REGULAR ARTICLE

Simple powerful robust tests based on sign depth

Kevin Leckey[1] ⓘ · **Dennis Malcherczyk[1]** ⓘ · **Melanie Horn[1]** ·
Christine H. Müller[1] ⓘ

Abstract

Up to now, powerful outlier robust tests for linear models are based on M-estimators and are quite complicated. On the other hand, the simple robust classical sign test usually provides very bad power for certain alternatives. We present a generalization of the sign test which is similarly easy to comprehend but much more powerful. It is based on K-sign depth, shortly denoted by K-depth. These so-called K-depth tests are motivated by simplicial regression depth, but are not restricted to regression problems. They can be applied as soon as the true model leads to independent residuals with median equal to zero. Moreover, general hypotheses on the unknown parameter vector can be tested. While the 2-depth test, i.e. the K-depth test for $K = 2$, is equivalent to the classical sign test, K-depth test with $K \geq 3$ turn out to be much more powerful in many applications. A drawback of the K-depth test is its fairly high computational effort when implemented naively. However, we show how this inherent computational complexity can be reduced. In order to see why K-depth tests with $K \geq 3$ are more powerful than the classical sign test, we discuss the asymptotic behavior of its test statistic for residual vectors with only few sign changes, which is in particular the case for some alternatives the classical sign test cannot reject. In contrast, we also consider residual vectors with alternating signs, representing models that fit the data very well. Finally, we demonstrate the good power of the K-depth tests for some examples including high-dimensional multiple regression.

✉ Kevin Leckey
  kevin.leckey@tu-dortmund.de

  Dennis Malcherczyk
  dennis.malcherczyk@tu-dortmund.de

  Melanie Horn
  mhorn@statistik.tu-dortmund.de

  Christine H. Müller
  cmueller@statistik.tu-dortmund.de

[1] Department of Statistics, TU Dortmund University, 44227 Dortmund, Germany

1 Introduction

Outlier robust tests for linear models are mainly given by Wald-type tests, likelihood ratio tests, and score-type tests based on M-estimators and related estimators as originally proposed by Schrader and Hettmansperger (1980), Markatou et al. (1991), Silvapulle (1992), Heritier and Ronchetti (1994). See also Hampel et al. (2011), Chapter 7, Huber and Ronchetti (2009), Chapter 13, and Maronna et al. (2019), Chapter 5. M-estimators have the disadvantage that they depend on score functions which must be specified. Moreover, they are not scale invariant so that the scale must be estimated simultaneously as this is done by the MM-estimators proposed by Yohai (1987). These MM-estimators are defined iteratively using the S-estimators for scale introduced by Rousseeuw and Yohai (1984). The robust tests given by lmRob in the R-packages robust and lmrob in robustbase are based on MM-estimators with special score functions where some efficient calculation is given by approaches of Koller and Stahel (2011, 2017). These tests are very powerful but complicated to compute since optimal regression and scale estimates are determined by an adaptive procedure.

We propose here powerful outlier robust tests which are much simpler since they are based only on signs of residuals. They can be used as soon as residuals $R_n(\theta)$, $n = 1, \ldots, N$, of a parametric model given by a parameter $\theta \in \Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}$, can be defined and which satisfy

$$P_\theta(R_n(\theta) > 0) = \frac{1}{2} = P_\theta(R_n(\theta) < 0). \tag{1}$$

Such residuals appear in linear or nonlinear regression models with realized regressors x_n where observations are of the form $Y_n = g(x_n, \theta) + E_n$ and the error variable E_n has a continuous distribution with median equal to zero. Then the residuals are given by $R_n(\theta) = Y_n - g(x_n, \theta)$. Generalized linear models are further examples of residuals satisfying (1) if the link function can be expressed by the median of the observations Y_n, i.e. if $\mathrm{med}(Y_n) = g(x_n, \theta)$; see, e.g., Leckey et al. (2020) for a load-sharing model which also leads to residuals satisfying (1). More examples are given by stochastic processes with i.i.d. increments E_n such as AR(p) processes given by $Y_n = g(Y_{n-1}, \ldots, Y_{n-p}, \theta) + E_n$. Realizations of $R_1(\theta), \ldots, R_N(\theta)$ are denoted by $r_1(\theta), \ldots, r_N(\theta)$.

The proposed tests are called K-depth tests and are based on the so-called K-sign depth, shortly denoted by K-depth. The K-depth of a parameter θ in a set of realized observations $y_1, \ldots, y_N$ is the relative amount of K-tuples $\{n_1, \ldots, n_K\} \subset \{1, \ldots, N\}$ with alternating signs of the residuals $r_{n_1}(\theta), \ldots, r_{n_K}(\theta)$. A hypothesis $H_0 : \theta \in \Theta_0$ is rejected if the K-depth of all $\theta \in \Theta_0$ is too small. The only hyperparameter which must be chosen is K. A good choice is often a value close to the dimension p of the parameter θ but other choices are possible.

For $K = 2$ and hypotheses of the form $H_0 : \theta = \theta^0$, the K-depth test is the classical sign test which counts the number of positive (or negative) residuals $r_1(\theta^0)$ and rejects the null hypothesis if the number of positive signs is too small or too large. In particular, it does not reject the null hypothesis if half of the residuals are positive and half of them are negative. However, this can also happen for alternatives with parameters far away from θ^0, see Fig. 1, p. 13. Therefore this simple sign-test is not powerful for many alternatives. However, the proposed K-depth test with $K \geq 3$ is much more powerful as we show in this paper.

The K-depth has its origin in simplicial regression depth. Simplicial regression depth is a modification of the regression depth introduced by Rousseeuw and Hubert (1999) to generalize the depth notion to regression. Originally, the halfspace depth of Tukey (1975) was used to obtain a generalization of the median for multivariate data. Liu (1988, 1990) extended this to simplicial depth. Simplicial depth can be expressed by counting the number of all $p + 1$-tupels of the p-dimensional data set with positive halfspace depth. Replacing halfspace depth by regression depth leads to simplicial regression depth. For the calculation of simplicial regression depth, Rousseeuw and Hubert (1999) and Müller (2005) noted that the regression depth of a p-dimensional parameter vector within $p + 1$ observations is greater than zero if and only if the residuals have alternating signs. Sufficient conditions for this equivalence and a proof of this property are given by Kustosz et al. (2016). This led to the idea to define the depth of a parameter θ directly via alternating signs of residuals in K-tuples.

It should be noted here that depth of a parameter in observations coming from a parametric model is treated only by few authors as Mizera and Müller (2004); Müller (2005); Denecke and Müller (2014); Paindaveine and Van Bever (2018); Wang (2019). Most of depth notion concern depth of data points in multivariate data sets as those of Zuo and Serfling (2000); Mosler (2002); Agostinelli and Romanazzi (2011); Lok and Lee (2011); Paindaveine and Van Bever (2013); Dong and Lee (2014); Claeskens et al. (2014); López-Pintado et al. (2014); Nagy and Ferraty (2019); Liu et al. (2020).

Any simplicial depth notion has the advantage that it is a U-statistics so that the asymptotic distribution can be derived by methods for U-statistics. For Liu's simplicial depth for multivariate data, this was used in Liu (1990); Dümbgen (1992); Arcones and Gine (1993). However, simplicial regression depth is often a degenerated U-statistic so that more effort is necessary to derive the asymptotic distribution, see Müller (2005); Wellmann et al. (2009); Wellmann and Müller (2010). The advantage of K-depth is that its distribution does not depend on the model and can be easily calculated for small sample sizes because only all 2^N combinations of signs must be considered. Moreover, its asymptotic distribution was derived in Kustosz et al. (2016) for $K = 3$ and in Malcherczyk et al. (2021) for general $K \geq 3$.

The derivation of the asymptotic distribution leads to an asymptotic equivalent variant of the K-depth which can be calculated in linear time $O(N)$ while a naive implementation has a complexity of $O(N^K)$ if N is the sample size. Studying especially the behavior of K-depth in the situation of few sign changes in the data leads to another implementation in this paper. This implementation is based on blocks of equal signs and is exact, of complexity $O(N)$, and much faster than the implementation based on the asymptotic form. This allows the application in multiple regression with high dimension where K should grow with the dimension. In multiple regres-

sion, an order of the residuals are needed. For this, we use new results of Horn and Müller (2020) - see also Horn (2021) - concerning optimal ordering of the multivariate explanatory variables.

In Sect. 2, we introduce the K-depth and the K-depth tests, discuss a relationship to runs test, and show how the computational complexity can be reduced by block implementation. Basic properties of the K-depth are derived in Sect. 3. This concerns a strong law of large numbers for the K-depth, the behavior at alternating signs of residuals and the behavior when only few sign changes occur. In particular, it is shown for all $K \geq 3$ that the expected value of the K-depth and its maximal value have the same limit as the number of observations tends to infinity. It is also shown that residuals with few sign changes have a K-depth that is strictly less than this limit which explains why the K-depth test has a high power at alternatives that tend to have few sign changes. A comparison between the K-depth tests for different values of K is given in Sect. 4. At first, for $K = 2$, the equivalence of the K-depth test and the classical sign test is derived formally. Afterwards, the p-values of K-depth tests with $K = 3, 4, 5, 6$ are compared in some worst case scenarios with few sign changes taken from Sect. 3. Sect. 5 demonstrates the good power of the K-depth tests via simulations for quadratic regression and for multiple regression with high dimensions. Finally, a discussion of the results and an outlook are given in Sect. 6. More details of the proofs and the block implementation as well as further simulation results are given in the supplementary material.

Notation. Throughout the article, $r_1(\theta), \ldots, r_N(\theta)$ denote realizations of the residuals $R_1(\theta), \ldots, R_N(\theta)$. If the choice of the parameter θ is clear, we also use the abbreviations $r_n := r_n(\theta)$ and $R_n := R_n(\theta)$ for $n = 1, \ldots, N$. The sign of a real number x is denoted by $\psi(x) = \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function. In some asymptotic calculations we make use of the O-Notation: For real-valued sequences $(a_n)_{n \geq 1}$ and $(b_n)_{n \geq 1}$, we write $a_n = O(b_n)$ if there is a constant $C > 0$ and an integer n_0 with $|a_n| \leq C|b_n|$ for all $n \geq n_0$. Furthermore, $a_n = \Theta(b_n)$ denotes that both $a_n = O(b_n)$ and $b_n = O(a_n)$.

2 *K*-depth tests and reduction of their computational complexity

In this section, we introduce the K-depth of a vector and how to use the K-depth notion as a test statistic. We also briefly discuss the issue of a fairly high computational complexity when working with K-depth tests. This issue can be resolved by using alternative representations of the original definition of the K-depth. The results are based on the following general assumption on the statistical models with unknown parameter $\theta \in \Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}$:

the residuals $R_1(\theta), \ldots, R_N(\theta)$ of N observations in $\mathbb{R}$ are independent

and satisfy (1) if θ is the true parameter.   (2)

2.1 *K*-depth and *K*-depth tests

The **K-sign depth** or shortly **K-depth** $d_K(r_1, \ldots, r_N)$ of $r_1, \ldots, r_N$ is the relative number of K-element subsets with alternating signs, i.e. for $K \geq 2$,

$$
d_K(r_1, \ldots, r_N) := \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < n_2 < \ldots < n_K \leq N} \Big(\prod_{k=1}^{K} \mathbb{1}\left\{(-1)^k r_{n_k} > 0\right\}
\tag{3}
$$
$$
+ \prod_{k=1}^{K} \mathbb{1}\left\{(-1)^k r_{n_k} < 0\right\} \Big).
$$

Remark 1 Note that the definition of the K-sign depth depends on the chosen order and therefore this choice is a crucial aspect. If $x_n \in \mathbb{R}^q$ for $q > 1$ then various multivariate orderings can be used. Not all of them provide powerful tests. Among the most promising approaches are orderings according to a shortest Hamiltonian path through the regressors $x_1, \ldots, x_N$. Luckily, also less computationally intensive approximations of such a path (such as the nearest neighbor approach) seem to perform similarly well. A detailed discussion on these and other orderings can be found in Horn and Müller (2020). See also Sect. 5 for an example. Moreover, note that under some conditions given by Kustosz et al. (2016), K-depth is equivalent to simplicial regression depth if $K = p + 1$ and $\theta \in \mathbb{R}^p$. Hence, an appropriate choice of K is a natural number close to $p + 1$. However, in contrast to simplicial regression depth, other choices than $K = p + 1$ are possible as well.

In order to obtain a non-degenerate limit distribution, the **K-depth test** is based on the following test statistic:

$$
T_K(\theta) := T_K(R_1(\theta), \ldots, R_N(\theta))
$$
$$
:= N \left(d_K(R_1(\theta), \ldots, R_N(\theta)) - \left(\frac{1}{2}\right)^{K-1} \right).
\tag{4}
$$

A test based on (4) requires the α-quantiles of the distribution of the test statistic. If N is small, the finite sample distribution for any K can be easily simulated since the determination of the K-depth with an underlying C++ algorithm computing Formula (3) is fairly fast for small N. For larger N, see Subsection 2.2.

With the quantiles at hand, the K-depth test, $K \geq 2$, is defined as in Müller (2005): A hypothesis of the form $H_0 : \theta \in \Theta^0$ shall be rejected if the K-depth $d_K(r_1(\theta), \ldots, r_N(\theta))$ of θ or $T_K(\theta)$ is too small for all $\theta \in \Theta^0$. Hence, if q_α is the α-quantile of the distribution of $T_K(\theta)$ under θ then the K-depth test for $H_0 : \theta \in \Theta^0$ is given by

$$
\text{reject } H_0 : \theta \in \Theta^0 \text{ if } \sup_{\theta \in \Theta^0} T_K(\theta) < q_\alpha.
\tag{5}
$$

Remark 2 The K-depth test can also be used in a two-sided version:

$$\text{reject } H_0 : \theta \in \Theta^0 \text{ if } \sup_{\theta \in \Theta^0} T_K(\theta) < q_{\alpha_1} \text{ or } \inf_{\theta \in \Theta^0} T_K(\theta) > q_{1-\alpha_2}$$

with $\alpha_1 + \alpha_2 = \alpha$, for example $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$. This test also rejects H_0 if too many sign changes occur in the residual vector, which is an indicator for negatively correlated residuals , for example in time series. While the one-sided version is mostly focused on detecting deviations from 0 in the median and can detect only strong positive correlation in the residuals, the two-sided version is the preferable choice when testing simultaneously whether the residuals are independent and have median zero. However, since our applications are mainly focused around tests on the median rather than on independence, this two-sided version will not be used subsequently. Nevertheless, note that a simplified version of the K-depth leads to a test which can be considered as a generalization of the runs test of Wald and Wolfowitz (1940) for testing the hypothesis of independent residuals, see e.g. Gibbons and Chakraborti (2003), pp. 78-86: This simplified K-depth uses only subsequent residuals and can be defined as in Kustosz et al. (2016) for $K \geq 2$:

$$d_K^S(r_1, \ldots, r_N) := \frac{1}{N - K + 1} \sum_{n=1}^{N-K+1} \Big(\prod_{k=1}^{K} \mathbb{1}\left\{ (-1)^k r_{n+k-1} > 0 \right\} \tag{6}$$
$$+ \prod_{k=1}^{K} \mathbb{1}\left\{ (-1)^k r_{n+k-1} < 0 \right\} \Big).$$

If $K = 2$ then this simplified K-depth counts the number of sign changes and thus the number of runs. Kustosz et al. (2016) used the simplified versions because they are faster to compute and their asymptotic behavior is easy to derive. However, since the simplified K-depth only considers $N - K + 1$ subsets instead of $\binom{N}{K}$, tests based on it are usually less powerful than tests based on the full K-depth, in particular if the independence of the residuals is ensured, see Kustosz et al. (2016) and Falkenau (2016).

2.2 Runtime and block-implementation

A major drawback of the K-depth test is its slow runtime when using an algorithm based on the definition (3). This definition requires the consideration of all increasing K-tuples in $\{1, \ldots, N\}$, hence leading to an algorithm with runtime $\Theta(N^K)$. Such an algorithm is clearly impractical in applications with fairly large sample sizes. Fortunately, the derivation of a limit theorem of the test statistic $T_K(\theta)$ leads to an asymptotically equivalent form of (4) which can be computed in linear time for all $K \geq 3$. More precisely, under the true parameter θ

$$T_K(R_1(\theta), \ldots, R_N(\theta)) = \Psi_K(\mathcal{W}_{\bullet}^N) + o_P(1), \quad K \geq 3,$$

where Ψ_K is a functional given in Malcherczyk et al. (2021), $o_P(1)$ is a random variable converging to zero in probability, and $\mathcal{W}_\bullet^N = (\mathcal{W}_t^N)_{t \in [0,1]}$ with

$$\mathcal{W}_t^N = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \psi\left(R_n(\theta)\right) \text{ with } \psi\left(R_n\right) = \mathbb{1}\{R_n > 0\} - \mathbb{1}\{R_n < 0\}.$$

Such a limit theorem is given in Kustosz et al. (2016) for $K = 3$. A generalization to all $K \geq 3$ as well as the resulting efficient algorithm can be found in Malcherczyk et al. (2021).

We will not go into detail on how this algorithm with runtime $\Theta(N)$ works since it requires a major part of the computation necessary to obtain the limit theorem and this is beyond this paper. Instead, we discuss a different approach which, when implemented carefully, even has a faster runtime than the algorithm from Malcherczyk et al. (2021). Moreover, the algorithm discussed below always yields the exact K-depth rather than an asymptotic approximation.

This section first provides the general idea of the algorithm that immediately results in an efficient procedure for residual vectors with only few sign changes. At the end of the section, a more careful implementation of this approach is sketched that leads to an efficient (linear time) algorithm to compute the exact K-depth. We refer to this approach as *block-implementation*. Aside from speeding up the implementation based on (3), this approach will be useful to derive some of the properties presented in Sect. 3.

Block-implementation. Let $r := (r_1, \ldots, r_N)$ be a vector of residuals and let $\psi(x)$ denote the sign of a real number x, i.e. $\psi(x) := \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$. The vector r is decomposed into *blocks* by letting a new block start at index j if and only if r_{j-1} and r_j have different signs. More formally, we define the number $B(r)$ of blocks and their starting positions $s_1(r), \ldots, s_{B(r)}(r)$ via $s_1(r) := 1$ and

$$B(r) := 1 + \sum_{n=2}^{N} \mathbb{1}\left\{\psi\left(r_{n-1}\right) \neq \psi\left(r_n\right)\right\},$$

$$s_b(r) := \min\left\{\ell > s_{b-1}(r); \ \psi\left(r_\ell\right) \neq \psi\left(r_{\ell-1}\right)\right\}, \quad b = 2, \ldots, B(r).$$

For convenience, we define $s_{B(r)+1}(r) := N + 1$. The *block sizes* are defined as

$$q_b(r) := s_{b+1}(r) - s_b(r), \quad b = 1, \ldots, B(r).$$

Example 1 The vector $r = (1, 2, 6, -1, 3, 2, -5, 2)$ consists of $B(r) = 5$ blocks

$$(\underbrace{1, 2, 6}_{\text{block 1}}, \underbrace{-1}_{\text{block 2}}, \underbrace{3, 2}_{\text{block 3}}, \underbrace{-5}_{\text{block 4}}, \underbrace{2}_{\text{block 5}}).$$

The block sizes are $q_1(r) = 3$, $q_3(r) = 2$ and $q_j(r) = 1$ for $j = 2, 4, 5$.

We say that the *n*th residual r_n *belongs to block j* if and only if $s_j(r) \leq n < s_{j+1}(r)$. The *sign of block j* is defined as the sign of the first (and thus any) element $r_{s_j(r)}$ belonging to that block. Blocks $j_1 < \ldots < j_k$ are called *alternating* if and only if the signs of the blocks are alternating, i.e. the signs of block j_i and j_{i+1} are different for all $i = 1, \ldots, k - 1$. Note that two blocks j_1 and j_2 have different signs if and only if j_1 is even and j_2 is odd or vice versa. In particular, the blocks $j_1 < \ldots < j_k$ are alternating if and only if $j_{i+1} - j_i$ is odd for all $i = 1, \ldots, k - 1$.

Example 2 Consider the block decomposition for the vector *r* from Example 1. In this decomposition, blocks 1, 3, 5 have positive signs and blocks 2, 4 have negative signs. Hence if $\mathcal{A}$ denotes the set of alternating triples of blocks then

$$\mathcal{A} = \{(1, 2, 3), \quad (1, 2, 5), \quad (1, 4, 5), \quad (2, 3, 4), \quad (3, 4, 5)\}.$$

Since a triple (r_i, r_j, r_k), $i < j < k$, of entries from *r* is alternating if and only if they belong to an alternating triple of blocks, we may count the number of triples in *r* with alternating signs by counting the corresponding combinations of elements from alternating blocks, i.e. in our example with *r* of length $N = 8$,

$$d_3(r) = \frac{1}{\binom{8}{3}} \sum_{(i,j,k)\in\mathcal{A}} q_i(r)q_j(r)q_k(r) = \frac{6+3+3+2+2}{\binom{8}{3}} = \frac{4}{14}.$$

More generally, we have the following alternative representation of (3):

Lemma 1 *Let* $\mathbb{O} := 2\mathbb{N}_0 + 1$ *denote the set of all odd positive integers and let*

$$\mathcal{A}_{K,B} := \left\{(i_1, \ldots, i_K) \in \{1, \ldots, B\}^K; \ i_k - i_{k-1} \in \mathbb{O} \text{ for } k = 2, \ldots, K\right\},$$

$$d_{K,N,B}(q_1, \ldots, q_B) := \frac{1}{\binom{N}{K}} \sum_{(i_1,\ldots,i_K)\in\mathcal{A}_{K,B}} \prod_{k=1}^{K} q_{i_k}, \quad B \in \mathbb{N}, \ q_1, \ldots, q_B > 0.$$

Let $q_1(r), \ldots, q_{B(r)}(r)$ *be the block sizes of a vector* $r = (r_1, \ldots, r_N)$. *Then*

$$d_K(r_1, \ldots, r_N) = d_{K,N,B(r)}(q_1(r), \ldots, q_{B(r)}(r)). \tag{7}$$

Remark 3 Note that the size of $\mathcal{A}_{K,B}$ is $\Theta(B^K)$. Also note that the effort to compute the block sizes $q_1(r), \ldots, q_{B(r)}(r)$ of a vector $r = (r_1, \ldots, r_N)$ is $\Theta(N)$. Hence, a naive algorithm based on the expression in Lemma 1 has computational complexity $\Theta(N + B^K)$ if $B = B(r)$ is the number of blocks in *r*. With some additional effort, the computational costs can even be reduced to $\Theta(N + B)$ by properly storing all relevant terms during the computation. For simplicity, we only discuss $K = 3$ here, more details on general K can be found in the supplementary material and in Malcherczyk (2022). Note that factoring out the length q_{i_2} of the second block in the representation

from Lemma 1 yields

$$d_{3,N,B}(q_1, \ldots, q_B) = \frac{1}{\binom{N}{3}} \sum_{i_2=2}^{B-1} q_{i_2} \left(\sum_{\substack{i_1=1 \\ i_2-i_1 \text{ odd}}}^{i_2-1} q_{i_1} \right) \left(\sum_{\substack{i_3=i_2+1 \\ i_3-i_2 \text{ odd}}}^{B} q_{i_3} \right). \tag{8}$$

This representation can be computed in linear time complexity by deriving the values of the inner sums in advance: To this end, let

$$\mathcal{F}(i_2) = \sum_{\substack{i=1 \\ i_2-i \text{ odd}}}^{i_2-1} q_i, \qquad \mathcal{B}(i_2) = \sum_{\substack{i=i_2+1 \\ i-i_2 \text{ odd}}}^{B} q_i, \quad i_2 = 2, \ldots, B-1.$$

Note that all values $(\mathcal{F}(i_2), \mathcal{B}(i_2))$, $i_2 = 2, \ldots, B-1$, can be computed with a total complexity of $\Theta(B)$ similarly to the cumulative sum of a vector of length B. With these values stored, (8) can be computed in linear time since the product of the inner sums equals $\mathcal{F}(i_2) \cdot \mathcal{B}(i_2)$ which now can be computed in constant time. For $K \geq 4$, a similar approach leads to a representation with

$$\mathcal{B}_1(i_2) := \mathcal{B}(i_2), \quad \mathcal{B}_j(i_2) = \sum_{\substack{i=i_2+1 \\ i-i_2 \text{ odd}}}^{B-j+1} q_i \mathcal{B}_{j-1}(i), \quad j \geq 2, \quad i_2 = 2, \ldots, B-j,$$

and $d_{K,N,B}(q_1, \ldots, q_B) = \frac{1}{\binom{N}{K}} \sum_{i_2=2}^{B-K+2} q_{i_2} \mathcal{F}(i_2) \mathcal{B}_{K-2}(i_2)$. More details can be found in the supplementary material.

Remark 4 As a simulation study in the dissertation of Malcherczyk (2022) reveals, the efficient block implementation stated in Remark 3 is even faster than the asymptotic variant from Malcherczyk et al. (2021), even when considering residuals from the null hypothesis that have a large number of blocks. More details can be found in (Malcherczyk (2022), Chapter 5.3.)

3 Basic properties of the K-depth

This section contains some of the basic properties of the K-depth. In particular, we discuss the typical behavior in terms of a law of large numbers in Sect. 3.1. Sections 3.2 and 3.3 contain extremal cases where the test statistic is close to its maximal or minimal value, respectively.

3.1 Law of large numbers

Let $R_1 := R_1(\theta), \ldots, R_N := R_N(\theta)$ be independent random variables satisfying (1).
Then the expectation of the K-depth is given by

$$
\mathbb{E}_\theta \left(d_K(R_1(\theta), \ldots, R_N(\theta)) \right)
$$
$$
= \frac{1}{\binom{N}{K}} \sum_{1 \le n_1 < n_2 < \ldots < n_K \le N} \left(\left(\frac{1}{2}\right)^K + \left(\frac{1}{2}\right)^K \right) = \left(\frac{1}{2}\right)^{K-1}. \tag{9}
$$

A convergence of the K-depth towards this expectation can be shown by rewriting the
summands in (3) using the identity in the next lemma. In order to avoid triple indices,
we write $i(j)$ instead of i_j.

Lemma 2 *If $E_{n_1}, \ldots, E_{n_K}$ are random variables with $P(E_{n_i} \neq 0) = 1$ for $i = 1, \ldots, K$
and $K \in \mathbb{N} \setminus \{1\}$ then we have*

$$
\prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k < 0\} - \left(\frac{1}{2}\right)^{K-1}
$$
$$
= \frac{1}{2^{K-1}} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \le i(1) < \ldots < i(2L) \le K} \prod_{j=1}^{2L} (-1)^{i(j)} \psi \left(E_{n_{i(j)}} \right) \ P\text{-almost surely}, \tag{10}
$$

where $\psi(x) := \mathbb{1}\{x > 0\} - \mathbb{1}\{x < 0\}$.

Proof (Sketch) The proof is based on $\mathbb{1}\{x > 0\} = (\psi(x) + 1)/2$ and $\mathbb{1}\{x < 0\} = (-\psi(x) + 1)/2$ for $x \neq 0$ and on $\prod_{i=1}^{K}(a_i + 1) = \sum_{\ell=1}^{K} \sum_{1 \le i(1) < \cdots < i(\ell) \le K} \prod_{j=1}^{\ell} a_{i(j)} + 1$
for arbitrary $a_1, \ldots, a_K$. This implies

$$
\prod_{k=1}^{K} \mathbb{1}\{E_k(-1)^k > 0\} = \frac{1}{2^K} \prod_{k=1}^{K} \left((-1)^k \psi(E_k) + 1 \right)
$$
$$
= \frac{1}{2^K} \left(\sum_{\ell=1}^{K} \sum_{1 \le i(1) < \ldots < i(\ell) \le K} (-1)^{i(1)+\cdots+i(\ell)} \prod_{j=1}^{\ell} \psi \left(E_{i(j)} \right) + 1 \right)
$$

and a similar expression for $\prod_{k=1}^{K} \mathbb{1}\{E_k(-1)^k < 0\}$.

$\square$

Studying the variance of the expression (10) reveals that it converges to zero as $N \to \infty$. Hence Lemma 2 leads to a law of large numbers for K-depth:

Theorem 1 *Let $K \geq 2$. If $R_1(\theta), \ldots, R_N(\theta)$ are satisfying (1) then*

$$d_K(R_1(\theta), \ldots, R_N(\theta)) \longrightarrow \left(\frac{1}{2}\right)^{K-1}$$

P_θ-almost surely as $N \to \infty$.

Proof (Sketch) Set $R_n = R_n(\theta)$. The assertion follows from

$$\mathbb{E}_\theta(d_K(R_1, \ldots, R_N)) = \left(\frac{1}{2}\right)^{K-1}, \quad \text{var}_\theta(d_K(R_1, \ldots, R_N)) = O(N^{-2})$$

by using Chebyshev's inequality and the Borel-Cantelli Lemma. The bound on the variance can be deduced from the representation given in Lemma 2 by taking into account that $\psi(R_1), \ldots, \psi(R_N)$ are i.i.d. and uniformly distributed on $\{-1, 1\}$ and therefore

$$\mathbb{E}_\theta\left(\prod_{j=1}^{2L} \psi\left(R_{n_{i(j)}}\right) \prod_{j=1}^{2L} \psi\left(R_{\tilde{n}_{i(j)}}\right)\right) = \begin{cases} 1, & \text{if } n_{i(j)} = \tilde{n}_{i(j)} \text{ for } j = 1, \ldots, 2L, \\ 0, & \text{else.} \end{cases}$$

$\square$

3.2 *K*-depth for alternating signs

In this section we study the behavior of the K-depth of residuals with alternating signs, i.e. of residuals $r_1, \ldots, r_N$ with $\psi(r_n) = -\psi(r_{n+1})$ for $n = 1, \ldots, N-1$. Alternating signs indicate a good fit and the K-depth attains its maximum value in this situation. Therefore it is of interest what exactly this maximum value is. This is given by the following theorem. As usual, we use the convention $\binom{n}{k} = 0$ for $n < k$.

Theorem 2 *Suppose $r_1, \ldots, r_N$ have alternating signs. Then, for $2 \leq K \leq N$,*

$$d_K(r_1, \ldots, r_N) = \frac{1}{\binom{N}{K}}\left(\binom{\lfloor(N+K)/2\rfloor}{K} + \binom{\lceil(N+K-2)/2\rceil}{K}\right).$$

Proof (Sketch) Let $r_1, \ldots, r_N$ be residuals with alternating signs. Let $|\mathcal{A}_{K,N}|$ be the size of the set $\mathcal{A}_{K,N}$ from Lemma 1. Then $d_K(r_1, \ldots, r_N) = |\mathcal{A}_{K,N}|/\binom{N}{K}$. It therefore only remains to count the number of $1 \leq i_1 < \ldots < i_K \leq N$ for which $i_{j+1} - i_j$ is odd for all $j = 2, \ldots, K$. The supplementary material contains a combinatorial deduction of this number that is based on rewriting $i_{j+1} - i_j = 2a_j + 1$, $a_j \in \mathbb{N}$. $\square$

Note that Theorem 2 can also be used to determine the size of the index set $\mathcal{A}_{K,B}$ in the block-implementation:

Corollary 1 *Let B, $K \geq 2$ be integers and let $\mathcal{A}_{K,B}$ be as in Lemma 1. Then*

$$|\mathcal{A}_{K,B}| = \binom{\lfloor (B+K)/2 \rfloor}{K} + \binom{\lceil (B+K-2)/2 \rceil}{K},$$

where $|\mathcal{A}_{K,B}|$ denotes the size of $\mathcal{A}_{K,B}$.

Theorem 2 implies that the K-depth of residuals with alternating signs converges to the expected value $(1/2)^{K-1}$ as $N \to \infty$. In conjunction with Corollary 1, we may extend this property to the following more general class of alternating vectors:

Definition 1 Let $M \in \mathbb{N}$ and let $r = (r_1, \ldots, r_N)$ be a vector of residuals. The residuals $r_1, \ldots, r_N$ are *alternating in blocks of size M* if N is a multiple of M and if

$$q_j(r) = M \quad \text{for all } j = 1, \ldots, B(r),$$

where the number $B(r)$ of blocks and the size $q_j(r)$ of block j are defined in Sect. 2.2. In particular, residuals have *alternating signs* if they are alternating in blocks of size 1.

With Corollary 1, it is not hard to compute the K-depth of such residuals:

Lemma 3 *Let $M, N \in \mathbb{N}$ with $B := N/M \in \mathbb{N}$. Furthermore, let $\langle x \rangle_J = \prod_{j=0}^{J-1}(x-j)$ for $x \in \mathbb{N}$ and $x \geq J$. If $r_1, \ldots, r_N$ are alternating in blocks of size M and if $B \geq K$, then*

(a) $d_K(r_1, \ldots, r_N) = \dfrac{\langle \frac{B+K-2}{2} \rangle_{K-1}}{B^{K-1}} \cdot \dfrac{N^K}{\langle N \rangle_K}$ *if $K + B$ is even,*

(b) $d_K(r_1, \ldots, r_N) = \dfrac{2\langle \frac{B+K-1}{2} \rangle_K}{B^K} \cdot \dfrac{N^K}{\langle N \rangle_K}$ *if $K + B$ is odd.*

Proof (Sketch) A combination of Lemma 1 and Corollary 1 yield an explicit expression for the K-sign depth of a vector with blocks of equal size. The assertion follows at once for odd $B + K$ and after rearranging this expression for even $B + K$. □

An asymptotic analysis of the K-depth based on Lemma 3 reveals that the K-depth test statistic of residuals that alternate in blocks of size M converges to its maximal value:

Theorem 3 *Let M be a fixed integer. If the residuals $r_1, \ldots, r_N$ are alternating in blocks of size M, then*

$$\lim_{N \to \infty} N\left(d_K(r_1, \ldots, r_N) - \left(\frac{1}{2}\right)^{K-1}\right) = \frac{K(K-1)}{2^K}.$$

Proof (Sketch) The assertion follows from the explicit formula given in Lemma 3 by approximating the falling factorials up to their second order term using

$$\langle x + a \rangle_J = \prod_{j=0}^{J-1}(x+a-j) = x^J + J\left(a - \frac{J-1}{2}\right)x^{J-1} + O(x^{J-2}),$$

for $x = B/2, a = (K - 2)/2, J = K - 1$ and $x = B/2, a = (K - 1)/2, J = K$ and $x = N, a = 0, J = K$, respectively. □

Remark 5

(a) Theorem 3 yields that the maximal value of the test statistic (i.e. the value for residuals with alternating signs) is asymptotically $K(K - 1)/2^K$. Since the minimal K-depth is zero, the minimal value of the test statistic is $-N/2^{K-1}$ which diverges as $N \to \infty$. Hence the (asymptotic) distribution of the test statistic $T_K(\theta)$ is bounded from above but unbounded from below. In particular, its distribution is not symmetric.

(b) Since the test statistic converges to its maximal value if the residuals are alternating in blocks of size $M \geq 1$, the (one-sided) K-depth test will not reject the model when such residuals are observed and N is sufficiently large. This can often be desirable in practice where alternating residuals indicate a good fit and a systematic alternation (in blocks of fixed size) can be caused by some vibration behavior which is difficult to filter out.

(c) If the independence of the residuals is questionable and of additional interest then alternating residuals are indicating dependence. In such situations, the two-sided K-depth test as proposed in Remark 2 can be used. Since alternating residuals yield the maximal possible value, the two-sided test will always reject the model when such residuals are observed and N is sufficiently large.

3.3 Behavior in situations of few sign changes

Residual vectors with only few sign changes usually indicate a bad choice for the modeling parameter, see, e.g., Fig. 1 for so-called nonfits in a quadratic regression model. A nonfit is defined as in Rousseeuw and Hubert (1999):
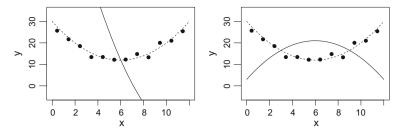
Definition 2 A parameter θ is called a *nonfit* if there exists another parameter $\tilde{\theta}$ such that $|r_n(\tilde{\theta})| < |r_n(\theta)|$ for all $n = 1, \ldots, N$.

The 2-depth test can struggle rejecting such bad choices since this test, as we will formally show in Sect. 4.1, is equivalent to the classical sign test. In particular, it does not reject the model if nearly half of the residuals are positive, regardless of how many sign changes the residuals have. K-depth tests with $K \geq 3$ are much more powerful in this regard since they immediately reject models that lead to few sign changes. More precisely, the following lemma is easy to show for residuals vectors $r = (r_1, \ldots, r_N)$ where the number $B(r)$ of blocks (see Sect. 2.2) is small:

Lemma 4 *Let $K \geq 3$. Then $d_K(r_1, \ldots, r_N) = 0$ if and only if $B(r) \leq K - 1$.*

Note that a K-depth of zero is the smallest possible value of the K-depth. Hence this will always lead to a rejection of the null hypothesis by the K-depth test if the sample size is high enough that a rejection at level α is possible. Usually a nonfit of a p-dimensional parameter is expressed by at most $p - 1$ sign changes. Hence a K-depth test with $K = p + 1$ will protect against bad power at nonfits, see also Kustosz et al. (2016). However, choices $K < p + 1$ can also lead to a good power of the

Fig. 1 12 observations generated by $Y_n = g(x_n, \theta^0) + E_n$ with $g(x, \theta^0) = 30 - 6x + 0.5 x^2$ (dashed line, $\theta^0 = (30, -6, 0.5)^\top$) and $E_n \sim \mathcal{N}(0, 1.5^2)$. The solid lines correspond to parameters that yield alternatives with either one or two sign changes: $\theta^{(1)} = (120, -24, 1)^\top$ yielding $g(x, \theta^{(1)}) = 120 - 24x + x^2$ on the left hand side and $\theta^{(2)} = (3, 6, -0.5)^\top$ yielding $g(x, \theta^{(2)}) = 3 + 6x - 0.5 x^2$ on the right hand side

K-depth test at alternatives for which the expected depth of $(1/2)^{K-1}$ is not reached. More precisely, since all α-quantiles of the asymptotic distribution of the K-depth test statistic $T_K(\theta)$ are fixed values greater than $-\infty$, we have the following property for growing sample size N: The strict inequality

$$\lim_{N \to \infty} \sup_{\theta \in \Theta^0} d_K(r_1(\theta), \ldots, r_N(\theta)) < \left(\frac{1}{2}\right)^{K-1} \tag{11}$$

implies $\lim_{N \to \infty} \sup_{\theta \in \Theta^0} T_K(\theta) = -\infty$ so that $H_0 : \theta \in \Theta^0$ is rejected if N is sufficiently large.

Condition (11) is in particular satisfied if the relative number of either the positive or negative residuals is tending to 1. This is often the case when the region of explanatory variables is growing to infinity as N converges to infinity. This was used in Kustosz et al. (2016) to show the consistency of a test based on simplicial depth for explosive AR(1) regression.

Assuming a bounded, fixed support for the explanatory variables, the relative number of positive/negative residuals usually does not tend to one for alternatives, e.g. in polynomial regression. However, one at least expects only few sign changes then; see Fig. 1 for examples with only one or two sign changes. We therefore end the section with a discussion on the K-depth of residual vectors where the number of blocks/sign changes is bounded.

For the remainder of the section, we will use the alternative representation of the K-depth based on the block-implementation (see Sect. 2.2). Recall that the K-depth of residuals $r_1, \ldots, r_N$ with B blocks and block sizes $q_1, \ldots, q_B$ is given by

$$d_{K,N,B}(q_1, \ldots, q_B) = \frac{1}{\binom{N}{K}} \sum_{(i_1, \ldots, i_K) \in \mathcal{A}_{K,B}} \prod_{k=1}^{K} q_{i_k}.$$

Although $q_1, \ldots, q_B$ are integers in practice, it will be more convenient in the subsequent analysis to let $q_1, \ldots, q_B$ be positive real numbers. In order to see that the K-depth test always rejects the null hypothesis if B is sufficiently small, we need to consider the input $q_1, \ldots, q_B$ with maximal K-depth. While it is arguably quite

intuitive to assume that this maximum is attained at $q_j = N/B$ for all $j = 1, \ldots, N$, a formal proof to determine the maximum is challenging. We therefore state the following conjecture which we only checked for some particular choices of K and B and could only prove for $K = 3$ completely. The proof is based on an optimization via Lagrange multipliers which, in particular, requires to show the uniqueness of its critical point. However, transforming the system of equations to deduce the uniqueness becomes very complicated for larger K, see the supplementary material for the proof and the main problem for the case $K \geq 4$:

Conjecture 1 *Let $K \geq 3$, $B \geq K$ and $N \geq B$. Consider the set*

$$\mathcal{M}_{K,N,B} := \arg\max \left\{ d_{K,N,B}(q_1, \ldots, q_B); \ (q_1, \ldots, q_B) \in (0, N)^B, \ \sum_{b=1}^{B} q_b = N \right\}.$$

Then the following holds: (a) If $K + B$ is even then

$$\mathcal{M}_{K,N,B} = \left\{ \left(\frac{N}{B}, \ldots, \frac{N}{B} \right) \right\}.$$

(b) If $K + B$ is odd then

$$\mathcal{M}_{K,N,B} = \left\{ \left(\frac{\beta N}{B-1}, \frac{N}{B-1}, \ldots, \frac{N}{B-1}, \frac{(1-\beta)N}{B-1} \right); \ \beta \in (0, 1) \right\}.$$

The necessity of a case distinction between $K + B$ even/odd might be a bit surprising at first. But in fact it is not hard to check that the function $d_{K,N,B}$ has the following property:

Lemma 5 *Let $K \geq 2$ and $B \geq K$. If $K + B$ is odd then*

$$d_{K,N,B}(q_1, \ldots, q_B) = d_{K,N,B-1}(q_1 + q_B, q_2, \ldots, q_{B-1}).$$

Proof (Sketch) Assume that $K + B$ is odd. The key observation to prove the lemma is that, for any $(i_2, \ldots, i_{K-1}) \in \{2, \ldots, B-1\}^{K-1}$, the vector $(1, i_2, \ldots, i_{K-1})$ is in $\mathcal{A}_{K,B}$ if and only $(i_2, \ldots, i_{K-1}, B) \in \mathcal{A}_{K,B}$. Hence, summands in $d_{K,N,B}(q_1, \ldots, q_B)$ where $i_1 = 1$ can be merged with those where $i_K = B$, resulting in a rearranged sum equal to $d_{K,N,B}(q_1 + q_B, q_2, \ldots, q_{B-1})$ as claimed. □

Hence we may assume w.l.o.g. that $K + B$ is even and use Lemma 5 to cover the odd case. Before stating the general result, we consider the special cases $B = K$ and $B = K + 1$. In these cases, Conjecture 1 is easy to verify since, by definition,

$$d_{K,N,K}(q_1, \ldots, q_K) = \frac{1}{\binom{N}{K}} \prod_{j=1}^{K} q_j, \tag{12}$$

$$d_{K,N,K+1}(q_1, \ldots, q_{K+1}) = \frac{1}{\binom{N}{K}}(q_1 + q_{K+1}) \prod_{j=2}^{K} q_j.$$

In particular, we have the following theorem for the maximal K-depth among all valid block sizes $q_1, \ldots, q_B$. The set of these valid block sizes is denoted by

$$\mathcal{Q}_{N,B} := \left\{ (q_1, \ldots, q_B) \in \mathbb{N}^B; \sum_{j=1}^{B} q_j = N \right\}, \quad N, B \in \mathbb{N}. \tag{13}$$

Theorem 4 *Let $K \geq 2$, $B \in \{K, K+1\}$ and let $\mathcal{Q}_{N,B}$ be as above. Then*

$$\lim_{N \to \infty} \sup \left\{ d_{K,N,B}(q_1, \ldots, q_B); \ (q_1, \ldots, q_B) \in \mathcal{Q}_{N,B} \right\}$$

$$= \frac{K!}{K^K} \leq \left(\frac{1}{2} \right)^{K-1}, \tag{14}$$

where the inequality in (14) is strict for $K \geq 3$.

Proof (Sketch) For $B = K$, one needs to compute the global maximum of the function given in (12) with the side conditions $q_1, \ldots, q_K \in \mathbb{N}$ and $\sum_{k=1}^{N} q_k = N$. When disregarding the integer condition, this can easily be done, e.g., by using Lagrange multipliers. This reveals a unique global maximum at $q_k = N/K$ for all $k = 1, \ldots, B$ which coincides with the integer maximum whenever $N/K \in \mathbb{N}$. The case $B = K+1$ follows from the case $B = K$ and Lemma 5. □

For the general case $B \geq K+2$, we will only consider the input $q_1 = \ldots = q_B = N/B$ since this is assumed to yield the maximal depth according to Conjecture 1 if $K + B$ is even. Lemma 3 yields the following result on the asymptotic K-depth.

Theorem 5 *Let $K \geq 2$ and $B \geq K$ be fixed. If $K + B$ is even then*

$$\lim_{N \to \infty} d_{K,N,B}\left(\frac{N}{B}, \ldots, \frac{N}{B} \right) = \frac{\prod_{k=1}^{K-1} \left(\frac{B+K}{2} - k \right)}{B^{K-1}} \leq \left(\frac{1}{2} \right)^{K-1}. \tag{15}$$

The inequality in (15) is strict for $K \geq 3$.

Proof (Sketch) The equality follows from Lemma 3 since $N^K / \langle N \rangle_K \to 1$ as $N \to \infty$. For the upper bound, let $g(x) = ((B+K)/2 - x)((B-K)/2 + x)$ and rewrite

$$\prod_{k=1}^{K-1} \left(\frac{B+K}{2} - k \right) = \varepsilon_{K,B} \prod_{k=1}^{\lfloor (K-1)/2 \rfloor} g(k), \quad \varepsilon_{K,B} = \begin{cases} 1, & \text{if } K \text{ is odd,} \\ B/2, & \text{if } K \text{ is even.} \end{cases}$$

Then the bound follows since g has a unique global maximum at $x = K/2$. □

Remark 6 If $K + B$ is odd then Lemma 5 and Theorem 5 yield for all $\beta \in (0, 1)$

$$\lim_{N \to \infty} d_{K,N,B} \left(\frac{\beta N}{B - 1}, \frac{N}{B - 1}, \dots, \frac{N}{B - 1}, \frac{(1 - \beta)N}{B - 1} \right)$$
$$= \frac{1}{2^{K-1}} \frac{\prod_{k=1}^{K-1}(B - 1 + K - 2k)}{(B - 1)^{K-1}} \le \left(\frac{1}{2} \right)^{K-1} \quad (16)$$

with a strict inequality for $K \ge 3$. Moreover, if we assume that Conjecture 1 is true, then (15) and (16) imply for any fixed number B of blocks

$$\lim_{N \to \infty} \sup \left\{ d_{K,N,B}(q); \ q \in \mathcal{Q}_{N,B} \right\}$$
$$= \frac{1}{2^{K-1}} \frac{\prod_{k=1}^{K-1}(B - \mathbb{1}\{K + B \text{ odd}\} + K - 2k)}{(B - \mathbb{1}\{K + B \text{ odd}\})^{K-1}} \le \left(\frac{1}{2} \right)^{K-1}$$

with $\mathcal{Q}_{N,B}$ defined as in (13). Moreover, the inequality above is strict for $K \ge 3$. Hence, $H_0 : \theta \in \Theta^0$ is rejected at an alternative for sufficiently large sample sizes N if the number of blocks in $(r_1(\theta), \dots, r_N(\theta))$ is uniformly bounded for all $\theta \in \Theta^0$ as $N \to \infty$.

4 Comparison of K-depth tests for different K

A proper choice for K is a crucial aspect to obtain a K-depth test with high power. This section contains some basic observations for the cases $K \le 6$, in particular in terms of power when only few sign changes are observed. A more profound comparison in applications will be done in Sect. 5.

As we will see in Sect. 4.1, the 2-depth test is usually a bad choice since it is equivalent to the classical sign test. This test struggles to reject the null hypothesis at alternatives that lead to a nearly equal amount of positive and negative residuals. K-depth tests with $K \ge 3$ can correctly identify and reject such alternatives as long as the number of sign changes in the residual vector is fairly low. A discussion on the p-values of the K-depth tests, $K = 3, \dots, 6$, for several different sample sizes can be found in Sect. 4.2.

4.1 Equivalence of the 2-depth test and the classical sign test

The test statistic of the classical sign test is given by

$$T_{\text{sign}}(\theta) := \frac{N_+(\theta) - N/2}{\sqrt{N}/2} \quad \text{where} \quad N_+(\theta) := \sum_{n=1}^{N} \mathbb{1}\{R_n(\theta) > 0\}$$

denotes the number of residuals with positive signs among the residual vector $(R_1(\theta), \dots, R_n(\theta))$. Assuming (1), this test statistic converges in distribution to the

standard normal distribution. Hence the classical sign test (in its asymptotic version) is defined via

$$\text{reject } H_0 : \theta \in \Theta^0 \text{ if for all } \theta \in \Theta^0 : T_{\text{sign}}(\theta) < u_{\frac{\alpha}{2}} \text{ or } T_{\text{sign}}(\theta) > u_{1-\frac{\alpha}{2}},$$

where u_α denotes the α-quantile of the standard normal distribution. Equivalently, one can define the classical sign test via

$$\text{reject } H_0 : \theta \in \Theta^0 \text{ if } \inf_{\theta \in \Theta^0} T_{\text{sign}}(\theta)^2 > \chi^2_{1,1-\alpha},$$

where $\chi^2_{1,\alpha}$ is the α-quantile of the χ^2_1 distribution. Note that $T_{\text{sign}}(\theta)^2$ is minimized if $N_+(\theta) = N/2$. Hence the test will not reject the null hypothesis if half of the residuals are positive.

To see the relationship to the 2-depth test, note that a pair of residuals has alternating signs if and only if one of them is positive and the other one is negative. Since we have $N_+(\theta)$ positive and $N - N_+(\theta)$ negative residuals (assuming $R_n(\theta) \neq 0$ P_θ-almost surely for all $n = 1, \ldots, N$), the 2-depth satisfies P_θ-almost surely:

$$d_2(R_1(\theta), \ldots, R_N(\theta)) = \frac{1}{\binom{N}{2}} N_+(\theta)(N - N_+(\theta)).$$

The 2-depth can be transformed into $T_{\text{sign}}(\theta)$ by using the identity

$$x(N - x) = -(x - N/2)^2 + N^2/4, \quad x \in \mathbb{R},$$

for $x = N_+(\theta)$. A straightforward calculation based on this identity reveals that the test statistic (4) satisfies for $K = 2$,

$$T_2(\theta) = \frac{N}{2(N-1)} - \frac{N}{2(N-1)} T_{\text{sign}}(\theta)^2 \quad P_\theta\text{-almost surely.}$$

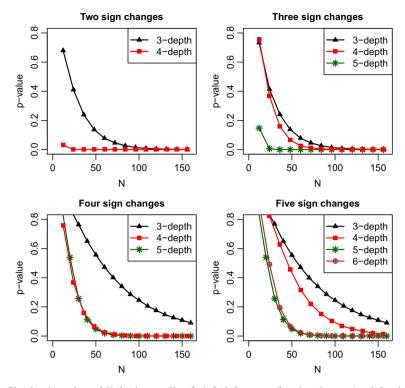Hence the 2-depth test and the classical sign test are equivalent.

4.2 Comparison of K-depth tests for $K \geq 3$

As we have seen in Sect. 3.3, K-depth tests with $K \geq 3$ are capable of rejecting nonfits that lead to a small number of sign changes, at least as long as the sample size N is sufficiently large. We will now take a closer look at the performance for small samples sizes up to $N = 160$.

Recall that, according to Conjecture 1, we assume that the maximal K-depth of a residual vector $r = (r_1, \ldots, r_N)$ with B blocks is given by

$$\eta_{K,N,B} := \begin{cases} d_{K,N,B}\left(\frac{N}{B}, \ldots, \frac{N}{B}\right), & \text{if } K + B \text{ is even,} \\ d_{K,N,B}\left(\frac{N}{2(B-1)}, \frac{N}{B-1}, \ldots, \frac{N}{B-1}, \frac{N}{2(B-1)}\right), & \text{if } K + B \text{ is odd.} \end{cases}$$

Fig. 2 Simulated p-values of K-depth tests, $K = 3, 4, 5, 6$, for two to five sign changes (top left to bottom right) at different sample sizes N

Hence, the test statistic (4) for a residual vector with B blocks can be at most

$$\widetilde{\eta}_{K,N,B} := N\left(\eta_{K,N,B} - \left(\frac{1}{2}\right)^{K-1}\right).$$

Figure 2 contains the p-values when observing a value of $\widetilde{\eta}_{K,N,B}$ for $B = 3, 4, 5, 6$ blocks or 2, 3, 4, 5 sign changes, respectively, i.e. the probabilities

$$P_\theta\left(T_K(R_1(\theta), \ldots, R_N(\theta)) \leq \widetilde{\eta}_{K,N,B}\right)$$

are plotted for samples sizes N between 10 and 160 and $K = 3, 4, 5, 6$.

Recall that if a residual vector has B block, i.e. $B - 1$ sign changes, then K-depth tests with $K > B$ will automatically reject the null hypothesis as soon as the sample size is large enough to make a rejection possible for the test. Figure 2 thus only contains K-depth tests with $K \leq 4$ for situations with two sign changes to highlight that the p-value of the 4-depth test indeed becomes 0 if N is sufficiently large. The same applies to the 5-depth test when three sign changes occur. The other two plots (four and five sign changes) do not contain the corresponding 6- and 7-depth tests since their p-values behave similarly.

All four subfigures of Fig. 2 indicate that the p-values of all considered K-depth tests are decreasing to zero for growing sample size. They decrease more slowly for $K = 3, 4$ than for $K = 5, 6$, but even the p-value of the 3-depth test reaches 0.1 for a sample size greater than $N = 150$. It is remarkable that the p-values of the K-depth tests with $K = B - 1$ and $K = B$ are always very similar for all $B - 1 = 3, 4, 5$ sign changes we considered. However, this does not hold for $B - 1 = 2$ since the 2-depth test is the classical sign test which always has a p-value of 1 in the case of two blocks of equal size.

5 Applications

The high power of 3-depth tests in the case of two unknown parameters was already shown for explosive AR(1) models, namely in Kustosz et al. (2016) for linear AR(1)-models given by $Y_n = \theta_0 + \theta_1 Y_{n-1} + E_n$ and in Kustosz et al. (2016) for nonlinear AR(1)-models given by $Y_n = Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + E_n$, see also Falkenau (2016). In particular these results showed for normally distributed errors E_n that 3-depth tests possess similarly high power compared to classical tests based on least squares.

Other results for the quadratic regression model, a nonlinear AR(1)-model and an explosive AR(2)-model, each with three unknown parameters, can be found in the supplementary material. These examples show that there is not much difference in the power of the 3-depth test, the 4-depth test, and the classical F- and t-test, respectively, if the sample size is large enough, which means close to 100. There are only relevant differences if the sample size is small. See for example Fig. 3 for testing $H_0 : \theta = (1, 0, 1)^\top$ in a quadratic regression model given by $Y_n = \theta_0 + \theta_1 x_n + \theta_2 x_n^2 + E_n$ with $\theta = (\theta_0, \theta_1, \theta_2)^\top$. This example concerns normally distributed errors, but the results are very similar for Cauchy distributed errors. The only exception is the F-test which loses much power if the errors have a Cauchy distribution. See the supplementary material for the behavior of Cauchy distributed errors and for other alternatives.

Additionally, we demonstrate here the good power of K-depth tests with $K = 21$ and $K = 38$ for a high-dimensional multiple regression model given by $Y_n = \sum_{d=1}^{D} \theta_d x_{nd} + E_n$ with $D \in \{10, 20, 40, 80\}$ and $N = 100$. The regressors are ordered by computing a shortest path through the multidimensional data. This is done here by the Shortest Hamiltonian Path (SHP), see for example Applegate et al. (2006). Horn and Müller (2020) show that this ordering is superior to other possibilities for ordering. The SHP belongs to the NP-hard problems. In particular, any known exact algorithm to compute it has exponential time complexity in the number of data points in the worst case. However, empirically the runtimes are quite small for medium numbers of observations, see Horn and Müller (2020) or Horn (2021).

The tested hypothesis is $H_0 : \theta_d = 0 \; \forall d = 1, \ldots, D$ vs. $H_1 : \exists d = 1, \ldots D : \theta_d \neq 0$. The 21-depth test and the 38-depth test are compared with the classical F-test and the sign test as well as a robust Wald test and a robust score test. For the Wald test, estimators of the parameters and covariance matrix of an MM-regression obtained by the function `lmRob()` from the R-package `robust` (Wang et al. 2019) are used. For the robust score test, a self implemented R-function is used based on a high-dimensional version of the procedure from Khan and Yunus (2014). The scores are computed by the
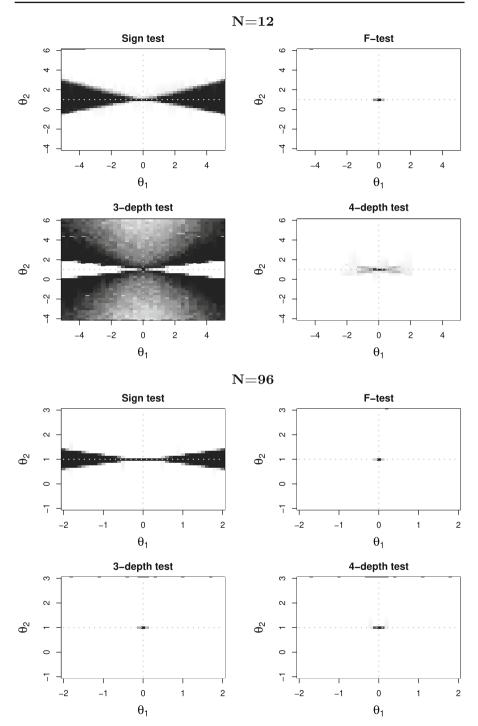
Fig. 3 Simulated power of the sign test, the F-test, the 3-depth test, and the 4-depth test in the quadratic regression model with normally distributed errors with sample sizes $N = 12$ (upper part) and $N = 96$ (lower part), where the component θ_0 is fixed to 1 (20 gray levels were used, where black corresponds to [0, 0.05] and white to (0.95, 1])
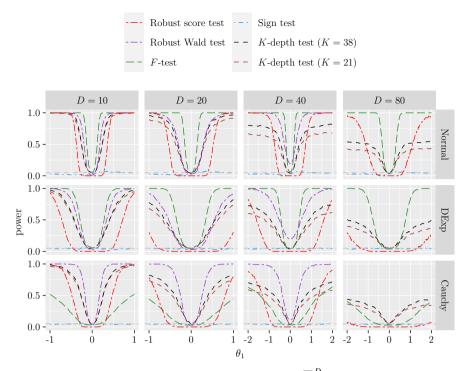
Fig. 4 Extracts of the simulated power functions for the model $Y_n = \sum_{d=1}^{D} \theta_d x_{nd} + E_n$. Here, the power functions are only shown for $\theta_1 \in [-1, 1]$ or $[-2, 2]$, all other values of θ are zero. The K-depth tests are conducted with an ordering according to the exact solution of the Shortest Hamiltonian Path problem. The gray dashed line shows the level of the test $\alpha = 0.05$

R-function `psi.weight()` using the setting `ips = 4` from the package `robust` and the scale factor is estimated by `lmRob.S()` from the package `robustbase`. The performance of the tests is measured in three different situations: Firstly with normally distributed errors E_n, secondly with double exponentially distributed errors, and thirdly with Cauchy distributed errors. Because of the high dimensionality, the complete power functions cannot be shown, but only some aspects. Here, it is looked at the aspect $\lambda(\theta) = \theta_1 \in [-1, 1]$, where all other components of θ are set to zero. The power was simulated with 1000 repetitions for the K-depth test, F-test and sign test and with 500 repetitions for the robust Wald test and robust score test at 101 or 201 equidistant points within $[-1, 1]$ or $[-2, 2]$, respectively. Because of the symmetry of the model in θ, the power functions look the same for all aspects $\lambda(\theta) = \theta_d$, $d = 1, \ldots, D$. Similar results are obtained if other alternatives like $\theta_1 = \ldots = \theta_D = \gamma, \gamma \in [-1, 1]$, are considered, see the supplementary material.

Figure 4 shows the extracts of the simulated power functions for the considered aspect. Firstly, this figure shows that the K-depth test performs better for higher K, e.g., $K = 38$ performs better than for $K = 21$ in Fig. 4 or $K = 5$ in the supplementary material. In general, it holds that K should have at least the same magnitude as D to reach good results of the K-depth test. It can be nicely seen in Fig. 4 that the power

of the K-depth test for $K = 21$ is satisfying for $D = 10$ and $D = 20$, whereas it is worse for $D = 40$ and $D = 80$ compared to the case $K = 38$. Secondly, the robust Wald test performs well for $D = 10$ and $D = 20$, but for $D = 40$, the level is not maintained, i.e., the power values are much larger than $\alpha = 0.05$ at H_0, and for $D = 80$, the robust Wald test cannot be carried out at all. Indeed, it is not the Wald test itself which causes the problems, but the underlying MM-estimation. In our simulations, the R-function `lmRob()` always threw an error when trying to calculate the estimator for $D = 80$ dimensions and $N = 100$ data points stating that internally a matrix cannot be inverted because of numerical singularity. Similar problems appear when using `lmrob()` from the R-package `robustbase` instead. In contrast to this, the K-depth tests or the robust score test remain computable for such high dimensions although the power of the K-depth test is not very good due to values of K much smaller than D. The robust score test has a very small power if θ_1 is closer to zero but the power function increases more strongly for higher deviations.

Furthermore, Fig. 4 shows that of course the F-test performs best when having normally distributed errors. But for Cauchy distributed errors, the K-depth test is better than the F-test. The classical sign test performs poorly regardless of what the dimension D is. Its power is always about 0.05. Furthermore, the cases $D = 10$ and $K = 21$ or $D = 20$ and $K = 38$ show that the K-depth test can keep up with the robust Wald test, the robust score test and the F-test (for normally distributed errors) when K is sufficient large in comparison to D. Unfortunately, the parameter K of the K-depth test cannot be chosen arbitrarily high for fixed N, since otherwise the test is unable to reject at all due to the circumstance that the α-quantile can then coincide with its minimal value. Some benchmarks how high the parameter K can be chosen for given N are given in (Malcherczyk (2022), Chapter 6.3.2). For larger sample sizes, the power of the K-depth tests increases significantly for all considered dimensions D, but still do not reach the power of the robust Wald test which is computable then. See the supplementary material for $N = 500$.

The results in this section were computed with the help of the R-package `GSignTest` (Horn 2020). For computing the SHP, the package `TSP` (Hahsler and Hornik 2019) and the "Concorde"-solver (Applegate et al. 2004) were used. Graphics were made with the help of the packages `rgl` (Adler and Murdoch 2020) and `ggplot2` (Wickham 2016).

Supplementary material. A file with full proofs, more details on the block implementation and further simulation results can be found under the following link: https://doi.org/10.1007s00362-022-01337-5.

6 Discussion and outlook

K-sign depth can be used to define simple robust tests which we refer to as K-depth tests. While the parameter choice $K = 2$ essentially leads to the classical sign test and thus has several limitations in rejecting alternatives, K-depth tests for $K \geq 3$ are fairly powerful. They are not as powerful as the complicated robust Wald tests based on MM-estimators but can outperform classical approaches such as the F-test, in particular in the presence of outliers. The K-depth tests are not very well-suited

for small sample sizes and models where the number of sign changes in the residual vector is likely to exceed $K-1$ at alternatives. However, the K-depth tests perform very well in our examples once the sample size is sufficiently large.

The K-depth test can also be used when having no inherent order in the data, like for example for multiple regression. For this, ordering the regressors according to a Shortest Hamiltonian Path leads to very good power of the test for rather low dimensions. In higher dimensions, the parameter K should be of the same magnitude as the number of dimensions. When this is not possible, the power of the K-depth test decreases. However, in contrast to the robust Wald test based on MM-estimators, it still works without any errors caused by numerical issues.

To reduce the runtime of $\Theta(N^K)$ of the definition of the K-depth, a faster block implementation is presented which leads to an algorithm with linear runtime. A linear runtime of an asymptotically equivalent form can also be obtained by the derivation of the asymptotic distribution of the K-depth for $K \geq 3$, see Malcherczyk et al. (2021).

Although the simulation study in this article only deals with one-point hypotheses, the K-sign depth can also be used to test general hypotheses of the form $H_0 : \theta \in \Theta_0$. In this case, the maximal value of the test statistic in Θ_0 must be computed. However, more research is necessary to find an efficient algorithm for this maximum.

Moreover, this paper is mainly focused on the one-sided version of K-depth test to detect shifts in the medians of the residuals. A two-sided version of the K-depth test can also detect dependence structures within the residuals and may be useful for stationary AR-models and other stationary processes. Once again, further research is necessary to compare the two-sided K-depth test with other approaches when testing simultaneously whether residuals are independent and have medians equal to zero.

A possible extension of the presented approach to multivariate observations might be possible. In particular, multivariate sign changes based on the multivariate spatial sign of Möttönen and Oja (1995) could be used as in Paindaveine (2009) for counting the K-tuples with $K-1$ sign changes. This would lead to a multivariate K-sign depth. However, it is not clear how to transfer the concept of blocks as used in this paper.

References

Adler D, Murdoch D (2020) rgl: 3D visualization using OpenGL. R package version 0.100.54

Agostinelli C, Romanazzi M (2011) Local depth. J Stat Plan Inference 141:817–830

Applegate D, Bixby R, Chvátal V, Cook W (2004) Concorde tsp solver

Applegate D, Bixby R, Chvátal V, Cook W (2006) The traveling salesman problem. Applied mathematics. Princton University Press, Princton

Arcones MA, Gine E (1993) Limit theorems for U-processes. Ann Probab 21(3):1494–1542

Claeskens G, Hubert M, Slaets L, Vakili K (2014) Multivariate functional halfspace depth. J Am Stat Assoc 109:411–423

Denecke L, Müller CH (2014) Consistency of the likelihood depth estimator for the correlation coefficient. Stat Pap 55:3–13

Dong Y, Lee S (2014) Depth functions as measures of representativeness. Stat Pap 55:1079–1105

Dümbgen L (1992) Limit theorems for the simplicial depth. Stat Probab Lett 14:119–128

Falkenau CP (2016) Depth based estimators and tests for autoregressive processes with application on crack growth and oil prices. Dissertation, TU Dortmund. https://doi.org/10.17877/DE290R-17269

Gibbons J, Chakraborti S (2003) Nonparametric statistical inference. Statistics, textbooks and monographs. Marcel Dekker Incorporated, New York

Hahsler M, Hornik K (2019). TSP: traveling salesperson problem (TSP). R package version 1.1-7

Hampel F, Ronchetti E, Rousseeuw P, Stahel W (2011) Robust statistics: the approach based on influence functions. Wiley, New York

Heritier S, Ronchetti E (1994) Robust bounded-influence tests in general parametric models. J Am Stat Assoc 89(427):897–904

Horn M (2020) GSignTest: robust tests for regression-parameters via sign depth. R package version 1.0.5, https://github.com/melaniehorn/GSignTest

Horn M (2021) Sign depth for parameter tests in multiple regression. Dissertation, TU Dortmund. http://dx.doi.org/10.17877/DE290R-22355

Horn M, Müller CH (2020) Tests based on sign depth for multiple regression. SFB Discussion Paper 07/20. https://doi.org/10.17877/DE290R-20984

Huber P, Ronchetti E (2009) Robust statistics. Wiley, New York

Khan S, Yunus R (2014) Score test in robust m-procedure. J Appl Probab Stat 9:55–75

Koller M, Stahel WA (2011) Sharpening Wald-type inference in robust regression for small samples. Comput Stat Data Anal 55(8):2504–2515

Koller M, Stahel WA (2017) Nonsingular subsampling for regression s estimators with categorical predictors. Comput Stat 32(2):631–646

Kustosz CP, Leucht A, Müller CH (2016) Tests based on simplicial depth for AR(1) models with explosion. J Time Ser Anal 37:763–784

Kustosz CP, Müller CH, Wendler M (2016) Simplified simplicial depth for regression and autoregressive growth processes. J Stat Plan Inference 173:125–146

Leckey K, Müller CH, Szugat S, Maurer R (2020) Prediction intervals for load-sharing systems in accelerated life testing. Qual Reliab Eng Int 36(6):1895–1915

Liu RY (1988) On a notion of simplicial depth. Proc Natl Acad Sci USA 85:1732–1734

Liu RY (1990) On a notion of data depth based on random simplices. Ann Stat 18:405–414

Liu X, Luo S, Zuo Y (2020) Some results on the computing of Tukey's halfspace median. Stat Pap 61:303–316

Lok WS, Lee SMS (2011) A new statistical depth function with application to multimodal data. J Nonparametric Stat 23:617–631

López-Pintado S, Sun Y, Lin JK, Genton MG (2014) Simplicial band depth for multivariate functional data. Adv Data Anal Classif 8(3):321–338

Malcherczyk D (2022) K-sign depth: asymptotic distribution, efficient computation and applications. Dissertation, TU Dortmund. https://doi.org/10.17877/DE290R-22644

Malcherczyk D, Leckey K, Müller CH (2021) K-sign depth: from asymptotics to efficient implementation. J Stat Plan Inference 215:344–355

Markatou M, Stahel WA, Ronchetti E (1991) Robust M-type testing procedures for linear models. In: Stahel W, Weisberg S (eds) Directions in robust statistics and diagnostics: Part I. Springer, New York, pp 201–220

Maronna R, Martin D, Yohai V, Salibián-Barrera M (2019) Robust statistics: theory and methods (with R). Wiley, New York

Mizera I, Müller CH (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, vol 165. Springer, New York

Möttönen J, Oja H (1995) Multivariate spatial sign and rank methods. J Nonparametric Stat 5(2):201–213

Müller CH (2005) Depth estimators and tests based on the likelihood principle with application to regression. J Multivar Anal 95(1):153–181

Nagy S, Ferraty F (2019) Data depth for measurable noisy random functions. J Multivar Anal 170:95–114

Paindaveine D (2009) On multivariate runs tests for randomness. J Am Stat Assoc 104(488):1525–1538

Paindaveine D, Van Bever G (2013) From depth to local depth: a focus on centrality. J Am Stat Assoc 108:1105–1119

Paindaveine D, Van Bever G (2018) Halfspace depths for scatter, concentration and shape matrices. Ann Stat 46(6B):3276–3307

Rousseeuw P, Yohai V (1984) Robust regression by means of S-estimators. In: Franke J, Härdle W, Martin D (eds) Robust and nonlinear time series analysis. Springer, New York, pp 256–272

Rousseeuw PJ, Hubert M (1999) Regression depth. J Am Stat Assoc 94(446):388–402

Schrader RM, Hettmansperger TP (1980) Robust analysis of variance based upon a likelihood ratio criterion. Biometrika 67(1):93–101

Silvapulle MJ (1992) Robust tests of inequality constraints and one-sided hypotheses in the linear model. Biometrika 79(3):621–630

Tukey JW (1975) Mathematics and the picturing of data. Proc Int Congress Math 2:523–531

Wald A, Wolfowitz J (1940) On a test whether two samples are from the same population. Ann Math Stat 11(2):147–162

Wang J (2019) Asymptotics of generalized depth-based spread processes and applications. J Multivar Anal 169:363–380

Wang J, Zamar R, Marazzi A, Yohai V, Salibian-Barrera M, Maronna R, Zivot E, Rocke D, Martin D, Maechler M,Konis K (2019) Robust: Port of the S+ "Robust Library". R package version 0.4-18.2

Wellmann R, Harmand P, Müller CH (2009) Distribution-free tests for polynomial regression based on simplicial depth. J Multivar Anal 100(4):622–635

Wellmann R, Müller CH (2010) Tests for multiple regression based on simplicial depth. J Multivar Anal 101(4):824–838

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, New York

Yohai V (1987) High breakdown-point and high efficiency robust estimates for regression. Ann Stat 15:642–656

Zuo Y, Serfling R (2000) General notions of statistical depth function. Ann Stat 28:461–482