# Depth notions for orthogonal regression

by Robin Wellmann, Christine H. Müller *
University of Kassel

May 7, 2008

Running Headline: Depth notions for regression

**Abstract**

Global depth, tangent depth and simplicial depths for classical and orthogonal regression are compared in examles and properies that are usefull for calculations are derived. Algorithms for the calculation of depths for orthogonal regression are proposed and tests for multiple regression are transfered to orthogonal regression. These tests are distribution free in the case of bivariate observations. For a particular test problem, the power of tests that are based on simplicial depth and tangent depth are compared by simulations.

**Keywords:** orthogonal regression, tangent depth, global depth, simplicial depth, asymptotic tests.
**AMS Subject classification:** Primary 62G05, 62G10; secondary 62J05, 62J12, 62G20.

## 1 Introduction

Daniels (1954) introduced a regression depth for simple linear regression, which he called a score. He derived a test for the regression parameters that is based on the distribution of the score. Rousseeuw and Hubert (1999) gave a more appealing characterization and extended it to multiple regression, see also Van Aelst et al (2002). They also worked out the analogy to Tukey's half space depth. Mizera (2002) introduced extensions of

*Research supported by the SFB/TR TRR30 Project D6

these depth notions to general parametrical models and named them global depth $d_G$, local depth $d_{loc}$, and tangent depth $d_T$, where in general $d_G \leq d_{loc} \leq d_T$. He gave sufficient conditions for their equality and showed that these depth notions are unequal for orthogonal regression.

Mizera and Müller (2004) studied these depth notions in the location scale model and Müller (2005) proposed asymptotic tests for linear and quadratic regression that are based on an extension of Liu's simplicial depth, which is nothing but the U-Statistic with a modified tangent depth as the kernel function. The tests are based on the asymptotic distribution of the simplicial depth, which is a degenerated U-Statistic in the most important cases.

Under general assumptions, Wellmann et al (2008a) derived the asymptotic distribution for polynomial regression with polynomials of arbitrary degree and thus provided distribution free tests for testing all hypothesis of the form $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \notin \Theta_0$, where $\Theta_0$ is an arbitrary subset of the parameter space. Wellmann et al (2008a) also showed that the asymptotic distribution for extended linear regression can be obtained by calculating the spectral decomposition of a function which depends only on the probability law of the vector product of the regressors. In Wellmann et al. (2008b), this result is used to derive the asymptotic distribution also for multiple regression. This paper extends these results to orthogonal regression.

Orthogonal regression means that a fit of a regression line or plane is measured by the perpendicular distance of the observations to the line or plane. This is different to classical regression, where the distance is measured parallel to the y-axis. Orthogonal regression shall be used in particular when the role of the x- and the y-axis can be exchanged or when the regression line shall be rotation invariant as it should be in images.

In Section 2, global depth for classical and orthogonal regression are introduced according to Mizera (2002). But here the depths are based on the regression planes and not on the parameters. Section 2 also provides an algorithm for calculating the global depth for orthogonal regression in the case of three bivariate observations, which is needed to calculate the simplicial depth for an arbitrary number of observations.

Section 3 introduces tangent depths for classical and orthogonal regression and studies their interrelation in Theorem 1. Tangent depth for orthogonal regression was already studied by Mizera (2002). But he characterized the orthogonal regression depth only for two dimensions. We give a characterization for any dimension based on a different parameterization. This parameterization leads to graphical representations of domains with constant depth. A plot of these domains is analogue to the dual plot of Rousseeuw and Hubert (1999) for classical linear regression, but with the advantage that the observations itselves can be included in the plot. The domains of constant depth, given in Theorem 2, are used for the calculation of maximum depth estimates.

In Section 4, generalized simplicial depths based on tangent depth and global depth for orthogonal regression are introduced and compared in examples. It turns out that simplicial depth based on global depth is more useful for estimation while simplicial depth based on tangent depth is more appropriate for testing. The tests based on the simplical depth with the tangent depth are treated in Section 5 in more detail. These tests are distribution free in the case of bivariate observations. In examples, the test for testing that the true regression line is horizontal showed to have a better power than the test of Daniels (1954), which can also be transfered to orthogonal regression.

# 2 Global depth

In this section, global depths for classical and orthogonal regression are introduced.

While the global depth for classical regression depends on the absolute residuals of observations $z_1, ..., z_N \in I\!R^q$, the global depth for orthogonal regression depends on the distances between the observations and the regression function.

Thereby, the **absolute residual** of an observation $z_n = \binom{x_n}{y_n} \in I\!R^q$ with $x_n \in I\!R^{q-1}$ with respect to a function $g : I\!R^{q-1} \to I\!R$ is defined as

$$\text{res}(g, z_n) = |y_n - g(x_n)|,$$

whereas the minimum **distance** between an observation $z_n \in I\!R^q$ and $g \subset I\!R^q$ is defined as

$$\text{dist}(g, z_n) = \inf_{z \in g} ||z - z_n||.$$

In this paper, we consider only the case that $g$ is a hyperplane.

**Definition 1** *The* ***global depth for classical regression*** $d_G^R(g, z)$ *of a hyperplane* $g$ *with respect to given observations* $z = (z_1, ..., z_N)$, $z_n \in I\!R^q$, *is the smallest number* $m$ *of observations* $z_{i_1}, \ldots, z_{i_m}$ *that needs to be removed such that there is a hyperplane* $\tilde{g} \subset I\!R^q$ *with*

$$\text{res}(\tilde{g}, z_n) < \text{res}(g, z_n)$$

*for all* $n \in \{1, ..., N\} \backslash \{z_{i_1}, \ldots, z_{i_m}\}$.

Note that this global depth is defined only for hyperplanes that are not orthogonal to the x-plane and can thus be considered as the graph of a function. The definition of global depth for orthogonal regression is similar, but here the distances are considered and not the absolute residuals, so that the following definition holds for arbitrary hyperplanes.

**Definition 2** *The **global depth for orthogonal regression** $d_G^R(g,z)$ of a hyperplane $g$ with respect to given observations $z_1,...,z_N \in I\!R^q$ is the smallest number $m$ of observations $z_{i_1},\ldots,z_{i_m}$ that needs to be removed such that there is a hyperplane $\tilde{g} \subset I\!R^q$ with*

$$\text{dist}(\tilde{g}, z_n) < \text{dist}(g, z_n)$$

*for all $n \in \{1,...,N\} \backslash \{z_{i_1},\ldots,z_{i_m}\}$.*

The global depth for classical regression coincides with the regression depth of Rousseeuw and Hubert (1999), whereas the global depth for orthogonal regression coincides with the corresponding depth in Mizera (2002). The definitions are illustrated by some examples for $q = 2$.

In Figure 1, all observations are on the same side of $g$. For classical regression the absolute residuals have to be considered. In such a case there is always a line $\tilde{g}$ parallel to $g$ for which all absolute residuals are smaller. No observation needs to be removed, so that $d_G^R(g,z) = 0$.

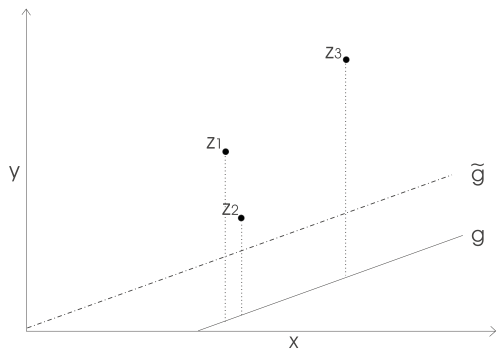


Figure 1: $d_G^R(g,z) = 0$

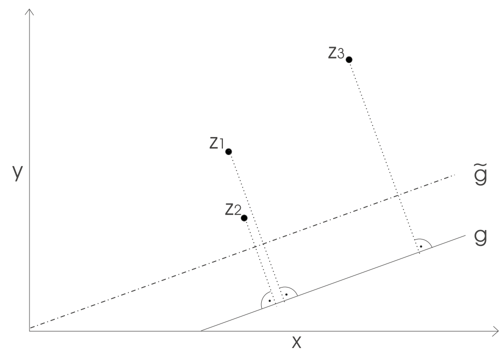


Figure 2: $d_G^o(g,z) = 0$

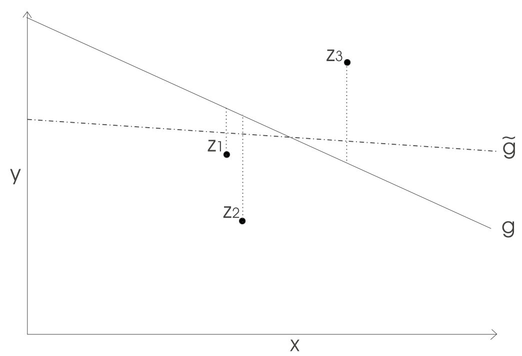


Figure 3: $d_G^R(g,z) = 0$

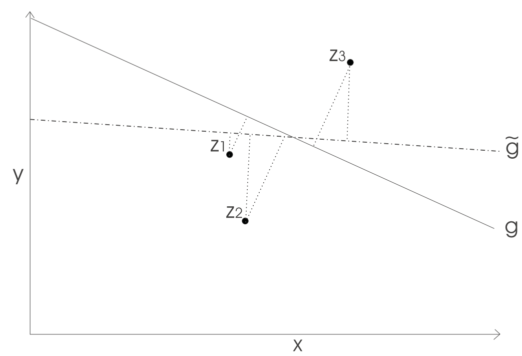


Figure 4: $d_G^o(g,z) = 0$

For orthogonal regression we have to consider not the absolute residuals but the minimum distances, which are the distances in orthogonal direction. Figure 2 shows that there is a line $\tilde{g}$ parallel to $g$ for which all distances are smaller, so that $d_G^o(g,z) = 0$.

In Figure 3 the residuals change their sign, which means that the first two observations are below the regression line and the third observation is above the line. In this case one can choose a point between $z_2$ and $z_3$ and rotate the line somewhat. In this way a line $\tilde{g}$ is obtained for which all absolute residuals are smaller, so that $d_G^R(g, z) = 0$.

Figure 4 shows for this example that $\tilde{g}$ is also closer to all observations with respect to the minimum distance, so that also $d_G^o(g, z) = 0$.

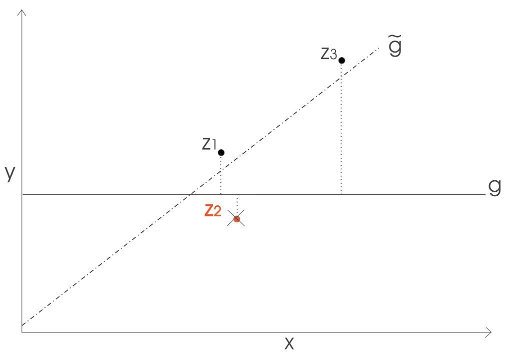


Figure 5: $d_G^R(g, z) = 1$

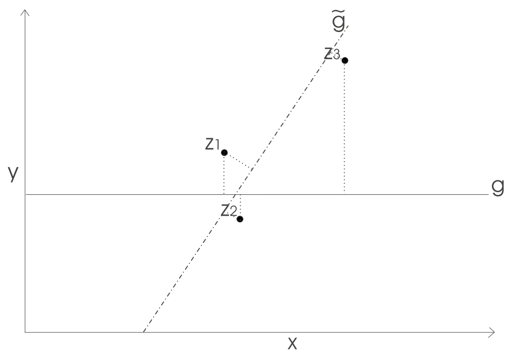


Figure 6: $d_G^o(g, z) = 0$

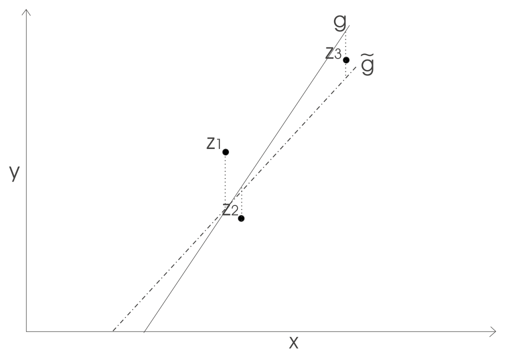


Figure 7: $d_G^R(g, z) = 0$

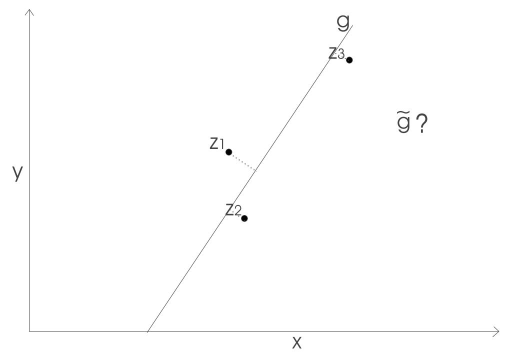


Figure 8: $d_G^o(g, z) = 1$

In the third example the residuals given by Figure 5 are alternating which means that the first observation is above the line, the second is below and the third is above. It is easy to see that there is no line $\tilde{g}$ for which all absolute residuals are smaller. But if we remove an observation, namely $z_2$, then there is a line for which all remaining absolute residuals are smaller, so that $d_G^R(g, z) = 1$.

Now we consider the minimum distances. The line $\tilde{g}$ in Figure 6 is closer to all observations so that $d_G^o(g, z) = 0$. We see that alternating residuals are not sufficient for the orthogonal depth to be one.

We can pose the question if the depth for orthogonal regression is always smaller than the depth for classical regression. The next example shows that this is not the case. In Figure 7 the absolute residuals change their sign, so that $d_G^R(g, z) = 0$. But there is no line which is closer to all observations, so that $d_G^o(g, z) = 1$ (see Figure 8).

### 2.1 Calculation of the global depth for orthogonal regression

Now we give a characterization of global depth for orthogonal regression that leads to an algorithm for the calculation of the depth for three observations in the case $q = 2$. We do not give a formal proof of this algorithm, because we think that it is clear from the pictures. For a proof see Wellmann (2008).

In the general case with an arbitrary number of observations we can create an open circle around each observation so that the radius is the distance between the observation and the line $g$ (see Figure 9). Then the global depth is nothing but the number of circles which must be removed such that there is a line $\tilde{g}$ which intersects all remaining circles.

In the case of 3 observations we need to remove at most one circle, so that the depth is at most 1, provided that the observations do not belong to the line. Furthermore, the depth is 1 if and only if there is a line which intersects all circles. Now we propose an algorithm for checking this condition.

If all observations belong to the same side of $g$ then there is a line $\tilde{g}$ parallel to $g$ which intersects all circles, so that the depth is 0 (see Figure 10).

Now we consider the case that exactly two observations belong to the same side of $g$ and that their circles have a nonempty intersection. There is a line $\tilde{g}$ that intersects this intersection and the remaining circle (see Figure 11). This line intersects all circles and thus the depth of $g$ is 0.

Finally we have to consider the case that exactly two circles belong to the same side of $g$, but they have an empty intersection. In this case the union of all lines that intersect both circles is bounded by the 4 tangents on both circles. In Figure 12, this union is given by the grey area. If the remaining circle intersects this area then there is a line $\tilde{g}$ which intersects all circles, so that the depth is 0. Otherwise there is no such line, so that the depth is 1.

## 3 Tangent depth

The definition of tangent depth depends on the parameterization of the regression function, so that this section starts with an overview on possible parameterizations.

For classical linear regression typical parameters of a hyperplane are the intercept and

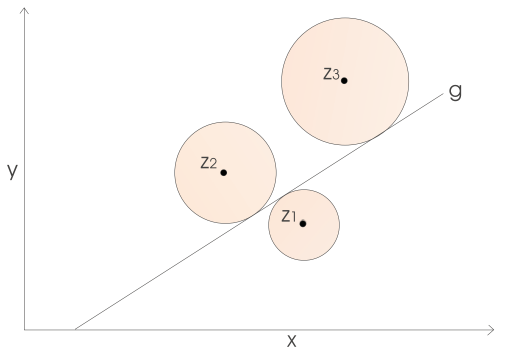


Figure 9: Calculation

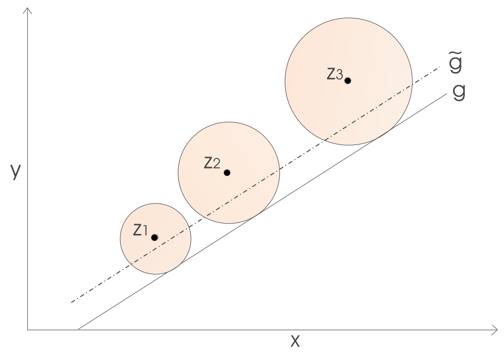


Figure 10: Step 1

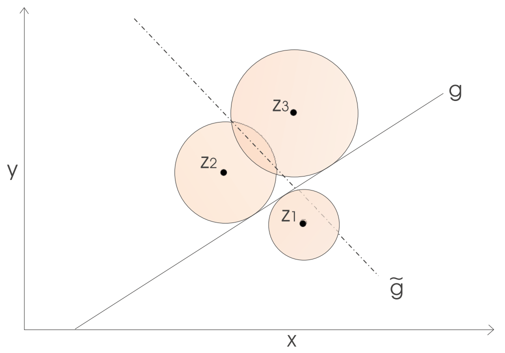


Figure 11: Step 2

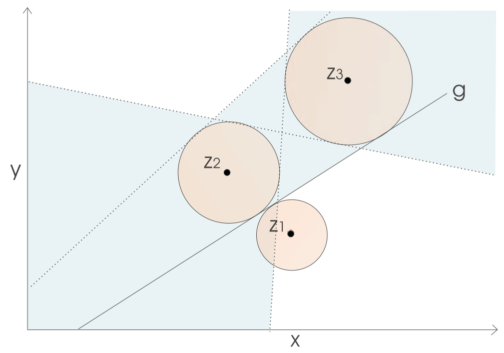


Figure 12: Step 3

the slopes (see Figure 13). The hyperplane $\mathrm{g}_\beta$ with $\beta = (\beta_1, ..., \beta_q)^T$ is defined as

$$\mathrm{g}_\beta := \{(x_1, ..., x_{q-1}, y)^T : y = \beta_1 + \beta_2 x_1 + ... + \beta_q x_{q-1}\}.$$

Note that this parameterization excludes vertical hyperplanes.

For orthogonal regression there exists no canonical parameterization. In Mizera (2002) a hyperplane was parameterized by the vector $(s, b^T)^T \in I\!R^{q+1}$, where $b \in I\!R^q$ is a unit vector orthogonal to the hyperplane and $sb$ is the intersection of the hyperplane with the linear space generated by $b$.

In this paper, we choose the point of the hyperplane with minimum distance to the origin as the parameter (see Figure 14). This point $\xi \in I\!R^q$ belongs to the hyperplane and is orthogonal to it. Then the hyperplane $g_\xi$ is given by

$$g_\xi := \{z : (z - \xi)^T \xi = 0\}.$$

Note that this parameterization must exclude hyperplanes through the origin. But, this is more a technical restriction, because Theorem 1 will show that the resulting tangent depth can be extended to cover also hyperplanes through the origin.

Hyperplanes with parameterization for orthogonal regression are denoted in italic as $g_\xi$, whereas hyperplanes with parameterization for classical regression are denotet as $\mathrm{g}_\beta$ throughout the paper.

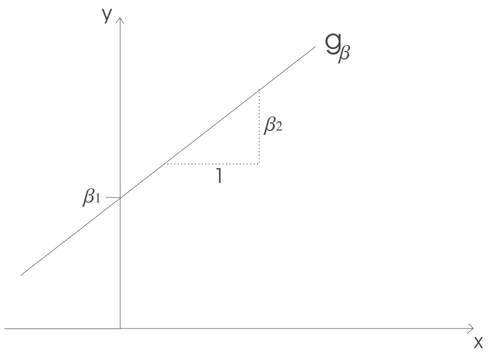


Figure 13: classical regression

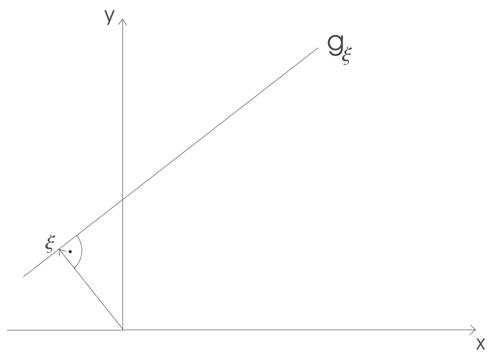


Figure 14: orthogonal regression

**Definition 3 (Tangent depth for classical regression)** *The* ***tangent depth*** *for classical regression of a hyperplane* $\mathrm{g}_\beta$, $\beta \in I\!R^q$ *with respect to given observations* $z = (z_1, ..., z_N)$ *is:*

$$d_T^R(\mathrm{g}_\beta, z) = \min_{\substack{H \subset I\!R^q \text{ half space} \\ 0 \in H}} \#\{n : \tfrac{\partial}{\partial \beta} \mathrm{res}(\mathrm{g}_\beta, z_n)^2 \in H\}.$$

Thereby, $\frac{\partial}{\partial \beta}\mathrm{res}(\mathrm{g}_\beta, z_n)^2$ denotes the gradient of the squared residual of $z_n$ at $\beta$. Roughly speaking, the gradient for an observation $z_n$ at $\beta$ is a direction in which the parameters are worse than $\beta$. The tangent depth is the minimum number of directional vectors that belong to a closed half space which contains the origin.

In this way, a tangent depth can be assigned to each global depth. Note that the parameterization that is used for the definition of tangent depth does not nessecarily have to coincide with the parameterization of the statistical model. However, this parameterization is appropriate, because Mizera (2002) showed that this tangent depth coincides with the global depth for classical regression, that is,

$$d_T^R(\mathrm{g}_\beta, z) = d_G^R(\mathrm{g}_\beta, z).$$

Tangent depth for orthogonal regression is defined in the same way:

**Definition 4 (Tangent Depth for orthogonal regression)** *The* ***tangent depth*** *for orthogonal regression of a hyperplane $g_\xi$, $\xi \in \Xi = I\!R^q \backslash \{0\}$ with respect to given observations $z = (z_1, ..., z_N)$ is:*

$$d_T^o(g_\xi, z) = \min_{\substack{H \subset I\!R^2 \text{ half space} \\ 0 \in H}} \#\{n : \tfrac{\partial}{\partial \xi} \text{dist}(g_\xi, z_n)^2 \in H\}.$$

The squared distance, considered as a function of $\xi$, is indeed differentiable since it is well known that

$$\text{dist}(g_\xi, z_n) = \frac{|\xi^T (z_n - \xi)|}{||\xi||}$$

for all $\xi \neq 0$. We will see in Figure 18 that the tangent depth for orthogonal regression does not coincide with the corresponding global depth. Mizera (2002) showed in general, that a global depth is always smaller than or equal to the corresponding tangent depth, so that

$$d_T^o(g_\xi, z) \geq d_G^o(g_\xi, z).$$

He also showed that the tangent depth for classical regression has the following characterization:

$$d_T^R(\mathrm{g}_\beta, z) = \min_{u \neq 0} \#\{n : \text{sign}(y_n - \mathrm{g}_\beta(x_n))\ u^T \begin{pmatrix} 1 \\ x_n \end{pmatrix} \geq 0\},$$

where $z_n = \binom{x_n}{y_n}$ with $x_n \in I\!R^{q-1}$. A similar formula holds for orthogonal regression:

**Lemma 1** *For all observations $z_1, ..., z_N \in I\!R^q$ and all $\xi \in I\!R^q \backslash \{0\}$, the tangent depth for orthogonal regression is given by*

$$d_T^o(g_\xi, z) = \min_{u \neq 0} \#\{n : \text{sign}\left(\frac{\xi^T (z_n - \xi)}{||\xi||}\right)\ u^T \left(\frac{\xi^T (z_n + \xi)}{\xi^T \xi} \xi - z_n\right) \geq 0\}.$$

**Proof**

$$
\begin{aligned}
\frac{\partial}{\partial \xi}\mathrm{dist}(g_\xi, z_n)^2 &= \frac{\partial}{\partial \xi}\frac{(\xi^T(z_n-\xi))^2}{\xi^T\xi} \\
&= \frac{\partial}{\partial \xi}(\xi^T z_n - \xi^T\xi)^2 \cdot (\xi^T\xi)^{-1} \\
&= (\xi^T z_n - \xi^T\xi)^2 \cdot \left(-\frac{1}{(\xi^T\xi)^2}\right) \cdot 2\xi + 2(\xi^T z_n - \xi^T\xi)(z_n - 2\xi)\cdot(\xi^T\xi)^{-1} \\
&= -2\left(\frac{\xi^T(z_n-\xi)}{\xi^T\xi}\right)^2 \xi + 2\frac{\xi^T(z_n-\xi)}{\xi^T\xi}(z_n - 2\xi) \\
&= -2\frac{\xi^T(z_n-\xi)}{\xi^T\xi}\left(\frac{\xi^T(z_n-\xi)}{\xi^T\xi}\xi + 2\xi - z_n\right) \\
&= -2\frac{\xi^T(z_n-\xi)}{\xi^T\xi}\left(\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\xi - z_n\right). \\
&= -2\frac{|\xi^T(z_n-\xi)|}{\xi^T\xi}\ \mathrm{sign}\left(\xi^T(z_n-\xi)\right)\left(\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\xi - z_n\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
d_T^o(g_\xi, z) &= \min_{u\neq 0} \#\{n : u^T \tfrac{\partial}{\partial\xi}\mathrm{dist}(g_\xi, z_n)^2 \geq 0\} \\
&= \min_{u\neq 0} \#\{n : \mathrm{sign}\left(\frac{\xi^T(z_n-\xi)}{||\xi||}\right)\ u^T\left(\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\xi - z_n\right) \geq 0\}. \qquad \Box
\end{aligned}
$$

Although this formula provides a possibility to calculate the depth, the formula seems not as simple and useful as the formula for classic regression. However, the next Theorem shows that the tangent depth for orthogonal regression of a hyperplane $g_\xi$ is nothing but the tangent depth for classical regression of the x-plane with respect to transformed observations.

**Theorem 1** *Let $\xi \in I\!R^q\backslash\{0\}$ and let $D := \frac{1}{||\xi||}\begin{pmatrix} B \\ \xi^T \end{pmatrix}$ such that the rows of $B \in I\!R^{(q-1)\times q}$ are a basis of $(I\!R\xi)^\perp$.*

*Then for all $z_1, ..., z_N \in I\!R^q$ we have*

$$
d_T^o(g_\xi, (z_1, ..., z_N)) = d_T^R(\mathrm{g}_0, (D(z_n - \xi))_{n=1,...,N}).
$$

**Proof**

For $n = 1, ..., N$ let $x_n \in I\!R^{q-1}, y_n \in I\!R$ such that

$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} = D(z_n - \xi) = \frac{1}{||\xi||}\begin{pmatrix} Bz_n \\ \xi^T z_n \end{pmatrix} - \frac{1}{||\xi||}\begin{pmatrix} B\xi \\ \xi^T\xi \end{pmatrix}$$
$$= \frac{1}{||\xi||}\begin{pmatrix} Bz_n - 0 \\ \xi^T z_n - \xi^T\xi \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{||\xi||}Bz_n \\ \frac{\xi^T(z_n-\xi)}{||\xi||} \end{pmatrix}.$$

With $A := \frac{1}{||\xi||}\begin{pmatrix} \frac{1}{||\xi||} & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \xi^T \\ -B \end{pmatrix}$ we obtain

$$A\left(\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\xi - z_n\right) = \frac{1}{||\xi||}\begin{pmatrix} \frac{1}{||\xi||} & 0 \\ 0 & I \end{pmatrix}\left[\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\begin{pmatrix} \xi^T\xi \\ -B\xi \end{pmatrix} - \begin{pmatrix} \xi^T z_n \\ -Bz_n \end{pmatrix}\right]$$
$$= \frac{1}{||\xi||}\begin{pmatrix} \frac{1}{||\xi||} & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \xi^T(z_n+\xi) - \xi^T z_n \\ 0 + Bz_n \end{pmatrix}$$
$$= \frac{1}{||\xi||}\begin{pmatrix} \frac{1}{||\xi||} & 0 \\ 0 & I \end{pmatrix}\begin{pmatrix} \xi^T\xi \\ Bz_n \end{pmatrix}$$
$$= \begin{pmatrix} 1 \\ \frac{1}{||\xi||}Bz_n \end{pmatrix}$$

Since the matrix $A$ is invertible and does not depend on the observations, we obtain with Lemma 1:

$$d_T^o(g_\xi, (z_1, ..., z_N)) = \min_{u\neq 0}\{n : \text{sign}\left(\frac{\xi^T(z_n-\xi)}{||\xi||}\right) u^T\left(\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\xi - z_n\right) \geq 0\}$$
$$= \min_{u\neq 0}\{n : \text{sign}(y_n)(A^T u)^T\left(\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\xi - z_n\right) \geq 0\}$$
$$= \min_{u\neq 0}\{n : \text{sign}(y_n)u^T A\left(\frac{\xi^T(z_n+\xi)}{\xi^T\xi}\xi - z_n\right) \geq 0\}$$
$$= \min_{u\neq 0}\{n : \text{sign}(y_n)u^T\begin{pmatrix} 1 \\ \frac{1}{||\xi||}Bz_n \end{pmatrix} \geq 0\}$$
$$= \min_{u\neq 0}\{n : \text{sign}(y_n)u^T\begin{pmatrix} 1 \\ x_n \end{pmatrix} \geq 0\}$$
$$= d_T^R(\mathrm{g}_0, (D(z_n-\xi))_{n=1,...,N}).$$

□

Note that for any hyperplane $g_\xi$ we can choose $D = D_\xi$ as a rotation matrix with $D_\xi\xi = \binom{0}{||\xi||}$. In this case, the transformation

$$T_{g_\xi}(z_n) := D_\xi(z_n - \xi)$$

shifts and rotates the observations and the regression function such that the regression function becomes the x-plane. This means that the tangent depth for orthogonal regression of $g_\xi$ is nothing but the tangent depth for classical regression of the x-plane $\mathrm{g}_0$ with respect to the shifted and rotated observations $T_{g_\xi}(z_1), ..., T_{g_\xi}(z_N)$. We extend this transformation $T$ canonically to cover also hyperplanes through the origin. If $g$ is a hyperplane through the origin, i.e. cannot expressed by $g_\xi$ with $\xi \neq 0$, then we use only a rotation $T_g$ which rotates $g$ to $\mathrm{g}_0$ in the x-plane. Hence we define the tangent depth of arbitrary hyperplanes $g \subset \mathbb{R}^q$ by

$$d_T^o(g, (z_1, ..., z_N)) := d_T^R(\mathrm{g}_0, (T_g(z_1), ..., T_g(z_N))).$$

Thus, algorithms for the calculation of tangent depth for classical regression can be used to calculate tangent depth for orthogonal regression. However, for the calculation of maximum depth estimators, also the level sets of the tangent depth need to be known.

Since the tangent depth is the hafspace depth of 0 with respect to the gradients, a sufficient small shift of their positions would not change the depth, provided that the gradients are in general position (for a proof see Wellmann, 2008). Thus, a parameter $\xi$ can be on the boundary of a level set only if $q$ gradients at $\xi$ are linearly dependent. It follows that the set

$$\mathrm{Border}(z) := \bigcup_{\{n_1,...,n_q\} \subset \{1,...,N\}} \left\{ \xi : \det \left( \frac{\partial}{\partial \xi} \mathrm{dist}(g_\xi, z_{n_1})^2, ..., \frac{\partial}{\partial \xi} \mathrm{dist}(g_\xi, z_{n_q})^2 \right) = 0 \right\}$$

divides the parameter space into domains with constant depth. The next theorem gives a simple representation of this set.

**Theorem 2** *Let $z_1, ..., z_N \in \mathbb{R}^q$, such that for all $\{n_1, ..., n_q\} \subset \{1, ..., N\}$ the vectors $z_{n_1} - z_{n_q}$, ..., $z_{n_{q-1}} - z_{n_q}$ are linearly independent. In the case $q = 2$ this means that the observations are pairwise different. The set*

$$\mathrm{Border}(z) = \bigcup_{n=1}^{N} \partial \mathbb{B} \left( \frac{z_n}{2}, \|\frac{z_n}{2}\| \right) \cup \bigcup_{\{n_1,...,n_q\} \subset \{1,...,N\}} \sum_{j<q} \mathbb{R}(z_{n_j} - z_{n_q})$$

*divides $\xi = \mathbb{R}^q \backslash \{0\}$ into domains with constant tangent depth for orthogonal regression. Thereby, $\mathbb{B}(x, r)$ denotes the open ball with center $x$ and radius $r$.*

**Proof**

Let $n_1, ..., n_q \in \{1, ..., N\}$ be pairwise different and $\xi \neq 0$. With the formula for the gradients, given in the proof of Lemma 1, we obtain

$$\begin{aligned} 0 &= \det\left(\frac{\partial}{\partial \xi}\mathrm{dist}(g_\xi, z_{n_1})^2, ..., \frac{\partial}{\partial \xi}\mathrm{dist}(g_\xi, z_{n_q})^2\right) \\ &= \left(\frac{-2}{\xi^T\xi}\right)^q \left(\prod_{j=1}^{q} \xi^T(z_{n_j} - \xi)\right) \det\left(\frac{\xi^T(z_{n_1}+\xi)}{\xi^T\xi}\xi - z_{n_1}, ..., \frac{\xi^T(z_{n_q}+\xi)}{\xi^T\xi}\xi - z_{n_q}\right) \end{aligned}$$

if and only if the determinant on the right hand side is equal to 0, or if there is a $j \in \{1, ..., q\}$ with $\xi^T(z_{n_j} - \xi) = 0$.

Let the matrices $A$ and $B$ be as in the proof of Theorem 1. Then

$$\begin{aligned} 0 &= \det\left(\frac{\xi^T(z_{n_1}+\xi)}{\xi^T\xi}\xi - z_{n_1}, ..., \frac{\xi^T(z_{n_q}+\xi)}{\xi^T\xi}\xi - z_{n_q}\right) \\ &= \det\left(A^{-1}\begin{pmatrix} 1 & \cdots & 1 \\ \frac{1}{||\xi||}Bz_{n_1} & \cdots & \frac{1}{||\xi||}Bz_{n_q} \end{pmatrix}\right) \\ &= \frac{\det(A^{-1})}{||\xi||^{q-1}} \det\begin{pmatrix} 1 & \cdots & 1 \\ Bz_{n_1} & \cdots & Bz_{n_q} \end{pmatrix} \\ &= \frac{\det(A^{-1})}{||\xi||^{q-1}} \det\begin{pmatrix} 0 & \cdots & 0 & 1 \\ Bz_{n_1} - Bz_{n_q} & \cdots & Bz_{n_{q-1}} - Bz_{n_q} & Bz_{n_q} \end{pmatrix} \\ &= \frac{\det(A^{-1})}{(-||\xi||)^{q-1}} \det\left(B(z_{n_1} - z_{n_q}), ..., B(z_{n_{q-1}} - z_{n_q})\right) \end{aligned}$$

if and only if there is a $\lambda \in I\!R^{q-1}$, $\lambda \neq 0$ with $0 = \sum_{j<q} \lambda_j B(z_{n_j} - z_{n_q}) = B\sum_{j<q} \lambda_j (z_{n_j} - z_{n_q})$. Since the rows of $B$ are a basis of $(I\!R\xi)^\perp$ and since $z_{n_1} - z_{n_q}, ..., z_{n_{q-1}} - z_{n_q}$ are linearly independent, this means that $0 \neq \sum_{j<q} \lambda_j (z_{n_j} - z_{n_q}) \in ((I\!R\xi)^\perp)^\perp = I\!R\xi$, so that

$$\sum_{j<q} I\!R(z_{n_j} - z_{n_q}) \cap I\!R\xi \neq \{0\}.$$

But this holds if and only if $\xi \in \sum_{j<q} I\!R(z_{n_j} - z_{n_q})$.

Since

$$\xi^T(z_{n_j} - \xi) = 0 \Leftrightarrow ||\xi - \frac{z_{n_j}}{2}|| = ||\frac{z_{n_j}}{2}|| \Leftrightarrow \xi \in \partial I\!B\left(\frac{z_{n_j}}{2}, ||\frac{z_{n_j}}{2}||\right)$$

we obtain

$$\begin{aligned} &\left\{\xi : \det\left(\frac{\partial}{\partial \xi}\mathrm{dist}(g_\xi, z_{n_1})^2, ..., \frac{\partial}{\partial \xi}\mathrm{dist}(g_\xi, z_{n_q})^2\right) = 0\right\} \\ &= \bigcup_{j=1}^{q} \partial I\!B\left(\frac{z_{n_j}}{2}, ||\frac{z_{n_j}}{2}||\right) \cup \sum_{j<q} I\!R(z_{n_j} - z_{n_q}) \end{aligned}$$

and the claim follows. □

For calculating domains with constant depth, Rousseeuw and Hubert (1999) developed a concept of duality for classical linear regression. This means in particular that each observation corresponds to a line in the parameter space. For orthogonal regression, we have no such duality. Here also, each observation corresponds not to a line, but to a circle. But additionally we have to consider lines which correspond to pairs of observations.

An Example with three observations for $q = 2$ may illustrate the domains of constant depth for orthogonal regression. Figure 15 shows according to Theorem 2 how the parameter space for orthogonal regression may be divided up into domains with constant depth by circles and lines. Each circle corresponds to one observation.

The observations can be plotted into the same diagram (see Figure 16). For each observation $z_n$ we obtain one circle in the parameter space. This circle contains the observation and the origin and has centre $\frac{z_n}{2}$. Furthermore, for each pair of observations we obtain one line through the origin within the parameter space. The directional vector of this line is the difference between the corresponding observations.

Figure 17 shows for a particular parameter $\xi$ the corresponding regression line. Theorem 1 shows that for the calculation of tangent depth for orthogonal regression, we have to imagine that the regression line is the x-axis and then to calculate the tangent depth for classical degression, which is equal to the global depth for classical regression. For three observations, this depth is 1 if and only if the residuals are alternating. In our case, the first observation would be below the line, the second one above, and the third observation is below, so that the residuals are alternating and thus, the depth of the parameter is 1.

In Figure 18, all domains with depth one are coloured black and the remaining domains have depth 0. Two regression lines are plotted into the diagram for which the tangent depth is 1, as can be seen with a right angle triangular ruler. The increasing line has depth 1, although all observations are clother to the decreasing line. This shows, that global depth and tangent depth do not coincide for orthogonal regression, because the global depth of the increasing line is 0. Moreover, the tangent depth for orthogonal regression is 1, if the residuals in orthogonal direction are alternating, no matter how far the observations are away from the regression line. This is clearly not a desirable property of a depth function, so that we expect that global depth is more appropriate for parameter estimation.

Figure 19 and Figure 20 show the level sets of tangent depth and global depth for orthogonal regression with respect to 3 other observations. Again, the parameter space is divided up into domains with constant tangent depth by circles and lines. However, for global depth, the regions with depth one are much smaller and the boundaries of the domains are not completely contained in the union of circles and lines.

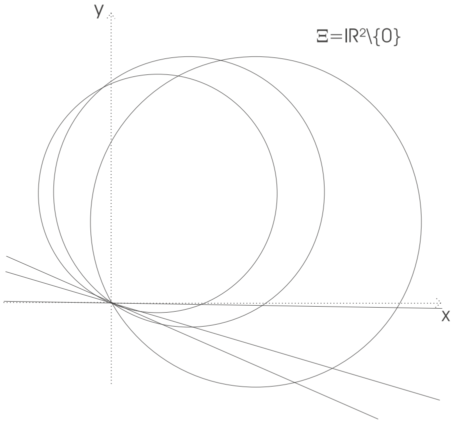


Figure 15: The border

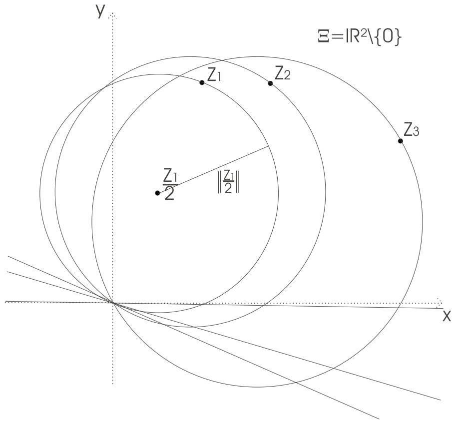


Figure 16: Observations included

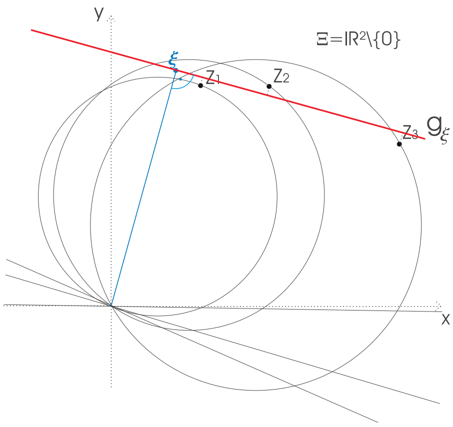


Figure 17: Parameterization

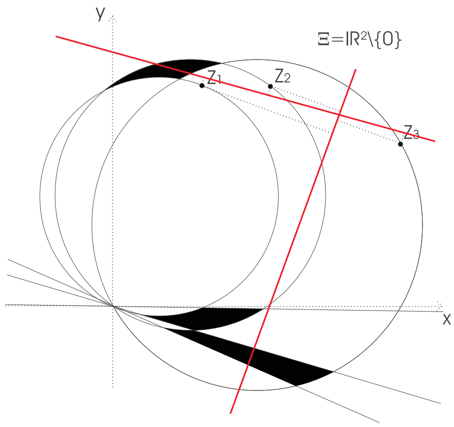


Figure 18: Level sets

# 4 Simplicial depth

This section introduces a third depth notion, namely an extension of Liu's simplicial depth for multivariate location. Each depth notion $d$ that is intruduced in the previous sections gives reason to the definition of a different simplicial depth:

**Definition 5 (Simplicial depth)** *The **simplicial depth** of a hyperplane $g \subset \mathbb{R}^q$ and observations $z = (z_1, ..., z_N)$ that is based on a depth notion $d$ is defined as*

$$S_d(g, z) = \binom{N}{q+1}^{-1} \sum_{1 \le n_1 < ... < n_{q+1} \le N} d(g, (z_{n_1}, ..., z_{n_{q+1}})).$$

That is, for a given hyperplane $g$ and given observations we calculate the depth $d$ for each subset of $q+1$ observations and the simplicial depth is defined as the mean of theses

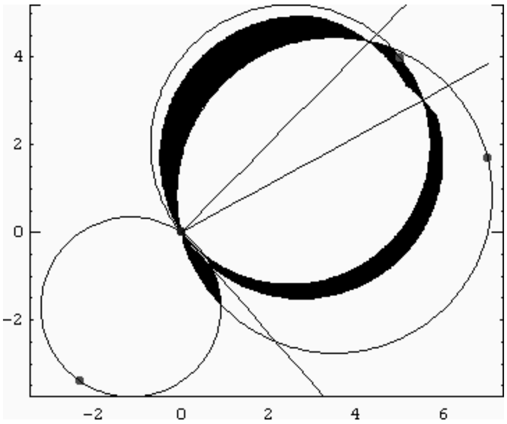


Figure 19: global depth

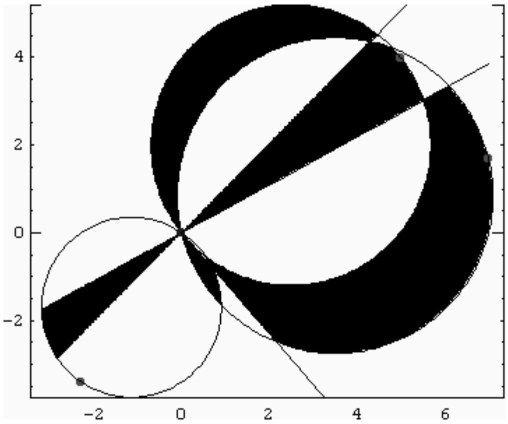


Figure 20: tangent depth

values.

We obtain a total of 3 different simplicial depths. One is based on the global depth for orthogonal regression, one on the tangent depth for orthogonal regression, and one is based on the global depth for classical regression.

Figure 21 and Figure 22 compare the simplicial depths for orthogonal regression by an example with 15 observations and $q = 2$. The Figures show the parameter space and the grey level of each parameter corresponds to the depth of this parameter. The black parameters have maximum depth. Also the observations and the true regression line are plotted into the diagrams. The used parameterization for orthogonal regression yield that the parameter of the true regression line is the point of the line with minimum distance to the origin. Note that the point in the diagram which marks the origin is not an observation.

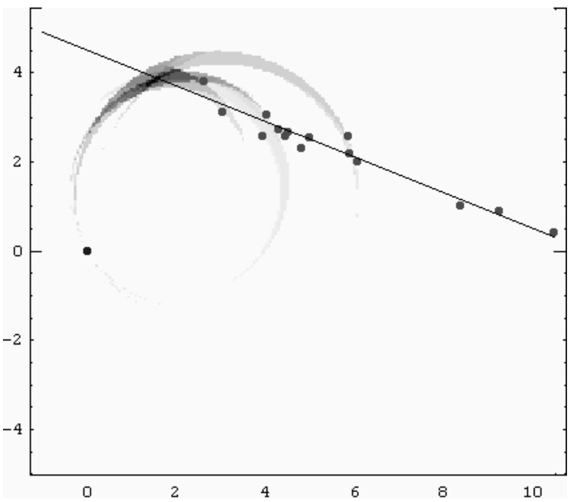


Figure 21: simplicial global depth

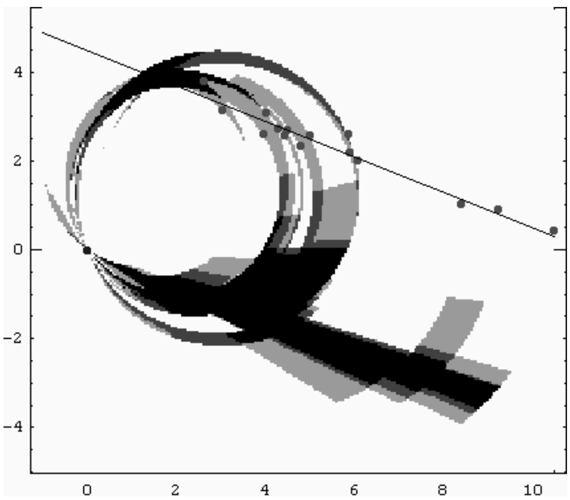


Figure 22: simplicial tangent depth

If the simplicial depth is based on global depth (Figure 21), then the true parameter belongs to the black area, as can be seen with a right-angle triangular ruler. Thus, the simplicial depth based on global depth estimates the true parameter well.

If the simplicial depth is based on tangent depth (Figure 22), then not only the true parameter has a high depth but also parameters far away from the true parameter. This shows that the simplicial depth based on tangent depth is not appropriate for parameter estimation. But it can be seen, that there are also many parameters with very small depth and this shows that tangent depth could be appropriate for tests, if under the alternative all parameters (resp. hyperplanes) from the null hypothesis have small depths.

Indeed the simplicial depth based on tangent depth may be more appropriate for tests than the depth which is based on global depth, because simulations showed that the distribution of the maximum depth depends much on the underlying distribution of the observations if the simplicial depth is based on global depth. This is not the case for simplicial depth based on tangent depth since in this case the asymptotic behaviour under the null hypothesis is the same as for simplicial depth for classical regression and this does not depend on the unknown parameters. This is shown in the next section.

# 5 Tests

Take $\Theta$ to be the parameter space of the statistical model. To allow also for semiparametrical models we do not assume that $\Theta$ is finite dimensional. Let $\mathcal{G}$ be the set of all hyperplanes in $I\!R^q$ and for $\theta \in \Theta$ let $\tilde{g}(\theta) \in \mathcal{G}$.

Let the $q$-variate random vectors $Z_1, ..., Z_N$ be independent and identically distributed and for $\theta \in \Theta$ suppose that

- $P_\theta(Y_n > 0 | X_n) = \frac{1}{2}$,
- $P_\theta(Y_n < 0 | X_n) = \frac{1}{2}$,
- $X_n$ has a continuous distribution,

where $\binom{X_n}{Y_n} = T_{\tilde{g}(\theta)}(Z_n)$ and $Y_n$ is a one dimensional random variable. In the case $q > 2$ we assume additionally that $X_n$ has a multivariate Cauchy distribution.

For orthogonal regression, usually an error-in-variable model is assumed. The following error-in-variable model is a special case of the general model, defined above:

**Example 1** Let $\theta \in \Theta$ and let $Z_1, ..., Z_N$ be i.i.d. bivariate, continuous distributed random variables such that

$$Z_n = V_n + E_n,$$

where $V_n : \Omega \to \tilde{g}(\theta)$ and the error $E_n$ is radially symmetric distributed given $V_n$.

This model satisfies the assumptions, given above.

**Proof**

We can write the transformation $T_{\tilde{g}(\theta)}$ as $T_{\tilde{g}(\theta)}(z) = D(z - w)$ with a rotation matrix $D$. Let $\binom{S_n}{0} = T_{\tilde{g}(\theta)}(V_n)$ and $\binom{U_n}{Y_n} = DE_n$. Note that $T_{\tilde{g}(\theta)}(Z_n) = \binom{X_n}{Y_n}$, where $X_n = U_n + S_n$. Since $\binom{U_n}{Y_n}$ is radially symmetric distributed given $V_n$, we have $P_\theta(Y_n > 0|V_n, U_n) = \frac{1}{2}$. Since $X_n$ can be written as a function of $V_n$ and $U_n$, it follows that $P_\theta(Y_n > 0|X_n) = \frac{1}{2}$.

The second assumption of the general model follows immediately and the third assumption holds by definition. □

Because of Theorem 1, the asymptotic distribution of the simplicial depth for orthogonal regression which is based on tangent depth is equal to the asymptotic distribution of the simplicial depth for classical regression, given in Wellmann et al. (2008), so that tests for testing

$$H_0 : \tilde{g}(\theta) \in \mathcal{G}_0 \subset \mathcal{G} \text{ against } H_1 : \tilde{g}(\theta) \notin \mathcal{G}_0$$

can be based on on the test statistic

$$T(Z) \quad = \quad N\left(\sup_{g\in\mathcal{G}_0} S_{d_T^o}(g, Z) - \frac{1}{2^q}\right). \tag{1}$$

$H_0$ is rejected, if the test statistic is less than the $\alpha$-quantile of the asymptotic distribution of the simplicial depth. This test is indeed an asymptotic $\alpha$-level test, since for any $c \in I\!R$ and all $\theta \in \Theta$ with $\tilde{g}(\theta) \in \mathcal{G}_0$ we have

$$P_\theta\left(\sup_{g\in\mathcal{G}_0} S_{d_T^o}(g, Z) \leq c\right) \quad \leq \quad P_\theta\left(S_{d_T^o}(\tilde{g}(\theta), Z) \leq c\right). \tag{2}$$

In the case $q = 2$ the distribution of the tangent depth under the above assumptions is given in Daniels (1954), see also Van Aelst et. al (2002) so that alternatively, the test could be based on the test statistic

$$T(Z) = \sup_{g\in\mathcal{G}_0} d_T^o(g, Z).$$

and $H_0$ is rejected, if the test statistic is less than the $\alpha$-quantile of the distribution of the tangent depth.

## 5.1 Power comparison with simulated data

We compared the power of both tests in the case $q = 2$, where the null hypothesis is tested that the true regression line is horizontal, so that $\mathcal{G}_0$ consists on all horizontal lines. We simulated true observations $V_1, ..., V_N$ on a line $g$ with $\text{dist}(g, 0) = 1$ for which the angle between $g$ and the $x$-axis is $0 \leq \gamma \leq \frac{\pi}{4}$. That is, we tested $H_0 : \gamma = 0$ against $H_1 : 0 < \gamma \leq \frac{\pi}{4}$.

We used the Cauchy distribution for power comparisons in order to simulate outliers. The true observations are simulated such that $T_g(V_n) = \binom{S_n}{0}$ where $S_n$ is Cauchy distributed with location parameter 0 and scale parameter 4.

The observations $Z_1, ..., Z_N$ satisfy

$$Z_n = V_n + E_n \text{ with } E_n \sim \text{Cauchy}_2(0, I),$$

which means that the Error $E_n$ has a centered, bivariate Cauchy Distribution with the identity matrix as the scatter matrix.

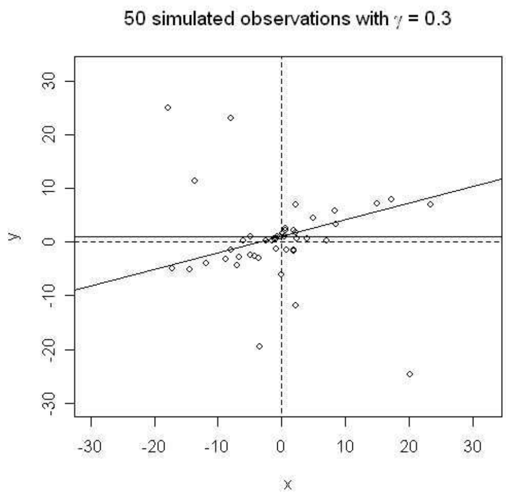


Figure 23: Simulated observations

All tests are performed to the level $\alpha = 0.05$. Figure 23 shows 50 simulated observations, the true regression line with $\gamma = 0.3$, and a horizontal line from the null hypothesis. In this example the null hypothesis was rejected with the simplicial depth test. Figures 24 and 25 show the probability to make the $\beta$-error for different values of $\gamma$ and different sample sizes.

It can be seen that both tests are indeed $\alpha$-level tests with $\alpha = 0.05$ because for $\gamma = 0$, both lines are above 0.95. It can be seen also that the simplicial depth test is better in

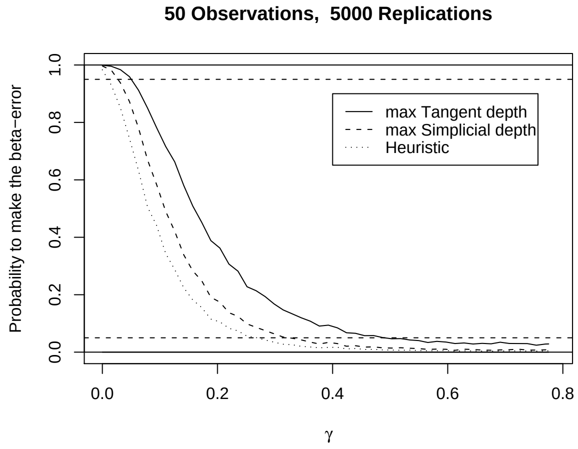


Figure 24: 50 observations

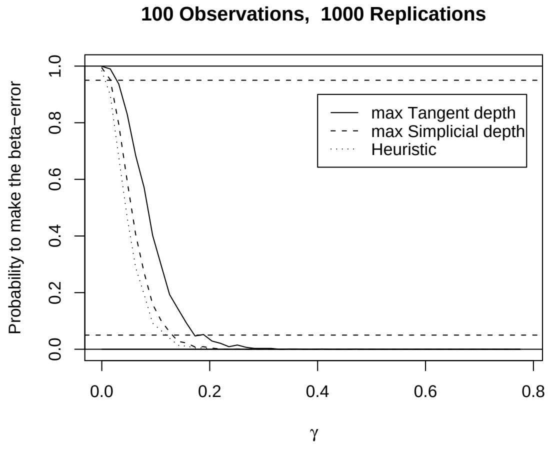


Figure 25: 100 observations

both examples than the tangent depth test because the curve of the simplicial depth test is below the other one. For 50 observations simulated with $0.3 \leq \gamma \leq \frac{\pi}{4}$ the nullhypothesis was rejected in nearly all cases by the simplicial depth test.

Since in Equation (1), the depth is maximized over several parameter values, the true level of the test is smaller than $\alpha$ (see Equation 2), so that the power is not very good for small $\gamma$. As a heuristic, we propose to use not (1) as the test statistic, but

$$T(Z) = N\left(S_{d_T^0}(\widehat{g}(Z), Z) - \frac{1}{2^q}\right),$$

where $\widehat{g}(Z)$ is the horizontal line with intercept $\text{med}(Z_{1,2}, ..., Z_{N,2})$. The resulting power function is also given in Figures (24) and (25). Further improvement for this particular test could be achieved by calculating the exact distribution of $S_{d_T^0}(\widehat{g}(Z), Z)$.

# References


[1] Daniels, H.E. (1954). A distribution-free test for regression parameters, *Ann. Math. Statist.* **25**, 499-513.

[2] Liu, R.Y. (1988). On a notion of simplicial depth. *Proc. Nat. Acad. Sci. USA* **85**, 1732-1734.

[3] Liu, R.Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18**, 405-414.

[4] Mizera, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30**, 1681-1736.

[5] Mizera, I. and Müller, Ch.H. (2004). Location-scale depth. *Journal of the American Statistical Association* **99**, 949-966. With discussion.

[6] Müller, Ch.H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *Journal of Multivariate Analysis* **95**, 153-181.

[7] Rousseeuw, P.J. and Hubert, M. (1999). Regression depth. *J. Amer. Statist. Assoc.* **94**, 388-433.

[8] Tukey, J.W. (1975). Mathematics and the picturing of data. In *Proc. International Congress of Mathematicians*, Vancouver 1974, **2**, 523- 531.

[9] Van Aelst, S., Rousseeuw, P.J., Hubert, M. and Struyf, A. (2002). The deepest regression method. *J. Multivariate Anal.* **81**, 138-166.

[10] Wellmann, R. (2008). On data depth with application to regression models and tests. *Ph.D. thesis.* University of Kassel, Germany.

[11] Wellmann, R., Katina, S. and Müller, Ch.H. (2007a). Calculation of simplicial depth estimators for polynomial regression with applications. *Computational Statistics and Data Analysis* **51**, 5025-5040.

[12] Wellmann, R., Harmand, P. and Müller, Ch.H. (2008a). Distribution free tests for polynomial regression based on simplicial depth. *Submitted.*

[13] Wellmann, R., Müller, Ch.H. (2008b). Tests for multiple regression based on simplicial depth. *Submitted.*

wellmann@mathematik.uni-kassel.de
Department of Mathematics
University of Kassel
D-34109 Kassel
Germany