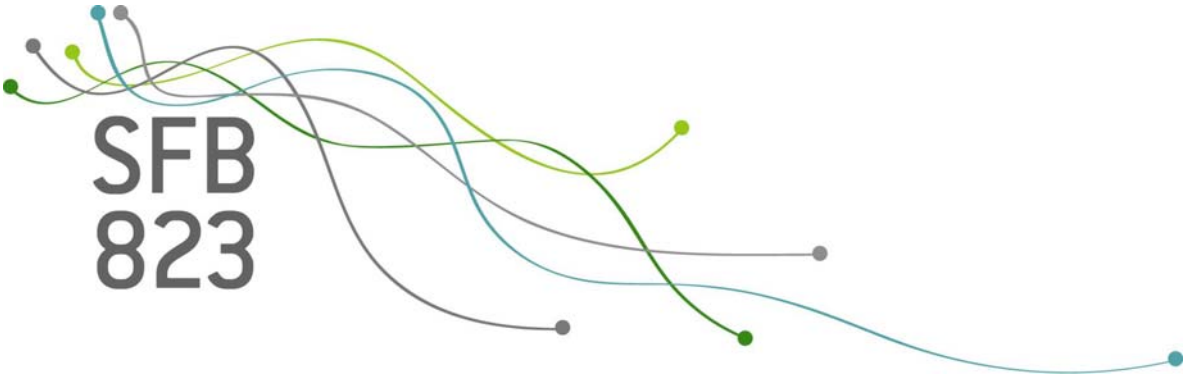
SFB 823

Discussion Paper

Tests based on sign depth for multiple regression

Melanie Horn, Christine H. Müller

Nr. 7/2020



Tests Based on Sign Depth for Multiple Regression

Melanie Horn · Christine H. Müller

Abstract The extension of simplicial depth to robust regression, the so-called simplicial regression depth, provides an outlier robust test for the parameter vector of regression models. Since simplicial regression depth often reduces to counting the subsets with alternating signs of the residuals, this led recently to the notion of sign depth and sign depth test. Thereby sign depth tests generalize the classical sign tests.

Since sign depth depends on the order of the residuals, one generally assumes that the D-dimensional regressors (explanatory variables) can be ordered with respect to an inherent order. While the one-dimensional real space possesses such a natural order, one cannot order these regressors that easily for $D > 1$ because there exists no canonical order of the data in most cases.

For this scenario, we present orderings according to the Shortest Hamiltonian Path and an approximation of it. We compare them with more naive approaches like taking the order in the data set or ordering on the basis of a single quantity of the regressor. The comparison bases on the computational runtime, stability of the order when transforming the data, as well as on the power of the resulting sign depth tests for testing the parameter vector of different multiple regression models. Moreover, we compare the power of our new tests with the power of the classical sign test and the F-test. Thereby, the sign depth tests based on our distance based approaches show similar power as the F-test for normally distributed residuals with the additional benefit of being much more robust against outliers.

Keywords Sign test · data depth · multiple regression · Shortest Hamiltonian Path · Traveling Salesman Problem · nearest neighbor method

Melanie Horn (Corresponding Author)
Faculty of Statistics, TU Dortmund University
Tel.: +49-231-7553114
E-mail: mhorn@statistik.tu-dortmund.de

Christine H. Müller
Faculty of Statistics, TU Dortmund University
Tel.: +49-231-7554238
E-mail: cmueller@statistik.tu-dortmund.de

1 Introduction

In many regression models, only a single outlier can distort the parameters of the model. When testing these parameters, often the parameters of the true model get rejected if only one outlier exists in the data set. A well-known example is the F-test in least-square regression. The literature provides many robust estimators and some robust tests for parameters of regression models, see e.g. the books of Rieder (1994), Jurečková and Sen (1996), Müller (1997), Rousseeuw and Leroy (2003), Huber (2004), Jurečková and Picek (2005), Maronna et al. (2006), Hampel et al. (2011), Clarke (2018). Most of the developed robust regression methods have the disadvantage that a tuning parameter must be chosen. This is not the case for an approach based on data depth introduced by Rousseeuw and Hubert (1999) extending the half space depth of Tukey (1975) to regression depth.

The half space depth has many extensions. Most of them concern multivariate data; see, for example simplicial depth of Liu (1990), zonoid depth treated in Mosler (2002), Mahalanobis depth as used in Hu et al. (2011) or functional data as in López-Pintado and Romo (2009), Claeskens et al. (2014), Agostinelli (2018). Not all of them provide robust methods and most of them concern only estimation.

There are not many extensions of the half space depth to regression and in particular to robust testing. Müller (2005) showed that simplicial regression depth, an extension of Liu's simplicial depth to the regression depth of Rousseeuw and Hubert (1999), can be used for testing the parameters of a regression model robustly. Kustosz et al. (2016b) proved that the simplicial regression depth often reduces to counting the subsets with alternating signs of the residuals so that Leckey et al. (2019) defined *K-sign depth* via the relative number of subsets of K observations with alternating signs of the residuals. This leads to *K-sign depth tests* which generalizes the classical sign test and can be used as soon as the residuals are independent with median zero under the null hypothesis.

The asymptotic distribution of the 3-sign depth was derived in Kustosz et al. (2016a) for AR(1)-regression, however the proof holds for any 3-sign depth if the residuals are independent with median equal to zero. For small sample sizes and not too large K, the distribution of the K-sign depth can be easily simulated. Using such distributions, Kustosz et al. (2016a), Kustosz et al. (2016b), and Leckey et al. (2019) show for several examples that 3-sign depth tests and 4-sign depth tests are much more powerful than the classical sign test and reach the power of the classical nonrobust tests based on least squares of the residuals. However these examples are restricted to the case where the regressor (explanatory variable) is one dimensional so that there is a natural ordering of residuals. Thereby, the ordering is crucial for determining the relative number of subsets of alternating signs of residuals.

Therefore, it is not clear how to extend this approach for the case of multivariate regressors. We propose two types of distance based ordering of the regressors, namely a method based on the Shortest Hamiltonian Path and an approximation of it. These distance based methods have also the advantage that they can be used for ordinal regressors as soon as a distance measure is given. We compare these approaches with more naive methods like using the order of the data set, random order or ordering based on projections or norms of the regressors.

This paper is organized as follows: In Section 2, the sign depth and the corresponding sign depth tests are defined. Section 3 presents the used approaches for ordering multidimensional regressors and compares these approaches by the computational runtime and stability of the order when transforming the data. In Section 4, a simulation study is presented which compares the different approaches of Section 3 with respect to the power of the resulting K-sign depth tests for testing the parameter vector of different multiple regression models. In this section, also the power of these approaches are compared with the power obtained by the classical sign test and the F-test. A conclusion and outlook is given in Section 5.

2 Test Based on Sign Depth in the Context of Regression

In this paper, we focus on multiple linear regression. Hence, at first the notation of a linear regression model is introduced, before the sign depth and the corresponding test is introduced.

2.1 Notation for Multiple Linear Regression

Let $\boldsymbol{y} = (y_1, \ldots, y_N)^\top \in \mathbb{R}^N$, $\boldsymbol{e} = (e_1, \ldots, e_N)^\top \in \mathbb{R}^N$, $\boldsymbol{\theta}, \boldsymbol{x}_n \in \mathbb{R}^D$, and $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_N)^\top \in \mathbb{R}^{N \times D}$ with N, $D \in \mathbb{N}$. Then, a multiple linear regression model is given by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{e}.$$

Here, the errors $e_1, \ldots, e_N$ are realizations of stochastically independent random variables $E_1, \ldots, E_N$ with $P(E_n < 0) = P(E_n > 0) = 0.5$. The latter assumption in particular says that the probability of an error to be zero is zero. These assumptions hold for example for every error distribution which is continuous and has a median equal to zero, like the standardized normal distribution or the standardized Cauchy distribution. With the definition of $\boldsymbol{z}_n := (y_n, \boldsymbol{x}_n^\top)^\top \in \mathbb{R}^{D+1}$ and $\boldsymbol{Z} := (\boldsymbol{z}_1, ..., \boldsymbol{z}_N)^\top$, the residual relating to $\boldsymbol{z}_n$ and a parameter vector $\boldsymbol{\theta}$ is given by $res(\boldsymbol{z}_n, \boldsymbol{\theta}) := y_n - \boldsymbol{x}_n^\top \boldsymbol{\theta}$.

2.2 The Sign Depth

As Rousseeuw and Hubert (1999) noticed for simple linear regression and Kustosz et al. (2016b) proved in the context of linear and nonlinear regression and autoregression, the simplicial regression depth often reduces to what we call the K-sign depth. Hence, the K-sign depth, with $K \geq 2$, can be seen as a generalization of the simplicial regression depth introduced by Müller (2005) for generalized linear models. We define the K-sign depth as follows:

Definition 1 Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_N$ be ordered according to some criteria and $K \in \mathbb{N}$ with $2 \leq K < N$. Then the **K-sign depth** d_S^K of parameter $\boldsymbol{\theta}$ in the data set $\boldsymbol{Z}$ is defined via:

$$\begin{aligned} s_{n_k}^{(1)} &:= \prod_{k=1}^{K} \mathbb{1}\{res(\boldsymbol{z}_{n_k}, \boldsymbol{\theta}) \cdot (-1)^k > 0\} \\ s_{n_k}^{(2)} &:= \prod_{k=1}^{K} \mathbb{1}\{res(\boldsymbol{z}_{n_k}, \boldsymbol{\theta}) \cdot (-1)^k < 0\} \\ d_S^K(\boldsymbol{\theta}, \boldsymbol{Z}) &:= \binom{N}{K}^{-1} \sum_{1 \leq n_1 < n_2 < \cdots < n_K \leq N} (s_{n_k}^{(1)} + s_{n_k}^{(2)}) \end{aligned}$$

where $\mathbb{1}\{...\}$ denotes the indicator function, which is one if the condition in the curly braces holds and zero otherwise.

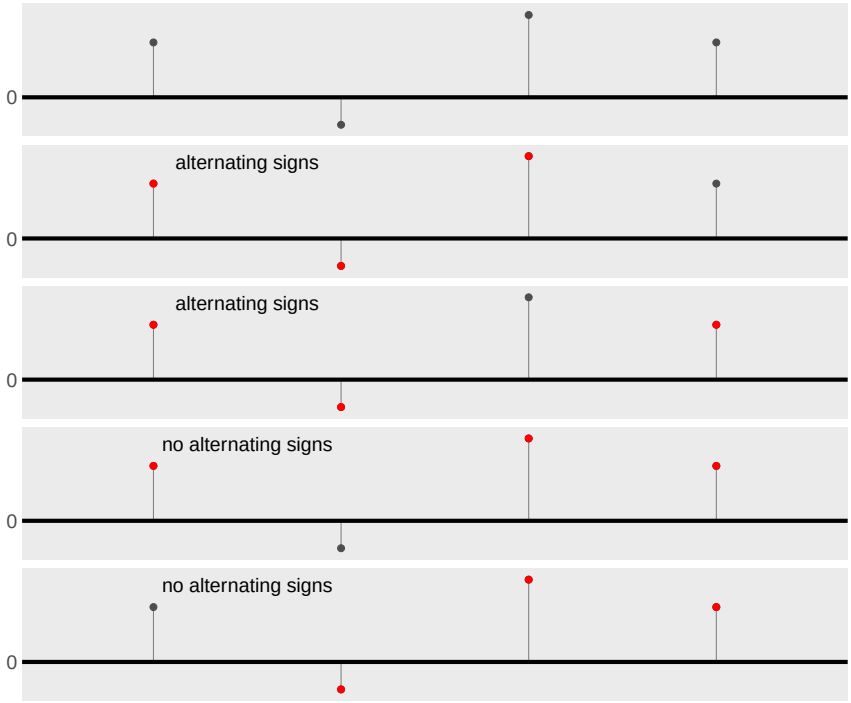


Fig. 1 Example of calculating the K-sign depth with parameter $K = 3$. The topmost figure shows four residuals of an arbitrary model, the lower figures show the $\binom{4}{3} = 4$ combinations of 3-tuples of these residuals. Here, two out of four 3-tuples have alternating signs, so the 3-sign depth is 0.5

In short, the K-sign depth is the relative number of ordered K-tuples with alternating signs of the residuals. It can be used as soon as a given ordering of the residuals is available. An example of calculating the 3-sign depth can be found in Figure 1.

Note that the value of the K-sign depth depends on the ordering of the regressors. In simple linear regression with only one explanatory variable (and possibly an intercept), there exists a distinct order of the regressors according to the values of the explanatory variable. But for regression with more than one explanatory variable, one can think of many different criteria for ordering the regressors. Our aim for this paper is to propose some possible ordering criteria and compare them with regard to theoretical properties, like stability of the order when transforming the data, and practical behavior in a simulation study. An example why the ordering criterion has a crucial effect on the sign depth is given in Example 1 below.

Example 1 Let $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{e}$ be a linear regression model with $\boldsymbol{x}_n = (1, x_{n1}, x_{n2})^\top$, $x_{n1}, x_{n2} \in [-1, 1]$, $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^\top$ and E_n *i.i.d.* $\mathcal{N}(0, 0.01)$. The true model has the parameter vector $\boldsymbol{\theta}_{true} = (0, 1, 1)^\top$. $N = 16$ observations of this model are given. To these data points, a regression model with the parameter vector $\hat{\boldsymbol{\theta}} = (-0.05, -0.5, 1)^\top$ is fitted. A visualization of the data points and the fitted model can be found below in Figure 2. It can be seen that the fitted model produces eight positive residuals and eight negative residuals.

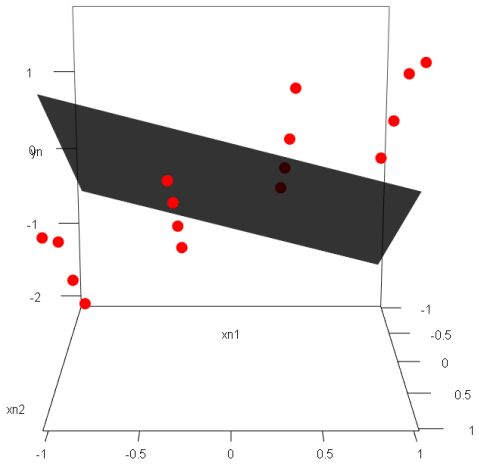
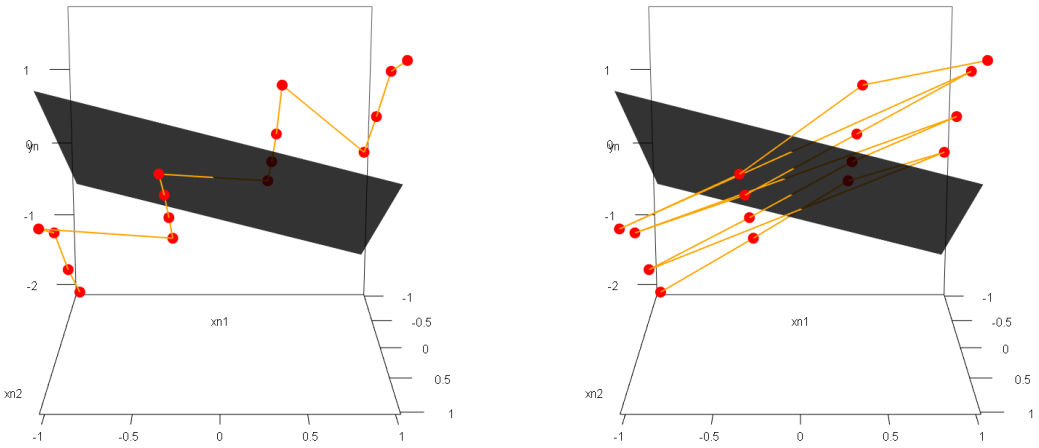


Fig. 2 Visualization of the 16 data points (red) and the fitted model (dark grey)

A possible method for ordering the regressors is to sort the regressors according to the values of only one explanatory variable and neglect the values of all other (see Section 3.2.) In this example, this would lead to two possible orders which are shown below in Figure 3.



(a) Ordering the regressors according to x_{n1}

(b) Ordering the regressors according to x_{n2}

Fig. 3 Visualization of two possible orders for the 16 regressors. The order is marked as orange line

As discussed below, these two orders lead to very different values of the K-sign depth. When ordering according to the values of x_{n1}, only one sign change in the residuals occurs. Ordering according to the values of x_{n2} leads to seven sign changes. This has crucial effect on all the K-sign depths with $K \geq 3$, as it can be seen in Table 1.

Table 1 K-sign depth with $K \in \{2, 3, 4, 5\}$ for the orders (a) and (b) of Figure 3

	$d_S^2(\hat{\boldsymbol{\theta}}, \boldsymbol{Z})$	$d_S^3(\hat{\boldsymbol{\theta}}, \boldsymbol{Z})$	$d_S^4(\hat{\boldsymbol{\theta}}, \boldsymbol{Z})$	$d_S^5(\hat{\boldsymbol{\theta}}, \boldsymbol{Z})$
(a)	0.533	0	0	0
(b)	0.533	0.286	0.176	0.088

In this paper, we will focus on the case $K = 3$ and $K = 4$ because the 3-sign depth and 4-sign depth are equivalent with the simplicial regression depth for a regression model with a parameter vector $\boldsymbol{\theta}$ which is two-dimensional and three-dimensional, respectively (Kustosz et al., 2016b). For a small number of observations (up to approximately $N = 25$), the distributions can be calculated exactly. For this, the 3-sign depth and the 4-sign depth of all 2^N possible sign combinations of the residuals have to be calculated. For a larger number of observations, simulation of the distributions is possible by using not all possible sign combinations but only a random subset. When the number of observations is sufficiently large, an asymptotic distribution of the standardized 3-sign depth $N \cdot (d_S^3(\boldsymbol{\theta}, \boldsymbol{Z}) - 0.25)$ was shown in (Kustosz et al., 2016a). The derivation of the asymptotic distribution of the general K-sign depth is in progress. This work indicates that K-sign depth can be calculated in linear time using an asymptotically equivalent form which is exact for 3-sign depth and 4-sign depth.

2.3 The Sign Depth Test

On the basis of the K-sign depth, testing the parameters of a regression model can be done. When the fitted parameter vector of a model is correct, the residuals scatter independently of each other around zero. In contrast, if the fitted parameter vector is not correct, the independence of the signs of the residuals is often violated and the residuals do not scatter around zero. This leads to fewer K-tuples with alternating signs and hence to a smaller K-sign depth. When the value of the K-sign depth is smaller than the α-quantile of the distribution of the corresponding K-sign depth, it is shown significantly to the level $\alpha \in (0, 1)$ that the considered model does not fit the data. The K-sign depth test can be written as follows:

Theorem 1 ***K-Sign Depth Test***

Let $\alpha \in (0, 1)$ and $K \in \mathbb{N}$ be fixed with $2 \leq K < N$, where N is the number of observations in the dataset $\boldsymbol{Z}$. Let $q_\alpha^{K,N}$ denote the α-quantile of the distribution of the K-sign depth for N observations. Then, the test

$$\varphi^K(\boldsymbol{\theta}, \boldsymbol{Z}) = \mathbb{1}\left\{ \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} d_S^K(\boldsymbol{\theta}, \boldsymbol{Z}) < q_\alpha^{K,N} \right\}$$

is a test to the level α for $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ vs. $H_1 : \boldsymbol{\theta} \notin \boldsymbol{\Theta}_0$.

Hence, the K-sign depth test rejects the null hypothesis $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ if the K-sign depth of all parameters $\boldsymbol{\theta}$ of the null hypothesis set $\boldsymbol{\Theta}_0$ is too small. For a proof of this simple test strategy based on data depth, see Müller (2005).

The K-sign depth test is a generalization of the classical sign test since the 2-sign depth test is equivalent with the classical sign test, see Leckey et al. (2019). For testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs. $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, the classical sign test is given by

$$\varphi(\boldsymbol{\theta}, \boldsymbol{Z}) = \mathbb{1}\{N_+ \notin [b_{N,0.5,\alpha/2}, b_{N,0.5,1-\alpha/2}]\},$$

where $N_+ \leq N$ is the number of positive residuals and $b_{N,0.5,\alpha/2}$ and $b_{N,0.5,1-\alpha/2}$ denote the $\alpha/2$-quantile and $(1-\alpha/2)$-quantile of the binomial distribution with parameters N and 0.5.

3 Methods for Ordering Multidimensional Data

As Example 1 shows, the order of the regressors is of high importance for the value of the sign depth and the result of the sign depth test. In the next subsections some possible criteria for ordering the regressors are presented, where not only the ordering methods are studied, but also their computational runtime and their behavior when transforming the data.

All computational results in this section and Section 4 were obtained with the software `R` (R Core Team, 2018) and our self-written `R`-package `GSignTest` (Horn, 2019), in which all presented methods of this paper are implemented, including the ordering methods, the K-sign depth, the K-sign depth tests, the classical sign test and the F-test for regression models. The package is freely available on Github. The graphics in this paper were created using the packages `ggplot2` (Wickham, 2009), `gridExtra` (Auguie, 2017) and `rgl` (Adler et al., 2018). The simulations were performed on the High Performance Cluster *LiDO3*[1] of the TU Dortmund University.

In this section, we use computational results to demonstrate the differences between the methods for ordering multidimensional data. To show the main concepts of the ordering methods, we use two data sets where each contains 20 two-dimensional regressors in $[-1, 1]^2$. The first data set, called LHS, is obtained by a Latin Hypercube Sampling (LHS) with maximizing the minimal distance between two points (Stein, 1987). The regressors of the second one, called Spiral, are arranged as a spiral with little noise. To compare the runtimes of the methods, we simulated 200 times a data set with $100 \times k$, $k = 1, \ldots, 10$, two-, three- and four-dimensional regressors using uniformly distributed random values in $[-1, 1]$ for every dimension of the regressors.

3.1 Naive Methods

The simplest methods for ordering multidimensional data are taking the order the observations appear in the data set or taking a random order. These two methods are equivalent if the order of the observations in the data set is already a random order. Because the order does not depend on the observations itself but on the index of the observation, these two methods are stable when applying transformation on the data, which may be important in some situations.

Figure 4 shows the obtained order by these methods with the help of the LHS and Spiral data sets. It can be seen that the two methods are equivalent for the LHS data because the used `R`-package `lhs` (Carnell, 2019) provides the regressors already in a random order. On the other hand, the two methods differ very much on the Spiral data set because the used `R`-package `mlbench` (Leisch and Dimitriadou, 2010) provides the regressors in a natural order.

[1] `https://lido.tu-dortmund.de/cms/de/LiDO3/index.html`

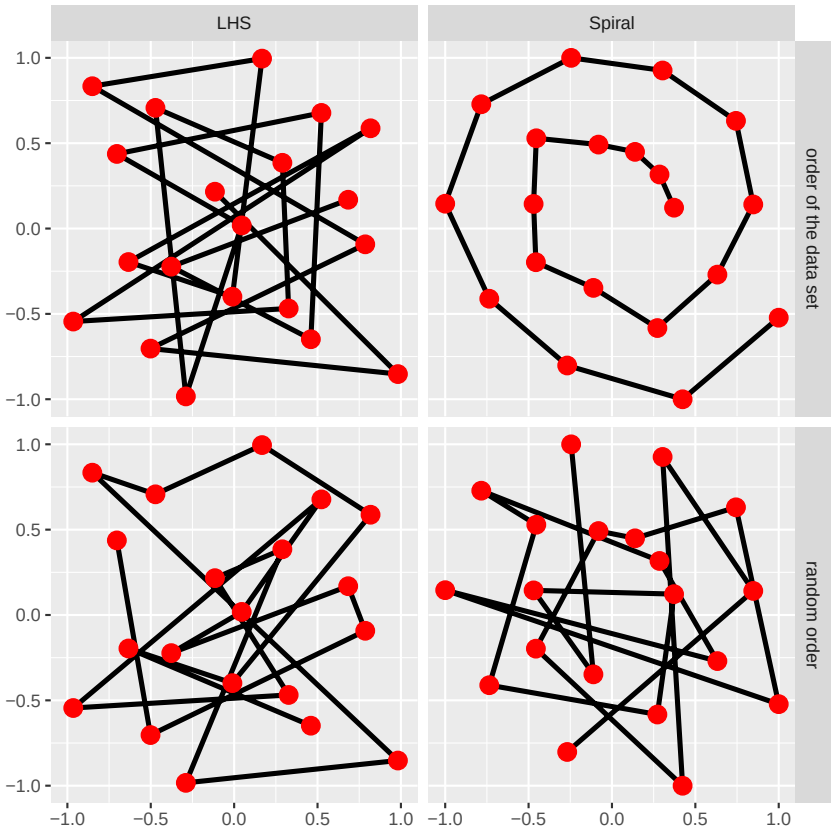


Fig. 4 Example of the two naive ordering methods described in Section 3.1: according to the data set (upper row) and random ordering (lower row). The left figures provide regressors of a two-dimensional Latin Hypercube Sampling (LHS) and the right figure's regressors are arranged as a two-dimensional spiral

Another advantage of these two methods is the runtime of the ordering process. When no ordering is done, the runtime obviously is $\mathcal{O}(1)$. When a random order is used, the Fisher-Yates-algorithm for randomly permuting data, which is used in `R`, has a runtime of $\mathcal{O}(N)$ (Durstenfeld, 1964). This behavior can also be seen in Figure 5. There, the empirical runtimes of sorting different regressor and dimension numbers are shown. As the theoretical runtimes say, the runtime of both methods does not depend on the number of dimensions. In addition, it can be seen that the runtimes are very small, so that only little time in the whole testing process has to be spent on the ordering process here.

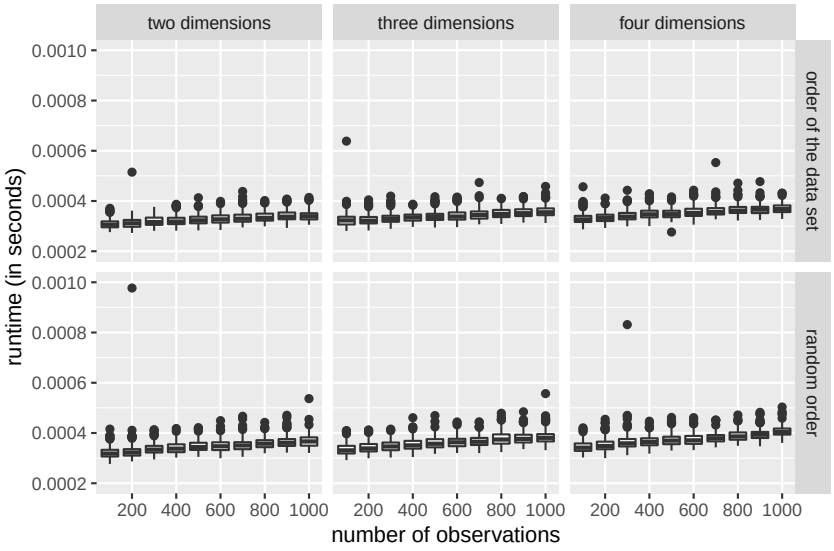


Fig. 5 Empirical runtimes of the two naive ordering methods described in Section 3.1: according to the data set (upper row) and random ordering (lower row)

3.2 Scalarization Based Methods

Another approach for ordering multidimensional data is ordering the regressors on the basis of scalarized values of each regressor. One possible scalarization is taking an arbitrary vector norm. Another method is projecting all regressors orthogonal on a line, for example the bisecting line. A special case of the orthogonal projection is ordering the regressors on the basis of the values of only one dimension, like in Example 1. Here, the regressors are projected on a line parallel to a coordinate axis. A disadvantage of all scalarizations is that the order directly depends on the values of the regressors. When transforming the data, the order might get completely different which is unwanted in most cases. An example of this behavior can be found in Example 1: When flipping the coordinate axes (or changing the projection the ordering is based on), the sign depth test can come to completely contrary results. When taking the norm as the scalarization, the ordering highly depends on the origin of the coordinate system because every vector norm produces positive values, no matter if the values in the vector are positive or negative. Figure 6 shows an example of the ordering on the same two data sets like in Section 3.1.

Besides the disadvantage of the non-stable order by data transformation, an advantage of the scalarization based methods is the rather short runtime for computing the order. Many scalarization methods, like calculating an arbitrary vector norm as well as projecting the regressor on an arbitrary line, need linear runtime. Because a single regressor has dimension $D \in \mathbb{N}$, the runtime for computing the scalarized value of one regressor is $\mathcal{O}(D)$. Doing this for all $N \in \mathbb{N}$ regressors, this leads to a runtime of $\mathcal{O}(ND)$. Sorting these values can be done in approximately $\mathcal{O}(N)$, when using *Radix Sort* (Knuth, 1998), which is used in `R` by default for 200 up to 100 000 numeric values. Overall, the runtime for ordering multidimensional regressors according to a scalarized value scales linearly in the number of regressors and dimensions, i.e. $\mathcal{O}(ND)$. The special case of the projection method, using only the values of one dimension for ordering,

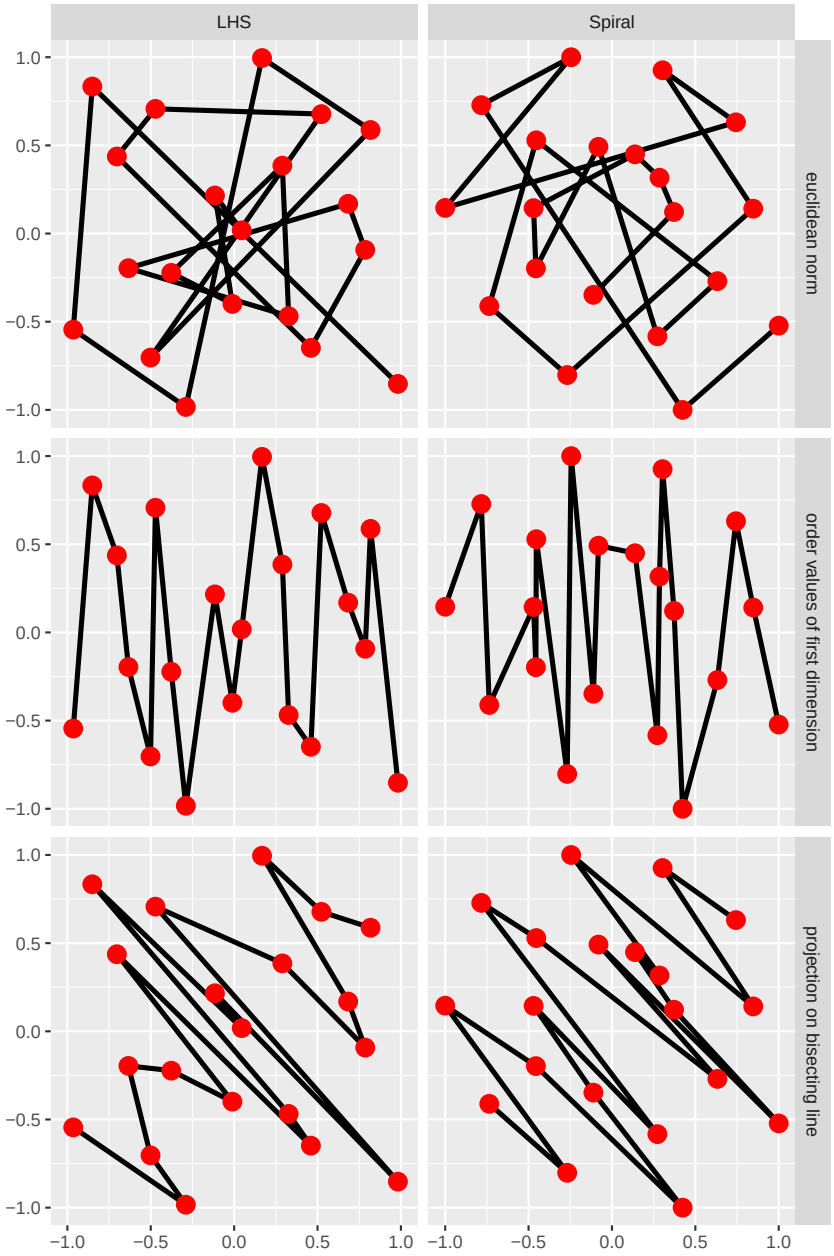


Fig. 6 Example of three scalarization based ordering methods described in Section 3.2: ordering according to the euclidean norm (upper row), according to the projection on the first dimension (middle row), and according to the projection on the bisecting line (lower row)

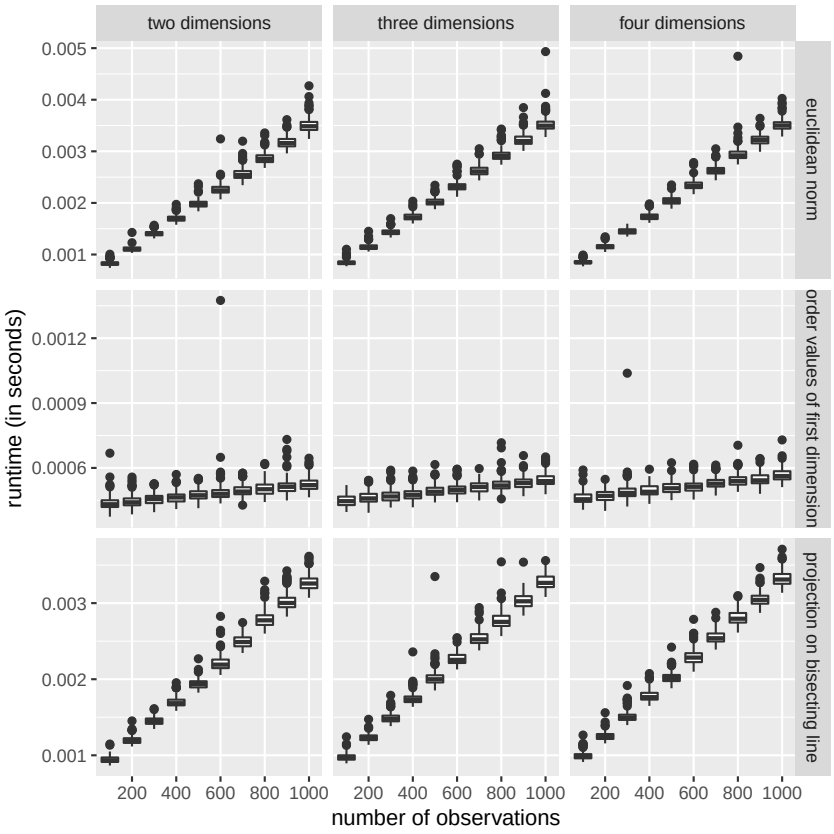


Fig. 7 Empirical runtimes of the three scalarization based ordering methods described in Section 3.2: ordering according to the euclidean norm (upper row), according to the projection on the first dimension (middle row), and according to the projection on the bisecting line (lower row)

can be sped up by using directly the radix sort algorithm on the values of the considered dimension. This leads to an approximative linear runtime which is independent of the number of dimensions, i.e. $\mathcal{O}(N)$. When looking at Figure 7 with the empirical runtimes of both methods including the special case, it can be seen, that all scalarization based methods seem to scale linearly in the number of regressors. In contrast, the linear trend in the number of dimensions is neglectable. On the other hand, it can be seen that using only the values of one dimension for sorting is much faster than its generalization. But overall, the runtimes of these ordering methods are very small.

3.3 Distance Based Methods

Taking the values of the regressors for ordering has some disadvantages as explained in Section 3.2. For this, an alternative can be to compute an order based on the pairwise distances of the regressors. Then,

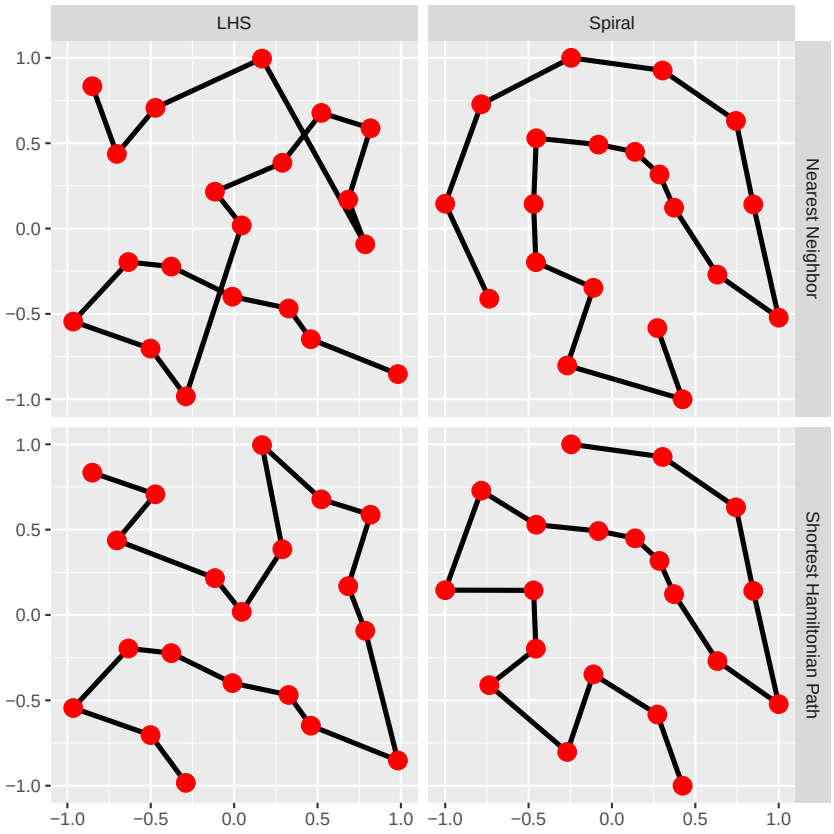


Fig. 8 Example of the two distance based ordering methods described in Section 3.3: the approximative nearest neighbor algorithm (upper row) and the exact solution of the SHP (lower row)

the ordering does not get influenced by any data transformation that does not change the relation of the distances, like flipping coordinate axes or shifting the origin of the coordinate system. Another advantage of distance based methods is that these methods can also be used for ordinal data, if the distance between two ordinal values can be expressed by a numeric value.

Ordering based on the pairwise distances of the regressors means that regressors that have a small distance to each other get ordered near to each other and regressors that have large distance are far away from each other in the ordering. Here, the term distance can mean an arbitrary distance measure, although in this paper always the euclidean distance is used. This aim for ordering can be rewritten as the problem to find a shortest path through the data. This so called *Shortest Hamiltonian Path* (SHP) is a simple transformation of the well-known *Traveling Salesman Problem* (TSP) (Lawler et al., 1985). Here, a solution for a complete graph has to be computed. Because the TSP (and also the SHP) belong to the so called NP-hard problems (which means in the worst case the exact solution cannot be computed in polynomial runtime), many approximative algorithms with faster runtime exist. One approximative algorithm is a *nearest neighbor* approach. The idea is to start at an arbitrary point and always proceed

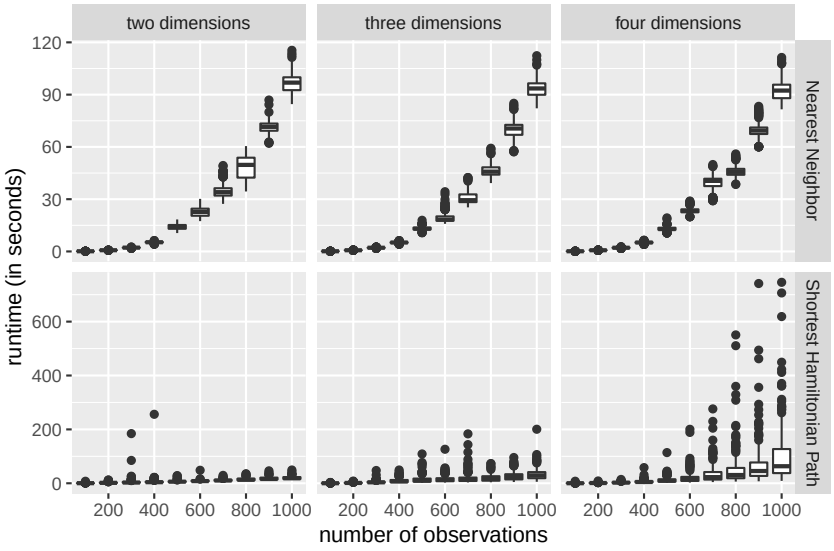


Fig. 9 Empirical runtimes of the two distance based ordering methods described in Section 3.3: the approximative nearest neighbor algorithm (upper row) and the exact solution of the SHP (lower row)

to the nearest not yet chosen point until all points have been visited. This has to be done N times (every regressor is starting point once) and the ordering with the shortest obtained path is chosen. It can be proven that the obtained path length of the nearest neighbor approach is at most longer than the optimal path of the exact solution of the SHP by a factor of $\frac{1}{2}\lceil\log_2(N)\rceil + \frac{1}{2}$ (Rosenkrantz et al., 1977). Two examples of the exact and the approximative solution of the shortest path problem are shown in Figure 8. The exact solution has be computed with the help of the TSP-solver *Concorde* (Applegate et al., 2004), which is a *branch-and-cut* algorithm implemented in `C`. Connecting the code to `R` has been done with the help of the `R`-package `TSP` (Hahsler and Hornik, 2017), which has also been used for computing the nearest neighbor solutions.

For calculating the runtime of the distance based methods, first the runtime of computing a full distance matrix has to be known. For most distance measures, calculating the distance of two D-dimensional points needs looking at every entry of the vectors once, so that calculating a distance needs runtime $\mathcal{O}(D)$. Because $\frac{N(N-1)}{2}$ distances have to be computed, the overall runtime of calculating the distance matrix is $\mathcal{O}(DN^2)$. For calculating the exact solution of the shortest path problem, the best known worst case runtime is $\mathcal{O}(N^2 2^N)$, see for example Applegate et al. (2006). For the approximative nearest neighbor algorithm calculating a tour from a fixed starting point needs $\mathcal{O}(N^2)$ (Applegate et al., 2006). Doing this once for every possible starting point obviously leads to an overall runtime of $\mathcal{O}(N^3)$. Assuming that $N \geq D$, the runtime for calculating the distance matrix can be neglected, so that overall the exact algorithm has runtime $\mathcal{O}(N^2 2^N)$ and the approximative algorithm has runtime $\mathcal{O}(N^3)$. The empirical runtimes of both approaches can be found in Figure 9. The cubic runtime of the nearest neighbor approach can be nicely seen. It can also be seen that these two methods take a lot more time than the methods in the previous sections whereas the approximative algorithm needs less time than the exact one in the worst case.

4 Simulation Study

To compare the sign depth tests based on the ordering methods described in Section 3 with each other and with the classical sign test and F-test, some simulations of the power functions of these tests were made. In this section, at first the set-up of the simulation study is described and afterwards the results will be shown.

4.1 Set-up of the Simulation Study

As stated before, in this paper we will focus on tests based on the 3-sign depth and the 4-sign depth. The simulation study will compare the simulated power functions based on five different linear models: One standard linear model with two explanatory variables without intercept, the same with intercept, a linear model with two explanatory variables with interaction between the explanatory variables, and two linear models without intercept with three and four explanatory variables:

(a) $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + e_n$,
(b) $y_n = \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + e_n$,
(c) $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n1} x_{n2} + e_n$,
(d) $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n3} + e_n$,
(e) $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n3} + \theta_4 x_{n4} + e_n$.

As it can be seen, the parameter vector $\boldsymbol{\theta}$ is two-dimensional in Model (a), three dimensional in Models (b), (c) and (d), and four-dimensional in Model (e). The entries of the error vector $\boldsymbol{E} = (E_1, \ldots, E_N)^\top$ are independent and identically distributed with median zero. For this, two different distributions have been used:

(1) normal distribution: $E_n \sim N(0, \sigma^2)$,
(2) Cauchy distribution: $E_n \sim Cau(0, s)$,

where $s = u_{0.75}\sigma$ with $u_{0.75}$ denoting the 75%-quantile of the standard normal distribution and σ is chosen to be 0.2. Taking this choice of s implies that the interquartile range of both error distributions is the same.

For these $5 \times 2 = 10$ combinations of experiments, the power function of the test hypotheses $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ are simulated. For this, the explanatory variables x_{n1}, x_{n2}, x_{n3}, x_{n4} are independently sampled from a continuous uniform distribution on the interval $[-1, 1]$. The number of regressors N is chosen to be 100 and the tests are performed to the level of $\alpha = 0.05$. The critical values of the depth tests based on 3-sign depth and 4-sign depth are obtained by choosing 1 000 000 times randomly one of the 2^{100} possible sign combinations of the residuals. The value of the power function at a given point $\boldsymbol{\theta}$ is estimated by calculating the relative number of rejections of H_0. For this, 1 000 independent replications of the test at each $\boldsymbol{\theta}$ are made, where $\boldsymbol{\theta}$ is chosen from an equidistant grid from $[-2, 2]^2$ and $[-2, 2]^3$, respectively, with 51 point in every dimension for the Models (a), (b), and (c). For the comparison of the Models (a), (d), and (e) in Subsection 4.6, grids from $[-2, 2]^k$, with 51 points for $k = 2$ and $k = 3$ and 19 points for $k = 4$ in every dimension are used.

In Subsections 4.2, 4.3, 4.4, and 4.5, at first we study the behavior of the different orderings for depth tests based on 3-sign depth. The two best ordering methods are then compared in depth tests based on 3-sign depth and 4-sign depth in Subsection 4.6.

4.2 Results for the model with two explanatory variables and without intercept

Starting with the results of the experiments for Model (a), we will look at an extract of the simulated power functions. Figure 10 shows these power functions on the interval $\boldsymbol{\theta} \in [-1, 1]^2$ for the sign depth test with $K = 3$ on the basis of the seven ordering methods described in Section 3. As reference methods, the power functions of the classical sign test and the F-test are also shown.

It can be seen in Figure 10 that independent of the error distribution the naive methods fail to build acceptable power functions. The power is always about 0.05, which means that the power values are not clearly larger than α in the area of the alternative. This result shows that taking a random order may lead to bad power of the depth test. Taking the order of the data set seems to lead to the same result, but note that here the order of the data set is already a random order, so that this result may not hold when the data set contains an inherent order.

Looking at the scalarization based methods in Figure 10, it can be seen that ordering the regressors according to their values of the euclidean norm does also not lead to an acceptable power function. In contrast, projecting the regressors orthogonal on a line leads to power functions with better characteristics. It can be seen that the power converges to one in most directions. Only the direction orthogonal to the direction of the projection line gets no acceptable power values. The reason for this behavior can be seen in Subfigure 3 (b) in Example 1 on Page 4: when projecting the regressors on a line, there exists a direction where the residuals are alternating and the sign depth test cannot reject the null hypothesis.

In contrary, the distance based methods seem to build very good power functions. The power functions are only at $\boldsymbol{\theta} = \mathbf{0}$ less or equal to 0.05 and converge in every direction towards one. It can be seen that the exact solution of the Shortest Hamiltonian Path problem achieves slightly better power values than its approximation. But remembering the clear shorter runtime of the approximation algorithm, the results of both methods are satisfying.

Looking at Figure 10, it can be seen that the results of all sign depth tests do not depend very much on the distribution of the errors. This behavior makes sense because the depth test only depends on the signs of the residuals and not on the values of the residuals. This is different to the F-test. Here, the results are not surprising: when taking normally distributed errors the F-test cannot be beaten by the sign depth tests. But when taking Cauchy distributed errors the F-test gets clearly worse than the distance based methods. The classical sign test can (nearly) never reject the null hypothesis, but this comes from the fact that the examined model has no intercept and the regressors are chosen symmetrically around zero. Without intercept the fitted model always goes through the origin and so, always about half of the residuals are positive.

4.3 Results for the model with two explanatory variables and intercept

Modeling a linear model with two explanatory variables and an intercept, of course leads to similar results when the intercept θ_0 is zero. But, as it can be seen in Figure 11, as soon as the intercept slightly differs from zero, the sign depth test rejects the null hypothesis more often when applying the naive ordering methods. It also can be seen that shifting the intercept leads to higher power values when θ_1 and θ_2 are close to zero. As soon as the slope parameters θ_1 and θ_2 differ too much from zero (and so the slope is different from the slope of the null hypothesis), the fitted model produces more sign changes in the residuals and the power gets smaller. When regarding larger values for θ_1 and θ_2 than shown in Figure 11, one could see that the power function gets close to α. In Figure 11, only the power functions for ordering

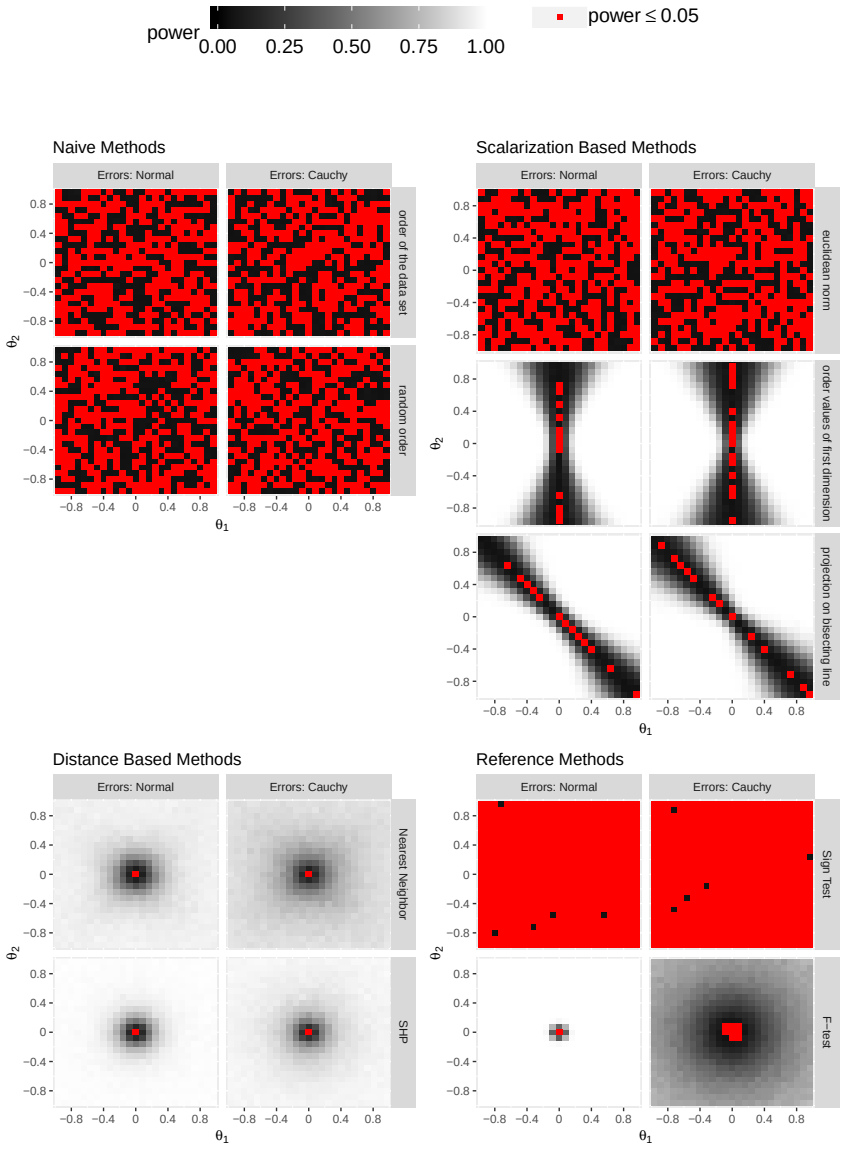


Fig. 10 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (a): $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + e_n$. Here, the depth tests with $K = 3$ and different ordering methods are compared with the reference methods, the classical sign test and the F-test. The areas, where the power is smaller than α, are marked in red

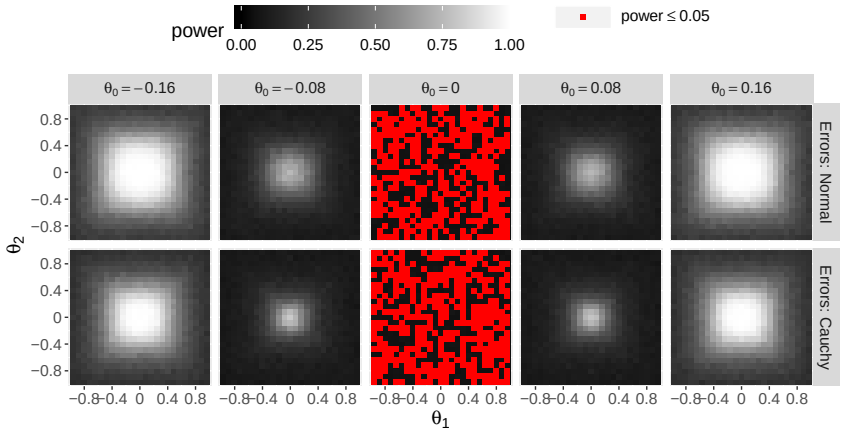


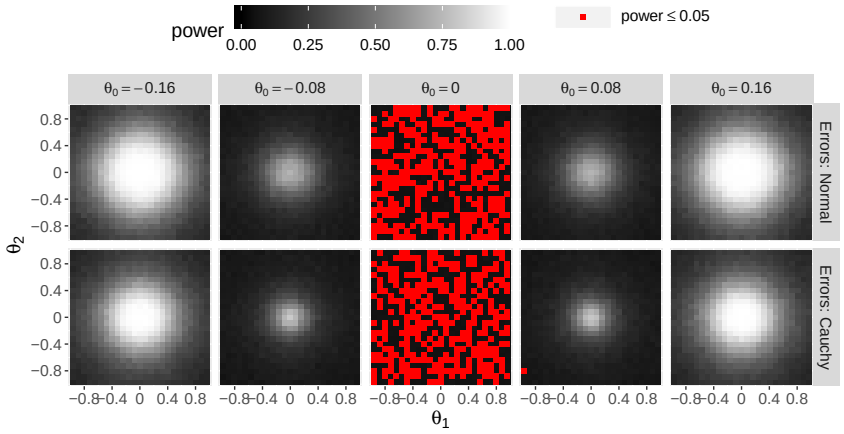
Fig. 11 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (b): $y_n = \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + e_n$. Here, the depth test with $K = 3$ and ordering the regressors according to their appearance in the data set is applied. The areas, where the power is smaller than α, are marked in red

the regressors according to their appearance in the data set are shown. As explained above, taking a random order would lead to the same results.

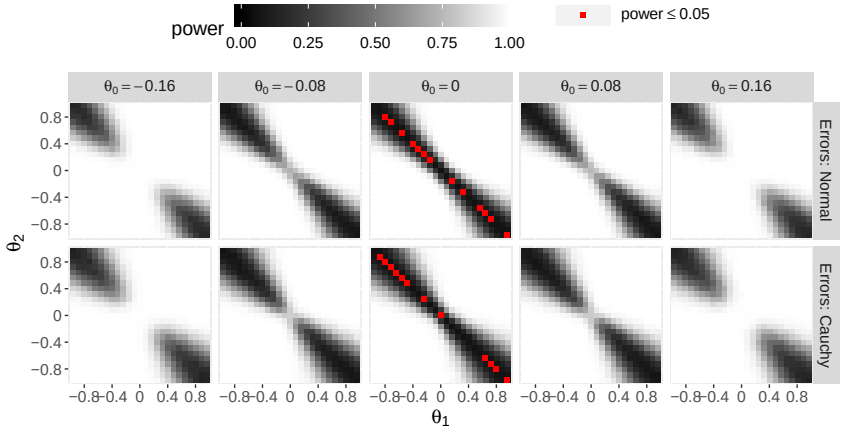
Subfigure 12 (a) shows results similar to the ones in Figure 11. Ordering the regressors on the basis of their values of the euclidean norm leads to no good results when the intercept is zero. When the intercept differs from zero, the null hypothesis gets rejected more often, but the more the slopes differ from zero, the smaller the power gets. The second scalarization based ordering method also accepts the null hypothesis only when the intercept is zero. In contrast to scalarization based on the euclidean norm, the power functions based on scalarization by an orthogonal projection have larger areas with large power values as can be seen in Subfigure 12 (b). As already mentioned above, the power values are small only in the orthogonal direction of the projection line. But with larger values of the intercept, the power values on this line also get larger, especially for slopes close to zero. The results are the same for ordering the regressors on the basis of the values of the first dimension because they are equivalent to the results of projecting the regressors on the bisecting line.

Looking at the power functions of the distance based methods in Figure 13, it can be seen that these power functions have desirable characteristics. The power function only gets close to α when $\boldsymbol{\theta}$ is $\mathbf{0}$ and with increasing values of any component of $\boldsymbol{\theta}$ the power values get larger and converge to one. This result holds for both methods, the ordering on the basis of the exact Shortest Hamiltonian Path algorithm as well as for the ordering on the basis of the approximative nearest neighbor algorithm, but the power values of the exact algorithm are slightly larger in the alternative.

The reference methods in Figure 14 show the results which can be expected by the results in Figure 10. The F-test cannot be beaten by the sign depth test in case of normally distributed errors, but has bad performance in case of Cauchy distributed errors. The classical sign test behaves more like the sign depth test with naive ordering. Here again, a shift of the intercept leads to a significant change of the number of positive residuals.



(a) Ordering the regressors according to their values of the euclidean norm

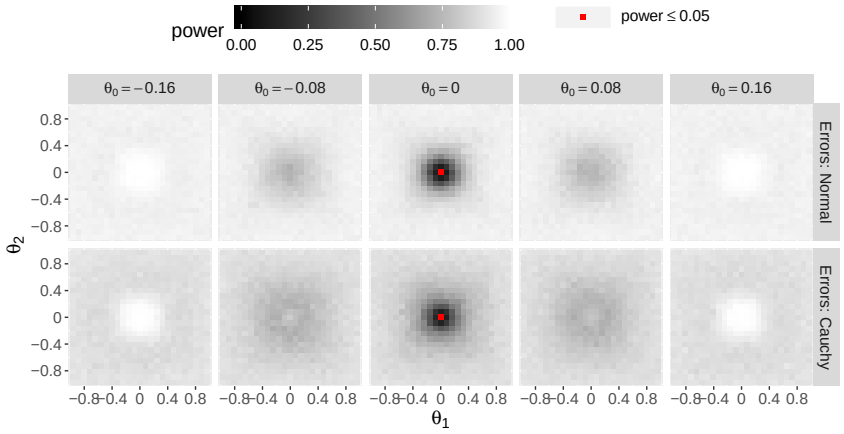


(b) Ordering the regressors according to their projected values on the bisecting line

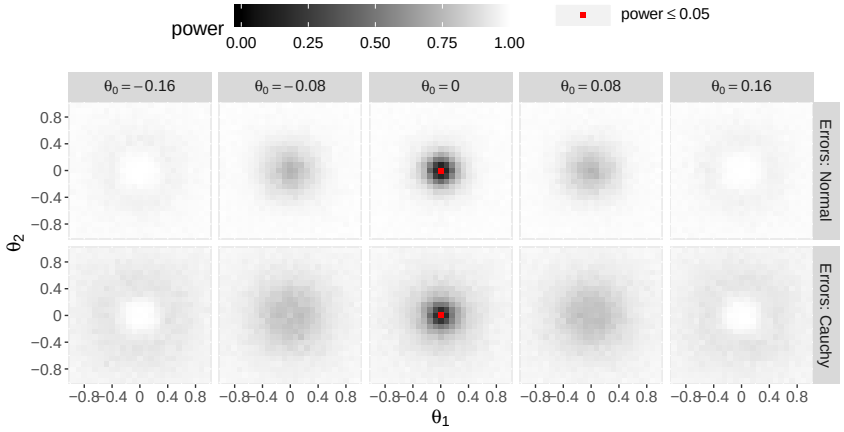
Fig. 12 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (b): $y_n = \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + e_n$. Here, the depth tests with $K = 3$ and ordering the regressors according to scalarizations are applied. The areas, where the power is smaller than α, are marked in red

4.4 Results for the model with two explanatory variables and interaction

The power functions of Model (c), $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n1} x_{n2} + e_n$, do not have much more new information. The corresponding Figures 17 - 20 can be found in the appendix. These power functions are similar to the ones for Models (a) and (b): the naive methods and ordering the regressors according to the values of the euclidean norm perform rather bad, whereas the distance based methods yield very good power functions, especially the exact shortest path algorithm.



(a) Ordering the regressors with the help of the nearest neighbor algorithm

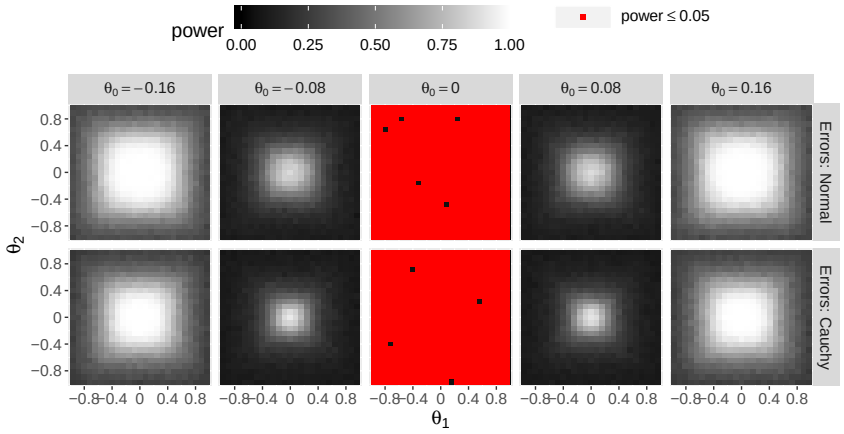


(b) Ordering the regressors with the help of the exact shortest path algorithm

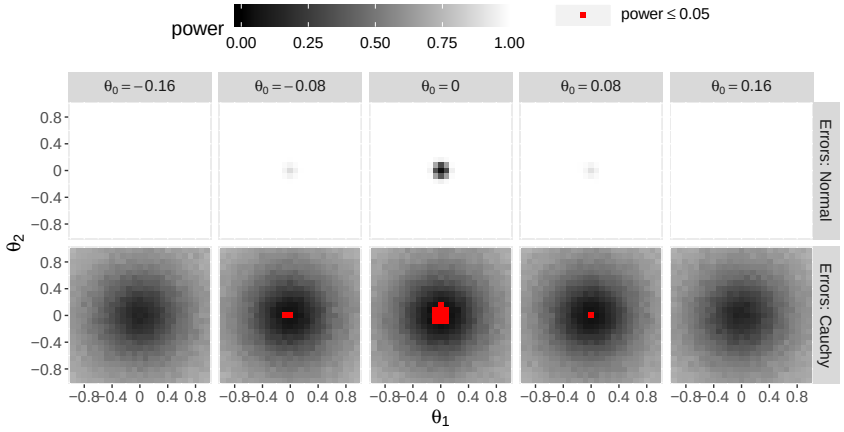
Fig. 13 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (b): $y_n = \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + e_n$. Here, the depth tests with $K = 3$ and ordering the regressors on the basis of their pairwise distances are applied. The areas, where the power is smaller than α, are marked in red

4.5 Comparison of the models with two explanatory variables

The main difference between Model (b) and (c) is that all methods except the F-test with normally distributed errors have worse power for Model (c) than for Model (b). In particular, the power is much worse for the naive methods, the method based on the euclidean norm and the classical sign test. This can be also seen from Figure 15.



(a) The sign test



(b) The F-test

Fig. 14 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (b): $y_n = \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + e_n$. Here, the reference methods, classical sign test and F-test are applied. The areas, where the power is smaller than α, are marked in red

To compare the power functions of the different ordering methods more easily, Figure 15 shows the percentage of grid points in the area of the alternative where the power values are larger than a certain threshold. The larger the value is, the better performs the method because power functions should converge to one as quick as possible in the alternative. The left column of Figure 15 shows the percentage where the power values are at least 0.05, i.e. the level of the test. In this case, it would be desirable that all values are one because this would mean that the corresponding test has a power greater than α for every $\boldsymbol{\theta}$ of the alternative. But, as it can be seen, especially the values of the naive methods including the euclidean norm and the values of the classical sign test are smaller. The percentage of

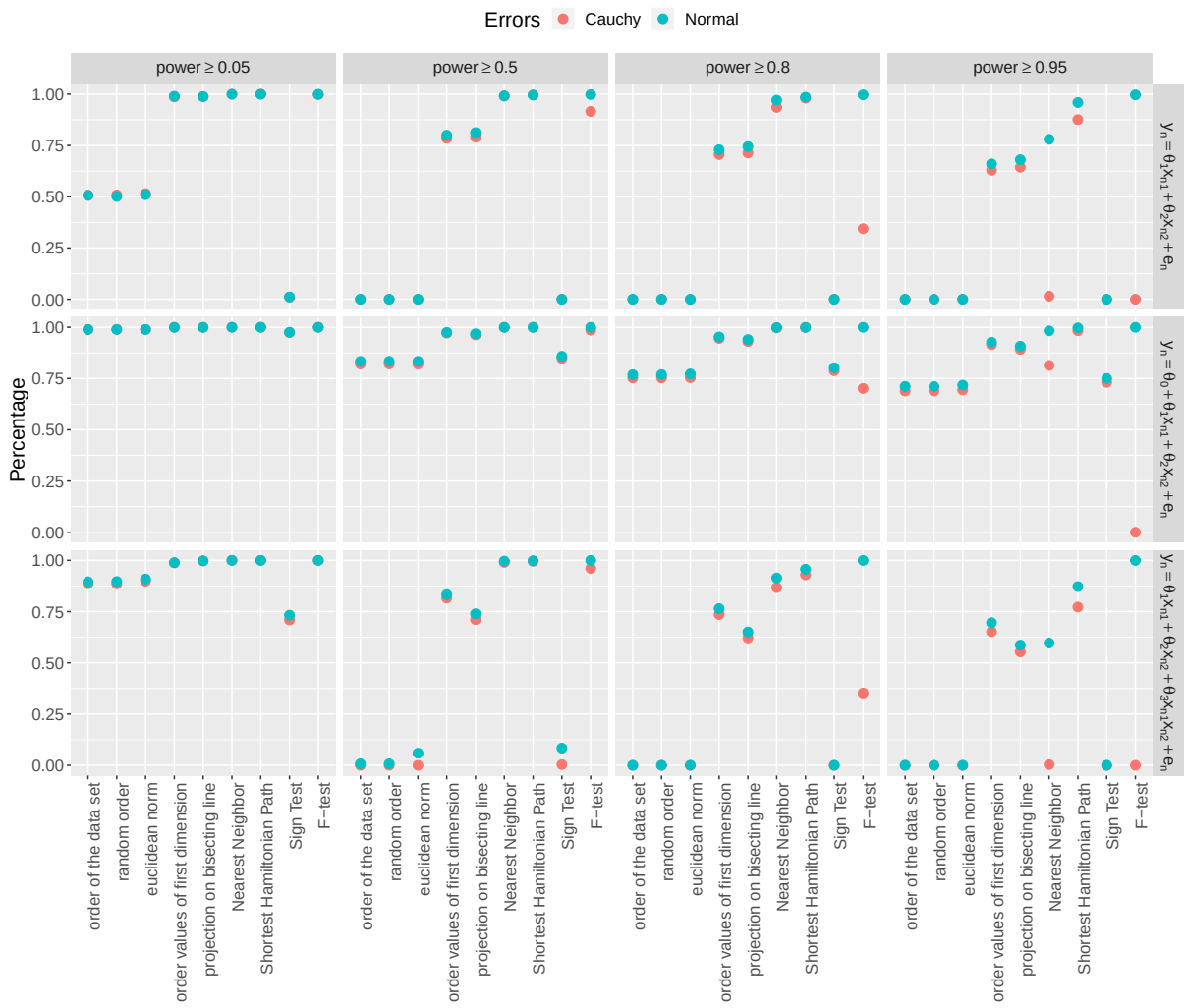


Fig. 15 Percentage of the power values which are at least 0.05, 0.5, 0.8 and 0.95 for the Models (a), (b), and (c) using 3-sign depth for the depth tests. If the percentages for Cauchy and normally distributed errors are the same, then only the symbol for the normal distribution, i.e. the green symbol, is shown. For every component of $\boldsymbol{\theta}$, the interval of $[-2, 2]$ with 51 equidistant grid points is considered

power values which is larger than 0.5 is nearly zero for these methods in Models (a) and (c), whereas the distance based methods still have values close to one. Although the values of the projecting methods decrease, they are still relatively large. The same results hold for the percentage of the power functions which have values at least 0.8. However, in this column of the figure, a large difference of the values of the F-test for normally and Cauchy distributed errors is considerable. Most of the red points in Figure

15 for Cauchy distribution have (nearly) the same value as the green points for normal distribution and thus are invisible. Although the sign depth tests with Cauchy distributed errors never performs better than for normally distributed errors, in most cases the difference is very small. This is different to the F-test, which has best performance for normally distributed errors and very bad power values for Cauchy distributed errors. The percentage of power values for the F-test which is larger or equal than 0.95 is zero for Cauchy distributed errors and one for normally distributed errors. In the most right column of Figure 15 also the performance difference of the two distance based methods can be seen. The exact shortest path algorithm has the largest values over all methods except the F-test with normally distributed errors, whereas the approximative nearest neighbor algorithm has lower values, especially for Cauchy distributed errors. It can also be seen that often the values for the Cauchy distribution are slightly worse than the values for the normal distribution. This is caused by the larger tails of the Cauchy distribution. The larger tails provide larger deviations from the values of the null hypothesis, so that the signs of the residuals change only for more extreme values of the alternative.

As an overall result, it can be said that the distance based methods perform best independent of the distribution of the errors. When the data set is small enough for taking the exact algorithm for the Shortest Hamiltonian Path, this should be done because its performance is better than the performance of the approximative algorithm.

4.6 Comparison of the models with two, three, and four explanatory variables

Finally, we compare the power of the depth tests based on 3-sign depth and 4-sign depth and the distance based orderings with the F-test for the Models (a), (d), and (e) which contain two, three and four explanatory variables. Figure 16 shows the percentage of power values which are at least 0.05, 0.5, 0.8 and 0.95. The five tests differ not very much in the percentage of power values larger than 0.05 and 0.5. Clear differences appear only for the percentage of power values larger than 0.8 and 0.95. In particular, the depth tests based on the nearest neighbor ordering behave worse than the depth tests using the exact Shortest Path algorithm. However, the depth tests based on the exact Shortest Path algorithm cannot reach the power of the F-test any more for normally distributed errors in the model with four explanatory variables. This happens for the depth test based on the 3-sign depth already in the model with three explanatory variables. Hence the conjecture is that a depth test based on the 5-sign depth should be used for the model with four explanatory variables to get a similar power as the F-test for normally distributed errors. More generally: a depth test based on $(K+1)$-sign depth should be used for the model with K explanatory variables.

5 Conclusion and Outlook

In this paper, we presented an extension of the sign depth test based on K-sign depth presented by Leckey et al. (2019) for multiple regression. The sign depth test depends on an ordering of the regressors in the data set. We studied different possibilities to order the regressors in multiple regression. Besides naive ordering methods like taking the order of the data set, we considered orders based on scalarization of the regressors and distance based orders of the regressors. We performed a large simulation study on artificial data with two, three, and four explanatory variables and five different linear models and compared the sign depth tests with different ordering methods. This is done mainly using the 3-sign depth and also occasionally using the 4-sign depth. Additionally, we compared the sign depth test with two classical

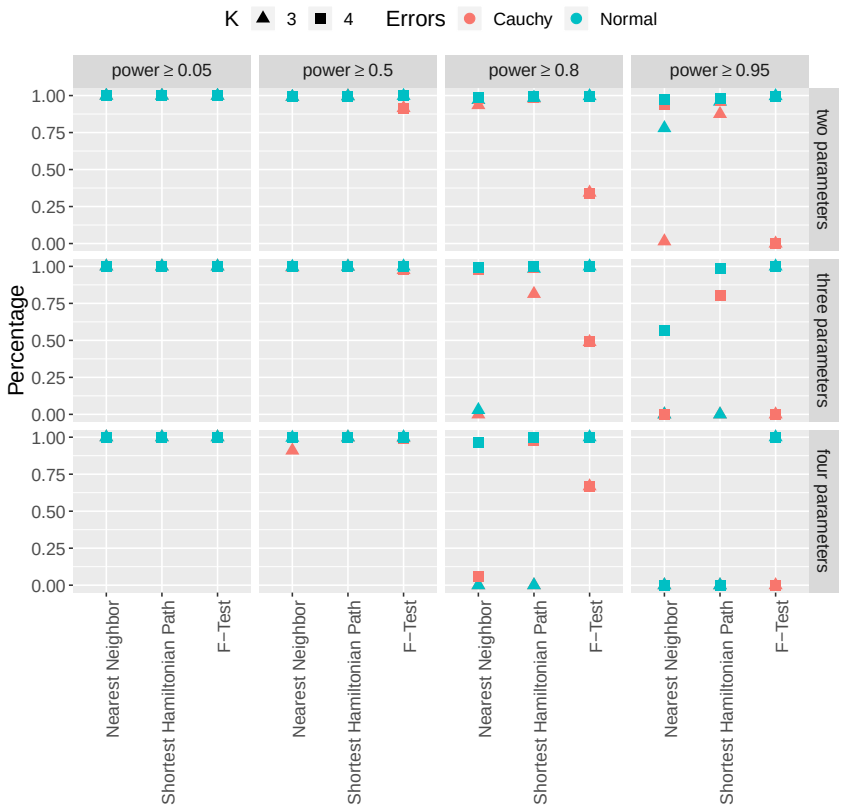


Fig. 16 Percentage of the power values which are larger than 0.05, 0.5, 0.8 and 0.95 for the Models (a), (d), and (e) using 3-sign depth and 4-sign depth for the depth tests with distance based ordering. If the percentages for Cauchy and normally distributed errors are the same, then only the symbol for the normal distribution, i.e. the green symbol, is shown. For $\boldsymbol{\theta}$ of the alternative, grids from $[-2, 2]^k$ with 51 points for $k = 2$ and $k = 3$ and 19 points for $k = 4$ in every dimension are used

tests for regression parameters, the classical sign test and the F-test. The generated simulated power functions showed that ordering on the basis of the pairwise distances of the regressors performs best. The idea behind this ordering is to find a shortest path through all data points. Because this so-called *Shortest Hamiltonian Path* problem is a transformation of the well known Traveling Salesman Problem, the exact algorithm for finding this shortest path has exponentially runtime in the worst case. For this reason, also an approximative algorithm has been used in the simulation study. The results in Section 4 and especially Figures 15 and 16 on Pages 21 and 23, respectively, showed that the exact algorithm performs better than the approximative one and should be used when the data set is not too large. But both distance based methods performed better than the other considered methods, including the classical methods, when the errors of the underlying model are not normally distributed. For normally distributed errors, the F-test of course is best, but the distance based methods have nearly the same power if the

number of explanatory variables is two or three. For a larger number of explanatory variables, it seems that a larger K for the K-sign depth should be used. For that, an algorithm for calculating K-sign depth in linear time, which is in progress, would be useful. Nevertheless, we can conclude here already that the sign depth test is a good choice for testing parameters of regression models and can now also be used for multiple regression.

Of course, in the field of sign depth tests for multiple regression, further research has to be done. In this paper, only artificial data sets with random values for the explanatory variables have been used. In the next step, real data sets including categorical data could be regarded. Therefore, a reasonable distance measure between categorical data and numerical data has to be used for our proposed distance based methods. Also, the influence of the parameter K of the sign depth test can be regarded. In this paper, only $K = 3$ and $K = 4$ have been used, but larger values for K should be regarded as well. Furthermore, our proposed method is not limited to linear models, but could be also used for non-linear models or time series models.

Acknowledgments

The authors gratefully acknowledge support from the Collaborative Research Center "Statistical Modelling of Nonlinear Dynamic Processes" (SFB 823, B5) of the German Research Foundation (DFG).

A Appendix

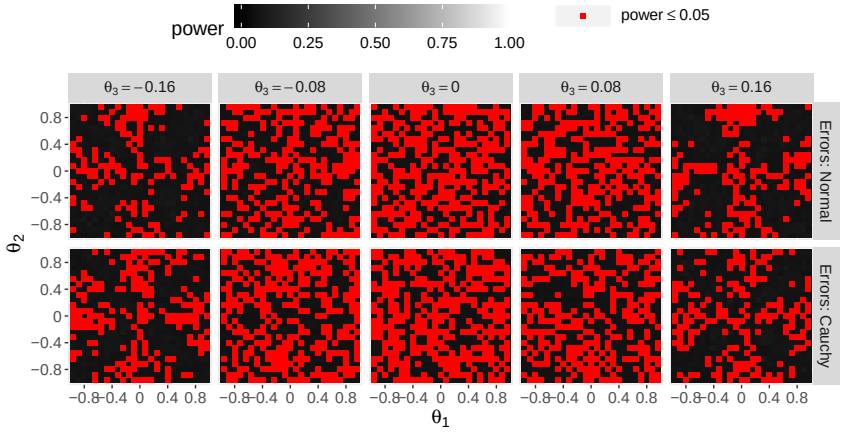
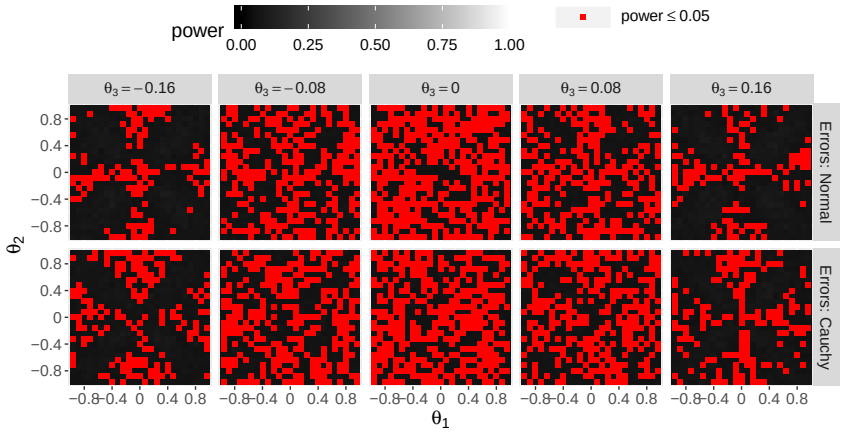
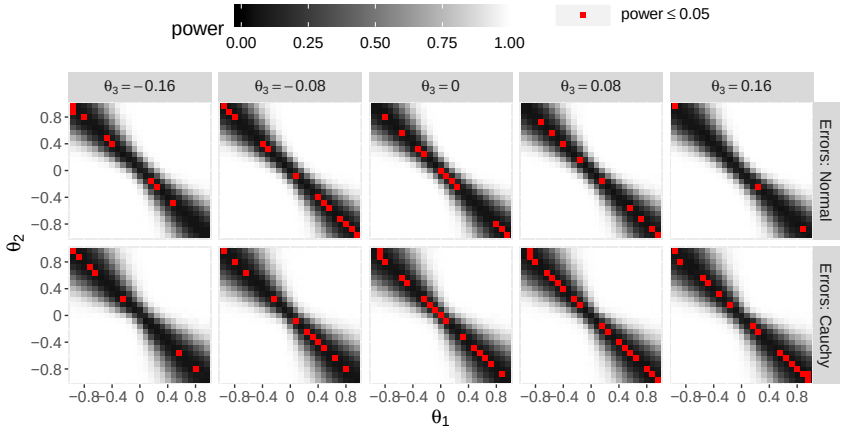


Fig. 17 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (c): $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n1} x_{n2} + e_n$. Here, the depth test with $K = 3$ and ordering the regressors according to their appearance in the data set is applied. The areas, where the power is smaller than α, are marked in red

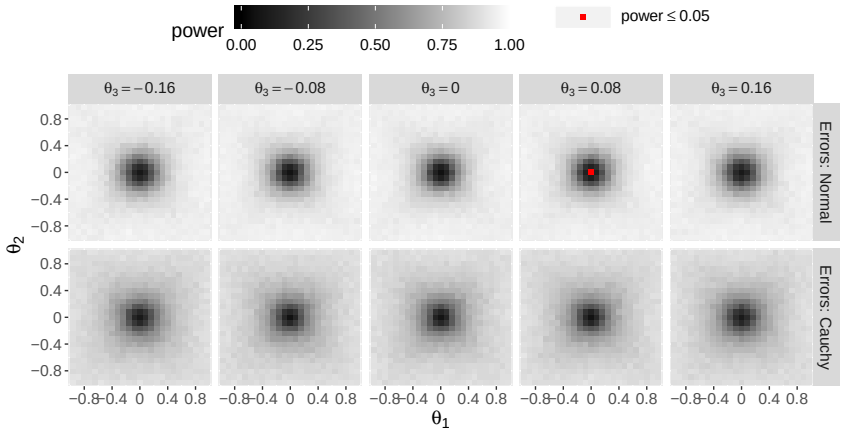


(a) Ordering the regressors according to their values of the euclidean norm

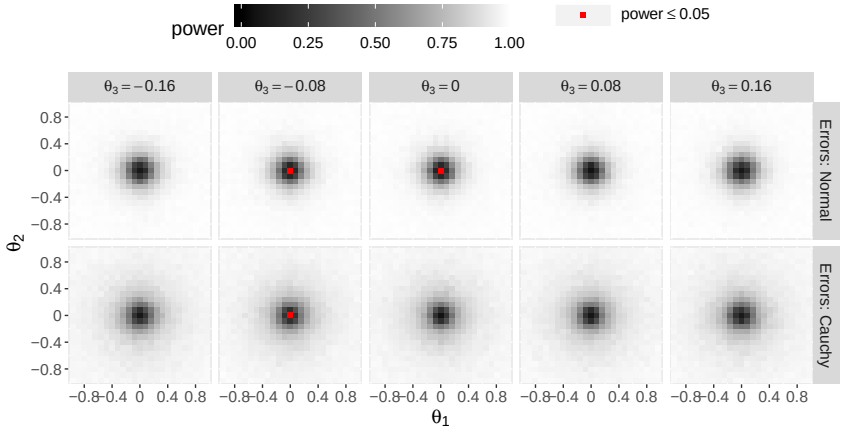


(b) Ordering the regressors according to their projected values on the bisecting line

Fig. 18 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (c): $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n1} x_{n2} + e_n$. Here, the depth tests with $K = 3$ and ordering the regressors according to scalarizations are applied. The areas, where the power is smaller than α, are marked in red

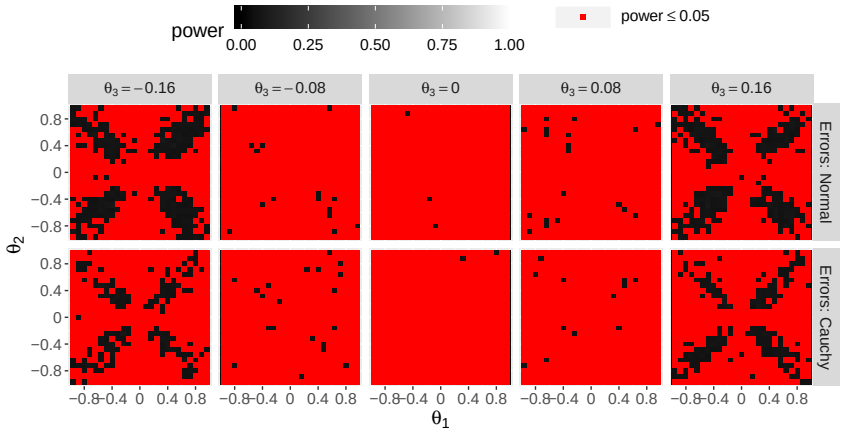


(a) Ordering the regressors with the help of the nearest neighbor algorithm

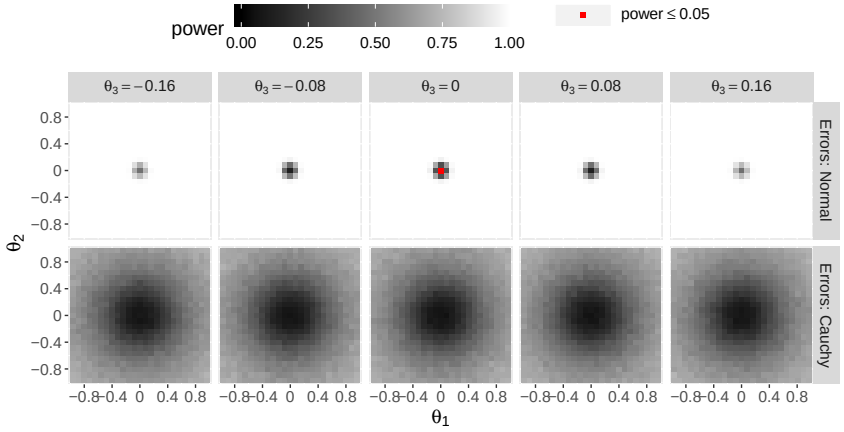


(b) Ordering the regressors with the help of the exact shortest path algorithm

Fig. 19 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (c): $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n1} x_{n2} + e_n$. Here, the depth tests with $K = 3$ and ordering the regressors on the basis of their pairwise distances are applied. The areas, where the power is smaller than α, are marked in red



(a) The sign test



(b) The F-test

Fig. 20 Simulated power functions of the test hypotheses for $H_0 : \boldsymbol{\theta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\theta} \neq \mathbf{0}$ to the level $\alpha = 0.05$ in Model (c): $y_n = \theta_1 x_{n1} + \theta_2 x_{n2} + \theta_3 x_{n1} x_{n2} + e_n$. Here, the reference methods, the classical sign test and the F-test, are applied. The areas, where the power is smaller than α, are marked in red

References

Adler D, Murdoch D, others (2018) rgl: 3D Visualization Using OpenGL. URL `https://CRAN.R-project.org/package=rgl`, R package version 0.99.16

Agostinelli C (2018) Local half-region depth for functional data. Journal of Multivariate Analysis 163:67–79

Applegate D, Bixby R, Chvátal V, Cook W (2004) Concorde. URL `www.tsp.gatech.edu`

Applegate D, Bixby R, Chvátal V, Cook W (2006) The Traveling Salesman Problem. Applied Mathematics, Princton University Press

Auguie B (2017) gridExtra: Miscellaneous Functions for "Grid" Graphics. URL `https://CRAN.R-project.org/package=gridExtra`, R package version 2.3

Carnell R (2019) lhs: Latin Hypercube Samples. URL `https://CRAN.R-project.org/package=lhs`, R package version 1.0.1

Claeskens G, Hubert M, Slaets L, Vakili K (2014) Multivariate functional halfspace depth. Journal of the American Statistical Association 109(505):411–423

Clarke B (2018) Robustness Theory and Application. Wiley Series in Probability and Statistics, Wiley

Durstenfeld R (1964) Algorithm 235: Random Permutation. Commun ACM 7(7):420

Hahsler M, Hornik K (2017) TSP: Traveling Salesperson Problem (TSP). URL `https://CRAN.R-project.org/package=TSP`, R package version 1.1-5

Hampel F, Ronchetti E, Rousseeuw P, Stahel W (2011) Robust Statistics: The Approach Based on Influence Functions. Wiley Series in Probability and Statistics, Wiley

Horn M (2019) GSignTest: Robust Tests for Regression-Parameters via Sign Depth. URL `https://github.com/melaniehorn/GSignTest`, R package version 1.0.3

Hu Y, Wang Y, Wu Y, Li Q, Hou C (2011) Generalized mahalanobis depth in the reproducing kernel Hilbert space. Statistical Papers 52(3):511–522

Huber P (2004) Robust Statistics. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series, Wiley

Jurečková J, Picek J (2005) Robust Statistical Methods with R. CRC Press

Jurečková J, Sen P (1996) Robust Statistical Procedures: Asymptotics and Interrelations. A Wiley Interscience publication, Wiley

Knuth DE (1998) The Art of Computer Programming, Volume 3: (2Nd Ed.) Sorting and Searching. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA

Kustosz C, Leucht A, Müller CH (2016a) Tests based on simplicial depth for AR(1) models with explosion. Journal of Time Series Analysis

Kustosz C, Müller CH, Wendler M (2016b) Simplified simplicial depth for regression and autoregressive growth processes. Journal of Statistical Planning and Inference 173:125–146

Lawler E, Lenstra J, Rinnooy Kan A, Shmoys D (1985) The Travelling Salesman Problem. Wiley-Interscience series in discrete mathematics and optimization, John Wiley & sons

Leckey K, Malcherczyk D, Müller CH (2019) Powerful generalized sign tests based on sign depth. (Submitted)

Leisch F, Dimitriadou E (2010) mlbench: Machine Learning Benchmark Problems. R package version 2.1-1

Liu R (1990) On a notion of data depth based on random simplices. Annals of Statistics 18:405–414

López-Pintado S, Romo J (2009) On the concept of depth for functional data. Journal of the American Statistical Association 104(486):718–734

Maronna R, Martin D, Yohai V (2006) Robust Statistics: Theory and Methods. Wiley Series in Probability and Statistics, Wiley

Mosler K (2002) Multivariate dispersion, central regions and depth. The lift zonoid approach., vol 165. New York, NY: Springer

Müller CH (1997) Robust Planning and Analysis of Experiments, vol 124. Berlin: Springer

Müller CH (2005) Depth estimators and tests based on the likelihood principle with application to regression. Journal of Multivariate Analysis 95:153–181

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`

Rieder H (1994) Robust asymptotic statistics. New York, NY: Springer-Verlag

Rosenkrantz D, Stearns R, Lewis II P (1977) An Analysis of Several Heuristics for the Traveling Salesman Problem. SIAM Journal on Computing 6(3):563–581

Rousseeuw P, Hubert M (1999) Regression depth. Journal of the American Statistical Association 94:388–433

Rousseeuw P, Leroy A (2003) Robust Regression and Outlier Detection. Wiley Series in Probability and Statistics, Wiley

Stein M (1987) Large Sample Properties of Simulations Using Latin Hypercube Sampling. Technometrics 29:143–151

Tukey JW (1975) Mathematics and the picturing of data. Proc. int. Congr. Math., Vancouver 1974, Vol. 2, 523-531 (1975)

Wickham H (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, URL `http://ggplot2.org`