SIMPLIFIED SIMPLICIAL DEPTH FOR REGRESSION AND AUTOREGRESSIVE GROWTH PROCESSES

Christoph P. Kustosz * † and Christine H. Müller *
Technische Universität Dortmund
Fakultät Statistik
Vogelpothsweg 87
D-44221 Dortmund
Germany
kustosz@statistik.tu-dortmund.de, cmueller@statistik.tu-dortmund.de

Martin Wendler*
Ruhr-Universität Bochum
Fakultät für Mathematik
Universitätsstraße 150
D-44801 Bochum
Germany
Martin.Wendler@ruhr-uni-bochum.de

Abstract

We simplify simplicial depth in two directions for regression and autoregressive growth processes. At first we show that simplicial tangent depth often reduces to counting the subsets with alternating signs of the residuals if the regressors are ordered. The second simplification is given by not regarding all subsets of residuals. By consideration of only special subsets of residuals, the asymptotic distributions of the simplified simplicial depth notions are normal distributions so that tests and confidence intervals can be derived easily. We propose two simplifications for the general case and a third simplification for the special case where two parameters are unknown. Additionally, we derive conditions for the consistency of the tests. We show that the simplified depth notions can be used for polynomial regression, for several nonlinear regression models, and for several autoregressive growth processes. We compare the efficiency and robustness of the different simplified versions by a simulation study concerning the Michaelis-Menten model and a nonlinear autoregressive process of order one and provide an application on crack growth.

Keywords and Phrases: Alternating sign, distribution-free test, robustness, tangent depth, asymptotic distribution, consistency.

---

* Research supported by DFG via the Collaborative Research Centre SFB 823 'Statistical modelling of nonlinear dynamic processes' (Project B5).

† Corresponding Author, E-Mail: kustosz@statistik.tu-dortmund.de, Tel. +492317553114

1. INTRODUCTION

Data depth is a possibility to generalize the median and ranks to complex situations. Starting with the halfspace depth of Tukey (1975) for multivariate data, meanwhile many depth notions were proposed. There exist depth notions for regression as in Rousseeuw and Hubert (1999), for generalized linear models as in Müller (2005), for estimation equations as in Lin and Chen (2006), for functional data as in López-Pintado and Romo (2009) or Claeskens et al. (2014), for copulas as in Denecke and Müller (2011), and for correlation as in Denecke and Müller (2014). Further depth can be used to estimate quantiles, also in regression, as discussed by Hallin et al. (2010). To describe multimodal densities, Paindaveine and van Bever (2013) and Agostinelli and Romanazzi (2011) proposed local versions of depth and Lok and Lee (2011) introduced a depth function based on interpoint distances. Depth is also applied to analyze distributions as shown by Kong and Zuo (2010), Mizera and Müller (2004) and Rousseeuw and Ruts (1999) or for classification as presented by Dutta and Ghosh (2012) and Li et al. (2012). See also the book of Mosler (2002) and the general approaches of Zuo and Serfling (2000a,b) or Mizera (2002).

Important for the statistical applicability of a depth notion beyond estimation is that at least an asymptotic distribution is known. However, it is very difficult for many depth notions to derive the asymptotic distribution. One general approach is to use simplicial depth, since simplicial depth is a U-statistic and the asymptotic distribution for U-statistics is in principle known.

Simplicial depth was originally introduced by Liu (1988, 1990) as an extension of the halfspace depth of Tukey (1975). If the data are K-dimensional then the simplicial depth of a parameter $\mu \in \mathbb{R}^K$ is the relative number of simplices spanned by $K+1$ data points which contain μ. Thereby, μ is contained in a simplex spanned by $K+1$ points if its halfspace depth with respect to these $K+1$ points is greater than 0. This is the key to generalize simplicial depth to many situations. As soon as a depth notion $d(\theta,(z_1,\ldots,z_{K+1}))$ of a K-dimensional parameter θ and a specific model is known for any data set $(z_1,\ldots,z_{K+1})$, then simplicial depth of θ in a sample $z_* = (z_1,\ldots,z_N)$ is defined as

$$d_S(\theta, z_*) := \frac{1}{\binom{N}{K+1}} \sum_{1\le n_1<n_2<\ldots<n_{K+1}\le N} \mathbb{1}\{d(\theta,(z_{n_1},z_{n_2},\ldots,z_{n_{K+1}}))>0\}, \tag{1}$$

where $\mathbb{1}\{h(z)>0\}$ denotes the indicator function $\mathbb{1}_A(z)$ with $A=\{\tilde{z};\ h(\tilde{z})>0\}$ for any function h. Thereby, $d_S(\theta,z_*)$ should be large if θ is the correct parameter of the model and should be small if θ is not the correct parameter. Hence a simple rule for testing $H_0:\theta\in\Theta_0$ is the following: reject H_0 if $\sup_{\theta\in\Theta_0} d_S(\theta,z_*)$ is smaller than a critical value c, as e.g. proposed by Müller (2005). The critical value must be determined by the distribution $d_S(\theta,z_*)$ or at least by the asymptotic distribution of $d_S(\theta,z_*)$, if θ is the underlying parameter.

Although $d_S(\theta,z_*)$ is a U-statistic, it is only in few cases not a degenerated U-statistic, see Denecke and Müller (2011, 2012, 2013, 2014). In most cases, $d_S(\theta,z_*)$ is a degenerated U-statistic and its asymptotic distribution must be determined by a spectral decomposition of the conditional expectation. If the unknown parameter is one-dimensional, then the spectral decomposition is still simple as shown for linear regression through the origin in Müller (2005) and for a linear AR(1) model without intercept in Kustosz and Müller (2014). However, it becomes more complicated if more than one parameter is unknown. Dümbgen (1992) derives a functional limit theorem for simplicial depth for general distributions under some restrictions and applies it to location models. Other results were presented for linear and quadratic regression in Müller (2005), for polynomial regression in Wellmann et al. (2009), for multiple regression in Wellmann and Müller (2010a), and for orthogonal regression in Wellmann and Müller (2010b). Thereby, not only the derivation is complicated but also the resulting asymptotic distributions are. In case of specific models the limit distributions can be derived exactly. For example, the asymptotic distribution is given by an infinite sum of Chi-squared distributed random variables for polynomial regression. For AR(1) processes with intercept, it is even worse. Here the asymptotic distribution is given by an integrated squared Gaussian process as shown in Kustosz et al. (2015).

Therefore here, we provide simplified versions of the simplicial depth. At first, we prove in Section 2 that the calculation of the depth $d(\theta, (z_{n_1}, z_{n_2}, \ldots, z_{n_{K+1}}))$ of $K + 1$ data points reduces in many cases to a check whether residuals at the ordered data set have alternating signs. This property of simplicial depth was already noticed by Rousseeuw and Hubert (1999) for linear regression and used by Müller (2005) for polynomial regression. However, a complete proof for this property was not given. Here we provide general sufficient conditions for this property which are satisfied not only by polynomial regression but by many other models like nonlinear models or autoregressive models. As soon as these sufficient conditions are satisfied, the simplicial depth can be easily calculated by counting the subsets with $K + 1$ points with alternating signs. Hence the simplicial depth is a modification of the simple sign test where only subsets with one data point are considered. Since the tests are based on the signs of the residuals only, they are robust against outliers.

However, even checking the simple criterion of alternating signs can be computationally intensive if N and K are large since $\binom{N}{K+1}$ subsets have to be analyzed. Additionally, the above mentioned problem of deriving the asymptotic distribution remains. Therefore, we propose simplified versions of the simplicial depth in Section 3 by not regarding all $\binom{N}{K+1}$ subsets. Instead, we propose only subsequent subsets. The subsets are nonoverlapping in the first version and overlapping in the second version. Additionally, a third version is introduced for the case of two unknown parameters, i.e. $K = 2$. All versions have a computational complexity of N instead of $\binom{N}{K+1}$ and it is proven that the asymptotic distribution is always the normal distribution.

In Section 4, sufficient conditions for the consistency of tests based on these simplified simplicial depth statistics are proven. Section 5 contains several examples, where the conditions used in Sections 2, 3, and 4 are satisfied. These examples include polynomial regression, several nonlinear models and several autoregressive growth processes with two and three unknown parameters. Finally, Section 6 provides a simulation study for the Michaelis-Menten model and a nonlinear autoregressive process with two parameters and Section 7 an application on crack growth.

All proofs are given in the Appendix.

2. DATA DEPTH VIA ALTERNATING SIGNS

We consider a general model of the form

$$y_n = g(x_n, \theta) + e_n, \quad \text{for } n \in \{1, \ldots, N\},$$

where $\theta \in I\!R^K$ is the unknown parameter vector and $z_n = (y_n, x_n) \in I\!R^2$, $n \in \{1, \ldots, N\}$, are the data points. The errors $e_1, \ldots, e_N$ are realizations of independent and identically distributed random variables $E_1, \ldots, E_N$, so that $y_1, \ldots, y_N$ and $z_1, \ldots, z_N$ are realizations of random variables $Y_1, \ldots, Y_N$ and $Z_1, \ldots, Z_N$, respectively. Although the regressors $x_1, \ldots, x_N$ are fixed for regression, we always interpret them as realizations of random variables $X_1, \ldots, X_N$ which satisfy

$$X_1 < X_2 < \ldots < X_N \tag{2}$$

almost surely. In particular, we have $X_n = Y_{n-1}$ for autoregressive processes so that condition (2) implies that the autoregressive process is strictly increasing, i.e. it is a growth process.

To provide a characterization of the depth of θ at subsets with $K+1$ data points in this section, we regard here only $N = K+1$. Moreover, we do not need the random variables here, but they are important for the asymptotic normality shown in Section 3.

General depth notions are global and tangent depth introduced by Mizera (2002). Mizera proposed these depth notions for an arbitrary quality function. Here the quality function shall be given by the squared residuals so that global depth coincides with tangent depth in many cases. Although the interpretation of tangent depth is less obvious, it is computationally more feasible. Therefore, we use tangent depth here. The tangent depth of θ in $z_* = (z_1, \ldots, z_{K+1})$ is defined as

$$d_T(\theta, z_*) := \frac{1}{K+1} \min_{u \in I\!R^K} \sharp \left\{ n \in \{1, \ldots, K+1\};\ u^\top \frac{\partial}{\partial \theta} \operatorname{res}(z_n, \theta)^2 \leq 0 \right\}$$

where $\operatorname{res}(z_n, \theta) := y_n - g(x_n, \theta)$, $n \in \{1, \ldots, N\}$, are the residuals and $\sharp A$ denotes the cardinality of a set A. Setting

$$v(x_n, \theta) := \frac{\partial}{\partial \theta} g(x_n, \theta)$$

tangent depth can be also written as

$$d_T(\theta, z_*) = \frac{1}{K+1} \min_{u \in I\!R^K} \sharp \left\{ n \in \{1, \ldots, K+1\};\ u^\top v(x_n, \theta) \operatorname{res}(z_n, \theta) \leq 0 \right\}.$$

Our first theorem proves the relation between $d_T(\theta, z_*) > 0$ and alternating signs of the residuals. For that, we need some definitions.

Definition 1.
a) Let sgn(y) denote the sign of a number $y \in I\!R$, i.e. sgn$(y) = 1$ if $y > 0$, sgn$(y) = -1$ if $y < 0$, and sgn$(y) = 0$ if $y = 0$.
b) A vector $s = (s_1, \ldots, s_{K+1})^\top \in I\!R^{K+1}$ has alternating signs if sgn$(s_k) = -$sgn$(s_{k+1}) \neq 0$ for all $k \in \{1, \ldots, K\}$ is satisfied. If s has alternating signs, then it has K sign changes.
c) A function $f : [a, b] \to I\!R$ has K sign changes on the interval $[a, b] \subset I\!R$ if there exist $x_1 < x_2 < \ldots < x_{K+1}$ with $x_k \in [a, b]$ for $k \in \{1, \ldots, K+1\}$ and sgn$(f(x_k)) = -$sgn$(f(x_{k+1})) \neq 0$ for $k \in \{1, \ldots, K\}$ and no $x_1 < x_2 < \ldots < x_{L+1}$ for $L > K$ so that $x_l \in [a, b]$ for $l \in \{1, \ldots, L+1\}$ and sgn$(f(x_l)) = -$sgn$(f(x_{l+1})) \neq 0$ for $l \in \{1, \ldots, L\}$.

Theorem 1. *Let be $x_1 < x_2 < \ldots < x_{K+1} \in I\!R$ and assume the following conditions for $w_u : [x_1, x_{K+1}] \to I\!R$ given by $w_u(x) = u^\top v(x, \theta)$:*

A) w_u *has at most* $K-1$ *sign changes on* $[x_1, x_{K+1}]$ *for all* $u \in I\!R^K$*,*

B) For any $s \in \{-1,1\}^{K+1}$ *with at most* $K-1$ *sign changes, there exists* $u_0 \in I\!R^K$ *with* *sgn*$(w_{u_0}(x_n)) = s_n$ *for* $n \in \{1,\ldots,K+1\}$*.*

Then $d_T(\theta, z_*) > 0$ *holds if and only if* $(\textit{res}(z_1,\theta),\ldots,\textit{res}(z_{K+1},\theta))^\top$ *has alternating signs or at least one of the residuals is zero.*

The examples in Section 5 show that the conditions of Theorem 1 are usually satisfied for well known models. Hence in these cases, simplicial depth based on d_T reduces to

$$d_S(\theta, z_*) = \frac{1}{\binom{N}{K+1}} \sum\nolimits_{1\le n_1<n_2<\ldots<n_{K+1}\le N} \left(\prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(z_{n_k},\theta)(-1)^k > 0\right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(z_{n_k},\theta)(-1)^{k+1} > 0\right\} + 1 - \prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(z_{n_k},\theta) \neq 0\right\} \right). \tag{3}$$

3. SIMPLIFIED SIMPLICIAL DEPTH

If we assume that the residuals have continuous distributions, then they are not equal to zero with probability one. Under this assumption and the assumptions of Theorem 1, we have with probability one that $d_T(\theta,(z_{n_1},\ldots,z_{n_{K+1}})) > 0$ holds if and only if the residuals $\text{res}(z_{n_1},\theta),\ldots,\text{res}(z_{n_{K+1}},\theta)$ have alternating signs, whereby $n_i \in \{1,...,N\}$ and $n_i > n_j$ if $i > j$. Hence the simplicial depth (3) of θ in $z_* = (z_1,\ldots,z_N)$ is given almost surely by

$$d_S(\theta, z_*) = \frac{1}{\binom{N}{K+1}} \sum\nolimits_{1\le n_1<n_2<\ldots<n_{K+1}\le N} \left(\prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(z_{n_k},\theta)(-1)^k > 0\right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(z_{n_k},\theta)(-1)^{k+1} > 0\right\} \right). \tag{4}$$

The asymptotic distribution of d_S given by (4) is only known for $K \in \{1,2\}$. For $K=2$ it is for example given by an integrated two-dimensional Gaussian process as shown in Kustosz et al. (2015). To obtain more simple asymptotic results and to avoid the consideration of all $\binom{N}{K+1}$

subsets of the data set, we define the following simplified simplicial depth notions:

$$d_S^1(\theta, z_*) := \frac{1}{\left\lfloor \frac{N}{K+1} \right\rfloor} \sum_{n=1}^{\left\lfloor \frac{N}{K+1} \right\rfloor} \left(\prod_{k=1}^{K+1} \mathbb{1}\left\{ \text{res}(z_{(K+1)(n-1)+k}, \theta)(-1)^k > 0 \right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{ \text{res}(z_{(K+1)(n-1)+k}, \theta)(-1)^{k+1} > 0 \right\} \right), \tag{5}$$

$$d_S^2(\theta, z_*) := \frac{1}{N-K} \sum_{n=1}^{N-K} \left(\prod_{k=1}^{K+1} \mathbb{1}\left\{ \text{res}(z_{n-1+k}, \theta)(-1)^k > 0 \right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{ \text{res}(z_{n-1+k}, \theta)(-1)^{k+1} > 0 \right\} \right). \tag{6}$$

The depth $d_S^1(\theta, z_*)$ uses only nonoverlapping subsets, while the subsets used in $d_S^2(\theta, z_*)$ are overlapping. In the case $K = 2$, we also consider

$$d_S^3(\theta, z_*) := \frac{1}{\left\lfloor \frac{N-1}{2} \right\rfloor} \sum_{n=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \Big(\mathbb{1}\left\{ \text{res}(z_n, \theta) > 0 \right\} \mathbb{1}\left\{ \text{res}(z_{\lfloor \frac{N+1}{2} \rfloor}, \theta) < 0 \right\} \mathbb{1}\left\{ \text{res}(z_{N-n+1}, \theta) > 0 \right\} + \mathbb{1}\left\{ \text{res}(z_n, \theta) < 0 \right\} \mathbb{1}\left\{ \text{res}(z_{\lfloor \frac{N+1}{2} \rfloor}, \theta) > 0 \right\} \mathbb{1}\left\{ \text{res}(z_{N-n+1}, \theta) < 0 \right\} \Big). \tag{7}$$

For testing, the definitions given by (4), (5), (6), and (7) are sufficient, so that we use them to simplify the notation. However to obtain connected confidence intervals and depth contours, the terms accounting for residuals equal to zero as appearing in (3) should be added.

Theorem 2. *If θ is the underlying parameter with $P_\theta(\text{res}(Z_n, \theta) > 0) = P_\theta(\text{res}(Z_n, \theta) < 0) = \frac{1}{2}$ for all $n \in \{1, \ldots, N\}$, then*

a) $$T_N^1(\theta) := \sqrt{\left\lfloor \frac{N}{K+1} \right\rfloor} \frac{d_S^1(\theta, Z_*) - \left(\frac{1}{2}\right)^K}{\sqrt{\left(\frac{1}{2}\right)^K \left(1 - \left(\frac{1}{2}\right)^K\right)}} \longrightarrow \mathcal{N}(0,1),$$

b) $$T_N^2(\theta) := \sqrt{N-K} \frac{d_S^2(\theta, Z_*) - \left(\frac{1}{2}\right)^K}{\sqrt{(\frac{1}{2})^K \cdot [3 - (\frac{1}{2})^{K-1} \cdot K - 3 \cdot (\frac{1}{2})^K]}} \longrightarrow \mathcal{N}(0,1),$$

c) $$T_N^3(\theta) := \sqrt{\left\lfloor \frac{N-1}{2} \right\rfloor} \frac{d_S^3(\theta, Z_*) - \frac{1}{4}}{\sqrt{\frac{3}{16}}} \longrightarrow \mathcal{N}(0,1),$$

in distribution for $N \to \infty$.

Note that the only assumption needed here for asymptotic normality is that the median of the residuals is zero. The proofs are based on appropriate central limit theorems.

An asymptotic α-level test for a general null hypothesis of the form $H_0 : \theta \in \Theta_0$ is then for any $i \in \{1, 2, 3\}$:

$$\text{reject } H_0 \text{ if } \sup_{\theta \in \Theta_0} T_N^i(\theta) < q_\alpha, \tag{8}$$

where q_α is the α-quantile of the standard normal distribution.

4. CONSISTENCY OF THE TESTS BASED ON SIMPLIFIED SIMPLICIAL DEPTH

Since confidence sets can be constructed from tests for point hypotheses $H_0 : \theta = \theta^0$, we now show the consistency of the tests given by (8) for the case $\Theta_0 = \{\theta^0\}$. Thereby, a test for $H_0 : \theta = \theta^0$ based on $T_N^i(\theta^0)$, $i \in \{1, 2, 3\}$, is consistent at $\theta^* \neq \theta^0$ if

$$\lim_{N \to \infty} P_{\theta^*}\left(T_N^i(\theta^0) < q_\alpha\right) = 1.$$

For linear and nonlinear regression we can consider two different asymptotic scenarios.

> Scenario (A) with finite horizon: There exist $a, b \in I\!R$ such that $a \leq x_{1N} < x_{2N} < \ldots < x_{NN} \leq b$ is satisfied for all $N \in I\!N$. Usually, it is satisfied that $x_{1N} < x_{2N} < \ldots < x_{NN}$ are equidistant points or the deviation from equidistant points is small.
>
> Scenario (B) with infinite horizon: For all $b \in I\!R$, there exists $N_0 \in I\!N$ such that $X_n \geq b$ holds almost surely for $n \geq N_0$. In particular $X_1 < X_2 < \ldots < X_N \longrightarrow \infty$ is satisfied for $N \to \infty$.

For autoregression, only Scenario (B) makes sense.

Note, that the simple sign test, see e.g. Huggins (1989), is usually consistent under Scenario (B) but has consistency problems under Scenario (A) when, for example, the signs of the residuals are positive in the first half of the interval $[a, b]$ and negative in the second half.

Let be $\mathcal{M}_{iN}$ the set of indexes $(n_1, \ldots, n_{K+1})$ used in the simplified simplicial depth d_S^i, for example $\mathcal{M}_{2N} = \{(1, \ldots, K + 1), (2, \ldots, K + 2), \ldots, (N - K, \ldots, N)\}$.

To show consistency of a test for $H_0 : \theta = \theta^0$ at $\theta^* \neq \theta^0$ based on $T_N^i(\theta^0)$, we must show that there exists $\beta < \left(\frac{1}{2}\right)^K$ with $\mathrm{E}_{\theta^*}(d_S^i(\theta^0, Z_*)) \leq \beta$ for almost all N. Sufficient for this is that

$$\mathrm{E}_{\theta^*}\left(\prod_{k=1}^{K+1} \mathbb{1}\left\{\mathrm{res}(Z_{n_k N}, \theta^0)(-1)^k > 0\right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{\mathrm{res}(Z_{n_k N}, \theta^0)(-1)^{k+1} > 0\right\}\right) \tag{9}$$

is less or equal β for almost all subsets $(n_1, \ldots, n_{K+1}) \in \mathcal{M}_{iN}$. Because of

$$\begin{aligned} &\mathrm{res}(Z_n, \theta^0) \lessgtr 0 \Leftrightarrow Y_n - g(X_n, \theta^0) \lessgtr 0 \\ &\quad \Leftrightarrow \;\; Y_n - g(X_n, \theta^*) \lessgtr g(X_n, \theta^0) - g(X_n, \theta^*) \Leftrightarrow E_n \lessgtr g(X_n, \theta^0) - g(X_n, \theta^*), \end{aligned} \tag{10}$$

the behavior of $g(\cdot,\theta^0) - g(\cdot,\theta^*)$ is crucial. However, there are situations depending on the $g(\cdot,\theta^0) - g(\cdot,\theta^*)$ so that the quantity in (9) is not less than $\left(\frac{1}{2}\right)^K$. To see this, consider quadratic regression without linear term given by $Y_n = g(x_n,\theta) + E_n = \theta_0 + \theta_1 x_n^2 + E_n$ with $\theta = (\theta_0,\theta_1)^\top \in I\!R^2$, $K = 2$, and $\theta^0 = (1,0)^\top$, $\theta^* = (1,1)^\top$. Then with (10) the expectation (9) equals

$$P_{\theta^*}(E_{n_1N} < -x_{n_1N}^2, E_{n_2N} > -x_{n_2N}^2, E_{n_3N} < -x_{n_3N}^2) \\ +P_{\theta^*}(E_{n_1N} > -x_{n_1N}^2, E_{n_2N} < -x_{n_2N}^2, E_{n_3N} > -x_{n_3N}^2). \tag{11}$$

This becomes arbitrary small for almost all $(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN}$ in Scenario (B). But in Scenario (A), we could have $x_{n_2N} = 0$, $-x_{n_1N} = x_{n_3N} > 0$ infinitely often. Assuming symmetric errors and setting $P_{\theta^*}(E_{n_1N} > -x_{n_1N}^2) = \frac{1}{2} + \zeta$, probability (11) equals

$$\left(\frac{1}{2}-\zeta\right)\frac{1}{2}\left(\frac{1}{2}-\zeta\right) + \left(\frac{1}{2}+\zeta\right)\frac{1}{2}\left(\frac{1}{2}+\zeta\right) = \frac{1}{4}+\zeta^2 > \left(\frac{1}{2}\right)^K.$$

However, the situation $x_{n_2N} = 0$, $-x_{n_1N} = x_{n_3N} > 0$ happens at most for one subset (n_1,n_2,n_3) for tests based on T_N^1 and T_N^2. Moreover under a monotonicity assumption for $g(x_{n_1N},\theta^0) - g(x_{n_1N},\theta^*),\ldots,g(x_{n_{K+1}N},\theta^0) - g(x_{n_{K+1}N},\theta^*)$, we can prove that the expectation in (9) is bounded by $\beta \le \left(\frac{1}{2}\right)^K$, where in several cases $\beta < \left(\frac{1}{2}\right)^K$ is satisfied. This is shown by the following lemma.

Lemma 1. *Assume Scenario (A) for regression and that the errors $E_1,\ldots,E_N$ have continuous and symmetric distributions around zero. If $c_k := g(x_{n_kN},\theta^0) - g(x_{n_kN},\theta^*)$ for $k \in \{1,\ldots,K+1\}$, $K \in I\!N$, satisfy*

$$c_1 \le c_2 \le \ldots \le c_{K+1} \quad \textit{or} \quad c_1 \ge c_2 \ge \ldots \ge c_{K+1}$$

and $|c_k| \ge c_0 \ge 0$ for all $k \in \{1,\ldots,K+1\}$, then the expectation in (9) is bounded by $\left(\frac{1}{2}\right)^K - \left(\frac{1}{2}\right)^{K-2}\zeta^2$ where $\zeta := \frac{1}{2} - P_{\theta^*}(E_n > c_0) \ge 0$.

Hence for Scenario (A), we have only to ensure that the monotonicity of the c_k in Lemma 1 is satisfied for almost all subsets $(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN}$.

Theorem 3.
a) Assume $i \in \{1,2\}$, Scenario (A) holds for regression, the errors $E_1,\ldots,E_N$ have continuous and symmetric distributions around zero with support given by $I\!R$ and the existence of $c > 0$, $\delta > 0$, and a finite partition $[a,b) = \bigcup_{l=1}^L [a_l,b_l)$ of $[a,b)$ so that $g(\cdot,\theta^0) - g(\cdot,\theta^*)$ is monotone on $[a_l,b_l)$ and $|g(\cdot,\theta^0) - g(\cdot,\theta^*)| \ge c$ on $[a_l+\delta, b_l-\delta) \neq \emptyset$ for $l = 1,\ldots,L$. Then the test for $H_0: \theta = \theta^0$ based on $T_N^i(\theta^0)$ is consistent at θ^*.
b) If Scenario (B) holds for regression or autoregression and there exists $c \neq 0$ with $g(X_n,\theta^0) - g(X_n,\theta^*) = c$ for all $n \in \{1,\ldots,N\}$ and $N \in I\!N$ and $P_{\theta^*}(E_n > c) \neq \frac{1}{2}$, then the test for $H_0: \theta = \theta^0$ based on T_N^i is consistent at θ^* for $i \in \{1,2\}$.
c) If Scenario (B) holds for regression or autoregression and $\lim_{b\to\infty} g(b,\theta^0) - g(b,\theta^*) = \infty$ or $\lim_{b\to\infty} g(b,\theta^0) - g(b,\theta^*) = -\infty$, then the test for $H_0: \theta = \theta^0$ based on T_N^i is consistent at θ^* for $i \in \{1,2,3\}$.

The examples in Section 5 show that the conditions a), b) and c) of Theorem 3 are often satisfied. In particular condition b) is satisfies if θ^0 and θ^* differs only with respect to the intercept of the regression function.

5.1 Polynomial regression

Consider the model

$$y_n = \theta_0 + \theta_1 x_n + \theta_2 x_n^2 + \ldots + \theta_p x_n^p + e_n$$

so that $\theta = (\theta_0, \theta_1, \ldots, \theta_p)^\top \in I\!R^{p+1} = I\!R^K$ with $K = p+1$ and $g(x,\theta) = \theta_0 + \theta_1 x + \theta_2 x^2 + \ldots + \theta_p x^p$. Then

$$v(x,\theta) \;\; = \;\; (1, x, x^2, \ldots, x^p)^\top \in I\!R^{p+1},$$

and

$$w_u(x) = u_1 + u_2 x + u_3 x^2 + \ldots + u_{p+1} x^p \tag{12}$$

is again a polynomial of order p. It is well known that a polynomial of order p has at most p roots so that it has at most $p = K-1$ sign changes and Condition A) of Theorem 1 is satisfied. The roots are determined by $(u_1, \ldots, u_{p+1})^\top$. In particular $(x-\omega_1)\cdot(x-\omega_2)\cdot\ldots\cdot(x-\omega_p)$ is a polynomial of order p with roots at $\omega_1, \ldots, \omega_p$. Hence the roots can be placed at arbitrary locations so that sign changes happen at these roots. This means that also Condition B) of Theorem 1 is satisfied.

Now consider $\theta^* = (\theta_0^*, \theta_1^*, \ldots, \theta_p^*)^\top \neq (\theta_0^0, \theta_1^0, \ldots, \theta_p^0)^\top = \theta^0$. If $\theta_k^0 = \theta_k^*$ for $k \in \{1, \ldots, p\}$, then $\theta_0^* \neq \theta_0^0$ so that the assumptions of Theorem 3 b) are satisfied. Since a constant function is also monotone, also the assumptions of Theorem 3 a) are satisfied. If $\theta_0^* = \theta_0^0$, then $\theta_k^0 \neq \theta_k^*$ for at least one $k \in \{1, \ldots, p\}$ so that $g(\cdot, \theta^0) - g(\cdot, \theta^*)$ is an unbounded function implying that the assumptions of Theorem 3 c) are satisfied. Moreover, the function $g(\cdot,\theta^0) - g(\cdot,\theta^*)$ consists of a finite number of monotone pieces and a finite number of roots on an interval $[a,b]$ implying that the assumptions of Theorem 3 a) are satisfied for this case as well. Hence the test for $H_0 : \theta = \theta^0$ based on T_N^i is consistent at all $\theta^* \neq \theta^0$ for $i \in \{1,2\}$ in Scenario (A) and (B). The test based on T_N^3 is consistent at any θ^* with $\theta_k^0 \neq \theta_k^*$ for at least one $k \in \{1, \ldots, p\}$ for Scenario (B).

5.2 Michaelis-Menten model

The Michaelis-Menten model is a widely used model for enzyme kinetics. In this model, the explanatory variable is the concentration $x_n > 0$ of a substrate and the dependent variable is the reaction rate, denoted by y_n. Assuming independent measurements errors, the model is given by

$$y_n = \frac{\theta_0 x_n}{\theta_1 + x_n} + e_n$$

so that $\theta = (\theta_0, \theta_1)^\top \in (0,\infty)^2 = (0,\infty)^K$ with $K = 2$ and $g(x,\theta) = \frac{\theta_0 x}{\theta_1 + x}$. Data depth for the Michaelis-Menten model already was studied by Van Aelst et al. (2002). However, their depth notion is different from the depth notions used here and no test was provided. Here, we obtain

$$v(x,\theta) = \frac{x}{\theta_1 + x} \left(1, \frac{-\theta_0}{\theta_1 + x}\right)^\top = \frac{x}{\theta_1 + x} \; (1, \tilde{x})^\top$$

with $\tilde{x} = \frac{-\theta_0}{\theta_1+x}$. Since always $\frac{x}{\theta_1+x} > 0$, this factor has no influence on the sign changes of $w_u(x) = u^\top v(x,\theta)$ so that we can consider $\tilde{w}_u(\tilde{x}) = u^\top \tilde{v}(\tilde{x},\theta)$ with $\tilde{v}(\tilde{x},\theta) = (1,\tilde{x})^\top$. Hence $\tilde{w}_u(\tilde{x})$ is of form (12) for $p = 1$, i.e. as for polynomial regression of order 1, so that Conditions A) and B) of Theorem 1 are satisfied for $\tilde{w}_u(\tilde{x})$ according to Section 5.1. But this means that the conditions are also satisfied for $w_u(x)$.

Since $g(\cdot,\theta)$ is a strictly increasing function bounded by θ_0, the assumptions of Theorem 3 c) are not satisfied. Moreover, $g(x_n,\theta^0) - g(x_n,\theta^*) = c$ cannot be satisfied for all $n \in \{1,\ldots,N\}$ if $\theta^* = (\theta_0^*,\theta_1^*)^\top \neq (\theta_0^0,\theta_1^0)^\top = \theta^0$. However, the function $g(\cdot,\theta^0) - g(\cdot,\theta^*)$ consists of a finite number of monotone pieces and a finite number of roots on an interval $[a,b]$ so that the assumptions of Theorem 3 a) follow as for polynomial regression. Hence any test for $H_0 : \theta = \theta^0$ based on T_N^i is consistent at all $\theta^* \neq \theta^0$ for $i \in \{1,2\}$ for Scenario (A).

5.3 Exponential model

Another widely used nonlinear model is the exponential model given by

$$y_n = \theta_1 e^{\theta_2 x_n} + e_n$$

so that $\theta = (\theta_1,\theta_2)^\top \in I\!R^2 = I\!R^K$ with $K = 2$, $g(x,\theta) = \theta_1 e^{\theta_2 x}$, and

$$v(x,\theta) \;=\; e^{\theta_2 x}\;(1\,,\,\theta_1 x)^\top \in I\!R^2 = e^{\theta_2 x}\;\tilde{v}(x,\theta)$$

with $\tilde{v}(x,\theta) = (1\,,\,\theta_1 x)^\top$. Since always $e^{\theta_2 x_n} > 0$ we again can work with $\tilde{v}(x,\theta)$ instead of $v(x,\theta)$. Then we get

$$w_u(x) = u_1 + u_2\theta_1 x = \tilde{u}_1 + \tilde{u}_2 x = \tilde{w}_{\tilde{u}}(x)$$

with $\tilde{u}_1 = u_1$ and $\tilde{u}_2 = u_2\theta_1$. Since $\tilde{w}_{\tilde{u}}(x)$ is the $w_u(x)$ in (12) for $p = 1$, i.e. as for polynomial regression of order 1, the Conditions A) and B) of Theorem 1 again follow from Section 5.1.

The consistency of a test for $H_0 : \theta = \theta^0$ based on T_N^i at all $\theta^* \neq \theta^0$ and $i \in \{1,2\}$ for Scenario (A) follows as for polynomial regression and the Michaelis-Menten model. The assumptions of Theorem 3 b) are never satisfied. However, the assumptions of Theorem 3 c) hold if $\theta_2^* > 0$ or $\theta_2^0 > 0$ is satisfied so that a test for $H_0 : \theta = \theta^0$ based on T_N^i is consistent at all $\theta^* = (\theta_1^*,\theta_2^*)^\top \neq (\theta_1^0,\theta_2^0)^\top = \theta^0$ with $\theta_2^* > 0$ or $\theta_2^0 > 0$ for $i \in \{1,2,3\}$ in Scenario (B).

5.4 Nonlinear polynomial model I

The derivation of Conditions A) and B) of Theorem 1 is not always possible via the polynomial regression model treated in Section 5.1. An example is the nonlinear polynomial model given by

$$y_n = \theta_0 + \theta_1 x_n^{\theta_2} + e_n$$

so that $\theta = (\theta_0,\theta_1,\theta_2)^\top \in I\!R^3 = I\!R^K$ with $K = 3$, $g(x,\theta) = \theta_0 + \theta_1 x^{\theta_2}$, and

$$v(x,\theta) \;=\; (1, x^{\theta_2}, \theta_1 x^{\theta_2}\,\log(x))^\top \in I\!R^3.$$

This leads to

$$w_u(x) = u_1 + u_2 x^{\theta_2} + u_3\,\theta_1 x^{\theta_2}\,\log(x). \tag{13}$$

For deriving the Conditions A) and B) of Theorem 1, the following lemma is necessary.

Lemma 2. *If $\theta_1 \neq 0, \theta_2 \neq 0$, then $w_u : [0,\infty) \to I\!R$ given by (13) has the following properties:*
a) w_u has exactly one extremum at $x = exp\left(-\frac{1}{\theta_2} - \frac{u_2}{u_3\theta_1}\right)$ for all $u = (u_1, u_2, u_3)^\top \in I\!R^3$ with $u_3 \neq 0$.
b) For all $0 < \xi_1 < \xi_2$, there exists a vector $u_+ \in I\!R^3$ with $w_{u_+}(\xi_1) = w_{u_+}(\xi_2) = 0$ and $w_{u_+}(x) > 0$ for all $x \in (\xi_1, \xi_2)$ and a vector $u_- \in I\!R^3$ with $w_{u_-}(\xi_1) = w_{u_-}(\xi_2) = 0$ and $w_{u_-}(x) < 0$ for all $x \in (\xi_1, \xi_2)$.

If $\theta_1 = 0$ or $\theta_2 = 0$, then clearly w_u has at most one sign change on $[0,\infty)$ for all $u \in I\!R^3$. If $\theta_1 \neq 0, \theta_2 \neq 0$, then w_u has exactly one extremum according to Lemma 2 a), which means that w_u can have at most $2 = K-1$ sign changes on $[0,\infty)$. Hence Condition A) of Theorem 1 is satisfied for all $\theta \in I\!R^3$.

However, to show Condition B) of Theorem 1, we must exclude the cases $\theta_1 = 0$ and $\theta_2 = 0$. But this excludes only the case of a constant function, i.e. the model $y_n = \theta_0 + e_n$. Hence we assume $\theta_1 \neq 0, \theta_2 \neq 0$.

Now regard any $0 \leq x_1 < x_2 < x_3 < x_4$ and any $s \in \{-1,1\}^4$ with at most $K-1 = 2$ sign changes. If $s = (s_1, s_2, s_3, s_4)^\top$ has $K-1 = 2$ sign changes, the missing possible third sign change is between s_k and s_{k+1} with $k=1$, $k=2$, or $k=3$.

For $k=1$, set $\xi_1 \in (x_2, x_3)$, $\xi_2 \in (x_3, x_4)$. Then $x_3 \in (\xi_1, \xi_2)$ and according to Lemma 2 b), there exists $u_0 \in I\!R^3$ with $w_{u_0}(\xi_1) = 0 = w_{u_0}(\xi_2)$ and sgn$(w_{u_0}(x_3)) = s_3$. Since w_{u_0} has exactly one extremum according to Lemma 2 a), w_{u_0} has only sign changes at ξ_1 and ξ_2 so that there is a sign change between $w_{u_0}(x_2)$ and $w_{u_0}(x_3)$ as well as between $w_{u_0}(x_3)$ and $w_{u_0}(x_4)$, and no sign change between $w_{u_0}(x_1)$ and $w_{u_0}(x_2)$ so that sgn$(w_{u_0}(x_k)) = s_k$ for $k \in \{1,2,3,4\}$.

Using $\xi_1 \in (x_1, x_2)$ and $\xi_2 \in (x_3, x_4)$ for $k=2$ and $\xi_1 \in (x_1, x_2)$ and $\xi_2 \in (x_2, x_3)$ for $k=3$ provides, with the same arguments as for $k=1$, the existence of u_0 with sgn$(w_{u_0}(x_k)) = s_k$ for $k \in \{1,2,3,4\}$. The case that s has less than $K-1 = 2$ sign changes can be treated with similar arguments. Hence Condition B) of Theorem 1 is satisfied.

If $\theta_0^* \neq \theta_0^0$ and $\theta_k^* = \theta_k^0$ for $k \in \{1,2\}$, then the assumptions of Theorem 3 a) and b) are satisfied. Otherwise we have the same situation as for the exponential model. In particular the assumptions of Theorem 3 c) are only satisfied for $\theta^* = (\theta_0^*, \theta_1^*, \theta_2^*)^\top$ and $\theta^0 = (\theta_0^0, \theta_1^0, \theta_2^0)^\top$ with $\theta_2^* > 0$ or $\theta_2^0 > 0$.

5.5 Nonlinear polynomial model II

Another nonlinear polynomial model is given by

$$y_n = \theta_1(\theta_2 - x_n)^{\theta_3} + e_n$$

so that $\theta = (\theta_1, \theta_2, \theta_3)^\top \in I\!R^3 = I\!R^K$ with $K = 3$, $g(x, \theta) = \theta_1(\theta_2 - x)^{\theta_3}$, and

$$v(x, \theta) \;=\; (\theta_2 - x)^{\theta_3 - 1} \; (\theta_2 - x\,,\, \theta_1\,\theta_3\,,\, \theta_1\,(\theta_2 - x)\,\log(\theta_2 - x))^\top \in I\!R^3 = (\theta_2 - x)^{\theta_3 - 1}\,\tilde{v}(x, \theta)$$

with $\tilde{v}(x, \theta) = (\theta_2 - x\,,\, \theta_1\,\theta_3\,,\, \theta_1\,(\theta_2 - x)\,\log(\theta_2 - x))$. We here assume that $0 \le x_n < \theta_2$, where in particular $(\theta_2 - x_n)^{\theta_3 - 1} > 0$ holds for all $n \in \{1, \ldots, K + 1\}$. Hence we can work with $\tilde{v}(x, \theta)$ instead of $v(x, \theta)$ so that

$$w_u(x) = u_1(\theta_2 - x) + u_2\theta_1\,\theta_3 + u_3\,\theta_1\,(\theta_2 - x)\,\log(\theta_2 - x) = \tilde{w}_u(\tilde{x})$$

with $\tilde{x} = \theta_2 - x \in (0, \theta_2]$ and $\tilde{w}_u(\tilde{x}) = u_1\tilde{x} + u_2\theta_1\,\theta_3 + u_3\,\theta_1\,\tilde{x}\,\log(\tilde{x})$. This $\tilde{w}_u$ is of the form of w_u in (13) with $\theta_2 = 1$ so that the result in Section 5.4 provides that the Conditions A) and B) of Theorem 1 are also satisfied here.

Again, the assumptions of Theorem 3 a) are always satisfied. The specialty of this model is that we have explosion for $\theta_3 < 0$ when $x_N \to \theta_2$. This is not Scenario (B). However, the proof of Theorem 3 c) holds also for this case so that consistency of a test for $H_0 : \theta = \theta^0$ based on T_N^i follows at all $\theta^* = (\theta_1^*, \theta_2^*, \theta_3^*)^\top \neq (\theta_1^0, \theta_2^0, \theta_3^0)^\top = \theta^0$ with $\theta_3^* < 0$ or $\theta_3^0 < 0$ for $i \in \{1, 2, 3\}$.

5.6 AR(1) growth models

General linear and nonlinear AR(1) growth processes are given by

$$y_n = g(y_{n-1}, \theta) + e_n$$

or

$$y_n = y_{n-1} + g(y_{n-1}, \theta) + e_n \tag{14}$$

with $y_0 < y_1 < \ldots < y_{p+1}$, where $g(x, \theta)$ is of the same form as in the examples of Sections 5.1 to 5.5. Setting $\tilde{y}_n = y_n - y_{n-1}$ in the second case and $x_n = y_{n-1}$, the Conditions A) and B) of Theorem 1 are also satisfied according to Sections 5.1 to 5.5. The results for consistency under Scenario (B) of Sections 5.1 to 5.5 transfer to these processes as well.

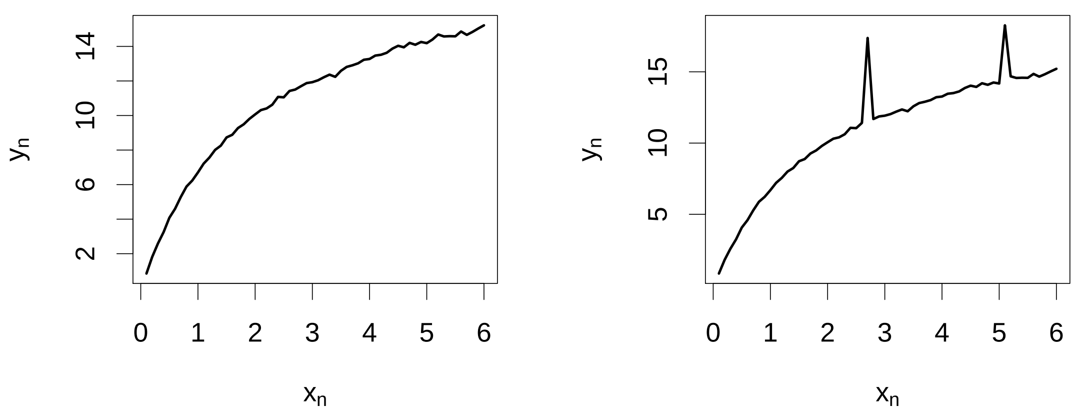
Model (14) appears in particular when the Euler-Maruyama approximation is used for stochastic differential equations, see e.g. Iacus (2008). An example of strictly increasing observations y_n is crack growth where the function $g(x, \theta)$ of Section 5.4 provides a stochastic version of the Paris-Erdogan equation which is widely used in engineering sciences, see Pook (2000). In Kustosz and Müller (2014) the one-parameter case with $\theta_0 = 0$ and $\theta_3 = 1$ was studied. Here the two-parameter case of the nonlinear function of Section 5.4 with $\theta_0 = 0$ is considered in the simulation study of Section 6.2 and in the application of Section 7. A simulation study for the linear version with $\theta_3 = 1$ is provided by Kustosz et al. (2015).

6. SIMULATION STUDY

In this section, the finite sample behavior of the simplified simplicial depth tests for $H_0 : \theta = \theta^0$ is studied in two models. All calculations were performed in *R*, see R Core Team (2014). In all examples, the simplified depth tests are compared with the simple sign test which is based on the numbers of positive residuals. For a comparison with other tests see Kustosz et al. (2015), where the test based on the full simplicial depth based, see (4), and the test defined by the Ordinary Least Squares estimator are considered. Further two types of errors, namely $N(0, 0.1)$ error variables, to evaluate the tests under standard assumptions, and errors defined by $(1-\epsilon)N(0, 0.1) + \epsilon N(5, 1)$ with $\epsilon = 0.05$, to simulate a skewed and contaminated error distribution, are considered.

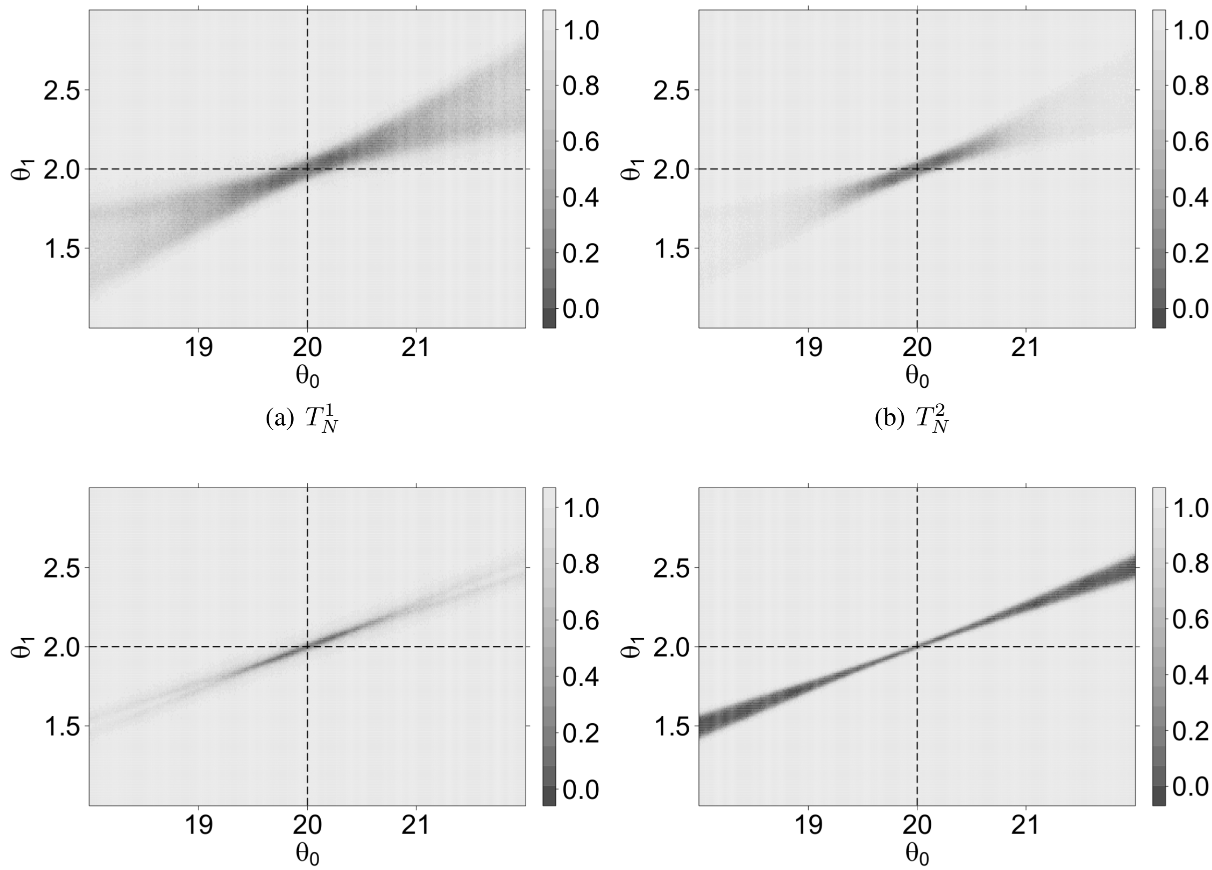
6.1 Michaelis-Menten model

The first example evaluates the resulting depth based tests for a Michaelis-Menten model. In Figure 1 (a) an example process with parameter $\theta^0 = (\theta^0_0, \theta^0_1)^\top = (20, 2)^\top$ is depicted. An exemplary process with contaminated errors is illustrated in Figure 1 (b). To satisfy the conditions of Scenario (A), x_n is fixed by $x_1 = 0 < x_2 = x_1 + 0.1 < ... < x_{60} = 6$, so that $N = 61$. To analyze the power, the tests are evaluated for $H_0 : \theta = \theta^0$ and processes with parameters on a grid defined by $[18, 22] \times [1, 3]$ for θ_0 and θ_1 with step width 0.01 are simulated. The tests are evaluated 100 times for each parameter on the grid. The simulated power is then defined by the relative number of rejections. The results for $N(0, 0.1)$ errors on a level of $\alpha = 0.05$ are shown in Figure 2.



(a) $N(0, 0.1)$ Errors (b) Contaminated $N(0, 0.1)$ Errors

Figure 1. Simulated Processes from the Michaelis-Menten Model. Two exemplary simulations from the Michaelis-Menten model with $\theta_0 = 20, \theta_1 = 2$ observed at $x = (0, 0.1, 0.2, ..., 5.9, 6)$ with two different error distributions.

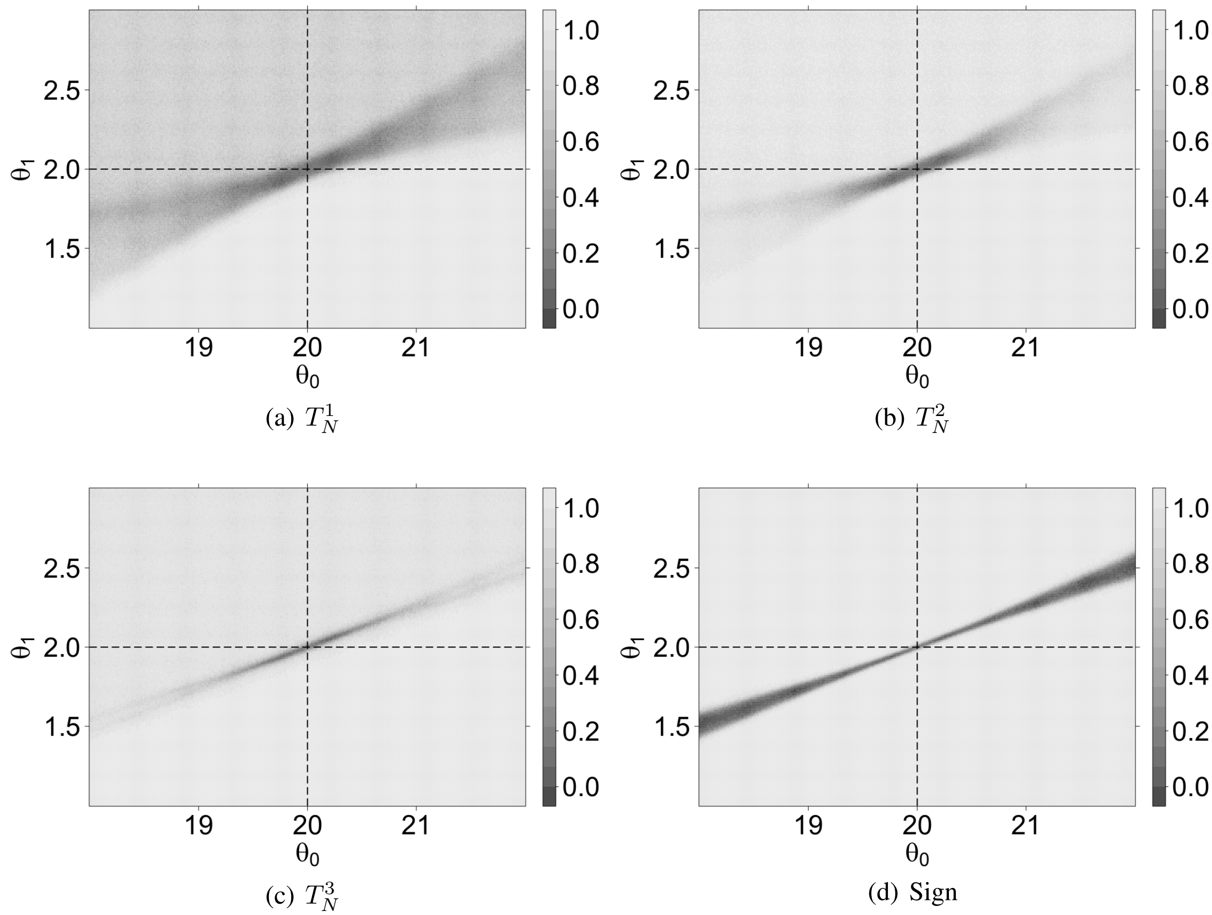


(a) T_N^1 (b) T_N^2

(c) T_N^3 (d) Sign

Figure 2. Simulated Power for Different Tests with $N(0, 0.1)$ Errors under the Michaelis-Menten Model. The simulated relative number of rejections of $H_0 : \theta = (20, 2)$ based on different values of $\theta = (\theta_0, \theta_1)$ is depicted. The errors are simulated as $N(0, 0.1)$ random variables. The parameters for the null hypothesis are marked by the dashed lines.

The simplicial depth tests have power functions, which are increasing when the parameter deviates from $H_0 : \theta = (20, 2)$. This is also true for the sign test. However, due to the fixed sample size the sign test has a wide range of not rejected parameters if half of the data is under and half of the data is overestimated. Since these fits can be arbitrary chosen when one residual is fixed in the middle of the dataset, there is an unbounded set of such parameters. This problem does not appear for the simplicial depth statistics. In a direct comparison, the T_N^3 test performs best measured by the number of 11 parameters on the grid with simulated power below $\alpha = 0.05$, followed by the T_N^2 with 67 and T_N^1 with 196 parameters with power below of the 5% level. When counting the number of grid points on which the tests are uniquely best, measured by their power, we have 8093 points for T_N^3 and 3627 points for T_N^2 followed by just 1 point for T_N^1. This shows, that T_N^3 outperforms T_N^2 when parameters with larger distance to H_0 are tested. Note, that although we were not able to prove the consistency of the T_N^3 test it appears to be a valid test for $H_0 : \theta = \theta^0$ as well.
We also evaluate the tests in case of skewed and nonnormal errors. The resulting power functions are given in Figure 3.



(a) T_N^1 (b) T_N^2 (c) T_N^3 (d) Sign

Figure 3. Simulated Power for Contaminated $N(0, 0.1)$ Errors under the Michaelis-Menten Model. The simulated relative number of rejections of $H_0 : \theta = (20, 2)$ based on different values of $\theta = (\theta_0, \theta_1)$ is depicted. The errors are simulated as contaminated $N(0, 0.1)$ random variables, whereby in a fraction of 5% $N(5, 1)$ variables are added. The parameters for the null hypothesis are marked by the dashed lines.

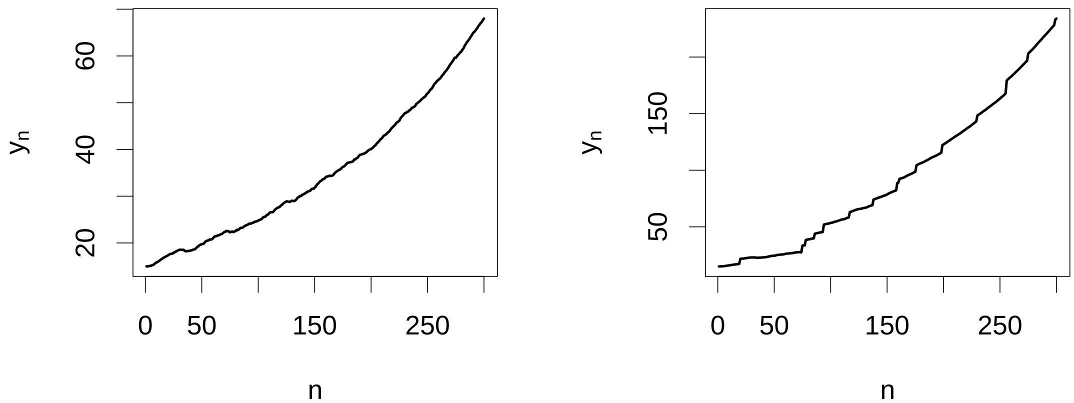
The results are similar to the noncontaminated case. Due to the contamination the power decreases slightly for values of $\theta_2 > \theta_2^0$ but still leads to rejection in most cases. The number of grid points with power below of the 5% level is 17 for T_N^3, 75 for T_N^2 and 229 for T_N^1. Here the number of grid points with unique superior power shows another result. The T_N^2 test is best at 33825 grid points, followed by the T_N^3 test with 10784 and the T_N^1 with 1 point. The sign test shows systematic problems again but is less influenced by the contamination.

6.2 Nonlinear AR(1) growth model

In the second example, the model

$$y_n = y_{n-1} + \theta_1 y_{n-1}^{\theta_2} + e_n,$$

which plays an important role in modeling crack growth, see Section 5.6 and Section 7, is considered. Here the null hypothesis $H_0 : \theta = (\theta_1^0, \theta_2^0)$ with $\theta_1^0 = 0.005$ and $\theta_2^0 = 1.002$ is tested. Processes with a starting value $y_0 = 15$ and $N = 500$ observations are examined. Example processes under H_0 are presented in Figure 4.

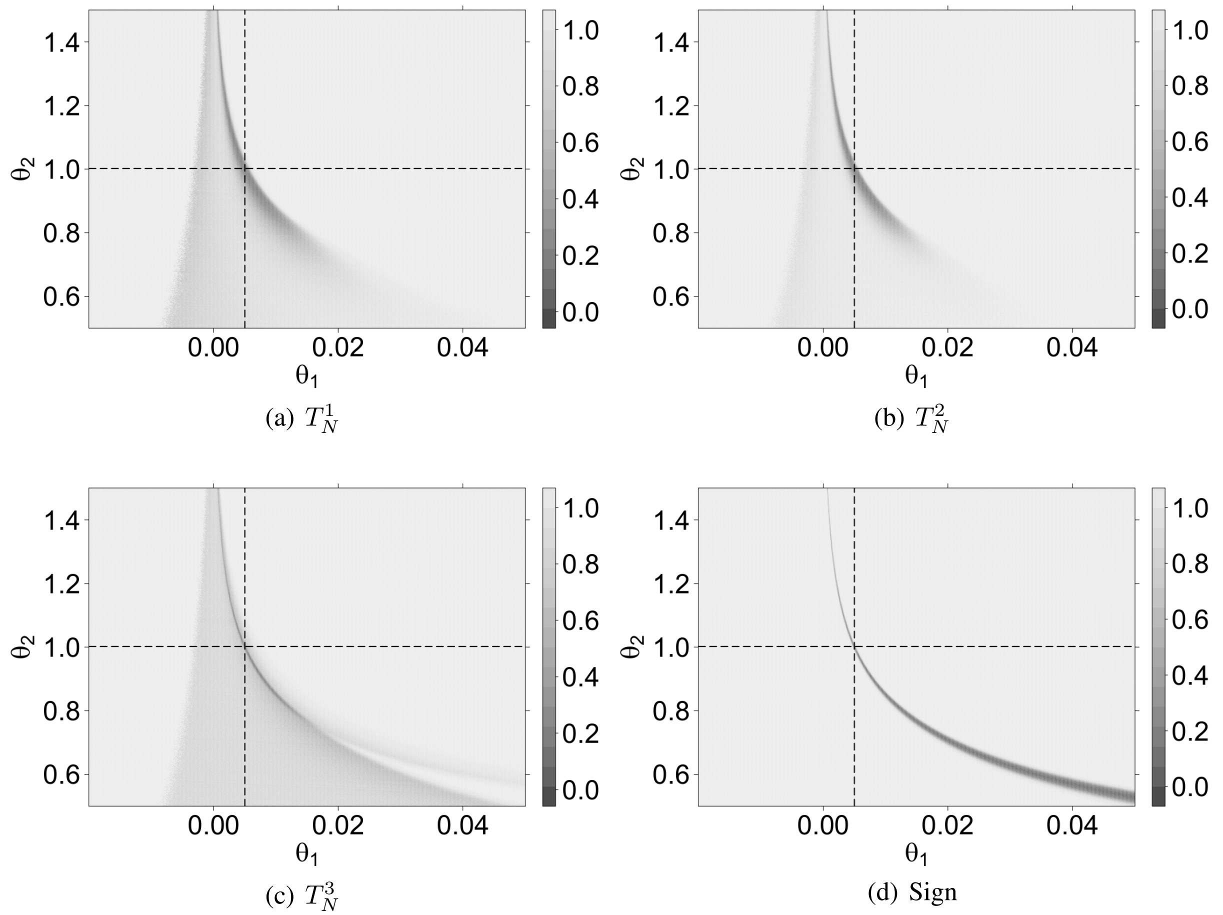


(a) $N(0, 0.1)$ Errors

(b) Contaminated $N(0, 0.1)$ Errors

Figure 4. Simulated Nonlinear AR(1) Processes. Realizations of nonlinear AR(1) processes with $\theta_1 = 0.005, \theta_2 = 1.002, y_0 = 15$ and two different error distributions are depicted.

To evaluate the power of tests for $H_0 : \theta = \theta^0 := (\theta_1^0, \theta_2^0)^\top$, a grid defined by $[-0.02, 0.05] \times [0.5, 1.5]$ with a step width of 0.0001 for θ_1 and 0.001 for θ_2 is considered. On each grid point the processes are generated 100 times to simulate the power of the test at a 5% level for processes with a length of $N = 500$ observations. The resulting power functions for normal errors are depicted in Figure 5.



(a) T_N^1 (b) T_N^2

(c) T_N^3 (d) Sign

Figure 5. Simulated Power for $N(0, 0.1)$ Errors under the Nonlinear AR(1) Model. The simulated relative number of rejections of $H_0 : \theta = (0.005, 1.002)$ based on different values of $\theta = (\theta_0, \theta_1)$ is depicted. The errors are simulated as $N(0, 0.1)$ random variables. The parameters for the null hypothesis are marked by the dashed lines.

One can observe, that the depth based tests have power functions, which are increasing to one when the parameter deviates from $H_0 : \theta = (0.005, 1.002)$. Due to the model, the power functions are not symmetric. It is hard to compare, which test is best, but by counting the number of parameters with power below of a 5% level we see that the T_N^3 with 81 points outperforms the T_N^1 test with 270 points followed by the T_N^2 version with 303 points. By counting the fractions with unique best power we have 145263 points for T_N^2, 9475 points for T_N^3 and 120 points for T_N^1. By consideration of a wider parameter range, a systematic shortcoming of the sign test gets obvious. The sign test again does not reject parameters, for which half of the residuals are negative and half are positive, even if the model fit is poor. The residuals of a process $\tilde{Y}$ defined by $\theta = (\theta_1, \theta_2) \neq (\theta_1^0, \theta_2^0) = \theta^0$ are given by $r_n(\theta^0, \tilde{Y}) = E_n + \theta_1 \tilde{Y}_{n-1}^{\theta_2} - \theta_1^0 \tilde{Y}_{n-1}^{\theta_2^0}$. If the errors are assumed to be approximately zero, then $r_n(\theta^0, \tilde{Y}) \lesseqgtr 0$ holds approximately if and only if $\theta_1 \lesseqgtr \theta_1^0 \tilde{Y}_{n-1}^{\theta_2^0 - \theta_2}$. Since $\tilde{Y}$ is strictly increasing, we obtain for $\theta_2 < \theta_2^0$ that

$$\theta_1^0 \tilde{Y}_0^{\theta_2^0 - \theta_2} < \dots < \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor - 1}^{\theta_2^0 - \theta_2} < \theta_1 < \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor}^{\theta_2^0 - \theta_2} < \dots < \theta_1^0 \tilde{Y}_N^{\theta_2^0 - \theta_2}$$

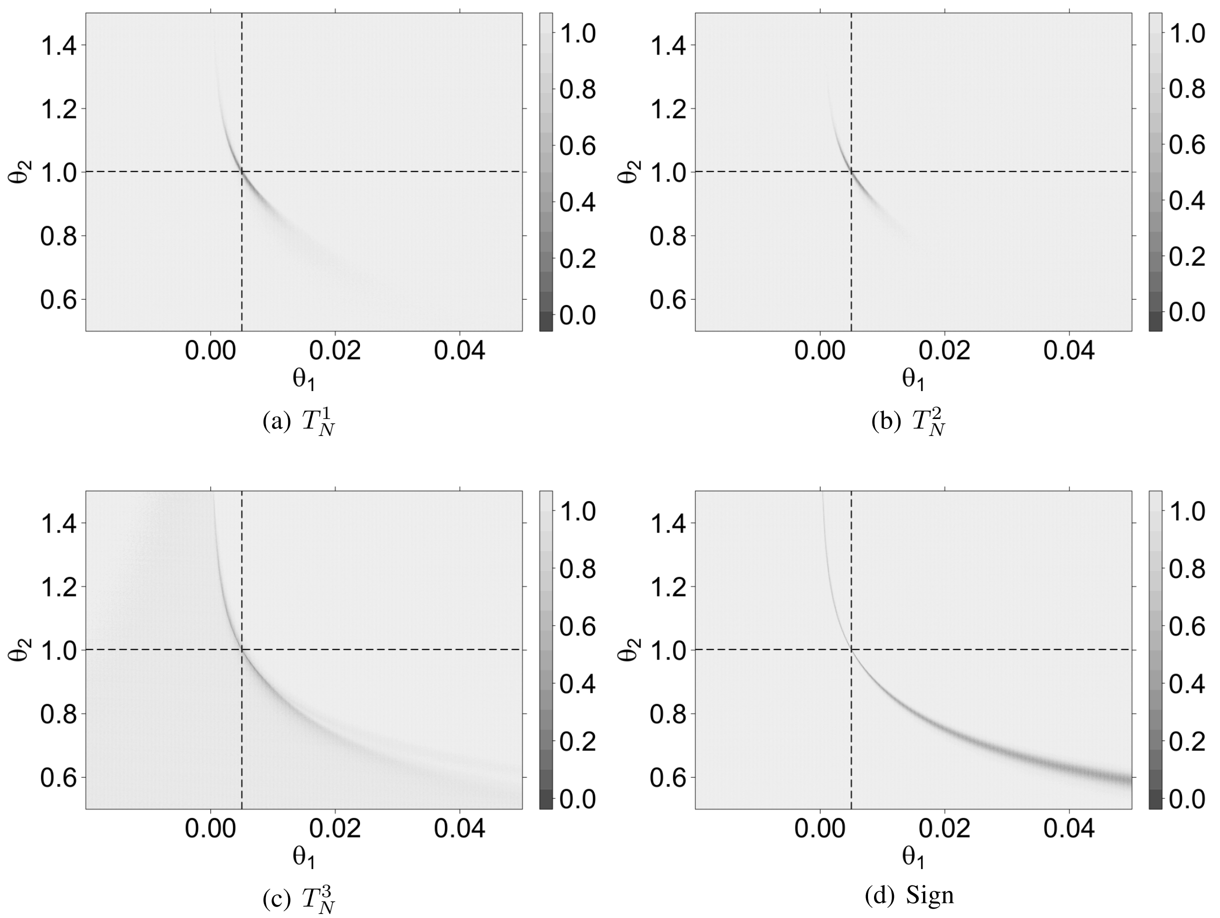
implies $r_n(\theta^0, \tilde{Y}) > 0$ for $n \in \{1, \dots, \lfloor N/2 \rfloor\}$ and $r_n(\theta^0, \tilde{Y}) < 0$ for $n \in \{\lfloor N/2 \rfloor + 1, \dots, N\}$.

Similarly, if $\theta_2 > \theta_2^0$ then

$$\theta_1^0 \tilde{Y}_0^{\theta_2^0-\theta_2} > ... > \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor -1}^{\theta_2^0-\theta_2} > \theta_1 > \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor}^{\theta_2^0-\theta_2} > ... > \theta_1^0 \tilde{Y}_N^{\theta_2^0-\theta_2}$$

implies $r_n(\theta^0, \tilde{Y}) < 0$ for $n \in \{1, ..., \lfloor N/2 \rfloor\}$ and $r_n(\theta^0, \tilde{Y}) > 0$ for $n \in \{\lfloor N/2 \rfloor + 1, ..., N\}$. For $\theta_2 \to \infty$, the interval $\left[\theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor}^{\theta_2^0-\theta_2}, \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor -1}^{\theta_2^0-\theta_2}\right]$ reduces to one point, so that only few θ_1 can satisfy $\theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor -1}^{\theta_2^0-\theta_2} > \theta_1 > \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor}^{\theta_2^0-\theta_2}$ for large θ_2. The opposite is the case for $\theta_2 \to 0$, where the interval $\left[\theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor -1}^{\theta_2^0-\theta_2}, \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor}^{\theta_2^0-\theta_2}\right]$ becomes larger, explaining the widening of the area with low power of the sign test for small θ_2.

For an error distribution which is contaminated with positive outliers in 5% of all cases, the resulting power functions are presented in Figure 6. As in the noncontaminated case the region of the depth based tests with low power is bounded while the sign test shows a systematic problem for a range of parameters with small θ_2. In general the power functions are steeper, since the jumps lead to a faster growing process, what is exploited by the proposed tests. The number of parameters below the 5% level are 2 for T_N^3 and 21 for T_N^1 and T_N^2. The best power is achieved at 18891 points for T_N^2, 683 points for T_N^3 and 37 points for T_N^1.



(a) T_N^1 (b) T_N^2

(c) T_N^3 (d) Sign

Figure 6. Simulated Power for Contaminated $N(0, 0.1)$ Errors under the Nonlinear AR(1) Model. The simulated relative number of rejections of $H_0 : \theta = (0.005, 1.002)$ based on different values of $\theta = (\theta_0, \theta_1)$ is depicted. The errors are simulated as contaminated $N(0, 0.1)$ random variables, whereby in a fraction of 5% variables with a $N(5, 1)$ distribution are added. The parameters for the null hypothesis are marked by the dashed lines.

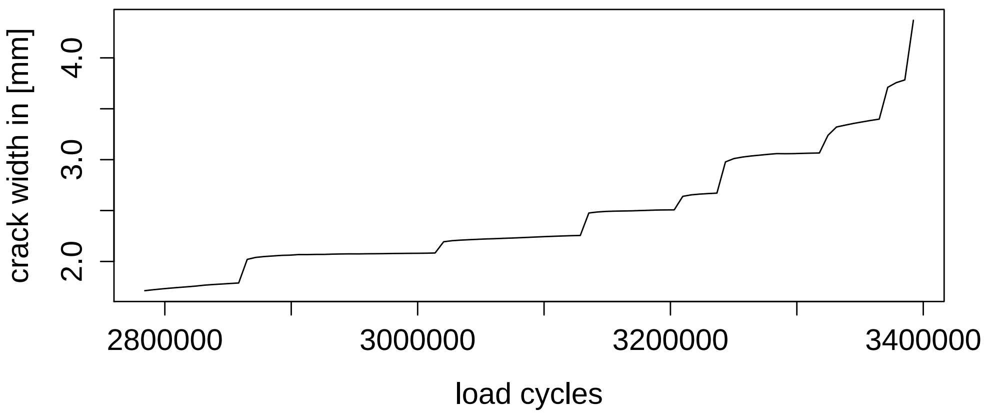


Figure 7. Observed crack growth in prestressed concrete in mm. The observations are recorded at $n = 1, ..., 91$ steps with differences of 6756 load cycles.

Figure 7 shows the growth of the crack width y_n in prestressed concrete in an experiment conducted by Maurer and Heeke (2010), where n denotes the observation index of $n = 1, ..., 91$ discretely observed values at load cycles from 2784104 to 3392189 in steps of 6765. The jumps, visible in the crack growth process, are caused by the breaking of the tension wires and can be considered as outliers although they are innovation outliers. In Kustosz and Müller (2014), the AR(1) model

$$y_n = y_{n-1} + \theta_1 y_{n-1}^{\theta_2} + e_n,$$

with known $\theta_2 = 1$ was used to model the growth of the crack width y_n without the jumps. However, a more realistic description of the process should allow θ_2 to be unknown. Figure 8 illustrates y_n dependent on y_{n-1}. As can be seen in this Figure, the fit of the nonlinear least squares (NLS) estimator given by $(\widehat{\theta}_1, \widehat{\theta}_2) = (1.2 \cdot 10^{-8}, 12.05)$ is quite well for low increments but also influenced by the large increments caused by jumps. Confidence regions based on d_S^1 and d_S^2 can be computed by grid search. Fits constructed by parameters in the 95$\%$ confidence region from the d_S^1 test statistic are given by the dashed lines in Figure 8. Due to a structural change of the dynamics, this confidence region consists of two parts which are not connected, namely a part around $(0.0001, 5.97350)$ and a part around $(0.000115, 2.654)$. However the two parts are so small, that the corresponding curves only appear as two slightly different curves in Figure 8. Both lines are not influenced by the jumps and show good fits for large increments as well.

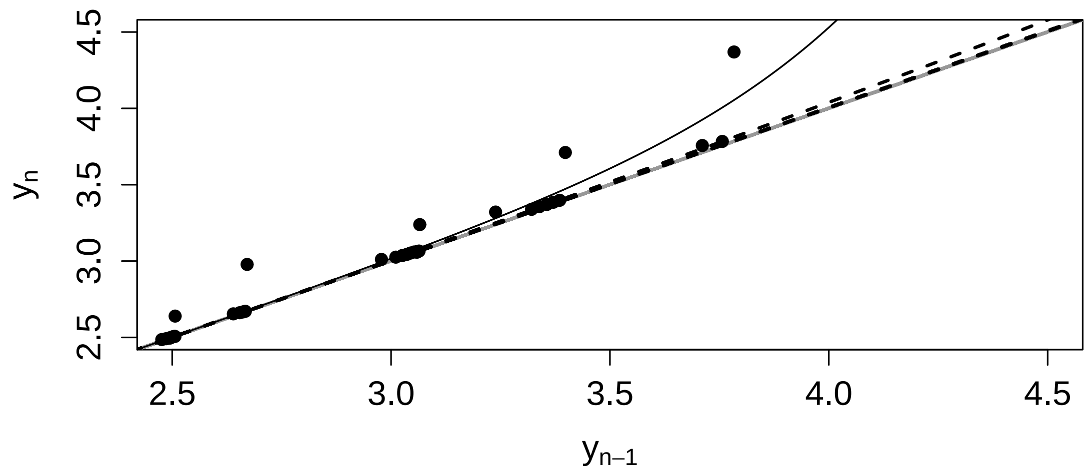


Figure 8. Fit of y_n against y_{n-1} plotted for $y_n \geq 2.25$. The dots represent the observed pairs (y_n, y_{n-1}). The solid black line is the fit based on the NLS estimator. The dashed black lines show fits resulting from parameters in the 95% confidence region based on d_S^1 and the grey line shows the fit calculated by the parameters from the 99.9% confidence region based on d_S^2.

Although the depth statistics improve the robustness of the parameter estimates for the crack growth process, they also react to a nonhomogeneous structure of the data. The d_S^2 test statistic is in this sense more sensible to local deviations of the model assumptions, so that the corresponding confidence regions do not contain any parameters up to a level of 99.9%. The resulting fit from parameters in the 99.9% confidence set based on d_S^2 is presented by the light grey curve in Figure 8. Again the confidence set, an area around $(0.00074, 0)$, is so small that the corresponding curves appear as one line. This line is also not influenced by the jumps.

ACKNOWLEDGEMENTS

We gratefully thank Prof. Dr.-Ing. Reinhard Maurer and Dipl.-Ing. Guido Heeke form the Chair of Concrete Construction of the TU Dortmund University for providing the data for the application in Section 7.

References

Agostinelli, C., and Romanazzi, M. (2011), "Local Depth," *Journal of Statistical Planning and Inference*, 141, 817-830.

Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014), "Multivariate Functional Halfspace Depth," *Journal of the American Statistical Association*, 109, 411-423.

Denecke, L., and Müller, Ch. H. (2011), "Robust Estimators and Tests for Copulas Based on Likelihood Depth," *Computational Statistics and Data Analysis*, 55, 2724-2738.

——(2012), "Consistency and Robustness of Tests and Estimators Based on Depth," *Journal Statistical Planning and Inference*, 142, 2501-2517.

——(2013), "New Robust Tests for the Parameters of the Weibull Distribution for Complete and Censored Data," *Metrika*, DOI:10.1007/s00184-013-0454-8.

——(2014), "Consistency of the Likelihood Depth Estimator for the Correlation Coefficient," *Statistical Papers*, 55, 3-13.

Dümbgen, L. (1992), "Limit Theorems for the Simplicial Depth," *Statistics and Probability Letters*, 14, 119-128.

Dutta, S., and Ghosh, A. K. (2012), "On Robust Classification Using Projection Depth," *Annals of the Institute of Statistical Mathematics*, 64, 657-676.

Hallin, M., Paindaveine, D., and Siman, M. (2010), "Multivariate Quantiles and Multiple-Output Regression Quantiles: From L1 Optimization to Halfspace Depth," (with discussion), *The Annals of Statistics*, 38, 635-669.

Huggins, R. (1989), "The Sign Test for Stochastic Processes," *Australian and New Zealand Journal of Statistics*, 31, 153-165.

Hoeffding, W., and Robbins, H. (1948), "The Central Limit Theorem for Dependent Random Variables," *Duke Mathematical Journal*, 15, 773-780.

Iacus, S. (2008), *Simulation and Inference for Stochastic Differential Equations*, Springer, New York.

Kong, L., and Zuo, Y. (2010), "Smooth Depth Contours Characterize the Underlying Distribution," *Journal of Multivariate Analysis*, 101, 2222-2226.

Kustosz, C. P., and Müller, Ch. H. (2014), "Analysis of Crack Growth with Robust Distribution-Free Estimators and Tests for Nonstationary Autoregressive Processes," *Statistical Papers*, 55, 125-140.

Kustosz, C. P., Leucht, A. and Müller, Ch. H. (2015), "Tests Based on Simplicial Depth for AR(1) Models with Explosion,". Submitted.

Li, J., Cuesta-Albertos, J. A., and Liu, R. Y. (2012), "DD-Classifier: Nonparametric Classification Procedure Based on DD-plot," *Journal of the American Statistical Association*, 107, 737-753.

Lin, L., and Chen, M. (2006), "Robust Estimating Equation Based on Statistical Depth," *Statistical Papers*, 47, 263-278.

Liu, R. Y. (1988), "On a Notion of Simplicial Depth," *Proceedings of the National Academy of Sciences of the United States of America*, 85, 1732-1734.

——(1990), "On a Notion of Data Depth Based on Random Simplices," *The Annals of Statistics*, 18, 405-414.

Lok, W. S., and Lee, S. M. S. (2011), "A New Statistical Depth Function With Application to Multimodal Data," *Journal of Nonparametric Statistics*, 23, 617-631.

López-Pintado, S., and Romo, J. (2009), "On the Concept of Depth for Functional Data," *Journal of the American Statistical Association*, 104, 718-734.

Maurer, R., and Heeke, G. (2010), "Ermüdungsfestigkeit von Spannstählen aus einer älteren Spannbetonbrücke," *Bericht für Landesbetrieb Straßenbau NRW*. Technical report, TU Dortmund.

Mizera, I. (2002), "On Depth and Deep Points: A Calculus," *The Annals of Statistics*, 30, 1681-1736.

Mizera, I., and Müller, Ch. H. (2004), "Location-Scale Depth" (with discussion), *Journal of the American Statistical Association*, 99, 949-966.

Mosler, K. (2002), *Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach*, Lecture Notes in Statistics, 165, Springer, New York.

Müller, Ch. H. (2005), "Depth Estimators and Tests Based on the Likelihood Principle with Application to Regression," *Journal of Multivariate Analysis*, 95, 153-181.

Paindaveine, D., and van Bever, G. (2013), "From Depth to Local Depth: A Focus on Centrality," *Journal of the American Statistical Association*, 108, 1105-1119.

Pook, L. (2000), *Linear Elastic Fracture Mechanics for Engineers: Theory and Application*, WIT Press, Southampton.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.

Rousseeuw, P. J., and Hubert, M. (1999), "Regression Depth," *Journal of the American Statistical Association*, 94, 388-433.

Rousseeuw, P. J., and Ruts, I. (1999), "The Depth Function of a Population Distribution," *Metrika*, 49, 213-244.

Tukey, J. W. (1975), "Mathematics and the Picturing of Data," in *Proceedings ot the International Congress of Mathematicians*, 2, 523- 531.

Van Aelst, S., Rousseeuw, P. J., Hubert, M., and Struyf, A. (2002), "The Deepest Regression Method," *Journal of Multivariate Analysis*, 81, 138-166.

Wellmann, R., Harmand, P., and Müller, Ch. H. (2009), ”Distribution-Free Tests for Polynomial Regression Based on Simplicial Depth,” *Journal of Multivariate Analysis*, 100, 622-635.

Wellmann, R., and Müller, Ch. H. (2010a), ”Tests for Multiple Regression Based on Simplicial Depth,” *Journal of Multivariate Analysis*, 101, 824-838.

——(2010b), ”Depth Notions for Orthogonal Regression,” *Journal of Multivariate Analysis*, 101, 2358-2371.

Zuo, Y., and Serfling, R. (2000a), ”General Notions of Statistical Depth Function,” *The Annals of Statistics*, 28, 461-482.

——(2000b), ”Structural Properties and Convergence Results for Contours of Sample Statistical Depth Functions,” *The Annals of Statistics*, 28, 483-499.

8. APPENDIX: PROOFS

Proof of Theorem 1:
Clearly, $d_T(\theta, z_*) > 0$ if $\text{res}(z_n, \theta) = 0$ for some $n \in \{1, \dots, K+1\}$. Therefore, we only have to consider the situation where $\text{res}(z_n, \theta) \neq 0$ for all $n \in \{1, \dots, K+1\}$.

Assume that $(\text{res}(z_1, \theta), \dots, \text{res}(z_{K+1}, \theta))^\top$ does not have alternating signs. This means that there exists $k \in \{1, \dots, K\}$ with $\text{sgn}(\text{res}(z_k, \theta)) = \text{sgn}(\text{res}(z_{k+1}, \theta))$. Set $s_n = \text{sgn}(\text{res}(z_n, \theta))$ for $n \in \{1, \dots, K+1\}$ and $s = (s_1, \dots, s_{K+1})^\top$. Then $s \in \{-1, 1\}^{K+1}$ and s has at most $K-1$ sign changes. According to Condition B), there exists $u_0 \in I\!R^K$ with $\text{sgn}(w_{u_0}(x_n)) = s_n$ for $n \in \{1, \dots, K+1\}$. But this implies

$$\text{sgn}(w_{u_0}(x_n))\,\text{sgn}(\text{res}(z_n, \theta)) = 1 \;\text{ for }\; n \in \{1, \dots, K+1\}$$

and thus

$$u_0^\top v(x_n, \theta)\,\text{res}(z_n, \theta) = w_{u_0}(x_n)\,\text{res}(z_n, \theta) > 0 \;\text{ for }\; n \in \{1, \dots, K+1\}$$

so that $d_T(z_*, \theta) = 0$.

Conversely, assume $d_T(\theta, z_*) = 0$. Then there exists $u \in I\!R^K$ with

$$u^\top v(x_n, \theta)\,\text{res}(z_n, \theta) = w_u(x_n)\,\text{res}(z_n, \theta) > 0 \;\text{ for }\; n \in \{1, \dots, K+1\}. \tag{15}$$

Since w_u has at most $K-1$ sign changes on $[x_1, x_{K+1}]$ according to Condition A), there exists $k \in \{1, \dots, K\}$ with

$$\text{sgn}(w_u(x_k)) = \text{sgn}(w_u(x_{k+1})).$$

This means with (15) that $\text{sgn}(\text{res}(z_k, \theta)) = \text{sgn}(\text{res}(z_{k+1}, \theta))$ so that $(\text{res}(z_1, \theta), \dots, \text{res}(z_{K+1}, \theta))^\top$ does not have alternating signs. □

Before we start the proof of Theorem 2 we recall the definition of m-dependence for random variables.

Definition 2. *A sequence of random variables $X_1, X_2, \ldots$ is m-dependent for $m \geq 0$, if $(X_i, \ldots, X_{i+n})$ is independent of $(X_{i+j}, \ldots, X_{i+j+n})$ for all $j > m$ and $i, n \in I\!N$.*

Proof of Theorem 2:
First note, that $res(\theta, Z_n) = E_n$ holds if θ is the underlying parameter.
a) Set

$$V_n := \prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(Z_{(K+1)(n-1)+k}, \theta)(-1)^k > 0\right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(Z_{(K+1)(n-1)+k}, \theta)(-1)^{k+1} > 0\right\}.$$

Then $V_n,\ n \in \{1, \ldots, \lfloor \frac{N}{K+1} \rfloor\}$, are independent variables with Bernoulli distribution satisfying $P(V_n = 1) = (1/2)^K$, so that the assertion follows from the CLT.
b) Set

$$V_n := \prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(Z_{n-1+k}, \theta)(-1)^k > 0\right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{\text{res}(Z_{n-1+k}, \theta)(-1)^{k+1} > 0\right\}.$$

Then $V_n, n \in \{1, ..., N-K\}$, are also Bernoulli variables with $P(V_n = 1) = (1/2)^K$. By centering to $X_n = V_n - (\frac{1}{2})^K$ we get a series of stationary random variables with $E[X_n] = 0$ and $E[|X_n|^3] < \infty$. So the limit theorem of Hoeffding and Robbins (1948) for m-dependent random variables can be applied, since X_n and X_m are dependent if and only if the corresponding index sets are overlapping. This implies, that $V_1, V_2, ...$ is K-dependent. To calculate the variance component in the limit distribution we need to calculate $E(X_1 X_d)$ for $d \in \{1, ..., K+1\}$ and get

$$A = E[X_1^2] + \sum\nolimits_{d=2}^{K+1} 2 \cdot E[X_1 X_d].$$

For $d > K+1$ the terms are zero, since the underlying events are independent.
For $d \in \{1, ..., K+1\}$ we have

$$\begin{aligned} E[X_1 X_d] &= E\left[\left(V_1 - \left(\frac{1}{2}\right)^K\right)\left(V_d - \left(\frac{1}{2}\right)^K\right)\right] \\ &= E\left[V_1 V_d\right] - \left(\frac{1}{2}\right)^{2K} = \left(\frac{1}{2}\right)^{K+d-1} - \left(\frac{1}{2}\right)^{2\cdot K}. \end{aligned}$$

By insertion of the explicit expressions for the expected values, A can be calculated by

$$\begin{aligned} A &= \sum\nolimits_{d=2}^{K+1} 2 \cdot \left[\left(\frac{1}{2}\right)^{K+d-1} - \left(\frac{1}{2}\right)^{2K}\right] + \left(\frac{1}{2}\right)^K \left(1 - \left(\frac{1}{2}\right)^K\right) \\ &= \left(\frac{1}{2}\right)^K \left[\sum\nolimits_{d=0}^{K-1} \left(\frac{1}{2}\right)^d - K\left(\frac{1}{2}\right)^{K-1} + 1 - \left(\frac{1}{2}\right)^K\right] \\ &= \left(\frac{1}{2}\right)^K \left[3 - \left(\frac{1}{2}\right)^{K-1} \cdot K - 3 \cdot \left(\frac{1}{2}\right)^K\right]. \end{aligned}$$

c) Set

$$\begin{aligned} V_n = {} & \mathbb{1}\left\{\text{res}(Z_n, \theta) > 0\right\} \mathbb{1}\left\{\text{res}(Z_{\lfloor \frac{N+1}{2} \rfloor}, \theta) < 0\right\} \mathbb{1}\left\{\text{res}(Z_{N-n+1}, \theta) > 0\right\} \\ & + \mathbb{1}\left\{\text{res}(Z_n, \theta) < 0\right\} \mathbb{1}\left\{\text{res}(Z_{\lfloor \frac{N+1}{2} \rfloor}, \theta) > 0\right\} \mathbb{1}\left\{\text{res}(Z_{N-n+1}, \theta) < 0\right\}. \end{aligned}$$

Again V_n are Bernoulli variables, here with $P(V_n = 1) = 1/4$. To apply the CLT we need to assure independence of $V_1, ..., V_{\lfloor \frac{N-1}{2} \rfloor}$. At first note that

$$\begin{aligned} & P\left(V_n = 0 | E_{\lfloor \frac{N+1}{2} \rfloor} > 0\right) \\ = & P\left(E_n > 0, E_{N-n+1} > 0\right) + P\left(E_n > 0, E_{N-n+1} < 0\right) + P\left(E_n < 0, E_{N-n+1} > 0\right) \\ = & \frac{3}{4} = P(V_n = 0), \end{aligned}$$

since $E_1, ..., E_N$ are independent. Analogously we obtain

$$P\left(V_n = 0 | E_{\lfloor \frac{N+1}{2} \rfloor} < 0\right) = \frac{3}{4} = P(V_n = 0)$$

and

$$P\left(V_n = 1|E_{\lfloor\frac{N+1}{2}\rfloor} < 0\right) = P\left(V_n = 0|E_{\lfloor\frac{N+1}{2}\rfloor} > 0\right) = \frac{1}{4} = P(V_n = 1).$$

Therefore independence of $E_1, ..., E_N$ implies that V_n and V_m, with $n < m < \lfloor\frac{N+1}{2}\rfloor$ are conditionally independent given $E_{\lfloor\frac{N+1}{2}\rfloor}$, so that

$$\begin{aligned}
&P(V_n = k, V_m = l) \\
=&P\left(V_n = k, V_m = l|E_{\lfloor\frac{N+1}{2}\rfloor} > 0\right) P\left(E_{\lfloor\frac{N+1}{2}\rfloor} > 0\right) \\
+&P\left(V_n = k, V_m = l|E_{\lfloor\frac{N+1}{2}\rfloor} < 0\right) P\left(E_{\lfloor\frac{N+1}{2}\rfloor} < 0\right) \\
=&P\left(V_n = k|E_{\lfloor\frac{N+1}{2}\rfloor} > 0\right) P\left(V_m = l|E_{\lfloor\frac{N+1}{2}\rfloor} > 0\right) \cdot \frac{1}{2} \\
+&P\left(V_n = k|E_{\lfloor\frac{N+1}{2}\rfloor} < 0\right) P\left(V_m = l|E_{\lfloor\frac{N+1}{2}\rfloor} < 0\right) \cdot \frac{1}{2} \\
=&P(V_n = k)P(V_m = l),
\end{aligned}$$

for $k, l \in \{0, 1\}$. Hence V_n and V_m are independent. Similarly, we obtain the independence of $V_1, ..., V_{\lfloor\frac{N-1}{2}\rfloor}$. □

Proof of Lemma 1: Without loss of generality, assume $0 \leq c_0 = c_1 \leq c_2 \leq \ldots \leq c_{K+1}$. Since the distribution of E_n is continuous and symmetric around 0, we have

$$\zeta_k := \frac{1}{2} - P_{\theta^*}(E_{n_k} > c_k) = \frac{1}{2} - P_{\theta^*}(E_{n_k} < -c_k) > 0 \quad \text{for all} \quad k \in \{1, \ldots, K+1\},$$

and in particular $P_{\theta^*}(E_{n_k} > c_k) = \frac{1}{2} - \zeta_k$, $P_{\theta^*}(E_{n_k} < c_k) = \frac{1}{2} + \zeta_k$, and
$0 \leq \zeta := \zeta_1 \leq \zeta_2 \leq \ldots \leq \zeta_{K+1} \leq \frac{1}{2}$. This implies with (10) that the expectation (9) equals

$$\prod_{k=1}^{K+1}\left(\frac{1}{2} + (-1)^{k+1}\zeta_k\right) + \prod_{k=1}^{K+1}\left(\frac{1}{2} + (-1)^{k}\zeta_k\right). \tag{16}$$

To prove the assertion, we have to show that (16) is bounded by $(\frac{1}{2})^K - (\frac{1}{2})^{K-2}\zeta^2$, which is equivalent to

$$\prod_{k=1}^{K+1}\left(1 + (-1)^{k+1}x_k\right) + \prod_{k=1}^{K+1}\left(1 + (-1)^{k}x_k\right) \leq 2 - 2x_1^2, \tag{17}$$

whereby $x_k := 2 \cdot \zeta_k$ and $0 \leq x_1 \leq x_2 \leq ... \leq x_{K+1} \leq 1$.
Note that

$$(1 + x_1)(1 - x_2) + (1 - x_1)(1 + x_2) = 2 - 2x_1x_2 \leq 2 - 2x_1^2,$$

since $x_1 \geq 0$. To conclude by induction first note, that from $x_{K+1} \geq x_K$

$$(1 - x_K)(1 + x_{K+1}) \geq (1 + x_K)(1 - x_{K+1})$$

and

$$(1 + x_{K-1})(1 - x_K)(1 + x_{K+1}) \geq (1 - x_{K-1})(1 + x_K)(1 - x_{K+1})$$

follows. By a successive application of this inequality we get

$$\begin{aligned} b :=& (1 \pm x_1)(1 \mp x_2) \cdot \ldots \cdot (1 - x_K)(1 + x_{K+1}) \\ \geq& (1 \mp x_1)(1 \pm x_2) \cdot \ldots \cdot (1 + x_K)(1 - x_{K+1}) =: a. \end{aligned} \tag{18}$$

By (18) and $ad + bc \leq ac + bd$ for $0 < a \leq b, 0 < c \leq d$ we get

$$\begin{aligned} &(1 \mp x_1)(1 \pm x_2) \cdot \ldots \cdot (1 + x_K)(1 - x_{K+1})(1 + x_{K+2}) \\ +&(1 \pm x_1)(1 \mp x_2) \cdot \ldots \cdot (1 - x_K)(1 + x_{K+1})(1 - x_{K+2}) \\ \leq&(1 \mp x_1)(1 \pm x_2) \cdot \ldots \cdot (1 + x_K)(1 - x_{K+1})(1 - x_{K+2}) \\ +&(1 \pm x_1)(1 \mp x_2) \cdot \ldots \cdot (1 - x_K)(1 + x_{K+1})(1 + x_{K+2}), \end{aligned} \tag{19}$$

by setting $d = (1 + x_{K+2}), c = (1 - x_{K+2})$. (17) now follows from

$$\begin{aligned} &\prod_{k=1}^{K+2} \left(1 + (-1)^{k+1} x_k\right) + \prod_{k=1}^{K+2} \left(1 + (-1)^k x_k\right) \\ =& \prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) (1 + (-1)^{K+3} x_{K+2}) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) (1 + (-1)^{K+2} x_{K+2}) \\ =& \frac{1}{2} \Bigg[(\prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) (1 + (-1)^{K+3} x_{K+2}) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) (1 + (-1)^{K+2} x_{K+2}) \\ &+ (\prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) (1 + (-1)^{K+3} x_{K+2}) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) (1 + (-1)^{K+2} x_{K+2}) \Bigg] \\ \overset{(19)}{\leq}& \frac{1}{2} \Bigg[(\prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) (1 + (-1)^{K+3} x_{K+2}) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) (1 + (-1)^{K+2} x_{K+2}) \\ &+ (\prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) (1 + (-1)^{K+2} x_{K+2}) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) (1 + (-1)^{K+3} x_{K+2}) \Bigg] \\ =& \frac{1}{2} \Bigg[\left(\prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) \right) \\ &\quad \cdot \left((1 + (-1)^{K+3} x_{K+2}) + (1 + (-1)^{K+2} x_{K+2}) \right) \Bigg] \\ =& \frac{1}{2} \left[\left(\prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) \right) ((1 + x_{K+2}) + (1 - x_{K+2})) \right] \\ =& \frac{1}{2} \left[\left(\prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) \right) \cdot 2 \right] \\ =& \prod_{k=1}^{K+1} \left(1 + (-1)^{k+1} x_k\right) + \prod_{k=1}^{K+1} \left(1 + (-1)^k x_k\right) \leq 2 - 2x_1^2 \end{aligned}$$

by induction. □

Proof of Theorem 3:
Let be $\epsilon > 0$ arbitrary. If we can show that $N_* \in I\!N$, $\delta_* > 0$, and a statistic $\widetilde{T}_N^i$ with $T_N^i(\theta^0) \leq \widetilde{T}_N^i$ exist such that

$$\mathrm{E}_{\theta^*}\left(\widetilde{T}_N^i\right) \leq -\sqrt{N}\left(\delta_* - \frac{q_\alpha}{\sqrt{N_*}}\right) \tag{20}$$

and

$$\mathrm{var}_{\theta^*}\left(\widetilde{T}_N^i\right) \leq \epsilon N \delta_*^2 \tag{21}$$

for all $N \geq N_*$, then Chebyshev's inequality provides for all $N \geq N_*$ using $\frac{q_\alpha}{\sqrt{N}} \geq \frac{q_\alpha}{\sqrt{N_*}}$

$$\begin{aligned}
P_{\theta^*}\left(T_N^i(\theta^0) \;\geq\; q_\alpha\right) &\leq P_{\theta^*}\left(\widetilde{T}_N^i \geq q_\alpha\right) \leq P_{\theta^*}\left(|\widetilde{T}_N^i - E_{\theta^*}(\widetilde{T}_N^i)| \geq q_\alpha - E_{\theta^*}(\widetilde{T}_N^i)\right)\\
&\leq P_{\theta^*}\left(|\widetilde{T}_N^i - E_{\theta^*}(\widetilde{T}_N^i)| \geq \sqrt{N}\frac{q_\alpha}{\sqrt{N}} + \sqrt{N}\left(\delta_* - \frac{q_\alpha}{\sqrt{N_*}}\right)\right)\\
&\leq P_{\theta^*}\left(|\widetilde{T}_N^i - E_{\theta^*}(\widetilde{T}_N^i)| \geq \sqrt{N}\,\delta_*\right) \leq \frac{\epsilon N \delta_*^2}{N\,\delta_*^2} = \epsilon.
\end{aligned}$$

a) Set

$$\begin{aligned}
\mathcal{M}_{iN}^1 &:= \left\{(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN};\ \exists l = 1,\ldots,L \text{ with } x_{n_1N},\ldots,x_{n_{K+1}N} \in [a_l+\delta, b_l-\delta)\right\}\\
\mathcal{M}_{iN}^2 &:= \left\{(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN} \setminus \mathcal{M}_{iN}^1;\ \exists l = 1,\ldots,L \text{ with } x_{n_1N},\ldots,x_{n_{K+1}N} \in [a_l, b_l)\right\}\\
\mathcal{M}_{iN}^3 &:= \mathcal{M}_{iN} \setminus (\mathcal{M}_{iN}^1 \cup \mathcal{M}_{iN}^2)
\end{aligned}$$

The assumptions of Theorem 3 a) imply that the conditions of Lemma 1 are satisfied with $c_0 = c > 0$ for all $(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN}^1$ so that the expectation (9) is bounded by $\left(\frac{1}{2}\right)^K - \left(\frac{1}{2}\right)^{K-2}\zeta^2$ with $\zeta := \frac{1}{2} - P_{\theta^*}(E_n > c) > 0$ for all $(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN}^1$. For all $(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN}^2$, the conditions of Lemma 1 are satisfied with $c_0 = 0$ so that (9) is bounded by $\left(\frac{1}{2}\right)^K$ for all $(n_1,\ldots,n_{K+1}) \in \mathcal{M}_{iN}^2$. Since only subsequent residuals are used in T_N^i, the elements of $\mathcal{M}_{iN}$ have the form $(m, m+1,\ldots,m+K)$ with $m \in \{1,\ldots,N-K\}$ so that $\mathcal{M}_{iN}^3$ contains at most $(K+1)(L+1)$ elements for all $N \in I\!N$. This means

$$\frac{\sharp\mathcal{M}_{iN}^1}{\sharp\mathcal{M}_{iN}} \to p,\ \frac{\sharp\mathcal{M}_{iN}^2}{\sharp\mathcal{M}_{iN}} \to 1-p,\ \frac{\sharp\mathcal{M}_{iN}^3}{\sharp\mathcal{M}_{iN}} \to 0,$$

for $N \to \infty$, where $p \in (0,1)$. For example, if $x_{1N},\ldots,x_{NN}$ are equidistant points in $[a,b]$ then p is the length of $\bigcup_{l=1}^L [a_l+\delta, b_l-\delta)$ divided by the length of $[a,b]$. Since 2 is a general upper bound of (9), we obtain

$$\begin{aligned}
\mathrm{E}_{\theta^*}(d_S^i(\theta^0, Z_*)) &\leq \frac{\sharp\mathcal{M}_{iN}^1}{\sharp\mathcal{M}_{iN}}\left(\left(\frac{1}{2}\right)^K - \left(\frac{1}{2}\right)^{K-2}\zeta^2\right) + \frac{\sharp\mathcal{M}_{iN}^2}{\sharp\mathcal{M}_{iN}}\left(\frac{1}{2}\right)^K + \frac{\sharp\mathcal{M}_{iN}^3}{\sharp\mathcal{M}_{iN}}\,2\\
&\longrightarrow p\left(\left(\frac{1}{2}\right)^K - \left(\frac{1}{2}\right)^{K-2}\zeta^2\right) + (1-p)\left(\frac{1}{2}\right)^K = \left(\frac{1}{2}\right)^K - p\left(\frac{1}{2}\right)^{K-2}\zeta^2 < \left(\frac{1}{2}\right)^K.
\end{aligned}$$

Hence there exists $\gamma < 0$, $N_0 \in I\!N$ such that for all $N \geq N_0$

$$\mathrm{E}_{\theta_*}\left(d_S^i(\theta^0, Z_*) - \left(\tfrac{1}{2}\right)^K\right) \leq \gamma < 0.$$

If $N \geq 2K$ then $N - K \geq \frac{N}{2} \geq \frac{N}{2(K+1)}$ and $\lfloor \frac{N}{K+1} \rfloor \geq \frac{N}{2(K+1)}$, so that

$$E_{\theta^*}(T_N^i(\theta^0)) \leq \sqrt{\frac{N}{2(K+1)}}\frac{\gamma}{v_i} =: \sqrt{N}\gamma_i,$$

where v_i is the denominator of $T_N^i(\theta^0)$ and $\gamma_i = \frac{\gamma}{\sqrt{2(K+1)}v_i}$.
Setting $\sqrt{N_i} > \frac{q_\alpha}{\gamma_i}$ delivers $\delta_i := \frac{q_\alpha}{\sqrt{N_i}} - \gamma_i > 0$. Since $\mathrm{res}(Z_1, \theta^0), \ldots, \mathrm{res}(Z_N, \theta^0)$ remain independent under P_{θ^*} for regression, the summands of $T_N^1(\theta^0)$ are independent and the summands of $T_N^2(\theta^0)$ are K-dependent, so that

$$var_{\theta^*}(T_N^i(\theta^0)) = \sigma_i^2 \leq \epsilon N \delta_i^2,$$

if $N \geq \frac{\sigma_i^2}{\epsilon\delta_i^2}$. Set $N_* := max\{\sqrt{N_i}, \frac{\sigma_i^2}{\epsilon\delta_i^2}, N_0\}$ and $\delta_* := \frac{q_\alpha}{\sqrt{N_*}} - \gamma_i$, then

$$\delta_* \geq \frac{q_\alpha}{\sqrt{N_i}} - \gamma_i = \delta_i > 0.$$

Therefore

$$E_{\theta^*}(T_N^i(\theta^0)) \leq \sqrt{N}\gamma_i = -\sqrt{N}\left(\delta_* - \frac{q_\alpha}{\sqrt{N_*}}\right)$$

and

$$var_{\theta^*}(T_N^i(\theta^0)) = \sigma_i^2 \leq N_*\epsilon\delta_i^2 \leq N_*\epsilon\delta_*^2 \leq N\epsilon\delta_*^2$$

for all $N \geq N_*$, so that conditions (20) and (21) are satisfied for $\tilde{T}_N^i = T_N^i(\theta_0)$ and $i \in \{1, 2\}$.

b) Under the assumptions of b), the simplified simplicial depths are given by

$$\begin{aligned} & d_S^i(\theta^0, Z_*) \\ = \; & \frac{1}{\sharp\mathcal{M}_{iN}} \sum\nolimits_{(n_1,\ldots,n_{K+1})\in\mathcal{M}_{iN}} \left(\prod_{k=1}^{K+1} \mathbb{1}\left\{E_{n_k}(-1)^k > c\right\} + \prod_{k=1}^{K+1} \mathbb{1}\left\{E_{n_k}(-1)^{k+1} > c\right\}\right). \end{aligned}$$

Setting $p := P_{\theta^*}(E_n > c)$, we get $p \neq \frac{1}{2}$ and obtain

$$\mathrm{E}_{\theta^*}(d_S^1(\theta, Z_*)) = \mathrm{E}_{\theta^*}(d_S^2(\theta, Z_*)) = \begin{cases} 2\,p^{\frac{K+1}{2}}\,(1-p)^{\frac{K+1}{2}} < \left(\frac{1}{2}\right)^K, & \text{if } K \text{ is odd,} \\ p^{\frac{K}{2}}\,(1-p)^{\frac{K}{2}} < \left(\frac{1}{2}\right)^K, & \text{if } K \text{ is even,} \end{cases}$$

so that condition (20) and (21) are satisfied as in a).
c) We consider only the case $\lim_{b\to\infty} g(b, \theta^0) - g(b, \theta^*) = \infty$ since the proof for the other case is completely analogous. Because of $\lim_{b\to\infty} g(b, \theta^0) - g(b, \theta^*) = \infty$, there exists $b_0 > 0$, $\gamma > 0$, and $\beta < \left(\frac{1}{2}\right)^K$ with $g(b, \theta^0) - g(b, \theta^*) > \gamma$ for all $b > b_0$ and $P_{\theta^*}(E_n > \gamma) \leq \frac{\beta}{2}$. According to

Scenario (B), there exists $N_0 \in I\!N$ so that $X_n > b_0$ almost surely for all $n \geq N_0$. Then we can work with the following upper bounds

$$\begin{aligned}
d_S^1(\theta^0, Z_*) \quad &\leq \\
\frac{1}{\lfloor \frac{N}{K+1} \rfloor} & \left(N_0 \;+ \sum\nolimits_{n=N_0}^{\lfloor \frac{N}{K+1} \rfloor} \left(\mathbb{1}\left\{ E_{(K+1)(n-1)+2} > \gamma \right\} + \mathbb{1}\left\{ E_{(K+1)(n-1)+1} > \gamma \right\} \right) \right) =: \widetilde{d}_S^1 \\
d_S^2(\theta^0, Z_*) \quad &\leq \quad \frac{1}{N-K} \left(N_0 + \sum\nolimits_{n=N_0}^{N-K} \left(\mathbb{1}\left\{ E_n > \gamma \right\} + \mathbb{1}\left\{ E_{n+1} > \gamma \right\} \right) \right) =: \widetilde{d}_S^2 \\
d_S^3(\theta^0, Z_*) \quad &\leq \quad \frac{1}{\lfloor \frac{N-1}{2} \rfloor} \left(N_0 + \sum\nolimits_{n=N_0}^{\lfloor \frac{N-1}{2} \rfloor} \left(\mathbb{1}\left\{ E_{\lfloor \frac{N+1}{2} \rfloor} > \gamma \right\} + \mathbb{1}\left\{ E_{N-n+1} > \gamma \right\} \right) \right) =: \widetilde{d}_S^3.
\end{aligned}$$

Set $\widetilde{T}_N^i := \sqrt{N} \lambda_i \left(\widetilde{d}_S^i - \left(\frac{1}{2}\right)^K \right)$, where λ_i is an appropriately chosen constant, then $\widetilde{T}_N^i$ also is an upper bound of $T_N^i(\theta^0)$. Then there exists $N_* > N_0$ such that (20) is satisfied for $\widetilde{T}_N^i$ for $i \in \{1, 2, 3\}$ and $N \geq N_*$ as in a). Since the summands of $\widetilde{T}_N^1$ are independent and the summands of $\widetilde{T}_N^2$ are 1-dependent, also condition (21) is satisfied for $\widetilde{T}_N^1$ and $\widetilde{T}_N^2$. To show that (21) also is satisfied for $\widetilde{T}_N^3$, use γ so large that $P_{\theta^*}(E_N > \gamma)^2 \leq P_{\theta^*}(E_n > \gamma) \leq \frac{\epsilon\, \delta_*^2}{\lambda_3^2 4}$ is satisfied as well. Since the errors E_n are independent and identically distributed, we obtain then

$$\begin{aligned}
&\mathrm{var}_{\theta^*}\left(\widetilde{T}_N^i \right) \\
&\leq \quad N \lambda_3^2 \, \mathrm{E}_{\theta^*} \left[\left(\mathbb{1}\left\{ E_{\lfloor \frac{N+1}{2} \rfloor} > \gamma \right\} + \mathbb{1}\left\{ E_{N-N_0+1} > \gamma \right\} \right)^2 \right] \\
&= \quad 2N\lambda_3^2 \left(P_{\theta^*}(E_N > \gamma) + P_{\theta^*}(E_N > \gamma)^2 \right) \leq 2N\lambda_3^2 \, 2 \, \frac{\epsilon\, \delta_*^2}{\lambda_3^2 \, 4} \;=\; \epsilon N \delta_*^2. \qquad \square
\end{aligned}$$

Proof of Lemma 2:

a) Because of

$$\begin{aligned}
\frac{\partial}{\partial t} w_u(x) &= u_2\theta_2 x^{\theta_2 - 1} + u_3\, \theta_1\theta_2 x^{\theta_2-1} \, \log(x) + u_3\, \theta_1 x^{\theta_2} \, \frac{1}{x} \\
&= \quad x^{\theta_2-1} \left(u_2\theta_2 + u_3\, \theta_1\theta_2 \; \log(x) + u_3\, \theta_1 \right) \geq 0 \\
&\Longleftrightarrow \quad u_2\theta_2 + u_3\, \theta_1\theta_2 \; \log(x) + u_3\, \theta_1 \geq 0 \\
&\Longleftrightarrow \quad u_3\, \theta_1\theta_2 \; \log(x) \geq -u_2\theta_2 - u_3\, \theta_1 \\
&\Longleftrightarrow \quad \log(x) \geq -\frac{u_2}{u_3\theta_1} - \frac{1}{\theta_2} \text{ if } u_3\, \theta_1\theta_2 > 0, \log(x) \leq -\frac{u_2}{u_3\theta_1} - \frac{1}{\theta_2} \text{ if } u_3\, \theta_1\theta_2 < 0 \\
&\Longleftrightarrow \quad x \geq \exp\left(-\frac{u_2}{u_3\theta_1} - \frac{1}{\theta_2} \right) \text{ if } u_3\, \theta_1\theta_2 > 0, x \leq \exp\left(-\frac{u_2}{u_3\theta_1} - \frac{1}{\theta_2} \right) \text{ if } u_3\, \theta_1\theta_2 < 0,
\end{aligned}$$

w_u has a minimum at $x = exp\left(-\frac{1}{\theta_2} - \frac{u_2}{u_3\theta_1}\right)$ if $u_3\, \theta_1\theta_2 > 0$ and a maximum at $x = exp\left(-\frac{1}{\theta_2} - \frac{u_2}{u_3\theta_1}\right)$ if $u_3\, \theta_1\theta_2 < 0$.

b) Let be $0 < \xi_1 < \xi_2$ arbitrary. The equation system

$$\begin{pmatrix} \xi_1^{\theta_2} & \theta_1 \xi_1^{\theta_2} \; \log(\xi_1) \\ \xi_2^{\theta_2} & \theta_1 \xi_2^{\theta_2} \; \log(\xi_2) \end{pmatrix} \begin{pmatrix} v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} v_2 \xi_1^{\theta_2} + v_3\, \theta_1 \xi_1^{\theta_2} \; \log(\xi_1) \\ v_2 \xi_2^{\theta_2} + v_3\, \theta_1 \xi_2^{\theta_2} \; \log(\xi_2) \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

has exact one solution $(v_2, v_3)^\top$ since

$$\begin{aligned}&\det\begin{pmatrix} \xi_1^{\theta_2} & \theta_1\xi_1^{\theta_2}\,\log(\xi_1) \\ \xi_2^{\theta_2} & \theta_1\xi_2^{\theta_2}\,\log(\xi_2)\end{pmatrix} \\ =\;& \xi_1^{\theta_2}\,\theta_1\xi_2^{\theta_2}\,\log(\xi_2) - \xi_2^{\theta_2}\,\theta_1\xi_1^{\theta_2}\,\log(\xi_1) = \xi_1^{\theta_2}\,\theta_1\xi_2^{\theta_2}\,(\log(\xi_2) - \log(\xi_1)) \neq 0.\end{aligned}$$

For this solution $(v_2, v_3)^\top$, it holds

$$\begin{aligned} w_{1,v_2,v_3}(\xi_1) = 1 + v_2\xi_1^{\theta_2} + v_3\,\theta_1\xi_1^{\theta_2}\,\log(\xi_1) &= 0, \\ w_{1,v_2,v_3}(\xi_2) = 1 + v_2\xi_2^{\theta_2} + v_3\,\theta_1\xi_2^{\theta_2}\,\log(\xi_2) &= 0.\end{aligned}$$

Since w_{1,v_2,v_3} has at most one extremum according to a), this extremum must be attained in (ξ_1, ξ_2) and the extreme value is not equal to zero. It is negative if it is a minimum and positive if it is a maximum. The use of $u_0 = (-1, -v_2, -v_3)^\top$ changes a negative minimal value for w_{u_*} with $u_* = (1, v_2, v_3)^\top$ to a positive maximal value for w_{u_0} and vice versa. Denote u_0 or u_*, respectively, by u_+ if the extreme value is positive and by u_- if the extreme value is negative. If the extreme value is positive, then $w_{u_+}(x) > 0$ for all $x \in (\xi_1, \xi_2)$ because only one extremum exists. The same argument provides $w_{u_-}(x) < 0$ for all $x \in (\xi_1, \xi_2)$ if the extreme value is negative. $\square$