, z) = 1\\
&\Leftrightarrow \mathrm{sign}_\theta(z_n) \neq 0 \quad \text{for } n = 1, \ldots, q+1 \quad \text{and} \quad d_H(0, (s_1(\theta)x_1, \ldots, s_{q+1}(\theta)x_{q+1})) = 1\\
&\overset{\text{Lemma 1}}{\Leftrightarrow} \mathrm{sign}_\theta(z_n) \neq 0 \quad \text{for } n = 1, \ldots, q+1 \quad \text{and} \quad \mathrm{sign}_\theta(z_1)x_1 \in \mathbb{R}_{\le 0}\mathrm{sign}_\theta(z_2)x_2 + \cdots + \mathbb{R}_{\le 0}\mathrm{sign}_\theta(z_{q+1})x_{q+1}\\
&\Leftrightarrow \forall n = 2, \ldots, q+1 : \mathrm{sign}_\theta(z_n) = \sigma_{\tilde{x}}(n)\\
&\overset{(9)}{\Leftrightarrow} \forall n = m+1, \ldots, q+1 : \mathrm{sign}_\theta(z_n) = \sigma_{\tilde{x}}(n).
\end{aligned}$$

Hence,

$$\begin{aligned}
E(\psi_\theta(Z')) &= P(\forall n = m+1, \ldots, q+1 : \mathrm{sign}_\theta(Z_n) = \sigma_{\tilde{X}}(n), \tilde{X} \in \boldsymbol{X}_{ex})\\
&= \int_{\boldsymbol{X}_{ex}} P(\forall n = m+1, \ldots, q+1 : \mathrm{sign}_\theta(Z_n) = \sigma_{\tilde{x}}(n) | \tilde{X} = \tilde{x}) \mathrm{d}P^{\tilde{X}}(\tilde{x}).
\end{aligned}$$

Since $(\mathrm{sign}_\theta(Z_{m+1}), \ldots, \mathrm{sign}_\theta(Z_{q+1}))$ and $\tilde{X}$ are independent, it follows, that

$$\begin{aligned}
E(\psi_\theta(Z')) &= \int_{\boldsymbol{X}_{ex}} P(\forall n = m+1, \ldots, q+1 : \mathrm{sign}_\theta(Z_n) = \sigma_{\tilde{x}}(n)) \mathrm{d}P^{\tilde{X}}(\tilde{x})\\
&= \int_{\boldsymbol{X}_{ex}} \prod_{n=m+1}^{q+1} P(\mathrm{sign}_\theta(Z_n) = \sigma_{\tilde{x}}(n)) \mathrm{d}P^{\tilde{X}}(\tilde{x})\\
&= \int_{\boldsymbol{X}_{ex}} \left(\frac{1}{2}\right)^{q+1-m} \mathrm{d}P^{\tilde{X}}(\tilde{x})\\
&= \left(\frac{1}{2}\right)^{q+1-m} P(\tilde{X} \in \boldsymbol{X}_{ex}).
\end{aligned}$$

For $m = 1$ we have $\boldsymbol{X}_{gp} \subset \boldsymbol{X}_{ex}$ and thus $\psi_\theta^1(z_1) = \left(\frac{1}{2}\right)^{q+1-1} P(\tilde{X} \in \boldsymbol{X}_{ex}) = \frac{1}{2^q}$.
It remains to prove the second equation. Therefore, let $m = 2$.

Let $x_3, \ldots, x_{q+1} \in \mathcal{X}$, such that $(x_3, \ldots, x_{q+1}) \in \boldsymbol{X}_{gp}$ and let $w := x_3 \times \cdots \times x_{q+1}$. Then we have

$$\begin{aligned}
&(x_1, \ldots, x_{q+1}) \in \boldsymbol{X}_{ex}\\
&\overset{\text{Def.}}{\Leftrightarrow} \exists s_3, \ldots, s_{q+1} \in \{-1, 1\} : d_H(0, (s_1x_1, \ldots, s_{q+1}x_{q+1})) = 1\\
&\overset{\text{Lemma 1}}{\Leftrightarrow} \exists s_3, \ldots, s_{q+1} \in \{-1, 1\} : s_1x_1 \in \mathbb{R}_{<0}s_2x_2 + \cdots + \mathbb{R}_{<0}s_{q+1}x_{q+1}\\
&\Leftrightarrow \exists \alpha, \beta > 0, \exists \lambda \in \mathbb{R}^q, \lambda \neq 0 : (\alpha s_1x_1 + \beta s_2x_2, x_3, \ldots, x_{q+1})\lambda = 0\\
&\Leftrightarrow \exists \alpha, \beta > 0 : \det(\alpha s_1x_1 + \beta s_2x_2, x_3, \ldots, x_{q+1}) = 0\\
&\Leftrightarrow \exists \alpha, \beta > 0 : (\alpha s_1x_1 + \beta s_2x_2)^T w = 0\\
&\Leftrightarrow \exists \alpha, \beta > 0 : \alpha s_1x_1^T w + \beta s_2x_2^T w = 0\\
&\Leftrightarrow \mathrm{sign}(s_1x_1^T w) = -\mathrm{sign}(s_2x_2^T w)\\
&\Leftrightarrow s_1s_2x_1^T w x_2^T w < 0.
\end{aligned}$$

Note, that the equation

$$P(sU < 0) = sP(U < 0) + \frac{1-s}{2}$$

holds for each $\mathbb{R}$-valued random variable U with $P(U = 0) = 0$ and $s \in \{-1, 1\}$. It follows that

$$\begin{aligned}
\psi_\theta^2(z_1, z_2) &= E(\psi_\theta(Z')) - E(\psi_\theta)\\
&= \left(\frac{1}{2}\right)^{q+1-2} P(\tilde{X} \in \boldsymbol{X}_{ex}) - \frac{1}{2^q}\\
&= \left(\frac{1}{2}\right)^{q-1} P(s_1s_2x_1^T W x_2^T W < 0) - \frac{1}{2^q}
\end{aligned}$$

$$= \left(\frac{1}{2}\right)^{q-1}\left(s_1s_2P(x_1^TWx_2^TW < 0) + \frac{1-s_1s_2}{2}\right) - \frac{1}{2^q}$$
$$= \frac{s_1s_2(P(x_1^TWx_2^TW < 0) - \frac{1}{2})}{2^{q-1}}.$$

Proof of Proposition 2. Note, that the equation

$$P\left(\prod_{j=1}^{N} U_j < 0\right) = \frac{1}{2} - \frac{1}{2}(1 - 2P(U_1 < 0))^N$$

holds for $N \in \mathbb{N}$ and i.i.d. $\mathbb{R}$-valued random variables $U_1, \ldots, U_N$ with $P(U_1 = 0) = 0$. Since the occurring determinants are Vandermonde determinants, we have for all $t_1, t_2 \in \mathbb{R}$:

$$\begin{aligned}
\mathcal{K}(x(t_1), x(t_2)) &= P(x(t_1)^T(X_3 \times \cdots \times X_{q+1})x(t_2)^T(X_3 \times \cdots \times X_{q+1}) < 0) - \frac{1}{2} \\
&= P(\det(x(t_1), x(T_3), \ldots, x(T_{q+1})) \cdot \det(x(t_2), x(T_3), \ldots, x(T_{q+1})) < 0) - \frac{1}{2} \\
&= P\left(\prod_{j\geq 3}(T_j - t_1)\prod_{3\leq i<j\leq q+1}(T_j - T_i) \cdot \prod_{j\geq 3}(T_j - t_2)\prod_{3\leq i<j\leq q+1}(T_j - T_i) < 0\right) - \frac{1}{2} \\
&= P\left(\prod_{j=3}^{q+1}(T_j - t_1)(T_j - t_2) < 0\right) - \frac{1}{2} \\
&= \frac{1}{2} - \frac{1}{2}(1 - 2P((T_1 - t_1)(T_1 - t_2) < 0))^{q-1} - \frac{1}{2} \\
&= -\frac{1}{2}(1 - 2|F^{T_1}(t_1) - F^{T_1}(t_2)|)^{q-1}. \quad \square
\end{aligned}$$

Proof of Proposition 3. At first we derive the Fourier series representation of f^r where f is given by

$$f : [-1, 1] \ni z \longrightarrow f(z) = \frac{1}{2} - |z| \in \left[-\frac{1}{2}, \frac{1}{2}\right].$$

Since

$$\left\{\frac{1}{\sqrt{2}}, \cos(l\pi\cdot), \sin(l\pi\cdot); l \in \mathbb{N}\right\}$$

is an orthonormal basis of $\mathbb{L}_2[-1, 1]$ and f^r is with f an even function, f^r can be represented only by $\frac{1}{\sqrt{2}}$ and the cosine functions, i.e.

$$f^r(z) = \alpha_0^{(r)} \cdot \frac{1}{\sqrt{2}} + \sum_{l=1}^{\infty}\alpha_l^{(r)} \cdot \cos(l\pi z).$$

Since f^r is continuous and piecewise differentiable, the series is uniformly convergent so that

$$\alpha_0^{(r)} = \int_{-1}^{1} f^r(z) \cdot \frac{1}{\sqrt{2}}\mathrm{d}z = \sqrt{2}\int_0^1 f^r(z)\mathrm{d}z$$

and for $l \geq 1$

$$\alpha_l^{(r)} = \int_{-1}^{1} f^r(z) \cdot \cos(l\pi z)\mathrm{d}z = 2\int_0^1 f^r(z) \cdot \cos(l\pi z)\mathrm{d}z.$$

This implies for $r = 1$

$$\alpha_0^{(1)} = 0,$$

and for $l \geq 1$

$$\alpha_l^{(1)} = 2\int_0^1 \left(\frac{1}{2} - z\right) \cdot \cos(l\pi z)\mathrm{d}z = \begin{cases} 0, & \text{if } l \text{ is even,} \\ \dfrac{4}{l^2\pi^2}, & \text{if } l \text{ is odd.} \end{cases} \tag{10}$$

For $r=2$, we obtain

$$\alpha_0^{(2)} = 2\int_0^1 \left(\frac{1}{2}-z\right)^2 \cdot \frac{1}{\sqrt{2}}\mathrm{d}z = \frac{\sqrt{2}}{12},$$

and for $l \geq 1$

$$\alpha_l^{(2)} = 2\int_0^1 \left(\frac{1}{2}-z\right)^2 \cdot \cos(l\pi z)\mathrm{d}z = \begin{cases} \dfrac{4}{l^2\pi^2}, & \text{if } l \text{ is even,} \\ 0, & \text{if } l \text{ is odd.} \end{cases} \tag{11}$$

For $r > 2$, we have

$$\alpha_0^{(r)} = \frac{2}{\sqrt{2}}\int_0^1 \left(\frac{1}{2}-z\right)^r \mathrm{d}z = \frac{2}{\sqrt{2}} \begin{cases} 0, & \text{if } r \text{ is odd,} \\ \dfrac{1}{(r+1)2^r}, & \text{if } r \text{ is even,} \end{cases}$$

and for $l \geq 1$, partial integration provides the following recursion formula for $\alpha_l^{(r)}$

$$\begin{aligned}
\alpha_l^{(r)} &= 2\int_0^1 \left(\frac{1}{2}-z\right)^r \cdot \cos(l\pi z)\mathrm{d}z \\
&= 2\frac{1}{l\pi}\sin(l\pi z)\left(\frac{1}{2}-z\right)^r\Bigg|_0^1 + \frac{2r}{l\pi}\int_0^1 \left(\frac{1}{2}-z\right)^{r-1} \cdot \sin(l\pi z)\mathrm{d}z \\
&= \frac{2r}{l\pi}\int_0^1 \left(\frac{1}{2}-z\right)^{r-1} \cdot \sin(l\pi z)\mathrm{d}z \\
&= -\frac{2r}{l\pi}\frac{1}{l\pi}\cos(l\pi z)\left(\frac{1}{2}-z\right)^{r-1}\Bigg|_0^1 - \frac{2r}{l\pi}\frac{(r-1)}{l\pi}\int_0^1 \left(\frac{1}{2}-z\right)^{r-2} \cdot \cos(l\pi z)\mathrm{d}z \\
&= -\frac{2r}{l\pi}\frac{1}{l\pi}\left[(-1)^l\left(-\frac{1}{2}\right)^{r-1} - \left(\frac{1}{2}\right)^{r-1}\right] - \frac{2r}{l\pi}\frac{(r-1)}{l\pi}\frac{1}{2}\alpha_l^{(r-2)} \\
&= -\frac{r}{l^2\pi^2 2^{r-2}}\left[(-1)^{l+r-1}-1\right] - \frac{r(r-1)}{l^2\pi^2}\alpha_l^{(r-2)} \\
&= \begin{cases} -\dfrac{r(r-1)}{l^2\pi^2}\alpha_l^{(r-2)}, & \text{if } l+r \text{ is odd,} \\ \dfrac{r}{l^2\pi^2}\left[\dfrac{1}{2^{r-3}} - (r-1)\alpha_l^{(r-2)}\right], & \text{if } l+r \text{ is even.} \end{cases}
\end{aligned}$$

Since $\alpha_l^{(1)} = 0$ if l is even and $\alpha_l^{(2)} = 0$ if l is odd, we obtain $\alpha_l^{(r)} = 0$ if $r+l$ is odd. If $r+l$ is even and $l \geq 1$, then we have

$$\alpha_l^{(r)} = -\sum_{\substack{k\in\{1,\dots,r\} \\ k \text{ odd}}} \frac{r!}{2^{r-k-2}(r-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}}.$$

This can be seen by induction over r: for $r=1$ and l odd, it holds according to (10)

$$-\sum_{\substack{k\in\{1,\dots,r\} \\ k \text{ odd}}} \frac{r!}{2^{r-k-2}(r-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}} = -\frac{1!}{2^{-2}0!}(-l^2\pi^2)^{-1} = \frac{4}{l^2\pi^2} = \alpha_l^{(1)},$$

and for $r=2$ and l even, it holds according to (11)

$$-\sum_{\substack{k\in\{1,\dots,r\} \\ k \text{ odd}}} \frac{r!}{2^{r-k-2}(r-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}} = -\frac{2!}{2^{-1}1!}(-l^2\pi^2)^{-1} = \frac{4}{l^2\pi^2} = \alpha_l^{(2)}.$$

The induction step is done from r to $r+2$, that is:

$$\begin{aligned}
\alpha_l^{(r+2)} &= \frac{r+2}{l^2\pi^2}\left[\frac{1}{2^{r-1}} - (r+1)\alpha_l^{(r)}\right] \\
&= \frac{r+2}{l^2\pi^2}\left[\frac{1}{2^{r-1}} + (r+1)\sum_{\substack{k\in\{1,\dots,r\} \\ k\text{odd}}} \frac{r!}{2^{r-k-2}(r-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}}\right]
\end{aligned}$$

$$= \frac{(r+2)!}{2^{r+2-3}(r+2-1)!}(l^2\pi^2)^{-1} - \sum_{\substack{k\in\{1,\ldots,r\}\\ k \text{ odd}}} \frac{(r+2)!}{2^{r+2-(k+2)-2}(r+2-(k+2))!}(-l^2\pi^2)^{-\frac{k+2+1}{2}}$$

$$= -\sum_{\substack{k\in\{1,\ldots,r+2\}\\ k \text{ odd}}} \frac{(r+2)!}{2^{r+2-k-2}(r+2-k)!}(-l^2\pi^2)^{-\frac{k+1}{2}}.$$

Hence, we always have $\alpha_0^{(r)} = \sqrt{2}\gamma_0^{(r)}$ and $\alpha_l^{(r)} = 2\gamma_l^{(r)}$ for $l \geq 1$, where $\gamma_l^{(r)}$ are the quantities of Proposition 3.

To finish the proof, we transfer the Fourier series representation of $f^r(z)$ on $[-1, 1]$ to that of $g^r(s, t) = f^r(s-t)$ on $[0, 1]^2$. This provides

$$\left(\frac{1}{2} - |s-t|\right)^r = f^r(s-t) = \alpha_0^{(r)} \cdot \frac{1}{\sqrt{2}} + \sum_{l=1}^{\infty} \alpha_l^{(r)} \cdot \cos(l\pi(s-t))$$

$$= \alpha_0^{(r)} \cdot \frac{1}{\sqrt{2}} + \sum_{l=1}^{\infty} \alpha_l^{(r)} \cdot [\cos(l\pi s)\cdot\cos(l\pi t) + \sin(l\pi s)\cdot\sin(l\pi t)]$$

which is the representation given by Proposition 3 using the relation between $\alpha_l^{(r)}$ and $\gamma_l^{(r)}$. The quantities $\gamma_l^{(r)}$ are used in Proposition 3 since only

$$\mathcal{S} = \left\{1, \sqrt{2}\cos(l\pi\,\cdot), \sqrt{2}\sin(l\pi\,\cdot); l \in \mathbb{N}\right\}$$

are normalized functions of $\mathbb{L}_2[0, 1]$. However, $\mathcal{S}$ is not an orthonormal system of $\mathbb{L}_2[0, 1]$. But, since the quantities $\gamma_l^{(r)}$ are zero as soon as $r + l$ is odd, only the systems

$$\left\{\sqrt{2}\cos(l\pi\,\cdot), \sqrt{2}\sin(l\pi\,\cdot); l \in \mathbb{N} \text{ and } l \text{ is odd}\right\} \quad \text{for } r \text{ odd},$$

$$\left\{1, \sqrt{2}\cos(l\pi\,\cdot), \sqrt{2}\sin(l\pi\,\cdot); l \in \mathbb{N} \text{ and } l \text{ is even}\right\} \quad \text{for } r \text{ even},$$

are relevant and these are orthonormal systems of $\mathbb{L}_2[0, 1]$. $\square$

Acknowledgments

We thank the associate editor and the reviewers for their helpful comments and suggestions that improved the paper a lot. The data base for the example was created by Stanislav Katina. We thank him for the permission to include them.

References

[1] M.A. Arcones, Z. Chen, E. Giné, Estimators related to U-processes with applications to multivariate medians: Asymptotic normality, Ann. Statist. 22 (1994) 1460–1477.
[2] Z.-D. Bai, X. He, Asymptotic distributions of the maximal depth estimators for regression and multivariate location, Ann. Statist. 27 (1999) 1616–1637.
[3] A. Christmann, P.J. Rousseeuw, Measuring overlap in logistic regression, Comput. Statist. Data Anal. 28 (2001) 65–75.
[4] H.E. Daniels, A distribution-free test for regression parameters, Ann. Math. Statist. 25 (1954) 499–513.
[5] L. Dümbgen, Limit theorems for the simplicial depth, Statist. Probab. Lett. 14 (1992) 119–128.
[6] A.J. Lee, U-Statistics. Theory and Practice, Marcel Dekker, New York, 1990.
[7] R.Y. Liu, On a notion of simplicial depth, Proc. Natl. Acad. Sci. USA 85 (1988) 1732–1734.
[8] R.Y. Liu, On a notion of data depth based on random simplices, Ann. Statist. 18 (1990) 405–414.
[9] R.Y. Liu, Data depth and multivariate rank tests, in: Y. Dodge (Ed.), L_1-Statistical Analysis and Related Methods, North-Holland, Amsterdam, 1992, pp. 279–294.
[10] R.Y. Liu, K. Singh, A quality index based on data depth and multivariate rank tests, J. Amer. Statist. Assoc. 88 (1993) 252–260.
[11] I. Mizera, On depth and deep points: A calculus, Ann. Statist. 30 (2002) 1681–1736.
[12] I. Mizera, Ch.H. Müller, Location-scale depth, J. Amer. Statist. Assoc. 99 (2004) 949–966 (with discussion).
[13] K. Mosler, Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach, in: Lecture Notes in Statistics, vol. 165, Springer, New York, 2002.
[14] Ch.H. Müller, Depth estimators and tests based on the likelihood principle with application to regression, J. Multivariate Anal. 95 (2005) 153–181.
[15] P.J. Rousseeuw, M. Hubert, Regression depth (with discussion), J. Amer. Statist. Assoc. 94 (1999) 388–433.
[16] J.W. Tukey, Mathematics and the picturing of data, in: Proc. International Congress of Mathematicians, Vancouver 1974, 2, 1975, pp. 523–531.
[17] S. Van Aelst, P.J. Rousseeuw, M. Hubert, A. Struyf, The deepest regression method, J. Multivariate Anal. 81 (2002) 138–166.
[18] R. Wellmann, S. Katina, Ch.H. Müller, Calculation of simplicial depth estimators for polynomial regression with applications, Comput. Statist. Data Anal. 51 (2007) 5025–5040.
[19] R. Wellmann, Ch.H. Müller, Tests for multiple regression based on simplicial depth, 2008 (submitted for publication).
[20] R. Wellmann, On data depth with application to regression and tests. Ph.D. Thesis, University of Kassel, Germany, 2007.
[21] H. Witting, U. Müller-Funk, Mathematische Statistik II. Teubner, Stuttgart, 1995.
[22] Y. Zuo, R. Serfling, General notions of statistical depth function, Ann. Statist. 28 (2000) 461–482.
[23] Y. Zuo, R. Serfling, Structural properties and convergence results for contours of sample statistical depth functions, Ann. Statist. 28 (2000) 483–499.