Supplement of "Simple powerful robust tests based on sign depth"

Kevin Leckey, Dennis A. Malcherczyk, Melanie Horn, and Christine H. Müller *

March 22, 2021

We provide further results for the block implementation in Section 1, more details for the proofs in Section 2, and further simulation results in Section 3. We repeat some parts of the main article to improve the readability. As an overview, the following details were added to each section:

- Most of Section 1 is copied from the main article. The only new details concern Remark 1. It now also covers the linear implementation for $K = 4$ to showcase how the idea can be generalized from $K = 3$ to arbitrary K.
- Section 2 contains all theorems, lemmas and other statements from the main article and added a full proof to each statement. It also contains a proof to the conjecture for the special case $K = 3$ as well as a short summary of the missing gaps to prove it for $K \geq 4$.
- Finally, Section 3 contains additional simulation results such as a quadratic regression model, an AR(2)-model, a nonlinear AR(1)-model and some more results on the multiple regression model from the main article.

Before presenting these extended results, first recall that we consider residuals $R_1(\theta), \ldots, R_N(\theta)$ from a parametric model and assume that these residuals satisfy for every possible model parameter $\theta \in \Theta$ and every $n = 1, \ldots, N$ that

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

*Department of Statistics, TU Dortmund University, D-44227 Dortmund, Germany, kevin.leckey@tu-dortmund.de, dennis.malcherczyk@tu-dortmund.de, mhorn@statistik.tu-dortmund.de, cmueller@statistik.tu-dortmund.de

Also recall that the K-depth of a vector $(r_1, \ldots, r_N) \in \mathbb{R}^N$ is defined as

$$
\begin{aligned}
d_K(r_1, \ldots, r_N) := \frac{1}{\binom{N}{K}} \sum_{1 \le n_1 < n_2 < \ldots < n_K \le N} \Bigg(& \prod_{k=1}^{K} \mathbb{1}\left\{(-1)^k r_{n_k} > 0\right\} \\
& + \prod_{k=1}^{K} \mathbb{1}\left\{(-1)^k r_{n_k} < 0\right\} \Bigg).
\end{aligned} \tag{2}
$$

1 The 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

$$
\begin{aligned}
B(r) &:= 1 + \sum_{n=2}^{N} \mathbb{1}\left\{\psi(r_{n-1}) \neq \psi(r_n)\right\}, \\
s_b(r) &:= \min\left\{\ell > s_{b-1}(r);\, \psi(r_\ell) \neq \psi(r_{\ell-1})\right\}, \quad b = 2, \ldots, B(r).
\end{aligned}
$$

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 nth residual r_n *belongs to block* j if and only if $s_j(r) \le n < s_{j+1}(r)$. 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 (2):

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

$$d_K(r_1,\ldots,r_N) = d_{K,N,B(r)}(q_1(r),\ldots,q_{B(r)}(r)). \tag{3}$$

Remark 1. 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 first consider the implementation for $K \in \{3, 4\}$. For $K = 3$, consider $d_{3,N,B}(q_1,\ldots,q_B)$ from Lemma 1 and note that factoring out the length q_{i_2} of the second block yields

$$\begin{aligned} d_{3,N,B}(q_1,\ldots,q_B) =& \frac{1}{\binom{N}{3}} \sum_{(i_1,i_2,i_3)\in\mathcal{A}_{3,B}} q_{i_1}q_{i_2}q_{i_3} \\ =& \frac{1}{\binom{N}{3}} \sum_{i_2=2}^{B-1} q_{i_2} \left(\sum_{\substack{i_1=1\\ i_2-i_1 \text{ odd}}}^{i_2-1} q_{i_1}\right)\left(\sum_{\substack{i_3=i_2+1\\ i_3-i_2 \text{ odd}}}^{B} q_{i_3}\right). \end{aligned} \tag{4}$$

Next compute the following forward- and backward cumulative sums:

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

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, (4) 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 $K = 4$, we have

$$d_{4,N,B}(q_1,\ldots,q_B) = \frac{1}{\binom{N}{4}} \sum_{i_2=2}^{B-2} q_{i_2} \left(\sum_{\substack{i_1=1\\ i_2-i_1 \text{ odd}}}^{i_2-1} q_{i_1}\right)\left(\sum_{\substack{i_3=i_2+1\\ i_3-i_2 \text{ odd}}}^{B-1} q_{i_3} \sum_{\substack{i_4=i_3+1\\ i_4-i_3 \text{ odd}}}^{B} q_{i_4}\right). \tag{5}$$

Once again, the formula above can be computed in linear time if the inner sums are computed in advance. Since the first of the inner sums equals $\mathcal{F}(i_2)$ from the previous case, it only remains to efficiently compute

$$\mathcal{B}_2(i_2) := \sum_{\substack{i_3=i_2+1\\ i_3-i_2 \text{ odd}}}^{B-1} q_{i_3} \sum_{\substack{i_4=i_3+1\\ i_4-i_3 \text{ odd}}}^{B} q_{i_4} = \sum_{\substack{i_3=i_2+1\\ i_3-i_2 \text{ odd}}}^{B-1} q_{i_3}\mathcal{B}(i_3), \quad i_2 = 2,\dots,B-1$$

where $\mathcal{B}(i)$, $i = 2,\dots,B-1$, is defined as in the previous case and can be computed in linear time. Hence, by computing these values in advance we may also compute $\mathcal{B}_2(i_2)$, $i_2 = 2,\dots,B-1$ in linear time. Therefore we may also compute (5) with a total of $\Theta(B)$ operations. The other cases $K \geq 5$ essentially only require an iterative computation of the terms

$$\mathcal{B}_j(i_2) = \sum_{\substack{i_3=i_2+1\\ i_3-i_2 \text{ odd}}}^{B-j+1} q_{i_3}\mathcal{B}_{j-1}(i_3), \quad i_2 = 2,\dots,B-j,$$

for $j = 2,\dots,K-2$ where $\mathcal{B}_1(i_2)$ equals $\mathcal{B}(i_2)$ from the case $K = 3$. Computing these terms require a total of $\Theta(KB)$ operations, which remains linear in B for any constant K. Then the K-depth is given by

$$d_{K,N,B}(q_1,\dots,q_B) = \frac{1}{\binom{N}{K}} \sum_{i_2=2}^{B-K+2} q_{i_2}\mathcal{F}(i_2)\mathcal{B}_{K-2}(i_2).$$

2 Proofs

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

$$\begin{aligned}
&\prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k > 0\} + \prod_{k=1}^{K} \mathbb{1}\{E_{n_k}(-1)^k < 0\} - \left(\frac{1}{2}\right)^{K-1} \\
&= \frac{1}{2^{K-1}} \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1\leq i(1)<\dots<i(2L)\leq K} \prod_{j=1}^{2L} (-1)^{i(j)}\psi\left(E_{n_{i(j)}}\right) \quad P\text{-almost surely,}
\end{aligned} \tag{6}$$

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

Proof of Lemma 2. In order to simplify the notation, we assume $(n_1,\dots,n_K) = (1,\dots,K)$. Note for $x \neq 0$

$$\mathbb{1}\{x > 0\} = \frac{1}{2}(\psi(x) + 1), \qquad \mathbb{1}\{x < 0\} = \frac{1}{2}(-\psi(x) + 1).$$

It is straightforward to check $\prod_{i=1}^{K}(a_i + 1) = \sum_{\ell=1}^{K} \sum_{1\leq i(1)<\dots<i(\ell)\leq K} \prod_{j=1}^{\ell} a_{i(j)} + 1$ for arbitrary $a_1,\dots,a_K$. This implies P-almost surely

$$\prod_{k=1}^{K} \mathbb{1}\{E_k(-1)^k > 0\} = \frac{1}{2^K} \prod_{k=1}^{K} \left((-1)^k\psi(E_k) + 1\right)$$

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

Similarly

$$\begin{aligned}
&\prod_{k=1}^{K}\mathbb{1}\{E_k(-1)^k<0\}\\
&=\frac{1}{2^K}\left(\sum_{\ell=1}^{K}\sum_{1\le i(1)<\ldots<i(\ell)\le K}(-1)^{i(1)+\cdots+i(\ell)+\ell}\prod_{j=1}^{\ell}\psi\left(E_{i(j)}\right)+1\right)\\
&=\frac{1}{2^K}\left(\sum_{\substack{\ell=1,\ldots,K\\ \ell\text{ even}}}\sum_{1\le i(1)<\ldots<i(\ell)\le K}(-1)^{i(1)+\cdots+i(\ell)}\prod_{j=1}^{\ell}\psi\left(E_{i(j)}\right)+1\right)\\
&\quad-\frac{1}{2^K}\sum_{\substack{\ell=1,\ldots,K\\ \ell\text{ odd}}}\sum_{1\le i(1)<\ldots<i(\ell)\le K}(-1)^{i(1)+\cdots+i(\ell)}\prod_{j=1}^{\ell}\psi\left(E_{i(j)}\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\prod_{k=1}^{K}\mathbb{1}\{E_k(-1)^k>0\}+\prod_{k=1}^{K}\mathbb{1}\{E_k(-1)^k<0\}\\
&=\frac{1}{2^{K-1}}\left(\sum_{\substack{\ell=1,\ldots,K\\ \ell\text{ even}}}\sum_{1\le i(1)<\ldots<i(\ell)\le K}(-1)^{i(1)+\cdots+i(\ell)}\prod_{j=1}^{\ell}\psi\left(E_{i(j)}\right)+1\right)\\
&=\left(\frac{1}{2}\right)^{K-1}+\frac{1}{2^{K-1}}\sum_{L=1}^{\left\lfloor\frac{K}{2}\right\rfloor}\sum_{1\le i(1)<\ldots<i(2L)\le K}(-1)^{i(1)+\cdots+i(2L)}\prod_{j=1}^{2L}\psi(E_{i(j)})
\end{aligned}$$

and the assertion follows. □

Theorem 1. *Let $K\ge 2$. 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 of Theorem 1. Set $R_n=R_n(\theta)$. Lemma 2 yields

$$\begin{aligned}
&d_K(R_1,\ldots,R_N)-\left(\frac{1}{2}\right)^{K-1}\\
&=\frac{1}{\binom{N}{K}}\sum_{1\le n_1<n_2<\ldots<n_K\le N}\frac{1}{2^{K-1}}\sum_{L=1}^{\left\lfloor\frac{K}{2}\right\rfloor}\sum_{1\le i(1)<\ldots<i(2L)\le K}\prod_{j=1}^{2L}(-1)^{i(j)}\psi\left(R_{n_{i(j)}}\right)
\end{aligned}$$

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

$$v := \sum_{L=1}^{\lfloor \frac{K}{2} \rfloor} \sum_{1 \le i(1) < \ldots < i(2L) \le K} 1$$

for the number of summands in the representation of K alternating signs given by Lemma 2. This number depends only on K and not on N. First of all, we show that each of these v summands is converging in probability to zero.

To this end, let $L = 1, \ldots, \lfloor \frac{K}{2} \rfloor$ and $1 \le i(1) < \ldots < i(2L) \le K$ be arbitrary. We consider the summand multiplied by the factor 2^{K-1}. Because $\mathbb{E}_\theta(\psi(R_n)) = 0$ and $R_1, \ldots, R_N$ are independent, we get at once for this summand

$$\mathbb{E}_\theta \left(\frac{1}{\binom{N}{K}} \sum_{1 \le n_1 < n_2 < \ldots < n_K \le N} (-1)^{i(1)+\ldots+i(2L)} \prod_{j=1}^{2L} \psi\left(R_{n_{i(j)}}\right) \right) = 0.$$

Moreover, $\psi(R_n)^2 = 1$ P_θ-almost surely implies

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

Hence

$$\begin{aligned}
&\operatorname{var}_\theta \left(\frac{1}{\binom{N}{K}} \sum_{1 \le n_1 < \ldots < n_K \le N} (-1)^{i(1)+\ldots+i(2L)} \prod_{j=1}^{2L} \psi\left(R_{n_{i(j)}}\right) \right) \\
&= \frac{1}{\binom{N}{K}^2} \sum_{1 \le n_1 < \ldots < n_K \le N} \sum_{1 \le \tilde{n}_1 < \ldots < \tilde{n}_K \le N} \mathbb{E}_\theta \left(\prod_{j=1}^{2L} \psi\left(R_{n_{i(j)}}\right) \prod_{j=1}^{2L} \psi\left(R_{\tilde{n}_{i(j)}}\right) \right) \\
&= \frac{1}{\binom{N}{K}^2} \sum_{\substack{1 \le n_1 < \ldots < n_K \le N,\ 1 \le \tilde{n}_1 < \ldots < \tilde{n}_K \le N \\ n_{i(j)} = \tilde{n}_{i(j)} \text{ for } j=1,\ldots,2L}} 1 \\
&\le \frac{1}{\binom{N}{K}^2} \sum_{1 \le n_1 < \ldots < n_{2L} \le N} \sum_{n_{2L+1}, \ldots, n_K \in \{1,\ldots,N\}} \sum_{\tilde{n}_{2L+1}, \ldots, \tilde{n}_K \in \{1,\ldots,N\}} 1 \\
&= \frac{\binom{N}{2L} N^{K-2L} N^{K-2L}}{\binom{N}{K}^2} \le \frac{(K!)^2}{(2L)!} \frac{N^{2L+2K-4L}}{(N-(K+1))^{2K}} \\
&= \frac{(K!)^2}{(2L)!} \frac{1}{N^{2L}} \frac{1}{\left(1 - \frac{K+1}{N}\right)^{2K}} \longrightarrow 0
\end{aligned}$$

for $N \to \infty$ so that Chebyshev inequality provides the convergence in probability to zero. Furthermore, the convergence in probability is sufficiently quick of order $O(N^{-2L})$ so that the Borel-Cantelli lemma implies the convergence to zero P_θ-almost surely. □

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

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

Proof of Theorem 2. Let $r_1, \dots, r_N$ be residuals with alternating signs. First note that $r_{i_1}, \dots, r_{i_K}$ are alternating if and only if $(i_1, \dots, i_K) \in \mathcal{A}_{K,N}$ with $\mathcal{A}_{K,N}$ defined as in Lemma 1. Hence

$$d_K(r_1, \dots, r_N) = \frac{|\mathcal{A}_{K,N}|}{\binom{N}{K}}$$

where $|\mathcal{A}_{K,N}|$ denotes the size of $\mathcal{A}_{K,N}$. Thus it only remains to determine this size.

In the subsequent analysis, we write $\mathbb{O}$ for the set of all odd positive integers, i.e. $\mathbb{O} = 2\mathbb{N}_0 + 1$. For a vector $(i_1, \dots, i_K)$ let $\Delta_1 := i_1$ and $\Delta_k := i_k - i_{k-1}$ for $k = 2, \dots, K$. Note that $(i_1, \dots, i_K) \in \mathcal{A}_{K,N}$ if and only if $(\Delta_1, \dots, \Delta_K)$ is part of the set

$$\mathcal{D}_{K,N} := \left\{ (\Delta_1, \dots, \Delta_K) \in \mathbb{N} \times \mathbb{O}^{K-1};\ \sum_{k=1}^{K} \Delta_k \leq N \right\}.$$

Hence $|\mathcal{A}_{K,N}| = |\mathcal{D}_{K,N}|$. In order to remove the additional condition $\Delta_k \in \mathbb{O}$ for $k \geq 2$, we will use the transformation $\widetilde{\Delta}_k = (\Delta_k + 1)/2$. Since this transformation for $k = 1$ only provides an integer if Δ_1 is odd, we additionally split the set into the two parts

$$\mathcal{D}^-_{K,N} := \{(\Delta_1, \dots, \Delta_K) \in \mathcal{D}_{K,N};\ \Delta_1 \in \mathbb{O}\},$$
$$\mathcal{D}^+_{K,N} := \{(\Delta_1, \dots, \Delta_K) \in \mathcal{D}_{K,N};\ \Delta_1 \notin \mathbb{O}\}.$$

The elements of $\mathcal{D}^-_{K,N}$ can be counted by noting that $(\Delta_1, \dots, \Delta_K) \in \mathcal{D}^-_{K,N}$ if and only if

$$\left(\widetilde{\Delta}_1, \dots, \widetilde{\Delta}_K\right) \in \widetilde{\mathcal{D}}^-_{K,N} := \left\{ (n_1, \dots, n_K) \in \mathbb{N}^K;\ \sum_{k=1}^{K} n_k \leq \frac{N+K}{2} \right\}$$

with $\widetilde{\Delta}_k = (\Delta_k + 1)/2$ for $k = 1, \dots, K$. Similarly, $(\Delta_1, \dots, \Delta_K) \in \mathcal{D}^+_{K,N}$ if and only if

$$\left(\frac{\Delta_1}{2}, \widetilde{\Delta}_2, \dots, \widetilde{\Delta}_K\right) \in \widetilde{\mathcal{D}}^+_{K,N} := \left\{ (n_1, \dots, n_K) \in \mathbb{N}^K;\ \sum_{k=1}^{K} n_k \leq \frac{N+K-1}{2} \right\}$$

with $\widetilde{\Delta}_k$ as above. In summary, the (bijective) transformations discussed above yield

$$|\mathcal{A}_{K,N}| = \left|\widetilde{\mathcal{D}}^-_{K,N}\right| + \left|\widetilde{\mathcal{D}}^+_{K,N}\right|. \tag{7}$$

The sizes of the remaining sets can easily be determined by noting that each element $(n_1, \dots, n_K)$ in $\widetilde{\mathcal{D}}^-_{K,N}$ corresponds to a K-element subset $\{m_1, \dots, m_K\}$ of the set $\{1, 2, \dots, \lfloor (N+K)/2 \rfloor\}$ by letting

$$m_k := \sum_{i=1}^{k} n_i \quad \text{for } k = 1, \dots, K.$$

Hence

$$\left|\widetilde{\mathcal{D}}^-_{K,N}\right| = \binom{\lfloor (N+K)/2 \rfloor}{K}.$$

Essentially the same arguments yield

$$\left|\widetilde{\mathcal{D}}^+_{K,N}\right| = \binom{\lfloor (N+K-1)/2 \rfloor}{K}.$$

The assertion follows after rewriting $\lfloor (N+K-1)/2 \rfloor = \lceil (N+K-2)/2 \rceil$ and by plugging the sizes of the sets back into (7). $\square$

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 $\mathcal{A}_{K,B}$.

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 of Lemma 3. First note that if $(r_1, \ldots, r_N)$ consists of B blocks and each block has size $M = N/B$, then Lemma 1 and Corollary 1 yield

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

Since binomial coefficients satisfy $\dbinom{x}{K} = \dfrac{\langle x \rangle_K}{K!}$ for $x \geq K$, this can be simplified to

$$d_K(r_1, \ldots, r_N) = \frac{N^K}{B^K \langle N \rangle_K} \left(\langle \lfloor (B+K)/2 \rfloor \rangle_K + \langle \lceil (B+K)/2 \rceil - 1 \rangle_K \right). \tag{8}$$

If $K+B$ is odd, the assertion follows since $\lfloor (B+K)/2 \rfloor = (B+K-1)/2 = \lceil (B+K)/2 \rceil - 1$. It only remains to consider $K+B$ even. For this case, let $x = (B+K)/2$. Then

$$\begin{aligned} \langle \lfloor x \rfloor \rangle_K + \langle \lceil x \rceil - 1 \rangle_K &= x \langle x-1 \rangle_{K-1} + \langle x-1 \rangle_{K-1}(x-K) \\ &= (2x-K)\langle x-1 \rangle_{K-1}. \end{aligned}$$

Since $2x - K = B$, the assertion follows after plugging this equality back into (8). $\square$

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 of Theorem 3. The proof is based on the formula given in Lemma 3. Let $B = N/M$ be the number of blocks and recall that M is fixed and thus $B = \Theta(N)$. The key observation to derive the asymptotic value of the test statistic for residuals which alternate in blocks of size M is the following asymptotic expansion: For any fixed a, J and as $x \to \infty$,

$$\langle x+a\rangle_J = x^J + J\left(a - \frac{J-1}{2}\right)x^{J-1} + O(x^{J-2}). \tag{9}$$

This equality is based on expanding the product in the definition of the falling factorial:

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

which yields (9) using the well-known formula $\sum_{j=0}^{J-1} j = J(J-1)/2$. Hence, Lemma 3(a) and (9) with $x = B/2$, $a = (K-2)/2$, $J = K-1$ yield for even $K+B$ that

$$d_K(r_1,\ldots,r_N) = \frac{\langle\frac{B+K-2}{2}\rangle_{K-1}}{B^{K-1}} \cdot \frac{N^K}{\langle N\rangle_K} = \left(\left(\frac{1}{2}\right)^{K-1} + O(N^{-2})\right)\frac{N^K}{\langle N\rangle_K}. \tag{10}$$

Applying (9) for $x = N$, $a = 0$ and $J = K$ yields

$$\frac{N^K}{\langle N\rangle_K} = \frac{1}{1 - \frac{K(K-1)}{2N} + O(N^{-2})} = 1 + \frac{K(K-1)}{2N} + O(N^{-2}),$$

where the second equality holds since $1/(1-x) = \sum_{j=0}^{\infty} x^j = 1 + x + O(x^2)$ as $x \to 0$. Plugging this asymptotic expansion back into (10) yields for even $K+B$ that

$$\begin{aligned} d_K(r_1,\ldots,r_N) &= \left(\left(\frac{1}{2}\right)^{K-1} + O(N^{-2})\right)\left(1 + \frac{K(K-1)}{2N} + O(N^{-2})\right) \\ &= \left(\frac{1}{2}\right)^{K-1} + \left(\frac{1}{2}\right)^{K-1}\frac{K(K-1)}{2N} + O(N^{-2}). \end{aligned}$$

The case that $K+B$ is odd can be treated in a similar fashion and leads to the same asymptotic expansion. Hence the K-depth of $r_1, \ldots, r_N$ satisfies

$$N \cdot \left(d_K(r_1,\ldots,r_N) - \left(\frac{1}{2}\right)^{K-1}\right) = \frac{K(K-1)}{2^K} + O(N^{-1})$$

and the assertion follows by taking the limit $N \to \infty$. $\square$

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

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

Then the following holds:

(a) If $K + B$ is even then

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

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

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

Proof of Conjecture 1 for $K = 3$. Recall that according to Lemma 5 we may assume w.l.o.g. that B is odd. Let $\widetilde{B} = \lfloor B/2 \rfloor = (B-1)/2$. For convenience, we will subsequently ignore the scaling factor $\binom{N}{3}$ and instead maximize

$$\begin{aligned} D_{3,B}(q_1, q_2, \dots, q_B) &:= \binom{N}{3} d_{3,N,B}(q_1, q_2, \dots, q_B) \\ &= \sum_{i=1}^{\widetilde{B}} q_{2i-1} \sum_{j=i}^{\widetilde{B}} q_{2j} \sum_{l=j}^{\widetilde{B}} q_{2l+1} + \sum_{i=1}^{\widetilde{B}-1} q_{2i} \sum_{j=i}^{\widetilde{B}-1} q_{2j+1} \sum_{l=j}^{\widetilde{B}-1} q_{2l+2} \end{aligned} \tag{11}$$

where the maximum is considered over all $q = (q_1, \dots, q_B)$ from the set

$$Q'_{N,B} := \left\{ (q_1, \dots, q_B) \in [0, N]^B;\ \sum_{j=1}^{B} q_j = N \right\}.$$

To this end, note that $Q'_{N,B}$ is a compact set with respect to the standard subspace topology of the $\mathbb{R}^B$-subset

$$\mathcal{R}_{N,B} := \left\{ (q_1, \dots, q_B) \in \mathbb{R}^B;\ \sum_{j=1}^{B} q_j = N \right\}.$$

Also note that all possible local maxima of $D_{3,B}$ within $\mathcal{R}_{N,B}$ can be computed by using Lagrange multiplier, i.e. by determining the stationary points of

$$F(q_1, \dots, q_B, \lambda) = D_{3,B}(q_1, \dots, q_B) + \lambda \left(N - \sum_{b=1}^{B} q_b \right).$$

Hence, a maximum q of $D_{3,B}$ within $Q'_{N,B}$ has to either be part of a stationary point (q, λ) of F or has to lie within the boundary

$$\partial Q'_{N,B} = \left\{ (q_1, \dots, q_B) \in Q'_{N,B};\ q_i = 0 \text{ for some } i = 1, \dots, B \right\}.$$

Therefore, the assertion follows if we prove the following two claims:

(a) The only stationary point of $D_{3,B}$ in $Q'_{N,B} \setminus \partial Q'_{N,B}$ is

$$q^* = q^*(B) = \left(\frac{N}{B}, \frac{N}{B}, \ldots, \frac{N}{B}\right) \in Q'_{N,B}.$$

(b) All $q \in \partial Q'_{N,B}$ satisfy $D_{3,B}(q) < D_{3,B}(q^*)$.

For part (a) first note that the partial derivatives of F are given by

$$\begin{aligned}\frac{\partial}{\partial q_{2m-1}} F(q,\lambda) = &\sum_{j=m}^{\widetilde{B}} q_{2j} \sum_{l=j}^{\widetilde{B}} q_{2l+1} + \sum_{i=1}^{m-1} q_{2i-1} \sum_{j=i}^{m-1} q_{2j} \\ &+ \sum_{i=1}^{m-1} q_{2i} \sum_{l=m-1}^{\widetilde{B}-1} q_{2l+2} - \lambda, \quad m = 1, \ldots, \widetilde{B}+1, \\ \frac{\partial}{\partial q_{2m}} F(q,\lambda) = &\sum_{i=1}^{m} q_{2i-1} \sum_{l=m}^{\widetilde{B}} q_{2l+1} + \sum_{j=m}^{\widetilde{B}-1} q_{2j+1} \sum_{l=j}^{\widetilde{B}-1} q_{2l+2} \\ &+ \sum_{i=1}^{m-1} q_{2i} \sum_{j=i}^{m-1} q_{2j+1} - \lambda, \quad m = 1, \ldots, \widetilde{B},\end{aligned}$$

with the convention $\sum_{i=a}^{b} x_i = 0$ for all $a > b$ and any sequence $(x_i)_{i \geq 1}$. In particular, a simple calculation reveals that these partial derivatives satisfy for all $q = (q_1, \ldots, q_B) \in (\mathbb{R} \setminus \{0\})^B$, $\lambda \in \mathbb{R}$ and $j = 3, \ldots, B-1$

$$\begin{aligned}&\frac{1}{q_j}\left(\frac{\partial}{\partial q_{j+1}} F(q,\lambda) - \frac{\partial}{\partial q_{j-1}} F(q,\lambda)\right) + \frac{1}{q_{j-1}}\left(\frac{\partial}{\partial q_j} F(q,\lambda) - \frac{\partial}{\partial q_{j-2}} F(q,\lambda)\right) \\ =& q_{j-1} - q_j.\end{aligned}$$

Hence, if (q,λ) is a stationary point then the equations above have to equal zero for all $j = 3, \ldots, B-1$ and thus

$$q_2 = q_3 = \ldots = q_{B-1}. \tag{12}$$

With the additional assumption of (12), one can further show that, e.g.,

$$0 = \frac{1}{q_2}\left(\frac{\partial}{\partial q_3} F(q,\lambda) - \frac{\partial}{\partial q_1} F(q,\lambda)\right) = q_1 - q_B$$

and hence $q_1 = q_B$ for any stationary point. Finally, (12) and $q_1 = q_B$ imply

$$0 = \frac{\partial}{\partial q_2} F(q,\lambda) - \frac{\partial}{\partial q_1} F(q,\lambda) = q_1(q_1 - q_2)$$

and therefore $q_1 = q_2 = \ldots = q_B$ if q is a stationary point without any zeros. The additional condition $\sum_j q_j = N$ therefore yields $q = q^*$ as claimed in (a).

Part (b) can be shown by induction on B after noting that setting $q_i = 0$ is equivalent to removing block i and merging blocks $i-1$ and $i+1$, that is if $q = (q_1, \dots, q_B)$ satisfies $q_i = 0$ then

$$D_{3,B}(q) = \begin{cases} D_{3,B-1}(q_2, \dots, q_B), & \text{if } i = 1, \\ D_{3,B-2}(q_1, \dots, q_{i-1} + q_{i+1}, \dots, q_B), & \text{if } i \in \{2, \dots, B-1\}, \\ D_{3,B-1}(q_1, \dots, q_{B-1}), & \text{if } i = B. \end{cases}$$

In order to only consider odd B for the induction, the cases $i \in \{1, B\}$ can further be reduced by applying Lemma 5 to obtain

$$D_{3,B-1}(q_2, \dots, q_B) = D_{3,B-2}(q_2 + q_B, q_3, \dots, q_{B-1}),$$
$$D_{3,B-1}(q_1, \dots, q_{B-1}) = D_{3,B-2}(q_1 + q_{B-1}, q_2, \dots, q_{B-2}).$$

Hence, assuming $q^*(B)$ is the maximum of $D_{3,B}$ by induction hypothesis, then the reduction above yields for all $q \in \partial Q'_{N,B+2}$

$$D_{3,B+2}(q) \leq D_{3,B}(q^*(B)) < D_{3,B+2}(q^*(B+2))$$

in which the last inequality follows by observing that, according to Lemma 3,

$$D_{3,B}(q^*(B)) = \binom{N}{3} \frac{N^3}{\langle N \rangle_3} \frac{\left\langle \frac{B+1}{2} \right\rangle_2}{B^2} = \frac{N^3}{24} \left(1 - \frac{1}{B^2}\right)$$

which is strictly increasing in B. □

Possible extensions of the proof to $K \geq 4$. It is possible to show for all $K \geq 4$ that $q^* = (N/B, \dots, N/B)$ indeed is a stationary point of $d_{K,N,B}$ if $K + B$ is even. This essentially only requires to show that the number of K-tuples in $\mathcal{A}_{K,B}$ in which index i appears (i.e. the number of summands in the K-depth with q_i in it) is the same for all $i \in \{1, \dots, B\}$. However, we did not manage to prove that this stationary point is unique for $K \geq 4$. Note that the argument in part (b) of the proof above also works for arbitrary K. Hence, the only missing gap for a proof of Conjecture 1 is the uniqueness of q^*.

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

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

Proof of Lemma 5. For $x \in \mathbb{R}$ and $w = (w_1, \dots, w_J) \in \mathbb{R}^J$ let $(x, w) = (x, w_1, \dots, w_J)$ and let $(w, x) = (w_1, \dots, w_J, x)$. Let $\mathcal{A}_{K,B}$ and $d_{K,N,B}(q_1, \dots, q_K)$ be as in Lemma 1. The key observations to prove Lemma 5 are the following: If $K + B$ is odd then, for every $i \in \{2, \dots, B-1\}^{K-1}$,

(a) $(1, i) \in \mathcal{A}_{K,B}$ if and only if $(i, B) \in \mathcal{A}_{K,B}$,

(b) there is no vector $j \in \{2, \dots, B-1\}^{K-2}$ with $(1, j, B) \in \mathcal{A}_{K,B}$.

Both (a) and (b) are not hard to check, details are given at the end of the proof. Based on these properties, we can split the sum in $d_{K,N,B}(q_1,\dots,q_K)$ in the following way: Let

$$\begin{aligned}\mathcal{B}_{K,B} &= \left\{i\in\{2,\dots,B-1\}^{K-1};\ (1,i)\in\mathcal{A}_{K,B}\right\},\\ \mathcal{C}_{K,B} &= \mathcal{A}_{K,B}\cap\{2,\dots,B-1\}^K.\end{aligned}$$

We may now split $\mathcal{A}_{K,B}$ into three parts: The first one contains vectors $v=(v_1,\dots,v_K)$ in $\mathcal{A}_{K,B}$ with $v_1=1$, the second one contains vectors with $v_K=B$ and the third part contains vectors with $v_1\neq 1$ and $v_K\neq B$ (vectors with $v_1=$ and $v_K=B$ are impossible according to (b)). Then (a) implies $\mathcal{A}_{K,B}=(\{1\}\times\mathcal{B}_{K,B})\cup(\mathcal{B}_{K,B}\times\{B\})\cup\mathcal{C}_{K,B}$. Hence

$$\sum_{(i_1,\dots,i_K)\in\mathcal{A}_{K,B}}\prod_{k=1}^{K}q_{i_k}=(q_1+q_B)\sum_{(i_1,\dots,i_{K-1})\in\mathcal{B}_{K,B}}\prod_{k=1}^{K-1}q_{i_k}+\sum_{(i_1,\dots,i_K)\in\mathcal{C}_{K,B}}\prod_{k=1}^{K}q_{i_k}.$$

Furthermore, note that $\mathcal{A}_{K,B-1}=(\{1\}\times\mathcal{B}_{K,B})\cup\mathcal{C}_{K,B}$ once again by splitting the set into two parts based to whether $v_1=1$ or not. In particular, if $\widetilde{q}_1=q_1+q_B$ and $\widetilde{q}_j=q_j$ for $j=2,\dots,B-1$, then

$$\sum_{(i_1,\dots,i_K)\in\mathcal{A}_{K,B-1}}\prod_{k=1}^{K}\widetilde{q}_{i_k}=\widetilde{q}_1\sum_{(i_1,\dots,i_{K-1})\in\mathcal{B}_{K,B}}\prod_{k=1}^{K-1}\widetilde{q}_{i_k}+\sum_{(i_1,\dots,i_K)\in\mathcal{C}_{K,B}}\prod_{k=1}^{K}\widetilde{q}_{i_k}.$$

Hence $d_{K,N,B}(q_1,\dots,q_B)=d_{K,N,B-1}(\widetilde{q}_1,\dots,\widetilde{q}_{B-1})$, which is the assertion.

Proof of (a) and (b). For simplicity, we will subsequently assume that K is odd and B is even. The other case can be treated similarly. For (a) note that $(1,i)\in\mathcal{A}_{K,B}$ requires $i=(i_1,\dots,i_{K-1})$ to start with an even index i_1 and continue alternating between odd and even in the subsequent indices. Since the length $K-1$ of i is even, the last index i_{K-1} of the vector has to be odd. Since B is even, this means that i_{K-1} and B indeed alternate between odd and even. Hence $(i,B)\in\mathcal{A}_{K,B}$. Similarly, $(i,B)\in\mathcal{A}_{K,B}$ requires i_{K-1} to be odd and subsequent indices in the vector to alternate between odd/even. Hence i_1 has to be even and thus $(1,i)\in\mathcal{A}_{K,B}$. For part (b) assume for the sake of contradiction that $(1,j,B)\in\mathcal{A}_{K,B}$ for a vector $j=(j_1,\dots,j_{K-2})\in\{2,\dots,B-1\}^{K-2}$. Since 1 is odd, this in particular means that j_1 is even. Since $K-2$ is odd, j_1 and j_{K-2} have the same parity in a vector j with entries that alternate between even/odd. Hence j_{K-2} is even. However, since B is even, j_{K-2} has to be odd in order to have $(1,j,B)\in\mathcal{A}_{K,B}$, which leads to a contradiction. □

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

$$\lim_{N\to\infty}\sup\left\{d_{K,N,B}(q_1,\dots,q_B);\ (q_1,\dots,q_B)\in\mathcal{Q}_{N,B}\right\}=\frac{K!}{K^K}\leq\left(\frac{1}{2}\right)^{K-1},\tag{13}$$

where the inequality in (13) is strict for $K\geq 3$.

Before proving Theorem 4, we start with a Lemma that yields the inequalities in Theorem 4 and Theorem 5.

Lemma 6. *Let K, B be integers with $B \geq K \geq 2$. Then*

$$\frac{\prod_{k=1}^{K-1}\left(\frac{B+K}{2}-k\right)}{B^{K-1}} \leq \left(\frac{1}{2}\right)^{K-1}$$

with equality if and only if $K = 2$.

Proof of Lemma 6. First note that by rearranging the order of the product one obtains

$$\prod_{k=1}^{K-1}\left(\frac{B+K}{2}-k\right) = \varepsilon_{K,B}\prod_{k=1}^{\lfloor (K-1)/2\rfloor}\left(\frac{B+K}{2}-k\right)\left(\frac{B+K}{2}-(K-k)\right), \tag{14}$$

$$\text{with} \quad \varepsilon_{K,B} = \begin{cases} 1, & \text{if } K \text{ is odd,} \\ B/2, & \text{if } K \text{ is even.} \end{cases}$$

Next note that the quadratic function $g(x) = ((B+K)/2 - x)((B-K)/2 + x)$ has a unique global maximum at $x = K/2$ and that $g(K/2) = B^2/4$. Hence

$$\prod_{k=1}^{\lfloor (K-1)/2\rfloor}\left(\frac{B+K}{2}-k\right)\left(\frac{B-K}{2}+k\right) \leq \left(\frac{B^2}{4}\right)^{\lfloor (K-1)/2\rfloor}$$

in which the inequality is strict if there is at least one factor with $k \neq K/2$, i.e. if $K \geq 3$. In combination with (14), this upper bound yields

$$\frac{\prod_{k=1}^{K-1}\left(\frac{B+K}{2}-k\right)}{B^{K-1}} \leq \frac{\varepsilon_{K,B}}{B^{K-1}}\left(\frac{B^2}{4}\right)^{\lfloor (K-1)/2\rfloor} = \left(\frac{1}{2}\right)^{K-1}$$

where the last equality can easily be checked by a case distinction between K even/odd. The assertion follows since this inequality is strict for $K \geq 3$. □

Proof of Theorem 4. We first consider the case $K = B$, i.e. the aim is to compute the maximum of the function

$$(q_1, \ldots, q_K) \mapsto d_{K,K}(q_1, \ldots, q_K) = \frac{1}{\binom{N}{K}}\prod_{k=1}^{K} q_k$$

under the side condition $(q_1, \ldots, q_K) \in \mathcal{Q}_{N,K}$, that is $q_1, \ldots, q_K \in \mathbb{N}$ and $\sum_{k=1}^{K} q_k = N$. When disregarding the condition $q_1, \ldots, q_K \in \mathbb{N}$, this can easily be done, e.g., by using Lagrange multipliers (considering the function $\ln(d_{K,K}(\cdot))$ instead of $d_{K,K}(\cdot)$ simplifies the calculations), which reveals a global maximum at

$$q_1 = \ldots = q_K = \frac{N}{K}.$$

Hence,

$$\sup\left\{d_{K,K}(q_1, \ldots, q_K);\ (q_1, \ldots, q_K) \in \mathcal{Q}_{N,K}\right\}$$

$$\leq d_{K,K}\left(\frac{N}{K},\ldots,\frac{N}{K}\right) = \frac{1}{\binom{N}{K}}\left(\frac{N}{K}\right)^{K}$$

with equality if $N/K \in \mathbb{N}$. Thus the limit values of the maximal depth of residual vectors with K blocks is given by

$$\lim_{N\to\infty} \frac{1}{\binom{N}{K}}\left(\frac{N}{K}\right)^{K} = \frac{K!}{K^K}.$$

The case $B = K+1$ can be treated in a similar fashion or can be deduced from $B = K$ and Lemma 5. In particular, the maximal value is attained at $q_1 + q_{K+1} = q_2 = \ldots = q_K$ and its limit value remains $K!/K^K$. The remaining inequality

$$\frac{K!}{K^K} < \left(\frac{1}{2}\right)^{K-1} \quad \text{for all } K \geq 3$$

follows from Lemma 6 with $B = K$. Hence the assertion follows. □

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 of Theorem 5. The identity for the limit of the test statistic follows from Lemma 3 since $N^K/\langle N\rangle_K \to 1$ for fixed K as $N \to \infty$. The inequality in (15) follows from Lemma 6. □

3 Further simulation results

At first, we present here further simulation for models with three unknown parameter as quadratic regression, AR(2)-regression and nonlinear AR(1)-regression with intercept. In these examples, all simulations were done with 100 repetitions since there were no visible difference when 500 repetitions were used. For getting the simulated p-values for the 3-depth test and the 4-depth test for $N = 96$, 3-depth and 4-depth were simulated 10 000 times. For $N = 12$, the exact distribution was used.

At last we consider the multiple regression case as considered in the paper, but here additionally with sample size $N = 500$ and for another aspect of alternatives.

3.1 Quadratic regression

In the quadratic regression model given by

$$Y_n = \theta_0 + \theta_1 x_n + \theta_2 x_n^2 + E_n, \quad n = 1,\ldots,N, \quad \theta = (\theta_0,\theta_1,\theta_2)^\top,$$

we consider the problem of testing the null hypothesis $H_0 : \theta = (1, 0, 1)^\top$ with a test with level $\alpha = 0.05$.

Figures 1 and 2 show the simulated power of the sign test, the F test, the 3-depth test, and the 4-depth test for $N = 12$ and $x_1 = -5.5$, $x_2 = -4.5$, ... $x_6 = -0.5$, $x_7 = 0.5$, ... $x_{12} = 5.5$ where E_n has a standard normal distribution. For each simulation, a 41×41 grid of alternatives and 100 repetitions were used. Moreover, the exact distributions for the 3-depth and the 4-depth were used to obtain the p-values. The parameter of the null hypothesis is given by the intersection of the two dotted lines.

In Figure 1, where the the component θ_2 was fixed to 1, the power of the 3-depth test is slightly worse than the power of the F test and better than the power of the 4-depth test. However, the power of the 3-depth tests is much worse in Figure 2, where the component θ_1 was fixed to 0 in the upper part and the component θ_0 was fixed to 1 in the lower part. In both cases, the 3-depth test is even worse than the sign test while only the 4-depth test is slightly worse than the F test. Both the 3-depth test and the sign test have an unbounded area of power less or equal $\alpha = 0.05$. For the 3-depth test, this is due to the fact that in these cases often two sign changes appear and only vectors with zero 3-depth can be rejected due to the small sample size. In particular, the maximum 3-depth for two sign changes is $\frac{4^3 \cdot 6}{12 \cdot 11 \cdot 10} = 0.291$ providing a p-value of 0.758 so that a rejection of the null hypothesis is not possible. Moreover, alternatives which tend to lead to two sign changes also tend to have a large difference between positive and negative signs, which is why the sign test perform better at these alternatives. Since the 4-depth is zero for two sign changes, the power of the 4-depth test is similar to the F test.

However the power of the 3-depth test becomes much better for $N = 96$. For this sample size, $x_1 = -5.9375$, $x_2 = -5.875$, ... $x_{48} = -0.0625$, $x_{49} = 0.0625$, ... $x_{96} = 5.9375$ were used as design points and the distribution of the 3-depth and the 4-depth was simulated 10 000 times to obtain relative exact p-values. Again, a 41×41 grid of alternatives was used for each simulation and each scenario was repeated 100 times. Figures 3, 4, and 5 show the results for the cases where at first component θ_2 was fixed to 1 (Figure 3), then component θ_1 was fixed to 0 (Figure 4), and at last component θ_0 was fixed to 1 (Figure 5). Each of the three figures provides the results for errors with standard normal distribution in the upper part and the results for errors with standard Cauchy distribution in the lower part.

For the normal distribution, the area of small power of the 3-depth test is now bounded. Only in the case where θ_1 is fixed to zero, this area is much larger than the area of small power of the F test. But in the two other cases, the 3-depth test behaves similarly to the F test. The 4-depth test behaves similarly to the F test in all three cases. The improved performance of the 3-depth test can be explained by the p-values discussed in the main article. In particular, the maximum depth for two sign changes is $\frac{32^3 \cdot 6}{96 \cdot 95 \cdot 94} = 0.229$ providing a p-value of 0.014 which is smaller than the significance level $\alpha = 0.05$.

If the errors follow a Cauchy distribution, then the power of the F test becomes very bad while the power functions of the sign test, the 3-depth test and the 4-depth test are only slightly changed. Although Figure 4 may indicate that the area of small power of the 3-depth test and the F test is unbounded, this is not correct. This is only caused by the small range of θ chosen for the plot. Increasing the range to the same area as in Figures 1 and 2 reveals that the area of small power is indeed bounded for both tests. However, the area of small power is still much larger for the F test than for the 3-depth test. Hence the 3-depth test and the 4-depth test are much more robust against outliers than the F test.

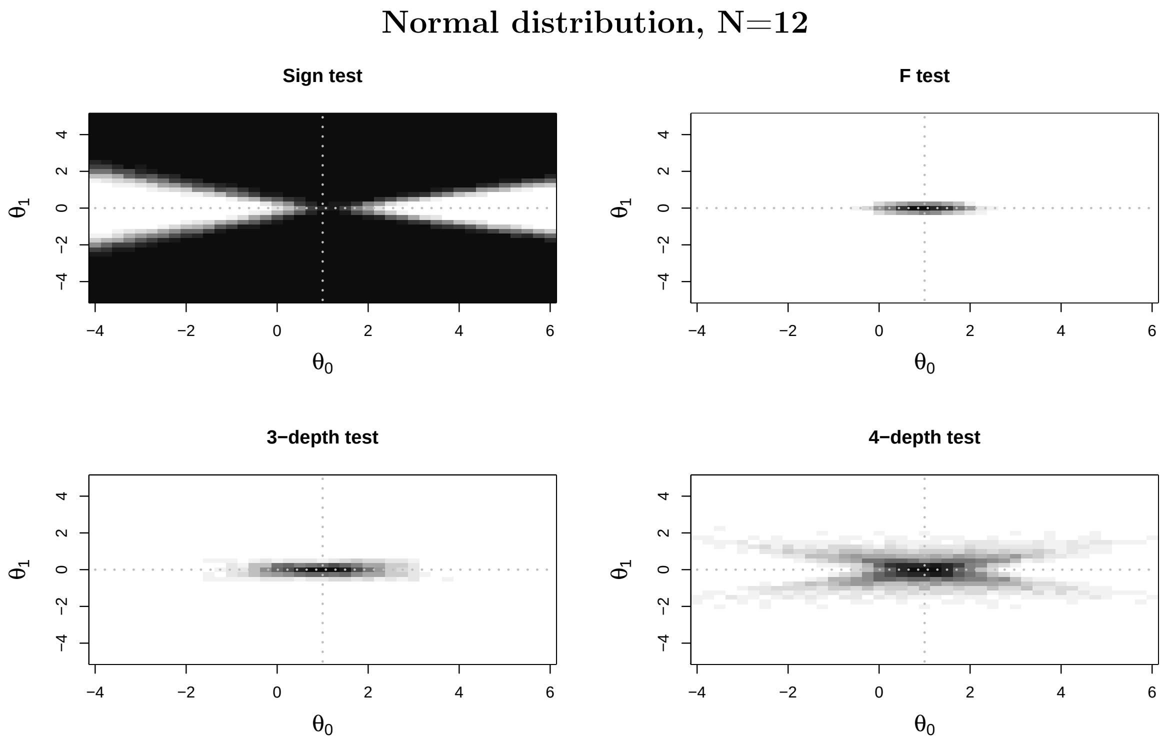

Figure 1: Simulated power of the sign test, the F test, the 3-depth test, and the 4-depth test for normally distributed errors for sample size $N = 12$, where component θ_2 is fixed to 1 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

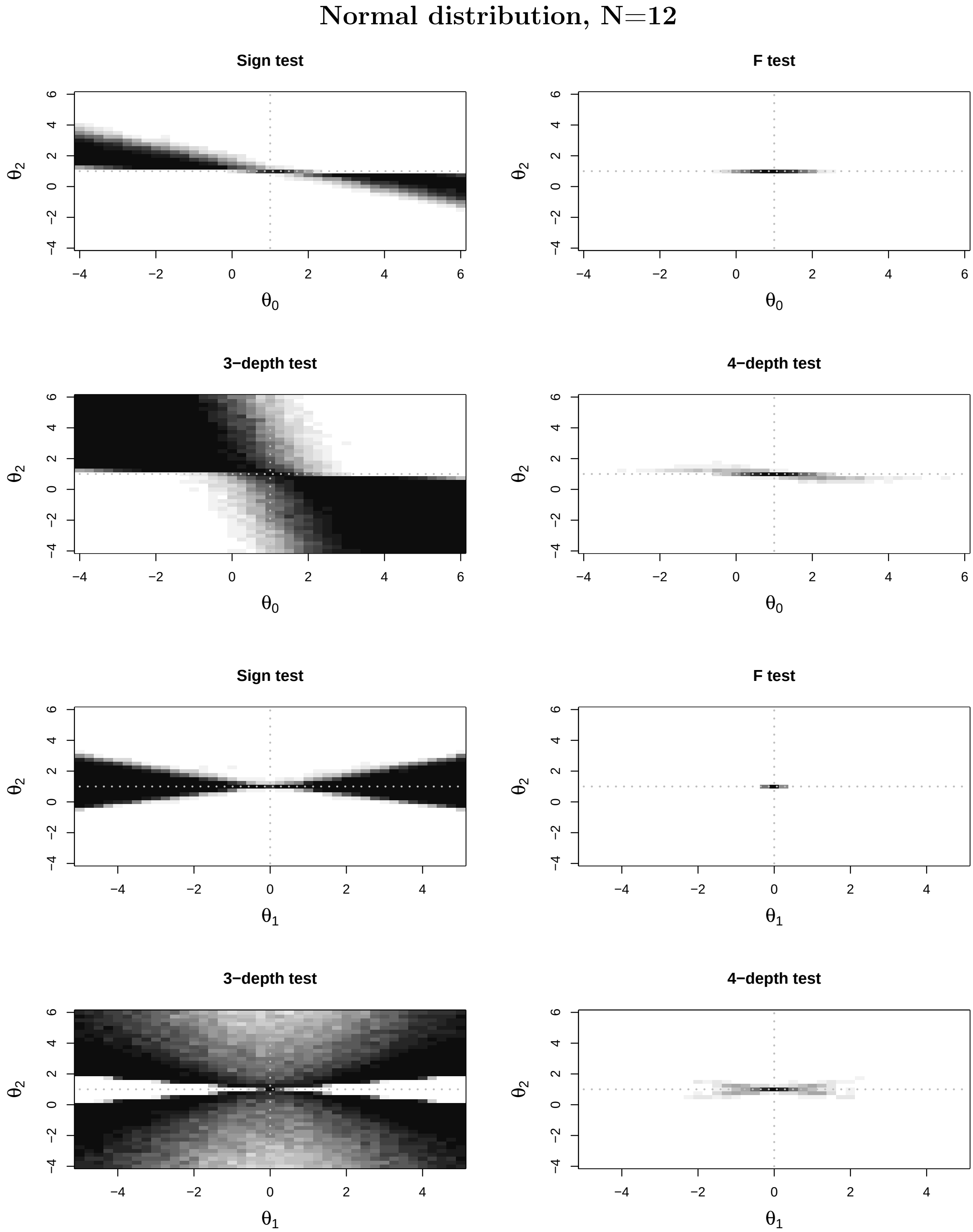

Figure 2: Simulated power of the sign test, the F test, the 3-depth test, and the 4-depth test for normally distributed errors for sample size $N = 12$, where the component θ_1 is fixed to 0 in the upper part and the the component θ_0 is fixed to 1 in the lower part (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

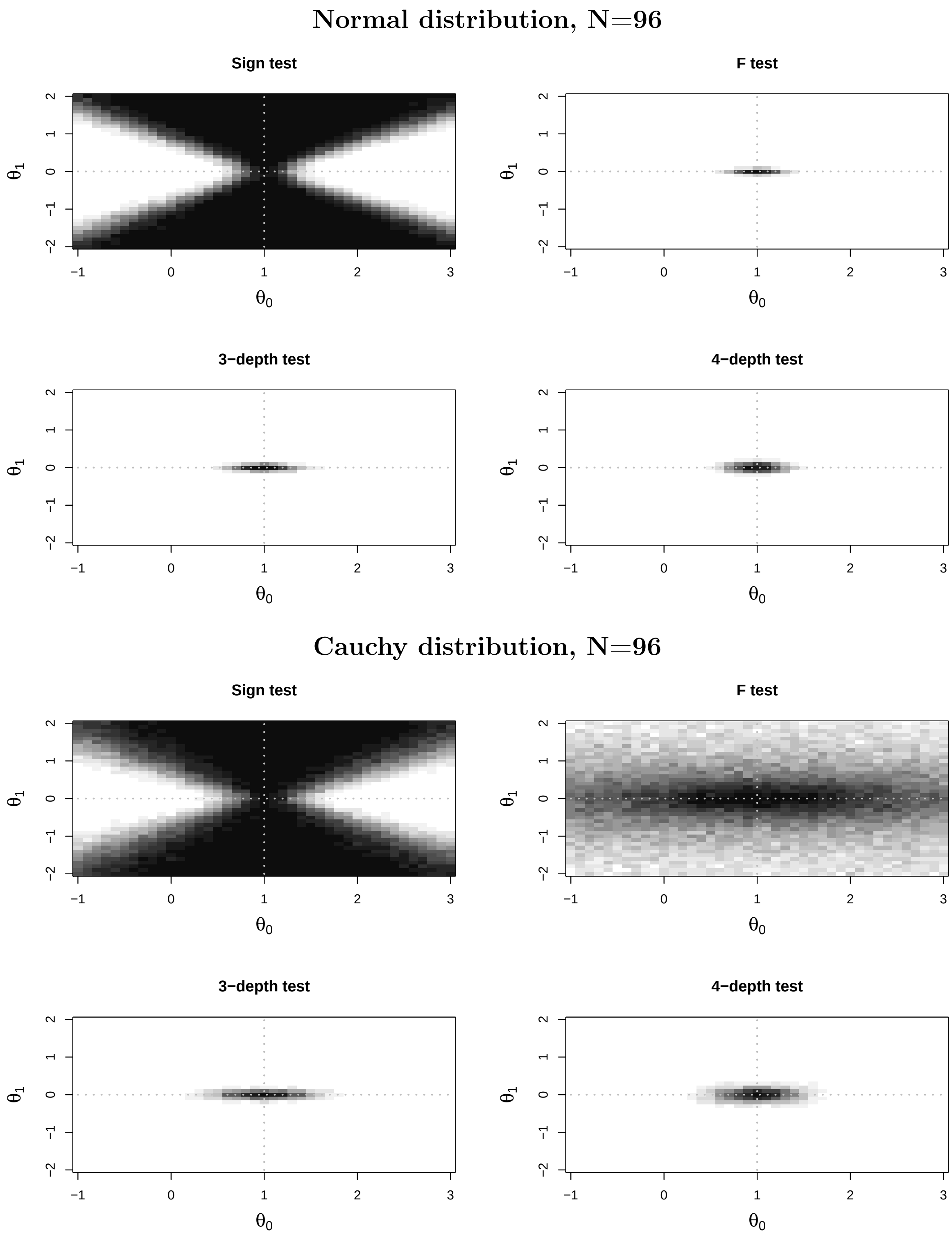

Figure 3: Simulated power of the sign test, the F test, the 3-depth test, and the 4-depth test for errors with normal distribution (upper part) and with Cauchy distribution (lower part) for sample size $N = 96$, where the component θ_2 is fixed to 1 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

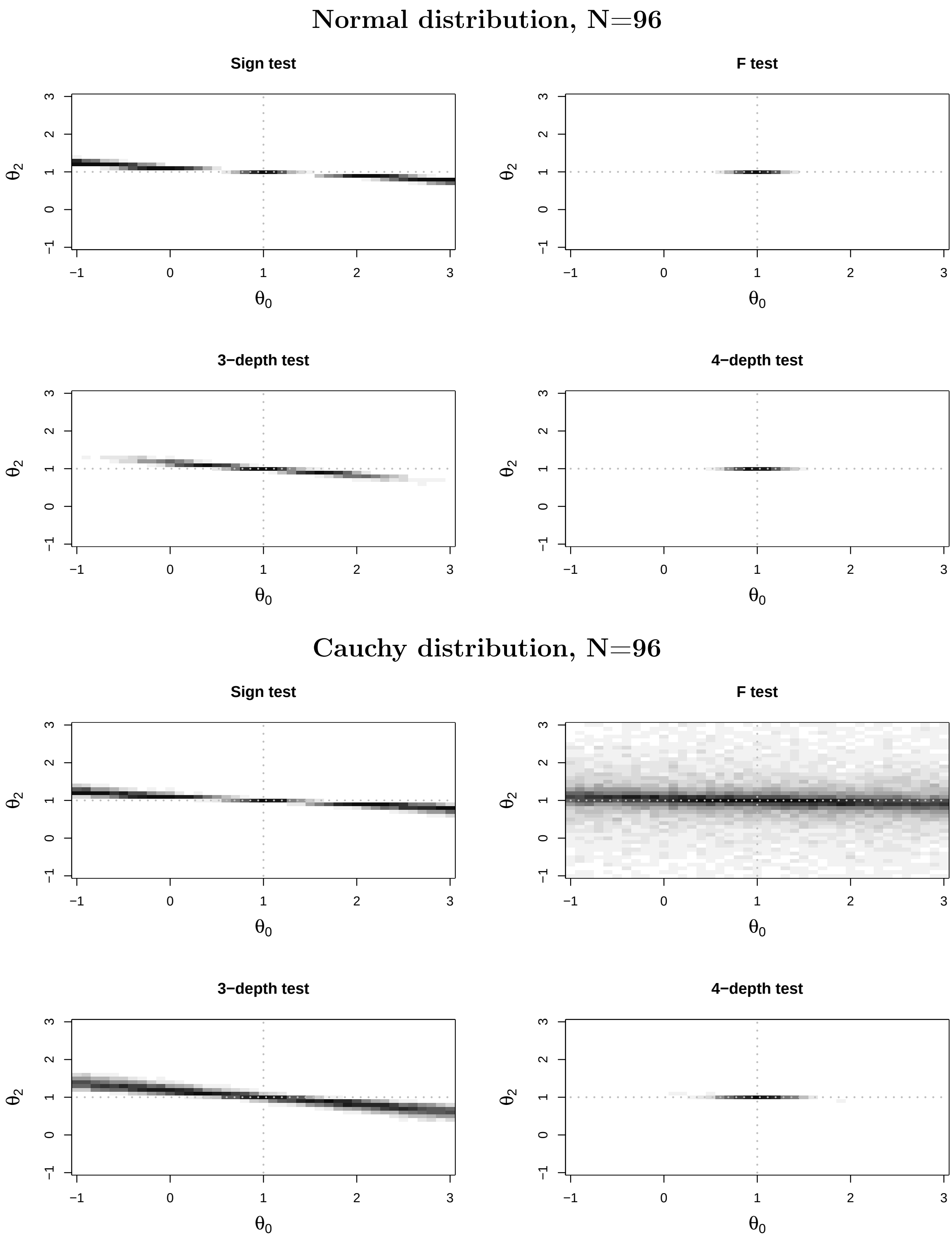

Figure 4: Simulated power of the sign test, the F test, the 3-depth test, and the 4-depth test for errors with normal distribution (upper part) and with Cauchy distribution (lower part) for sample size $N = 96$, where the component θ_1 is fixed to 0 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1])$.

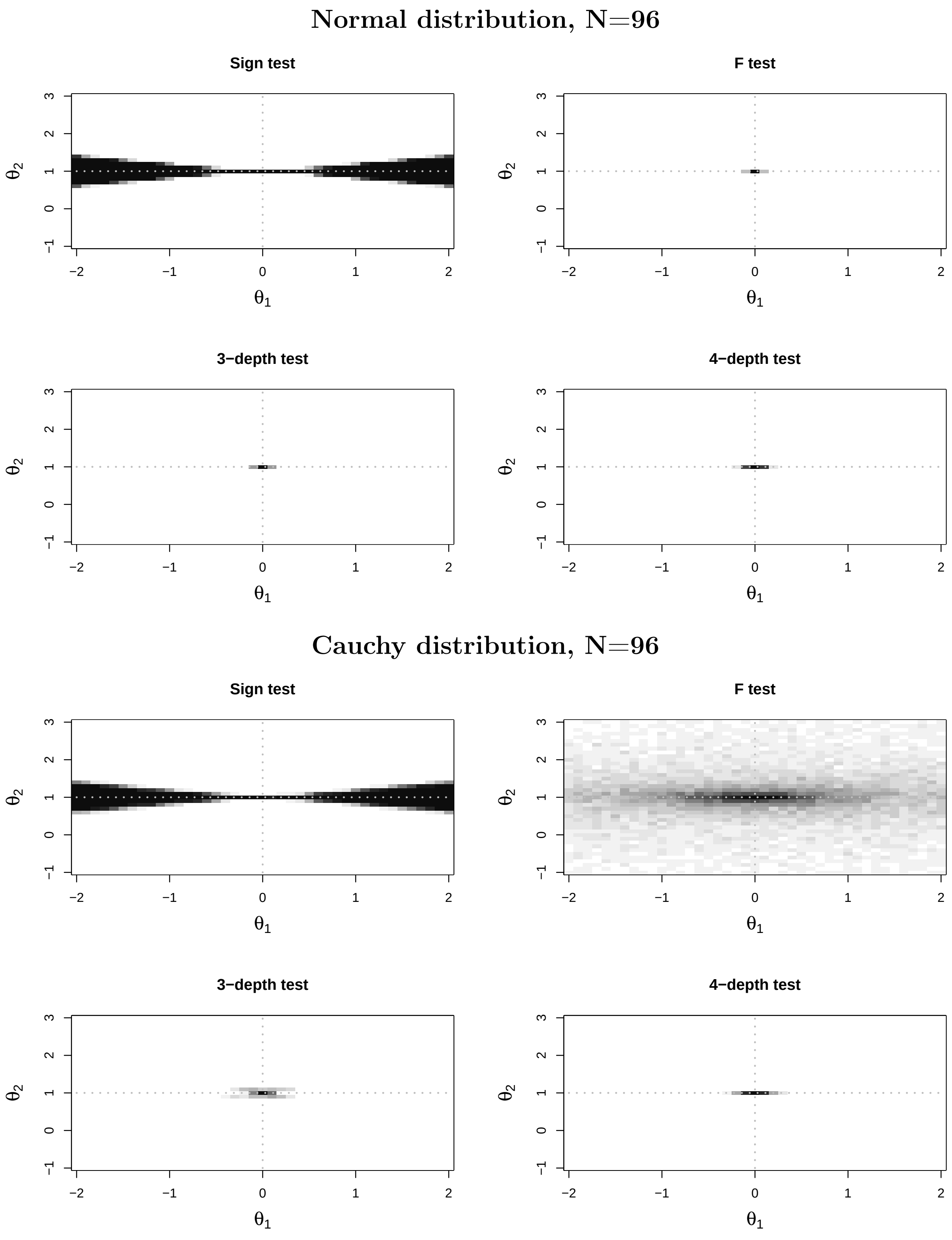

Figure 5: Simulated power of the sign test, the F test, the 3-depth test, and the 4-depth test for errors with normal distribution (upper part) and with Cauchy distribution (lower part) for sample size $N = 96$, where the component θ_0 is fixed to 1 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

3.2 AR(2)-model

Here we consider the autoregressive model given by

$$Y_n = \theta_0 + \theta_1 Y_{n-1} + \theta_2 Y_{n-2} + E_n,\ n = 1, \ldots, N,\ \ \theta = (\theta_0, \theta_1, \theta_2)^\top,$$

with $Y_{-1} = Y_0 = 5$. In contrast to most approaches for autoregressive models, we assume here only $P(E_n > 0) = P(E_n < 0) = \frac{1}{2}$. The residuals are $R_n(\theta) = Y_n - \theta_0 - \theta_1 Y_{n-1} - \theta_2 Y_{n-2}$. In the simulation, we used a normal distribution with mean 0 and standard deviation 0.01. A comparison of the sign test, the 3-depth test, the 4-depth test, and a t-test for testing $H_0 : \theta = (0.2, 0.8, 0.21)^\top$ for $N = 96$ with $\alpha = 0.05$ are given by the Figures 6, 7, and 8. Thereby a 41×41 grid of alternatives and 100 simulations for each alternative were used.

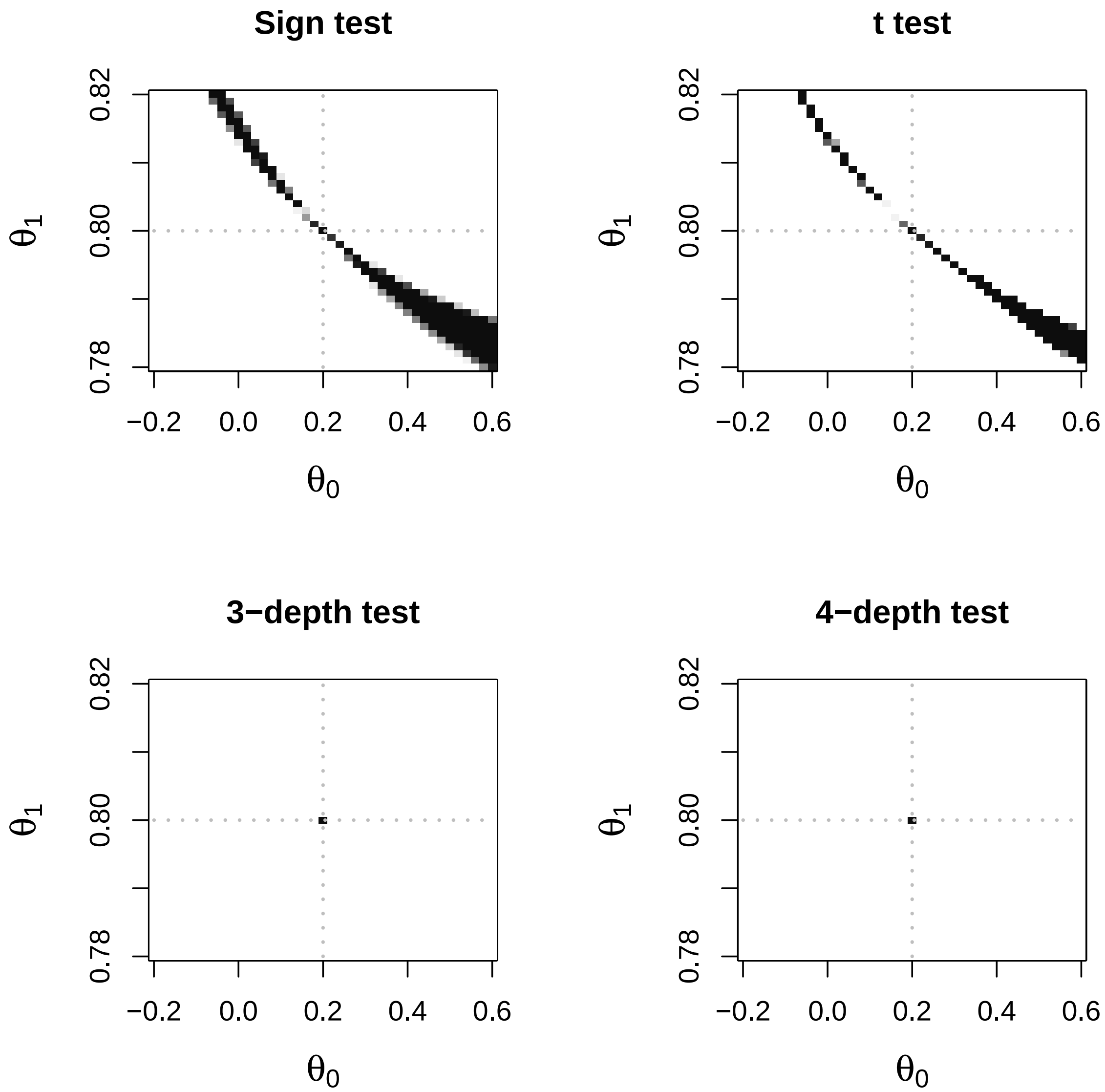

Figure 6: Simulated power of the sign test, 3-depth test, 4-depth test and t test for the AR(2)-model where θ_2 is fixed to 0.21 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

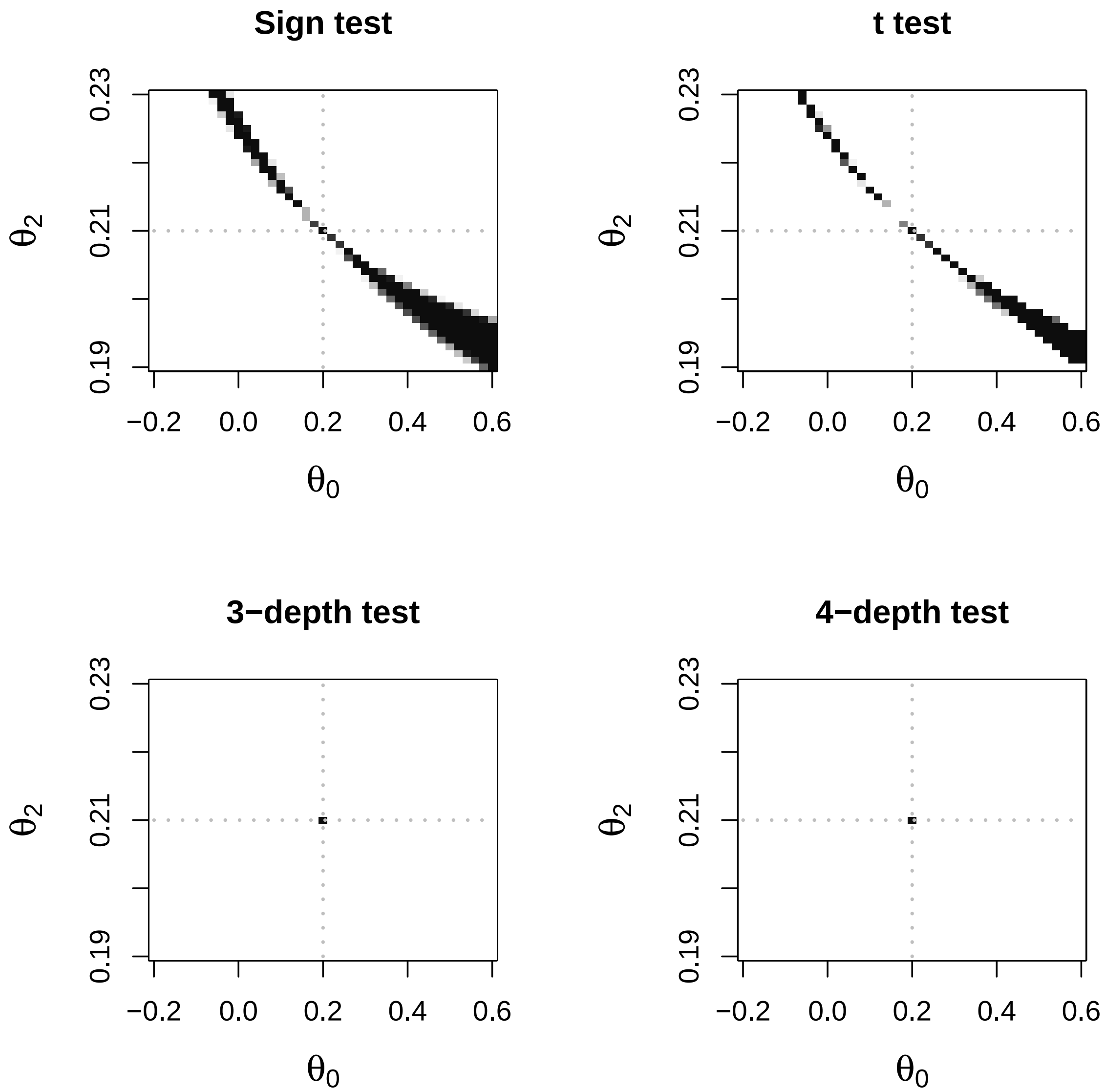

Figure 7: Simulated power of the sign test, 3-depth test, 4-depth test and t test for the AR(2)-model where θ_1 is fixed to 0.8 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

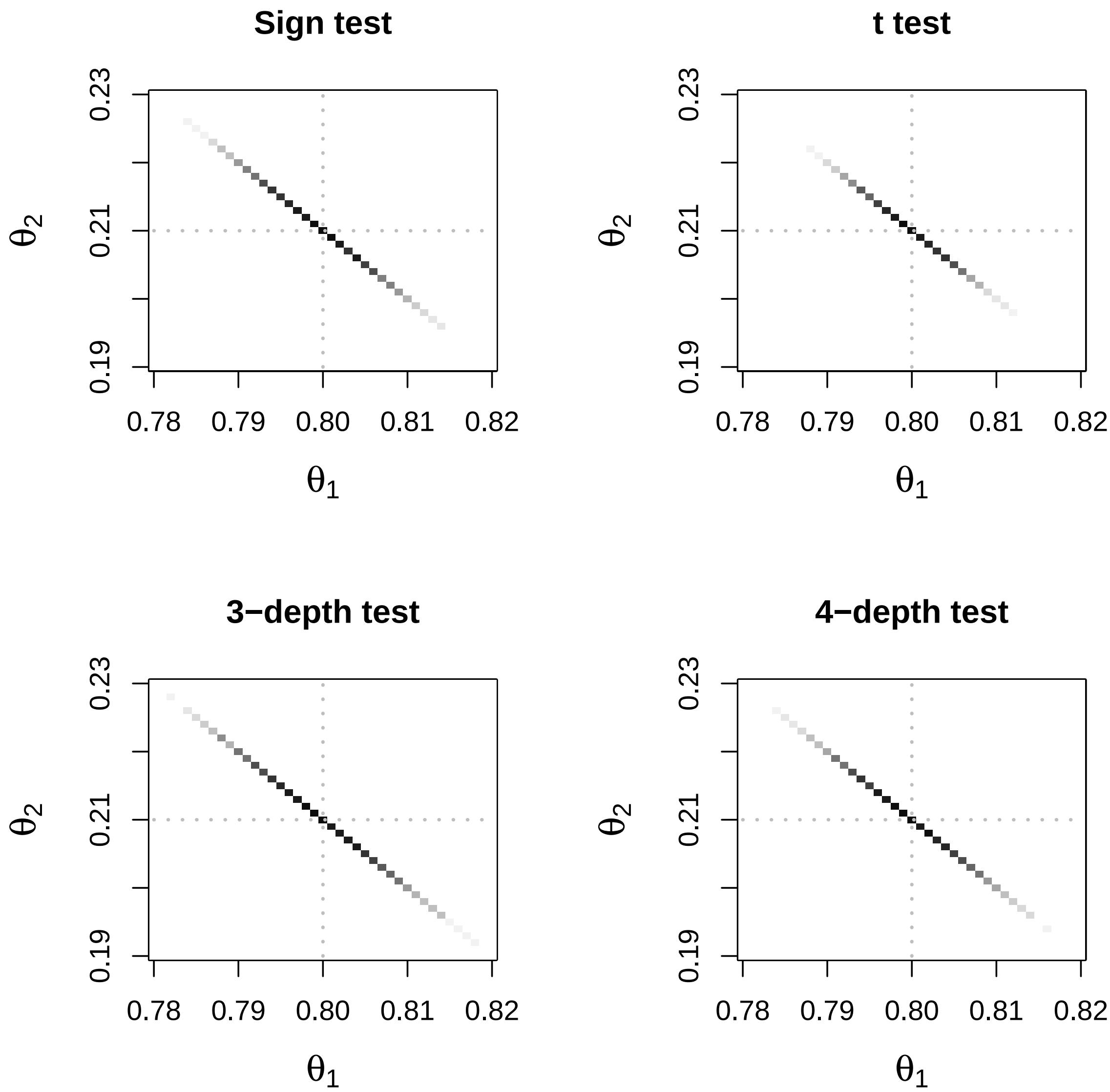

Figure 8: Simulated power of the sign test, 3-depth test, 4-depth test and t test for the AR(2)-model where θ_0 is fixed to 0.2 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

3.3 Nonlinear AR(1)-model

Here we consider the nonlinear autoregressive model given by

$$Y_n = \theta_0 + Y_{n-1} + \theta_1\, Y_{n-1}^{\theta_2} + E_n,\ n = 1, \dots, N,\ \ \theta = (\theta_0, \theta_1, \theta_2)^\top,$$

with $Y_0 = 15$. In contrast to most approaches for autoregressive models, we assume here only $P(E_n > 0) = P(E_n < 0) = \frac{1}{2}$. The residuals are $R_n(\theta) = Y_n - \theta_0 - Y_{n-1} - \theta_1\, Y_{n-1}^{\theta_2}$. In the simulation, we used a normal distribution with mean 0 and standard deviation 0.01. A comparison of the sign test, the 3-depth test, the 4-depth test, and a t-test for testing $H_0 : \theta = (0.01, 0.005, 1.002)^\top$ for $N = 96$ with $\alpha = 0.05$ are given by the Figures 9, 10, and 11. Thereby a 81×71 grid was used for the presentation of alternatives in θ_0 and θ_1, a 81×101 grid for the presentation of alternatives in θ_0 and θ_2, and a 65×74 grid of for the presentation of alternatives in θ_1 and θ_2, and 100 simulations for each alternative were used.

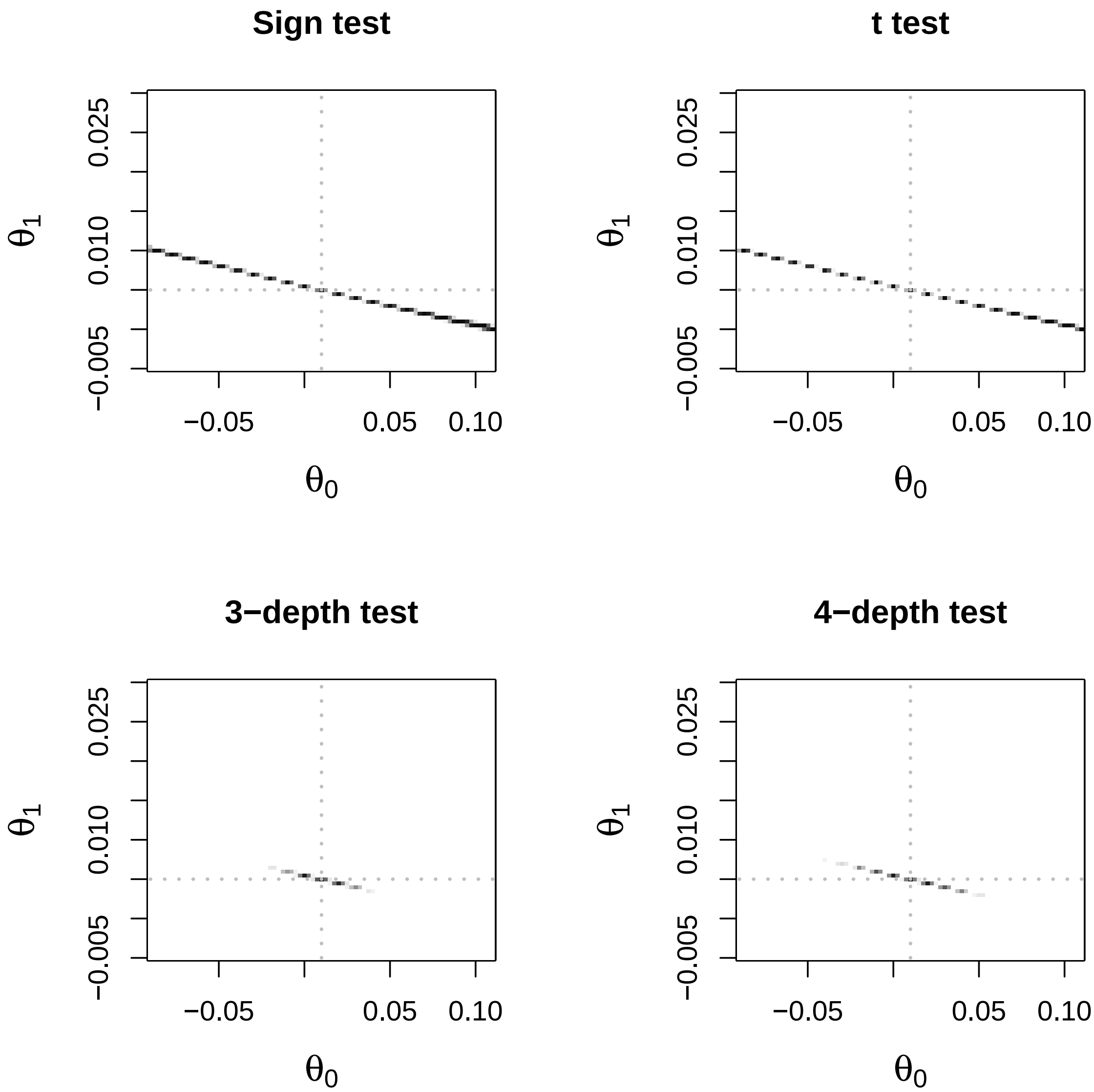

Figure 9: Simulated power of the sign test, 3-depth test, 4-depth test and t test for the nonlinear AR(1)-model where θ_2 is fixed to 1.002 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

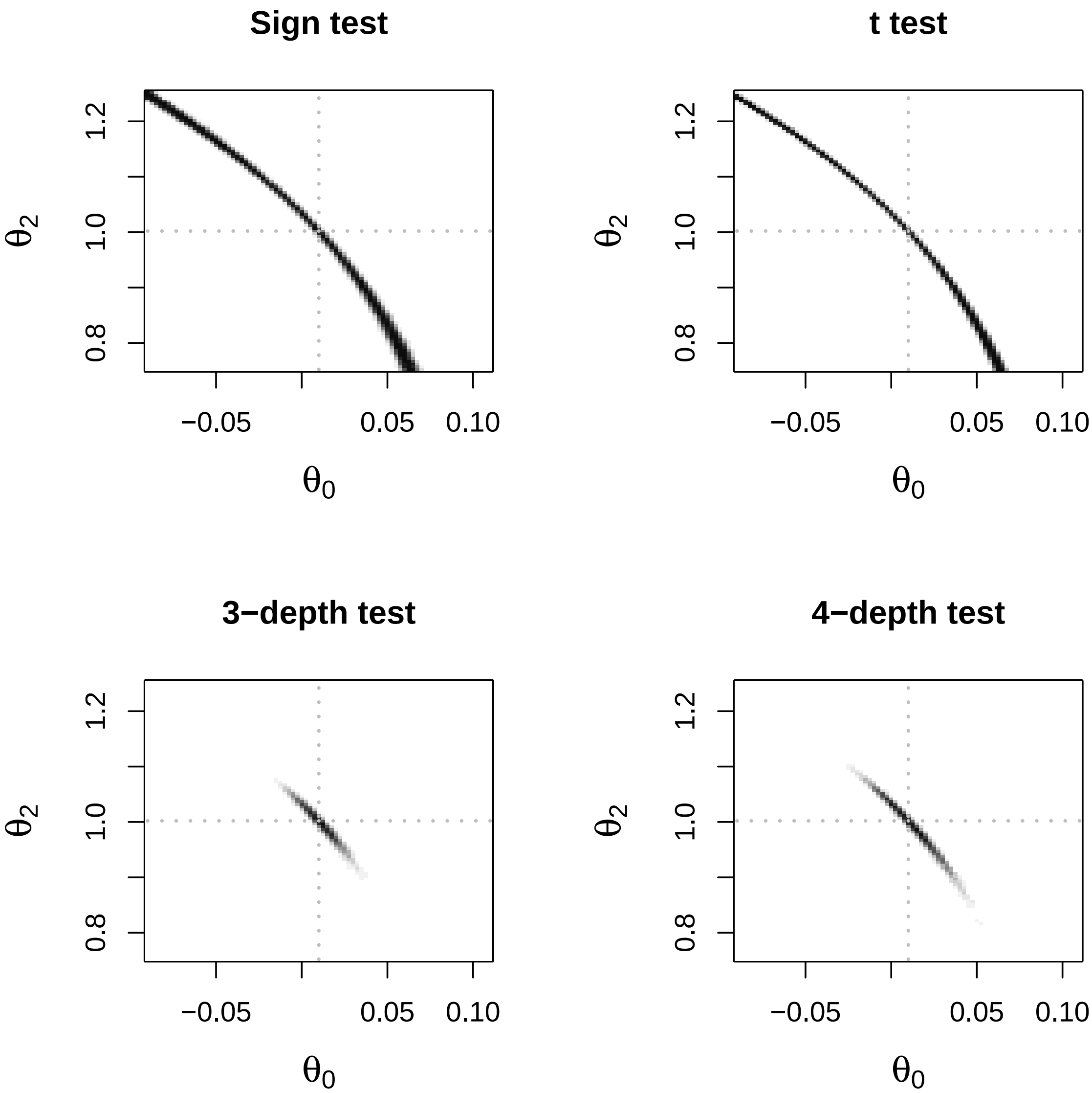

Figure 10: Simulated power of the sign test, 3-depth test, 4-depth test and t test for the nonlinear AR(1)-model where θ_1 is fixed to 0.005 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

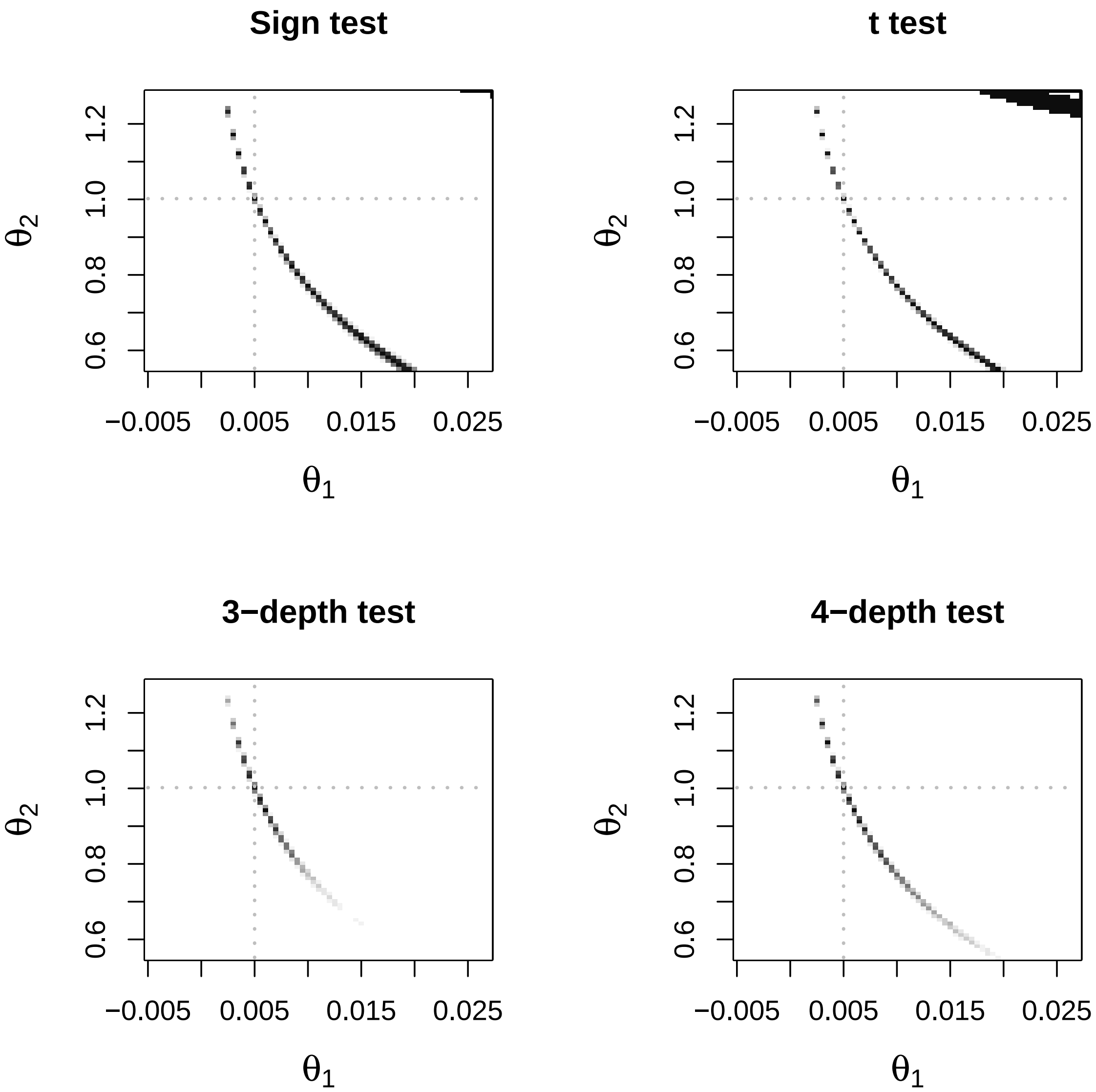

Figure 11: Simulated power of the sign test, 3-depth test, 4-depth test and t test for the nonlinear AR(1)-model where θ_0 is fixed to 0.01 (20 gray levels were used, where black corresponds to $[0, 0.05]$ and white to $(0.95, 1]$).

3.4 Multiple regression

Figures 12 shows the results for multiple regression as considered in the paper but here with sample size $N = 500$. For comparison, Figure 13 displays also the situation for $N = 100$ again. With the enlarged sample size $N = 500$, the power of the K-depth tests increases in all cases compared with $N = 100$. Now for $N = 500$, the robust Wald can be calculated also for $D = 40$ and $D = 80$. However, it is very conservative close to the null hypotheses then. In these close neighbourhoods of the null hypotheses, the K-depth tests with $K = 5$ and $K = 21$ behave better than the robust Wald test. However, for other alternatives, the Wald test remains to be more powerful.

Finally, in Figure 14 we present the simulated power for $N = 100$ in the multiple regression along alternatives of the form $\theta = \gamma \cdot \mathbf{1}_D$, where $\mathbf{1}_D$ is the D-dimensional vector consisting of ones and $\gamma \in [-1, 1]$.

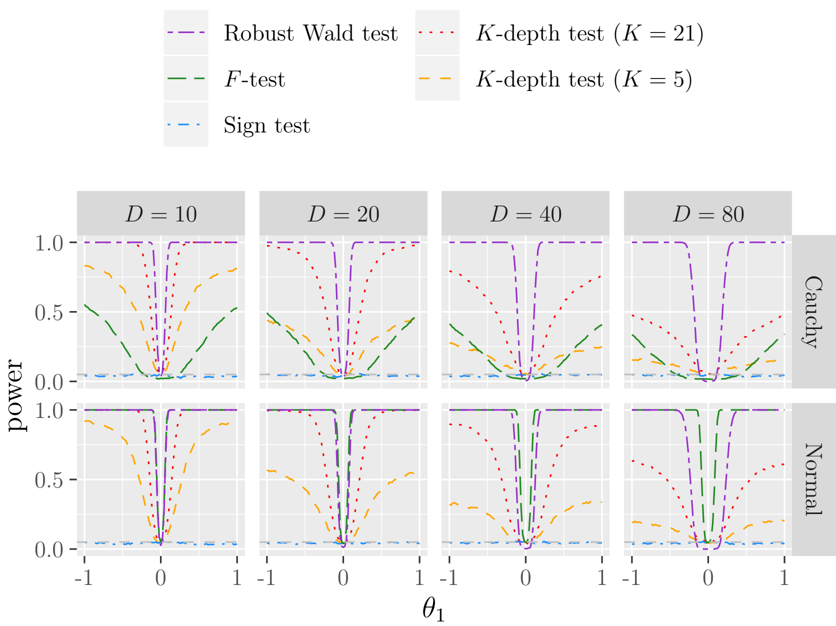

Figure 12: Extracts of the simulated power functions for the model $Y_n = \sum_{d=1}^{D} \theta_d x_{nd} + E_n$ with $N = 500$. Here, the power functions are only shown for $\theta_1 \in [-1,\, 1]$, all other values of θ are zero. The K-depth tests are conducted with an ordering according to the exact solution of the Shortest Hamiltonian Path problem. The gray dashed line shows the level of the test $\alpha = 0.05$.

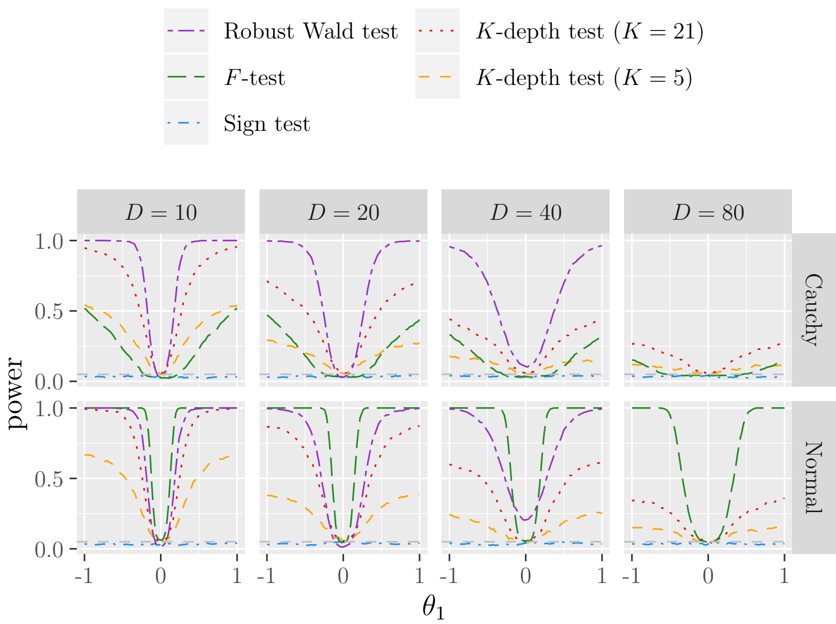

Figure 13: Extracts of the simulated power functions for the model $Y_n = \sum_{d=1}^{D} \theta_d x_{nd} + E_n$ with $N = 100$. Here, the power functions are only shown for $\theta_1 \in [-1, 1]$, all other values of θ are zero. The K-depth tests are conducted with an ordering according to the exact solution of the Shortest Hamiltonian Path problem. The gray dashed line shows the level of the test $\alpha = 0.05$.

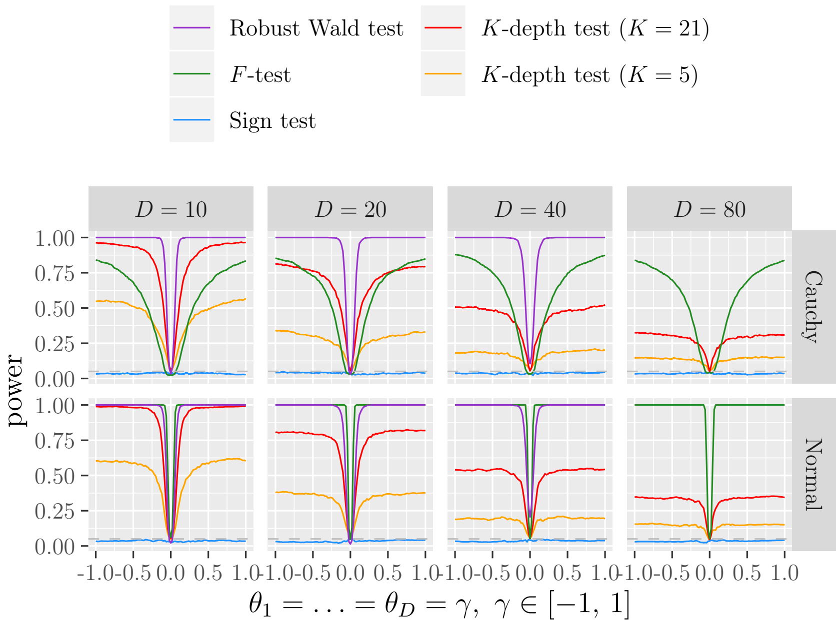

Figure 14: Extracts of the simulated power functions for the model $Y_n = \sum_{d=1}^{D} \theta_d x_{nd} + E_n$ with $N = 100$. Here, the power functions are only shown for alternatives of the form $\theta = \gamma \cdot \mathbf{1}_D$ with $\gamma \in [-1, 1]$. The K-depth tests are conducted with an ordering according to the exact solution of the Shortest Hamiltonian Path problem. The gray line shows the level of the test $\alpha = 0.05$.