Dissertation

zur Erlangung des akademischen Grades

eines Doktors der Naturwissenschaften (rer. nat.)

der Fakultät Statistik der Technischen Universität Dortmund

Depth based estimators and tests for autoregressive processes with application on crack growth and oil prices

by

Dipl.-Math. Christoph Falkenau

Vorgelegt: Dortmund, 25. Februar 2016

Tag der mündlichen Prüfung: 20. Juli 2016

Erstgutachter: Prof. Dr. Christine Müller

Zweitgutachter: Prof. Dr. Roland Fried

Kommissionsvorsitz: Prof. Dr. Jörg Rahnenführer

Acknowledgement

I gratefully thank the Prof. Dr. Reinhard Maurer and Dipl.-Ing. Guido Heeke for providing the data for this thesis.

The thesis was partially supported by the DFG Special Collaborative Centre SFB 823 'Statistical Modelling of Nonlinear Dynamic Processes' (Project B5).

This thesis was written and defended previous to my wedding, when I was named Christoph Kustosz. Since September, 3rd 2016 my name changed to Christoph Falkenau, under which the thesis now has been published.

Contents

1 Introduction

This thesis introduces new statistical methods for a class of autoregressive (AR) processes based on data depth. In particular, we introduce methods to allow statistical inference for explosive and non-linear AR processes. The motivation for our proposals is due to experiments of Maurer and Heeke (2010). The main aim of these experiments is a fundamental research of the properties of crack growth in prestressed concrete under low loading. By application of a physical formula, the class of non-linear AR processes given by

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + \theta_3 + E_n$$

defines a reasonable model choice. However, this full model, applied to our data, leads to identification problems. Hence, we consider models given by

$$\begin{aligned} Y_n &= Y_{n-1} + \theta_1 Y_{n-1} + E_n, \\ Y_n &= Y_{n-1} + \theta_1 Y_{n-1} + \theta_3 + E_n, \\ Y_n &= Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + E_n, \end{aligned}$$

here. The experiments imply specific properties on the processes and errors which cannot be covered by standard assumptions. As a consequence, classical AR estimators and tests cannot be applied directly. Therefore, we propose methods which can deal with this specific properties based on data depth. These properties are summarised as follows:

- The observed processes are explosive.
- The observed processes have upward jumps.
- The underlying processes are inhomogeneous in term of parameter changes.
- The observed processes are potentially right-censored.

In particular, we derive outlier robust estimators, tests and confidence sets for the parameters of the underlying crack growth process. Further, we define a method to analyse the changes in the parameters of this process which allows us to propose change point detection algorithms. To allow a calculation of the uncertainty of the failure times, we define a prediction method for the future development of the observed processes. Further, we introduce confidence sets for the values of the process at specific times and at different stress levels to extrapolate the censored

series. In combination, this allows us to calculate a S-N curve, relating the lifetime and stress, based on the available data. We also discuss the efficient implementation of the proposed methods and illustrate their advantages in comparison with standard methods by simulation studies.
To illustrate further fields of application, we also discuss a change point detection for oil prices which are typically modelled by autoregressive processes with parameters close to the unit-root case.

Specific Features of Crack Growth in Prestressed Concrete

The data from the Maurer and Heeke (2010) experiments inherits features which should be taken into account in a statistical analysis. These features can be seen in Figure 1 where the data from one of the experiments is depicted. The first feature

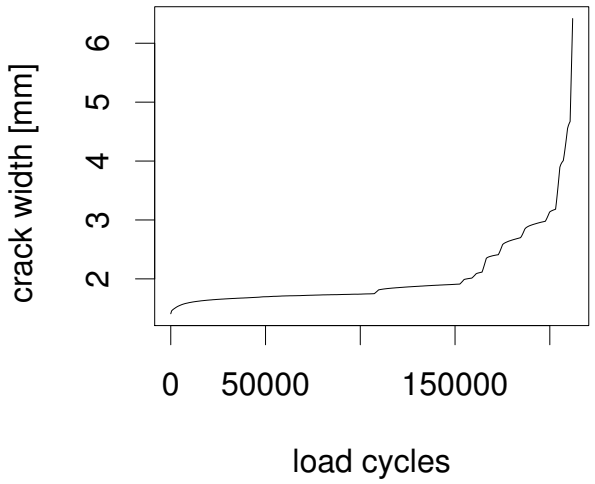


Fig. 1: Crack width in [mm] from the TR02 experiment by Maurer and Heeke (2010).

we observe is that the process shows explosive or exponential growth. This is not surprising, since we observe crack growth under cyclic loading without any maintenance. Hence the crack monotonically grows in time. Further, by the accumulation of damage, the speed of growth increases, too. Statistically, this leads to our assumption of explosive processes to model crack growth.
The second feature are jumps in the process. Since the complete tension wire, incorporated into the beam, consists of 35 singular twisted wires, we can see a jumping behaviour in the crack growth process every time one of the single wires breaks.

Typically, we do not observe a failure of all wires until total failure of the beam. However, some of these wires break and lead to upward jumps of the crack width process, since the beam loses a large amount of stability when such a wire breaks. Statistically, we treat these jumps as outliers, since a direct modelling of the occurrence times and jump heights is difficult for approximately eight to fifteen jumps in each experiment and five available experiments under non constant parameters only. These outliers are strictly positive and hence lead to asymmetric error assumptions, since the observed jumps only lead to increasing crack widths.
For the next feature, we consider the process without jumps, as depicted in Figure 2. We generated this process by consideration of the increments and manually removing the increments around points in time when tension wires break. These times are known, since they can be verified by measurements of microphones. In

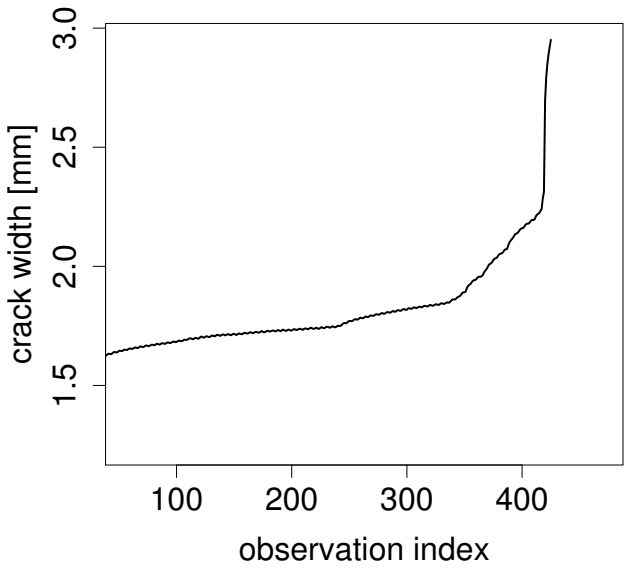


Fig. 2: Adjusted crack width in [mm] from the TR02 experiment by Maurer and Heeke (2010). We subtracted the jumps by clearing the data at the recorded jump times.

the adjusted crack width series we see that the process dynamics, in particular the growth parameter, changes within the experiment. This means that we have to deal with different phases of the dynamics in the underlying process. Hence, we also need to assure that the proposed models and methods enable us to define a change point detection algorithm to deal with the observed crack width series.
The fourth feature of the crack growth series just applies censored series. Since the experiments are quite long, and hence expensive, we need to deal with censored

series. In censored cases the experiments were aborted before total failure could be observed. With the final aim to derive the relation between the applied loads and lifetime, we need to predict the failure times for these censored experiments. Unfortunately, censoring and low-loading is highly correlated, since low loads lead to slow crack growth what induces long experiments. Hence, this problem will appear more often, if we conduct experiments in the interesting load range. This means that we need a model and statistical methods which allow a prediction for censored series. Quite related to this task, we also have to derive methods to measure the uncertainty of the failure time of the fully observed series, to generate confidence intervals for the arrival times at critical crack widths.
The features are summarised as follows:

- Explosive processes.
- Upward jumps.
- Parameter changes.
- Potential right censoring.

The thesis will discuss models and statistical methods to allow inference under these conditions.

Outline

Section 2 deals with the statistical methods which we derive in this thesis. We start with the formulation of the applied models by a short review of the engineering literature considering crack growth. Then, we give an overview about robust methods and robust regression and define test statistics for our models, based on data depth. For these statistics, we derive asymptotic distributions and propose tests as well as parameter confidence intervals. Finally, we present results on the consistency of the proposed tests. The implementation and examples of the introduced methods are shown in Section 4. We give details on the implementation of the test statistics and introduce algorithms to calculate confidence regions and estimates. Further, we propose approaches to detect parameter changes in the considered models and a prediction scheme for one of our models. To illustrate the performance of our methods, we afterwards present comparative simulation studies for the proposed methods in Section 5. The final analysis of the available crack growth series is presented in Section 6. We first show some results from the application of an unrestricted version of a crack growth model and discuss the limitations of these results. Then, we derive

a curve for the relationship of applied load and the time of failure for prestressed concrete by the application of the proposed methods. Since the theoretical results are not limited to crack growth series, we present another real world application in which our proposals can be applied. In particular, we give another example of the proposed change point detection method on oil price series. Section 7 gives an outlook and presents some further research topics related to this thesis.
In the Appendix, we present additional simulation results, poofs and a package which makes the proposed methods applicable in the programming language `R`, see RCoreTeam (2015). This package is available for installation via the online repository *github* at `https://github.com/ChrisKust/rexpar.git`.

2 Literature Overview

In this Section, we define the main statistical models and test statistics. Then, we derive some central properties to propose tests which can be applied to define phase change detection and prediction methods for growth processes.

2.1 Modelling of Crack Growth

In literature, various approaches to describe fatigue stochastically have been discussed. Thereby, different general ideas are applied to introduce physical assumptions and uncertainty. In this thesis, we focus on methods which aim to describe the stochastic crack length or width process directly. The basic formula for a large group of models was introduced by Paris and Erdogan (1963) who describe crack growth under cyclic loading by

$$\frac{da}{dN} = C \cdot \Delta K^m, \tag{1}$$

whereby a denotes the crack length or width, N is the number of load cycles and ΔK is the load intensity factor. C and m are constants which are related to the experimental setting and material. The load intensity factor is given by $\Delta K = \Delta\sigma \cdot G \cdot \sqrt{\pi \cdot a}$, where G is a geometry parameter and $\Delta\sigma$ is the actual load intensity applied in the experiment. This equation is called **Paris-Erdogan equation** or **Paris-Law**. Intuitively, it states that the rate of crack growth depends on the actual crack size and the experimental constants. Hence, a crack is assumed to grow faster, if it is large. This, in particular, is in line with the observed behaviour in the Maurer and Heeke (2010) experiments, described in the previous section. Since we are interested in the statistical properties of the model, rather than in a physical interpretation, it is convenient to translate (1) to a time dependent relation and to reduce the parameters. This, in particular, avoids identification problems, if all parameters are unknown. Hence, we rewrite (1) to

$$\frac{da}{dN} = C \cdot \left(\Delta\sigma \cdot G \cdot \sqrt{\pi \cdot a}\right)^m = \tilde{C} \cdot a^{m/2} = \tilde{C} \cdot a^{C_2}, \tag{2}$$

with $\tilde{C} = C \cdot (\Delta\sigma G\sqrt{\pi})^m$ and $C_2 = \frac{m}{2}$. Further, we want to relate the crack growth to time instead of load cycles. Since we observe the load cycles in a constant frequency

and amplitide, we can adjust (2) by multiplication with a constant and get

$$\frac{da(t)}{dt} = C_1 \cdot a(t)^{C_2}. \tag{3}$$

In the modified Paris-Erdogan equation (3), crack growth is purely deterministic. Due to microstructures and technical limitations, it is more convincing to consider a stochastic component in the model. Therefore, we switch to a more general formulation to describe a random crack size $A(t)$ in time t by

$$\frac{dA(t)}{dt} = F(A(t), X(t)), \tag{4}$$

whereby $X(t)$ is a stochastic process and $F(\cdot, \cdot)$ is a deterministic function setting the stochastic process and the crack growth process in relation. This general formulation was, for example, proposed by Chiquet et al. (2009).
In literature, models defined by (4) differ in several ways. The function F can be varied and the process $X(t)$ can be defined differently.
A direct combination of the process and a stochastic component was proposed by Sobczyk (1987). Here, instead of a deterministic function, F is a functional. The process is defined by

$$\frac{dA(t)}{dt} = J(t)dN(t), \tag{5}$$

whereby $N(t)$ is a Poisson process and $J(t)$ can be reduced to $(J_n)_{n\in\mathbb{N}}$, a series of independent and identically distributed non-negative random variables, describing random jump heights at the points in time t_n, where the Poisson process jumps. By solving (5), we get $A(t) = \sum_{n=1}^{N(t)} J_n$. If the involved jump-height distributions are known, the density of the arrival times at fixed crack lengths can be derived to analyse the process in more detail, including estimation and prediction.
Often F is related to Paris-Erdogan type functions. Then, $X(t)$ is defined by processes which allow a solution of the arising Stochastic Differential Equation (SDE). Chiquet et al. (2009) apply their general proposal (4) by

$$\frac{dA(t)}{dt} = C_1 \cdot A(t)^{C_2} \cdot X(t),$$

whereby they limit $X(t)$ to a Markov process with finite state space. Based on this limitation, they derive a transformation which allows the estimation of the parameters.

Sobczyk (1987) also presents a model based on the modified Paris-Erdogan equation. Thereby, he uses a deterministic function with multiplicative errors and defines

$$\frac{dA(t)}{dt} = \tilde{F}(A(t)) \cdot X(t).$$

Reduced to $\tilde{F}(A(t)) = C_1 A(t)^{C_2}$, different approaches for $X(t)$ were discussed in literature. Spencer et al. (1989) define $X(t) = \exp(Z(t))$, whereby $\frac{dZ(t)}{dt} = -\xi Z(t) + W(t)$ is a white noise with drift. Sobczyk (1979) proposes $X(t)$ to be modelled as purely white noise and Lin et al. (1985) propose $X(t)$ to be defined by a Poisson process with random jump heights. The model of Yang and Manning (1990) proposes a very simple formulation of $X(t)$, assuming $X(t)$ to be defined by Log-Normal random variables. Their approach is used by Wu et al. (2001) to solve the resulting model and to allow the modelling of the probability for exceedances of critical crack sizes. Zio and Zoia (2009) also assume a Log-Normal process for $X(t)$ and propose an estimation based on Bayes methods.
By allowing $X(t)$ to be the sum of a deterministic process and a purely stochastic part, Sobczyk (1987) specifies (4) to

$$\frac{dA(t)}{dt} = C_1 \cdot A(t)^{C_2} + C_3 \cdot A(t)^{C_2} \cdot X(t). \tag{6}$$

Since we can modify $X(t)$ by more complicated deterministic functions, (6) can be expressed more generally by

$$\frac{dA(t)}{dt} = F_1(A(t)) + F_2(A(t)) \cdot X(t)$$

with two deterministic functions F_1 and F_2. Snaidy et al. (1998) combine the Poisson random driver with the linear extension of $X(t)$ and propose to apply (6) with

$$X(t) = \int_0^t M(x) e^{-\beta(t-x)} dN(x),$$

whereby $M(x) > 0$ is a function defining the amplitude of the error process and $\beta \in \mathbb{R}$ is an additional decay parameter. Snaidy et al. (1998) derive solutions of the respective SDE for specific choices of C_2 .

2.2 A Modification of the Snaidy et al. (1998) Model

In our approach, we also start with the modified Paris-Law (3). Similar to Snaidy et al. (1998), we extend the Law to

$$\frac{dA(t)}{dt} = \alpha_1 A(t)^{\alpha_2} + \frac{dV(t)}{dt}, \tag{7}$$

whereby $V(t)$ is a stochastic error process and $\alpha_1, \alpha_2 > 0$. In comparison to the formulation of Snaidy et al. (1998), we do not model an influence of the actual crack size on the uncertainty of the error process. Further, we express the error process by its increments, making a discretisation more feasible. By assuming that $V(t)$ is a Brownian motion, the analysis of this model can be performed by known methods from stochastic analysis, when for example $\alpha_2 = 1$. For more complicated settings, assumptions on the error process can also lead to theoretically and practically feasible processes. The properties of crack growth in prestressed concrete, as discussed in the previous section, make it necessary to restrict the growth process to be increasing and to allow $V(t)$ to include jumps or at least to be defined by a skewed distribution. This contradicts the assumption of normally distributed errors and Brownian error processes. Through the approximation of the process by the Euler-Maruyama scheme, see Kloeden and Platen (1992), the problem translates to a non-linear discrete time autoregressive (AR) process Y_n.

Lemma 1. *Let $t_n = t_0 + nh, n \in \mathbb{N}$ be an equidistant partition of time. If $V(t)$ has independent and identically distributed increments, the Euler-Maruyama approximation of a stochastic process defined by (7) is given by*

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + \theta_3 + E_n,$$

with $Y_n = A(t_n)$, $\theta_1 = h\alpha_1$, $\theta_2 = \alpha_2$ and $\theta_3 = med(V(t+h) - V(t))$. Further, $med(E_n) = 0$ holds.

Proof. By the Euler-Maruyama approximation of (7) we get

$$\begin{aligned}
&\frac{A(t+h)-A(t)}{h} \approx \alpha_1 A(t)^{\alpha_2} + \frac{V(t+h)-V(t)}{h} \\
\Leftrightarrow &A(t+h)-A(t) \approx h\alpha_1 A(t)^{\alpha_2} + V(t+h) - V(t) \\
\Leftrightarrow &A(t+h) \approx h\alpha_1 A(t)^{\alpha_2} + A(t) + V(t+h) - V(t) \\
\Leftrightarrow &A(t+h) \approx \theta_1 A(t)^{\theta_2} + A(t) + \tilde{E}(t+h) \\
\overset{t=t_0+nh}{\Leftrightarrow} &Y_{n+1} \approx \theta_1 Y_n^{\theta_2} + Y_n + \tilde{E}_{n+1} \\
\Leftrightarrow &Y_{n+1} \approx \theta_1 Y_n^{\theta_2} + Y_n + med(\tilde{E}_{n+1}) + \tilde{E}_{n+1} - med(\tilde{E}_{n+1}) \\
\Leftrightarrow &Y_{n+1} \approx \theta_1 Y_n^{\theta_2} + Y_n + \theta_3 + E_{n+1},
\end{aligned}$$

if h is fixed and the process is observed in equidistant times $t_n = t_0 + nh$ for $n \in \mathbb{N}$, $t_0 \in \mathbb{R}$. □

Remark 2. *Note that the Euler-Maruyama approximation is a simple and potentially biased approximation method. However, this thesis will focus on the resulting stochastic model to derive statistical methods for the class of resulting processes. More adequate approximations, like the Milstein scheme, sampling methods or local linearisation methods lead to more difficult stochastic models which are not considered in this thesis but could be an issue for further research.*

The assumptions on crack growth processes imply that $Y_0 = y_0 > 0$ for a non-random initial crack length and that $Y_n > Y_m$ for $n > m$. We further assume that $\tilde{E}_n \geq 0$ holds. To have $med(E_n) = 0$, we centre the errors. The $med(E_n) = 0$ assumption later will be a major key to allow robust estimation and asymptotic results for statistics based on our models. To guarantee a growth process including these errors, we assume errors which are bounded on their negative side. Otherwise the growth assumption can be violated.

Lemma 3. *Assume the conditions of Lemma 1, $\theta_1 > 0, \theta_2 > 1$. $Y_n > Y_{n-1}$ for all $n \in \mathbb{N}$ i.e. the process is strictly increasing, if and only if*

$$E_n > -\theta_1 y_0^{\theta_2} - \theta_3 \ for\ all\ n \in \mathbb{N}.$$

Proof. Solving $Y_n - Y_{n-1} > 0$ delivers

$$E_n > -\theta_1 Y_n^{\theta_2} - \theta_3.$$

Since we have $Y_n > Y_{n-1}$ and $\theta_1 > 0, \theta_2 > 1$, this implies

$$E_n > -\theta_1 y_0^{\theta_2} - \theta_3 \ \forall \ n \in \mathbb{N}.$$

For the converse consider

$$E_n > -\theta_1 y_0^{\theta_2} - \theta_3 \ \forall \ n \in \mathbb{N}.$$

Then

$$Y_1 - Y_0 = \theta_1 y_0^{\theta_2} + \theta_3 + E_1 > 0.$$

Hence, $Y_1 > Y_0$. We can assume that $Y_n > Y_{n-1}$ for $n \leq N$ and fixed and known N. Then, by induction

$$Y_{N+1} - Y_N = \theta_1 Y_N^{\theta_2} + \theta_3 + E_{N+1} \overset{ind.}{>} \theta_1 y_0^{\theta_2} + \theta_3 + E_{N+1} > 0.$$

□

In this thesis, we analyse some partial models with one and two parameters defined by

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1} + E_n, \tag{8}$$

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1} + \theta_3 + E_n, \tag{9}$$

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + E_n, \tag{10}$$

with $(E_n)_n$ being a series of i.i.d. errors satisfying $med(E_n) = 0, E_n > -\theta_1 y_0^{\theta_2} - \theta_3$ and $(Y_n)_n$ a growth process. The models (8) and (9) are autoregressive processes of order one without and with intercept. Model (10) is a non-linear autoregressive process. In contrast to usual assumptions, we observe growth processes, what induces some conditions on the parameters and errors. In particular, we have to deal with explosive autoregressive processes here. Further, low loading implies that the interesting experiments could approach the unit root case, leading to the analysis of mildly explosive processes.

Considering the data from our crack growth experiments, we assume that the errors can be asymmetric and possess heavy tails. This restricts the application of standard methods for estimation and testing for autoregressive processes. Hence, we will propose a depth based statistic for the discussed models which imposes very mild conditions on the errors and takes the growth assumption into account.

2.3 Data Depth and Regression

In this section, we give a short introduction to depth based estimation and testing before we propose the central statistics for this thesis. Data depth was originally introduced to generalise the median in a multivariate setting.

Depth to Generalise Location Measures

The first known depth notion was proposed by Tukey (1975). It is defined by

$$HD(z,\mu) = \min_{u\in\mathbb{R}^d,\|u\|=1} \#\{z_i \in \{z_1,...z_N\} | u^T z_i \leq u^T\mu\},$$

whereby $z = (z_1, ..., z_N)$ is a sample of N points in $\mathbb{R}^d$ and $\mu \in \mathbb{R}^d$ is a parameter, see Struyf and Rousseeuw (1999). For a one dimensional value $\mu \in \mathbb{R}$ and a sample $z = (z_1, ..., z_N), z_i \in \mathbb{R}$ half-space depth is

$$HD(z,\mu) = \min\{\#\{n|\mu \leq z_n\}, \#\{n|\mu \geq z_n\}\}.$$

Hence, the parameter with maximal depth coincides with the median in this case. Even if Tukey (1975) was the first to name generalisations of multivariate rank methods *depth*, nowadays there are several depth notions based on older results. The oldest depth notion goes back to Mahalanobis (1936). Here, a depth is defined by taking the inverse of the squared distances of a data point to the (multivariate) mean, weighted by the dispersion matrix. The sample version is defined by

$$MD(z,\mu) = (1 + (\mu - \bar{z})^T \hat{\Sigma}_z^{-1} (\mu - \bar{z}))^{-1},$$

whereby $z = (z_1, ..., z_N)$ is a given sample of values $z_i \in \mathbb{R}^d$ and $\mu \in \mathbb{R}^d$ is a parameter of interest. $\bar{z}$ is the mean vector and $\hat{\Sigma}_z$ is the empirical covariance matrix of the dataset z. Here, depth is measured by a distance of the considered point to the centre of a distribution. The Oja depth (see Oja, 1983) is defined by the expected volume of simplexes defined by $d < N$ variables and the considered point μ. The empirical version is given by

$$OD(z,\mu) = \binom{N}{d}^{-1} \left(1 + \sum_{1\leq n_1<...<n_d\leq N} vol(S(\mu, z_{n_1}, ..., z_{n_d}))\right)^{-1},$$

whereby $vol(S(z_1, ..., z_k))$ is the volume of the $k-1$ dimensional simplex $S(z_1, ..., z_k)$ with vertices $z_1, ..., z_k$, see Liu et al. (1999). Figuratively, the Oja depth measures the

depth of a point μ by summation of the areas of all simplexes which can be formed by combinations of d points from the sample and μ as fixed vertex. Another simplex based depth notion is the simplicial depth by Liu (1990). Here, depth is defined by the probability of a considered point to be covered by simplexes formed from $d+1$ dimensional subsets of the variables $(Z_1, ..., Z_N)$. In the sample version one can replace the probability by the empirical distribution. The empirical simplicial depth is given by

$$SD(z, \mu) = \binom{N}{d+1}^{-1} \sum_{1 \leq n_1 < ... < n_{d+1} \leq N} \mathbb{1}_{\{\mu \in S(z_{n_1}, ..., z_{n_{d+1}})\}}.$$

A generalised version of simplicial depth will be introduced in Definition 6, since it is an essential ingredient for our methods. A general overview about depth notions for multivariate data analysis is given in Liu et al. (1999).

Depth Based on Quality Functions and Regression

With respect to regression models, extensions of the depth concept exist. A general framework was defined by Rousseeuw and Hubert (1999) who defined regression depth. The general idea is to define depth as a function of the parameter θ evaluated at observations z_n. Then, a parameter with depth of zero can be defined as a non-fit which describes a situation where a parameter can be considered as arbitrary bad fit with respect to its qualitative fit within the set of observations. Regression depth then is defined as the number of observations which have to be removed to make a parameter θ a non-fit. In their work, Rousseeuw and Hubert (1999) present more precise formulas for specific regression models and show that a maximal depth concept can be defined. Lin and Chen (2006) also applied the method to generalised linear models under similar conditions.
We use a specific depth, following Mizera (2002), which allows a general application to arbitrary quality functions.

Definition 4. *For $z = (z_1, ..., z_N)$, with $z_n \in \mathbb{R}^k$, $\theta \in \mathbb{R}^d$ and a function $Q(\theta, z_n) : \mathbb{R}^{d \times k} \to \mathbb{R}$, we define **tangential depth** of θ with respect to an observation vector z and quality function Q as*

$$d_T^Q(\theta, z) = \frac{1}{N} \underset{u \in \mathbb{R}^d}{min} \# \left\{ n \,\middle|\, u^\top \frac{\partial Q(\theta, z_n)}{\partial \theta} \geq 0 \right\}. \tag{11}$$

This definition generalises well known depth notions like half space depth or local

depth for specific choices of Q.

Remark 5. *Note that the parameter with maximal depth is not necessarily unique.*

The advantage of the depth notion is that the concept easily can be translated to multidimensional parameters and more complex problems by appropriate quality functions.
A more convenient definition of tangential depth can be formulated, if one notices that taking the minimum over all $u \neq 0$ just means to check all half-spaces containing 0 on their boundary. Since the length of u has no influence on the sign, one can replace (11) as follows

$$d_T^Q(\theta, z) = \frac{1}{N} \min_{\|u\|=1, u \in \mathbb{R}^d} \# \left\{ n \left| u^\top \frac{\partial Q(\theta, z_n)}{\partial \theta} \geq 0 \right. \right\}.$$

As proposed by Müller (2005) for likelihood depth, Wellmann and Müller (2010) for orthogonal regression and Wellmann et al. (2009) for polynomial regression, a simplicial depth notion can be defined based on arbitrary depth notions.

Definition 6. ***Simplicial depth*** *for a parameter $\theta \in \mathbb{R}^d$, based on observations $z = (z_1, ..., z_N)$ and an arbitrary depth notion d_A is defined by*

$$d_S^{d_A}(\theta, z) = \frac{1}{\binom{N}{d+1}} \sum_{1 \leq n_1 < ... < n_{d+1} \leq N} \mathbb{1}_{\{d_A(\theta, (z_{n_1}, ..., z_{n_{d+1}}) > 0\}}.$$

Often simplicial depth is a U-statistic. Therefore, the indicator $\mathbb{1}_{\{d_A(\theta, (z_{n_1}, ..., z_{n_{d+1}}))\}}$ has to define a proper kernel function. This can typically be achieved by the choice of the involved depth notions. But just in few cases this U-statistic is not degenerated. For instance Denecke and Müller (2011, 2012, 2014b,a) present applications, where this is the case. In the degenerated and the non-degenerated case appropriate limit theorems, see for example Witting and Müller-Funk (1995), can be applied. These theorems imply that simplicial depth converges in distribution and the limit is often defined by functions of Normal distributions or of χ^2 distributions respectively. In the degenerated case the kernel of the U-statistic has to be decomposed to allow asymptotic results. This was for example used in Kustosz and Müller (2014).

3 Depth Based Estimators for Autoregressive Models

The application for our models is based on the work for regression models. While in the one parameter case the results will be similar to linear regression through the origin, the two parameter models require new methods to derive limit distributions. The general approach in this thesis can be summarised as follows.
Assume that a dataset $y = (y_1, ..., y_N)^\top \in \mathbb{R}^N, x = (x_1, ..., x_N)^\top \in \mathbb{R}^N$ is given. We denote the complete data by $z = (z_1, ..., z_N)^\top = ((x_1, y_1)^\top, ..., (x_N, y_N)^\top)^\top$ and analyse a model

$$y_n = f(x_n, \theta) + \epsilon_n$$

with parameter $\theta \in \mathbb{R}^d$ and $(\epsilon_n)_n$ a series of independent and identically distributed errors with unknown distribution, satisfying $\text{med}(\epsilon_n) = 0$. To measure the quality of the fit of a parameter, we first define a real valued quality function $Q(\theta, z_n)$. Then, we use tangential depth to define a depth notion for our models. Plugging the resulting expression into simplicial depth delivers a possibility to derive asymptotic distributions.
This allows us to define estimators, tests and confidence regions by the following statements.

Definition 7. *For an arbitrary depth notion $d_A(\theta, z)$, the* ***maximum depth estimate*** *is defined by*

$$\hat{\theta}(z) = \underset{\theta \in \mathbb{R}^d}{argmax}\ d_A(\theta, z).$$

If we can derive the asymptotic distribution for an appropriate transformation $T_A(\theta, z)$ of an arbitrary depth notion d_A, asymptotic tests can be defined.

Theorem 8. *(see Müller, 2005)*
If $T_A(\theta, Z)$ has a continuous limit distribution, an asymptotic α level test for $H_0 : \theta \in \Theta^0$ is defined by $\varphi : \mathbb{R}^{N\times 2} \to \{0, 1\}$ with

$$\varphi(z) = \begin{cases} 1 & , if \ \sup\limits_{\theta\in\Theta^0} T_A(\theta, z) \leq q_\alpha \\ 0 & else \end{cases},$$

whereby q_α is the α quantile of the limit distribution of $T_A(\theta, Z)$.

Proof. For a random sample Z,

$$\begin{aligned} & \lim_{N\to\infty} \mathbb{P}^Z_{\theta^0}\left(\left\{z \in \mathbb{R}^{N\times 2} | \varphi(z) = 1\right\}\right) \\ = & \lim_{N\to\infty} \mathbb{P}^Z_{\theta^0}\left(\left\{z \in \mathbb{R}^{N\times 2} \,\middle|\, \sup_{\theta\in\Theta^0} T_A(\theta, z) \leq q_\alpha\right\}\right) \\ \leq & \lim_{N\to\infty} \mathbb{P}^Z_{\theta^0}\left(\left\{z \in \mathbb{R}^{N\times 2} | T_A(\theta^0, z) \leq q_\alpha\right\}\right) \\ = & \lim_{N\to\infty} \mathbb{P}_{\theta^0}(T_A(\theta^0, Z) \leq q_\alpha) \\ = & \ \alpha \end{aligned}$$

holds for all $\theta^0 \in \Theta^0$. □

This theorem states that if we can define an appropriate test statistic $T_A(\theta, z)$ with known asymptotic distribution, a test for $H_0 : \theta \in \Theta^0$ is defined, by rejection, if all parameters in Θ^0 are rejected. To apply this theorem for the proposed simplicial depth statistics, we have to assure that $T_A(\theta, z) = h(d_A(\theta, z))$, whereby h is a strictly monotone function of the depth. By the link between one-point hypothesis tests and confidence regions, we can also define depth based confidence regions.

Corollary 9. *Consider α level tests $\varphi_{\theta^0}(z) = \mathbb{1}_{\{T_A(\theta^0,z)\leq q_\alpha\}}$ for $H_0 : \theta = \theta^0$ as defined in Theorem 8. Let Θ be an appropriate parameter space. If $T_A(\theta, Z)$ has a continuous limit distribution, an asymptotic $(1 - \alpha)$ confidence region for θ is given by*

$$\hat{\Theta}_\alpha(z) = \{\theta \in \Theta | T_A(\theta, z) > q_\alpha\}.$$

Proof. For all $\theta^0 \in \Theta$, we have

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}_{\theta^0}^Z \left(\left\{ z \in \mathbb{R}^{N\times 2} \,\middle|\, \theta^0 \in \hat{\Theta}_\alpha(z) \right\} \right) \\
=\ &\lim_{N\to\infty} \mathbb{P}_{\theta^0}^Z \left(\left\{ z \in \mathbb{R}^{N\times 2} | T_A(\theta^0, z) > q_\alpha \right\} \right) \\
=\ &1 - \lim_{N\to\infty} \mathbb{P}_{\theta^0}^Z \left(\left\{ z \in \mathbb{R}^{N\times 2} | T_A(\theta^0, z) \leq q_\alpha \right\} \right) \\
=\ &1 - \lim_{N\to\infty} \mathbb{P}_{\theta^0} \left(T_A(\theta^0, Z) \leq q_\alpha \right) \\
=\ &1 - \alpha
\end{aligned}$$

for $z = (z_1, ..., z_N)$ and $Z = (Z_1, ..., Z_N)$. □

3.1 Linear AR(1) Process without Intercept

The first model for which we propose a depth based statistic is the one parameter model defined by (8). The main results can be found in Kustosz and Müller (2014). As proposed in Section 2.3, we first have to define a quality function. Since the squared residuals appear as a good quality function in case of normally distributed errors in linear regression, we define

$$Q(\theta_1, z_n) = (y_n - y_{n-1} - \theta_1 y_{n-1})^2 = r_n(\theta, y)^2,$$

whereby $z_n = (y_n, y_{n-1}), n \in \{1, ..., N\}$ and $r_n(\theta, y) := (y_n - y_{n-1} - \theta_1 y_{n-1}), n \in \{1, ..., N\}$ are the residuals. Further, $y = (y_0, ..., y_N)^\top$ is an observation of the underlying autoregressive process and $y_0 > 0$ is fixed and known. With respect to the model, we consider $\theta := \theta_1$ in this section to simplify notation. The derivative of the quality function then is

$$\frac{\partial Q(\theta, z_n)}{\partial \theta} = -2(y_n - y_{n-1} - \theta_1 y_{n-1}) y_{n-1}.$$

Considering the error conditions from Lemma 3 and assuming $y_0 > 0$ delivers that $y_n > 0$ for all $n \in \mathbb{N}$. Inserting this into tangential depth for the AR(1) process

without intercept leads to

$$\begin{aligned} d_T^{AR}(\theta, z) =& \frac{1}{N} \underset{|u|=1}{min} \#\left\{n \,|u(-2(y_n - y_{n-1} - \theta_1 y_{n-1})y_{n-1}) \geq 0\right\} \\ =& \frac{1}{N} \underset{u\in\{-1,1\}}{min} \#\left\{n \,|u(-2(y_n - y_{n-1} - \theta_1 y_{n-1})y_{n-1}) \geq 0\right\} \\ =& \frac{1}{N} \underset{u\in\{-1,1\}}{min} \#\left\{n \,|u((y_n - y_{n-1} - \theta_1 y_{n-1})y_{n-1}) \geq 0\right\} \\ =& \frac{1}{N} min\left\{\#\left\{n \,|((y_n - y_{n-1} - \theta_1 y_{n-1})y_{n-1}) \geq 0\right\}\right. \\ & \qquad \left., \#\left\{n \,|((y_n - y_{n-1} - \theta_1 y_{n-1})y_{n-1}) \leq 0\right\}\right\} \\ =& \frac{1}{N} min\left\{\#\left\{n \,|(y_n - y_{n-1} - \theta_1 y_{n-1}) \geq 0\right\}\right. \\ & \qquad \left., \#\left\{n \,|(y_n - y_{n-1} - \theta_1 y_{n-1}) \leq 0\right\}\right\} \\ =& \frac{1}{N} min\left\{\#\left\{n \,|r_n(\theta, y) \geq 0\right\}, \#\left\{n \,|r_n(\theta, y) \leq 0\right\}\right\}. \end{aligned} \tag{12}$$

We see that tangential depth just depends on the signs of the residuals in this model. Even if the limit distribution of tangential depth under $H_0 : \theta = \theta^0$, with $med(E_n) = 0$ can be derived quite easily, the properties of this statistic are not appropriate to analyse growth processes. Tests based on tangential depth coincide with simple sign tests. These tests have a huge drawback at finite samples for growth processes. This drawback is discussed in Section 5 in more detail. Plugging tangential depth into simplicial depth and using the growth assumptions allows an improvement of the tangential depth based tests. Simplicial depth for AR(1) processes without intercept based on tangential depth can be defined by

$$d_S^{AR}(\theta, y) = \frac{1}{\binom{N}{2}} \sum_{1\leq n_1<n_2\leq N} \mathbb{1}_{\{d_T^{AR}(\theta,(z_{n_1},z_{n_2})>0\}}. \tag{13}$$

Note that the simplicial depth can be regarded as function of $y = (y_0, ..., y_N)^\top$, i.e. the underlying observed process, instead of $z = (z_1, ..., z_N)^\top$, since for autoregressive models $x_n = y_{n-1}$ holds. Tangential depth is larger than 0 for two tuples of observations z_{n_1}, z_{n_2} if and only if the respective residuals change signs or one of the residuals is zero. The depth statistic d_S^{AR} can be calculated more explicitly.

Theorem 10. *For the autoregressive process of order one without intercept defined by*

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1} + E_n,$$

we have

$$d_S^{AR}(\theta, y) = \frac{1}{\binom{N}{2}} \sum_{1 \le n_1 < n_2 \le N} \big(\mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0\}} + (1 - \mathbb{1}_{\{r_{n_1}(\theta,y)\neq 0\}} \mathbb{1}_{\{r_{n_2}(\theta,y)\neq 0\}}) \big) . \tag{14}$$

We refer to d_S^{AR} *as* ***complete simplicial depth*** *for the autoregressive process without intercept.*

Proof. By (12), $d_T(\theta, (z_{n_1}, z_{n_2})) > 0$ holds if and only if

$$min\left\{ \#\{n \in \{n_1, n_2\} | r_n(\theta, y) \ge 0\}, \#\{n \in \{n_1, n_2\} | r_n(\theta, y) \le 0\} \right\} > 0.$$

This minimum is zero, if $r_n(\theta, y) > 0$ for both $n \in \{n_1, n_2\}$ or, if $r_n(\theta, y) < 0$ for both $n \in \{n_1, n_2\}$, since then one of the sets consists of two and the other of zero elements. The expression is larger than zero, if one of the residuals is non-negative and the other residual is non-positive or, if at least one residual is zero. This happens in three cases which can be summarised by

$$\mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0\}} + 1 - \mathbb{1}_{\{r_{n_1}(\theta,y)\neq 0\}} \mathbb{1}_{\{r_{n_2}(\theta,y)\neq 0\}}.$$

Replacing the indicators over the tangential depth in (13) by the upper expression provides (14). □

In application, the terms where one residual is exactly zero can be neglected. From the theoretical point of view, these events have probability zero for continuous data. However, empirical depth deterministically assigns a value larger than zero to them, because they appear as edges of considered simplexes. Therefore, these points have a fixed positive value of depth, what contradicts the fact of singular points having a probability of zero in the limit. Additionally, the depth from these edges counts all simplexes starting at the edge, what introduces jumps in the depth shape. Example 11 illustrates the problem when a simple location model is considered. This property was already observed by Burr et al. (2004) who proposed another modification which gives additional weight to points in the interior of the simplexes and thereby overcomes some of the related problems.

Example 11. *Consider simulated data from a linear AR(1) process without intercept with* $\mathcal{N}(0, 0.1)$ *errors,* $\theta_1 = 1.1$*,* $N = 10$ *and* $y_0 = 1.1$ *as depicted in Figure 3. The ten available observations lead to nine residuals and nine parameters satisfying* $\frac{y_n}{y_{n-1}} - 1 = \theta_1$ *which coincides with* $r_n(\theta, y) = 0$*. If we evaluate* d_S^{AR} *as proposed in*

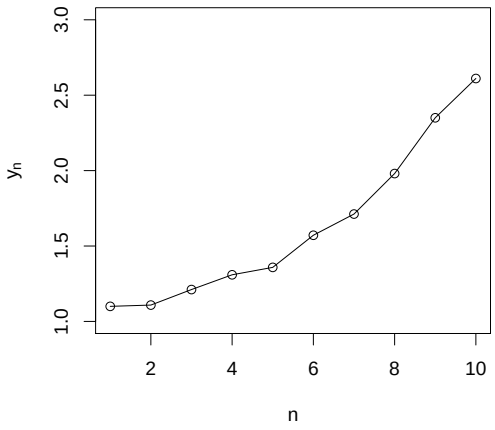


Fig. 3: Simulated AR process without intercept.

(14), we get an empirical depth shape as presented by the red function in Figure 4. We see that at the jump locations, which coincide with values of θ *given by* $\frac{y_n}{y_{n-1}} - 1$*, the empirical depth shape shows spikes. These spikes result from the fact that each parameter, given by* $\frac{y_n}{y_{n-1}} - 1$ *is the edge of several simplexes which are disjoint, except for this edge, but are contributing to the depth function. This inflates the value of the empirical simplicial depth. When we neglect these points and evaluate depth on a fine grid the resulting empirical depth is given by the dashed black lines.*

Remark 12. *In our application, the problem described above leads to regions of constant depth, surrounded by singular points of remarkably higher depth. This introduces discontinuities in the resulting confidence regions. Hence, we propose a version of simplicial depth, neglecting* $r_n(\theta, y) = 0$*, leading to a deflation of depth at these points. In particular, the term*

$$(1 - \mathbb{1}_{\{r_{n_1}(\theta,y)\neq 0\}}\mathbb{1}_{\{r_{n_2}(\theta,y)\neq 0\}})$$

is neglected. Note that in this case we still have spikes at these points, but with lower depth than in the surrounding parameters. However, convex hulls then lead to more reliable confidence regions.
Considering limit theory, this modification makes no difference for data with continuous distribution, since we have a probability of zero at these points. Further, the effect can be neglected, if the data is recorded at a higher accuracy than the parameter grid, since we do not evaluate at such points then. Hence, the problem can also be handled by a careful implementation of the statistic. In cases where the candidates

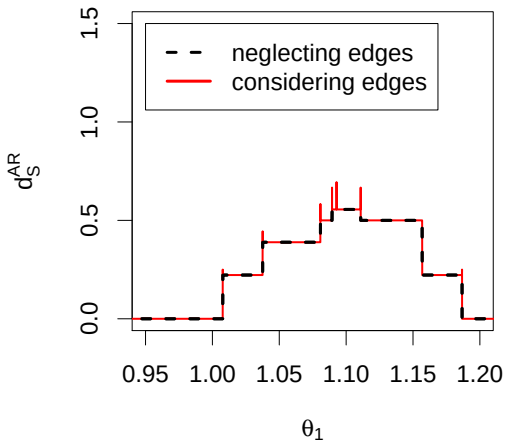


Fig. 4: Resulting depth by considering and neglecting edges for one simulated AR process.

coincide with the edges, the test statistic or the candidate set should be adjusted, for example as we propose in Section 4.1.

According to the definition of depth, a parameter implies a good model fit, if depth is maximal. This leads to the following definition.

Definition 13. *The **maximum simplicial depth estimator** for $\theta := \theta_1$ in model (8) is defined by*

$$\hat{\theta}^{AR}(y) = \underset{\theta>0}{argmax}\ d_S^{AR}(\theta, y).$$

Under the assumption that $\mathbb{P}(E_n = 0) = 0$, and considering Remark 12, we can treat simplicial depth as a U-statistic with kernel

$$\psi(x_1, x_2) = \mathbb{1}_{\{x_1<0\}}\mathbb{1}_{\{x_2>0\}} + \mathbb{1}_{\{x_1>0\}}\mathbb{1}_{\{x_2<0\}}. \tag{15}$$

By a spectral decomposition and calculation of the conditional expectation of the kernel, simplicial depth appears to be a degenerated U-statistic for which the Theorem of Hoeffding (see Witting and Müller-Funk, 1995) can be applied. This provides the following statement.

Theorem 14. *For the autoregressive process of order one without intercept defined by*

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1} + E_n,$$

with $\mathbb{P}(E_n \geq 0) = \frac{1}{2} = \mathbb{P}(E_n \leq 0)$ *and* $E_n \geq -\theta_1 y_0, \theta_1 > 0$*, simplicial depth under the true parameter* $\theta^0 := \theta_1^0$ *satisfies*

$$N\left(d_S^{AR}(\theta^0, (Y_1, ..., Y_N)) - \frac{1}{2}\right) \xrightarrow{d} \frac{1}{2} - \frac{1}{2}X^2,$$

for $N \to \infty$ *and* $X \sim \mathcal{N}(0,1)$*.*

Proof. See Kustosz and Müller (2014). The main argument is the theorem of Hoeffding (see Witting and Müller-Funk (1995), Satz 7.183, p. 650 as well as p. 155). Based on the observation that the simplicial depth statistic is a degenerated U-statistic, a spectral decomposition, as in Müller (2005), can be applied to derive the limit distribution. □

With these results, we can define tests and confidence intervals for explosive AR(1) processes without intercept based on simplicial depth.
To derive asymptotic tests for $H_0 : \theta \in \Theta^0$, Theorem 14 can be used.

Corollary 15. *An asymptotic* α*-level test for* $H_0 : \theta \in \Theta^0$ *for model (8) under the conditions of Theorem 14 is defined by* $\varphi : \mathbb{R}^{N+1} \to \{0,1\}$ *with*

$$\varphi(y) = \begin{cases} 1 & , if \ \sup\limits_{\theta \in \Theta^0} N(d_S^{AR}(\theta, y) - \frac{1}{2}) < \frac{1}{2} - \frac{1}{2} q_{\chi_1^2}(1-\alpha) \\ 0 & else \end{cases}, \tag{16}$$

whereby $q_{\chi_m^2}(\gamma)$ *is the* γ *quantile of the* χ^2 *distribution with* m *degrees of freedom.*

Proof of Corollary 15. (see Kustosz and Müller, 2014)
For $\theta^0 \in \Theta^0$ we have

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}_{\theta^0}\left(\sup_{\theta\in\Theta^0} N\left(d_S^{AR}(\theta, (Y_0, \ldots, Y_N)) - \frac{1}{2}\right) < \frac{1}{2} - \frac{1}{2}\chi_1^2(1-\alpha)\right) \\
&\leq \lim_{N\to\infty} \mathbb{P}_{\theta^0}\left(N\left(d_S^{AR}(\theta^0, (Y_0, \ldots, Y_N)) - \frac{1}{2}\right) < \frac{1}{2} - \frac{1}{2}\chi_1^2(1-\alpha)\right) \\
&= \mathbb{P}\left(-\frac{1}{2}(X^2 - 1) < \frac{1}{2} - \frac{1}{2}\chi_1^2(1-\alpha)\right) = \mathbb{P}(-X^2 + 1 < 1 - \chi_1^2(1-\alpha)) \\
&= \mathbb{P}(-X^2 < -\chi_1^2(1-\alpha)) = \mathbb{P}(X^2 > \chi_1^2(1-\alpha)) = \alpha.
\end{aligned}$$

□

By considering the link between tests and confidence intervals the next corollary immediately follows.

Corollary 16. *Under the assumptions of Theorem 14 an asymptotic* $(1-\alpha)$ *confidence interval for* θ_1 *is given by*

$$\hat{\Theta}^{AR}_{1-\alpha}(y) = \left\{ \theta > 0 \middle| N(d_S^{AR}(\theta, y) - \frac{1}{2}) \geq \frac{1}{2} - \frac{1}{2} q_{\chi^2_1}(1-\alpha) \right\}. \tag{17}$$

Proof of Corollary 16. The assertion follows from the application of Corollary 15 and the relation of point tests and confidence intervals in Corollary 9. □

The next statement shows how the empirical depth functions behave with respect to monotonicity. This statement later will be useful to compute depth and to understand the behaviour of the empirical depth under non-constant model parameters.

Lemma 17. *Let be* $\kappa_n := \frac{y_n}{y_{n-1}}, \kappa_n \neq \kappa_m$ *for* $n \neq m$*, with* $n, m \in \{1, ..., N\}$ *and* $\kappa_{(1)} < ... < \kappa_{(N)}$ *the ordered set of* $\kappa_1, ..., \kappa_N$ *for an observed autoregressive process without intercept* $y = (y_0, ..., y_N)$*. Then* $d_S^{AR}(\theta, y)$ *is maximal if and only if*

$$\theta \in \begin{cases} \left(\kappa_{(\frac{N}{2})} - 1, \kappa_{(\frac{N}{2})+1} - 1\right) & \text{for } N \in 2\mathbb{N} \\ \left(\kappa_{(\frac{N-1}{2})} - 1, \kappa_{(\frac{N+1}{2})} - 1\right) \cup \left(\kappa_{(\frac{N+1}{2})} - 1, \kappa_{(\frac{N+3}{2})} - 1\right) & \text{for } N \in 2\mathbb{N}+1 \end{cases}.$$

Remark 18. *We use the open sets to avoid the usage of parameters which coincide with roots of the residuals, since then depth is potentially deflated, as discussed in Remark 12.*

Proof of Lemma 17. At first, we observe that changes in $d_S^{AR}(\theta, y)$ can only appear, if the residuals $r_n(\theta, y)$ change signs. For an observed process $y = (y_0, ..., y_N)$ these sign changes can be determined by

$$\theta_n + 1 = \kappa_n := \frac{y_n}{y_{n-1}}.$$

Now, consider the ordered set of these locations of sign changes denoted by $\kappa_{(n)} = \theta_{(n)} + 1$. One can easily show that $d_S^{AR}(\theta, y) = 0$ for $\theta + 1 = \kappa < \kappa_{(1)}$ and for $\theta + 1 = \kappa > \kappa^*_{(N)}$, since then for any $\kappa < \kappa_{(1)}$, we have $\kappa < \kappa_{(n)}$ for all $n \in \{1, ..., N\}$ and hence for n_0 being the index of $\kappa_{(1)}$, i.e. $\kappa_{(1)} = \frac{y_{n_0}}{y_{n_0-1}}$,

$$y_n - \kappa y_{n-1} > y_n - \kappa_n y_{n-1} = y_n - \frac{y_n}{y_{n-1}} y_{n-1} = 0 \ \forall n \in \{1, ..., N\}.$$

Analogously, we get $y_n - \kappa y_{n-1} < 0 \; \forall \; n \in \{1, ..., N\}$ for $\kappa > \kappa_{(N)}$.
We now prove that there is a set defined by $[\theta_{(n_0)}, \theta_{(m_0)}]$ with $\theta_{(n_0)} = \kappa_{(n_0)} - 1$ and $\theta_{(m_0)} = \kappa_{(m_0)} - 1$, $n_0 < m_0$, so that

$$d_S^{AR}(\theta, y) > d_S^{AR}(\theta_{(m_0)+1}, y) \text{ and } d_S^{AR}(\theta, y) > d_S^{AR}(\theta_{(n_0)-1}, y)$$

for all $\theta \in [\theta_{(n_0)}, \theta_{(m_0)}]$. To see this, the order of the candidate parameters can be used. Due to the symmetry of $d_S^{AR}(\theta, y)$ in the residual signs, we can calculate depth directly based on the ordered parameters $\kappa_{(n)}$ which are determined by the observations.

Crossing $\kappa_{(1)}$ delivers the first contribution to the depth statistic. For $\kappa = \theta + 1 \in (\kappa_{(1)}, \kappa_{(2)})$ we have

$$y_{n_0} - y_{n_0-1} - \theta y_{n_0-1} = y_{n_0} - \kappa y_{n_0-1} < y_{n_0} - \kappa_{(1)} y_{n_0-1} = y_{n_0} - \kappa_{n_0} y_{n_0-1} = 0$$

for n_0 being the index of the observations defining the first candidate parameter, i.e. $\frac{y_{n_0}}{y_{n_0-1}} = \kappa_{(1)}$, and

$$y_n - y_{n-1} - \theta y_{n-1} = y_n - \kappa y_{n-1} > y_n - \kappa_{(2)} y_{n-1} \geq y_n - \kappa_n y_{n-1} = 0$$

$\forall \; n \in \{1, ..., N\} \setminus \{n_0\}$, since $\kappa_n \geq \kappa_{(2)}$. This means that $r_{n_0}(\theta, y) < 0$ and $r_n(\theta, y) > 0 \; \forall n \in \{1, ..., N\} \setminus \{n_0\}$. Now the symmetry of the terms in the statistic can be used to see that

$$\begin{aligned} d_S^{AR}(\theta, y) &= c(N) \sum_{n_1 < n_2} \mathbb{1}_{\{r_{n_1}(\theta,y)<0\}} \mathbb{1}_{\{r_{n_2}(\theta,y)>0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)>0\}} \mathbb{1}_{\{r_{n_2}(\theta,y)<0\}} \\ &= c(N) \sum_{n=1}^{N} \mathbb{1}_{\{r_n(\theta,y)<0\}} \sum_{m=1}^{N} \mathbb{1}_{\{r_m(\theta,y)>0\}}, \end{aligned} \tag{18}$$

whereby $c(N) = 1/\binom{N}{2}$ is the scaling constant. By (18), we see that $d_S^{AR}(\theta, y)$ just depends on the number of negative residuals which can be uniquely determined by the interval for κ due to the order $\kappa_{(n)}$. Hence, for example, for $\kappa \in (\kappa_{(1)}, \kappa_{(2)})$ we have

$$d_S^{AR}(\kappa - 1, y) = c(N) \cdot 1 \cdot (N - 1) = \frac{N - 1}{\binom{N}{2}}. \tag{19}$$

For a general θ or κ, we can compute

$$d_S^{AR}(\kappa - 1, y) = d_S^{AR}(\theta, y) = c(N) \cdot k \cdot (N - k) = c(N)(Nk - k^2), \tag{20}$$

with k being defined by the number of negative residuals for $\kappa \in (\kappa_{(k)}, \kappa_{(k+1)})$. Considering the function given by (20) as continuous in the number of negative residuals k, this can be used to show that a unique maximising interval exists, since the function

$$f(k) = (kN - k^2)c(N)$$

has a unique maximum in $k = \frac{N}{2}$ for $k \in \mathbb{R}$. Since we just have $k \in \{0, ..., N\}$, we need to take the discrete structure of the empirical depth function into account. Therefore, we need to differentiate the solution for odd and even numbers of residuals.

For $N \in 2\mathbb{N}$, we find one maximising interval given by $(\kappa_{(\frac{N}{2})}, \kappa_{(\frac{N}{2})+1})$. To show this, observe that an even number N of residuals gives an even number of candidates. Further $N/2$ is an integer. For $\kappa \in (\kappa_{(\frac{N}{2})}, \kappa_{(\frac{N}{2})+1})$, we then have $\kappa > \kappa_{(n)}$ for $n \in \{1, ..., \frac{N}{2}\}$ which are $N/2$ candidates and $\kappa < \kappa_{(n)}$ for $n \in \{\frac{N}{2} + 1, ..., N\}$ which are also $N/2$ candidates. This means that we have exactly $N/2$ negative and $N/2$ positive residuals. Crossing one of the boundaries of the interval turns one residual and therefore reduces depth.

Note that the maximal depth on this interval using (18) is given by $c(N) \cdot \frac{N}{2} \cdot \frac{N}{2} = c(N) \cdot \frac{N^2}{4}$, what is the same value as the maximum of f given by the continuous optimisation problem.

For $N \in 2\mathbb{N} + 1$, we need to combine two maximising intervals, since an equal separation of residuals with positive and negative signs is not possible. We now can have $\frac{N-1}{2}$ positive and $\frac{N+1}{2}$ negative residuals, and vice versa. The interval $(\kappa_{(\frac{N-1}{2})}, \kappa_{(\frac{N+1}{2})})$ leads to $\frac{N+1}{2}$ negative residuals and $\frac{N-1}{2}$ positive residuals. Considering the next possible interval by crossing $\kappa_{(\frac{N+1}{2})}$, defined by $(\kappa_{(\frac{N-1}{2})}, \kappa_{(\frac{N+1}{2})})$, turns one sign and we get $\frac{N-1}{2}$ negative residuals and $\frac{N+1}{2}$ positive residuals. For parameters outside of these intervals another sign change reduces the depth below its maximal possible value again.

Note that here the maximal depth using (18) is given by $c(N) \cdot \frac{N+1}{2} \cdot \frac{N-1}{2} = c(N) \cdot \frac{N^2-1}{4}$, what is below the theoretical value given by $c(N) \cdot \frac{N^2}{4}$.

□

Remark 19. *A closed form solution of the maximal empirical depth for the linear*

autoregressive model without intercept can be given by $c(N)(kN - k^2)$ *with* $k = \lfloor \frac{N}{2} \rfloor$.

Lemma 17 gives a simple method to calculate the parameter with maximal simplicial depth for the autoregressive model without intercept.

Corollary 20. *Using the notation from Lemma 17, the* ***maximum simplicial depth estimator*** *for* θ *in model (8) is*

$$\hat{\theta}^{AR}(y) \in \begin{cases} \left(\kappa_{(\frac{N}{2})} - 1, \kappa_{(\frac{N}{2})+1} - 1\right) & \text{for } N \in 2\mathbb{N}, \\ \left(\kappa_{(\frac{N-1}{2})} - 1, \kappa_{(\frac{N+1}{2})} - 1\right) \cup \left(\kappa_{(\frac{N+1}{2})} - 1, \kappa_{(\frac{N+3}{2})} - 1\right) & \text{for } N \in 2\mathbb{N} + 1. \end{cases}$$

An interesting observation is that this property also holds, if y_n is not strictly increasing or, if the process consists of phases defined by multiple autoregression parameters, as can be seen in the examples in Section 5.

3.2 Linear AR(1) Process with Intercept and Non-linear AR(1) Process

Extending the model to an AR(1) process with intercept is not straightforward. While the construction of simplicial depth can be performed similar to the one parameter case, the derivation of the limit distribution gets more complicated. We now consider the model defined by (9). In this section, we simplify notation by $\theta = (\theta_1, \theta_3)$ in the context of the linear model or $\theta = (\theta_1, \theta_2)$ in the context of the non-linear model. Remind that $z_n = (y_n, y_{n-1})^\top$ and that the final statistics can be regarded as functions in $y = (y_0, ..., y_N)^\top$, also. By the same intuition as in Section 3.1, we define the quality function by the model residuals via

$$Q(\theta, z_n) = (y_n - y_{n-1} - \theta_1 y_{n-1} - \theta_3)^2 = r_n(\theta, y)^2. \tag{21}$$

To define tangential depth, we now need to derive the quality function in two parameters. This leads to

$$\frac{\partial Q(\theta, z_n)}{\partial \theta_1} = -2 r_n(\theta, y) y_{n-1}$$
$$\frac{\partial Q(\theta, z_n)}{\partial \theta_3} = -2 r_n(\theta, y).$$

Hence, tangential depth for an AR(1) process with intercept is defined by

$$d_T^{ARi}(\theta, y) = \min_{\|u\|=1, u\in\mathbb{R}^2} \#\left\{n|(u_1, u_2)\cdot(y_{n-1}, 1)^\top \cdot r_n(\theta, y) \geq 0\right\}.$$

This expression is less trivial than the purely residual dependent expression in the one parameter case. Nevertheless, one can show that for growth processes simplicial depth, based on tangential depth, reduces to the evaluation of the residuals. Since the results for the non-linear model are identical to the linear model with intercept when the residuals are replaced, we formulate the main theorems for both cases here. Therefore, preliminary calculations for the non-linear model given by (10) have to be done.
The quality function for the non-linear model is

$$Q(\theta, z_n) = (y_n - y_{n-1} - \theta_1 y_{n-1}^{\theta_2})^2 = r_n(\theta, y)^2 \tag{22}$$

with derivatives

$$\begin{aligned}\frac{\partial Q(\theta, z_n)}{\partial \theta_1} &= -2r_n(\theta, y)y_{n-1}^{\theta_2}\\ \frac{\partial Q(\theta, z_n)}{\partial \theta_2} &= -2r_n(\theta, y)\theta_1 y_{n-1}^{\theta_2}\log(y_{n-1}).\end{aligned}$$

Hence, tangential depth here is

$$\begin{aligned}d_T^{nAR}(\theta, y) &= \min_{\|u\|=1, u\in\mathbb{R}^2} \#\left\{n|(u_1, u_2)\cdot(y_{n-1}^{\theta_2}, \theta_1 y_{n-1}^{\theta_2} log(y_{n-1}))^\top \cdot r_n(\theta, y) \geq 0\right\}\\ &= \min_{\|u\|=1, u\in\mathbb{R}^2} \#\left\{n|(u_1, u_2)\cdot(1, \theta_1 log(y_{n-1}))^\top \cdot r_n(\theta, y) \geq 0\right\}.\end{aligned}$$

Note that here a strictly positive process is not a simplification but necessary to allow the application of tangential depth.
Similar to the linear autoregressive model with intercept, it is possible to show that tangential depth is larger than zero if and only if a set of residuals based on three tuples of observations has alternating signs. Hence, simplicial depth for the non-linear autoregressive model reduces to the expression for the linear autoregressive model by replacing the residuals. To see this, we use the following result from Kustosz et al. (2016b).

Theorem 21. *For a model given by*

$$y_n = g(\theta, x_n) + E_n, n \in \{1, ..., N\}$$

with $z_n = (x_n, y_n) \in \mathbb{R}^2, \theta \in \mathbb{R}^K$ *and* $(E_n)_n$ *a series of independent and identically distributed errors, consider* $v(\theta, x_n) := \frac{\partial}{\partial\theta} g(\theta, x_n)$ *and tangential depth, given by the criterion* $u^\top \frac{\partial Q(\theta, z_n)}{\partial\theta} = u^\top \cdot v(\theta, x) \cdot r_n(\theta, z) \leq 0$ *with* $u \in \mathbb{R}^K$*. Let be* $x_1 < x_2 < \ldots < x_{K+1} \in I\!R$ *and assume the following conditions for* $w_u : [x_1, x_{K+1}] \to I\!R$ *given by* $w_u(x) = u^\top \cdot v(\theta, x)$*:*

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

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

Then $d_T(\theta, z) > 0$ *holds if and only if* $(r_1(\theta, z), \ldots, r_{K+1}(\theta, z))^\top$ *has alternating signs or at least one of the residuals is zero.*

Proof. The proof is given in Kustosz et al. (2016b). □

Theorem 22.
a) For model (9) and $y_n > y_{n-1}$ *for all* n*, simplicial depth reduces to*

$$d_S^{ARi}(\theta, y) = \frac{1}{\binom{N}{3}} \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \Big(\mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0, r_{n_3}(\theta,y)>0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0, r_{n_3}(\theta,y)<0\}} + 1 - \mathbb{1}_{\{r_{n_1}(\theta,y)\neq 0\}} \mathbb{1}_{\{r_{n_2}(\theta,y)\neq 0\}} \mathbb{1}_{\{r_{n_3}(\theta,y)\neq 0\}} \Big), \tag{23}$$

whereby r_{n_i} *are the residuals defined by (21).*
b) For model (10) and $y_n > y_{n-1}$ *for all* n*, simplicial depth reduces to*

$$d_S^{nAR}(\theta, y) = \frac{1}{\binom{N}{3}} \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \Big(\mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0, r_{n_3}(\theta,y)>0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0, r_{n_3}(\theta,y)<0\}} + 1 - \mathbb{1}_{\{r_{n_1}(\theta,y)\neq 0\}} \mathbb{1}_{\{r_{n_2}(\theta,y)\neq 0\}} \mathbb{1}_{\{r_{n_3}(\theta,y)\neq 0\}} \Big), \tag{24}$$

whereby r_{n_i} *are the residuals defined by (22).*
We refer to d_S^{ARi} *as* ***complete simplicial depth*** *for the autoregressive process with intercept and to* d_S^{nAR} *as* ***complete simplicial depth*** *for the non-linear autoregressive process.*

Proof. The proof is given in Kustosz et al. (2016b). The main idea is, to characterise $d_T^{ARi} > 0$ in (25) by the number of sign changes in the residuals with Theorem 21. The assertion then follows directly. For d_S^{nAR}, we have to consider $w_u(x) =$

$u_1 x^{\theta_2} + u_2\theta_1 x^{\theta_2} log(x)$ with respect to the possible sign changes. Therefore, it suffices to calculate the roots of w_u, assuming $u_1, u_2 > 0$, $x > 0$ and $\theta_1, \theta_2 \neq 0$. We get

$$\begin{aligned} & u_1 x^{\theta_2} + u_2\theta_1 x^{\theta_2} \log(x) = 0 \\ \Leftrightarrow & u_1 x^{\theta_2} = -u_2\theta_1 x^{\theta_2} \log(x) \\ \Leftrightarrow & u_1 = -u_2\theta_1 \log(x) \\ \Leftrightarrow & \frac{u_1}{-u_2\theta_1} = \log(x) \\ \Leftrightarrow & \exp(\frac{u_1}{-u_2\theta_1}) = x. \end{aligned}$$

Since one can in addition show that w_u has exactly one extremum at $x = \exp(-\frac{1}{\theta_2} - \frac{u_1}{u_2\theta_1})$ which is a minumum, if $u_2\theta_1\theta_2 > 0$ and a maximum, if $u_2\theta_1\theta_2 < 0$, see Kustosz et al. (2016b), proof of Lemma 2, pp. 31, u_2 can always be used to satisfy B) in Theorem 21. In particular, the choice of u_2 allows to flip the signs in the two regions of $w_u(x)$. This completes the assertion for the non-linear model. □

Definition 23. *Under the assumptions of Theorem 22 and by dismissing of the $r_{n_i}(\theta, y) = 0$ cases, see Remark 12, we redefine equations (23) and (24).*
a) For model (9) and $y_n > y_{n-1}$ for all n, simplicial depth reduces to

$$d_S^{ARi}(\theta, y) = \frac{1}{\binom{N}{3}} \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \big(\mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0, r_{n_3}(\theta,y)>0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0, r_{n_3}(\theta,y)<0\}} \big) \tag{25}$$

whereby r_{n_i} are the residuals defined by (21).
b) For model (10) and $y_n > y_{n-1}$ for all n, simplicial depth reduces to

$$d_S^{nAR}(\theta, y) = \frac{1}{\binom{N}{3}} \sum_{1 \leq n_1 < n_2 < n_3 \leq N} \big(\mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0, r_{n_3}(\theta,y)>0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0, r_{n_3}(\theta,y)<0\}} \big) \tag{26}$$

whereby r_{n_i} are the residuals defined by (22).

The definition of the simplicial depth estimator is similar to the one parameter model.

Definition 24. *For models (9) and (10)* ***maximum simplicial depth estimators*** *are defined by*

$$\hat{\theta}^{ARi}(y) \in \underset{\theta\in\mathbb{R}^2}{argmax}\ d_S^{ARi}(\theta, y)$$

and

$$\hat{\theta}^{nAR}(y) \in \underset{\theta\in\mathbb{R}^2}{argmax}\ d_S^{nAR}(\theta, y).$$

Again, simplicial depth has a form similar to a U-Statistic. The kernel is defined by

$$\psi(e_1, e_2, e_3) = \mathbb{1}_{\{e_1>0, e_2<0, e_3>0\}} + \mathbb{1}_{\{e_1<0, e_2>0, e_3<0\}}.$$

The limit theorem of Hoeffding cannot be applied directly, since the kernel is not symmetric. In particular, the limit distribution is a sum of integrated χ_1^2 processes.

Theorem 25. *For models (9) and (10) with* $\mathbb{P}(E_n \geq 0) = \frac{1}{2} = \mathbb{P}(E_n \leq 0)$, Y_n *almost surely strictly increasing,* $Y_0 = y_0, \theta_1 > 1$, *the respective simplicial depth under the true parameter* θ^0 *satisfies*

$$N\left(d_S^{mod}(\theta, Y) - \frac{1}{4}\right) \xrightarrow{d} \frac{3}{4} + \frac{3}{4}X_2^2(0) - \frac{3}{2}\int_{-2}^{2} X_1^2(t)dt,$$

whereby $X(t) = (X_1(t), X_2(t))^\top$ *is a centred Gaussian process on* $[-2, 2]$ *with continuous paths and covariance matrix*

$$\begin{aligned} &Cov(X(s), X(t)) \\ &= \begin{pmatrix} \int_0^1 \mathbb{1}_{(-0.5,0.5]}(x-s)\,\mathbb{1}_{(-0.5,0.5]}(x-t)\ dx & \int_0^1 \mathbb{1}_{(-0.5,0.5]}(x-s)\ drx \\ \int_0^1 \mathbb{1}_{(-0.5,0.5]}(x-t)\ dx & 1 \end{pmatrix}. \end{aligned} \tag{27}$$

Thereby $d_S^{mod}(\theta, y), mod \in \{ARi, nAR\}$ *are given by (25) and (26), respectively.*

Proof. The proof is given in Kustosz et al. (2016a) for the AR(1) process with intercept only. In this paper general residuals, satisfying $med(E_n) = 0$ under the null hypothesis, are considered. Hence, the result also holds for the non-linear autoregressive model used here. It suffices to observe that the resulting depth is given by (26) and just depends on the model residuals. Simplicial depth is related to a U-Statistic, but has an asymmetric kernel so that standard results for U-statistics cannot be used. By appropriate approximations and the median assumption on the errors, a construction similar to a proof for degenerated U-statistics can be

applied. □

Compared with the first limit theorem an analytic expression for the quantiles of this limit distribution is not available. Therefore, we present a method to approximate the limit distribution to generate tables of the distribution's quantiles for application, see Section 4.1. Analogously to the last section, tests and confidence intervals can be constructed.

Corollary 26. *An asymptotic α-level test for $H_0 : \theta \in \Theta^0 \subset \mathbb{R}^2$ for models (8) and (10) under the conditions of Theorem 25 is defined by $\varphi : \mathbb{R}^{N+1} \to \{0, 1\}$ with*

$$\varphi(y) := \begin{cases} 1 & , if \sup\limits_{\theta \in \Theta^0} (N(d_S^{mod}(\theta, y) - \frac{1}{4})) < q_G(\alpha) \\ 0 & else \end{cases}, \tag{28}$$

whereby $q_G(\gamma)$ is the γ quantile of the distribution of $W := \frac{3}{4} + \frac{3}{4} X_2^2(0) - \frac{3}{2} \int_{-2}^{2} X_1^2(t) dt$, with $X(t) = (X_1(t), X_2(t))^\top$ a centred Gaussian process on $[-2, 2]$ with continuous paths and covariance matrix defined by (27) and $d_S^{mod}(\theta, y), mod \in \{ARi, nAR\}$ is the appropriate simplicial depth for the considered model.

Finally, confidence intervals can be constructed.

Corollary 27. *Under the assumptions of Theorem 25 asymptotic $(1-\alpha)$ confidence regions for θ in models (9) and (10) are given by*

$$\hat{\Theta}^{mod}(y) = \left\{ \theta \in \mathbb{R}^2 \middle| N\left(d_S^{mod}(\theta, y) - \frac{1}{4} \right) \geq q_G(\alpha) \right\}. \tag{29}$$

Thereby $d_S^{mod}(\theta, y), mod \in \{ARi, nAR\}$ is the appropriate simplicial depth for the considered model.

In case of the two parameter model a property of neighbouring maximising regions, as in the autoregressive process without intercept, does not hold. Note that we use the term neighbouring instead of connected, since the straights dividing these regions are given by the roots of the residuals and do not have the same depth as the enclosed regions. A statement under extremely strict assumptions is presented in the following lemma. First, we introduce the convex hull following Luenberger (1969).

Definition 28. *(see Luenberger, 1969, p. 17 and p.18)*
i) A set $K \subseteq V$ in a linear vector space V is said to be convex, if given $x_1, x_2 \in K$,

all points of the form $\alpha x_1 + (1-\alpha)x_2$ *for* $\alpha \in [0,1]$ *are in* K.
ii) Let $S \subseteq V$ *be an arbitrary set in a linear vector space* V. *The convex hull of* S, *denoted by* $conv(S)$, *is the smallest convex set* K *containing* S. *In other words,* $conv(S)$ *is the intersection of all convex sets* $K \subseteq V$ *containing* S.

Lemma 29. *Let* $y = (y_0, ..., y_N)$ *be an observation from an autoregressive process with intercept given by (9). Assume that an intercept condition,* $y_n > y_m$ *for all* $n > m$, $\theta_1, \theta_3 > 0$, *an intersection condition,*

$$\frac{y_n - y_m}{y_{n-1} - y_{m-1}} > \frac{y_n - y_k}{y_{n-1} - y_{k-1}}$$

for $m < k < n$ *and a root condition*

$$\frac{y_n}{y_{n-1}} < \frac{y_m}{y_{m-1}}$$

for all $m < n$ *hold. Then* $d_S^{ARi}(\theta, y)$ *has a connected maximising parameter region given by*

$$\Theta_{max} \in int\ conv\left(\theta_{i,j}, i \in \left\{\frac{2N}{3}+1, \frac{2N}{3}\right\}, j \in \left\{\frac{N}{3}, \frac{N}{3}+1\right\}\right)$$

for $N \in 3\mathbb{N}$,

$$\begin{aligned}\Theta_{max} \in\ & int\ conv\left(\theta_{i,j}, i \in \left\{\frac{2N+1}{3}+1, \frac{2N+1}{3}\right\}, j \in \left\{\frac{N-1}{3}, \frac{N-1}{3}+1\right\}\right) \\ & \cup\ int\ conv\left(\theta_{i,j}, i \in \left\{\frac{2N+1}{3}+1, \frac{2N+1}{3}\right\}, j \in \left\{\frac{N+2}{3}, \frac{N+2}{3}+1\right\}\right) \\ & \cup\ int\ conv\left(\theta_{i,j}, i \in \left\{\frac{2N-2}{3}+1, \frac{2N-2}{3}\right\}, j \in \left\{\frac{N-1}{3}, \frac{N-1}{3}+1\right\}\right)\end{aligned}$$

for $N \in 3\mathbb{N}+1$,

$$\begin{aligned}\Theta_{max} \in\ & int\ conv\left(\theta_{i,j}, i \in \left\{\frac{2N-1}{3}+1, \frac{2N-1}{3}\right\}, j \in \left\{\frac{N+1}{3}, \frac{N+1}{3}+1\right\}\right) \\ & \cup\ int\ conv\left(\theta_{i,j}, i \in \left\{\frac{2N-1}{3}+1, \frac{2N-1}{3}\right\}, j \in \left\{\frac{N-2}{3}, \frac{N-2}{3}+1\right\}\right) \\ & \cup\ int\ conv\left(\theta_{i,j}, i \in \left\{\frac{2N+2}{3}+1, \frac{2N+2}{3}\right\}, j \in \left\{\frac{N+1}{3}, \frac{N+1}{3}+1\right\}\right)\end{aligned}$$

for $N \in 3\mathbb{N}+2$,

whereby $\theta_{i,j} = \{\theta | r_i(\theta, y) = r_j(\theta, y)\}$ *and* $int\ conv(x_1, ..., x_N)$ *denotes the interior of*

the convex hull, spanned by the points $x_1, ..., x_N$.

Proof. First, note that each tupelo of subsequent observations $z_n = (y_n, y_{n-1})$ defines a line $\theta_3 = a \cdot \theta_1 + b$ which separates parameters leading to negative and positive values of the respective residual $r_n(\theta, y)$. These lines can be calculated by solving $r_n(\theta, y) = 0$ for $n \in \{1, ..., N\}$ leading to

$$\begin{aligned} & y_n - y_{n-1} - \theta_1 y_{n-1} - \theta_3 = 0 \\ \Leftrightarrow & \theta_3 = y_n - y_{n-1} - \theta_1 y_{n-1} \\ \Leftrightarrow & \theta_3 = y_n - \kappa y_{n-1}, \end{aligned}$$

with $\kappa := \theta_1 + 1$. An example based on 3 residuals is depicted in Figure 5.

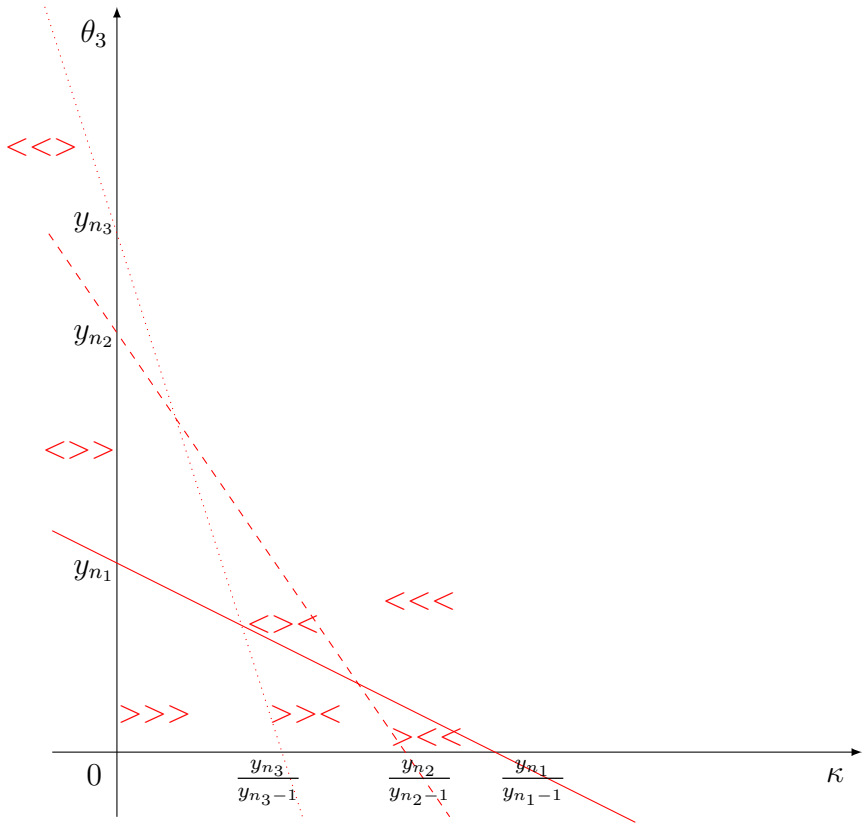


Fig. 5: Residual signs for the linear AR(1) growth model with intercept.

One can easily see that the lines intersect the θ_3 axis at y_n and the κ axis at $\frac{y_n}{y_{n-1}}$. Due to the growth assumption, the intersections on the θ_3 axis are ordered by the indexes of the respective residuals. Further, the intersections on the κ axis, have to be ordered in reverse, due to our assumptions. Based on these lines, we have

positive residuals for parameters below the respective line and negative residuals for parameters above. The general idea to find the maximising region is depicted in Figures 6,7 and 8. Since the lines are ordered with respect to their index n, we see that for parameters above all lines we just observe negative residuals. For parameters below all lines, we only have positive residuals. All other combinations appear, if the lines intersect. Due to the ordering, the first intersection of the N line appears, when the $r_N(\theta, y)$ line intersects the $r_{N-1}(\theta, y)$ line. This changes the last residual to a positive one, leaving all remaining residuals negative. By crossing the remaining lines $r_{N-2}(\theta, y)$,...,$r_1(\theta, y)$, the residuals are subsequently turned to positive ones. Due to the assumption

$$\frac{y_n - y_m}{y_{n-1} - y_{m-1}} > \frac{y_n - y_k}{y_{n-1} - y_{k-1}}$$

for $m < k < n$, $r_{N-1}(\theta, y)$ can just intersect the remaining lines $r_{N-2}(\theta, y), ..., r_1(\theta, y)$, if it crossed the $r_N(\theta, y)$ line already. Hence, the last residual is fixed with a positive sign for all further combinations caused by a sign change of $r_{N-1}(\theta, y)$. Continuing this over all residuals, the only possible sign structures for growth processes are given by k positive signs, followed by j negative signs and completed by $N-k-j$ positive signs again. This has two implications. First of all, there is a boundary region defined by the r_N line and the region where all residuals are positive, where no sign changes appear. Hence, on an infinite boundary region, depth is zero. Further, we can determine the regions where depth can be maximal by the sign change order. To do this, we maximise the function

$$f(k, j) = kj(N - k - j),$$

for $k, j \in \mathbb{N}$ and $k + j < N$. We start with

$$f(k, j) = kjN - k^2 j - kj^2$$

and the derivatives

$$\frac{\partial f}{\partial k} = jN - 2kj - j^2,$$
$$\frac{\partial f}{\partial j} = kN - 2kj - k^2.$$

These partial derivatives are 0, if and only if

$$\begin{aligned} jN - j^2 &= kN - k^2 \\ \Leftrightarrow j(N-j) &= k(N-k). \end{aligned}$$

Since $j, k, N \in \mathbb{N}$ this can only hold for $k = j$ or $k = (N - j)$. We can exclude $k = (N - j)$, since then $f(N - j, j) = 0$. Hence, we can reduce the function and maximise $\tilde{f}(k) = k^2(N - 2k)$ instead. We now have

$$\frac{\partial \tilde{f}}{\partial k} = 2kN - 6k^2,$$

what is 0, if and only if $k = 0$ or $k = \frac{N}{3}$. The second derivative proves that for $k = 0$ we have got a minimum and for $k = \frac{N}{3}$ a maximum. Since $k, j, (N - k - j)$ have to be natural numbers, the finite sample solution needs to by analysed with respect of the divisibility of N by 3.

The case $N \in 3\mathbb{N}$ is illustrated in Figure 6. Here the optimal solution is given by $k = \frac{N}{3} = j \in \mathbb{N}$. Hence $i = N - j - k = \frac{N}{3} \in \mathbb{N}$, too. We get an unique division of the residuals in a group of $N/3$ positive signs, followed by $N/3$ negative signs and completed by $N/3$ positive signs again. The respective region is bounded by the lines from the residuals at positions $\frac{2N}{3} + 1, \frac{2N}{3}, \frac{N}{3} + 1$ and $\frac{N}{3}$. Leaving this region violates the optimality criterion and reduces depth. This implies that we have got a neighbouring maximising region given by the claimed formula, since such a segmentation is only possible on this region under the stated assumptions.

In Figure 7, the situation for $N \in 3\mathbb{N} + 1$ is presented. Due to the discrete number of residuals, we cannot segment the residuals to groups with $N/3$ elements here. We either need to reduce k to the next integer or increase it likewise. Then possible divisions of residual signs are for example given by $(N-1)/3$ positive residuals followed by $(N + 2)/3$ negative residuals and completed by $(N - 1)/3$ residuals. However, each permutation of $\left(\frac{N-1}{3}, \frac{N+2}{3}, \frac{N-1}{3}\right)$ leads to the same depth. Further, all of these permutations can appear under our assumptions.

To clarify, if $k = j = \left\lfloor \frac{N}{3} \right\rfloor = (N-1)/3$ or $k = j = \left\lceil \frac{N}{3} \right\rceil = (N+2)/3$ should be used, we calculate depth in both cases.

For $k = j = (N - 1)/3$, depth is given by a constant scaling factor $c(N) = \left(3^3 \cdot \binom{N}{3}\right)^{-1}$ multiplied by $(N-1)^2 \cdot (N+2)$, since $i = N - k - j = (N + 2)/3$. For $k = j = (N + 2)/3$ depth is calculated by $c(N) \cdot (N + 2)^2(N - 4)$. Comparing the

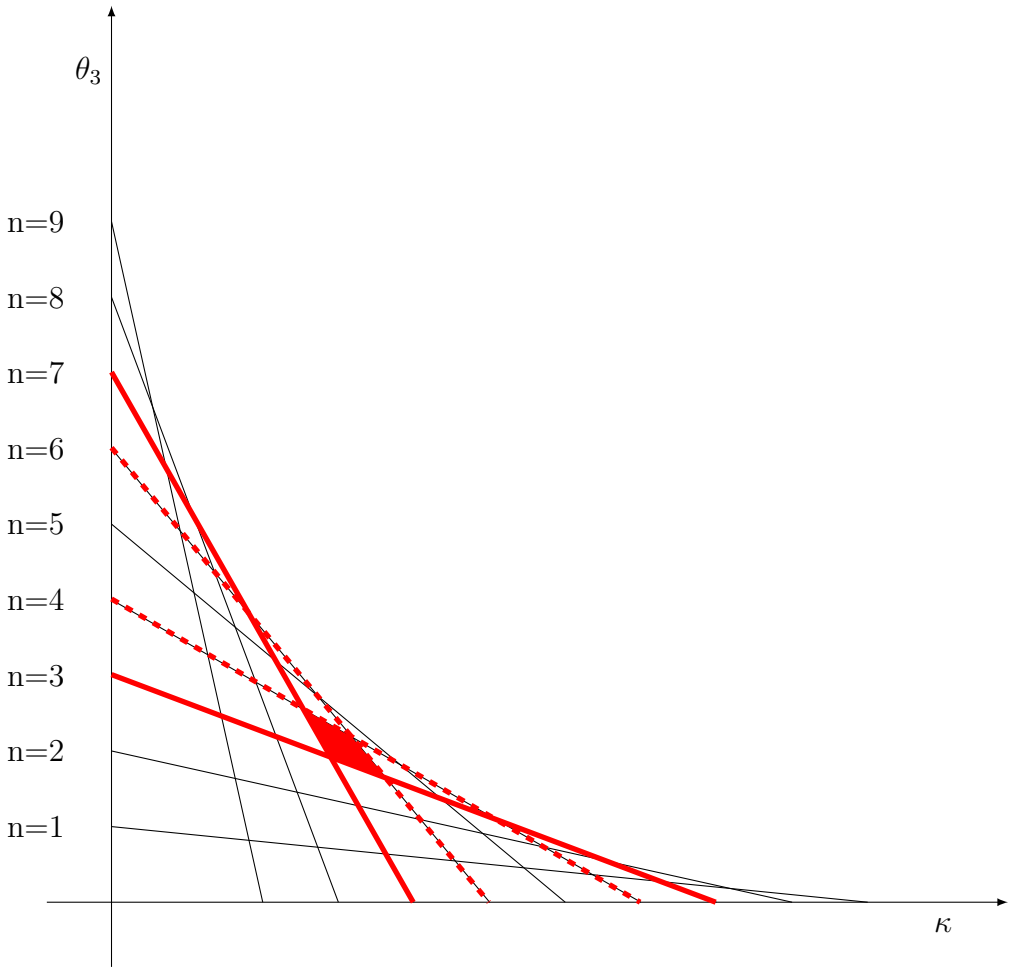


Fig. 6: Sign change example for $N = 9$.

polynomials gives

$$\begin{aligned}
&(N-1)^2(N+2) > (N+2)^2(N-4)\\
\Leftrightarrow &(N-1)^2 > (N+2)(N-4)\\
\Leftrightarrow &N^2-2N+1 > N^2-4N+2N-8\\
\Leftrightarrow &N^2-2N+1 > N^2-2N-8\\
\Leftrightarrow &1 > -8,
\end{aligned}$$

so that the solution given by $k = j = (N-1)/3$ is maximal for all $N \in \mathbb{N}$.
To analyse, if the resulting regions are neighbouring, we now observe that the maximising regions have to divide the residuals to permutations of $\left(\frac{N-1}{3}, \frac{N+2}{3}, \frac{N-1}{3}\right)$. There are three permutations and hence three possible sets which appear by evaluation of the intersects of the bounding residuals. For the first permutation $\left(\frac{N-1}{3}, \frac{N+2}{3}, \frac{N-1}{3}\right)$, the interesting residuals are $r_{\frac{N-1}{3}}(\theta, y), r_{\frac{N-1}{3}+1}(\theta, y)$, $r_{N-\frac{N-1}{3}+1}(\theta, y), r_{N-\frac{N-1}{3}}(\theta, y)$. Due to our assumptions, a sign change for the first two residuals from $r_{\frac{N-1}{3}}(\theta, y) > 0$ to $r_{\frac{N-1}{3}+1}(\theta, y) < 0$, combined with a sign from $r_{N-\frac{N-1}{3}}(\theta, y) < 0$ to $r_{N-\frac{N-1}{3}+1}(\theta, y) > 0$ is the only possible way to observe such a division. The region bounded by this restriction hence gives a parameter region with maximal depth.
Since we have two further permutations, the combinations of their bounding residuals give two more regions with maximal depth. These regions are automatically neighbouring, since the central region, given by the first permutation, can be reached from the two remaining regions by changing of one residual only.
The situation for $N \in 3\mathbb{N}+2$ is presented in Figure 8.
Similar calculations as in the $N \in 3\mathbb{N}+1$ case show that here the possible regions are defined by permutations of $\left(\frac{N+1}{3}, \frac{N-2}{3}, \frac{N+1}{3}\right)$. Again, three regions with maximal depth appear which are neighbouring through a central region.

□

Remark 30. *By solving the equations in θ, the values $\theta_{i,j}$ defining the edges of the maximising regions can be calculated explicitly. This is presented in Section 4.3. Further note that empirical depth coincides with the extremal value of the continuous optimisation function, if $N \in 3\mathbb{N}$ holds. Otherwise, maximal empirical depth is given by*

$$c(N) \cdot ((N-1) \cdot (N+2) \cdot (N-1)) \;\; for \;\; N \in 3\mathbb{N}+1$$

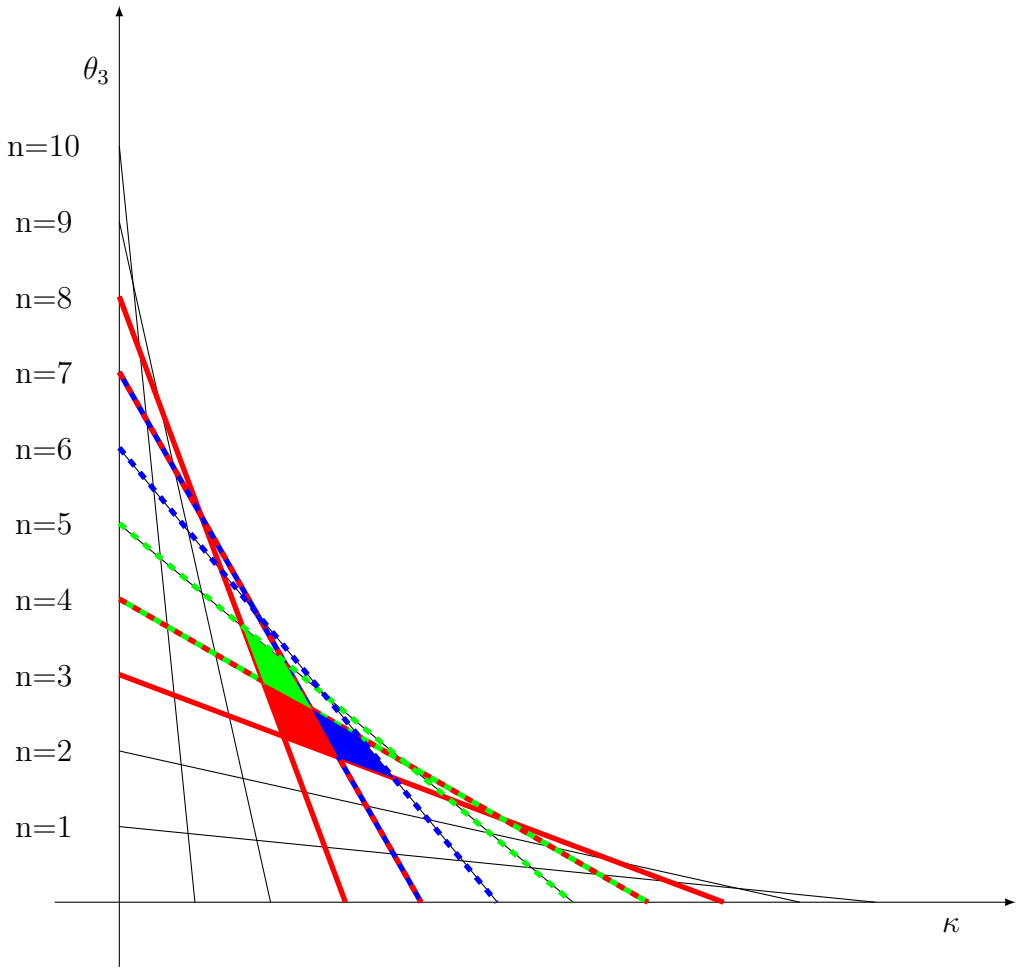


Fig. 7: Sign change example for $N = 10$.

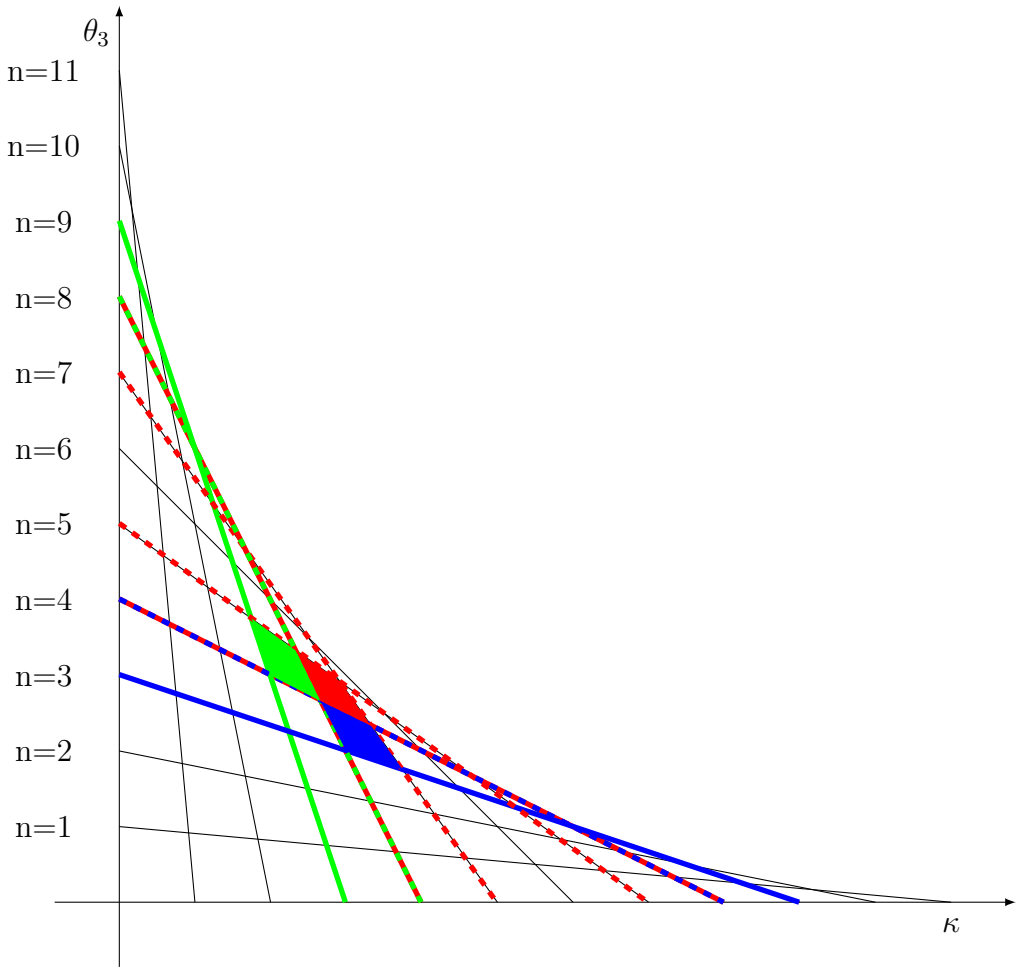


Fig. 8: Sign change example for $N = 11$.

or

$$c(N) \cdot ((N+1) \cdot (N-2) \cdot (N+1)) \;\; for \; N \in 3\mathbb{N}+2,$$

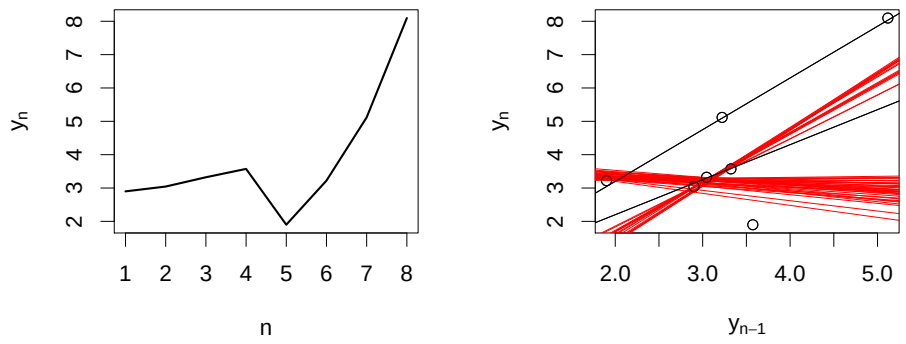
whereby $c(N) = \left(3^3 \cdot \binom{N}{3}\right)^{-1}$.

In contradiction to the one parameter case, where the order of the observation vector was not important, here the maximising regions can change, if the ordering does not hold. If the process is defined by multiple parameters, we also can observe multiple maximising regions. The second case for example appears, if the parameters change within the observed process.

Remark 31. *Simplicial depth for the autoregressive process with intercept, defined by (9), can have several local maxima, if the model assumptions are violated.*

The reason for the violation of this property is the additional degree of freedom introduced by the intercept parameter. In comparison to the one parameter model the fits can now be shifted in addition to a tilting. Further, the simplicial depth statistics are not symmetric. Hence, the effect of a change in θ does not only depend on simple intervals defined by the parameter but on regions defined by multiple parameters. The next example shows that in case of inhomogeneous models depth now shows multimodal contours. In particular, this shows that simplicial depth now is not a proper depth notion as defined by Zuo and Serfling (2000).

Example 32. *In this Example, we show the validity of Remark 31. Consider a realisation of an autoregressive process which changes its parameters from a growth parameter close to* $\kappa = 1.05$ *to a parameter close to* $\kappa = 1.55$ *and* $\theta_3 = 0.1$*, whereby* $\kappa = \theta_1 + 1$*, interrupted by a single downward jump between these two phases, as shown in Figure 9 (a).*
The simulated data is given by $y = (2.90, 3.04, 3.32, 3.57, 1.90, 3.22, 5.12, 8.10)^\top$*. The process starts in* $y_0 = 2.9$ *and jumps down to* $y_5 = 1.9$ *at the change point. For the independent errors, we used* $\mathcal{N}(0, 0.01)$ *distribution. In Figure 9 (b), we show the values of* y_n *against* y_{n-1}*. We observe two clusters of points. They can be identified by the parameters from the two different phases. Fits from these parameters,* $\theta_a = (1.05, 0.1)^\top$ *and* $\theta_b = (1.55, 0.1)^\top$ *are represented by the black lines in Figure 9 (b). The red lines are fits from 50 randomly drawn parameters which are in the regions with maximal depth.*



(a) Simulated process

(b) y_n against y_{n-1}

Fig. 9: Fits of an autoregressive process with two phases and downward jump. The left figure shows the observed process. The right figure depicts y_n versus y_{n-1}, the true parameter fits, represented by the black lines and fits from the regions with maximal depth, shown by the red lines.

In Figure 10, we show the lines defined by $r_n(\theta, y) = 0$, *i.e.* $\theta_3 = y_n - \kappa y_{n-1} = y_n - y_{n-1} - \theta_1 y_{n-1}$ *and evaluate the empirical depth on a grid defined by* $[-2, 4]^2$ *with accuracy* 0.001. *The lines, set by the roots of the residuals, are not ordered, since the process is not monotonically increasing. Further two potential regions with high depth appear.*

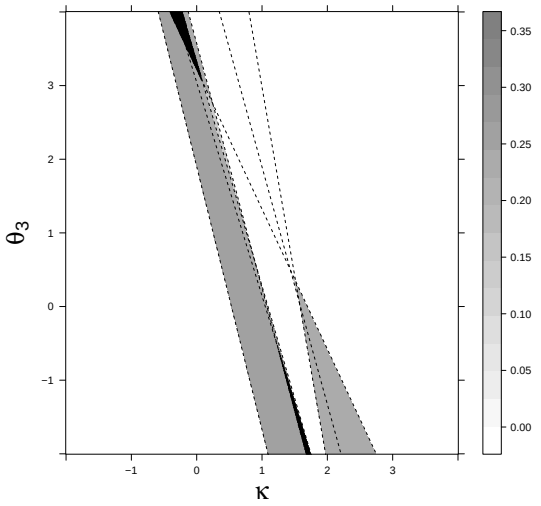


Fig. 10: Depth contour of the process depicted in Figure 9. The dashed lines show the parameter constellations for which the residuals are zero. The black regions are the regions with maximal depth.

Depth is maximised in two regions which are not connected. This happens, since the different phases and the downward jump change the residual order. This can in particular be seen in Figure 9 (b), where the red lines are given by $y_n = \kappa y_{n-1} + \theta_3$ *and* (κ, θ_3) *are randomly drawn parameters with maximal depth. It is interesting to mention that even if the regions are not connected the value of depth in this regions is* $0.3428571 = \left(\frac{7-1}{3} \cdot \frac{7+2}{3} \cdot \frac{7-1}{3}\right) \cdot \frac{1}{\binom{7}{3}}$ *and hence really maximal. Further, the parameter region with maximal depth for* $\theta_3 < 0$ *is not bounded, while the region for* $\theta_3 > 0$ *is.*

In case of the non-linear autoregressive process the results on a connected maximising region are even more restrictive than in the linear case. The statement can be found in the following Lemma.

Lemma 33. *The simplicial depth for the non-linear autoregressive process is maximal at a connected region, if* $y_n > y_{n-1}$, $y_m - y_{m-1} > y_n - y_{n-1}$ *for all* $m > n$ *and the conditions for the order of potential candidates hold for the linearisation*

$$log\left(\frac{y_n}{y_{n-1}} - 1\right) - \tilde{\theta}_2 log\left(y_{n-1}\right) = \tilde{\theta}_1,$$

where $\tilde{\theta}_1 = log(\theta_1)$ *and* $\tilde{\theta}_2 = (\theta_2 - 1)$. *Thereby, the conditions translate to an intercept condition*

$$log\left(\frac{y_n}{y_{n-1}} - 1\right) > log\left(\frac{y_m}{y_{m-1}} - 1\right) \ for\ all\ m < n,$$

an intersection condition

$$\frac{log\left(\frac{y_n}{y_{n-1}}-1\right)-log\left(\frac{y_m}{y_{m-1}}-1\right)}{log\,(y_{n-1})-log\,(y_{m-1})} > \frac{log\left(\frac{y_n}{y_{n-1}}-1\right)-log\left(\frac{y_k}{y_{k-1}}-1\right)}{log\,(y_{n-1})-log\,(y_{k-1})}\ for\ all\ m<k<n,$$

and a root condition

$$\frac{log\left(\frac{y_n}{y_{n-1}}-1\right)}{log\,(y_{n-1})} < \frac{log\left(\frac{y_m}{y_{m-1}}-1\right)}{log\,(y_{n-1})}\ for\ all\ m<n.$$

Proof. The proof simply follows by the arguments of Lemma 29 by observing that

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

can be translated to residual lines by

$$log\left(\frac{y_n}{y_{n-1}}-1\right) - (\theta_2 - 1)log(y_{n-1}) = log(\theta_1).$$

Under the growth conditions, we again introduce conditions, so that these lines are ordered by increasing intercepts and decreasing slopes, as in the linear case and hence can prove the assertion on neighbouring maximising regions. □

As in the linear case simple counterexamples for this property can be constructed, for example, if the underlying process has changing parameters.

Remark 34. *Even if we have got restricted possible sign structures for growth processes, all of our depth notions d_S^{AR}, d_S^{ARi} and d_S^{nAR} also indicate maximal depth for completely alternating signs. So far, we cannot prove this assertion, but in case of real world applications this case appears often. In these cases, we cannot claim that depth is maximised at a unique connected region, but empirical results show that this implication is very likely, as long as a process with long term growth and homogeneous parameters is observed. The proof remains an objective for future research.*

3.3 Simplified Depth Notions

Especially when constructing confidence regions, a problem in applying simplicial depth statistics is the computation time. A single evaluation of the full depth notion for the two parameter case is of order N^3. Since our confidence intervals are formulated in a general form, which requires the evaluation of depth on a parameter grid for an application, it is desirable to define a reasonable grid based on the

available observations. This restricts the evaluations of the depth statistic and improves the computational efficiency of our methods. A data based grid will be discussed in the Section 4.3 which deals with the implementation. In advance, in Section 4.1 we describe an algorithm, which is more efficient than a straightforward calculation of simplicial depth, to allow a fast computation of the statistics. From a theoretical point of view, we will reduce the computational costs from a third direction. We define simplified simplicial depth statistics which are closely related to the full simplicial depth expressions d_S^{AR}, d_S^{ARi}, and d_S^{nAR}. By derivation of the limit distributions, we propose estimators, tests and confidence intervals. Since the confidence intervals via full simplicial depth often are subsets of the intervals based on the simplified notions, we will be able to restrict the candidate set by a remarkably faster evaluation of the simplified statistics in advance of the calculation based on the full depth notions. The simplified depth notions will work for both cases, the one and two parameter models, as well as for a full model with $\theta \in \mathbb{R}^3$. In Kustosz et al. (2016b), we show that even more complex models can be handled by these depth notions.
After introducing the new depth notions d_S^1, d_S^2 and d_S^3, we will refer to them as **simplified simplicial depth** in this thesis. The formulas for the simplified depth notions are given in a general form by setting $Z_n = (X_n, Y_n), z_n = (x_n, y_n), \theta \in \mathbb{R}^K$ and $r_n(\theta, z)$ being the residuals of the respective models. This, for example, also allows the application of the statistics to general regression models. Remind that in case of autoregressive processes, we set $z_n = (y_n, y_{n-1})$. Further, $d_S^i(\theta, z)$ then can be considered as $d_S^i(\theta, y)$ for $i \in \{1, 2, 3\}$. However, to allow a more general application, we use the z notation in this section.

d_S^1 - Nonoverlapping Subsequent Subsets

The first simplified depth statistic can be defined by

$$d_S^1(\theta, z) := \frac{1}{\left\lfloor \frac{N}{K+1} \right\rfloor} \sum_{n=1}^{\left\lfloor \frac{N}{K+1} \right\rfloor} \left(\prod_{k=1}^{K+1} \mathbb{1}_{\left\{ r_{(K+1)(n-1)+k}(\theta,z)(-1)^k > 0 \right\}} + \prod_{k=1}^{K+1} \mathbb{1}_{\left\{ r_{(K+1)(n-1)+k}(\theta,z)(-1)^{k+1} > 0 \right\}} \right).$$

Instead of evaluating all possible permutations of the ascending indexed residuals, now just subsequent residuals without overlapping are evaluated. This, in particular, avoids dependent terms in the statistic. An example is given in Figure 11. This

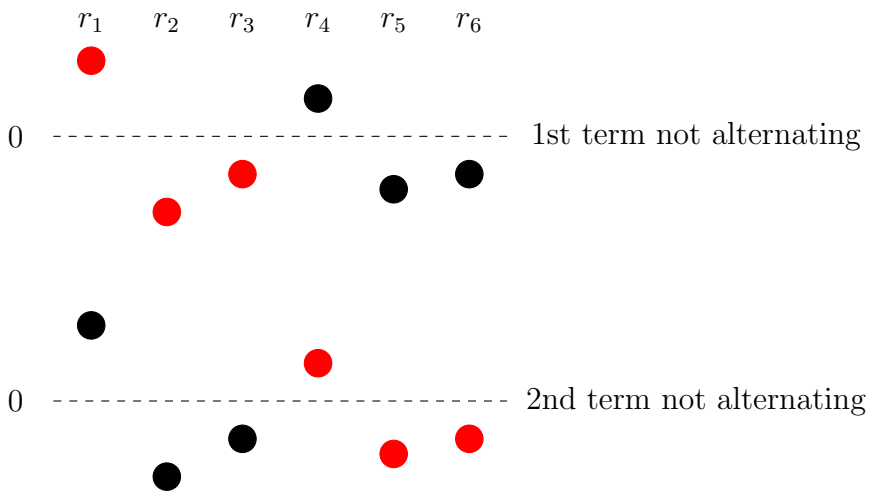


Fig. 11: Illustration of the d_S^1 statistic.

statistic obviously is a part of the full simplicial depth statistic. In the figure, the evaluation for a sample with seven observations for a two dimensional parameter is depicted. We hence have $N = 6$ residuals and $K = 2$. So just $\lfloor 6/3 \rfloor = 2$ terms have to be evaluated. Since the relevant groups have no sign changes, the resulting d_S^1 depth in this example is zero. A big advantage of this statistic is that the limit distribution under the true parameter can be derived easily when independent errors are assumed.

Theorem 35. *If $\theta^0 \in \mathbb{R}^K$ is the true parameter with $\mathbb{P}_{\theta^0}(r_n(\theta^0, Z) > 0) = \frac{1}{2} = \mathbb{P}_{\theta^0}(r_n(\theta^0, Z) < 0)$ for all $n = 1, ..., N$ then*

$$\sqrt{\left\lfloor \frac{N}{K+1} \right\rfloor} \frac{d_S^1(\theta^0, Z) - (\frac{1}{2})^K}{\sqrt{(\frac{1}{2})^K(1 - (\frac{1}{2})^K)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

for $N \to \infty$.

Proof. Note that $r_n(\theta, Z) = E_n$ holds, if θ is the underlying parameter.
Set

$$V_n := \prod_{k=1}^{K+1} \mathbb{1}_{\{r_{(K+1)(n-1)+k}(\theta,Z)(-1)^k > 0\}} + \prod_{k=1}^{K+1} \mathbb{1}_{\{r_{(K+1)(n-1)+k}(\theta,Z)(-1)^{k+1} > 0\}}.$$

Then V_n, $n \in \{1, \ldots, \lfloor \frac{N}{K+1} \rfloor\}$, are independent variables with Bernoulli distribution satisfying $\mathbb{P}(V_n = 1) = (1/2)^K$, so that the assertion follows from the central limit theorem (CLT), see e.g. van der Vaart (2007), Proposition 2.17, p. 16. □

Since the limit distribution is well known, we can define asymptotic tests and con-

fidence intervals. This test can be applied in all situations in which a simplicial depth based on a K dimensional parameter is appropriate. For our application, we just define the resulting test for the three autoregressive models. Note that in this situation we have $N+1$ observations and N residuals and z can be replaced by y.

Corollary 36. *An asymptotic α-level test for $H_0 : \theta \in \Theta^0$ for model (8) under the conditions of Theorem 14 or models (9) and (10) under the conditions of Theorem 25, is defined by $\varphi : \mathbb{R}^{N+1} \to \{0,1\}$ with*

$$\varphi(y) = \begin{cases} 1 & , if \ \sup\limits_{\theta \in \Theta^0} \sqrt{\left\lfloor \frac{N}{K+1} \right\rfloor} \frac{d_S^1(\theta, y) - (\frac{1}{2})^K}{\sqrt{(\frac{1}{2})^K (1 - (\frac{1}{2})^K)}} < q_{\mathcal{N}(0,1)(\alpha)} \\ 0 & else \end{cases}, \tag{30}$$

whereby $q_{\mathcal{N}(0,1)}(\gamma)$ is the γ quantile of the Standard Normal distribution.
Thereby, $r_n(\theta, y)$ is defined by the appropriate model residuals and $\Theta_0 \subset \mathbb{R}^K$.

The confidence intervals can be directly constructed by the following statement.

Corollary 37. *Under the assumptions of Corollary 36, an asymptotic $(1-\alpha)$ confidence region for θ is given by*

$$\hat{\Theta}^1(y) = \left\{ \theta \in \mathbb{R}^K \middle| \sqrt{\left\lfloor \frac{N}{K+1} \right\rfloor} \frac{d_S^1(\theta, y) - (\frac{1}{2})^K}{\sqrt{(\frac{1}{2})^K (1 - (\frac{1}{2})^K)}} \geq q_{\mathcal{N}(0,1)}(\alpha) \right\}, \tag{31}$$

whereby the residuals and the parameter dimension have to be defined as appropriate by the respective models.

It is also possible to calculate the exact distribution under H_0, if N is fixed and known. This simply follows from the arguments in the proof of Theorem 35. Instead of the application of the CLT, one can use the distribution of V_n to derive the distribution of the sum in the simplicial depth statistic.

Corollary 38. *If $\theta^0 \in \mathbb{R}^K$ is the true parameter with $\mathbb{P}_{\theta^0}(r_n(\theta, Z) > 0) = \frac{1}{2} = \mathbb{P}_{\theta^0}(r_n(\theta, Z) < 0)$ for all $n \in \{1, ..., N\}$ then*

$$\left\lfloor \frac{N}{K+1} \right\rfloor d_S^1(\theta, Z) \sim Bin\left(\left\lfloor \frac{N}{K+1} \right\rfloor, (1/2)^K \right).$$

Proof. After rescaling with $\left\lfloor \frac{N}{K+1} \right\rfloor$, we just have a sum of independent $Bin(1, (1/2)^K)$ random variables. Hence, the sum is $Bin\left(\left\lfloor \frac{N}{K+1} \right\rfloor, (1/2)^K \right)$ distributed. □

By application of this corollary, exact tests for the simplified depth notion can be constructed.

d_S^2 - Overlapping Subsequent Subsets

If it is desirable to account for dependency in the simplicial depth statistic, another version can be defined by

$$d_S^2(\theta, z) := \frac{1}{N-K} \sum_{n=1}^{N-K} \left(\prod_{k=1}^{K+1} \mathbb{1}_{\{r_{n-1+k}(\theta,z)(-1)^k > 0\}} + \prod_{k=1}^{K+1} \mathbb{1}_{\{r_{n-1+k}(\theta,z)(-1)^{k+1} > 0\}} \right).$$

This statistic again uses blocks of $K+1$ residuals for a K dimensional parameter but allows overlapping residual blocks. An example is given by Figure 12. In contrast to the first proposal, now a simple CLT is not applicable. But since we consider subsequent blocks of residuals the dependency is limited and hence a modified CLT can be applied.

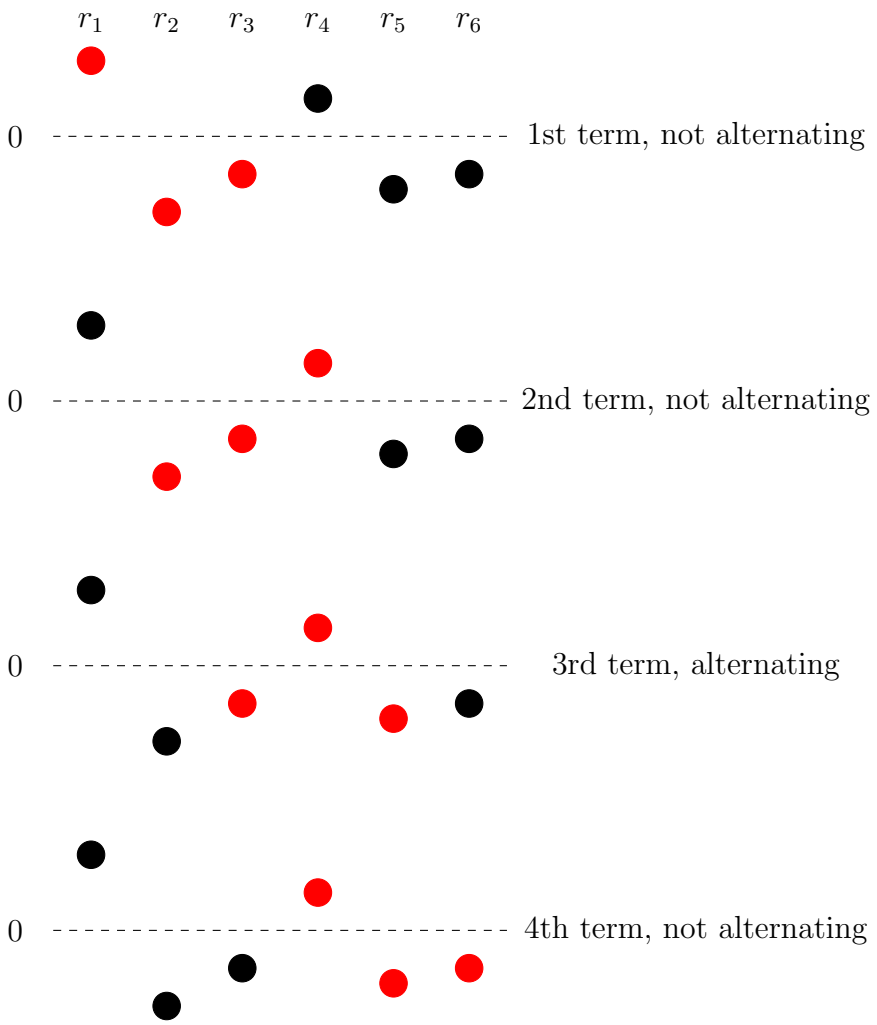


Fig. 12: Illustration of the d_S^2 statistic.

The example with $N = 6$ residuals shows that we now have $N - 2 = 4$ terms which are checked in the test statistic. The third term consists of residuals with alternating signs, so that we have a contribution to d_S^2 depth here. Hence, the resulting d_S^2 depth is $\frac{1}{4}$.

Theorem 39. *If $\theta^0 \in \mathbb{R}^K$ is the true parameter with $\mathbb{P}_{\theta^0}(r_n(\theta^0, Z) > 0) = \frac{1}{2} = \mathbb{P}_{\theta^0}(r_n(\theta^0, Z) < 0)$ then*

$$\sqrt{N-K}\frac{d_S^2(\theta^0, Z) - \left(\frac{1}{2}\right)^K}{\sqrt{(\frac{1}{2})^K \cdot [3 - (\frac{1}{2})^{K-1} \cdot K - 3 \cdot (\frac{1}{2})^K]}} \xrightarrow{d} \mathcal{N}(0,1),$$

for $N \to \infty$.

Proof. Note that $r_n(\theta, Z) = E_n$ holds, if θ is the underlying parameter.

Set

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

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

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

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

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

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

by

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

□

As for the first simplified statistic the limit distribution can be used to define tests and confidence regions for all considered models.

Corollary 40. *An asymptotic α-level test for $H_0 : \theta \in \Theta^0$ for model (8) under the conditions of Theorem 14 or models (8) and (10) under the conditions of Theorem 25 is defined by $\varphi : \mathbb{R}^N \to \{0, 1\}$ with*

$$\varphi(y) = \begin{cases} 1 & , if \ \sup\limits_{\theta \in \Theta^0} \sqrt{\left\lfloor \frac{1}{N-K} \right\rfloor} \frac{d_S^2(\theta, y) - \left(\frac{1}{2}\right)^K}{\sqrt{(\frac{1}{2})^K \cdot [3 - (\frac{1}{2})^{K-1} \cdot K - 3 \cdot (\frac{1}{2})^K]}} < q_{\mathcal{N}(0,1)(\alpha)} \\ 0 & else \end{cases}, \tag{32}$$

whereby $q_{\mathcal{N}(0,1)}(\gamma)$ is the γ quantile of the Standard Normal distribution. Thereby, $r_n(\theta, y)$ is defined by the appropriate model residuals and $\Theta \subset \mathbb{R}^K$.

Analogously, we get the confidence regions.

Corollary 41. *Under the assumptions of Corollary 40, an asymptotic $(1 - \alpha)$ confidence region for θ is given by*

$$\hat{\Theta}^2(y) = \left\{ \theta \in \mathbb{R}^K \middle| \sqrt{\left\lfloor \frac{1}{N-K} \right\rfloor} \frac{d_S^2(\theta, y) - \left(\frac{1}{2}\right)^K}{\sqrt{(\frac{1}{2})^K \cdot [3 - (\frac{1}{2})^{K-1} \cdot K - 3 \cdot (\frac{1}{2})^K]}} \geq q_{\mathcal{N}(0,1)}(\alpha) \right\}, \tag{33}$$

whereby the residuals and the parameter dimension has to be defined as appropriate.

Here, the calculation of an exact distribution of the test statistic is a harder task than for independent terms. We again can state that the terms in d_S^2 are Bernoulli variables, but the summation to Binomial variables is not straightforward, since the terms are overlapping and hence dependent. The main problem is a combinatorial task which increases in the number of observations.
A solution is so far not available. To allow tests based on the exact distribution, we hence deliver approximations of the exact distributions instead which are available in the `rexpar` package. This approximations are based on the simple fact that the residuals have median zero under H_0. We simulate residuals based on independent Bernoulli distributed variables on $\{-1, 1\}$ and calculate the test statistic based on these residuals. By a sufficient number of repetitions, the resulting values of the test statistic approximates the exact distribution of d_S^2 under H_0. This distribution clearly differs from a binomial distribution, as can be seen in Example 42.

Example 42. *We approximate the exact distribution of* $(N-2)d_S^2$ *for* $K = 2, N = 1000$. *Figure 13 shows the resulting approximaton and the Binomial distribution with* $N-2$ *and* $p = 1/4$ *which results, if we treated the terms in* d_S^2 *as independent.*

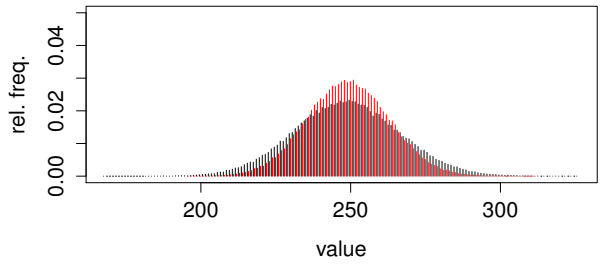


Fig. 13: Simulated exact distribution of the sum in the d_S^2 statistic for $N = 1000$ under H_0, depicted by the black lines and a $Bin(N-2, 1/4)$ distribution, depicted by the red lines.

Both distributions are centred around 250 what coincides with the expected value of $1/4$ *for the rescaled statistic. The exact statistic is flatter than a binomial distribution. Hence, a significant effect on the distribution caused by dependence is apparent. In particular, the critical values are influenced by the dependency.*

d_S^3 - Subsets with Fixed Point in Centre

The third simplification will be restricted to the two parameter case. The residual blocks considered here consist of one residual of the first half of the data, one residual

of the second half and the residual in the middle. To get conditional independence, we use the central residual in all blocks and all remaining residuals just once. The statistic is defined by

$$d_S^3(\theta, z) := \frac{1}{\left\lfloor \frac{N-1}{2} \right\rfloor} \sum_{n=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \left(\mathbb{1}_{\{r_n(\theta,z)>0\}} \mathbb{1}_{\left\{ r_{\left\lfloor \frac{N+1}{2} \right\rfloor}(\theta,z)<0 \right\}} \mathbb{1}_{\{r_{N-n+1}(\theta,z)>0\}} + \mathbb{1}_{\{r_n(\theta,z)<0\}} \mathbb{1}_{\left\{ r_{\left\lfloor \frac{N+1}{2} \right\rfloor}(\theta,z)>0 \right\}} \mathbb{1}_{\{r_{N-n+1}(\theta,z)<0\}} \right).$$

Here, the limit distribution again follows based on the CLT. An illustration is given in Figure 14.

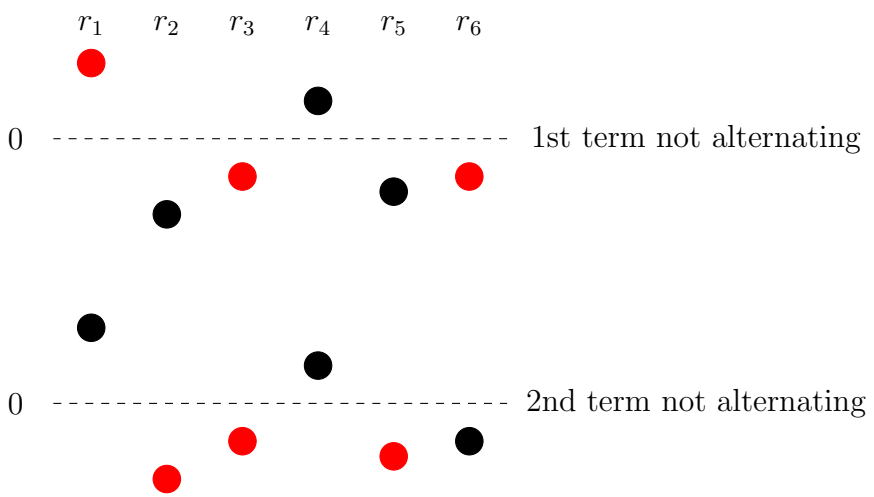


Fig. 14: Illustration of the d_S^3 statistic.

In the example, we have to evaluate $\lfloor 5/2 \rfloor = 2$ groups of residuals. Both groups do not have alternating signs. Hence, d_S^3 depth is zero.

Theorem 43. *If $\theta^0 \in \mathbb{R}^2$ is the true parameter with $\mathbb{P}_{\theta^0}(r_n(\theta^0, Z) > 0) = \frac{1}{2} = \mathbb{P}_{\theta^0}(r_n(\theta^0, Z) < 0)$ then*

$$\sqrt{\left\lfloor \frac{N-1}{2} \right\rfloor} \frac{d_S^3(\theta^0, Z) - \frac{1}{4}}{\sqrt{\frac{3}{16}}} \xrightarrow{d} \mathcal{N}(0,1)$$

for $N \to \infty$.

Proof of Theorem 43. Note that $r_n(\theta^0, Z) = E_n$ holds, if θ^0 is the true underlying parameter.
Set

$$\begin{aligned} V_n = &\mathbb{1}_{\{r_n(\theta,Z)>0\}} \mathbb{1}_{\left\{r_{\lfloor \frac{N+1}{2} \rfloor}(\theta,Z)<0\right\}} \mathbb{1}_{\{r_{N-n+1}(\theta,Z)>0\}} \\ &+\mathbb{1}_{\{r_n(\theta,Z)<0\}} \mathbb{1}_{\left\{r_{\lfloor \frac{N+1}{2} \rfloor}(\theta,Z)>0\right\}} \mathbb{1}_{\{r_{N-n+1}(\theta,Z)<0\}}. \end{aligned}$$

Again, V_n are Bernoulli variables, here with $\mathbb{P}_{\theta^0}(V_n = 1) = 1/4$. To apply the CLT, we need to assure independence of $V_1, ..., V_{\lfloor \frac{N-1}{2} \rfloor}$. At first, note that

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

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

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

and

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

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

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

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

The tests and confidence intervals can be formulated as follows.

Corollary 44. *An asymptotic α-level test for $H_0 : \theta \in \Theta^0$ for model (8) under the conditions of Theorem 14 or models (9) and (10) under the conditions of Theorem 25 is defined by $\varphi : \mathbb{R}^{N+1} \to \{0, 1\}$ with*

$$\varphi(y) = \begin{cases} 1 & , if \ \sup\limits_{\theta \in \Theta^0} \sqrt{\left\lfloor \frac{N-1}{2} \right\rfloor} \frac{d_S^3(\theta, y) - \frac{1}{4}}{\sqrt{\frac{3}{16}}} < q_{\mathcal{N}(0,1)(\alpha)} \\ 0 & else \end{cases}, \tag{34}$$

whereby $q_{\mathcal{N}(0,1)}(\gamma)$ is the γ quantile of the Standard Normal distribution.
Thereby, $r_n(\theta, z)$ is defined by the appropriate model residuals and $\Theta \subset \mathbb{R}^K$.

Remark 45. *Note that we just give the limit proof for $K = 2$ here. However, for each even K a similar statistic can be proposed. The proof of a limit theorem can be conducted analogously.*

The confidence sets can be constructed by the following corollary.

Corollary 46. *Under the assumptions of Corollary 44, an asymptotic $(1 - \alpha)$ confidence region for θ is given by*

$$\hat{\Theta}^3(y) = \left\{ \theta \in \mathbb{R}^K \middle| \sqrt{\left\lfloor \frac{N-1}{2} \right\rfloor} \frac{d_S^3(\theta, y) - \frac{1}{4}}{\sqrt{\frac{3}{16}}} \geq q_{\mathcal{N}(0,1)}(\alpha) \right\}, \tag{35}$$

whereby the residuals and the parameter dimension has to be defined as appropriate.

Again, we can calculate the exact distribution under H_0, if N is fixed and known by using the distribution of the terms in the statistic.

Corollary 47. *If $\theta^0 \in \mathbb{R}^K$ is the true parameter with $\mathbb{P}_{\theta^0}(r_n(\theta, Z) > 0) = \frac{1}{2} = \mathbb{P}_{\theta^0}(r_n(\theta, Z) < 0)$ for all $n \in \{1, ..., N\}$ then*

$$\left\lfloor \frac{N-1}{2} \right\rfloor d_S^3(\theta, Z) \sim Bin\left(\left\lfloor \frac{N-1}{2} \right\rfloor, (1/2)^K \right).$$

Proof. By the calculations in the proof of Theorem 43, we can rescale the statistic by $\left\lfloor \frac{N-1}{2} \right\rfloor$ to a sum of conditionally independent Bernoulli variables with identical distribution. Hence, the sum is a Binomial variable with the stated parameters. □

By application of this corollary, exact tests for the simplified depth notion can be constructed.

3.4 Consistency

In this section, we address the consistency of the proposed tests. Thereby, we will restrict on explosive processes, as necessary for our main application. The central results are discussed in Kustosz and Müller (2014), Kustosz et al. (2016a) and Kustosz et al. (2016b).
First of all, we need to define consistency for the depth based tests.

Definition 48. *A test $\varphi(\theta, y)$ for $H_0 : \theta \in \Theta^0$ is consistent at all $\theta^* \in \Theta \setminus \Theta^0$, if*

$$\lim_{N\to\infty} \mathbb{P}_{\theta^*}(\varphi(\theta^0, Y) = 1) = 1 \ \forall \theta^0 \in \Theta^0.$$

For full simplicial depth the consistency of the resulting tests can be derived as in Kustosz et al. (2016a). Therefore, we have to restrict the parameter space to the explosive case. We then can prove the assertion for specific one sided hypothesis and the one point null hypothesis. We start with the linear model without intercept defined by (8). Thereby, we consider $\kappa = \theta_1 + 1$ to simplify notation without loss of generality.

Theorem 49. *If the errors satisfy $E_n \geq y_0 - \tau y_0 + c \ \forall n$, $y_0 > 0, c > 0, \tau > 1, Y_0 = y_0, \Theta = [\tau, \infty), \theta_1^0 > \tau$ and we consider*

$$\Theta^0 = \{\theta_1 \in \Theta | \theta_1 \geq \theta_1^0\} \text{ or } \Theta^0 = \{\theta_1^0\},$$

then the test given by (16) is consistent at all $\theta^* \in \Theta \setminus \Theta^0$.

We need the following Lemma.

Lemma 50. *If the errors satisfy $E_n \geq y_0 - \kappa y_0 + c \ \forall n$, $y_0 > 0, c > 0, \theta > 1$ and $Y_0 = y_0$, then Y_n is strictly increasing with*

$$Y_n \geq \left(\sum_{k=0}^{n-1} \kappa^k \right) c + y_0$$

almost surely.

Proof. This easily follows by induction. For $n = 1$ we have

$$\begin{aligned} Y_1 =& \kappa Y_0 + E_1 \\ \geq& \kappa y_0 + y_0 - \kappa y_0 + c \\ =& c + y_0. \end{aligned}$$

Then

$$\begin{aligned}
Y_n =& \kappa Y_{n-1} + E_n \\
\geq& \kappa \left(\left(\sum_{k=0}^{n-2} \kappa^k \right) c + y_0 \right) + E_n \\
=& \left(\sum_{k=1}^{n-1} \kappa^k \right) c + \kappa y_0 + E_n \\
\geq& \left(\sum_{k=1}^{n-1} \kappa^k \right) c + \kappa y_0 + y_0 - \kappa y_0 + c \\
=& \left(\sum_{k=0}^{n-1} \kappa^k \right) c + y_0.
\end{aligned}$$

by induction. □

Now, we can prove the consistency.

Proof of Theorem 49. We have to show that

$$\lim_{N\to\infty} \mathbb{P}_{\theta^*} \left(\sup_{\theta\in\Theta^0} \left[N(d_S^{AR}(\theta, Y_*) - \frac{1}{2}) \right] \geq q(\alpha) \right) = 0,$$

whereby $q(\alpha)$ is the α quantile of the distribution given by $\frac{1}{2} - \frac{1}{2}X^2$ and $X \sim \mathcal{N}(0,1)$. $\mathbb{P}_{\theta^*}$ denotes the probability measure induced by the processes under θ^* and Y_* are the respective processes. We start with the one-sided hypothesis. The general idea is to apply the Chebychev inequality. To do this, some preliminary steps are necessary. First, note that

$$\begin{aligned}
&\sup_{\theta\in\Theta^0} \left[\mathbb{1}_{\{r_{n_1}(\theta,Y_*)>0, r_{n_2}(\theta,Y_*)<0\}} + \mathbb{1}_{\{r_{n_1}(\theta,Y_*)<0, r_{n_2}(\theta,Y_*)>0\}} \right] \\
\leq& \sup_{\theta\in\Theta^0} \left[\mathbb{1}_{\{r_{n_1}(\theta,Y_*)>0\}} + \mathbb{1}_{\{r_{n_2}(\theta,Y_*)>0\}} \right] \qquad (36) \\
\leq& \mathbb{1}_{\{r_{n_1}(\theta^*,Y_*)>(\theta^0-\theta^*)Y_{n_1-1}\}} + \mathbb{1}_{\{r_{n_2}(\theta^*,Y_*)>(\theta^0-\theta^*)Y_{n_2-1}\}}.
\end{aligned}$$

Under the conditions on the errors Y_n is a growth process. Hence, for each $\gamma > 0$, we can find N_0, such that

$$Y_n \geq (\sum_{k=0}^{n-1} \kappa^k)c + y_0 \geq \gamma$$

for all $n \geq N_0$,see Lemma 50. Since $\theta^0 > \theta^*$, we can select γ with $(\theta^0 - \theta^*)\gamma > \rho$ for

ρ satisfying $\mathbb{P}_{\theta^*}(r_n(\theta, Y_*) > \rho) < \frac{1}{4} - \delta$ for $\delta \in (0, \frac{1}{4})$. If we now define

$$H(r_{n_1}(\theta^*, Y_*), r_{n_2}(\theta^*, Y_*)) = \mathbb{1}_{\{r_{n_1}(\theta^*, Y_*) > \rho\}} + \mathbb{1}_{\{r_{n_2}(\theta^*, Y_*) > \rho\}},$$

we see that

$$\begin{aligned} &\sup_{\theta \in \Theta^0} \left[\mathbb{1}_{\{r_{n_1}(\theta, Y_*) > 0, r_{n_2}(\theta, Y_*) < 0\}} + \mathbb{1}_{\{r_{n_1}(\theta, Y_*) < 0, r_{n_2}(\theta, Y_*) > 0\}} \right] \\ \leq & H(r_{n_1}(\theta^*, Y_*), r_{n_2}(\theta^*, Y_*)) \end{aligned}$$

for all $n > N_0$. Further

$$\begin{aligned} & E_{\theta^*}[H(r_{n_1}(\theta^*, Y_*), r_{n_2}(\theta^*, Y_*))] \\ = & E_{\theta^*} \left[\mathbb{1}_{\{r_{n_1}(\theta^*, Y_*) > \rho\}} + \mathbb{1}_{\{r_{n_2}(\theta^*, Y_*) > \rho\}} \right] \\ = & \mathbb{P}(r_{n_1}(\theta^*, Y_*) > 0) + \mathbb{P}(r_{n_2}(\theta^*, Y_*) > 0) \\ < & \frac{1}{2} - 2\delta = \frac{1}{2} - \delta' \end{aligned} \tag{37}$$

for $n_1, n_2 > N_0$ and a $\delta, \delta' > 0$. To control the expected value and the variance term, we need to consider the terms for $n \leq N_0$ also. Hence, we define

$$M_0 = \{(n_1, n_2) | N_0 < n_1 < n_2\}$$

and use the fact that the terms of the full simplified depth statistic are already bounded by 1. Hence,

$$\begin{aligned} & \sup_{\theta \in \Theta^0} \left[N \left(d_S^{AR}(\theta, Y_*) - \frac{1}{2} \right) \right] \\ \leq & \frac{N}{\binom{N}{2}} \left(\left[\binom{N}{2} - \binom{N - N_0}{2} \right] + \sum_{(n_1, n_2) \in M_0} \left(H(r_{n_1}(\theta^*, Y_*), r_{n_2}(\theta^*, Z_*)) - \frac{1}{2} \right) \right) \\ := & T \end{aligned}$$

We now have

$$\begin{aligned} & E_{\theta^*}(T) \\ = & \frac{N}{\binom{N}{2}} \left(\left[\binom{N}{2} - \binom{N - N_0}{2} \right] + \sum_{(n_1, n_2) \in M_0} \left(E_{\theta^*}[H(r_{n_1}(\theta^*, Z_*), r_{n_2}(\theta^*, Z_*))] - \frac{1}{2} \right) \right). \end{aligned}$$

By (37), we can select $N_1 > N_0$ with

$$E_{\theta^*}(T) \leq -N\delta''$$

for a $\delta'' > 0$, since for N large enough

$$\frac{\binom{N-N_0}{2}}{\binom{N}{2}} \overset{N\to\infty}{\longrightarrow} 1$$

and

$$\frac{\binom{N-N_0}{2}}{\binom{N}{2} - \binom{N-N_0}{2}} \overset{N\to\infty}{\longrightarrow} 0.$$

In addition, we have

$$var_{\theta^*}(T) \leq \epsilon N^2 \left(\delta'' + \frac{q_\alpha}{N_2}\right)^2,$$

since H is bounded by 2 and for all $\epsilon > 0$ one can select $N_2 > N_1$, such that $\delta'' + \frac{q_\alpha}{N_2} > 0$ holds.
Now, we can apply the Chebychev inequality for all $N \geq N_2$ and get

$$\begin{aligned}
&P_{\theta^*}\left(\sup_{\theta\in\Theta_0}\left[N\left(d_S(\theta, Y_*) - \frac{1}{2}\right)\right] \geq q_\alpha\right) \\
&\leq \quad P_{\theta^*}\left(T \geq q_\alpha\right) \leq P_{\theta^*}\left(|T - E_{\theta^*}(T)| \geq q_\alpha - E_{\theta^*}(T)\right) \\
&\leq \quad P_{\theta^*}\left(|T - E_{\theta^*}(T)| \geq q_\alpha + N\,\delta''\right) \leq \frac{\epsilon N^2 \left(\delta'' + \frac{q_\alpha}{N_2}\right)^2}{N^2 \left(\delta'' + \frac{q_\alpha}{N}\right)^2} \leq \frac{\epsilon \left(\delta'' + \frac{q_\alpha}{N_2}\right)^2}{\left(\delta'' + \frac{q_\alpha}{N_2}\right)^2} = \epsilon.
\end{aligned}$$

For the point hypothesis the proof can be conducted analogously through defining H by

$$H(r_{n_1}(\theta^*, Y_*), r_{n_2}(\theta^*, Y_*)) = \mathbb{1}_{\{r_{n_1}(\theta^*,Y_*)>0, r_{n_2}(\theta^*,Y_*)<0\}} + \mathbb{1}_{\{r_{n_1}(\theta^*,Y_*)<0, r_{n_2}(\theta^*,Y_*)>0\}}$$

for $\theta^* > \theta^0$. □

Theorem 51. *If the errors satisfy $E_n \geq y_0 - \tau y_0 + c\ \forall n$, $y_0 > 0, c > 0, \tau > 1, Y_0 =$*

$y_0, \Theta = [0, \infty) \times [\tau, \infty), \theta_2^0 \geq 0, \theta_1^0 > \tau$ *and we consider*

$$\Theta^0 = \{(\theta_1, \theta_2)^\top \in \Theta | \theta_1 \geq \theta_1^0\} \text{ or } \Theta^0 = \{(\theta_1^0, \theta_2^0)\}$$

then the test given by (28) is consistent at all $\theta^* \in \Theta \setminus \Theta^0$.

Proof. The proof can be found in Kustosz et al. (2016a) and is analogue to the proof of Theorem 49. □

Since we want to work with the simplified statistics, the consistency of these tests has to be proven. In Kustosz et al. (2016b) a more general setting, allowing K-dimensional parameters and a wider range of models, is considered. The consistency for the methods in this thesis follows as corollary from these general results. We now regard a general model defined by

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

To assure consistency for explosive AR processes we have following theorem.

Theorem 52. *For each* $b \in \mathbb{R}$ *there exists* $N_0 \in \mathbb{N}$*, so that* $X_n \geq b \forall n \geq N_0$ *almost surely. Then, the tests for* $H_0 : \theta = \theta^0$ *for autoregressive processes based on the simplified depth notions* d_S^1 *and* d_S^2 *are consistent at all* $\theta^* \in \Theta \setminus \Theta^0$*, if one of the following conditions hold:*

(i) $\exists c \neq 0 : \; g(X_n, \theta^0) - g(X_n, \theta^*) = c \; \forall n = 1, .., N, N \in \mathbb{N}$.

(ii) $\lim_{n \to \infty} g(b, \theta^0) - g(b, \theta^*) \in \{-\infty, \infty\}$.
In case of (ii) d_S^3 *is consistent, too.*

Proof. The proof is presented in Kustosz et al. (2016b). □

Note that the condition on N_0 implies that $X_N \overset{N \to \infty}{\to} \infty$ in probability.

Lemma 53. *For general linear and non-linear autoregressive growth processes of order one defined by*

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

with $y_n > y_{n-1}$ *and* θ *assuring an explosive process, the tests defined on simplified simplicial depth are consistent.*

Proof. See Kustosz et al. (2016b), Examples 5.1 - 5.6. □

Corollary 54. *For the linear model with and without intercept and for the non-linear autoregressive model the tests (30), (32) and (34) are consistent under the assumptions of Theorem 52. Further, the simplified simplicial depth tests are consistent for growth models defined by* $Y_n = Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + \theta_3 + E_n$.

Proof of Corollary 54. Here, we just have to show that the conditions of Theorem 52 hold for autoregressive processes defined by our two parameter models. Since we consider growth processes the function $g(\cdot, \theta^0) - g(\cdot, \theta^*)$ is unbounded, so that condition (ii) of Theorem 52 follows directly.

□

In Kustosz et al. (2016b), we also show that the conditions of Theorem 52 hold for various other models. Further, the reduction to alternating signs of residuals is formalised in a more general way. This allows the application of simplicial depth based tests to a wide range of different growth models. The additional statement in the last corollary further allows us to apply the full simplified depth tests to the full model with $\theta \in \mathbb{R}^3$.

4 Computational Aspects

In this section, we discuss the implementation of the proposed tests and confidence sets for parameters of explosive autoregressive processes. Thereby, we will discuss some further statistical properties of the empirical versions of simplicial depth to reduce the computational costs. First, we will show how to implement the main statistics. Then, we will propose a data driven calculation scheme for confidence intervals and finally propose how to calculate full simplicial depth confidence intervals by using the simplified depth notions.

4.1 Calculation of Test Statistics

In this section, we discuss the implementation of the test statistics in `R` as applied in the package `rexpar` (RobustEXPlosiveAutoRegression) which is available at github at the current stage of development. The package can be downloaded from `https://github.com/ChrisKust/rexpar.git`.

Implementation of Simplicial Depth for One-Parameter Models

In case of the one parameter model, we use a simplification of (14), neglecting the $r_n(\theta, y) = 0$ terms, given by

$$\begin{aligned} d_S^{AR}(\theta, y) &= \frac{1}{\binom{N}{2}} \sum_{1 \le n_1 < n_2 \le N} \mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0\}} \qquad (38) \\ &= \frac{2}{N(N-1)} \left(\sum_{n_1=1}^{N} \mathbb{1}_{\{r_{n_1}(\theta,y)>0\}} \cdot \sum_{n_2=1}^{N} \mathbb{1}_{\{r_{n_2}(\theta,y)<0\}} \right). \end{aligned}$$

Therefore, we need to assume that $\mathbb{P}(E_n = 0) = 0$ holds, what does apply in our set of assumptions. Then, the simplification follows by straight forward calculations. The representation of d_S^{AR} by (38) makes the implementation easy, since the residuals have to be evaluated once and then just the indicator sums have to be multiplied. Hence, the implementation can be based on matrix and vector operations.

Algorithm 55.
Given: Observation $y = (y_0, ..., y_N)^\top$, *Parameter* $\theta \in \mathbb{R}$.
1. Define $y^{s1} = (y_1, ..., y_N)^\top$ *and* $y^{s2} = (y_0, ..., y_{N-1})^\top$.
2. Calculate $r = (r_1, ..., r_N)^\top = y^{s1} - \theta y^{s2}$ *by matrix operations.*
2. Calculate r_+ *as boolean vector with* $r_+ = \mathbb{1}(r > 0)$.
3. Calculate r_- *analogously with* $r_- = \mathbb{1}(r < 0)$.

4. Calculate the sums of r_+ and r_-.
5. Multiply the resulting sums and rescale by $\frac{2}{N(N-1)}$.

Thereby $\mathbb{1}(r > 0)$ is a vector of dimension of r with values 1 at positions n with $r_n > 0$ and values 0 otherwise.

Implementation of Simplicial Depth for Two-Parameter Models

For the two parameter statistics, we apply a different calculation method, since a simple multiplication of indices is not directly possible when N gets large. The main aim is to evaluate

$$\sum_{1\leq n_1<n_2<n_3\leq N} \left(\mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0, r_{n_3}(\theta,y)>0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0, r_{n_3}(\theta,y)<0\}}\right). \quad (39)$$

To allow this in reasonable time, we will use specific matrices and parallel computation. In general, we evaluate $\binom{N}{3}$ groups of residuals which leads to a N^3 complexity when simplicial depth is calculated. A simple loop based calculation therefore results in high computation times. To avoid loops in the two inner sums, we propose a matrix based calculation, when n_1 is fixed. Therefore, we define two matrices

$$M_1(r,k) := \begin{pmatrix} r_{k+1}(\theta,y) & \dots & r_{k+1}(\theta,y) & r_{k+1}(\theta,y) \\ r_{k+2}(\theta,y) & \dots & r_{k+2}(\theta,y) & 0 \\ \vdots & \iddots & & \vdots \\ r_{N-1}(\theta,y) & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{(N-k-1)\times(N-k-1)}$$

and

$$M_2(r,k) := \begin{pmatrix} r_{k+2}(\theta,y) & r_{k+3}(\theta,y) & r_{k+4}(\theta,y) & \dots & r_N(\theta,y) \\ r_{k+3}(\theta,y) & r_{k+4}(\theta,y) & \dots & r_N(\theta,y) & 0 \\ r_{k+4}(\theta,y) & \dots & \iddots & & \vdots \\ \vdots & \iddots & & & \vdots \\ r_N(\theta,y) & 0 & \dots & & 0 \end{pmatrix}$$

$$\in \mathbb{R}^{(N-k-1)\times(N-k-1)}$$

which can be evaluated relatively fast. The first matrix M_1 is calculated similar to a standard triangular matrix but with zeroes under the lower counter-diagonal. The second matrix is a so called Hankel matrix which allows the calculation of the

complete inner sums by matrix operations. Both matrices depend on the residuals vector $r = (r_1, ..., r_N)^\top$ and an index k. Since the matrices start at a specific index, redundant calculations by using more residuals then needed can be avoided and the calculation is increasingly fast depending on the first residual r_k, whereby k is defined by the actual value of the index n_1. This first residual is used in a remaining outer loop. To utilise modern computer capabilities this loop can additionally be divided by parallel computation. Below, a summary of the algorithm for an evaluation of (39) with an observed process $y = (y_0, ..., y_N)$ and $\theta \in \mathbb{R}^2$ fixed and known is presented.

Algorithm 56.
Given: Observation $y = (y_0, ..., y_N)^\top$, *Parameter* $\theta \in \mathbb{R}^2$.
1. Define $y^{s1} = (y_1, ..., y_N)^\top$ *and* $y^{s2} = (y_0, ..., y_{N-1})^\top$.
2. Calculate $r = (r_1, ..., r_N)^\top = y^{s1} - \theta_0 - \theta_1 y^{s2}$ *by matrix operations.*
3. Predefine a list of vectors starting at the k-th residual, with $k = 1, ..., N-3$. *A list element is given by* $(r_k, ..., r_N)$.
4. Parallel evaluate the terms of (39) for each list element from 3.

For fixed $k = 1, ..., N-3$ *proceed as follows:*
4.1. Generate $M_1(r, k)$, *i.e. based on* $(r_{k+1}, ..., r_{N-1})$.
4.2. Generate $M_2(r, k)$, *i.e. based on* $(r_{k+2}, ..., r_N)$.
4.3. Generate a matrix M by

$$\begin{aligned} M(r,k) :=& \mathbb{1}_{\{r_k>0\}} \cdot \mathbb{1}(M_1(r,k) < 0) \circ \mathbb{1}(M_2(r,k) > 0) \\ &+ \mathbb{1}_{\{r_k<0\}} \cdot \mathbb{1}(M_1(r,k) > 0) \circ \mathbb{1}(M_2(r,k) < 0), \end{aligned}$$

where $\mathbb{1}_{\{r_k \lesseqgtr 0\}}$ *is the simple one dimensional indicator and* $\mathbb{1}(M_l \lesseqgtr 0), l \in \{1, 2\}$ *is a component wise indicator, i.e.* $(\mathbb{1}(M_l(r,k) > 0))_{i,j} := \mathbb{1}_{\{M_{l_{i,j}}(r,k)>0\}}$. *Further* $\circ$ *defines a component wise multiplication, i.e.* $(M_l(r,k) \circ M_m(r,k)_{i,j} = (M_{l_{i,j}}(r,k) \cdot M_{m_{i,j}}(r,k)_{i,j})$. *Thereby the resulting element is a matrix with dimension* $(N-k-1) \times (N-k-1)$ *again.*
4.4. Calculate the sum of all elements of M.
Each solution of the parallel loop is the partial sum for $n_1 = k$ *fixed in (39).*
5. Calculate the sum of the parallel results and add the evaluation for the residual set (r_{N-2}, r_{N-1}, r_N).
6. Finally rescale the sum by $\binom{N}{3}$.

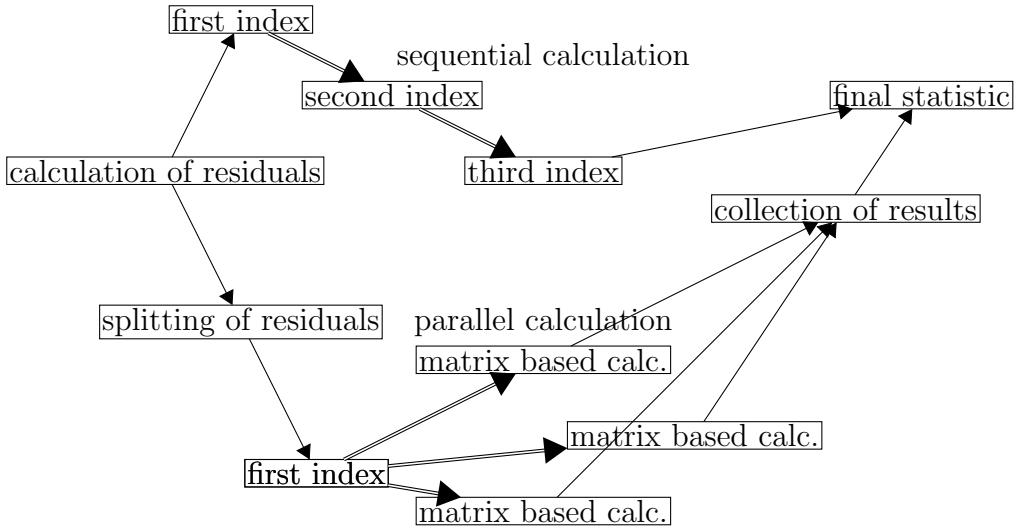


Fig. 15: Flow diagram of the parallel and loop based calculation.

Figure 15 illustrates a comparison of the sequential method with the parallel implementation. Both algorithms start with the calculation of residuals. The upper track to the final statistic shows the loop based calculation. The simple arrows show operations which are done with vector methods. The double struck arrows show steps where loops are applied. The sequential method uses two nested loops to calculate the sign changes for all relevant combinations of residuals. The lower track illustrates the parallel calculation. The parallel method splits the residuals and loops over the first index. Then the inner loops are replaced by vectorised calculations. The remaining loop can be split into chunks which are computed parallel. After a collection of the parallel results the final statistic can be computed.
In Section 5.6, a simulation study illustrating the performance gain by the matrix based calculation and parallel computation is presented.

Implementation of Simplified Depth Statistics

The calculation of the simplified statistics is similar to the one dimensional case. Here, we just use the fact that we have to evaluate the residuals in subsequent blocks. Then, the calculations reduce to matrix operations again. Since we do not have to consider the full set of all combinations of indices with increasing order the matrix methods work for reasonably large N.

Algorithm 57.
Given: Observation $y = (y_0, ..., y_N)^\top$, *Parameter* $\theta \in \mathbb{R}^2$.
1. Define $y^{s1} = (y_1, ..., y_N)^\top$ *and* $y^{s2} = (y_0, ..., y_{N-1})^\top$.
2. Calculate $r = (r_1, ..., r_N) = y^{s1} - \theta_0 - \theta_1 y^{s2}$ *by matrix operations.*

3. Calculate sets of shifted residuals by $r1 = r[from = 1, by = 3, to = N - 2]$, $r2 = r[from = 2, by = 3, to = N - 1]$ *and* $r3 = r[from = 3, by = 3, to = N]$.
4. Calculate the terms of the depth statistic by
$I = \mathbb{1}(r1 > 0) \cdot \mathbb{1}(r2 < 0) \cdot \mathbb{1}(r2 > 0) + \mathbb{1}(r1 < 0) \cdot \mathbb{1}(r2 > 0) \cdot \mathbb{1}(r2 < 0)$.
5. Calculate d_S^1 *by summation over* I *and rescaling with* $\frac{1}{\lfloor \frac{N}{3} \rfloor}$.

To calculate d_S^2 the vectors defining the residuals which we have to evaluate have to be replaced. Hence we use a modification of the algorithm for d_S^1.

Algorithm 58. *Replace 3. in Algorithm 57 by:*
3a. Calculate sets of shifted residuals by $r1 = r[from = 1, by = 1, to = N - 2]$, $r1 = r[from = 2, by = 1, to = N - 1]$ *and* $r3 = r[from = 3, by = 1, to = N]$.

For d_S^3 step 3. of Algorithm 57 has to be replaced by

Algorithm 59. *Replace 3. in Algorithm 57 by:*
3b. Calculate sets of shifted residuals by $r1 = r[from = 1, by = 1, to = \lfloor \frac{N+1}{2} \rfloor]$, $r2 = r[val = \lfloor \frac{N+1}{2} \rfloor, times = \lfloor \frac{N+1}{2} \rfloor]$ *and* $r3 = r[from = N, by = 1, to = \lfloor \frac{N+1}{2} \rfloor + 1]$.

By application of the implemented functions, the proposed tests can be evaluated. Therefore, we need to know the quantiles of the limit distributions. In case of χ^2 and Normal distributions they are known. In Section 4.2, we present a simulation method for the limit distribution of the simplicial depth statistics with $\theta \in \mathbb{R}^d, d > 1$.

Examples of the Implemented Simplicial Depth Statistics

To illustrate the resulting statistics, we give some examples based on simulated data.

Example 60. *Consider the data generating model given by*

$$Y_n = \theta_1 \cdot Y_{n-1} + Y_{n-1} + E_n,$$

with $\theta_1 = 0.01, y_0 = 15, E_n \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.1)$. *We simulate the process with* $N = 100$ *and thereby get a realisation as presented in Figure 16.*

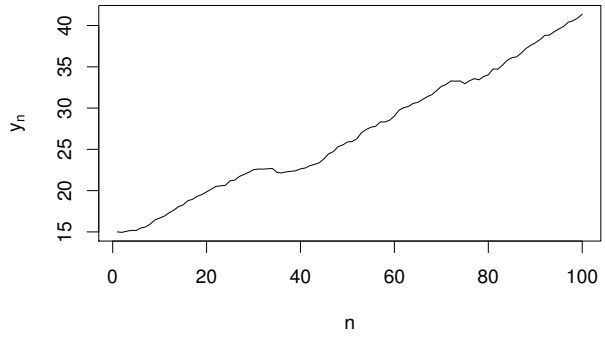


Fig. 16: Realisation of an explosive AR(1) process without intercept.

Now, depth can be evaluated on a grid. Since we know that the true parameter is 0.01*, we evaluate a parameter grid of* $M+1$ *parameters given by* $\theta_{1,0} = -0.1, \theta_{1,m+1} = \theta_{1,m} + 0.01$ *and* $\theta_{1,M} = 0.1$*. The resulting function* $d_S^{AR}(\theta, y)$ *with variable* θ *is depicted in Figure 17.*

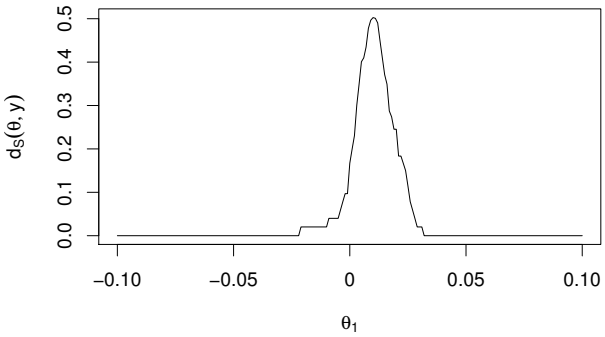


Fig. 17: Empirical depth curve of the simulated AR(1) process without intercept.

The maximal depth is achieved on the interval $[0.01023, 0.0106]$*. Due to the evaluation via signs of residuals no unique maximum exists. The maximal depth is attained at a connected subset of the parameter space.*

Example 61. *Consider the data generating model given by*

$$Y_n = \theta_1 \cdot Y_{n-1} + Y_{n-1} + \theta_3 + E_n,$$

with $\theta_1 = 0.01, \theta_3 = 0, y_0 = 15, E_n \sim \mathcal{N}(0, 0.1)$ *i.i.d. We simulate the process with* $N = 100$ *and thereby get a realisation as presented in Figure 16. Since we know*

that the true parameter is $(0.01, 0)$ *we evaluate a grid given by* $\theta_{1,0} = -0.1, \theta_{1,m+1} = \theta_{1,m} + 0.001, \theta_{1,M} = 0.1$ *for* θ_1 *and* $\theta_{3,0} = -1, \theta_{3,v+1} = \theta_{3,v} + 0.001$ *and* $\theta_{3,V} = 1$ *for* θ_3*. The resulting function* $d_S^{ARi}(\theta, y)$ *with variable* θ *is depicted in Figure 18.*

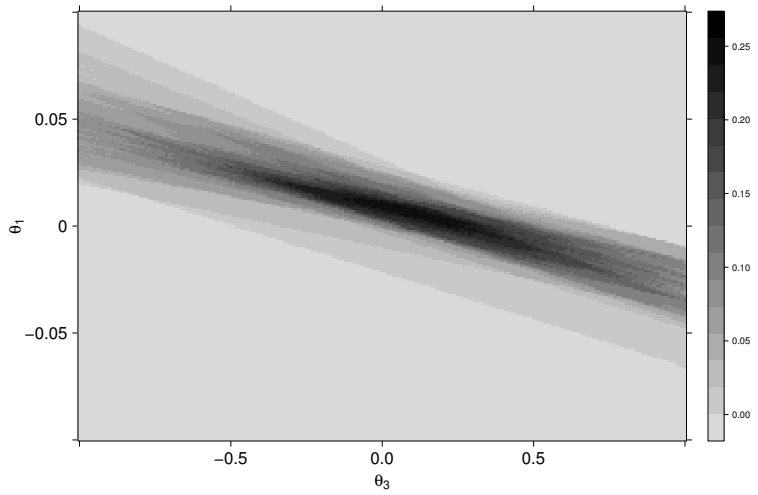


Fig. 18: Empirical depth curve of the simulated AR(1) process with intercept.

The maximal depth is achieved for the parameter $(0.006, 0.11)$*. The deviation to the true parameter is a random small sample effect. Depth is also quite large close to the true parameter.*

The most important observation is that empirical simplicial depth defines a piecewise constant function which is centred at a maximum that appears to be asymptotically located at the true parameter. This property unfortunately is just proven under growth assumptions, see Sections 3.1 and 3.2. Otherwise, simplicial depth for models with parameter dimension larger than one is not necessarily increasing with respect to a centre.

Remark 62. *Note that the last observation in particular contradicts the conditions for depth functions, as introduced by Zuo and Serfling (2000). Hence, formally simplicial depth is not necessarily a proper depth function when our assumptions are violated.*

4.2 Limit Distribution

Approximation of the Limit Process for Two Dimensional Parameters

An important step in this thesis is the calculation of the quantiles for the limit distribution of the simplicial depth with two dimensional parameters. In particular,

we need the quantiles of the distribution of $W := \frac{3}{4} + \frac{3}{4}X_2^2(0) - \frac{3}{2}\int_{-2}^{2} X_1^2(t)dt$, whereby $X(t) = (X_1(t), X_2(t)))^\top$ is a centred Gaussian process defined by the covariance matrix (27).

In Figure 19, a simulation of the paths of this bivariate process is depicted. The solid line represents the variable $X_1(t)$ which starts in 0 at $t = -0.5$ and returns to 0 at $t = 1.5$. The dashed line is a simulation of $X_2(t)$. This process is a draw from a $\mathcal{N}(0, 1)$ distribution and is constant over time. Note that the two processes meet at $t = 0.5$ due to the underlying covariance structure. To generate such simulations we

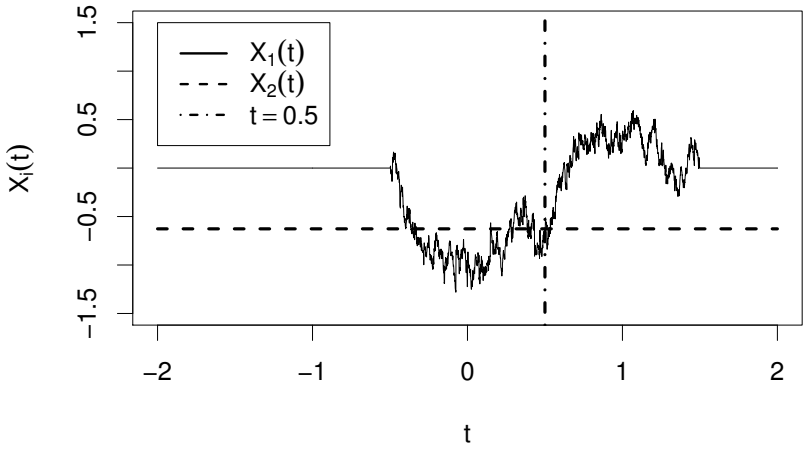


Fig. 19: Simulation of a path of the limit process.

rearrange the two-dimensional process and use the fact that $X_2(0)$ is determined by a $\mathcal{N}(0, 1)$ random variable. This can be seen in the reduced form representation of the limit process in the proof of Lemma 3.1 in Kustosz et al. (2016a). The component which defines $X_2(t)$ is $\frac{1}{N}\sum_{n=1}^{N} \Phi(E_n)$, whereby $\Phi(x) = \mathbb{1}_{\{x<0\}} - \mathbb{1}_{\{x>0\}}$ and $(E_n)_n$ are errors from a distribution with $med(E_n) = 0$. The central limit theorem assures that this component converges to a $\mathcal{N}(0, 1)$ distribution. In addition, we see that this component does not depend on t. With this observation, we can prove the following statement which is useful to simulate the limit distribution.

Theorem 63. *For a centred Gaussian process $(Y(t))_{t\in T}$ and $Y(t) = (Y^{(1)}(t), Y^{(2)}(t))^\top$, with covariance matrix given by*

$$cov(Y(s), Y(t)) = \begin{pmatrix} c_{st} & c_s \\ c_t & 1 \end{pmatrix},$$

where

$$c_{st} = \int_0^1 \mathbb{1}_{\{x-s\in(-0.5,0.5]\}} \mathbb{1}_{\{x-t\in(-0.5,0.5]\}} dx,$$
$$c_s = \int_0^1 \mathbb{1}_{\{x-s\in(-0.5,0.5]\}} dx,$$
$$c_t = \int_0^1 \mathbb{1}_{\{x-t\in(-0.5,0.5]\}} dx,$$

consider a grid $t = (t_1, ..., t_n)$. *Then* $Y^{(2)}(t_i) = Y^*$ *for all* $i \in \{1, ..., n\}$, *whereby* $Y^* \sim \mathcal{N}(0, 1)$ *and* $Y^{R,(1)} := (Y^{(1)}(t_1), ..., Y^{(1)}(t_n))^\top$ *conditioned on* $(Y^{(2)}(t_1), ..., Y^{(2)}(t_n))^\top = (y^*, ..., y^*)^\top$ *has a multivariate Normal distribution with mean vector*

$$\mu^R = \begin{pmatrix} c_{t_1} \\ \vdots \\ c_{t_n} \end{pmatrix} \cdot y^*$$

and covariance matrix

$$\Sigma^R = (c_{t_i t_j} - c_{t_i} \cdot c_{t_j})_{i,j\in\{1,...,n\}}.$$

Proof. The underlying process $Y(t) = (Y^{(1)}(t), Y^{(2)}(t))^\top$ is a bivariate Gaussian process. Hence, the process is also a Gaussian for all index sets $(t_1, ..., t_n)$. We set

$$Y := (Y^{(1)}(t_1), Y^{(2)}(t_1), Y^{(1)}(t_2), Y^{(2)}(t_2), ..., Y^{(1)}(t_n), Y^{(2)}(t_n))^\top$$

for an arbitrary but fixed index set $(t_1, ..., t_n)$. This random variable possesses a multivariate normal distribution with a zero mean function and a covariance matrix determined by $c_{t_i t_j}, c_{t_i}, c_{t_j}$.First, note that $c_{t_i} = c_{t_i t_i}$.
The mean vector of Y is given by μ. The covariance structure is defined by

$$\Sigma = \begin{pmatrix} \Sigma_{t_1 t_1} & \Sigma_{t_1 t_2} & ... & \Sigma_{t_1 t_n} \\ \Sigma_{t_2 t_1} & \Sigma_{t_2 t_2} & ... & \Sigma_{t_2 t_n} \\ ... & ... & ... & \\ \Sigma_{t_n t_1} & \Sigma_{t_n t_2} & ... & \Sigma_{t_n t_n} \end{pmatrix},$$

with

$$\Sigma_{t_i t_j} = \begin{pmatrix} c_{t_i t_j} & c_{t_i} \\ c_{t_j} & 1 \end{pmatrix}.$$

To allow the computation of the conditional distribution based on the second margin, we need to rearrange the process. This can be done by a linear transformation based on an invertible matrix C, defined by

$$C := \begin{pmatrix} 1 & 0 & 0 & 0 & ... & 0 & 0 \\ 0 & 0 & 1 & 0 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & ... & 1 & 0 \\ 0 & 1 & 0 & 0 & ... & 0 & 0 \\ 0 & 0 & 0 & 1 & ... & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & 0 & 1 \end{pmatrix}.$$

This matrix rearranges Y to

$$Y^R = CY = (Y^{(1)}(t_1), ..., Y^{(1)}(t_n), Y^{(2)}(t_1), ..., Y^{(2)}(t_n))^\top.$$

By Theorem 2.4.1. in Anderson (1958), Y^R is multivariate normally distributed with mean $\mu^R = C\mu$ and covariance matrix $\Sigma^R = C\Sigma C^\top$. Since $Y \sim \mathcal{N}(\overline{0}, \Sigma)$, we have $Y^R \sim \mathcal{N}(\overline{0}, \Sigma^R)$ and calculate

$$\Sigma^R = C\Sigma C^T = \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{pmatrix},$$

where

$$\tilde{\Sigma}_{11} = (c_{t_i t_j})_{i,j} = \begin{pmatrix} c_{t_1} & c_{t_1 t_2} & ... & c_{t_1 t_n} \\ c_{t_2 t_1} & c_{t_2} & ... & c_{t_2 t_n} \\ ... & ... & ... & ... \\ c_{t_n t_1} & c_{t_n t_2} & ... & c_{t_n} \end{pmatrix},$$

$$\tilde{\Sigma}_{12} = (c_{t_i})_{i,j} = \begin{pmatrix} c_{t_1} & c_{t_1} & ... & c_{t_1} \\ c_{t_2} & c_{t_2} & ... & c_{t_2} \\ ... & ... & ... & ... \\ c_{t_n} & c_{t_n} & ... & c_{t_n} \end{pmatrix},$$

$$\tilde{\Sigma}_{21} = (c_{t_j})_{i,j} = \tilde{\Sigma}_{12}^T = \begin{pmatrix} c_{t_1} & c_{t_2} & ... & c_{t_n} \\ c_{t_1} & c_{t_2} & ... & c_{t_n} \\ ... & ... & ... & ... \\ c_{t_1} & c_{t_2} & ... & c_{t_n} \end{pmatrix},$$

$$\tilde{\Sigma}_{22} = (1)_{i,j} = \begin{pmatrix} 1 & 1 & ... & 1 \\ 1 & 1 & ... & 1 \\ ... & ... & ... & ... \\ 1 & 1 & ... & 1 \end{pmatrix}.$$

All matrices have dimension $n \times n$ and hence Σ^R is a $2n \times 2n$ matrix. From Kustosz et al. (2016a) we know that $Y^{(2)}(t)$ is constant for all t and hence can set $Y^{(2)}(t) = Y^*$. Further, we know that $Y^{(2)}(t) = Y^* \sim \mathcal{N}(0,1)$. This is in line with $\tilde{\Sigma}_{22}$. By application of Theorem 2.5.1. from Anderson (1958), we can conclude that the distribution of Y^R conditioned on the last n components is multivariate Normal with mean μ^R and covariance which can be calculated by the parts of Σ^R.
Formally, for $Y^{R,(1)} := (Y^{(1)}(t_1), ..., Y^{(1)}(t_n))^T$ and $Y^{R,(2)} = (Y^{(2)}(t_1), ..., Y^{(2)}(t_n))^\top$, we know that

$$\begin{aligned} Y^{R,(1)} | \left(Y^{R,(2)} = (y^{(2)}(t_1), ..., y^{(2)}(t_n)) = (y^*, ..., y^*) \right) \\ \sim \mathcal{N}(\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^-(y^*, ..., y^*)^\top, \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^-\tilde{\Sigma}_{21}) \end{aligned}$$

holds.
Since $\tilde{\Sigma}_{22}$ is non-invertible, we have to use a generalised inverse here. A simple choice is $I = (\delta_{11}(i,j))_{i,j=1,...,n}$.
Hence, it is

$$\mu^R = \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^-(y^*, ..., y^*)^\top = (c_{t_1}, ..., c_{t_n})^T \cdot y^*$$

and

$$
\begin{aligned}
\Sigma^R &= \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-}\tilde{\Sigma}_{21} \\
&= \tilde{\Sigma}_{11} - \begin{pmatrix} c_{t_1} & c_{t_1} & ... & c_{t_1} \\ c_{t_2} & c_{t_2} & ... & c_{t_2} \\ ... & ... & ... & ... \\ c_{t_n} & c_{t_n} & ... & c_{t_n} \end{pmatrix} \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & ... & 0 \end{pmatrix} \begin{pmatrix} c_{t_1} & c_{t_2} & ... & c_{t_n} \\ c_{t_1} & c_{t_2} & ... & c_{t_n} \\ ... & ... & ... & ... \\ c_{t_1} & c_{t_2} & ... & c_{t_n} \end{pmatrix} \\
&= \tilde{\Sigma}_{11} - \begin{pmatrix} c_{t_1} & 0 & ... & 0 \\ c_{t_2} & 0 & ... & 0 \\ ... & ... & ... & ... \\ c_{t_n} & 0 & ... & 0 \end{pmatrix} \begin{pmatrix} c_{t_1} & c_{t_2} & ... & c_{t_n} \\ c_{t_1} & c_{t_2} & ... & c_{t_n} \\ ... & ... & ... & ... \\ c_{t_1} & c_{t_2} & ... & c_{t_n} \end{pmatrix} \\
&= \tilde{\Sigma}_{11} - \begin{pmatrix} c_{t_1}^2 & c_{t_1}c_{t_2} & ... & c_{t_1}c_{t_n} \\ c_{t_2}c_{t_1} & c_{t_2}^2 & ... & c_{t_2}c_{t_n} \\ ... & ... & ... & ... \\ c_{t_n}c_{t_1} & c_{t_n}c_{t_2} & ... & c_{t_n}^2 \end{pmatrix} \\
&= \begin{pmatrix} c_{t_1} & c_{t_1t_2} & ... & c_{t_1t_n} \\ c_{t_2t_1} & c_{t_2} & ... & c_{t_2t_n} \\ ... & ... & ... & ... \\ c_{t_nt_1} & c_{t_nt_2} & ... & c_{t_n} \end{pmatrix} - \begin{pmatrix} c_{t_1}^2 & c_{t_1}c_{t_2} & ... & c_{t_1}c_{t_n} \\ c_{t_2}c_{t_1} & c_{t_2}^2 & ... & c_{t_2}c_{t_n} \\ ... & ... & ... & ... \\ c_{t_n}c_{t_1} & c_{t_n}c_{t_2} & ... & c_{t_n}^2 \end{pmatrix} \\
&= \begin{pmatrix} c_{t_1} - c_{t_1}^2 & c_{t_1t_2} - c_{t_1}c_{t_2} & ... & c_{t_1t_n} - c_{t_1}c_{t_n} \\ c_{t_2t_1} - c_{t_2}c_{t_1} & c_{t_2} - c_{t_2}^2 & ... & c_{t_2t_n} - c_{t_2}c_{t_n} \\ ... & ... & ... & ... \\ c_{t_nt_1} - c_{t_n}c_{t_1} & c_{t_nt_2} - c_{t_n}c_{t_2} & ... & c_{t_n} - c_{t_n}^2 \end{pmatrix} \\
&= (c_{t_it_j} - c_{t_i}c_{t_j})_{i,j}.
\end{aligned}
$$

The last equality holds, since $c_{t_it_i} = c_{t_i}$. □

Theorem 63 is applied in the `rexpar` package to simulate the bivariate Gaussian limit process. Thereby, we can reduce the simulation on one random draw of the constant part y^*, and then just have to simulate a univariate Gaussian process, based on μ^R and Σ^R, as given by the Theorem. To approximate the distribution of the limit process which is of interest, we finally have to approximate the distribution of $W := \frac{3}{4} + \frac{3}{4}X_2^2(0) - \frac{3}{2}\int_{-2}^{2} X_1^2(t)dt$. This can be done by Riemann sums based on

the simulated processes given by

$$W_L := \frac{3}{4} + \frac{3}{2}x_2(t_0)^2 - \frac{3}{2}\sum_{l=0}^{L-1}\frac{1}{2}\left(x_1(t_{l+1})^2 + x_2(t_l)^2\right)(t_{l+1} - t_l), \tag{40}$$

whereby $(x_1(t_i), x_2(t_i))$ for $i = 1, ..., L$ are discrete realisations of the bivariate Gaussian process $X = (X_1, X_2)$ at L points in time $(t_1, ..., t_L)$.
In the supplementary `rexpar` package the quantiles of this distribution are given for probabilities at three decimal spaces in the `SimQuantiles` matrix. These quantiles were calculated based on 200000 simulated processes on a grid with step size 0.001 for the interval $[-0.5, 1.5]$. The remaining part of the process is constant and hence does not need to be simulated in the time domain. Taking the empirical quantiles yields the distribution presented in Figure 20 and the values in Table 1.

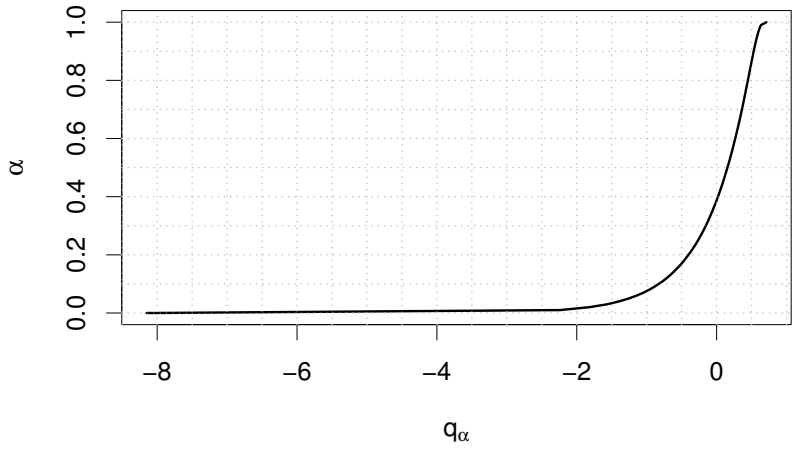


Fig. 20: Quantiles of the approximate distribution of W.

α	0.005	0.01	0.025	0.05	0.1	0.2
q_α	-2.6543216	-2.2403956	-1.6791727	-1.2545411	-0.8270908	-0.4032076

Tab. 1: Simulated quantiles of the integrated Gaussian process.

Approximation of the Limit Process for K Dimensional Parameters

A procedure which we later will use to check the limit distributions, see Section 5.7, is also quite useful for all cases where the limit distribution could not be derived,

yet. Since we assume that under H_0 the errors have a zero median, we simply can approximate the limit distribution by an appropriate simulation of the residual signs.
For the simplicial depth statistic with parameter $\theta \in \mathbb{R}^K$, we have to evaluate

$$\sum_{1\leq n_1<\ldots<n_{K+1}\leq N} \left(\prod_{k=1}^{K} \mathbb{1}_{\{r_{n_k}(\theta,y)(-1)^k>0\}} + \prod_{k=1}^{K} \mathbb{1}_{\{r_{n_k}(\theta,y)(-1)^{k+1}>0\}} \right).$$

Since under H_0 we have $med(E_n) = 0$, the limit is approximated by

$$\lim_{N\to\infty} \sum_{1\leq n_1<\ldots<n_{K+1}\leq N} \left(\prod_{k=1}^{K} \mathbb{1}_{\{E_{n_k}(-1)^k>0\}} + \prod_{k=1}^{K} \mathbb{1}_{\{E_{n_k}(-1)^{k+1}>0\}} \right).$$

Even if the actual error distribution is unknown, the $med(E_n) = 0$ assumption suffices to generate simulations, since we just need to evaluate the signs of the residuals. These signs can be simulated as Bernoulli variables, so that

$$\begin{aligned} &\lim_{N\to\infty} \sum_{1\leq n_1<\ldots<n_{K+1}\leq N} \left(\prod_{k=1}^{K} \mathbb{1}_{\{E_{n_k}(-1)^k>0\}} + \prod_{k=1}^{K} \mathbb{1}_{\{E_{n_k}(-1)^{k+1}>0\}} \right) \\ = &\lim_{N\to\infty} \sum_{1\leq n_1<\ldots<n_{K+1}\leq N} \left(\prod_{k=1}^{K} \mathbb{1}_{\{B_{n_k}(-1)^k>0\}} + \prod_{k=1}^{K} \mathbb{1}_{\{B_{n_k}(-1)^{k+1}>0\}} \right), \end{aligned} \tag{41}$$

whereby $(B_n)_{n\in\mathbb{N}}$ is a series of i.i.d. $Bin(1, 1/2)$ random variables on $\{-1, 1\}$. To generate a simulation of the depth statistic hence set

$$d_{Sim}(N,K) := \frac{1}{\binom{N}{K}} \sum_{1\leq n_1<\ldots<n_{K+1}\leq N} \left(\prod_{k=1}^{K} \mathbb{1}_{\{B_{n_k}(-1)^k>0\}} + \prod_{k=1}^{K} \mathbb{1}_{\{B_{n_k}(-1)^{k+1}>0\}} \right)$$

and define

$$Z_{Sim}(N,K) = N(d_{Sim}(N,K) - \mathbb{E}(d_S)).$$

The expected value of the full simplicial depth thereby can be calculated quite easily, since it just depends on the number of sign changes.
Due to the summation it is still computationally costly to generate one draw of this distribution for large K. Further a large N is necessary to approximate the limit. Nevertheless, this method delivers a possibility to achieve approximate limit distributions for $K > 2$, when the theoretical results are not available. In Section

5.7, we analyse the validity of the method in a simulation study.

4.3 Calculation of Confidence Regions for Linear AR(1) Models

To calculate confidence regions, we use the definitions (17), (29), (31), (33), and (35). Since the calculation of the test statistics still takes some time for long series, a grid based calculation which leads to many evaluations of the depth statistics is inefficient. Fortunately, the empirical depth allows a pre-selection of candidate points, since it is a piecewise constant function.

Candidates for the Linear Model Without Intercept

To define reasonable candidates, one first has to notice that each of the proposed depth notions just changes value, if $r_n(\theta, y)$ changes sign. For a fixed vector of observed data $y = (y_0, ...y_N)$ the sign changes can be expressed as functions of the parameter θ.
In model (9), we just have one parameter. Hence, each sign change of the residuals is associated to one parameter which can be calculated by

$$\begin{aligned} & y_n - \theta_1 y_{n-1} - y_{n-1} = 0 \\ \Leftrightarrow \ & \theta_1 = \frac{y_n}{y_{n-1}} - 1. \end{aligned} \tag{42}$$

More precise, a residual defined by (y_n, y_{n-1}) is negative, if $\theta_1 < \frac{y_n}{y_{n-1}} - 1$ and positive, if $\theta_1 > \frac{y_n}{y_{n-1}} - 1$. Based on tupelo of such parameters, one can specify candidates. From two residuals, one gets two sign change parameters $\theta_{1,1}, \theta_{1,2}$. Without loss of generality assume $\theta_{1,2} > \theta_{1,1}$. Since the residuals are just relevant if they change sign, one can also consider the centre of each interval, defined by

$$\theta_{1,C} := \frac{\theta_{1,2} + \theta_{1,1}}{2}.$$

By evaluation of all possible residuals, defined by the data y and $\theta_{1,n} = \frac{y_n}{y_{n-1}} - 1$ for $n \in \{1, ..., N\}$, one gets the candidate set

$$\Theta_{cand}^{d_S^{AR}} = \left\{ \frac{\theta_{1,i} + \theta_{1,j}}{2} \middle| i, j \in \{1, ..., N\}, \theta_{1,i} > \theta_{1,j} \right\}. \tag{43}$$

Allowing $\theta_{1,i} = \theta_{1,j}$ would evaluate depth in the jump points also.

Remark 64. *In (43), one can replace* $\frac{\theta_{1,i}+\theta_{1,j}}{2}$ *by* $\frac{\theta_{1,(i)}+\theta_{1,(i+1)}}{2}$*, with* $\theta_{1,(1)} < \theta_{1,(2)} < ... < \theta_{1,(N)}$ *being the ordered set of candidate parameters. Then, the resulting candidate set is smaller but allows the calculation of the depth at each constant part. However, for a construction of confidence sets it is more interesting to have points close to the jumps of the depth function. This can for example be achieved by additional candidates in the jump locations. But then, we have to notice that by definition of depth, we get spikes in these points. Further, it is useful to add* $\theta_{1,(1)}$ *and* $\theta_{1,(N)}$ *to the candidate set. Thereby, the boundaries of the depth shape are considered in a proper way.*

A simple alternative is to evaluate the points where the parameters change sign and to shift the parameter slightly to explore the complete depth contour. This avoids the calculation problems at the parameters with $r_n(\theta, y) = 0$ and approximates depth to a fixed precision. We define a second candidate set by

$$\Theta_{cand}^{d_S^{AR,\epsilon}} = \left\{ \theta_{1,n} \pm \epsilon \middle| n \in \{1, ..., N\} \right\} \tag{44}$$

for a fixed $\epsilon > 0$. This candidate set is even smaller than the first proposed set but relies on an ϵ which has to be small enough so that no relevant parameter ranges are skipped. This can be controlled by calculation of the differences resulting from the edges calculated by (42) and setting $\epsilon < min\{|\theta_{1,i} - \theta_{1,j}| \mid i \neq j, i, j \in \{1, ..., N\}\}$. We choose $\pm$ in the candidate set to have parameters on both sides of each jump of the depth function. For large samples it suffices to consider candidates given by (42) evaluated at all observations.

Example 65. *For a linear AR process without intercept defined by* $y_0 = 15, \theta_1 = 0.01$ *and* $\mathcal{N}(0, 0.1)$ *errors, a simulated process is presented in Figure 21.*

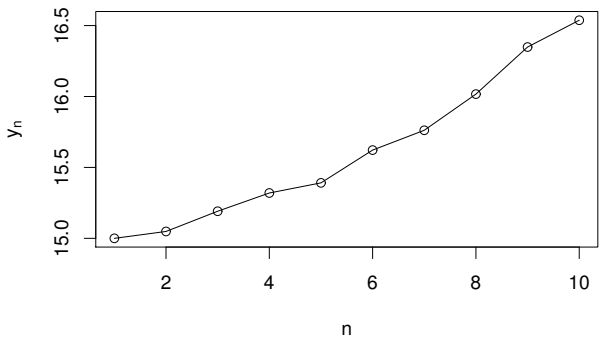


Fig. 21: Simulated AR(1) process, $\theta_1 = 0.01, y_0 = 15$ with $\mathcal{N}(0, 0.1)$ errors.

To evaluate the full depth function, we calculate d_S^{AR} *on a very fine grid defined by* $\{-0.1, -0.0999, -0.0998, ..., 0.0999, 0.1\}$. *Further, we calculate depth for the parameters defined by (43) marked by grey squares, for the approach proposed in Remark 64 marked by green dots, and for the second set (44) marked by red circles. The results can be seen in Figure 22.*

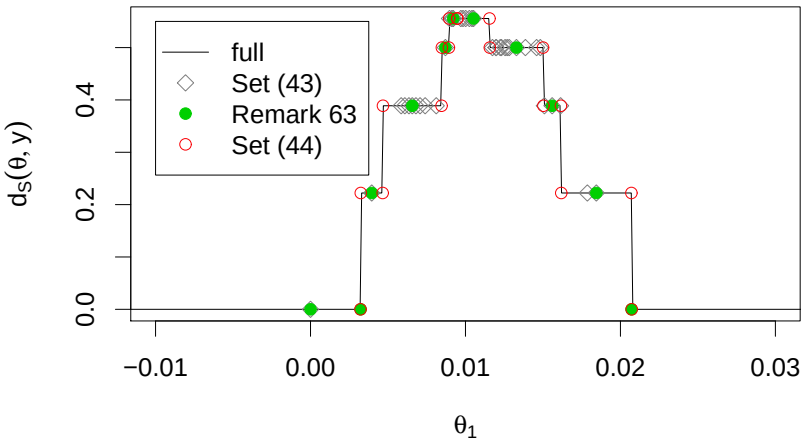


Fig. 22: Empirical depth for a simulated AR(1) process, $\theta = 0.01, y_0 = 15$ with $\mathcal{N}(0, 1)$ errors.

We see that for a reliable application of the first candidate set given by (43) it is necessary to include the minimal and maximal candidate, since they are not apparent, if just the differences are calculated. The set based on Remark 64 discovers every depth level, but does not allow a precise detection of the jump locations. The second

set explores the depth shape more efficient. We selected $\epsilon = 0.00001$. *While the set (43) results in many candidates, the* ϵ *based algorithm defined by set (44) detects the jumps in a very precise and efficient way.*

Candidates for the Linear Model With Intercept

A similar idea allows us to define candidates for the two parameter models. In the linear case, the sign changes are defined by straight lines set by parameters satisfying

$$r_n(\theta, y) = y_n - y_{n-1} - \theta_1 y_{n-1} - \theta_3 = 0 \tag{45}$$

for $n \in \{1, ..., N\}$. Hence, we now have half-space generating straights which are determined by the roots of the residuals. Based on two residuals, one can find the intersections of such straights.

Lemma 66. *Consider two tuples of observations* (y_n, y_{n-1}) *and* (y_m, y_{m-1}) *with roots of the residuals defined by*

$$\begin{aligned} r_n(\theta, y) &= y_n - y_{n-1} - \theta_1 y_{n-1} - \theta_3 = 0, \\ r_m(\theta, y) &= y_m - y_{m-1} - \theta_1 y_{m-1} - \theta_3 = 0. \end{aligned}$$

Then the lines, defining the sign changes of the residuals, intersect in (θ_1^*, θ_3^*) *with*

$$\theta_1^* = \frac{y_m - y_n}{y_{m-1} - y_{n-1}} - 1$$

and

$$\theta_3^* = \frac{y_n - y_m}{y_{m-1} - y_{n-1}} y_{n-1} + y_n.$$

Proof. To calculate θ_1^* set

$$\begin{aligned} y_n - y_{n-1} - \theta_1 y_{n-1} - \theta_3 &= y_m - y_{m-1} - \theta_1 y_{m-1} - \theta_3 \\ \Rightarrow y_n - y_{n-1} - \theta_1 y_{n-1} &= y_m - y_{m-1} - \theta_1 y_{m-1} \\ \Rightarrow \theta_1 (y_{m-1} - y_{n-1}) &= y_m - y_{m-1} - y_n + y_{n-1} \\ \Rightarrow \theta_1^* &= \frac{y_m - y_n}{y_{m-1} - y_{n-1}} - 1. \end{aligned}$$

By insertion of θ_1^* into the root equation for n, we get

$$\begin{aligned} & y_n - y_{n-1} - \theta_1^* y_{n-1} - \theta_3 = 0 \\ \Rightarrow\ & y_n - y_{n-1} - \left(\frac{y_m - y_n}{y_{m-1} - y_{n-1}} - 1 \right) y_{n-1} = \theta_3 \\ \Rightarrow\ & (-1) \cdot \frac{y_m - y_n}{y_{m-1} - y_{n-1}} y_{n-1} + y_n = \theta_3 \\ \Rightarrow\ & \frac{y_n - y_m}{y_{m-1} - y_{n-1}} y_{n-1} + y_n = \theta_3^*. \end{aligned}$$

□

Three residuals define a simplex with vertices defined by the intersection of the root equations. Since we know that the respective residual sign is constant at each half-space, a reasonable choice of candidate points is given by the points surrounded by simplexes formed by each set of three residuals. In particular, for explosive processes this is the only region where we can observe alternating signs. This can be seen for a set of three residual sign dividing straights under the assumption that we observe growth processes. In Figure 5 (see Section 3.2) this situation is depicted.
From equation (45), one can simply conclude that each residual sign dividing line can be interpreted as a function of θ_1 defining θ_3, if (y_n, y_{n-1}) is fixed. So, the regions defining all possible outcomes of signs for three residuals can be simply explored by examination of points around the three resulting vertices. Further, it follows that the respective residual has a negative sign, if parameter combinations above this straight are considered and a positive sign for parameter combinations below, as long as $y_n > y_{n-1}$ holds for all n. This can be seen by

$$\begin{aligned} r_n(\theta, y) =& y_n - (\theta_1 + 1) y_{n-1} - \theta_3 \lesseqgtr 0 \\ \Leftrightarrow & y_n - \kappa y_{n-1} \lesseqgtr \theta_3, \end{aligned}$$

with $\kappa = \theta_1 + 1$. If $y_n > y_{n-1} > 0$ holds, the half-space generating set is defined by a function $f_n(\kappa) = \theta_3$ which has an intercept y_n and a slope $(-1) \cdot y_{n-1}$ so that it is a decreasing linear function passing the points $(0, y_n)$ and $(\frac{y_n}{y_{n-1}}, 0)$. If $y_n > y_{n-1}$, then $\frac{y_n}{y_{n-1}} > 1$, so that the second intersection with the axis coincides with a positive value for κ. One then can see that alternating signs just appear, if we consider candidates inside of a simplex defined by the three data generated vertices (see Figure 5). We define $\theta^{n,m} = \left(\frac{y_m - y_n}{y_{m-1} - y_{n-1}} - 1, \frac{y_n - y_m}{y_{m-1} - y_{n-1}} y_{n-1} + y_n \right)$ and thereby get a first candidate

set by

$$\Theta_{cand}^{d_S^{ARi,\epsilon}} = \left\{ \theta^{n,m} \pm \epsilon \middle| n, m \in \{1, ..., N\}, n \neq m \right\}, \tag{46}$$

whereby here $\epsilon = (\epsilon_1, \epsilon_2)$ and $\pm$ denotes the usage of each combination of both signs. Therefore, a circular grid can be evaluated. Another candidate set for the linear two parameter model can be defined by the inner points of these simplexes via

$$\Theta_{cand}^{d_S^{ARi}} = \left\{ \frac{1}{3}(\theta^{i,j} + \theta^{i,k} + \theta^{j,k}) \middle| i, k, j \in \{1, ..., N\}, i \neq j, i \neq k, j \neq k \right\}. \tag{47}$$

For this set, we need to prove that the candidates are points in the simplex with vertices $\theta^{i,j}, \theta^{i,k}$ and $\theta^{j,k}$.

Lemma 67. *Consider a triangle with edges $A, B, C \in \mathbb{R}^2$ and define $S := \frac{1}{3}(A + B + C) \in \mathbb{R}^2$. Then $S \in \overline{conv(A, B, C)}$ holds.*

Proof. The proof can be found in Appendix A.4. □

Example 68. *In this example, we compare the candidate sets and grid based evaluation of depth in a two parameter example. Consider the model*

$$Y_n = \theta_1 Y_{n-1} + Y_{n-1} + \theta_3 + E_n,$$

with $y_0 = 1, \theta_1 = 1.05, \theta_3 = 0.2$ and $E_n \sim \mathcal{N}(0, 0.03)$ for $N = 5$ observations. We get a set of 4 resulting residuals. The candidate set (47) is then constructed from four different residuals. The vertices, defined by equation (45), are depicted as red crosses in Figure 23.

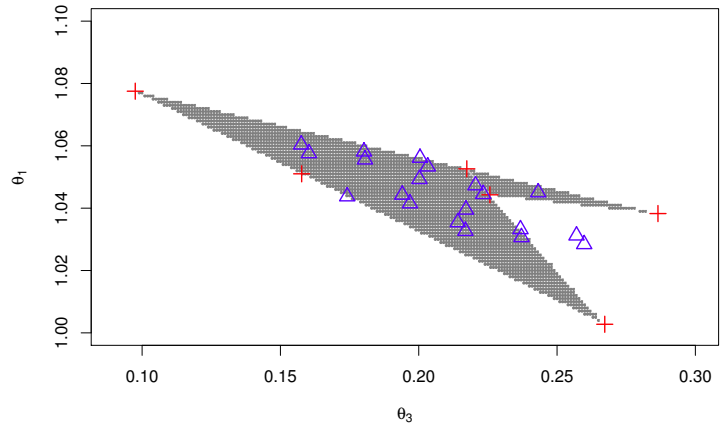


Fig. 23: Candidate set for a linear AR(1) model with intercept. The (△) show the candidates defined by the inner points of candidate simplexes ,see (47). The (+) are the candidates evaluated at the vertices of the data generated simplexes. The grey points are the parameters in the grid with depth larger than zero.

From these six vertices, we can calculate 20 interior points, marked by the triangles, as proposed in the candidate set (47). This already completely explores the depth function. Its values are given by coloured levels in the figure. Thereby, the dark points have a depth of $1/2$ *and few points at* θ_3 *between* 0.20 *and* 0.25 *have a depth of* $1/4$*. To calculate full depth for this example the depth function was evaluated on a grid with a precision of* 0.001*.*

For an implementation of the ϵ sets defined by (46), we use a construction based on each combination of three nearby vertices to reduce redundant candidates. First, we calculate all distances from the vertices used in (47). Then, for each candidate vertex $\theta^{i,j}$, we select the closest candidates $\theta_a^{i,j}, \theta_b^{i,j}$ based on euclidean distances. For this three points, we then calculate an interior ϵ candidate for $\theta^{i,j}$ by

$$\theta_{cand}^{i,j} = \theta^{i,j} + \epsilon \cdot (\theta_a^{i,j} + \theta_b^{i,j} - 2 \cdot \theta^{i,j}).$$

Thereby, ϵ has to be small enough to assure that $\theta_{cand}^{i,j}$ is in the simplex defined by $\{\theta^{i,j}, \theta_a^{i,j}, \theta_b^{i,j}\}$. To assure this, it is helpful to use a representation of $\theta_{cand}^{i,j}$ given by

$$\theta_{cand}^{i,j} = \theta^{i,j} + \epsilon' \cdot \left(\left(\frac{\theta_a^{i,j} + \theta_b^{i,j}}{2} \right) - \theta^{i,j} \right).$$

In this representation $\epsilon' \in (0, 1)$ automatically implies that the candidate point is

in the relevant simplex. We restrict $\epsilon' \in (-\frac{1}{3}, \frac{1}{3})$. Then the resulting point is in the centre of the triangle or outside of it. In our application, we will evaluate $\theta^{i,j}_{cand}$ with $\pm\epsilon'$ and $0 < \epsilon' < \frac{1}{3}$ fixed, to explore two directions of each candidate.

Example 69. *If we apply the ϵ candidate set, with $\epsilon = 0.001$, to Example 68, we get the candidate set presented in Figure 24.*

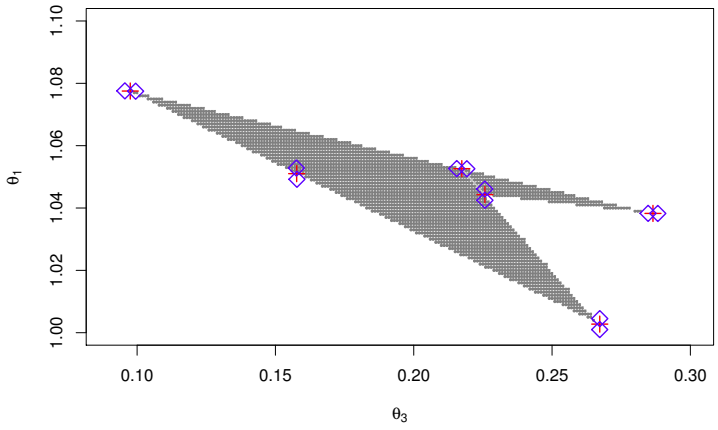


Fig. 24: Candidate set with $\epsilon = 0.001$ for a linear AR(1) model with intercept. The ($\Diamond$) show the points from the epsilon candidate set and (+) are the candidates at the interception of the root equations. The grey points are the parameters in the grid with depth larger than zero.

It is clearly visible that the candidate set now can be explored very efficiently by just few candidate points. The advantage to candidate set (46) is that we just need 12 *points in which depth has to be evaluated, while the set (47) results in* 40 *candidates.*

Parameter Confidence Set for the Linear Model Without Intercept

Having the candidates at hand, we can construct confidence regions for growth models by simplicial depth. Since we already know that depth is piecewise constant and we can identify all constant regions by candidate parameters close to the jump locations, confidence intervals can be constructed by an evaluation of the appropriately scaled and centred statistics on data generated candidate sets. Then, the convex hull delivers complete and simultaneous confidence regions for the model parameters.

Corollary 70. *An empirical $(1 - \alpha)$ confidence region based on the full simplicial*

depth for the AR(1) model without intercept is given by

$$\hat{\Theta}^{AR}_{1-\alpha}(y) = \left[min\{\theta \in \Theta^{d^{AR}_S}_{cand} | \theta \in \hat{\Theta}_{1-\alpha}(y)\}, max\{\theta \in \Theta^{d^{AR}_S}_{cand} | \theta \in \hat{\Theta}_{1-\alpha}(y)\} \right],$$

whereby $\hat{\Theta}(y)$ is defined as in (17). Alternatively, the confidence regions can be constructed by replacing $\Theta^{d^{AR}_S}_{cand}$ by $\Theta^{d^{AR},\epsilon}_{cand}$.

Proof. Since the candidate sets approximate the empirical depth shape, we asymptotically explore the complete depth function which is related to the asymptotic limit distribution to restrict the confidence set. Hence, the definition results in a finite sample approximation of the depth shape compared with the asymptotic distribution for testing. □

Note that the application of the ϵ method allows a finite sample correction by selecting ϵ very low. Then the bounds of the confidence region can be approximated more precise.

Example 71. *In the situation of Example 65 a 95% confidence interval is given by*

$$[0.00585, 0.01614]$$

for $\Theta^{d^{AR}_S}_{cand}$ candidate set and by

$$[0.00468, 0.01616]$$

for the $\Theta^{d^{AR},\epsilon}_{cand}$ set with $\epsilon = 0.00001$.

Parameter Confidence Set for the Linear Model With Intercept

To construct confidence sets for two parameter models, we use convex hulls of parameters which are the boundary of the two-dimensional non-rejection area. This leads to the following formula.

Corollary 72. *An empirical $(1-\alpha)$ confidence region based on the full simplicial depth for the AR(1) model with intercept is given by*

$$\hat{\Theta}^{ARi}_{1-\alpha}(y) = \overline{conv(\tilde{\Theta}_{1-\alpha}(y))}, \tag{48}$$

whereby $\tilde{\Theta}_{1-\alpha}(y)$ is the set of all candidates in $\Theta^{d^{ARi}_S}_{cand}$ or $\Theta^{d^{ARi},\epsilon}_{cand}$ which are in the confidence regions for the linear AR(1) model with intercept defined by (29).

We give another example for the constructed confidence regions based on the linear AR(1) model with intercept.

Example 73. *Consider a model given by*

$$Y_n = \theta_1 \cdot Y_{n-1} + Y_{n-1} + \theta_3 + E_n,$$

whereby $\theta_1 = 0.01, \theta_3 = 0.1$ *and* $E_n \sim \mathcal{N}(0, 0.1)$. *For* $y = (y_0, ...y_{100})$ *with* $y_0 = 15$, *we can construct* $1 - \alpha$ *confidence regions based on* d_S^{ARi}. *An example is depicted in Figure 25. Thereby, we used the candidate set (46) with* $\epsilon = 1 \cdot 10^{-19} \cdot (1, 1)^\top$. *The size of* ϵ *was set to a value below of the minimal euclidean distance of two candidate points.*

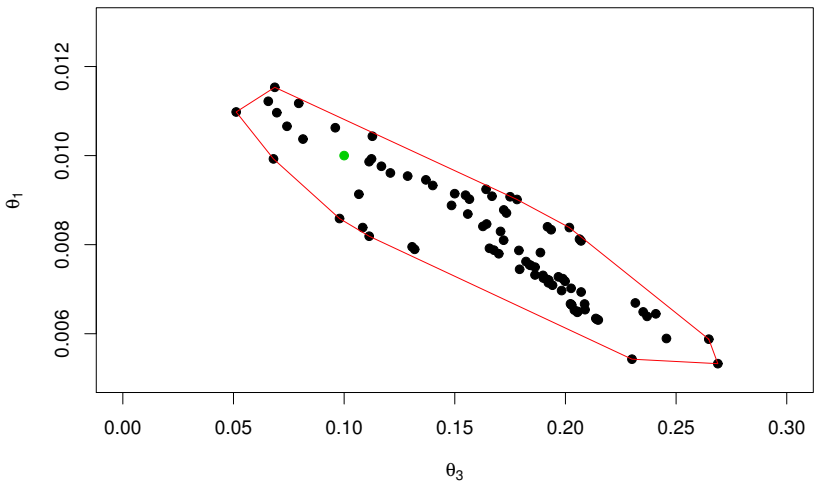


Fig. 25: Confidence region for a linear AR(1) process with intercept. The black dots are the candidate points which are contained in the confidence region. The red lines indicate the boundary of the interval. The green dot shows the true parameter.

Even if the candidate sets already reduce the number of evaluations of the depth statistics, the total number of evaluations depends on the number of observations. To allow an additional reduction of the candidates, one can exploit that the confidence sets based on the simplified depth notions are related to the full simplicial depth sets. Since the order of terms in the simplified statistics is N compared to N^K for the full depth statistics, the calculation of the simplified confidence regions can be used to restrict the candidate set. Hence, we define additional constructions for confidence regions derived from the simplified depth statistics. The empirical

confidence regions based on the simplified statistics are constructed similarly to the regions for the full simplicial depth.

Corollary 74. *An empirical* $(1-\alpha)$ *confidence region based on the simplified depth notions* d_S^i *for the AR(1) model with intercept is given by*

$$\hat{\Theta}_{1-\alpha}^{i,ARi}(y) = \overline{conv(\tilde{\Theta}_{1-\alpha}(y))},$$

whereby $\tilde{\Theta}_{1-\alpha}(y)$ *is the set of all candidates* $\Theta_{cand}^{d_S^{ARi}}$ *or* $\Theta_{cand}^{d_S^{ARi,\epsilon}}$ *which are in the confidence region based on* d_S^i *for the linear AR(1) model with intercept.*

Remark 75. *The candidate sets for the simplified depth notions can be reduced by consideration of intersections from relevant groups of residuals based on the definitions of the statistics. However, since the full candidate set includes these values, we do not analyse further candidate reductions here.*

Since the computation of the full depth regions is based on the calculation of the full depth statistic which computationally is more complicated than the calculation of the simplified notions, we propose another definition of confidence regions. We restrict the candidates for the full depth set to the set of values which are included in the confidence regions from the simplified notions. Based on the upper Corollary, we modify Corollaries 72 and 74 to the following reduced intervals.

Conjecture 76. *An empirical* $(1-\alpha)$ *confidence region for the AR(1) model with intercept is given by*

$$\hat{\Theta}_{1-\alpha}^{ARi}(y) = \overline{conv(\tilde{\Theta}_{1-\alpha}(y))},$$

whereby $conv(M)$ *denotes the convex hull of* M *and* $\bar{M}$ *is the closure of* M*. Further,* $\tilde{\Theta}_{1-\alpha}(y)$ *is the set of all candidates in a* d_S^i *confidence region with* $(1-\alpha)$ *level for the linear AR(1) model with intercept for* $i \in \{1,2,3\}$*.*

In Section 5.3, we show a simulation study which supports the conjecture.

4.4 Calculation of Confidence Regions for the Non-Linear AR Model

Candidates for the Non-Linear Model

To define a candidate set for the non-linear model, a similar reasoning as in the linear case can be applied. Again, the root equation of the residuals is a central

object. In the non-linear case it is given by

$$r_n(\theta, y) = y_n - y_{n-1} - \theta_1 y_{n-1}^{\theta_2} = 0. \tag{49}$$

By checking the intersections of three curves given by (49), one gets surrounded sets of parameters for which the residual signs alternate.

Lemma 77. *Consider two tupelo of observations* (y_n, y_{n-1}) *and* (y_m, y_{m-1}) *with roots of the residuals defined by*

$$r_n(\theta, y) = y_n - y_{n-1} - \theta_1 y_{n-1}^{\theta_2} = 0,$$
$$r_m(\theta, y) = y_m - y_{m-1} - \theta_1 y_{m-1}^{\theta_2} = 0.$$

Then the lines, defining the sign changes of the residuals, intersect in (θ_1^*, θ_2^*) *with*

$$\theta_2^* = \frac{log(\frac{y_n - y_{n-1}}{y_m - y_{m-1}})}{log(\frac{y_{n-1}}{y_{m-1}})}$$

and

$$\theta_1^* = \frac{y_n - y_{n-1}}{y_n^{\theta_2^*}}.$$

Proof. Here, we first solve for θ_2. We start with the root equations $r_n(\theta, y) = 0$ and $r_m(\theta, y) = 0$. Solving these for θ_1 delivers

$$\frac{y_n - y_{n-1}}{y_{n-1}^{\theta_2}} = \theta_1$$

and

$$\frac{y_m - y_{m-1}}{y_{m-1}^{\theta_2}} = \theta_1.$$

Hence, the lines intersect, if

$$\begin{aligned}
&\Leftrightarrow \frac{y_n - y_{n-1}}{y_{n-1}^{\theta_2}} = \frac{y_m - y_{m-1}}{y_{m-1}^{\theta_2}} \\
&\Leftrightarrow \frac{y_n - y_{n-1}}{y_m - y_{m-1}} = \frac{y_{n-1}^{\theta_2}}{y_{m-1}^{\theta_2}} \\
&\Leftrightarrow \frac{y_n - y_{n-1}}{y_m - y_{m-1}} = \left(\frac{y_{n-1}}{y_{m-1}}\right)^{\theta_2} \\
&\Leftrightarrow log\left(\frac{y_n - y_{n-1}}{y_m - y_{m-1}}\right) = \theta_2 \cdot log\left(\frac{y_{n-1}}{y_{m-1}}\right) \\
&\Leftrightarrow \frac{log\left(\frac{y_n - y_{n-1}}{y_m - y_{m-1}}\right)}{log\left(\frac{y_{n-1}}{y_{m-1}}\right)} = \theta_2^*.
\end{aligned}$$

For θ_1 we use θ_2^* and the root-equation for n. We get

$$\begin{aligned}
&y_n - y_{n-1} - \theta_1 y_{n-1}^{\theta_2^*} = 0 \\
\Rightarrow &y_n - y_{n-1} = \theta_1 y_{n-1}^{\theta_2^*} \\
\Rightarrow &\frac{y_n - y_{n-1}}{y_{n-1}^{\theta_2^*}} = \theta_1^*.
\end{aligned}$$

□

To calculate the candidates, we set

$$\gamma(n,m) := \frac{log\left(\frac{y_n - y_{n-1}}{y_m - y_{m-1}}\right)}{log\left(\frac{y_{n-1}}{y_{m-1}}\right)}$$

and define

$$\theta^{n,m} := \left(\frac{y_n - y_{n-1}}{y_n^{\gamma(n,m)}}, \gamma(n,m)\right).$$

Approximating the non-linear candidate sets by simplexes with edges, defined by the root-equations for the residuals, delivers the following candidate set:

$$\Theta_{cand}^{d_S^{nAR}} = \left\{ \frac{1}{3}(\theta^{i,j} + \theta^{i,k} + \theta^{j,k}) \middle| i,j,k \in \{1,...,N\}, i \neq j, i \neq k, j \neq k \right\}.$$

In Figure 26, the non-linear situation is depicted. For convenience, we express the root equations by $\theta_1 = f_n(\theta_2)$ with

$$\theta_1 = \frac{y_n - y_{n-1}}{y_{n-1}^{\theta_2}}$$

here. Thereby, it is easy to see that with growth assumptions on $(y_n)_n$ and $y_n > 0$, we get intercepts at $y_n - y_{n-1}$ for $\theta_2 = 0$ and asymptotes at 0 for θ_2 tending to infinity. One can also easily verify that the functions are strictly decreasing under growth assumptions.

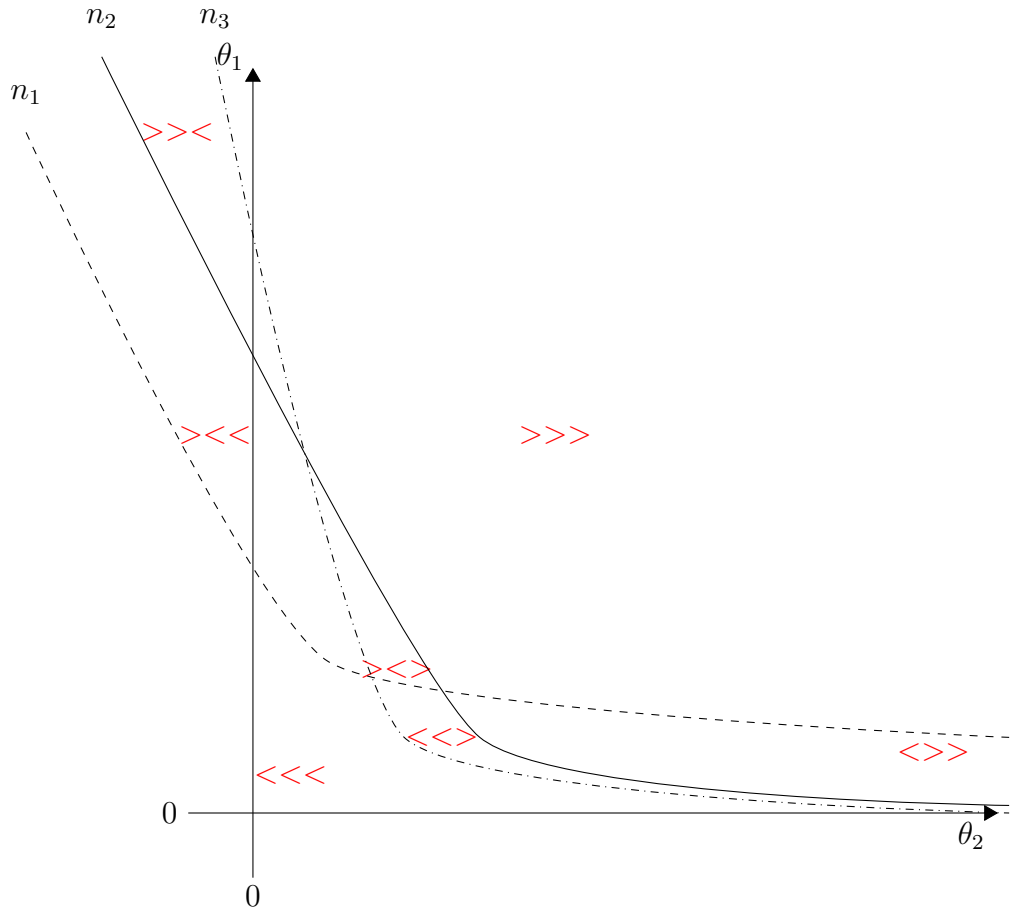


Fig. 26: Residual signs for the non-linear AR(1) growth model. The lines annotated with n_i are the curves given by the equation $y_{n_i} - y_{n_i-1} - \theta_1 y_{n_i-1}^{\theta_2} = 0$ with $n_1 < n_2 < n_3$ and $y_n < y_m$ for $n < m$. The signs in the enclosed regions show the signs of the respective residuals r_1, r_2, r_3.

Again, the region with alternating signs coincides with the enclosed region. By linear approximation, we can use the centre point of a triangle to find reasonable candidates. Otherwise, the alternative candidate set definition can be used. Analogously

to the previous models, an ϵ set for the non-linear model can be defined by

$$\Theta_{cand}^{d_S^{nAR,\epsilon}} = \left\{ \theta^{n,m} \pm \epsilon \middle| n, m \in \{1, ..., N\}, n \neq m \right\}.$$

For large sample sizes it suffices to check the vertices of the resulting simplexes. Hence the computation of empirical depth can be reduced to a countable set of candidates.

Parameter Confidence Set for the Non-Linear Model

For the non-linear model, we need to change the approximation method to define the bounds of the confidence regions based on the evaluated candidates. This is necessary, since in the linear model it is sufficient to consider a linear boundary representation which already describes equi-depth lines with respect to each parameter combination. For the non-linear model this equi-depth lines are curves given by the model equation. The calculation of these curves is computationally costly. On the other hand, convex hulls would not approximate the non-linear boundaries, defined by these curves. Since a full computation of these curves does not improve the results compared with its computational costs, we approximate the bounds by using so called Alpha-Shapes. Since α in this thesis is referring to the levels of tests and confidence intervals, we change the notation for the geometrical Alpha-Shapes to τ shapes here.
A τ shape is a generalised convex hull for scatter data. It was introduced by Edelsbrunner et al. (1983). First, we define a simplex.

Definition 78. *For $k+1$ vectors $u_0, ..., u_d \in \mathbb{R}^n$,$n \geq k+1$ with $u_1 - u_0, ..., u_k - u_0$ being linearly independent, a k dimensional simplex is defined by*

$$\Delta = \left\{ x = \sum_{i=0}^{k} \beta_i u_i \middle| \sum_{i=0}^{k} \beta_i = 1, \beta_i \geq 0 \forall i \in \{0, ..., k\} \right\}.$$

To define τ shapes, we consider a set of points denoted by $S \in \mathbb{R}^{d+1}$. Thereby, the points are assumed to be in general position. Then, for each subspace $T \subset S$ with dimension $dim(T) = k+1 < dim(S) = d+1$ and $|T| = k+1$, the convex hull Δ_T is a k dimensional simplex. This allows us to define τ shapes by τ balls and subsets of S.

Definition 79.
(i) For $\tau \in (0, \infty)$ the τ ball b_τ is defined as open ball with radius τ.

(ii) An τ ball b_τ is empty, if $b_\tau \cap S = \emptyset$.
(iii) A k dimensional simplex Δ_T is τ exposed, if there exists an empty τ ball b_τ with $T = \delta b_\tau \cap S$, whereby δb_τ is the boundary of b_τ.

Figuratively, a τ exposed subset of S is a k dimensional simplex which forms a τ ball that does not contain any other points of S in its interior.
To define a τ shape we define its boundary.

Definition 80. *For a set of points S with $dim(S) = d + 1$, the boundary of the τ shape S_τ is defined by*

$$\delta S_\tau = \{\Delta_T | T \subset S, dim(T) = |T| \leq d, \Delta_T \ \tau \ exposed\}.$$

So, the boundary of the τ shape is the union of all k dimensional simplexes with $0 \leq k \leq d$ which are τ exposed. Since the total boundary of S_τ consists of boundaries of simplexes one can show that δS_τ coincides with the boundary of a polytope. For us, τ shapes are useful due to the following statement.

Lemma 81. *For a set of points in general position, S with $dim(S) = d + 1$ and $\tau \in (0, \infty)$*

$$\lim_{\tau \to 0} S_\tau = S$$

and

$$\lim_{\tau \to \infty} S_\tau = conv(S)$$

holds.

Proof. For $\tau \to 0$, we can choose τ small enough, so that every singular point in S is τ exposed in a non-trivial simplex. Then the resulting shape just consists of the points in S and simplexes, shrinking to these points.
For $\tau \to \infty$, we can select τ larger than the radius necessary to include all points of S. Then no simplex in this ball can be τ exposed anymore. But also no further points can exclude $conv(S)$, so that the limit of the τ shape is the maximal simplex formation $conv(S)$. □

So, a τ shape can be used to explore a scatter structure more precise than a convex hull, but more coarse than by using all points. This helps us to construct confidence regions for the non-linear AR process based on the candidate points.

Corollary 82.
(i) An empirical $(1-\alpha)$ *confidence region based on full simplicial depth for the non-linear AR(1) model is given by*

$$\hat{\Theta}_{1-\alpha}^{nAR}(y) = \overline{conv_\tau(\tilde{\Theta}_{1-\alpha}(y))},$$

whereby $conv_\tau(M)$ *denotes the* τ *shape of* M *with an appropriate localisation parameter* τ*. Further,* $\tilde{\Theta}_{1-\alpha}(y)$ *is the set of all candidates in the* d_S^i *confidence region with* $1-\alpha$ *level for the non-linear AR(1) model or one of the unrestricted candidate regions* $\Theta_{cand}^{d_S^{nAR}}$ *or* $\Theta_{cand}^{d_S^{nAR},\epsilon}$*.*
(ii) The empirical $(1-\alpha)$ *confidence regions for the simplified notions are defined by replacement of* d_S^{nAR} *by the simplified notions* d_S^i*.*

4.5 Implementation of the Proposed Confidence Sets

Algorithms

We now want to present algorithms to implement the confidence regions. The algorithm for the one parameter confidence intervals is a straightforward implementation.

Algorithm 83.
Given: Observation $y = (y_0, ..., y_N)$*.*

1. *Calculate the candidate set given by (43) or (44).*
2. *Evaluate the test given by (16) on each candidate.*
3. *Mark all non-rejected candidates.*
4. *Define the confidence interval by the minimal and maximal non-rejected candidates.*

When two dimensional parameters are considered the construction is similar to the one parameter case for the simplified depth notions. For the full depth, we apply further steps to allow a faster calculation.

Algorithm 84.
Given: Observation $y = (y_0, ..., y_N)$*, Parameter* τ *in case of the non-linear model.*

1. *Calculate the candidate set given by (46) or (47).*
   If a simplified statistic was selected then proceed with:
   (a) *Evaluate the tests on each candidate.*

If the full simplicial depth is selected proceed with:

(a) Initialise a multi-core cluster, if available. Else use one core.

(b) Evaluate the test parallel on each candidate.

If the full simplicial depth with parameter preprocessing is selected proceed with:

(a) Evaluate the d_S^1 confidence region.

(b) Restrict the candidates to all candidates in the d_S^1 confidence region.

(c) Initialise a multi-core cluster, if available. Else use one core.

(d) Evaluate the test parallel on each candidate in the restricted set.

2. *Mark all non-rejected candidates.*

3. *Define the confidence region.*

   - *Based the convex hull for the linear model,*
   - *Based on the τ shape for the non-linear model,*

   with all non-rejected candidates.

Thereby, Algorithm 84 applies to the linear and the non-linear models by application of the appropriate formulas for the residuals and the appropriate candidate sets. Note that the preprocessing is not recommended for the non-linear model.
The proposed approach has some central advantages. Due to the candidate calculation, we have a data generated grid of relevant parameters. This avoids the evaluation of an unknown grid in $\mathbb{R}$ or $\mathbb{R}^2$ to calculate depth. Further, the simplified depth notions allow faster approximations of the relevant set for full depth calculation, since the candidates can be reduced by the evaluation of test statistics which are easier to calculate than the full depth. Finally, the algorithm introduces flexibility, since the test statistics can be applied to various models, if the residuals can be calculated. Then just some technical assumptions, like in our case the growth assumption, have to be checked to allow an application of the confidence set algorithm. In particular, we are able to construct parameter confidence sets without assumptions on the exact distributions of the data and the errors.

Examples of Implemented Confidence Sets

The first example compares the full depth confidence regions with the regions defined by the simplified statistics for the linear autoregressive model with intercept.

Example 85. *In continuation of Example 73, we calculate the* 95% *confidence regions for* d_S^{ARi} *and the three simplified notions* $d_S^i, i \in \{1,2,3\}$. *In Figure 25, the resulting regions are depicted. It is obvious that all regions are concentrated around the true parameter.*

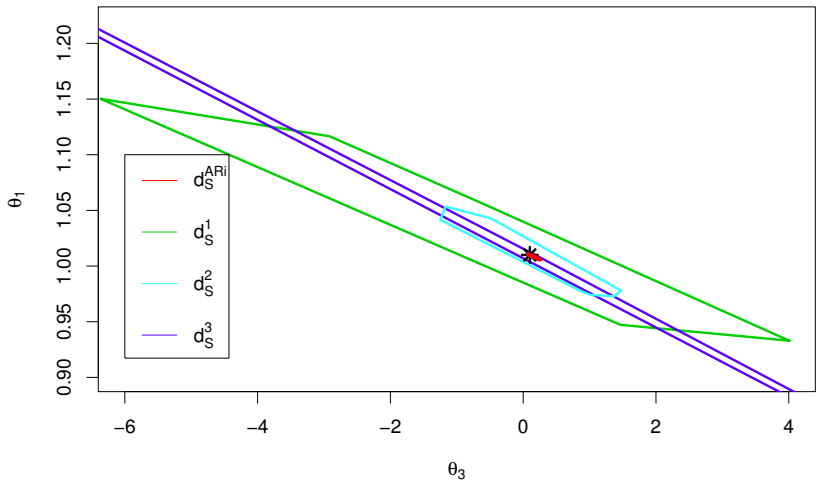


Fig. 27: Confidence regions for a linear AR(1) process with intercept based on different test statistics. $(*)$ is the true parameter.

The d_S^{ARi} *region defines the narrowest region. This is not surprising, since the* d_S^{ARi} *statistic also defines a much sharper test, as will be shown in the next sections. The* d_S^2 *confidence region appears to be the best region defined by the simplified statistics, followed by the* d_S^1 *version. The* d_S^3 *version leads to the widest region, since a large deviation on the diagonal can be observed.*
In Figure 28, full simplicial depth confidence regions based on different levels are depicted.

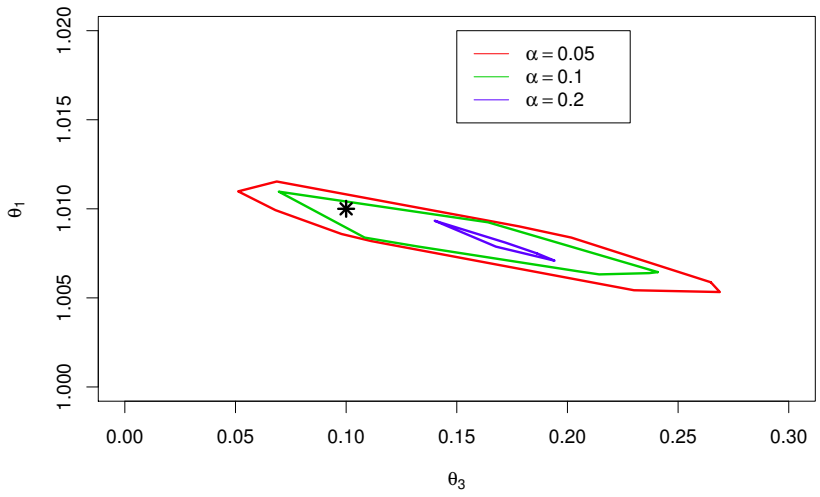


Fig. 28: d_S^{ARi} confidence regions for a linear AR(1) process with intercept based on different levels. $(*)$ is the true parameter.

We see that, as expected, a higher level $(1-\alpha)$ *produces larger confidence regions. In Figure 29, we present confidence regions for increasing sample sizes at a level of* 95%.

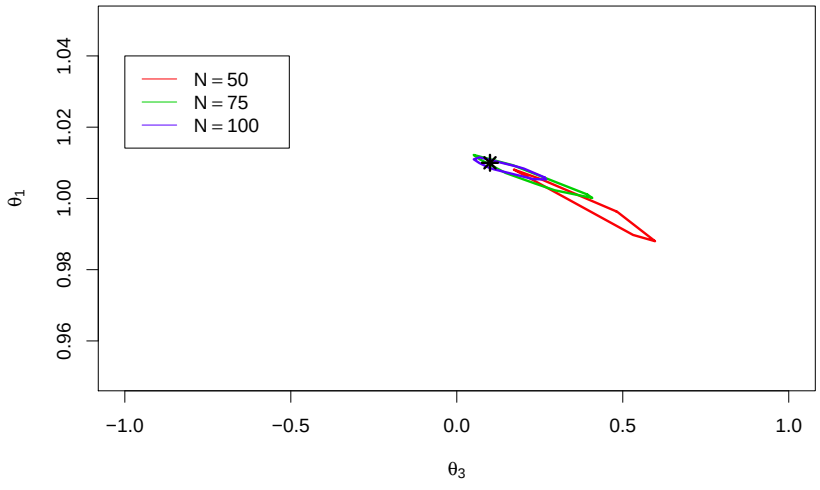


Fig. 29: d_S^{ARi} confidence regions for a linear AR(1) process with intercept based on different sample sizes N. $(*)$ is the true parameter.

Again, as expected, the confidence regions shrink to the true parameter, if the sample size increases. For low sample sizes, the confidence sets do not necessary include

the true parameter, due to random deviations.

The next example shows, how the confidence regions in the non-linear case can be applied.

Example 86. *Now, we consider a non-linear model defined by*

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + E_n,$$

with $\theta_1 = 0.01, \theta_2 = 1.01, E_n \sim \mathcal{N}(0, 0.1)$ *and* $y_0 = 15$. *First, we compare the confidence regions for a simulated process with* $N = 100$ *based on the different test statistics and a* 95% *level. The result can be found in Figure 30. We used* $\tau = 0.8$ *in all simulations.*

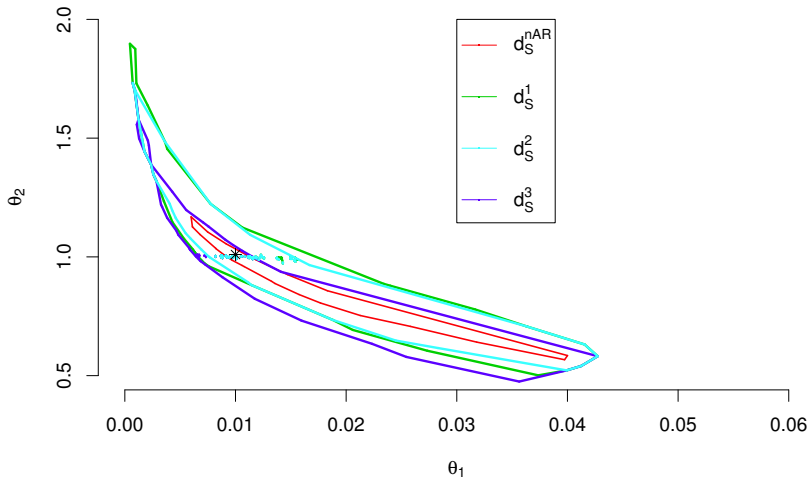


Fig. 30: Confidence regions for a non-linear AR(1) process based on different test statistics. $(*)$ is the true parameter.

We see that the confidence regions reflect the non-linearity of the data. The full depth defines the smallest confidence region and is surrounded by all simplified regions. In Figure 31, confidence regions for different levels α *based on full simplicial depth are depicted.*

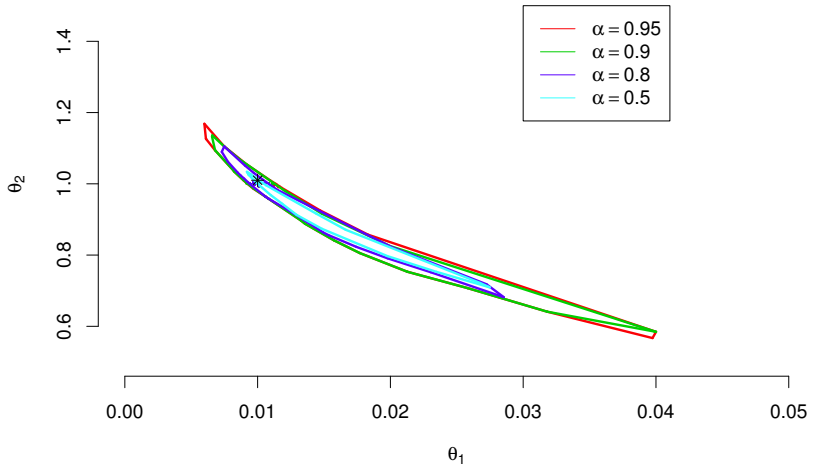


Fig. 31: d_S^{nAR} confidence regions based on full simplicial depth for a non-linear AR(1) process with different confidence levels α. $(*)$ is the true parameter.

The confidence regions are large when a high level is selected. In Figure 32, the confidence intervals for increasing series length on a level of 90% *are depicted.*

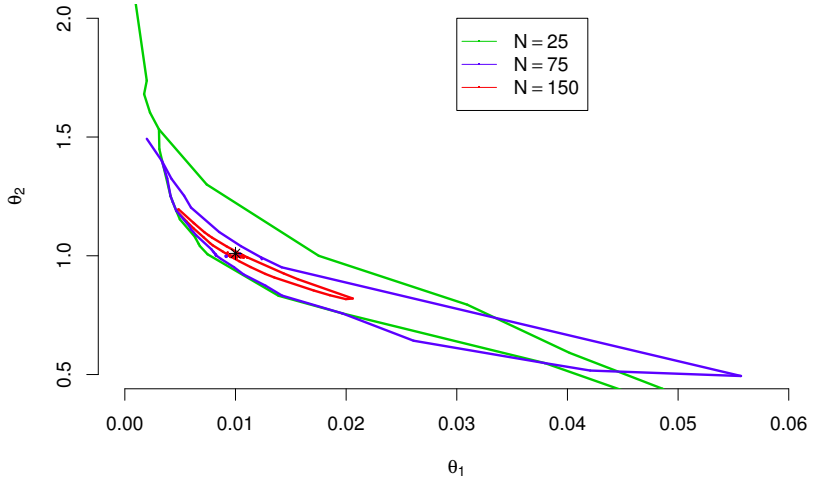


Fig. 32: d_S^{nAR} confidence regions based on full simplicial depth for a non-linear AR(1) process with different sample size N. $(*)$ is the true parameter.

Here, we see that the confidence regions shrink to the true parameter, if the sample size increases. This indicates a consistency of constructed confidence regions.

4.6 Estimation

Algorithms

The candidate set, proposed for the calculation of confidence regions, can also be used for parameter estimation. For estimation, we need to maximise depth in the parameter θ. Since we already know that depth just changes in the candidate points, we can restrict our maximisation to the candidate sets. Unfortunately, these sets define a very irregular grid and the objective function is piecewise constant. Hence, standard optimization procedures have difficulties to find global maxima.
We propose a data based optimisation method which uses the candidate set and appropriate distance measures to iterate. Thereby, the global maximiser of depth can be found relatively fast while the number of evaluations of the depth statistics can be held low.
For the one parameter model no real optimisation is necessary, since the candidates with maximal depth can simply be calculated as the median of the candidate points. From this inner interval all neighbouring candidates have to be checked until depth decreases. The resulting interval then defines the complete set of parameters with maximal depth.
The optimisation procedure for the linear model with two parameters is defined by the following algorithm.

Algorithm 87.
Given: Observation $y = (y_0, ...y_N)$.

1. *Calculate the candidate points.*
2. *Calculate a starting value* $\theta^{(0)}$ *based on the component wise medians of the candidates.*
3. *Calculate the distances of all candidates from the starting value* $\theta^{(0)}$ *(Thereby the norm on which the distances are based can be varied).*
4. *Restrict the candidates to values close to the starting value. Therefore, we define a maximal distance by an accuracy parameter* a.
5. *Calculate depth for all restricted candidates.*
6. *Begin an iterative optimisation until the maximum number of iteration is reached or the change of the maximum is below of a critical value* ϵ_{ter}. *For* $k \in \{1, ..., K_{max}\}$ *perform the following loop.*

(a) *Calculate candidate parameters* $\Theta_{cand} = \{\theta \in \Theta : \|\theta - \theta^{(k-1)}\| < a\}$

(b) *Calculate depth for the restricted candidates.*

(c) *Set* $\theta^{(k)} = \underset{\theta \in \Theta_{cand}}{argmax}\ d_S^{ARi}(\theta, y)$ *as the depth maximising parameter on the new restricted candidate set.*

(d) *If* $k = K_{max}$ *or* $\|\theta^{(k)} - \theta^{(k-1)}\| < \epsilon_{ter}$*, end iteration.*

7. *If multiple candidates have maximal depth, a point in the convex hull spanned by them is the resulting depth maximiser.*

Since this algorithm does not depend on the actual model, it can be applied to the linear and non-linear model by usage of the appropriate candidate sets and depth functions. The estimators are implemented in the *rexpar* package.

Examples of the Depth Based Estimators

Example 88. *In the situation of Example 65, the estimation function results in* $\hat{\theta}(y) = 0.009471$ *by using an accuracy parameter of* $a = 0.1$ *defining the search regions and* $\epsilon_{ter} = 0.000001$ *for the termination criterion. The result is unified by taking the mean of all candidates with maximal depth. The estimate comes from two considered candidates with maximal depth given by* 0.009470 *and* 0.009472*. It converges after two iterations. The result is in the full set of parameters with maximal depth. The complete region can be explored by using an accuracy parameter* $a = 1$ *in the algorithm. Thereby, all candidates are evaluated. Then, we get four candidates with maximal depth given by* $0.009472, 0.008941, 0.009470, 0.011560$ *and can conclude that depth is maximal on the interval* $[0.008941, 0.011560]$*. An unified solution, by taking the mean, then is* $\hat{\theta}(y) = 0.009861$*.*

Example 89. *To illustrate the usage of the two dimensional estimators, we consider a linear AR(1) model with intercept defined by* $(\theta_1, \theta_3) = (0.01, 0.2)$*, i.i.d.* $\mathcal{N}(0, 0.1)$ *errors and a series length of* $N = 25$*. Here, the simplicial depth estimate results in* $\hat{\theta}(y) = (0.023880561, -0.000350437)$ *as unified maximising candidate point. Thereby, the algorithm was applied on the vertices of the candidate simplexes only, to reduce the computational costs. The estimate results in* $d_S^{ARi}(\hat{\theta}(y), y) = 0.2281197$ *which is below the asymptotic maximal depth value of* $\frac{1}{4}$*. To check, if we missed relevant points by the reduction to the vertices of the candidate simplexes, we also estimate depth by consideration of larger candidate sets. When we use the inner points of all possible simplexes the size of the candidate set and hence the computation time increases. But we also get a better*

estimate $\hat{\theta}(y) = (0.02439084, 0.01530349)$ *with* $d_S(\hat{\theta}(y), y) = 0.3083004$. *By application of the* ϵ *candidate set, we can reduce the computational costs and get* $\hat{\theta}(y) = (0.02317518, 0.01259367)$ *with* $d_S^{ARi}(\hat{\theta}(y), y) = 0.3083004$. *So, both results are indifferent with respect to depth maximisation. Thereby, the estimator results are unified by selection of one maximising candidate which first component is the median of the first components from all maximising candidates and the second is the respective* θ_3 *to the maximising first component.*
To answer the question, if the methods really have converged to the true maximising region, we relax the search precision parameter, which was set to 0.1 *so far, to a value of* 1. *Then, all candidates are evaluated to maximise depth. In case of the* ϵ *candidate set, we now get* $\hat{\theta}(y) = (0.01998668, 0.08941891)$ *with* $d_S(\hat{\theta}(y), y) = 0.3094297$. *By using the inner points of all simplexes we get* $\hat{\theta}(y) = (0.01998668, 0.08941891)$ *with* $d_S(\hat{\theta}(y), y) = 0.3094297$. *This shows that the simplification methods work quite reliable, but we have to pay the cost of a lower precision.*

Remark 90. *The separate estimation procedure is only interesting, if the estimation of the region with maximal depth is of main interest. As far, as confidence regions are of interest, the region with maximal depth is included in the the resulting sets and a separate maximisation algorithm is not necessary. However, an improved version of the confidence set construction can use this fact in reverse. By a fast estimation of the region with maximal depth, the confidence regions can be explored from the inside. This can additionally reduce the computational costs of the confidence region construction, since the calculation can be stopped, if the set is explored completely. An implementation of this idea is future work.*

4.7 Change Point Detection

Algorithms for a Heuristic Change Point Detection

The basic idea to detect change points in growth processes, based on the proposed depth based statistics, is to compare the estimates and confidence intervals based on data ranges covering the period before and after a potential change point. By rolling windows, a complete series can be examined for change points. In an univariate setting the idea can be described as follows.

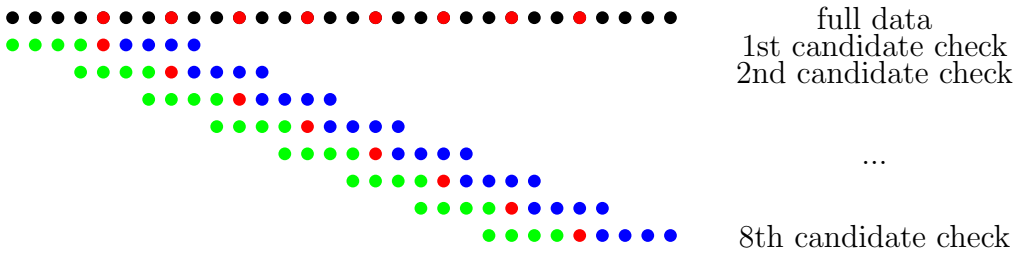


Fig. 33: Phase change candidates from $(x_1, ..., x_M)$. The black dots illustrate the complete data. The red dots are the considered candidate points. The green dots define the data for the respective pre-candidate window and the blue dots are set by the post-candidate window.

Under the assumption that the underlying process is observed at equidistant points in time and denoted by $y = (y_0, ..., y_N)$, we preselect candidates $x_1, ..., x_M, M < N$ from the observed series which we want to check as change points. These candidates can be all observations or a thinned set, for example to speed up computation.
In Figure 33, the full set of observations is depicted as the series of black dots on the top of the figure. The candidates, which are also observations, are marked as red dots. Now, for each candidate a left- and a right hand side window of observations is considered. In the first step, we see the first four observations from the left hand window and the observations six to ten form the right hand window with respect to the first candidate. Based on these two windows we check, if the first candidate is a potential change point by two alternative rules. The first rule just checks, if the left hand and right hand side parameter confidence regions overlap by calculation of the regions based on the two non overlapping observation windows. The second rule calculates the parameter confidence regions and the parameter estimates for each window and then checks, if the left hand side estimate is covered by the right hand side confidence region and vice versa. This can be quickly done by calculation of the estimates and applying the depth based tests to the estimates and the non-respective data. If both estimates are rejected, we identify the candidate point as a potential change point.
The algorithm then proceeds by checking the next candidate with the same method. After all candidates are checked, a series of marked potential change points is available. Since we assume that the real underlying process is continuous, it is possible that subsequent candidates are marked as change points. This for example happens, if there is a continuous transition between two phases. Therefore, we apply a post processing to collect subsequent change points and marking just one point at the median observation index as resulting change point. Another post processing step

robustifies the results to avoid fragments which are detected as change points when the errors are positive or negative in sequence.
We now give a more detailed definition of the detection algorithms. The first algorithm defines a change point by non-overlapping confidence regions.
Since theoretical results on the proposed change point detection algorithms are beyond the scope of this thesis, we define the methods as heuristic algorithms here.

Algorithm 91.
Given:

- $(x_1, ..., x_M)$ *set of candidate points as subset from the observation* $(y_0, , ..., y_N)$, $M < N$.
- *Bandwidth* $b \in \mathbb{N}$.
- *Level* $\alpha \in (0, 1)$.
- *Post processing parameters* m *and* m_p, *whereby* $m_p \in (0, 1)$ *is the percentage of length of* m*-th change point cluster as sensitivity parameter.*

With these parameters the algorithm is defined by:

1. *Select a candidate* x_i.
2. *Define a pre-*x_i *set* $y^- := (y_{j-b}, y_{j-b+1}, ..., y_{j-1})$, *with* j *denoting the index of the observation* y_j *which is related to the candidate* x_i.
3. *Define a post-*x_i *set* $y^+ = (y_{j+1}, y_{j+1}, ..., y_{j+b})$, *with* j *denoting the index of the observation* y_j *which is related to the candidate* x_i.
4. *Calculate a pre-*x_i *confidence interval based on* y^- *denoted by* $\hat{\Theta}^-_{\alpha,b,i}(y^-)$.
5. *Calculate a post-*x_i *confidence interval based on* y^+ *denoted by* $\hat{\Theta}^+_{\alpha,b,i}(y^+)$.
6. *Mark* j *as potential change point, if*

$$\hat{\Theta}^+_{\alpha,b,i}(y) \cap \hat{\Theta}^-_{\alpha,b,i}(y) = \emptyset.$$

7. *Loop over* i.
8. *Check marked possible change points for successive subsets with respect to* $(x_1, ..., x_M)$.

9. *Set a minimal amount of successive possible change points k to define a resulting change by m_p of the length of the m-longest detected successive set of possible change points.*

10. *Neglect possible change points, if less then k successive possible change points define the corresponding subset.*

11. *For remaining successive possible change points store the median of the indices forming a possible candidate group as change point.*

The second algorithm defines change points as candidates for which the estimates are not covered by the non respective confidence regions.

Algorithm 92.
Given:

- $(x_1, ..., x_M)$ *set of candidate points as subset from the observation* $(y_0, , ..., y_N)$, $M < N$.
- *Bandwidth* $b \in \mathbb{N}$.
- *Level* $\alpha \in (0, 1)$.
- *Percentage of length* m_p *of* m*-th change point cluster as sensitivity parameter.*

With these parameters the algorithm is defined by:

1. *Select a candidate* x_i.
2. *Define a pre-*x_i *set* $y^- := (y_{j-b}, y_{j-b+1}, ..., y_{j-1})$, *with* j *denoting the index of the observation* y_j *which is related to the candidate* x_i.
3. *Define a post-*x_i *set* $y^+ = (y_{j+1}, y_{j+1}, ..., y_{j+b})$, *with* j *denoting the index of the observation* y_j *which is related to the candidate* x_i.
4. *Calculate a pre-*x_i *estimate* $\hat{\theta}^-_{b,i}(y^-)$.
5. *Calculate a post-*x_i *estimate* $\hat{\theta}^+_{b,i}(y^+)$.
6. *Mark* j *as potential change point, if*

$$\varphi(\hat{\theta}^-_{b,i}(y^-))(y^+) \cdot \varphi(\hat{\theta}^+_{b,i}(y^+))(y^-) = 1$$

7. *Loop over* i.

8. *Check marked possible change points for successive subsets with respect to* $(x_1, ..., x_M)$.

9. *Set a minimal amount of successive possible change points* k *to define a resulting change by* m_p *of the length of the* m*-longest detected successive set of possible change points.*

10. *Neglect possible change points, if less then* k *successive possible change points define the corresponding subset.*

11. *For remaining successive possible change points store the median of the indices forming a possible candidate group as change point.*

Both detection algorithms do not crucially depend on the fact, if we consider the linear models or the non-linear one. Nevertheless, the implementation is slightly different when non-linear processes are considered. In particular, the calculation of confidence intervals and estimators has to be modified to allow a valid approximation as discussed in the previous section.
While Algorithm 91 depends on the calculation of confidence regions, Algorithm 92 just needs the result of the depth based estimator and the respective depth to compare the result with the quantile of the asymptotic distribution. Hence, change points can be computed faster by the second algorithm. The estimation of parameters by simplicial depth has been discussed in Section 4.6. A simulation study on the performance of the methods is presented in Section 5.
Especially in case of parameters with more than one dimension this approach is quite promising, since an interesting property of simplicial depth can be used.
While in the one dimensional case, simplicial depth for AR processes is automatically unimodal (see Lemma 29) due to the symmetry of the test statistic, this does not happen in the two parameter case (see Remark 31), if the parameter for example changes in the observed process. This observation introduces additional information for real world applications. If the resulting empirical depth, calculated by the full sample, is significantly lower than the theoretical maximal value or, if the empirical depth shape is not unimodal, one can assume phase changes. To have a preliminary estimate of the number of these changes, the empirical depth shape can be evaluated with variable θ. The number of modes then can be used to get a starting guess for the number of phases in the observed process. To detect the change locations, the phase change detection algorithms can be applied. Another possibility is the application of local depth, as proposed by Agostinelli and Romanazzi (2011) or Paindaveine and Van Bever (2013) to estimate the number of modes of the local depth distribution.

Here, appropriate versions of local simplicial depth have to be calculated. Then, the preliminary estimation by modes in the depth shape can even be applied to one parameter models.

Examples of the Implemented Change Point Detection Methods

We start with an example based on nearly deterministic time series for the linear model without intercept. For the first examples, we just evaluate the method based on Algorithm 92.

Example 93. *In Figure 34, the detection for a time series with little noise and true change points at observations 100 and 200 is depicted. In the upper subfigure, the series of estimates and confidence sets are presented. The blue dots show the left hand side estimates while the red dots illustrate right hand side estimates. Further, dashed lines illustrate the respective confidence sets. In the lower subfigure, the series is shown. The green lines indicate potential change points detected by our method. We can observe that two large sets of succeeding potential change points are detected. In addition, three sets of just few succeeding points exist.*

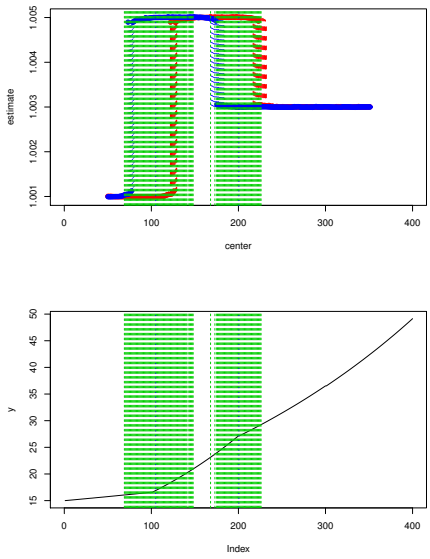


Fig. 34: Change point identification for nearly deterministic series.

To reduce the influence of random sequences from succeeding positive or negative residuals, we introduce a rule to neglect small sets of succeeding potential change points. The largest set of marked points defines the minimal number of succeeding

values to define a change point which can be detected by our method by a percentage m_p of its size. Since in our example the second block of potential change points is nearly as big as the largest, two sets of possible points define the resulting change points. To account for the continuity of the original series, we do not declare each candidate as change point, but the median of all indices forming a group of succeeding possible candidates. In the example, the first change point is detected at observation 105 and the second at observation 200.5.

The next examples show the performance of the method in repeated and more realistic situations.

Example 94. *In this example, we examine the method when the errors are normally distributed and one true change point exists. For all simulations, we use errors defined by $E_n \sim \mathcal{N}(0, 0.2)$ and a growth parameter $\kappa_1 = 1.001$ for observations 1 to 200 and $\kappa_1 = 1.005$ for observations 201 to 400. Hence, we have got one true change point at observation* 201*. In Figure 35 an example series with the identified change points is depicted. We apply $\alpha = 0.01, b = 49, m = 1, m_p = 0.5$ and evaluate every data point as possible change point (i.e. $(x_1, ..., x_M) := (y_{b+1}, ..., y_{N-b+1})$).*

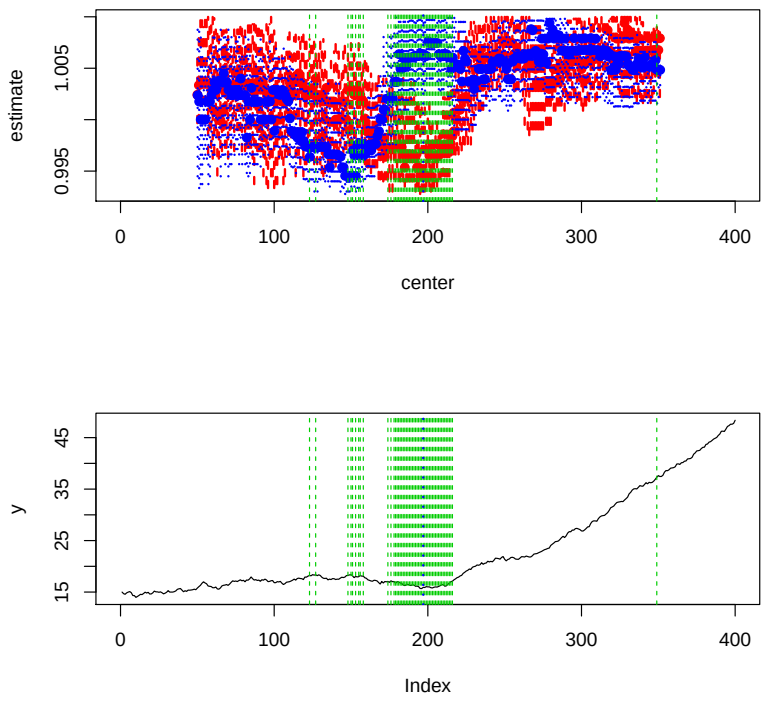


Fig. 35: Example series with identified change point, Normal case.

Now, the confidence sets and estimates are clearly visible in the upper subfigure. In particular, one can clearly observe the phases where the estimates and confidence sets do not overlap, leading to the identified potential change points. In the example, several fragments are identified as possible change points. Since the largest segment consists of many succeeding candidates, this series of succeeding candidate points results in a singular change point. This change point at observation 197 *is close to the true change point at observation* 200*.*
To compare the general performance of the method, we also evaluate the results from 1000 *simulated series. Here, we examine, if the true change point is detected and, if the amount of detected change points is correct. Therefore, we evaluate the locations of all detected change points from all simulations in one histogram and, in addition, the amount of detected change points per simulation in a second histogram. For the first example, the results are depicted in Figure 36. Thereby, the tuning parameters of detection method are selected as above.*

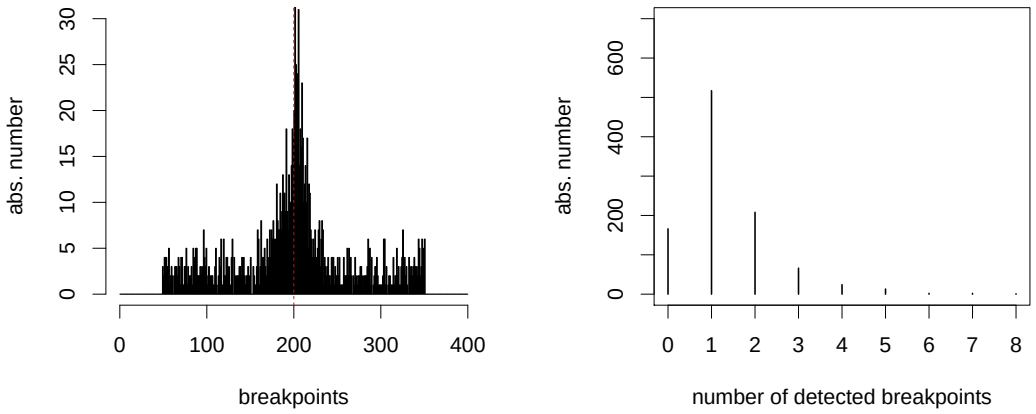


Fig. 36: Change points in 1000 series with Normal errors.

The left sub-figure shows the frequencies of the detected change points from all simulations. We see that in 180 of 1000 simulations no change point was detected. In all remaining simulations at least one change point could be discovered. The frequencies imply that the position of the change point can be estimated quite well. The detected change point locations clearly have a maximum at the true value and are symmetric with respect to this value. Nevertheless, the frequencies do not tend to 0, if points far away of the true change point are considered. A reason for this behaviour can

be found in the right sub-figure. This figure shows the frequency of the amount of detected points for each simulation. We see that in over 500 of 1000 simulations just one change point was identified. In nearly 200 simulations 2 change points were detected. In the remaining 120 observations 3 to 23 change points appeared. This shows that a part of the distribution problem in the left plot is caused by additionally identified change points. In this case, more carefully selected parameters could improve the method, since a wrong bandwidth or sensitivity m_p can lead to this false identifications.

The next example shows an application in case of atypical error distributions and multiple change points.

Example 95. *We simulate a process which starts with parameter $\kappa_1 = 1.003$ and a $\mathcal{N}(0, 0.2)$ error distribution for the first 100 observations. Then, the parameter switches to $\kappa_1 = 1.001$ and the error distribution to a contaminated $\mathcal{N}(0, 0.2)$ error distribution with jumps at $\mathcal{P}ois(1/100)$ distributed points in time and a $\mathcal{N}(5, 1)$ distributed jump height for 200 observations. The parameter changes again at observation 300 to $\kappa_1 = 1.005$ and the error distribution changes to a $\mathcal{F}(10, 1.928, -2)$ distribution. An example series is given in Figure 37.*

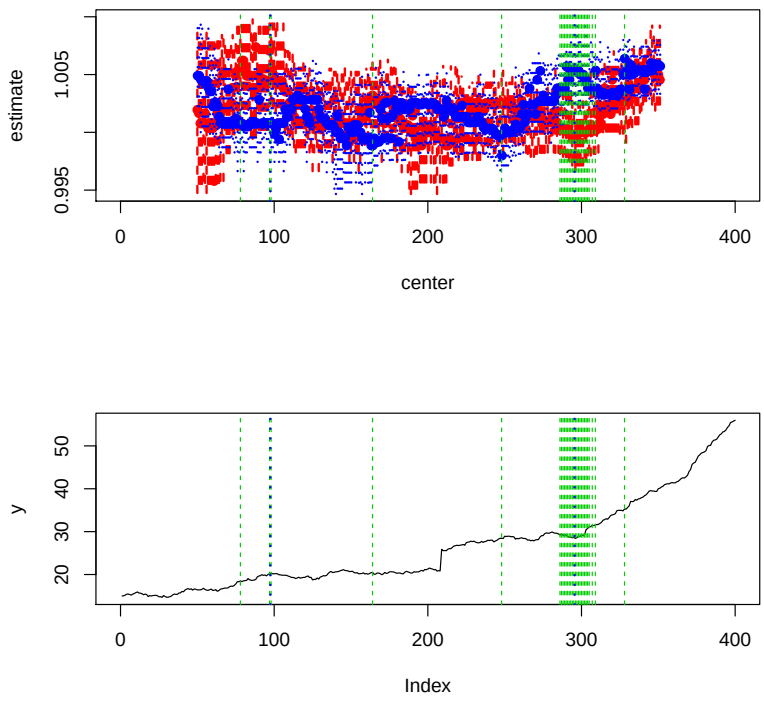


Fig. 37: Example series with identified change point, different errors.

The identified change points, with a bandwidth of $b = 49$ observations, a level of $\alpha = 0.01$ and a sensitivity parameter $m_p = 0.1$ for the threshold defining minimal clusters, are at observations 97.5 and 290. These values are close to the true change points. In the example, we can observe that the second cluster is remarkably larger then the first. This shows that the selection of m_p is crucial for a good performance of the change point detection.
In Figure 38, results form 1000 simulations with the same parameters, but with a shifted first change point to observation 200 , are shown.

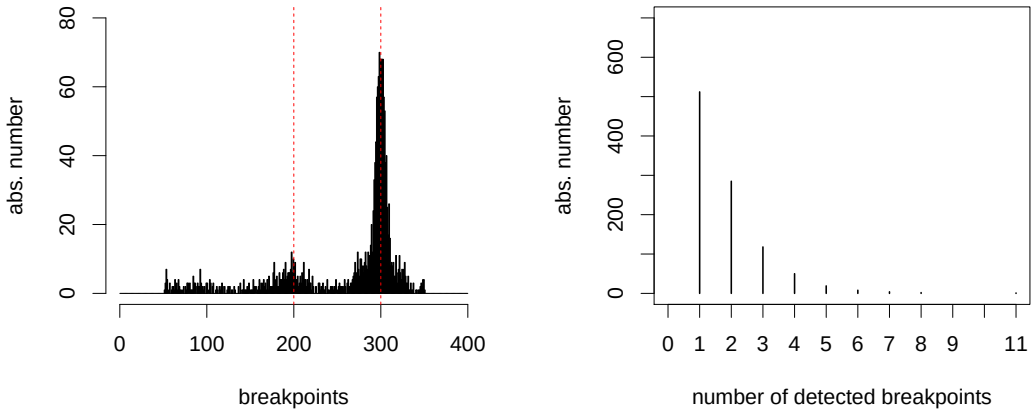


Fig. 38: Change points in 1000 series with varying errors.

In the histogram we see that the change point at 300 can be detected very well by our method. The change point at 200 can also be detected, but in a smaller proportion of all simulations. This is also reflected by the right sub-figure which shows that in 500 simulations just one change point is detected while two change points are indicated in 300 experiments. In the remaining 200 simulations, the method shows three to 11 change points with decreasing frequency.

Since the parameter m_p has to be user specified, it limits the flexibility of our algorithm. It seems very likely that an application of local depth, see Agostinelli and Romanazzi (2011) and Paindaveine and Van Bever (2013), is promising to compute estimates of the expected number of phase changes which then can be used to estimate m_p automatically from the data. This issue is future work.

4.8 Prediction as Phase-Wise Model

Since in fatigue experiments the point in time of failure is of particular interest, we propose a method which allows prediction for the class of growth processes discussed here. The general idea is to apply the depth based estimators and confidence intervals and to include the empirical depth function to consider properties of the parameter distribution. The errors then are incuded by a non-parametric bootstrapping.

General Idea of the Prediction Method

The prediction proposed here follows a simple heuristic. We do not have proofs considering the bootstrap validity so far. However, simulation studies imply that our construction results in approximately valid prediction intervals which can compete with pure bootstrapping approaches. The main modification of a pure bootstrapping thereby is that we correct the estimation of the model parameters by simplicial depth and, in addition, also correct the simulative distribution of the parameters by usage of the empirical depth.
We just define the prediction in the one parameter model. The growth parameter θ based on a dataset $y = (y_0, ..., y_N)$ can be estimated by maximising depth in θ_1. Depth defines a function of the parameter which is monotonically increasing from zero to its maximum achieved in the set of maximising parameter values and then is monotonically decreasing to zero again. A large deviation of θ to the true parameter leads to a value of the simplicial depth of zero, what coincides with a non-fit as defined by Rousseeuw and Hubert (1999). By consideration of an interval

$$[\theta_{min}, \theta_{max}]$$

as set of all parameter values for which depth is larger than zero, we get a set of parameters which are likely to have produced the dataset, as far as the supposed model is valid. This interval is obviously larger than a confidence interval for θ but bounded due to the non-fit definition. Furthermore, we know that a parameter gives a good fit, if d_S^{AR} is maximal. With this intuition, we can suppose that a parameter with a large value for $d_S^{AR}(\theta, y)$ is more probable to have produced the observed data, than a parameter with a low value for $d_S^{AR}(\theta, y)$. Hence, we can interpret $d_S^{AR}(\theta, y)$ as an function describing the probability of a good fit of θ given y.
For reliable prediction intervals, we need a density for the parameter. Since the

distribution of $\underset{\theta\in\mathbb{R}}{max}\ d_S^{AR}(\theta, y)$ is not available for arbitrary distributed data Y, we use the depth function as an alternative to model a parameter distribution.
We begin with the definition of a set of possible parameters by a depth based interval defined by

$$\theta_{min}(y) = \min\{\theta \in \mathbb{R} | d_S^{AR}(\theta, y) > 0\},$$
$$\theta_{max}(y) = \max\{\theta \in \mathbb{R} | d_S^{AR}(\theta, y) > 0\}.$$

Based on the data, we assume that parameters in $[\theta_{min}, \theta_{max}]^C$ posses a probability of zero. To allow parameter bootstrapping for prediction, we need to derive a proper parameter distribution. Therefore, we transform

$$T(\theta|y) = N\left(d_S^{AR}(\theta, y) - \frac{1}{2}\right)$$

to a density. We need to assure $f(\theta) \geq 0$ and $\int_\Theta f(\theta)d\theta = 1$. Hence, we define

$$\tilde{T}(\theta|y) = T(\theta|y) - \min_{\theta^*\in\mathbb{R}} T(\theta^*|y)$$

and

$$D(\theta|y) = \frac{\tilde{T}(\theta|y)}{\int_{\mathbb{R}} \tilde{T}(\theta^*|y)d\theta^*}.$$

Since the test statistic $\tilde{T}$ is piecewise constant for a finite observation vector y, we set

$$\int_{\mathbb{R}} \tilde{T}(\theta|y)d\theta = \sum_{\theta\in\tilde{\Theta}} \tilde{T}\left(\frac{\theta_{i+1} + \theta_i}{2}\Big|y\right)(\theta_{i+1} - \theta_i)$$

for a grid $\tilde{\Theta}$ covering $[\theta_{\min}, \theta_{\max}]$. This grid can be defined by the jump points of the empirical depth set calculated in (42). Finally, we calculate an empirical distribution by

$$F(\theta|y) = \sum_{\theta_i\in\tilde{\Theta}, \theta_i\leq\theta} D(\theta_i, y).$$

Implementation of the Predicion Method

Based on this empirical distribution, we can simulate the parameters. To construct prediction intervals, we proceed as follows.

Algorithm 96.

The aim is to generate R simulated continuations of an observed proces $y = (y_0, ..., y_N)$ up to index $N_f \in \mathbb{N}, N_f > N$.

1. *Calculate $F(\theta|y)$ based on $y = (y_0, ..., y_N)$.*
2. *Draw random parameters $\tilde{\theta}_r, r \in \{1, ..., R\}$ by application of the inversion method on F.*
3. *For each $\tilde{\theta}_r$ calculate the residuals $e_{1,r}(\tilde{\theta}_r, y), ..., e_{N,r}(\tilde{\theta}_r, y)$.*
4. *Continue the process $(y_n)_{n\in\mathbb{N}}$ by $y_{n+1,r} = \tilde{\theta}_r y_{n,r} + \tilde{e}_{n,r}$ for $n \in \{N, ..., N_f - 1\}$, whereby $\tilde{e}_{n,r}$ is a random draw from $(e_{1,r}(\tilde{\theta}_r, y), ..., e_{N,r}(\tilde{\theta}_r, y))$ with replacement.*

Remark 97. *For the construction from Algorithm 96, we so far cannot prove the validity of the prediction level theoretically. The main problem, compared to a pure bootstrapping approach is that we plug-in the empirical distribution $F(\theta|y)$. This on the one hand has an advantage for skewed error distributions, since it robustifies the estimate and simultaneously corrects the location of our empirical distribution. On the other hand it complicates the calculation of the level for the resulting prediction interval, since the residuals depend on the draws of the parameter and hence on F. A theoretical validation is future work and will not be addressed in this thesis. However, we present a simulation study in Section 5 which indicates a validity empirically.*

Now, the predicted value of the process at index N_f can be given. The construction is based on the empirical quantiles of the bootstrap continuations of the process following Efron (1979).

Definition 98.

(a) The predicted distribution of the process values at observation $N_f > N$ based on Algorithm 96 is given by the empirical distribution of

$$y^{N_f} = (y_{N_f,1}, ..., y_{N_f,r}).$$

A mean predicted process value for y_{N_f} then is

$$\hat{y}^{N_f} = \frac{1}{R} \sum_{r=1}^{R} \tilde{y}_{N_f,i}.$$

(b) The empirical distribution of the predicted time of arrival at a critical process value $C > max\{y_0, ..., y_N\}$ is given by the distribution of

$$n^C = (n_{C,1}, ..., n_{C,R}),$$

whereby

$$n_{C,r} = \inf\{n > N | y_{n,r} \geq C\}, r \in \{1, ..., R\}.$$

The mean predicted time of arrival at C then is

$$\hat{n}^C = \frac{1}{R} \sum_{r=1}^{R} n_{C,i}.$$

(c) The empirical prediction interval of the process value at observation N_f is given by the empirical quantiles of y^{N_f}. The empirical prediction interval of the time of arrival at C is given by the empirical quantiles of n^C.

Since this method defines very large confidence intervals, we propose two alternative approaches to construct F.

Construction 99.
The standard method for prediction is given by F setting

$$\theta_{\min} = \min\left\{\theta \in \mathbb{R} | F(\theta|y) > 0\right\} = \min\{\theta \in \mathbb{R} | d_S^{AR}(\theta, y) > 0\right\}$$

and

$$\theta_{\max} = \max\{\theta \in \mathbb{R} | F(\theta|y) > 0\} = \max\{\theta \in \mathbb{R} | d_S^{AR}(\theta, y) > 0\},$$

whereby these sets coincide by the construction of F as proposed in Algorithm 96. We also propose a variation by replacement of θ_{min} by

$$\min\left\{\theta \in \mathbb{R} \middle| N\left(d_S^{AR}(\theta, y) - \frac{1}{2}\right) > \frac{1}{2} - \frac{1}{2} q_{\chi_1^2}(1 - \alpha)\right\}$$

and θ_{max} by

$$\max\left\{\theta \in \mathbb{R} \middle| N\left(d_S^{AR}(\theta, y) - \frac{1}{2}\right) > \frac{1}{2} - \frac{1}{2} q_{\chi_1^2}(1 - \alpha)\right\},$$

where $q_{\chi_1^2}(\gamma)$ is the γ quantile of the χ^2 distribution with one degree of freedom.

Then $F(\theta|y)$ *is constructed, restricted on* $[\theta_{\min}, \theta_{\max}]$ *by setting* $d_S^{AR}(\theta|y) = 0$ *on* $[\theta_{\min}, \theta_{\max}]^C$.

The prediction intervals are much sharper by application of the restricted distribution function. However, the simulations in Section 5.4 show that they still hold the level.

Examples of the Proposed Predicion Method

Example 100. *To illustrate the prediction method, we consider a process defined by*

$$Y_n = \theta_1 \cdot Y_{n-1} + Y_{n-1} + E_n,$$

whereby $\theta_1 = 0.004, E_n \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.1)$ *and* $y_0 = 15$. *We observe a realisation with length* $N = 200$. *We consider the process* $y = (y_0, ..., y_{100})$ *as observed part and want to predict the process length at observation 200 and the time of arrival at a length of 30. The application of the prediction method with* $\alpha = 0.05$ *and* 1000 *simulated continuations is presented in Figure 39.*

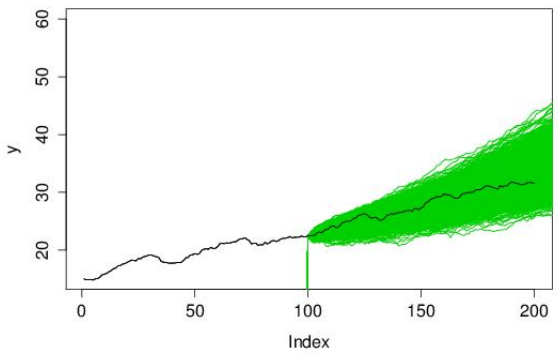


Fig. 39: Example of the prediction method. The black line shows the observed process and the green lines are continuations based on the bootstrap algorithm.

The value of the process at observation 200 is $y_{200} = 31.59334$. *The mean estimate of the process at observation* 200 *is* 33.65158 *and the median* 33.56702. *These values are close to the true value. Moreover the* 95% *prediction interval for the length at observation* 200 *is* $[27.72690, 39.82446]$.
For the time of arrival at a value of 30, *we get an mean estimate of* 175.098 *and a median estimate of* 172.5. *The prediction interval is given by* $[146, 223]$. *The true time of the first arrival at a process value of 30 is* 173.

Limitations of the Prediction Algorithm

In case of the two parameter models, an analogue algorithm seems promising. The central problem here is that we first have to transform the two dimensional empirical depth to a density and then want to generate random draws of two parameters. For this purpose it is not sufficient to draw from a uniform distribution, since the empirical distribution function itself then is two dimensional and the solution of $F(\theta) = u$ is not uniquely determined but consists of a set of parameters $\theta = (\theta_1, \theta_2)$ or $\theta = (\theta_1, \theta_3)$. This case is future work at this stage of our research and therefore is neither included in the theses nor in the supplementary `R` package.

4.9 Used R Packages

The main functions are written in the programming language `R`, see RCoreTeam (2015). In the implementation of our methods, we use of some `R` packages on matrix calculations, Alpha shapes and parallel computation, available in the `R` repository CRAN.

In the `rexpar` package, implementing the proposed methods for depth based analysis of growth processes, we use the following external packages.

The `matrixcalc` package was used to generate the matrices for the calculations of the depth statistics, see Novomestky (2012). To use parallel computation methods, we apply the parallel functions from the `parallel` package. Further, the calculation of the Alpha shapes is based on functions from the `alphahull` package, see Pateiro-Lopez and Rodriguez-Casal (2015). To simulate the limit distribution in the two parameter case, we further use functions from the `MASS` package, see Venables and Ripley (2002).

In the simulation studies, we use some additional packages which implement the known estimators. These packages were the `MASS`, `robustbase`, and `nlstools` packages, see Venables and Ripley (2002); Rousseeuw et al. (2015); Baty et al. (2015).

Further, the methods in the comparison of the change point detection are implemented in the `strucchange` and the `segmented` packages, see Zeileis et al. (2002); Muggeo (2008).

For the prediction comparisons, we use the `BootPR` package, see Jae (2014).

Since most of the calculations were performed at high performance clusters (HPCs), we use functions from the `snow` and `Rmpi` packages to allow parallel computation on these clusters, see Yu (2002); Tierney et al. (2013). Further, some parallel computations use the `BatchJobs` package, see Bischl et al. (2015) which allows an efficient parallel computation of identical tasks on HPCs.

All extensive calculations were performed on the cluster of the Faculty of Statistics and on the Linux High Performance Cluster (LiDOng) of the TU Dortmund University.

5 Simulation Studies

In this section, we present simulation results to analyse the performance of the proposed estimators, tests and further methods in comparison to existing approaches. Some of the results can be found in Kustosz and Müller (2014); Kustosz et al. (2016a,b). However, all comparisons were extended to a consistent set of methods and all additional studies illustrating the effect of the sample size and violations of the model are completely new in this thesis. Thereby, this thesis does not attempt to show that the new methods are a gold standard for growth models. The aim is to present the applicability of simplicial depth compared to standard methods, for example as used in engineering, to illustrate the robustness properties under restricted information on the error distribution.

5.1 Estimators

We present simulation studies to compare the performance of the proposed estimators with existing methods.

Linear Model Without Intercept

In the one parameter case, we compare the proposed estimators with the standard Ordinary Least Squares (OLS) estimator defined by

$$\underset{\theta_1 \in \mathbb{R}}{argmin} \sum_{n=1}^{N} (y_n - \theta_1 y_{n-1} - y_{n-1})^2,$$

which was proposed by Mann and Wald (1943) for autoregressive processes. Further, we take robust estimators into account. A natural choice in many applications is a M-estimator. This robust alternative for the least squares approach was introduced by Huber (1973). For autoregressive processes of order one without intercept this estimator is defined by

$$\underset{\theta_1 \in \mathbb{R}}{argmin} \sum_{n=1}^{N} \rho\left(\frac{y_n - \theta_1 y_{n-1} - y_{n-1}}{\hat{s}}\right),$$

whereby $\hat{s}$ is an estimate of the residuals standard deviation and ρ a robustifying function. In case of innovation outliers, which we consider in this thesis, Denby and Martin (1979) show that the M-estimator is very efficient under symmetric errors. The restriction to this kind of outliers follows from the fact, that we aim to model crack growth experiments, where the jumps result in innovation outliers only. In our application, we define ρ by the well known Huber function with truncation parameter $k = 1.345$.
Another quite simple robust estimator can be defined by the least absolute deviations of the model. This also is a specific M-estimator with $\rho(x) = |x|$. It is defined by

$$\underset{\theta_1 \in \mathbb{R}}{argmin} \sum_{n=1}^{N} |y_n - \theta_1 y_{n-1} - y_{n-1}|.$$

Here, all deviations of the residuals are weighted linearly. Hence the effect of outliers is reduced, if a $med(E_n) = 0$ assumption holds.
For more details on robust estimation in time series models, we refer to Maronna et al. (2006).
In all simulations, we generate 1000 AR(1) processes with the given parameters and apply seven different estimators to each simulated series. The first simulations show the performance of the estimators for explosive AR(1) processes with normally distributed errors and small sample size. In Figure 40, we see the results for seven estimators based on $E_n \sim \mathcal{N}(0, 0.1), \theta_1 = 0.01$ and $N = 10$ observations per process.

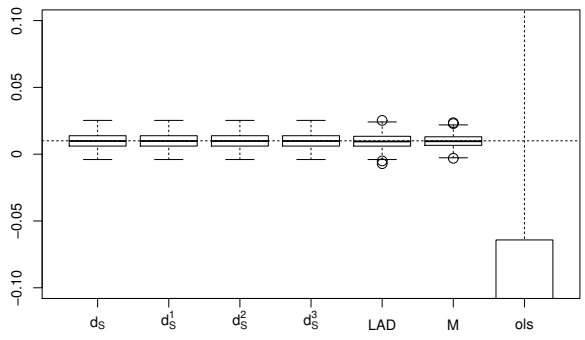


Fig. 40: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\theta_1 = 0.01$ based on $\mathcal{N}(0, 0.1)$ errors.

We see that all robust estimators perform very well. The OLS estimator shows a clear bias and hence systematically deviates from the true parameter value. The

bias of the OLS estimator also was reported by Kaufmann and Kruse (2013).
For larger sample sizes the bias of the OLS estimator decreases. In Figure 41, we see results for series with length $N = 250$.

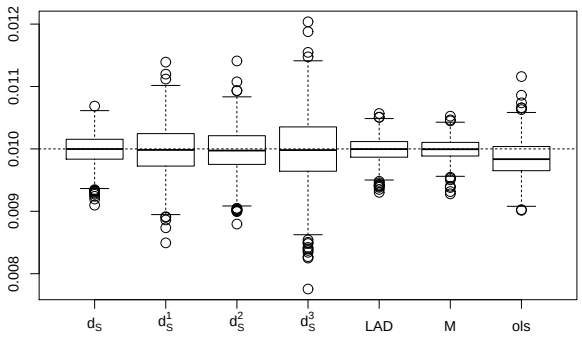


Fig. 41: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\theta_1 = 0.01$ based on $\mathcal{N}(0, 0.1)$ errors.

The OLS estimator still shows a small bias, while the robust alternatives perform very well. Now differences in the variance of the robust estimators can be observed. The M-estimator appears to be best in the set of our considered candidates followed by the LAD estimator. The full simplicial depth estimator shows a slightly worse performance. Not surprisingly, the simplified depth estimators are outperformed by the full simplicial depth version and hence by the other robust alternatives. In the set of the simplified depth estimators the d_S^1 and d_S^2 versions are comparably good while the d_S^3 estimator shows a worse performance. This can also be verified by the mean squared errors (MSE) of the estimators for our simulations presented in Table 2.

N	d_S	d_S^1	d_S^2	d_S^3	LAD	M	OLS
10	2784.85	2784.85	2784.85	2784.85	2677.30	1937.65	$8.76 \cdot 10^6$
250	6.30	14.85	12.08	30.96	4.02	2.79	10.41

Tab. 2: MSE $\cdot 10^8$ for linear AR processes without intercept with parameter $\theta_1 = 0.01$ based on $\mathcal{N}(0, 0.1)$ errors.

In Appendix A, we show the results for stationary processes and the unit root case. Since our primary aim was to propose estimators which can be applied, if the errors are not necessary normally distributed, the second set of simulations shows results from situations where the errors possess a normal distribution but are contaminated

by random outliers to certain amount. Therefore, we define the errors by

$$E_n = X_n + P_n \cdot J_n - c,$$

whereby $X_n \sim \mathcal{N}(0, 0.1), P_n \sim \mathcal{P}ois(\frac{5}{100}), J_n \sim \mathcal{N}(5, 1)$ and $c = 0.007127$ is a constant to correct the median of E_n. We denote this distribution by $\mathcal{CN}(0, 0.1)$. Further X_n, P_n and Y_n are assumed to be independent. Due to c we have $med(E_n) = 0$. E_n now has a skewed distribution, since we introduce jumps with a high probability for positive jump heights. By the parameter of the Poisson distribution, we further know that the contamination rate is approximately 5% in our simulations.
In Figure 42, the results for small samples are depicted.

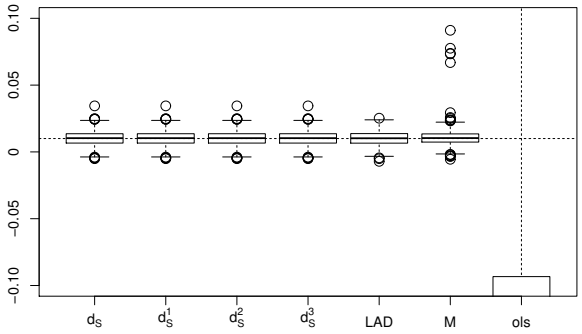


Fig. 42: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\theta_1 = 0.01$ based on $\mathcal{CN}(0, 0.1)$ errors.

The small sample results are similar to the non-contaminated case. All simplicial depth estimators show the same performance. Further, the LAD estimator slightly outperforms the simplicial depth estimators, while the M-estimator performs slightly worse due to some highly outlying estimates. The OLS estimator completely fails, since it is affected by the skewed error distribution. These results also are in line with the MSE presented in Table 3. In Figure 43, the results for larger samples are depicted.

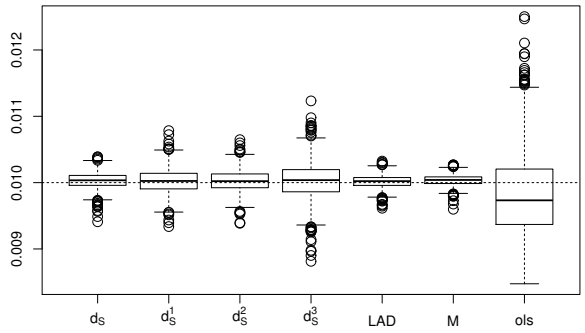


Fig. 43: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\theta_1 = 0.01$ based on $\mathcal{CN}(0, 0.1)$ errors.

Here, the LAD shows the best results. The M-estimator and the full simplicial depth estimator also preform quite well. The simplified depth notions are outperformed by the other robust estimators, but show reliable results. The OLS estimator clearly improves in the larger sample but is still more biased than the presented alternatives.

N	d_S	d_S^1	d_S^2	d_S^3	LAD	M	OLS
10	2808.80	2808.80	2808.80	2808.80	2635.43	6417.32	$13 \cdot 10^6$
250	1.59	3.56	2.71	8.06	0.96	0.78	42.61

Tab. 3: MSE $\cdot 10^8$ for linear AR processes without intercept with parameter $\theta_1 = 1.01$ based on $\mathcal{CN}(0, 0.1)$ errors.

In the last set of examples, we want to show how the estimators behave, if we drop the normal distribution completely. Hence, we define the errors by the Fréchet distribution with density

$$f_{\alpha,\beta,\gamma}(x) = \frac{\gamma}{\alpha}\left(\frac{x-\beta}{\alpha}\right)^{-1-\gamma} exp\left(-\left(\frac{x-\beta}{\alpha}\right)^{-\gamma}\right)$$

and parameters $\alpha = 1.928, \beta = -2, \gamma = 10$. So, again $med(E_n) = 0$ holds.
In Figure 44 the small sample results are depicted.

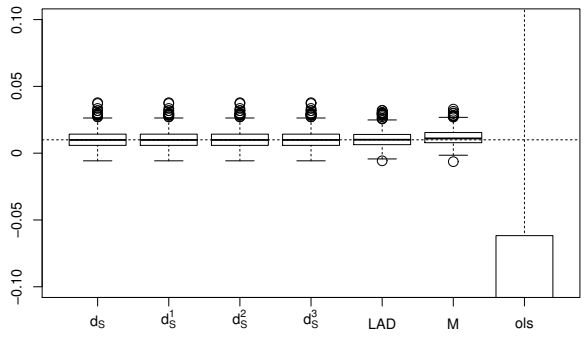


Fig. 44: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\theta_1 = 0.01$ based on $\mathcal{F}(1.928, -2, 10)$ errors.

We again can observe the superior performance of the robust estimators. Again, in the small sample no real differences are visible. The LAD is slightly better than the remaining robust alternatives and the M-estimator appears to show a small deviation in the median estimate compared to the other estimators. In Figure 45, the results for $N = 250$ observations are shown.

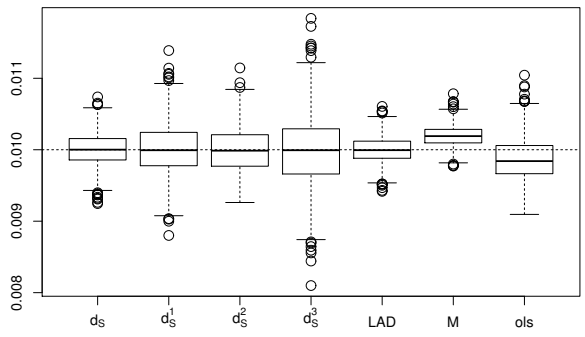


Fig. 45: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\theta_1 = 0.01$ based on $\mathcal{F}(1.928, -2, 10)$ errors.

Here, we see that the LAD estimator again performs best. The full simplicial depth estimator also performs quite well. The M-estimator has a clear bias due to the symmetric truncation based on the Huber function. The simplified depth notions define reliable, but weaker estimators than the full simplicial depth. Again, the OLS improves but still is clearly more inefficient than the alternative methods.

N	d_S	d_S^1	d_S^2	d_S^3	LAD	M	OLS
10	3788.73	3788.73	3788.73	3788.73	3361.19	3500.49	$6.99 \cdot 10^6$
250	5.12	12.07	9.71	24.04	3.22	5.94	10.64

Tab. 4: MSE $\cdot 10^8$ for linear AR processes without intercept with parameter $\theta_1 = 1.01$ based on $\mathcal{F}(1.928, -2, 10)$ errors.

The mean squared errors in Table 4 support the interpretation from the boxplots. As already remarked, Kaufmann and Kruse (2013) analysed the bias of the OLS estimator in AR models when the parameter is close to the unit root. This bias is also visible in the median absolute bias of the proposed estimators for the parameter $\kappa = \theta_1 + 1$ in Figure 46. Thereby the results are based on 100000 simulated estimates for autoregressive processes without intercept and parameters ranging from 0.9 to 1.1. The unit root case is $\kappa = 1$.

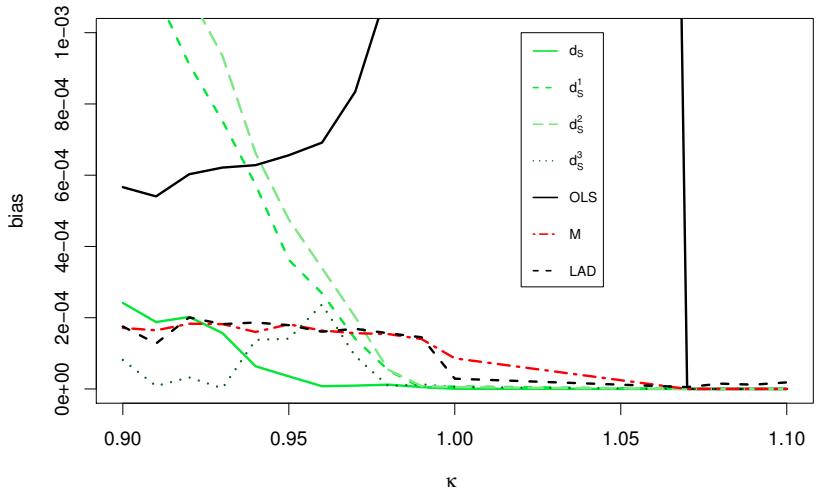


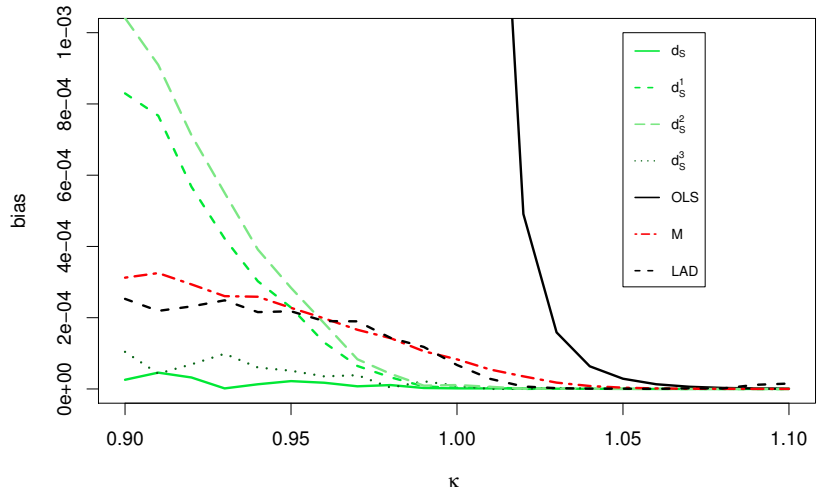
Fig. 46: Median bias of different estimators close to the unit root for linear autoregressive processes without intercept, $N = 100$ and $\mathcal{N}(0, 0.2)$ errors.

We see that the bias of the considered estimators decreases to zero for parameters tending to infinity. The bias of the OLS estimator has a maximum at the unit root and it quickly decreases to zero for values of the parameter which are larger than one. For parameters lower than one the decrease is slower. The M-estimator and the LAD estimator behave similar to each other. Close to the unit root their bias increases and remains quite stable for parameters lower than one. The full simplicial depth estimator shows a better performance. In particular, for values below one

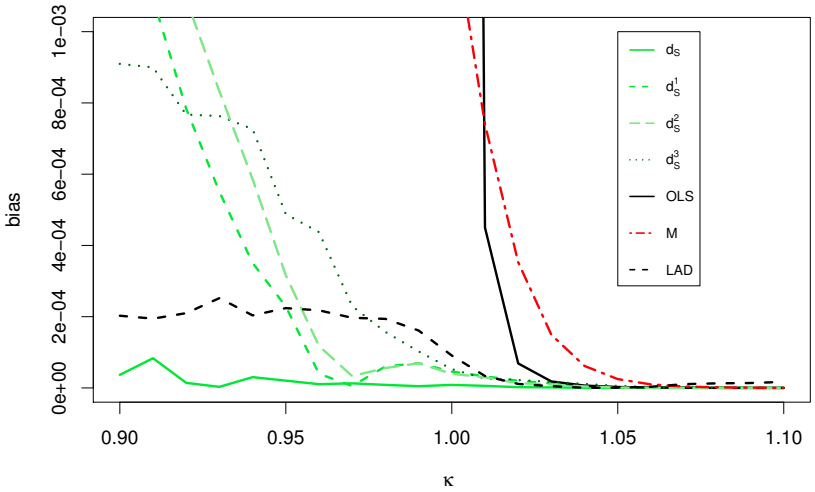
it outperforms all other considered estimators in terms of the median bias. This changes for parameters below 0.9. Then the M and LAD estimators perform better. The simplified depth estimators are also quite interesting. The d_S^1 and d_S^2 estimators have a quickly increasing bias for parameters below one, while the d_S^3 estimator shows a good performance for $\theta \approx 0.9$ as well. As can be seen in the previous simulation studies, this comes with a high variance.
Similar results can be seen for contaminated errors, as presented in Figure 47. Thereby, we use a contaminated Normal distribution and a Gumbel error distribution $\mathcal{G}(\alpha, \beta)$ defined by the continuous density function

$$f(x) = \frac{1}{\alpha} exp\left(-\left(\frac{x-\beta}{\alpha}\right) + exp\left(-\left(\frac{x-\beta}{\alpha}\right)\right)\right).$$

In our example, we apply $\alpha = 10$ and $\beta = -3.665129$. Thereby, the error distribution has an approximate median of zero, since the median of a Gumbel distribution is given by $\beta - \alpha ln(ln(2))$.



(a) $\mathcal{CN}(0, 0.2)$ errors



(b) $\mathcal{G}(10, -3.665129)$ errors

Fig. 47: Median bias of different estimators close to the unit root for linear autoregressive processes without intercept, $N = 100$.

We, in addition, can see that the M-estimator is biased when we consider a heavily asymmetric error distribution. On the other hand bias of the full simplicial depth reduces then. Further, the d_S^3 bias for $\kappa < 1$ is higher for asymmetric errors.
Summarising, we see that for the linear AR(1) model without intercept the OLS estimator shows remarkable problems close to the unit root. Hence, it should not be applied and robust estimators can be considered to reduce this bias. Thereby, not only our proposals are promising. Especially the LAD estimator shows a very good

performance. The major shortcoming for our purpose is that the limit distribution of the test statistic in this case depends on the error distribution (see e.g. Knight, 1998). Since we only want to assume that we have independent errors with $med(E_n) = 0$ we cannot derive tests for this estimator directly.

Linear Model With Intercept

Now, we want to compare our estimators for the linear AR(1) model with intercept. Again, the alternative methods are defined by the OLS estimator, the M-estimator and the LAD estimator. Therefore, we just have to adjust the residuals in the definitions for the extended model. The OLS estimator is now defined by

$$\underset{\theta=(\theta_1,\theta_3)\in\mathbb{R}^2}{argmin} \sum_{n=1}^{N} (y_n - \theta_1 y_{n-1} - y_{n-1} - \theta_3)^2.$$

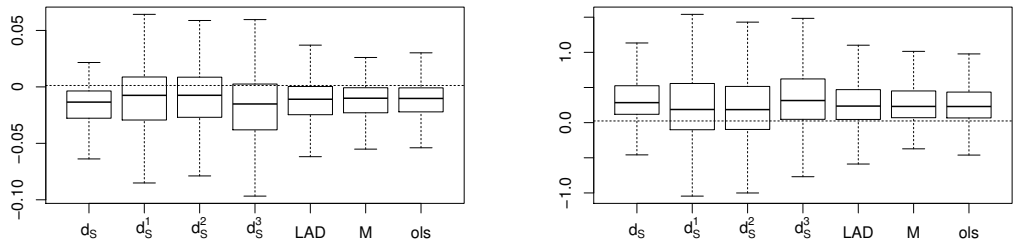
For the M-estimator we now use

$$\underset{\theta=(\theta_1,\theta_3)\in\mathbb{R}^2}{argmin} \sum_{n=1}^{N} \rho\left(\frac{y_n - \theta_1 y_{n-1} - y_{n-1} - \theta_3}{\hat{s}}\right).$$

The least absolute deviations estimator is defined by

$$\underset{\theta=(\theta_1,\theta_3)\in\mathbb{R}^2}{argmin} \sum_{n=1}^{N} |y_n - \theta_1 y_{n-1} - y_{n-1} - \theta_3|.$$

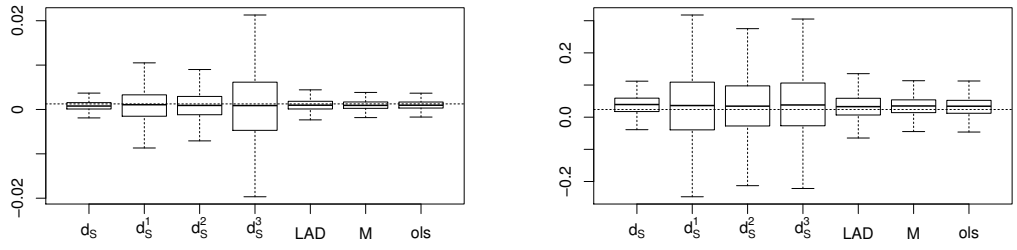
Since we observe two parameters now, the comparisons are based on higher sample sizes. We compare the estimators for $N \in \{100, 500\}$. Further, we again consider different error distributions. In Figure 48, the results of 1000 repeated estimations for a process with $\theta_1 = 0.00125368$, $\theta_3 = 0.02392$, $y_0 = 15$ and $\mathcal{N}(0, 0.2)$ errors are depicted. The parameters were randomly chosen and define a growth process.



(a) θ_1 (b) θ_3

Fig. 48: Estimator performance for linear AR processes with intercept and $N = 100$ observations based on $(\theta_1, \theta_3) = (0.00125368, 0.02392)$ with $\mathcal{N}(0, 0.2)$ errors.

We see that the OLS and M-estimators perform best. The LAD estimator shows a slightly worse performance followed by the simplified depth estimators which deliver reliable results but have a higher variance than the other proposals. The full simplicial depth estimator preforms as well as the LAD estimator. Due to the small sample size all estimators show deviations from the simulation parameters. If we increase the sample size to $N = 500$, the results are similar.

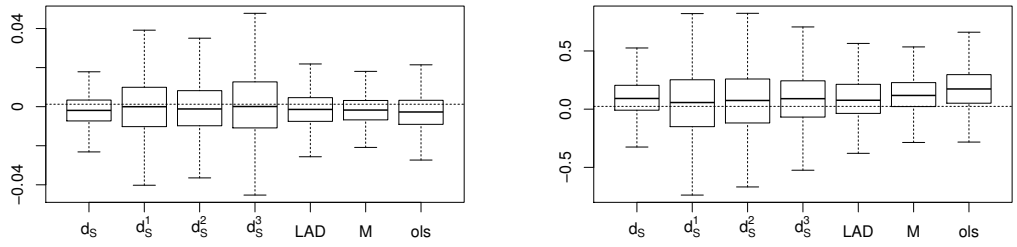


(a) θ_1 (b) θ_3

Fig. 49: Estimator performance for linear AR processes with intercept and $N = 500$ observations based on $(\theta_1, \theta_3) = (0.00125368, 0.02392)$ with $\mathcal{N}(0, 0.2)$ errors.

In Figure 49, we see that the OLS and M-estimators are performing very well and slightly better than the LAD estimator. All proposals do no show any remarkable bias. The depth based estimators show clearly higher variances than the remaining proposals. The full simplicial depth estimator thereby is comparable to the LAD estimator and is outperformed by the OLS and M-estimators.
If we again turn to error distributions which are skewed, the situation clearly

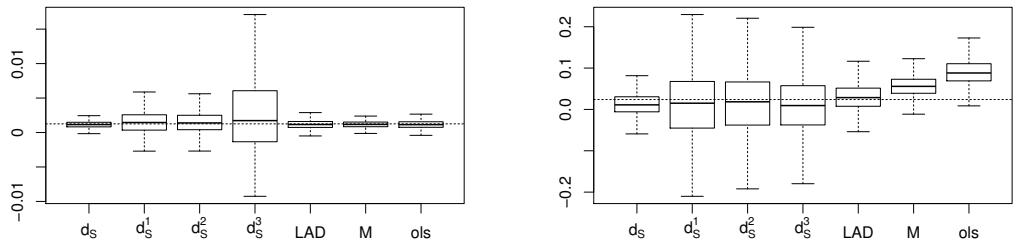
changes. Small sample results are shown in Figure 50.



(a) θ_1 (b) θ_3

Fig. 50: Estimator performance for linear AR processes with intercept and $N = 100$ observations based on $(\theta_1, \theta_3) = (0.00125368, 0.02392)$ with $\mathcal{G}(10, -3.665129)$ errors.

Now, the OLS estimator is biased in both parameters. The remaining estimators show a lower bias. The LAD and M-estimators also have larger distances to true values compared to the simplicial depth versions. The full simplicial depth estimator has a bias similar to the LAD estimator, but a slightly lower variance. In Figure 51 the results for a larger sample size are shown.

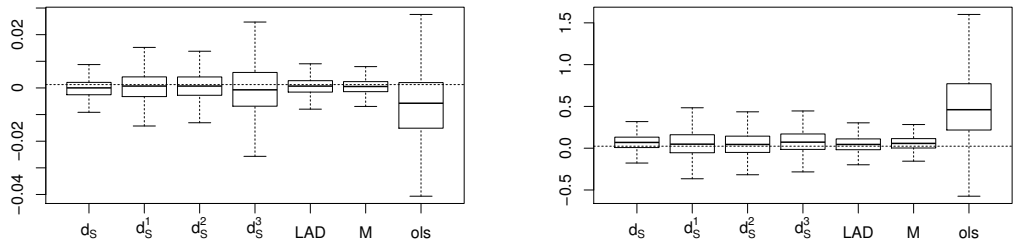


(a) θ_1 (b) θ_3

Fig. 51: Estimator performance for linear AR processes with intercept and $N = 500$ observations based on $(\theta_1, \theta_3) = (0.00125368, 0.02392)$ with $\mathcal{G}(10, -3.665129)$ errors.

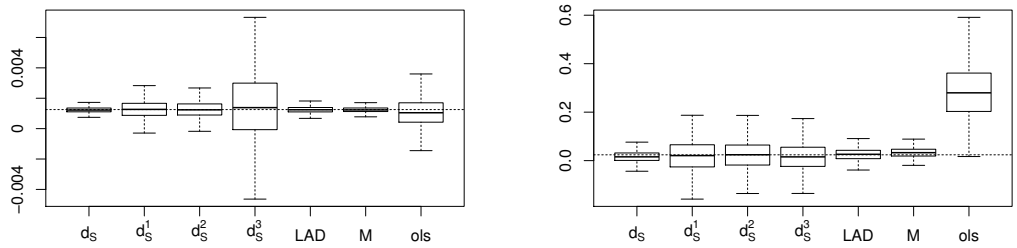
While the OLS estimator is consistent in the estimation of θ_1, as well as the M-estimator and both estimators show a remarkable low variance, they fail to estimate the intercept parameter θ_3 systematically. The reason for this problem comes from the asymmetric error distribution which leads to a systematic bias due to the $\mathbb{E}[E_n] = 0$ assumption which does not hold in this example. The LAD estimator

again performs best. The depth based estimators are also unbiased in both parameters but have higher variances than the LAD estimator. The full simplicial depth estimator shows a low variance, but a higher deviation for the estimate of the intercept than the LAD.
In Figures 52 and 53, the results for simulations with contaminated normally distributed errors are presented.



(a) θ_1 (b) θ_3

Fig. 52: Estimator performance for linear AR processes with intercept and $N = 100$ observations based on $(\theta_1, \theta_3) = (0.00125368, 0.02392)$ with $\mathcal{CN}(0, 0.2)$ errors.



(a) θ_1 (b) θ_3

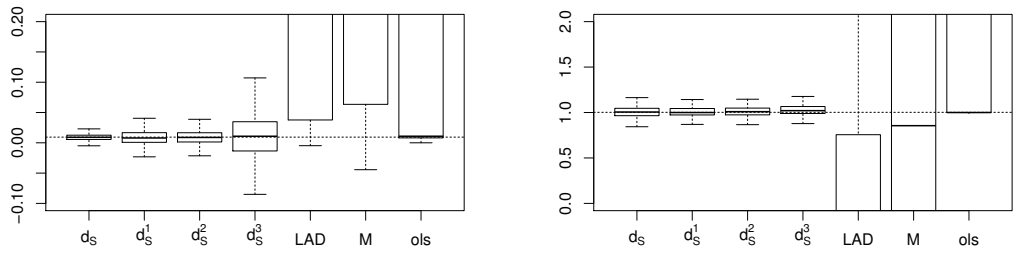
Fig. 53: Estimator performance for linear AR processes with intercept and $N = 500$ observations based on $(\theta_1, \theta_3) = (0.00125368, 0.02392)$ with $\mathcal{CN}(0, 0.2)$ errors.

Here, the OLS estimator again fails, due to the skewed error distribution. The M-estimator performs very well, as well as the LAD and d_S estimators. Again, the d_S estimator shows a slightly worse performance for the θ_1 estimate. The simplified estimators are ranked as in the previous studies.
Summarising, we see that the simplified depth based estimators allow to estimate

explosive AR processes with a lower bias than the standard methods, even if the error distributions are asymmetric. However, we cannot outperform the LAD estimator by the depth methods. Nevertheless, the depth based estimators have the advantage to allow a construction of simultaneous confidence regions and tests for the parameters of the model without knowledge of the exact error distribution, as proposed in Section 4.3.

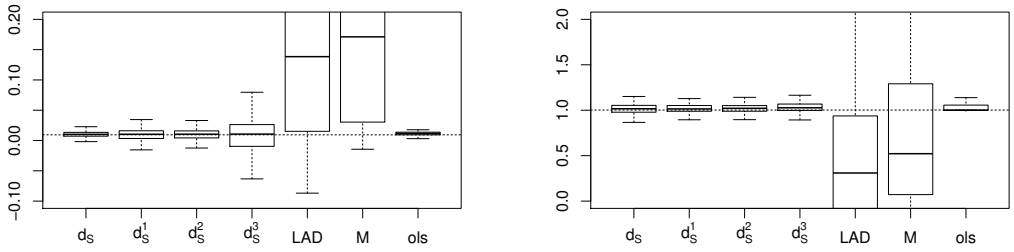
Non-Linear Model

In the non-linear case, we use the least squares approach, our depth based estimators and the least absolute deviation estimator as well as an M-estimator for non-linear models. Hence, we skip the definitions with the remark that it suffices to replace the residuals by the non-linear version. The results are quite similar to the two parameter model in the linear case.



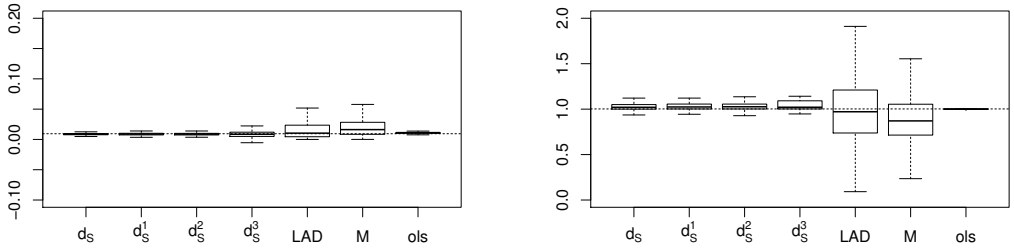
(a) θ_1 (b) θ_2

Fig. 54: Estimator performance for non-linear AR processes with $N = 100$ observations based on $(\theta_1, \theta_2) = (0.009392, 1.00225368)$ with $\mathcal{N}(0, 0.2)$ errors.



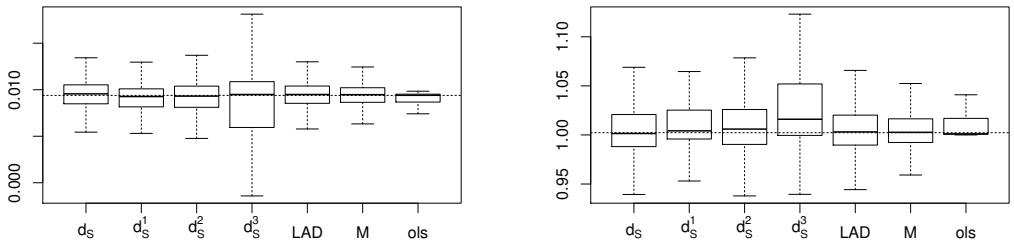
(a) θ_1 (b) θ_2

Fig. 55: Estimator performance for non-linear AR processes with $N = 100$ observations based on $(\theta_1, \theta_2) = (0.009392, 1.00225368)$ with $\mathcal{CN}(0, 0.2)$ errors.



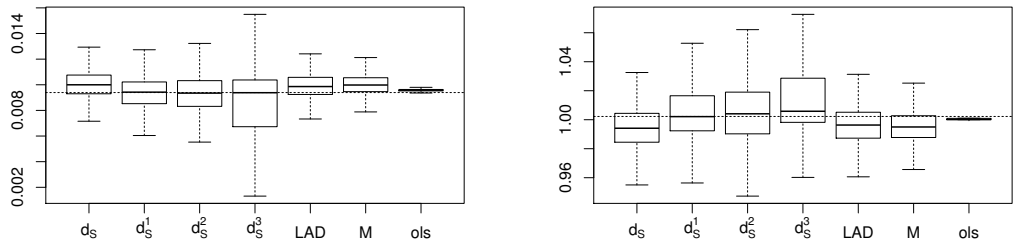
(a) θ_1 (b) θ_2

Fig. 56: Estimator performance for non-linear AR processes with $N = 100$ observations based on $(\theta_1, \theta_2) = (0.009392, 1.00225368)$ with $\mathcal{F}(1.928, -2, 10)$ errors.



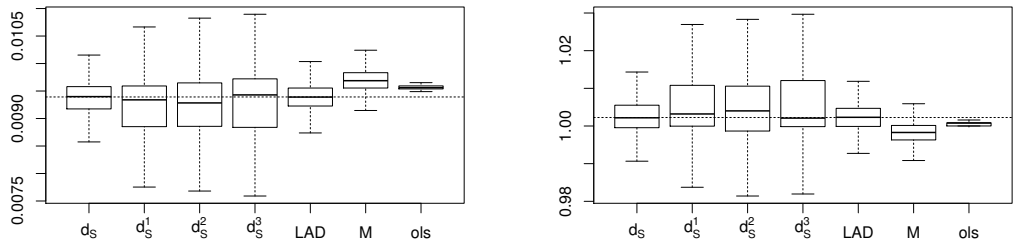
(a) θ_1 (b) θ_2

Fig. 57: Estimator performance for non-linear AR processes with $N = 500$ observations based on $(\theta_1, \theta_2) = (0.009392, 1.00225368)$ with $\mathcal{N}(0, 0.2)$ errors.



(a) θ_1 (b) θ_2

Fig. 58: Estimator performance for non-linear AR processes with $N = 500$ observations based on $(\theta_1, \theta_2) = (0.009392, 1.00225368)$ with $\mathcal{CN}(0, 0.2)$ errors.



(a) θ_1 (b) θ_2

Fig. 59: Estimator performance for non-linear AR processes with $N = 500$ observations based on $(\theta_1, \theta_2) = (0.009392, 1.00225368)$ with $\mathcal{F}(1.928, -2, 10)$ errors.

In Figures 54 to 59, we see that the depth based estimators are very reliable for all considered error distributions. Thereby, the simplified notions have higher variance than the full simplicial depth estimator and the LAD. The LAD and M-estimators have remarkable problems to estimate the parameters for low sample sizes. Further, the OLS estimator has a high variance in case of normal errors and is biased in case of asymmetric error distributions. For larger sample sizes, we clearly see the superiority of the full simplicial depth and LAD estimators. They are followed by the simplified depth estimators which are also consistent, but show a higher uncertainty. The M and OLS estimators are both heavily biased for skewed errors, but very competitive in case of $\mathcal{N}(0, 0.2)$ errors.

5.2 Tests

In this section, we compare the power of the proposed tests with standard methods.

Linear Model Without Intercept

For the linear model without intercept, we extend the comparisons presented in Kustosz and Müller (2014) by the power of the simplified depth tests. The definition of the tests based on simplicial depth was presented in Sections 3.1, 3.2 and 3.3. We compare our tests with a test based on

$$T_N(\theta) = (\hat{\theta} - \theta)\sqrt{\sum_{n=1}^{N} y_{n-1}^2},$$

with

$$\hat{\theta} = \frac{\sum_{n=1}^{N} y_n y_{n-1}}{\sum_{n=1}^{N} y_{n-1}^2} - 1$$

following Anderson (1959) who has shown that for normally distributed errors T_N has an approximate normal distribution with mean zero and known variance σ^2 under $H_0 : \theta = \theta^0$. Further, we compare the test with a simple sign test for AR processes studied by Huggins (1989). Here, the test statistic is given by

$$Q_N(\theta) = \sum_{n=1}^{N} sgn(Y_n - (1+\theta)Y_{n-1})$$

and has an exact distribution given by $2(B_N - \frac{N}{2})$, with B_N being binomial distributed with parameters N and $\frac{1}{2}$ under H_0.
In Figure 60, simulated power functions for processes with $\mathcal{N}(0, 0.2)$ errors and $N = 300$, tested for $H_0 : \theta = 0.002$ are presented. For each parameter the power function is based on 10000 simulated processes. In case of normally distributed errors the OLS test outperforms all proposed alternatives. This is not surprising, since it assumes the correct error distribution and therefore is not faced with any loss of efficiency by milder assumptions. The full simplicial depth test and the sign test are identical here. This happens, since we assume to observe growth processes, what reduces the test statistic of the full simplicial depth test to an evaluation of residuals. The full simplicial depth test can be improved by using the full derivatives of the quality function. In equation (12), we dropped the multiplication with y_{n-1}, since under H_0 the process (Y_n) is strictly positive. Hence, we defined full simplicial

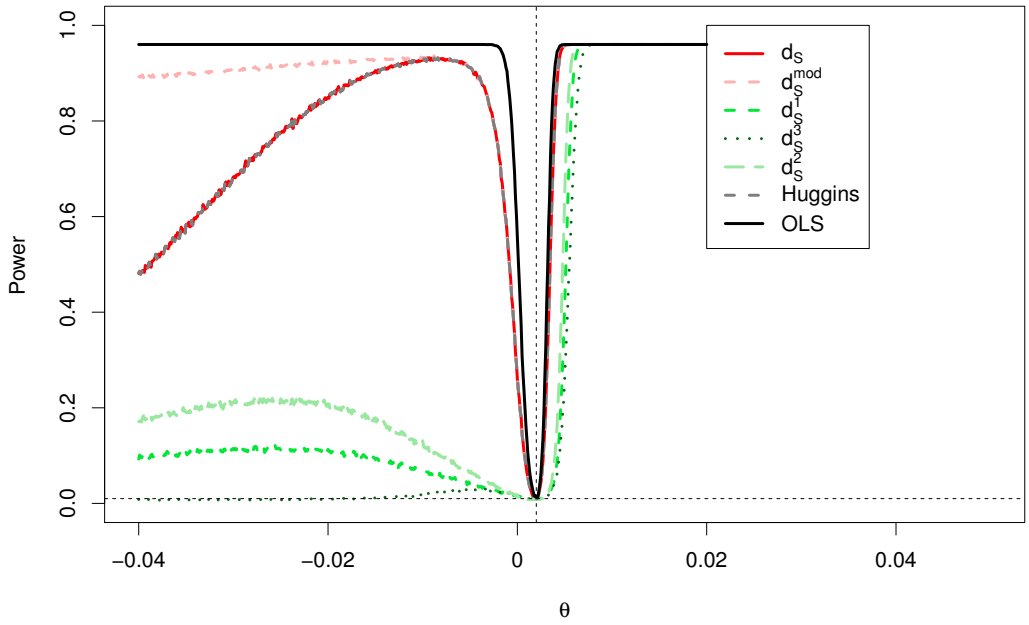


Fig. 60: Simulated power functions based on linear AR(1) processes without intercept, $\mathcal{N}(0, 0.2)$ errors and length of $N = 300$ tested with $H_0 : \theta = 0.002$.

depth based on the residuals only. Assuming a probability of zero for $E_n = 0$ and $Y_n = 0$ one can also use

$$d_S^{AR(1)}(\theta, y) = \frac{1}{\binom{N}{2}} \sum_{1 \le n_1 < n_2 \le N} \mathbb{1}_{\{r_{n_1}(\theta,y)y_{n_1-1} \ge 0, r_{n_2}(\theta,y)y_{n_2-1} \le 0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)y_{n_1-1} \le 0, r_{n_2}(\theta,y)y_{n_2-1} \ge 0\}}, \tag{50}$$

what changes the statistic, if y_n can change sign. Under H_1, especially for $\theta < 0$, this really can happen. We also examine the performance, when (50) is applied instead of (14). Note that under H_0 the limit does not change. The critical values remain the same. The power of the modified test is depicted as red dashed line in Figure 60. We see that the modified d_S test performs similar to the sign test and the full simplicial depth test, when $\theta > 0$ is considered, but outperforms these two tests for $\theta < 0$. This happens, since alternating signs and negative values of the process now are considered. The simplified depth tests show a good power for $\theta > \theta^0$ but behave relatively poor for $\theta < \theta^0$. This also happens due to potential negative values of y_n. Thereby the d_S^2 test is better than the d_S^1 test, followed by the d_S^3 test. In Figure 61, we see the power functions in case of contaminated normal errors. Now, the OLS

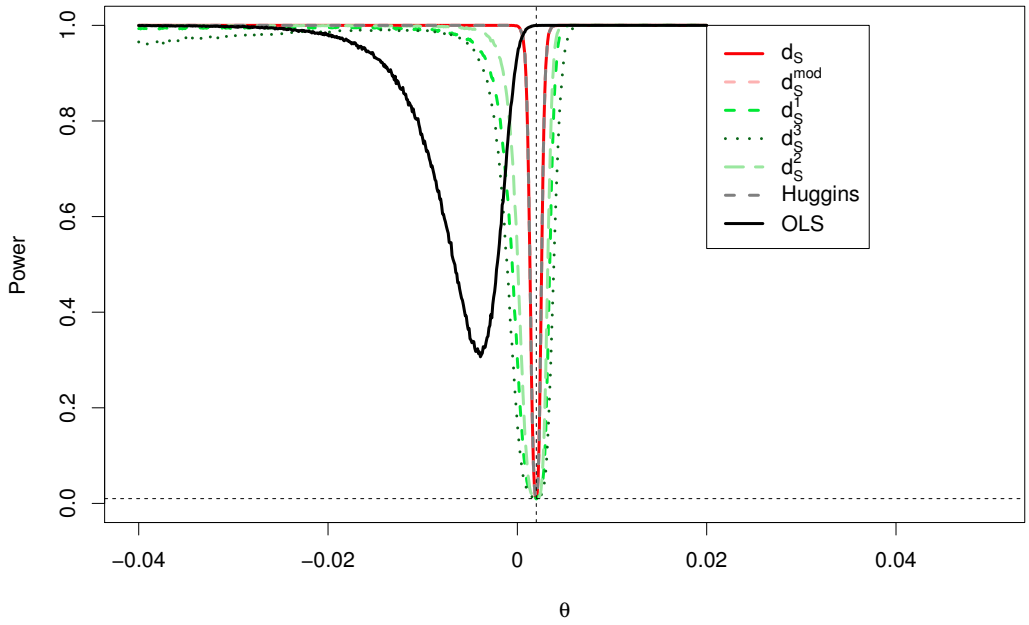


Fig. 61: Simulated power functions based on linear AR(1) processes without intercept, $\mathcal{CN}(0, 0.2)$ errors and length of $N = 300$ tested with $H_0 : \theta = 0.002$.

test completely fails, since it is biased by the asymmetric error distribution. The full simplicial depth test and the sign test again are identical. The modified simplicial depth test here just slightly outperforms the full simplicial depth test. In the Figure, this is not visible, since it happens in the range of $\theta < 0$. In case of contamination the simplified tests also perform well for $\theta < \theta^0$, since the process grows faster, due to positive jumps. Hence, values of $y_n < 0$ are less frequent. The ranking of the power from the simplified depth tests is the same as in the non contaminated case. When the error distribution is heavily skewed the modified depth test is superior to the other proposals, as shown in Figure 62. Here, we consider errors with a Gumbel distribution. We see that the sign and full simplicial depth tests behave identical and have low power for $\theta < \theta^0$. The modified test statistic improves the power in this parameter region. The OLS test completely fails due to the bias caused by the error distribution. Further, the simplified tests are not reliable for $\theta < \theta^0$, but perform well for $\theta > \theta^0$.
Figure 63 shows the results for processes with a sample size of $N = 1000$.

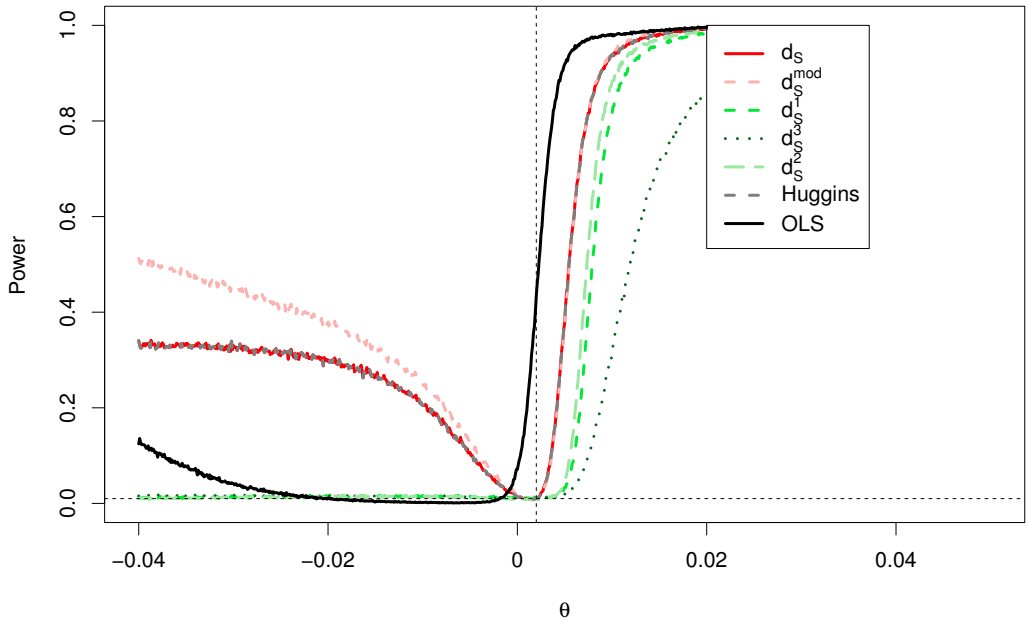
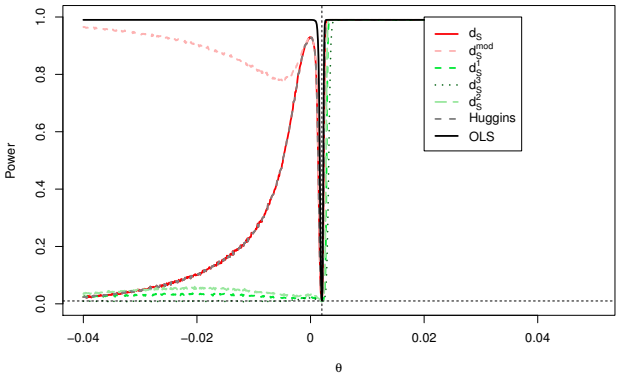
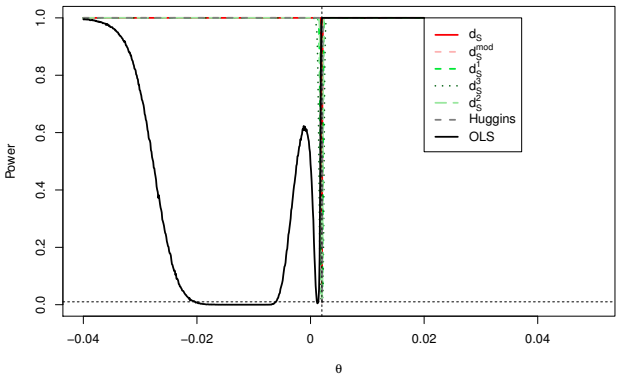


Fig. 62: Simulated power functions based on linear AR(1) processes without intercept, $\mathcal{G}(10, -3.665129)$ errors and length of $N = 300$ tested with $H_0 : \theta = 0.002$.

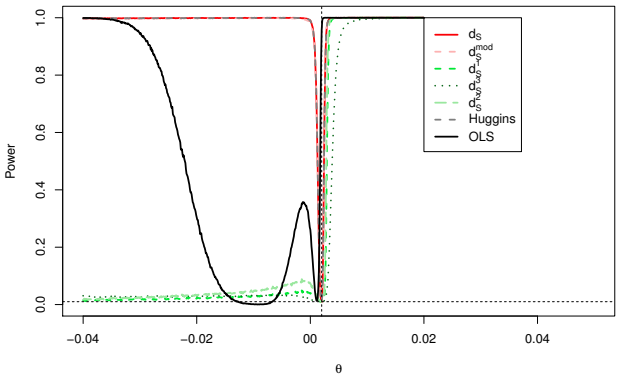
The results are the same as for the smaller sample size. The OLS test is just reliable for normally distributed errors. It is outperformed by the full simplicial depth, the modified simplicial depth and the sign test for the remaining error distributions. The modified simplicial depth can improve the test for $\theta < 0$, if y_n can be lower than zero. In this case the simplified depth tests break down for $\theta < \theta^0$. If the process is strictly positive, the simplified depth tests also perform well.



(a) Normal errors



(b) cont. Normal errors

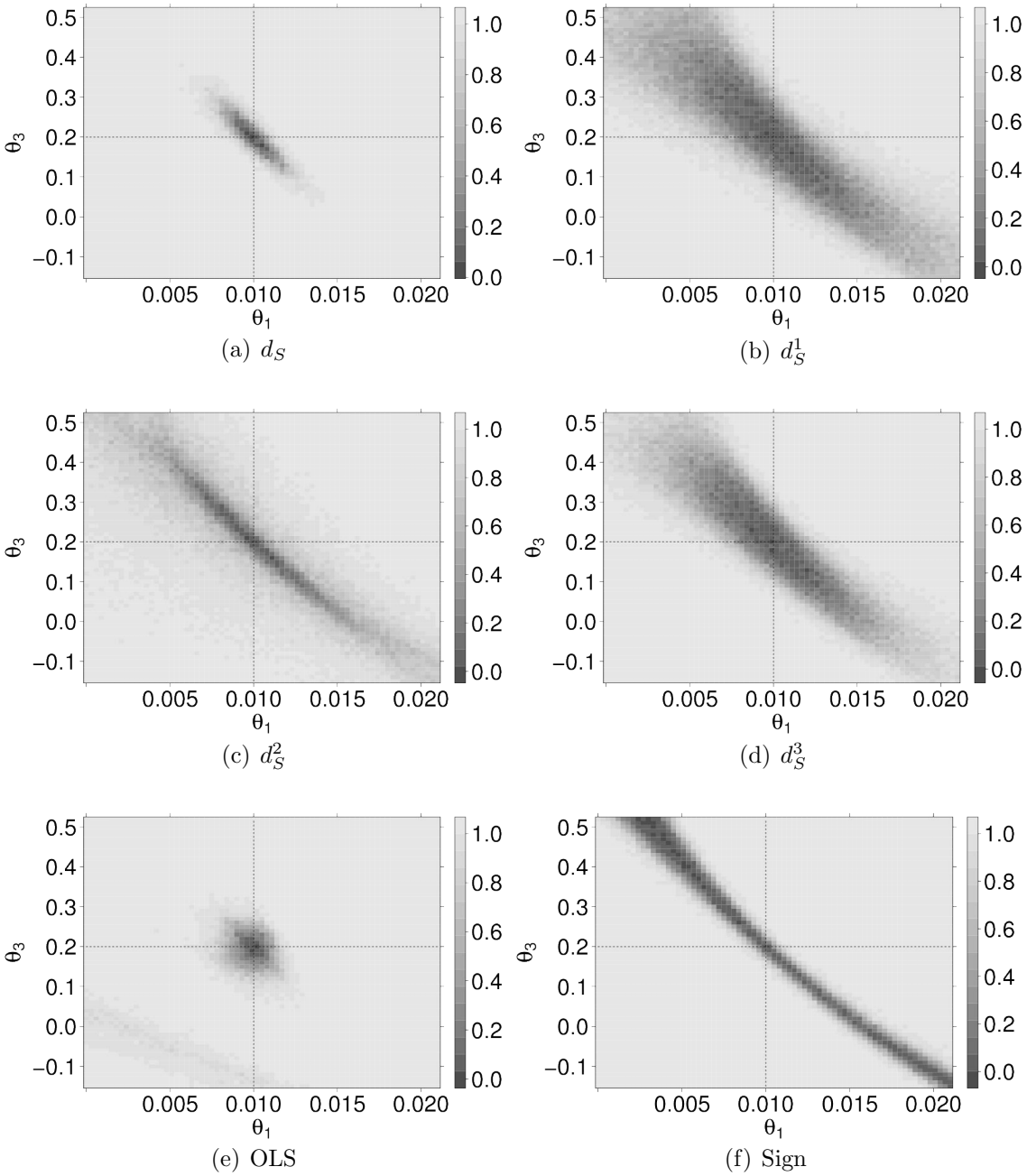


(c) Gumbel errors

Fig. 63: Power of the tests for $N = 1000$.

Linear Model With Intercept

To evaluate the power when an intercept is included, we extend the applied tests to the two parameter case by replacement of the residuals to the two parameter expressions. This can be simply done for the sign test and results in the same distribution under H_0. The results for the simplicial depth based tests were discussed in Section 3.3. For the OLS test, we need limit distributions for explosive AR processes. A derivation is presented by Wang and Yu (2013). Thereby, the critical values are based on an asymptotic independence of the marginal estimators. Since we assume that the exact error distribution is unknown, we apply the OLS test under the assumption of normally distributed errors in all examples. The following results were partially presented in Kustosz et al. (2016a). Here, we extend the simulations by the simplified depth notions.
We compare the test based on simplicial depth for $H_0 : \theta = \theta^0$ using $\theta^0 = (\theta_1^0, \theta_3^0)^\top = (0.01, 0.2)^\top$ with five other tests. We evaluate the power of the six tests on a grid defined by $\theta_3 \in [-0.15, 0.52]$ with mesh size 0.01 and $\theta_1 \in [0, 0.021]$ with mesh size 0.0003. For each grid point, we simulate $R = 100$ processes of length $N = 100$ with the underlying parameter combination and with starting value $y_0 = 15$. As in the one parameter case we consider three different distributions for the errors: a normal distribution with mean zero and variance 0.01, a contaminated normal distribution given by $A_n + P_n \cdot B_n$, whereby $A_n \sim \mathcal{N}(0, 0.1), B_n \sim \mathcal{N}(5, 1)$ and $P_n \sim Pois(5/100)$ are independent random variables for each n, and a Fréchet distribution with parameters $\alpha = 1.928, \beta = -2, \gamma = 10$.



(a) d_S (b) d_S^1

(c) d_S^2 (d) d_S^3

(e) OLS (f) Sign

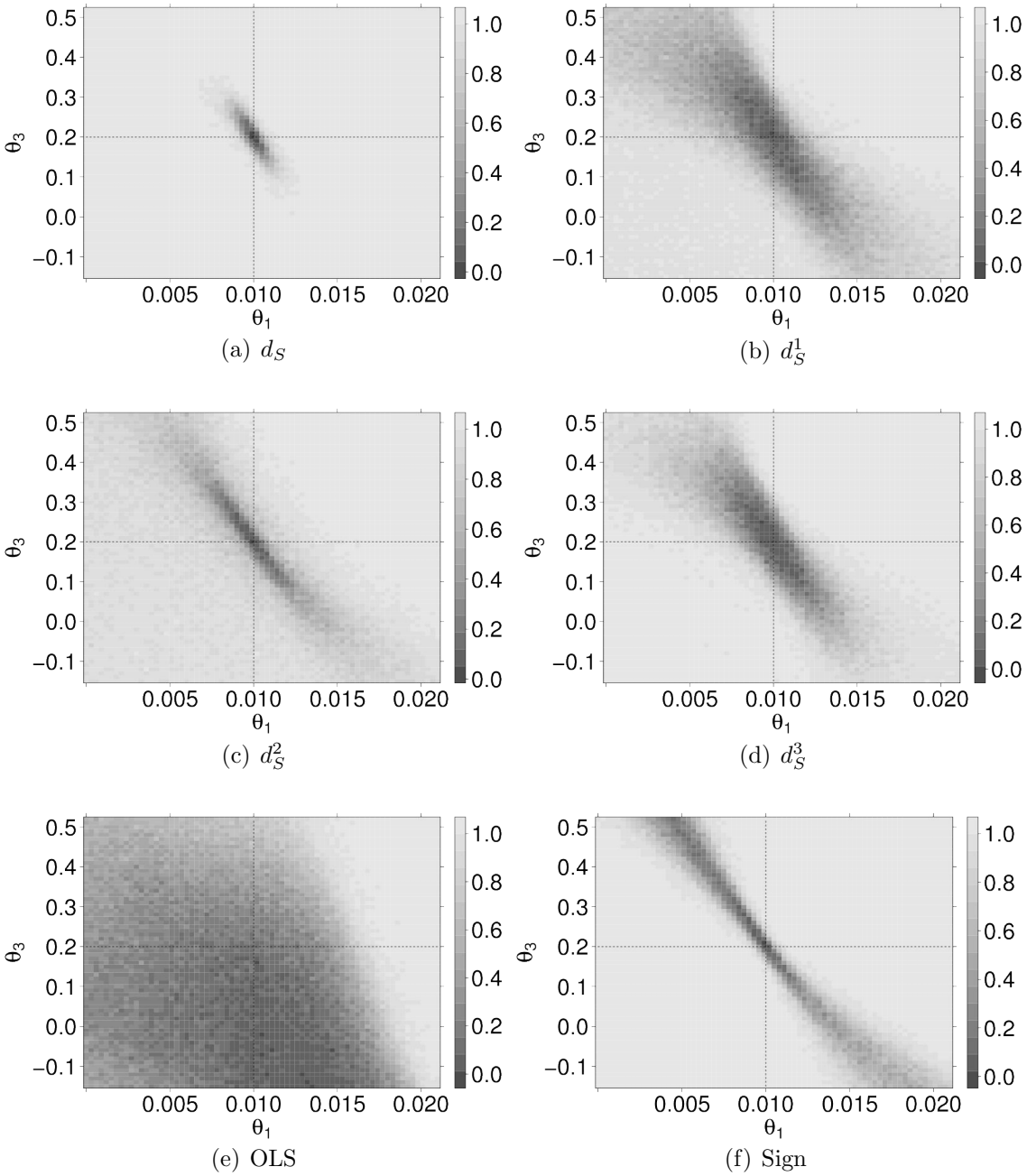
Fig. 64: Power of the tests based on normally distributed errors.

Figure 64 shows the power functions for normally distributed errors. Thereby, the horizontal and vertical lines denote the components of θ^0 so that their intersection is θ^0. One can clearly see that the OLS test performs best under normally distributed errors. This is not surprising, since it assumes the correct error distribution. The sign test behaves quite well close to the alternative. Unfortunately, in case of explosive processes the power also decreases when a combination of θ_3 and θ_1 leads to residuals which have a poor fit but have a median of zero. This for example happens,

if the first half of residuals is positive and the second half is negative. As a result, this test is very unstable in case of explosive AR(1) processes. The d_S test clearly outperforms the simplified depth tests. It also has a better performance than the OLS test in direction of a diagonal with positive slope, but accepts a wider range of values on a diagonal with negative slope.

In Figure 65, the comparison for errors with the contaminated normal distribution is depicted. Figure 66 provides the comparison for errors with the Féchet distribution. Now, the simplicial depth test performs clearly best. The OLS test suffers from heavy bias due to the skewed error distributions and the sign test still shows the identifiability problem. In Figures 114, 115, 116, which can be found in Appendix A, we compare the tests evaluated on the diagonal given by $\theta_3 = 50.7 - 50 \cdot (\theta_1 + 1)$, where the slope of the diagonal is negative. The straight line goes from $(-0.325, 1.0205)$ to $(0.725, 0.9995)$ through H_0 defined by $\tilde{\theta} = (\theta_3, \theta_1) = (0.2, 1.01)^T$ for a model expression in $\kappa = \theta_1 + 1$ and $\tilde{\theta}^0 = (\theta_3^0, \kappa^0)$. In the Figures, the x-axis is defined by the parameter $\lambda \in [0, 1]$ from the parametric form of the straight line given by $(0.725, 0.9995)^T + \lambda \cdot (-1.05, 0.021)^T$. On this line $\lambda = 0.5$ coincides with H_0. Here, the main advantage of the full simplicial depth compared to the sign test is clearly visible. Additionally, these figures show how the new test outperforms the OLS test in the case of non-normal errors where the OLS test in particular does not keep the level anymore.

Summarising, we see that the d_S test can be applied to explosive AR(1) processes under quite general conditions and does not suffer of systematic failure or heavy bias in case of skewed errors or outliers. Further, by the price of additional computational costs, the full simplicial depth statistic defines a test with higher power than the simplified statistics based on simplicial depth.

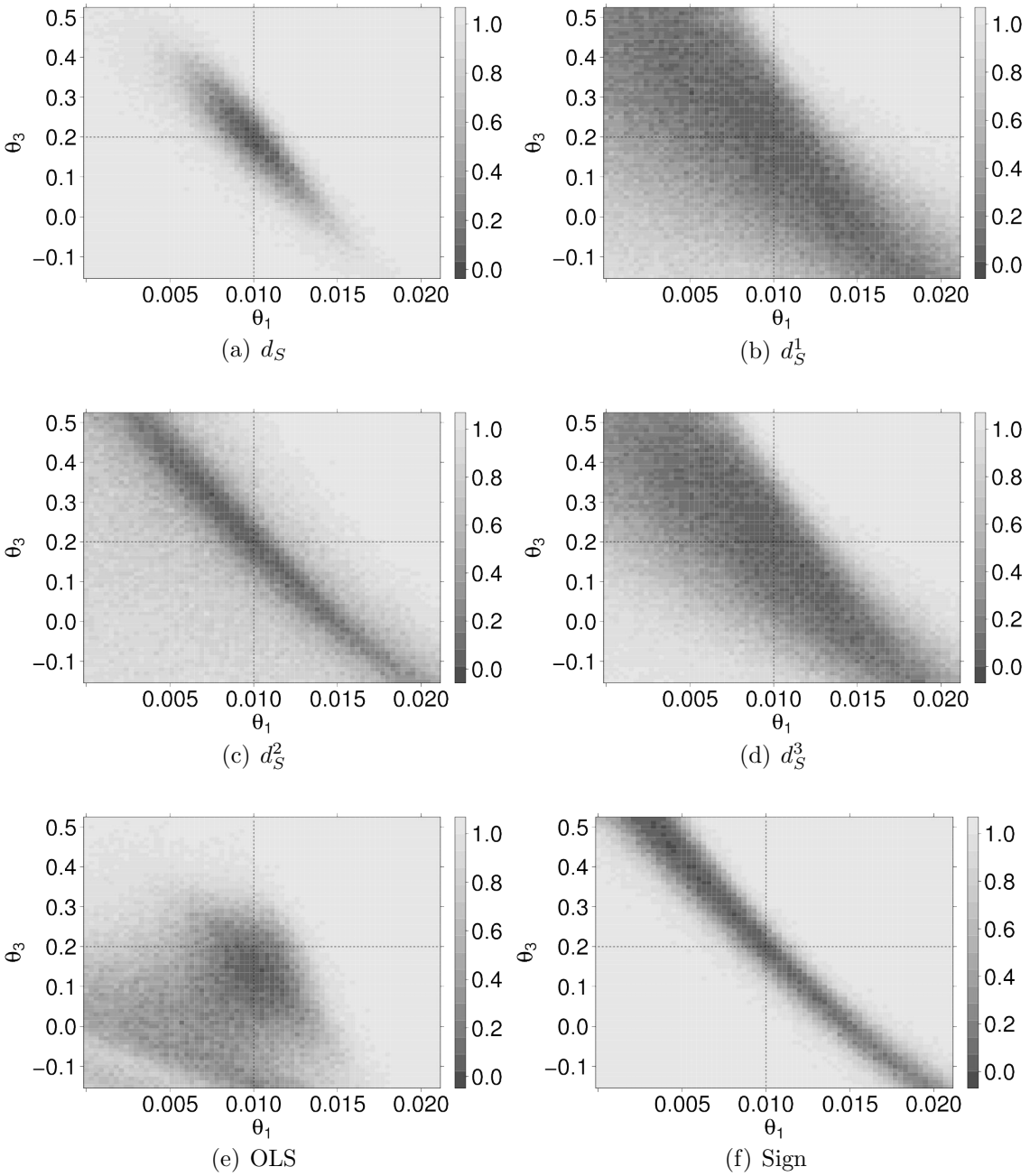


(a) d_S (b) d_S^1

(c) d_S^2 (d) d_S^3

(e) OLS (f) Sign

Fig. 65: Power of the tests based on errors with contaminated normal distribution.



(a) d_S (b) d_S^1

(c) d_S^2 (d) d_S^3

(e) OLS (f) Sign

Fig. 66: Power of the tests based on errors with Fréchet distribution.

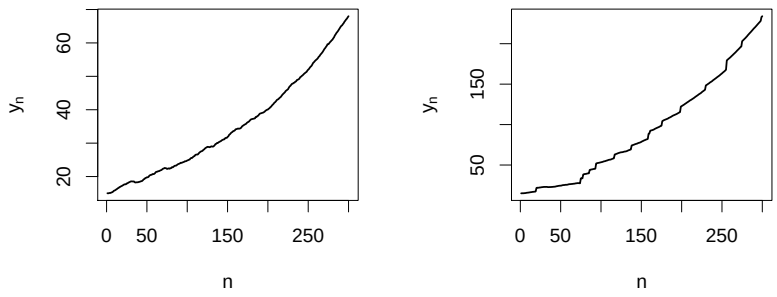
Non-Linear Model

For the non-linear model, we again use the appropriate versions of the proposed tests. The sign test again can be applied by replacement of the residuals. The same argument holds for the OLS test. To compute a test based on the assumption of normal errors, we use asymptotic confidence intervals, defined by Beale (1960) or Ritz and Streibig (2008). Thereby, we use the implementation form the `R` package

`nlstools` by Baty et al. (2015). The results of this thesis extend the results presented in Kustosz et al. (2016b).
We evaluate a process defined by

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + E_n.$$

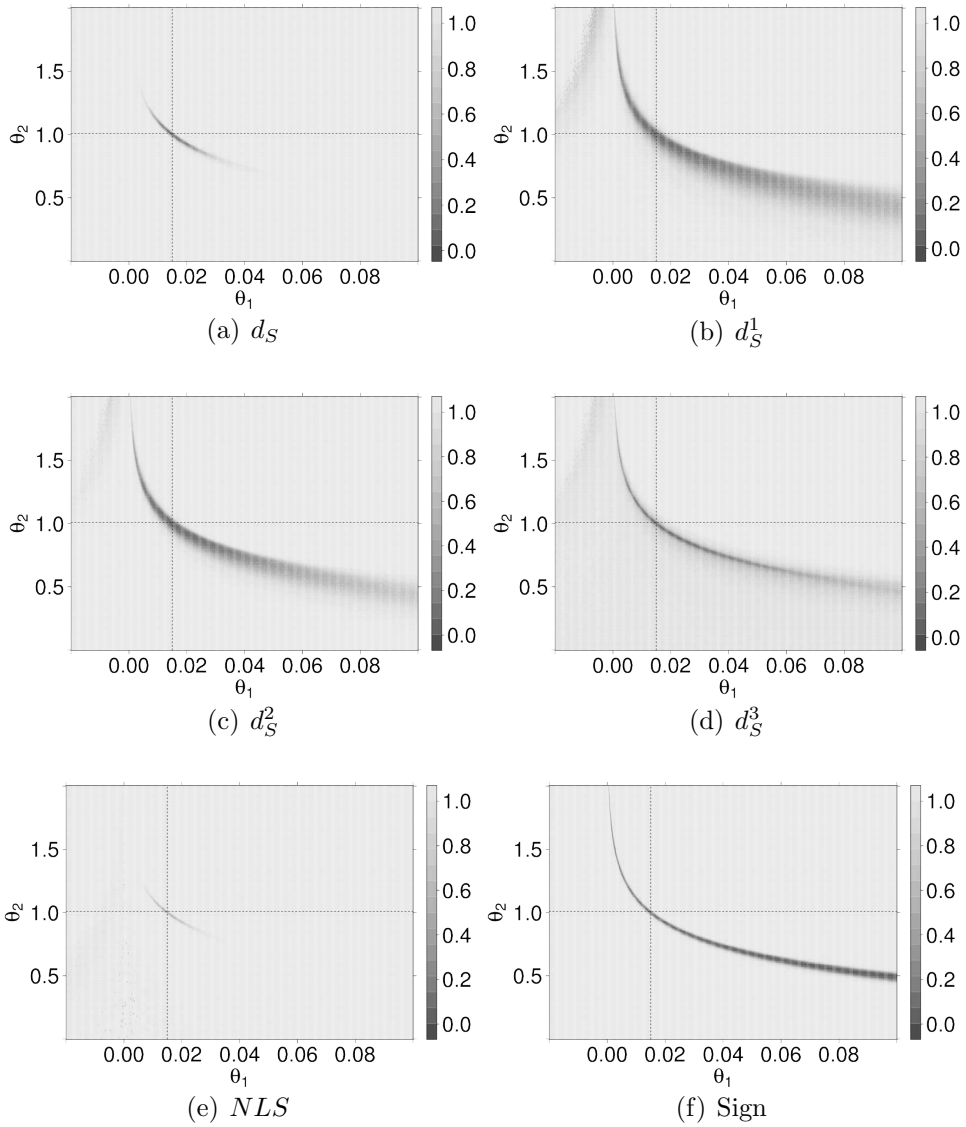
The null hypothesis is given by $H_0 : \theta = (\theta_1^0, \theta_2^0)$ with $\theta_1^0 = 0.005$ and $\theta_2^0 = 1.002$. We examine processes with a starting value of $y_0 = 15$ and a length of $N = 500$. Example processes are depicted in Figure 67.



(a) $\mathcal{N}(0, 0.1)$ Errors (b) Contaminated $\mathcal{N}(0, 0.1)$ Errors

Fig. 67: Realizations of non-linear AR(1) processes with $\theta_1 = 0.015, \theta_2 = 1.002, y_0 = 15$ and two different error distributions

To evaluate the power of tests for $H_0 : \theta = \theta^0 := (\theta_1^0, \theta_2^0)^\top$, a grid defined by $[-0.02, 0.1] \times [0, 2]$ with a step width of 0.0001 for θ_1 and 0.01 for θ_2 is considered. On each grid point the processes are generated 100 times to simulate the power of the test at a 5% level for processes with a length of $N = 500$ observations. The resulting power functions for normal errors are depicted in Figure 68.



(a) d_S (b) d_S^1

(c) d_S^2 (d) d_S^3

(e) NLS (f) Sign

Fig. 68: Simulated power for $\mathcal{N}(0, 0.1)$ errors under the non-linear AR(1) Model. The simulated relative number of rejections of $H_0 : \theta = (0.015, 1.002)$ based on different values of $\theta = (\theta_0, \theta_1)$ is depicted. The errors are simulated as $N(0, 0.1)$ random variables. The processes have a length of $N = 500$ observations. The parameters for the null hypothesis are marked by the dashed lines.

One can observe that the depth based tests have power functions which are increasing to one when the parameter deviates from $H_0 : \theta = (0.015, 1.002)$. Due to the model, the power functions are not symmetric. The NLS test shows the best results for normally distributed errors. It is directly followed by the d_S test. It appears, as if the d_S^3 test slightly outperforms the d_S^2 test followed by the d_S^1 version. By consideration of a wider parameter range, a systematic shortcoming of the sign test

gets obvious. The sign test again does not reject parameters, for which half of the residuals are negative and half are positive, even if the model fit is poor. This is discussed in more detail in the following remark 101. A closer look at the levels, as presented in Kustosz et al. (2016b), reveals a problem of the NLS test. Due to the necessary approximation of the non-linear model, the NLS test hardly holds the level at H_0 for normally distributed errors. In case of non-normal errors, it completely fails to keep the level at any evaluated point.

Remark 101. *The residuals of a process $\tilde{Y}$ defined by $\theta = (\theta_1, \theta_2) \neq (\theta_1^0, \theta_2^0) = \theta^0$ are given by $r_n(\theta^0, \tilde{Y}) = E_n + \theta_1 \tilde{Y}_{n-1}^{\theta_2} - \theta_1^0 \tilde{Y}_{n-1}^{\theta_2^0}$. If the errors are assumed to be approximately zero, then $r_n(\theta^0, \tilde{Y}) \lesseqgtr 0$ holds approximately if and only if $\theta_1 \lesseqgtr \theta_1^0 \tilde{Y}_{n-1}^{\theta_2^0 - \theta_2}$. Since $\tilde{Y}$ is strictly increasing, we obtain for $\theta_2 < \theta_2^0$ that*

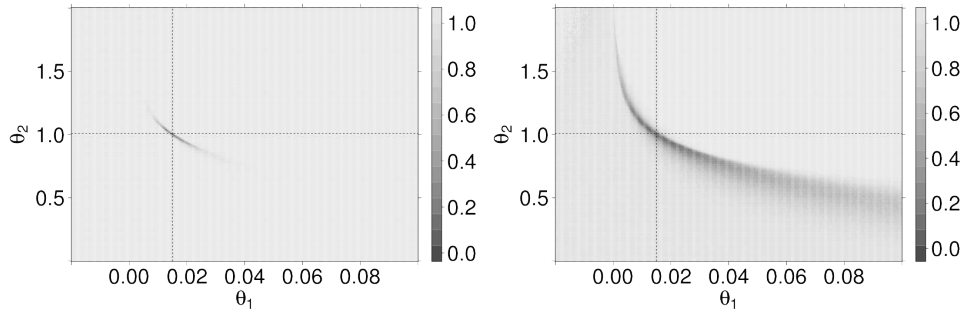
$$\theta_1^0 \tilde{Y}_0^{\theta_2^0 - \theta_2} < ... < \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor - 1}^{\theta_2^0 - \theta_2} < \theta_1 < \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor}^{\theta_2^0 - \theta_2} < ... < \theta_1^0 \tilde{Y}_N^{\theta_2^0 - \theta_2}$$

implies $r_n(\theta^0, \tilde{Y}) > 0$ for $n \in \{1, ..., \lfloor N/2 \rfloor\}$ and $r_n(\theta^0, \tilde{Y}) < 0$ for $n \in \{\lfloor N/2 \rfloor + 1, ..., N\}$. Similarly, if $\theta_2 > \theta_2^0$ then

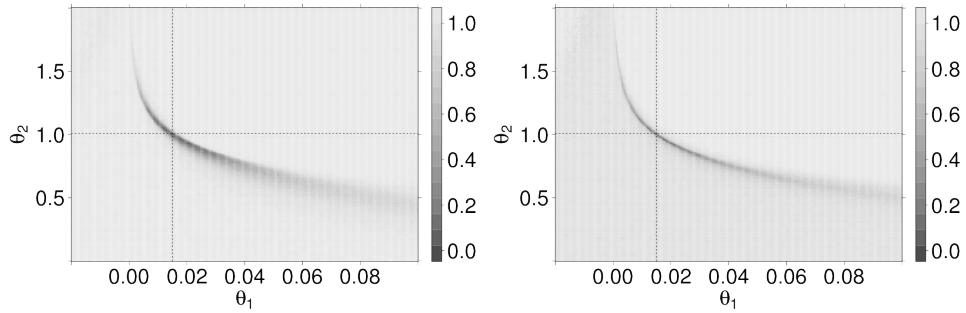
$$\theta_1^0 \tilde{Y}_0^{\theta_2^0 - \theta_2} > ... > \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor - 1}^{\theta_2^0 - \theta_2} > \theta_1 > \theta_1^0 \tilde{Y}_{\lfloor N/2 \rfloor}^{\theta_2^0 - \theta_2} > ... > \theta_1^0 \tilde{Y}_N^{\theta_2^0 - \theta_2}$$

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

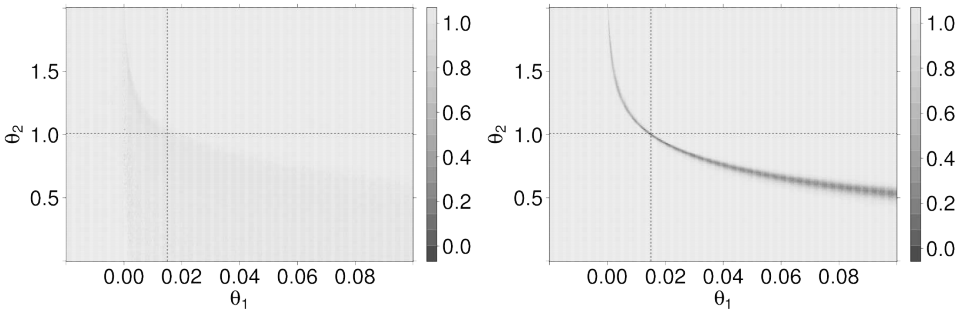
For an error distribution which is contaminated with positive outliers in 5% of all cases, the resulting power functions are presented in Figure 69. As in the non-contaminated case the region of the depth based tests with low power is bounded while the sign test shows a systematic problem for a range of parameters with small θ_2. In general, the power functions are steeper, since the jumps lead to a faster growing process, what is exploited by the proposed tests. The NLS test now does not hold the level and hence is not a $1 - \alpha$ level test anymore.



(a) d_S (b) d_S^1



(c) d_S^2 (d) d_S^3



(e) NLS (f) Sign

Fig. 69: Simulated power for $\mathcal{CN}(0, 0.1)$ Errors under the non-linear AR(1) Model. The simulated relative number of rejections of $H_0 : \theta = (0.015, 1.002)$ based on different values of $\theta = (\theta_0, \theta_1)$ is depicted. The errors are simulated as contaminated $N(0, 0.1)$ random variables, whereby in a fraction of 5% variables with a $N(5, 1)$ distribution are added. The parameters for the null hypothesis are marked by the dashed lines.

Summarising, we see that the full simplicial depth allows us to test the parameter of growth processes with a high power, independent of the exact error distribution and model. Hence, in case of sparse information on the model, e.g. when just few experiments are available and more experiments are very costly, the depth based tests deliver an interesting alternative to the standard approaches to test the pa-

rameters simultaneously. In addition, the simplified depth notions can be applied as less costly robust alternatives, when the d_S depth is appropriate.

5.3 Confidence Intervals

In this section, we demonstrate that the proposed parameter confidence intervals perform as expected by calculation of the empirical coverage rates in different situations.

Coverage Rates for the Linear Model Without Intercept

Starting with the linear model without intercept, we compare the coverage rates of the confidence intervals based on d_S and d_S^i for a growth parameter of $\theta_1 = 0.005$ and different error distributions, as well as sample sizes
$N \in \{10, 50, 100, 200, 250, 500, 750, 1000\}$. The first comparison shows the coverage rates when the errors are normally distributed with mean 0 and standard deviation 0.2. The results are depicted in Figure 70. Thereby, we simulate the confidence intervals $R = 10000$ times for each sample size N and calculate the coverage rates by the relative number of simulations in which the true parameter is located in the confidence interval.

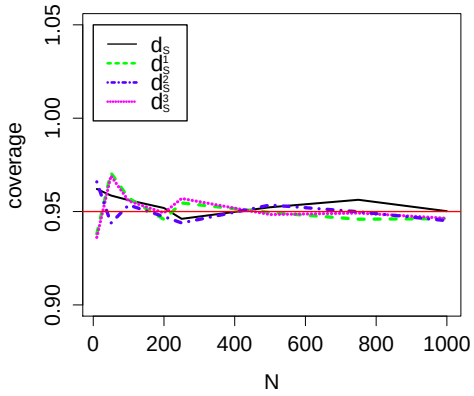


Fig. 70: Coverage rates for the depth based tests with different sample sizes for linear AR(1) processes without intercept and $\theta_1 = 0.005$, $E_n \sim \mathcal{N}(0, 0.2)$.

The coverage rates are close to the desired level of $\gamma = 95\%$.
Applying contaminated normal errors delivers Figure 71.

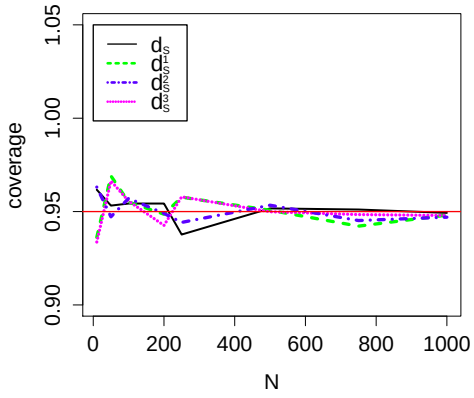


Fig. 71: Coverage rates for the depth based tests with different sample sizes for linear AR(1) processes without intercept and $\theta_1 = 0.005$, $E_n \sim \mathcal{CN}(0, 0.2)$.

The results are similar to the non-contaminated case. The full depth intervals are closer to the pre-set level for sample sizes larger than $N = 500$.
The last simulation, depicted in Figure 72, shows the results for Fréchet distributed errors, leading to the same conclusions.

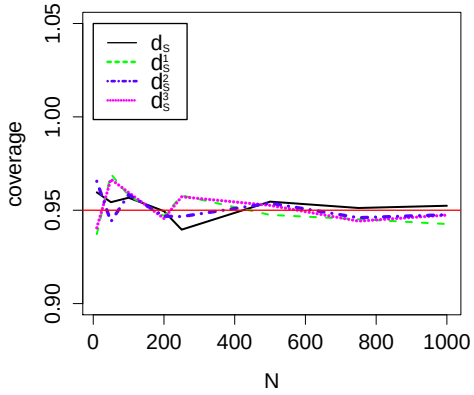
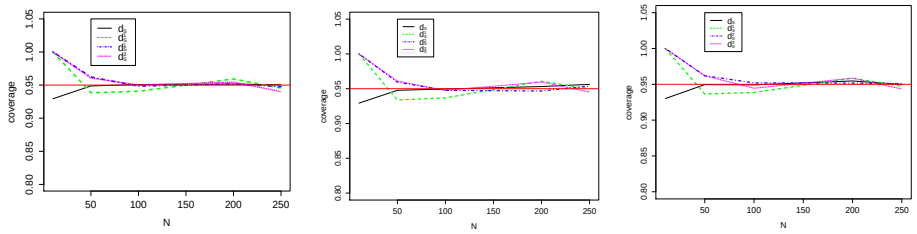


Fig. 72: Coverage rates for the depth based tests with different sample sizes for linear AR(1) processes without intercept and $\theta_1 = 0.005$, $E_n \sim \mathcal{F}(1.928, -2, 10)$.

We can sumarise that the full depth confidence intervals are more reliable than the simplified notions. Nevertheless, the simplified depth notions asymptotically also hold the level and just show small deviations of the pre-set values.

Coverage Rates for the Linear Model With Intercept

A similar study for the linear model with intercept is shown in Figure 73.

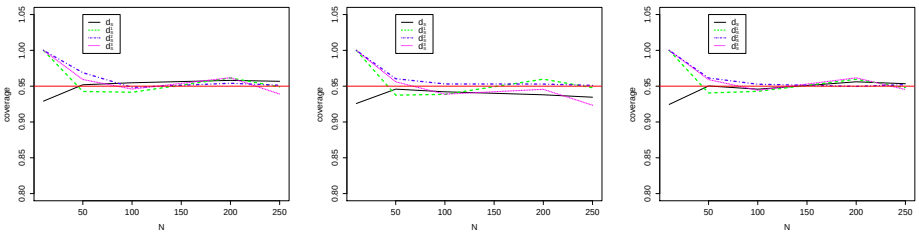


(a) $\mathcal{N}(0, 0.2)$ (b) $\mathcal{CN}(0, 0.2)$ (c) $\mathcal{F}(1.928, -2, 10)$

Fig. 73: Coverage rates for the depth based tests with different sample sizes for linear AR(1) processes without intercept and $\theta_1 = 1.005, \theta_3 = 0.1$.

Coverage Rates for the Non-Linear Model

Evaluations for the non-linear model are presented in Figure 74.



(a) $\mathcal{N}(0, 0.2)$ (b) $\mathcal{CN}(0, 0.2)$ (c) $\mathcal{F}(1.928, -2, 10)$

Fig. 74: Coverage rates for the depth based tests with different sample sizes for non-linear AR(1) processes and $\theta_1 = 0.01, \theta_2 = 0.005$.

We can conclude that the asymptotic results support our proofs and the method really delivers $(1 - \alpha)$ confidence sets for the parameters.

Relation Between Full and Simplified Depth Confidence Intervals

Reconsidering the examples in Section 4.5 leads to a proposal for faster calculation of full simplicial depth confidence sets. For all considered models, the d_S confidence regions were covered by the simplified versions in the presented examples. If this would be true in general, we could reduce the candidates for a calculation of the d_S confidence regions to candidates which are already included in the confidence set of

the smallest region from the simplified notions. Since we do not have theoretical results on the relation of the different confidence regions, we propose to base the improved algorithm to calculate d_S confidence intervals on the d_S^1 statistic, to assure the consideration of all relevant candidates. Since the d_S^1 confidence regions define the smallest confidence regions in our set of test statistics, but also are very likely to cover the full simplicial depth confidence regions completely. This can be also motivated empirically by the following simulation studies for the linear models with and without intercept, as well as for the non-linear model.

Relation for the Linear Model Without Intercept

	Rate of overlapment			Mean size of regions			
	d_S^1	d_S^2	d_S^3	d_S^1	d_S^2	d_S^3	d_S
$\mathcal{N}(0, 0.2)$	81.2%	83.0 %	83.8 %	0.0105	0.0085	0.1134	0.0041
$\mathcal{CN}(0, 0.2)$	81.5%	81.9 %	83.5 %	0.007	0.0068	0.0085	0.0027
$\mathcal{F}(1.928, -2, 10)$	81.9%	82.9 %	82.3 %	0.0108	0.0084	0.0112	0.004
results when we just check bounds							
$\mathcal{N}(0, 0.2)$	95.1%	91.3 %	96.2 %	0.0105	0.0085	0.1134	0.0041
$\mathcal{CN}(0, 0.2)$	94.7%	87.8%	97.0%	0.007	0.0068	0.0085	0.0027
$\mathcal{F}(1.928, -2, 10)$	94.7%	89.5%	96.7%	0.0108	0.0084	0.0112	0.004

Tab. 5: Covarage of the full depth confidence set by reduced depth based confidence sets for the linear model without intercept with $\theta_1 = 0.01, y_0 = 15, N = 100$, based on 1000 simulations.

Table 5 shows, how often the full simplicial depth confidence regions are completely covered by the confidence regions based on the simplified notions for the linear model without intercept. To generate these numbers, we simulated the underlying process 1000 times and calculated the four confidence sets for each simulation. Then, we counted the number of simulations for which the full depth regions are completely covered by the regions based on the simplified notions. In addition, we give the mean sizes of the resulting intervals. We see that the rates of the overlapping are quite high. The d_S^3 statistic has the highest rates for both considered sample sizes. This coincides with the largest intervals, measured by the mean distance of the endpoints, in all simulations. The lowest coverage rates are resulting from the d_S^2 statistic which also has the smallest confidence intervals in the set of simplified statistics. The full simplicial depth has the smallest confidence sets in total.
Another interesting observation for the linear process without intercept is that the results change when we just check, if the boundaries of the full depth intervals are

fully included in the maximal and minimal parameter from the simplified sets. In this case the rates rise remarkably. A closer look at the simulations reveals the following reason. Due to the evaluation of the exact candidate set, we deflate depth. Hence, the confidence set might be not connected. As a result, the intervals do not cover the full simplicial depth region in these cases, if all parameters are checked individually. This can be *corrected* by taking the minimal and maximal values. Then the discontinuity problem at the candidate points is avoided.
Our proposal for the linear model without intercept is, to evaluate the d_S^1 intervals to reduce the candidates for the construction of confidence sets based on full simplicial depth. This allows a precise and fast evaluation of the full depth interval in about 95% of all applications.

Relation for the Linear Model With Intercept

The relation of confidence sets for the linear model with intercept is studied in Table 6. Here, the d_S^2 statistic has the highest coverage rates. In addition the size of the intervals is low, compared to the d_S^1 and d_S^3 intervals. Hence, we propose to base the candidate preselection on the d_S^2 intervals in case of linear AR processes with intercept. However, the d_S^1 results do not appear much worse. This results in a reliable and fast construction of full depth confidence sets in about 90% of applications.

	Rate of overlapment			Mean size of regions			
	d_S^1	d_S^2	d_S^3	d_S^1	d_S^2	d_S^3	d_S
$\mathcal{N}(0, 0.2)$	89.2 %	92.2 %	77.5 %	31.44	8.27	5169.81	0.55
$\mathcal{CN}(0, 0.2)$	86.4 %	89.4 %	69.6 %	35.05	11.63	1841.30	0.36
$\mathcal{F}(1.928, -2, 10)$	88.1 %	89.8 %	75.0 %	17.04	7.28	752.95	0.50

Tab. 6: Covarage of the full depth confidence set by reduced depth based confidence sets for the linear model with intercept with $\theta_1 = 0.001, \theta_3 = 0.1, y_0 = 15, N = 100$, based on 1000 simulations.

Relation for the Non-Linear Model

In Table 7 the same study for the non-linear model is presented. Now, the coverage rates are clearly lower. The sizes are quite uninformative, since the non-linearity leads to large distances. Due to rates below 70 % for all simplified notions in the non-linear model, a preselection of candidates for full depth by pre-evaluation cannot be recommended.
Summarising, we can claim that in the linear models the simplified notions can be

used with some caution. For exact results, one should be aware that by a pre-evaluation of the simplified statistics potential candidates for the d_S regions could be neglected. For the non-linear model a pre-selection cannot be recommended.

	Rate of overlapment			Median size of regions			
	d_S^1	d_S^2	d_S^3	d_S^1	d_S^2	d_S^3	d_S
$\mathcal{N}(0, 0.2)$	67.5 %	69.2 %	25.7 %	$7.26 \cdot 10^{94}$	$9.15 \cdot 10^{74}$	$4.31 \cdot 10^{88}$	$5 \cdot 10^3$
$\mathcal{CN}(0, 0.2)$	63.8 %	68.2 %	23.9 %	$1.58 \cdot 10^{71}$	$1.63 \cdot 10^{53}$	$2.30 \cdot 10^{72}$	$4.98 \cdot 10^1$
$\mathcal{F}(1.928, -2, 10)$	66.2 %	69.7 %	24.1 %	9.39	5.17	4.08	1.37

Tab. 7: Coverage of the full depth confidence set by reduced depth based confidence sets for the non-linear model with $\theta_1 = 0.01, \theta_2 = 1.002, y_0 = 15, N = 100$.

5.4 Prediction

In this section, we show empirically that our bootstrapping procedure appears to be valid even if we are not able to prove this validity theoretically so far. At the same time, we compare the coverage rates of our prediction intervals with existing methods based on bias correction and bootstrapping as presented in Kim (2003, 2001); Kilian (1998) and Thombs and Schuchany (1990).

Alternative Prediction Methods

The Bootstrap-after-Bootstrap (BaB) prediction introduced by Kim (2001), following Kilian (1998), calculates a bias corrected parameter estimate by application of a bootstrapping scheme and then, based on the first bootstrap, computes bootstrapping confidence intervals as proposed by Efron (1979).
The second method follows Thombs and Schuchany (1990) who propose simple bootstrap prediction (BS) intervals based on the OLS estimator. A similar approach was proposed by Stute and Gründer (1993) for explosive autoregression explicitly. The third method (BC) uses the approach of Shaman and Stine (1988); Stine and Shaman (1989) who introduce a bias corrected estimator for the coefficient of AR processes. By the extension of Kim (2003) this leads to bootstrap prediction intervals based on mean unbiased estimators.

Comparative Simulation Study

Our simulation study covers three different error distributions and three sample sizes. We compare the size of the prediction intervals and the coverage rates of the true value. The results for $N = 100$ are presented in Tables 8, 9 and 10. Thereby,

in all simulations we performed 1000 repetitions.

	depth	BaB	BS	BC
coverage rate	95.5%	75.4 %	94.7%	90.5%
size of interval	3.91	4.94	4.58	4.45

Tab. 8: Prediction for the linear model without intercept for $N = 100$ observations, $\theta_1 = 0.01, y_0 = 15, E_n \sim \mathcal{N}(0, 0.2)$ and a prediction at $N_f = 115$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	93.8%	85.1 %	91.8%	87.4%
size of interval	18.84	20.28	26.51	20.50

Tab. 9: Prediction for the linear model without intercept for $N = 100$ observations, $\theta_1 = 0.01, y_0 = 15, E_n \sim \mathcal{CN}(0, 0.2)$ and a prediction at $N_f = 115$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	92.5%	75.2 %	93.1%	90.6%
size of interval	5.66	6.70	6.62	6.48

Tab. 10: Prediction for the linear model without intercept for $N = 100$ observations, $\theta_1 = 0.01, y_0 = 15, E_n \sim \mathcal{G}(10, -3.665129)$ and a prediction at $N_f = 115$, $\gamma = 0.95$.

For short term prediction, the simplicial depth based method shows the smallest prediction intervals with quite reliable coverage rates, see Tables 8, 9 and 10. The bootstrap prediction has similar coverage rates, but also larger prediction intervals. The remaining methods are always worse than the simplicial depth prediction and the bootstrap prediction.
For larger forecast horizons, we get the results presented in Tables 11, 12 and 13. While the coverage rates of the simplicial depth prediction fall just slightly below the rates from bootstrap prediction, the size of the simplified depth prediction intervals now increases. The bootstrap method is clearly superior here. Nevertheless, the simplified depth prediction is quite reliable and appears to be a valid prediction method. We also want to compare the effect of a larger sample size. The results are depicted in Tables 14, 15 and 16 for short term prediction.
In Tables 17, 18 and 19 results for a larger prediction horizon are presented.

	depth	BaB	BS	BC
coverage rate	91.1%	49.4 %	94.1%	73.1%
size of interval	11.55	14.21	4.59	14.88

Tab. 11: Prediction for the linear model without intercept for $N = 100$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{N}(0, 0.2)$ and a prediction at $N_f = 150$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	90.3%	68.1 %	91.1%	73.6%
size of interval	46.53	54.91	26.46	58.07

Tab. 12: Prediction for the linear model without intercept for $N = 100$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{CN}(0, 0.2)$ and a prediction at $N_f = 150$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	91.0%	51.1 %	93.7%	73.3%
size of interval	16.60	20.29	6.58	21.34

Tab. 13: Prediction for the linear model without intercept for $N = 100$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{G}(10, -3.665129)$ and a prediction at $N_f = 150$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	96.3%	0 %	93.4%	99.8%
size of interval	5.51	207.02	3.77	479.72

Tab. 14: Prediction for the linear model without intercept for $N = 500$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{N}(0, 0.2)$ and a prediction at $N_f = 515$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	96.1%	0 %	93.8%	99.7%
size of interval	20.26	416.91	21.48	1244.77

Tab. 15: Prediction for the linear model without intercept for $N = 500$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{CN}(0, 0.2)$ and a prediction at $N_f = 515$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	96.7%	0 %	93.5%	99.9%
size of interval	7.28	280.65	5.44	667.74

Tab. 16: Prediction for the linear model without intercept for $N = 500$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{G}(10, -3.665129)$ and a prediction at $N_f = 515$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	96.4%	0 %	94.4%	0%
size of interval	20.38	312.70	3.78	934.95

Tab. 17: Prediction for the linear model without intercept for $N = 500$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{N}(0, 0.2)$ and a prediction at $N_f = 550$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	96.6%	0 %	93.2%	0%
size of interval	51.47	743.76	21.38	2423.63

Tab. 18: Prediction for the linear model without intercept for $N = 500$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{CN}(0, 0.2)$ and a prediction at $N_f = 550$, $\gamma = 0.95$.

	depth	BaB	BS	BC
coverage rate	97.3%	0 %	94.3%	0%
size of interval	25.24	443.98	5.43	1310.82

Tab. 19: Prediction for the linear model without intercept for $N = 500$ observations, $\theta = 1.01, y_0 = 15, E_n \sim \mathcal{G}(10, -3.665129)$ and a prediction at $N_f = 550$, $\gamma = 0.95$.

We can conclude that in case of the short term forecast the simplicial depth prediction is at least comparable with the bootstrap prediction. It is more conservative, but therefore also holds the level. When a larger prediction horizon is considered, the intervals of the simplicial depth prediction are clearly larger than the bootstrap intervals. This again happens with coverage rates above γ, while the rates for the bootstrap prediction are below the desired level. The proposed prediction methods have not been developed for models with multiple parameters so far. The main issue to be solved in this context is an appropriate modification of the random parameter generation as discussed in Section 4.8. This is an aim for future research.

5.5 Change Points

In this chapter, we compare our heuristic change point detection procedures with existing methods.

Compared Change Point Detection Methods

To compare our methods with other available proposals, we use the `R` packages `strucchange` and `segmented`. These sources implement common change point detection methods for time series. Further, we compare both proposed variations of the depth based change point detection Algorithms 91 and 92. We refer to the method based on non-overlapping confidence sets, defined by Algorithm 91 as Method 1 and to the estimator based method from Algorithm 92 as Method 2.
The `strucchange` package implements results discussed by Bai and Perron (1998, 2003). The application is also addressed by Zeileis et al. (2003). The general approach to detect the change points is based on a model dependent residual sum of squares. Thereby this sum is minimised with respect to potential change points $(i_1, ..., i_m)$ in the observed data. The actual algorithm in the package uses the Bellman principle, as proposed by Bai and Perron (2003) to allow a dynamic programming algorithm.
The second proposal from literature, available via the `segmented` package, is based on the results of Muggeo (2003). Here, a break point detection for general regression models is presented. Based on the first order Taylor expansion, piecewise linear models are constructed for which an iterative breakpoint detection procedure is formulated.

Comparison for one Change Point and Normal Errors

In Figure 75 a simulated series with $N = 100$ observations and a breakpoint at $n = 50$ is depicted. The comparison of the detection methods, based on multiple realisations of this process, are presented in Figure 76. Thereby, we consider random draws from a linear autoregressive model with intercept and parameters $\theta_1 = 0.01, \theta_3 = 0$ for $n \in \{1, ...50\}$ and $\theta_1 = 0.02, \theta_3 = 0$ for $n \in \{51, ..., 100\}$. The errors are independent and identically $\mathcal{N}(0, 0.2)$ distributed. For the change point detection based on simplicial depth, we applied the tests based on the statistic for the AR parameter θ_1 only and the full test for a model with intercept θ_3. All methods are preformed on an $\alpha = 0.05$ level. For the post-processing, we neglected all potential change point clusters with a length below 75% of the longest detected cluster in the depth based algorithms.

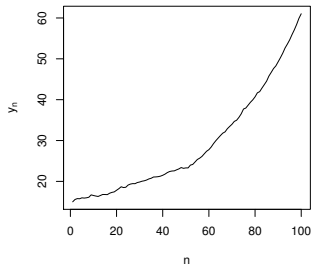
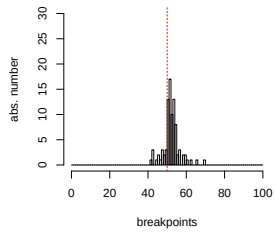


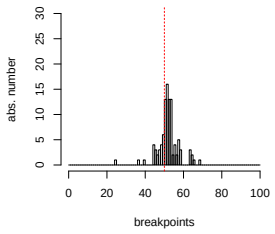
Fig. 75: Simulated AR(1) process with parameter $\theta_1 = 0.01$ at $n \in \{1, ..., 50\}$ and parameter $\theta_1 = 0.02$ for $n \in \{51, ..., 100\}$. $E_n \sim \mathcal{N}(0, 0.2)$ and $\theta_3 = 0$.

In Figure 76 (l), we see that the algorithm from the `segmented` package detects exactly one change point in each simulation. The reason is that the algorithm uses a preliminary estimate for a singular breakpoint and hence deterministically detects one change in each simulation. The distribution of the detected change points from this algorithm is centred around a value of 40, instead of the true value of 50, and shows an asymmetry with many values between 60 and 80. The `strucchange` algorithm shows a much lower deviation from the true value at 50 but it does not detect change points in every simulation. In particular, Figure 76 (k) shows that in about 40 of the 100 simulations no change point was detected. Further in few simulations multiple change points are detected by this method. The depth based methods also show a distribution of the detected change points which is concentrated around the true change point location at observation 50. Instead of a high rate of non-detection, Method 1 detects multiple change points more often,

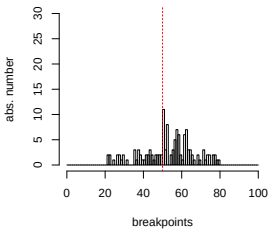
than the `strucchange` algorithm. The cases, where no change point is detected, are lower than the `strucchange` results for all proposals. However, the variance of the location estimates is higher than the variance from the `strucchange` estimates. The one-dimensional depth methods possess a lower variance than the two dimensional ones. This is not surprising, since the two dimensional proposals need to consider two parameters, while the one dimensional methods assume $\theta_3 = 0$, what is correct in our simulation study. All depth based methods also detect multiple change points in few cases.



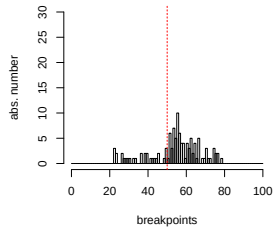
(a) 1 dim, Method 1



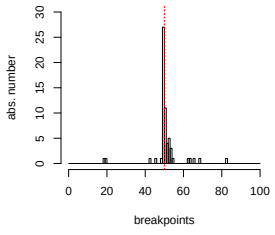
(b) 1 dim, Method 2



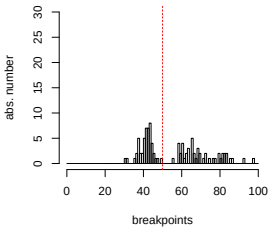
(c) 2 dim, Method 1



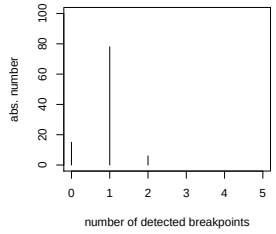
(d) 2 dim, Method 2



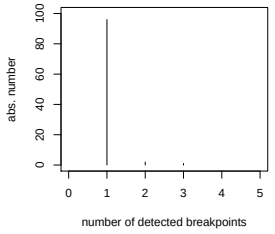
(e) strucchange



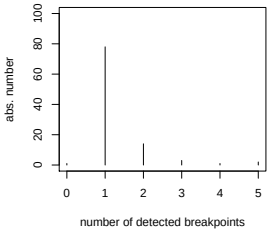
(f) segmented



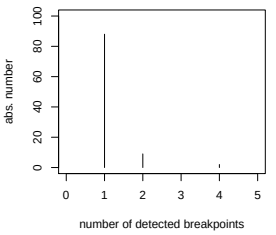
(g) 1 dim, Method 1



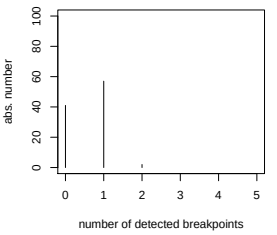
(h) 1 dim, Method 2



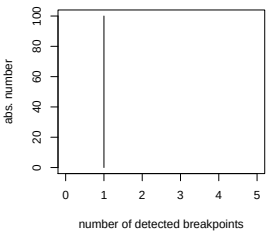
(i) 2 dim, Method 1



(j) 2 dim, Method 2



(k) strucchange

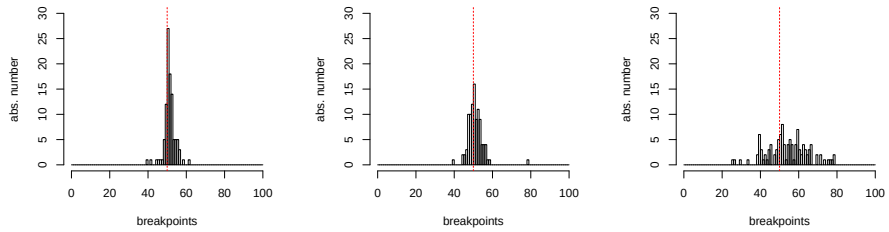


(l) segmented

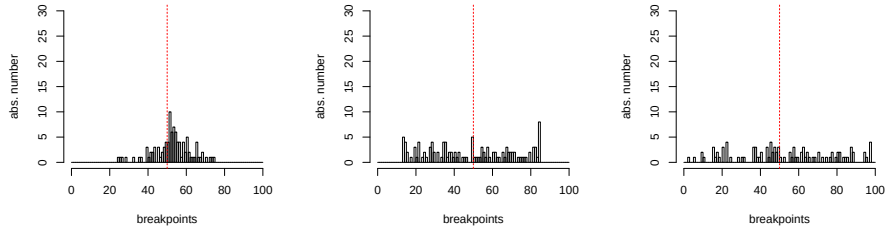
Fig. 76: Change points from $R = 100$ simulations of processes defined by $\theta_1 = 0.01$ at $n \in \{1, ..., 50\}$ and $\theta_1 = 0.02$ for $n \in \{51, ..., 100\}$, $E_n \sim \mathcal{N}(0, 0.2)$, $\theta_3 = 0$.

Comparison for one Change Point and Contaminated Normal Errors

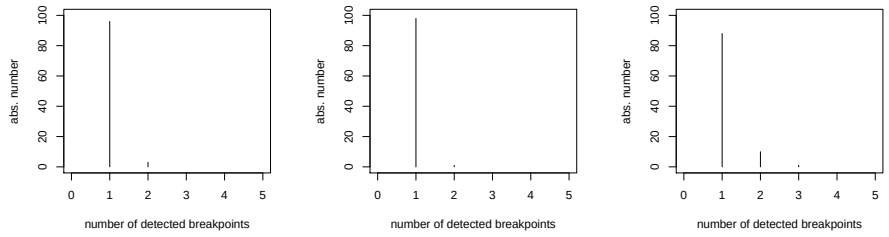
In Figure 77, the results based on contaminated normal errors are presented.



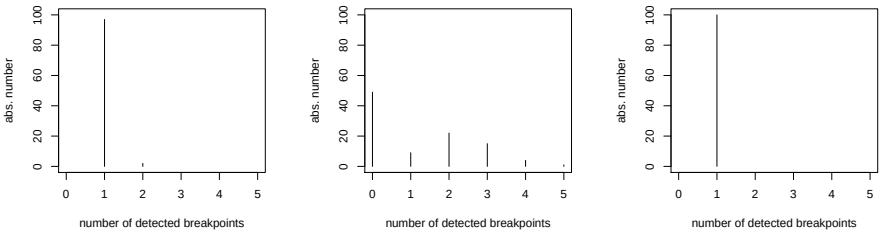
(a) 1 dim, Method 1 (b) 1 dim, Method 2 (c) 2 dim, Method 1



(d) 2 dim, Method 2 (e) strucchange (f) segmented



(g) 1 dim, Method 1 (h) 1 dim, Method 2 (i) 2 dim, Method 1



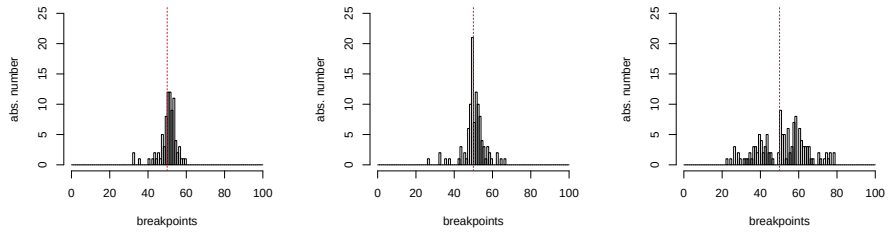
(j) 2 dim, Method 2 (k) strucchange (l) segmented

Fig. 77: Change points from $R = 100$ simulations of processes defined by $\theta_1 = 0.01$ at $n \in \{1, ..., 50\}$ and $\theta_1 = 0.02$ for $n \in \{51, ..., 100\}$, $E_n \sim \mathcal{CN}(0, 0.2)$, $\theta_3 = 0$.

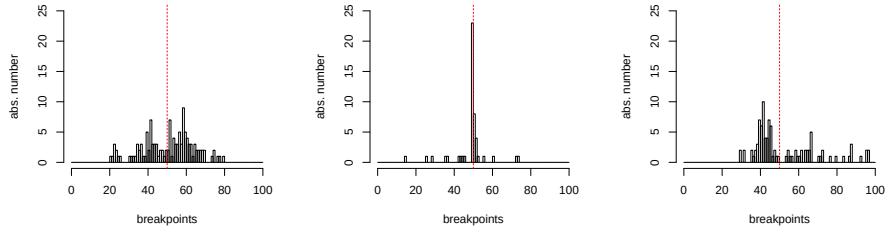
The depth based methods still show a good performance in estimating the location of the change point. Again, the variance based on one-dimensional models is lower than for two-dimensional models. The number of detected change points is one for nearly all of the simulations based on the depth methods. Multiple change points are just detected in very few simulations. The `strucchange` method is influenced by the jumps in the series introduced by the contamination. Hence, more than one change point is detected quite often, because the jumps caused by contamination are misspecified as change points. This also has an effect on the poor estimates of the change point location. The method available through the `segmented` package performs quite similar to the non-contaminated case. The true number of change points is again correct due to the parameters used for the algorithm. The variance increases due to the jumps what indicates that the contamination also influences the estimates in this approach.

Comparison for one Change Point and Fréchet Errors

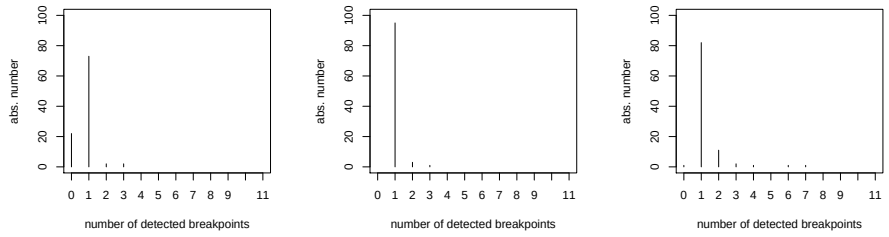
The next example, see Figure 78, shows an evaluation based on Fréchet errors.



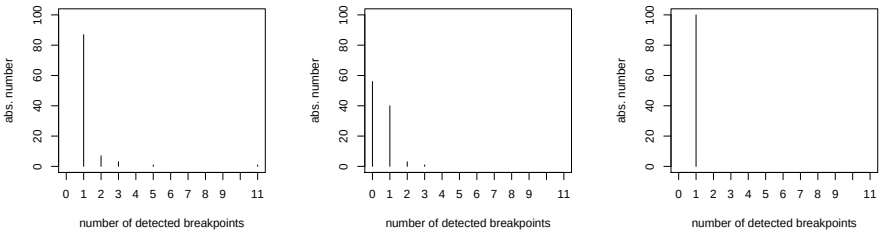
(a) 1 dim, Method 1 (b) 1 dim, Method 2 (c) 2 dim, Method 1



(d) 2 dim, Method 2 (e) strucchange (f) segmented



(g) 1 dim, Method 1 (h) 1 dim, Method 2 (i) 2 dim, Method 1



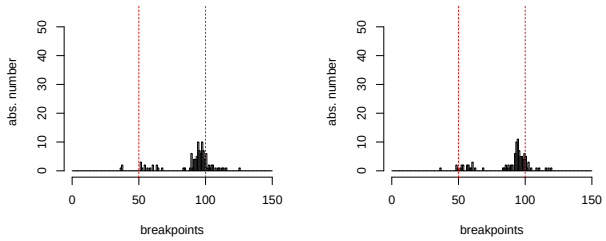
(j) 2 dim, Method 2 (k) strucchange (l) segmented

Fig. 78: Change points in $R = 100$ simulations of processes defined by $\theta_1 = 0.01$ at $n \in \{1, ..., 50\}$ and $\theta_1 = 0.02$ for $n \in \{51, ..., 100\}$, $E_n \sim \mathcal{F}(1.928, -2, 10)$, $\theta_3 = 0$.

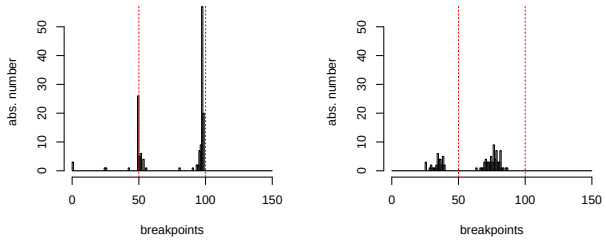
In this case, the robust methods preform very well again. The variances of the location estimates are comparable to the contaminated case. The number of detected multiple change points increases. The location of the true change point is also nearly perfectly estimated by the `strucchange` method. Unfortunately, this just happens in 40 % of the simulations. In more than 50 % no change can be detected. This ratio is remarkably lower for all of the depth methods. As expected, the `segmented` method also is detecting exactly one change point for Fréchet errors. The distribution of the estimated locations is quite similar to the simulations with normally distributed errors. In particular, it is again biased.

Comparison for Two Change Points and Normal Errors

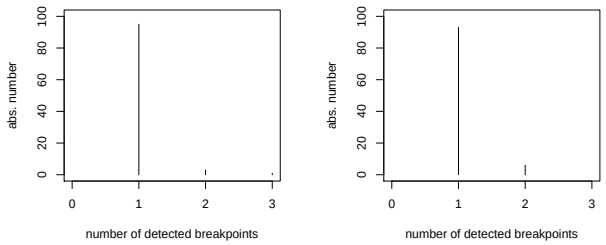
The last simulation study shows the performance of the methods, if multiple change points exist. Thereby, we reduce our attention to simulations with normal error distribution. To generate an example, we extend the process considered above by a third phase with parameters $\theta_1 = 0.01$ and $\theta_3 = 0.1$ for 50 observations. The results are depicted in Figure 79. We do not consider the one-dimensional models anymore, since we also assume a change in the intercept parameter θ_3.



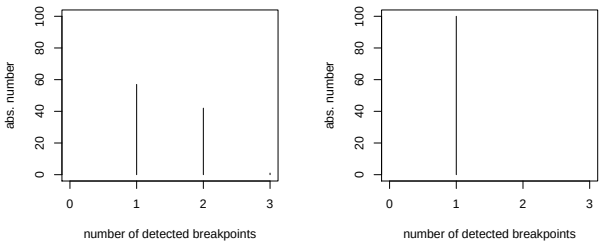
(a) Method 1 (b) Method 2



(c) strucchange (d) segmented



(e) Method 1 (f) Method 2



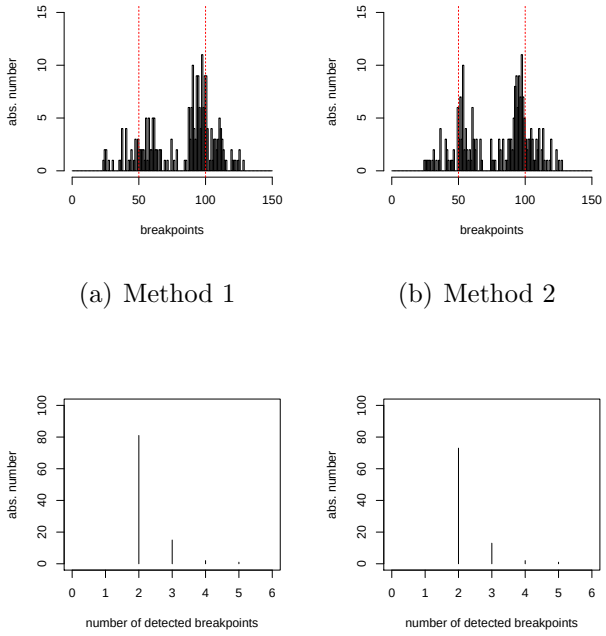
(g) strucchange (h) segmented

Fig. 79: Change points in $R = 100$ simulations processes defined by $(\theta_1, \theta_3) = (0.01, 0)$ at $n \in \{1, ..., 50\}$, $(\theta_1, \theta_3) = (0.02, 0)$ for $n \in \{51, ..., 100\}$ and $(\theta_1, \theta_3) = (0.01, 0.1)$ for $n \in \{101, ..., 150\}$, $E_n \sim \mathcal{N}(0, 0.2)$.

The `stucchange` method performs very well. It detects the locations of the change points with a high precision and further detects both changes in over 40% of the simulations. In about 55% at least one change can be detected. In few remaining simulations three changes are detected. The `segmented` method again shows biased location estimates. Further, by the preliminary guess of one change point, just one point is detected in each simulation. The depth based methods also detect the change point locations quite well, but fail to detect multiple change points automatically. In case of the depth based methods the detection fails, because the parameters are poorly tuned. Hence, in most of the simulations just one change point is detected. This in particular happens, since the length of the candidate clusters differs, what leads to neglected change points, due to our robustification step.

Parameter Tuning for Depth Methods

To illustrate the tuning problem in case of the depth methods, we have also evaluated the same processes with the information that we expect two change points. The results are depicted in Figure 80.



(a) Method 1 (b) Method 2

(c) Method 1 (d) Method 2

Fig. 80: Change points in $R = 100$ simulations of processes defined by $(\theta_1, \theta_3) = (0.01, 0)$ at $n \in \{1, ..., 50\}$, $(\theta_1, \theta_3) = (0.02, 0)$ for $n \in \{51, ..., 100\}$ and $(\theta_1, \theta_3) = (0.01, 0.1)$ for $n \in \{101, ..., 150\}$, $E_n \sim \mathcal{N}(0, 0.2)$, $m_p = 2$.

With the additional information, the change points are detected in all cases now. However, the numbers of detected change points is still quite low. The variance of our proposals is still higher than the variance of the `strucchange` results. The reason is that we detect two points in each simulation, and hence also have more observations of change point locations. The second method thereby results in slightly more additional change points but also has a higher precision with respect to the detected change point locations.
Summarising, we can state that the depth based methods deliver very flexible and comparative methods for change point detection in growth processes. The proposed methods are quite flexible, since they also can be applied to the non-linear model without any additional cost. In further work, we aim to develop a way for an optimal parameter selection. Thereby, local depth or the simplicial depth shape itself can be applied to preliminary estimate the expected number of change points. Further, a training for the bandwidth parameters has to be proposed. So far, we select the parameters based on a trade-off with respect to precision and computation time.

5.6 Runtimes

To illustrate the performance gain resulting from proposed the calculation method in Section 4.1 compared to a straightforward loop calculation, we perform a short simulation study. Therefore, we evaluate the simple algorithm to calculate the full simplicial depth for two-dimensional parameters programmed with nested loops and compare it to the proposed algorithm in case of a vectorisation without usage of parallel calculations and with parallel calculations on two cores. The study is preformed on a LENOVO L420 with Intel(R) Core(TM) i3.2310M CPU(2.10GHz), 4GB RAM, Windows 7 64bit. Thereby, we calculate the test statistic for series of length $N \in \{50, 100, 500, 1000\}$ and repeat the calculations 1000 times for each series length to estimate the average runtime. The resulting average runtimes for the evaluation of one depth statistic are presented in Table 20. It is clearly visible

N	50	100	500	1000
avg. time 1 CPU	0.049	0.20	14.67	117.30
avg. time 2 CPU	0.020	0.114	11.59	96.86
avg. time loop	0.187	1.491	181.64	n.a.

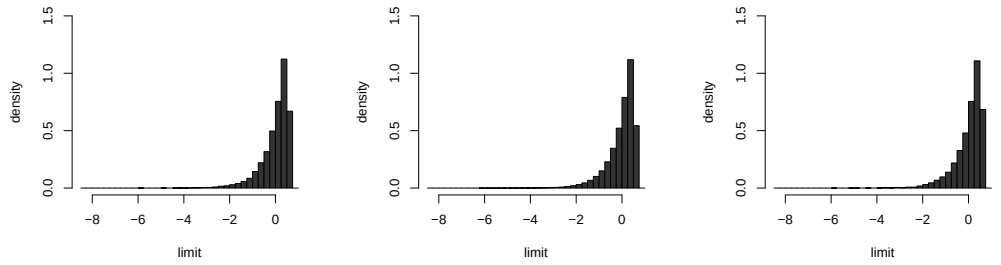
Tab. 20: Average runtime (seconds), 1000 repetitions for each scenario.

that the matrix based algorithm comes with a large increase of performance. Even for a short series of $N = 50$ observations the matrix based algorithm reduces the calculation time by 74% of the loop based method. The difference appears to be not linear in N. The runtime is reduced by 92% for a sample size of $N = 500$. The reduction based on the application of multiple cores results in a lower increase of efficiency. Of course the advantage of the matrix based calculation is clearly visible. For a sample size of $N = 50$ the computation time is reduced by 90% of the loop based algorithm and by 60% of the matrix based algorithm without multiple cores. But in case of $N = 500$ observations the reduction by parallel computing just results in 21% of the runtime from the simple matrix based method. Nevertheless, the time is reduced by 94% of the loop based algorithm. The reason for the decreasing improvement by parallel computation is a poor management of the memory in the current stage of development of our functions. So far, the residuals are exported to the utilised cores in total. Since this happens multiple times in the algorithm, depending on the sample size, the parallel method loses performance due to this operations with respect to memory usage. In further revisions of the package, we will address this issue, either by an adjusted approach to utilise the available cores in `R` directly or by `C` implementations which allow a more efficient memory management.

5.7 Limit Distribution

Validity of the Limit Distribution

To check the validity of the derived limit distribution, we present a short simulation based comparison under the null hypothesis. Therefore, we calculated $d_S^{ARi}(\theta_0, y)$ for 10000 processes of length $N = 1000$ and a parameter $\theta = (\theta_1, \theta_3) = (0.001, 0.02)$ for an linear AR(1) process with intercept to compare the empirical distribution of $N(d_S^{ARi}(\theta, y) - \frac{1}{4})$ with the approximative results from W_L, see (40). The simulated processes to calculate d_S^{ARi} were based on Normal errors and on Gumbel errors with $med(E_n) = 0$. A comparison is presented in Figure 81.



(a) Histogram of $N(d_S^{ARi}(\theta, y) - 1/4)$, for y based on Normal errors

(b) Histogram of the approximated theoretical limit distribution

(c) Histogram of $N(d_S^{ARi}(\theta, y) - 1/4)$, for y based on Gumbel errors

Fig. 81: Histograms of limit and statistic. The histograms from d_S are based on 10000 simulated statistics. The theoretical distribution is approximated by the bivariate Gaussian limit process based on 200000 simulated paths.

We see that the distributions look quite similar. Further, two sample Kolmogorov-Smirnov (KS) tests (Conover, 1971, pp. 309) show that equal distributions cannot be rejected, when we systematically test sub samples from the simulated distributions. The analysis is restricted to sub samples, since a KS test is too sensitive to small deviations in large sample comparisons. Therefore, we test 1000 sub samples, each with 50 randomly drawn simulations, against the full approximated limit distribution. The size of the sub samples is selected low enough, to allow an exact evaluation of the KS test. Then, we count the number of rejections based on an 5% level based on the KS test. The results are presented in Table 21.

Error distribution	Normal	Gumbel
Relative number of rejections	0.057	0.066

Tab. 21: Relative number of rejections of the KS test on a 5% level by comparison with the full approximated limit distribution based on 1000 repetitions.

We see that for nearly 95% of the sub samples equal distributions cannot be rejected in both comparisons. This strongly supports the validity of our limit distribution.

Reliability of the Simulation for Arbitrary Parameter Dimensions

To check the reliability of the simulation scheme for parameters with arbitrary parameter dimension K, introduced in Section 4.2, we compare quantiles of the approximated limit for the two dimensional processes with the results from a simulation based on Bernoulli variables. Therefore, we simulated 1000 values of the test statistics based on equation (41) with $N = 1000$ and $K = 2$ for each simulation. In Figure 82, we compare the histogram from the approximated asymptotic distribution with a histogram from the simulated and correctly rescaled limit based on Bernoulli variables.

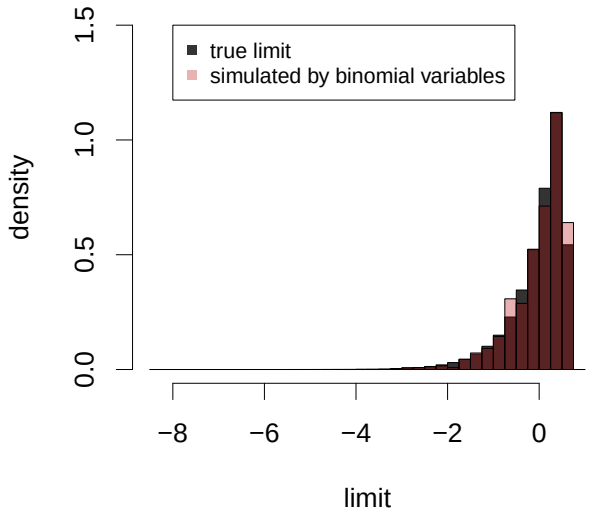


Fig. 82: Histograms of limit and simulated limit.

We see that there are some deviations of the distributions, but the general shape is quite similar. A KS test, comparing the resulting distributions, has a p-value of

0.2073. Hence, we cannot reject equal distributions here. This supports the validity of the simulation scheme.
For $K > 2$ the choice of N and the number of simulations to explore the limit distribution should be carefully checked, for example by an evaluation of the variation between sub-samples, to assure that the limit is approximated precise enough.

6 Real Data Examples

6.1 Maurer and Heeke Data - Model Choice

Technical Details and Assumptions

Maurer and Heeke (2010) conducted fatigue experiments on prestressed concrete under cyclic loading. Their main focus was to analyse the crack growth under low loading, since in this field no fundamental research results are available. One reason for this is that low loading directly leads to long experiments which are very costly. Hence, Maurer and Heeke (2010) just present a limited amount of potentially censored experimental results. We propose statistical methods which nevertheless allow inference on their data with respect to the relation of stress and lifetime.
The experimental specimen were prestressed concrete beams as presented in Figures 83 and 84. Prestressed concrete beams are solid concrete blocks, where a tension wire is incorporated to increase the stability of the beam. A typical wire consists of multiple twisted metallic wires. In our case 35 single wires were used per complete tension wire. The location of the tension wire in each beam is illustrated in Figure 84 by the dashed line.

Fig. 83: TR02 experiment of Maurer and Heeke (2010).

Thereby,material for the tension wires was won from wires of an Autobahn bridge at the BAB A1 near Hagen in Germany after its demolition. From the eleven meter tension wires, new five meter tension wires were produced and incorporated into new concrete beams. In total, five experimental beams could be produced. Details on the selection of the wires and the production can be found in Maurer and Heeke (2010).

To observe fatigue under experimental conditions a test station was designed. This test station is depicted in Figure 84.

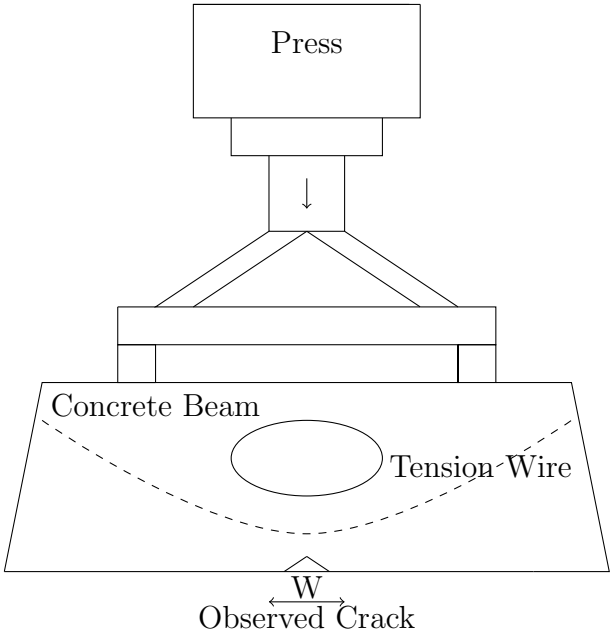
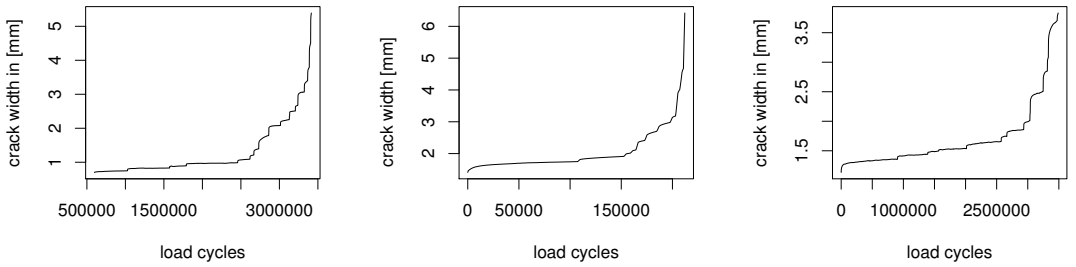


Fig. 84: Test station used for crack growth experiments.

By application of a cyclic load by the press, horizontal loads were induced to the test specimen so that a calculable load was applied to the tension wire at the crack tip. At the way recorder (see W in Figure 84) the width of this initiated crack was measured. Further, data from elongation recorders and microphones was collected. In our models, we just consider the recorded crack width data $a(t)$ at time t. Thereby, t is the time, recorded at equidistant points which are defined by the load cycles. In total, results from five experiments denoted by TR01, TR02, TR03, TR04 and TR05 are available. The experiments were conducted under different loads and technical settings. The most important parameters can be found in Table 22.

	Prestressing tension [kN]	Max. load F_{max} [kN]	Min. load F_{min} [kN]	Amplitude F_{Amp} [kN]	$\Delta\sigma$ $[N/mm^2]$	Sample Size
TR01	179.3	342	250	46	200	2806
TR02	179.3	454	250	102	455	502
TR03	234.0	477	388	44.5	200.3	7739
TR04	234.0	456	388	34	150	11748
TR05	234.0	432	388	22	98	12401

Tab. 22: Parameters of the Maurer and Heeke (2010) experiments.



(a) TR01 (b) TR02 (c) TR03

Fig. 85: Fully observed rack growth time series by Maurer and Heeke (2010).

While in experiments TR01, TR02 and TR03 the total failure of the concrete beams could be observed TR04 and TR05 are censored experiments. The duration of experiment TR01 was 29 days. TR02 just needed 2 days until total failure. For TR03 data was recorded for 26 days. The censored experiments were aborted after 42 and 89 days without a total failure.
To derive the relation between applied load and the number of load cycles until failure, including point-wise confidence intervals by the available data, we need a method to construct prediction intervals for the fully observed series and also a method to complete the censored experiments. The choice of these methods, which we introduce in Section 2, is limited by the properties of crack growth.
The fully observed series are depicted in Figure 85. Due to the experimental conditions, we can formulate central assumptions on the data generating process.

F1 Since we observe fatigue without maintenance, the long term process should be non-decreasing.

F2 Due to the incorporated tension wires, positive jumps in the process can be observed. These jumps appear, if one of the M tension wires, whereby M is fixed and known, breaks.

F3 Since we work under experimental conditions, the starting value y_0 is fixed and known.

F4 There are observation and measurement errors with median zero.

These assumptions induce that the experiments always result in growth processes, due to the fact F1. Since the observed processes are very long, compared to a low

number of tension wires and since we do not have experiments under repeated conditions, we are not able to estimate the frequency and height of the jumps introduced by tension wires, mentioned in F2. Hence, we combine the errors defined by F4 with the jumps which are just apparent for less than 2.5 % of the data for TR02 and less than 1% for TR01 and TR03. Therefore, we assume that we observe a skewed error distribution with a median equal to zero. Fact F3 allows us to apply our theory based on non-random starting values. Note that with random starting values, some of the arguments which guarantee growth processes need to be carefully checked. Our main aim is to derive a S-N curve, relating the applied loads $\Delta\sigma$, or S for stress in engineering, to the lifetime in load cycles, denoted by N.

Discussion of the Full Three Parameter Model

Since the complete model was defined by the three parameter discrete approximation of the stochastic Paris-Law, we first analyse the crack growth data for the full model given by

$$Y_n = \theta_3 + Y_{n-1} + \theta_1 Y_{n-1}^{\theta_2} + E_n, \tag{51}$$

with $med(E_n) = 0$. By the derivatives with respect to the parameters, we can calculate tangential depth. The derivatives for the usual quality function given by the squared residuals

$$Q(\theta, z) = (y_n - y_{n-1} - \theta_1 y_{n-1}^{\theta_2} - \theta_3)^2$$

are given by

$$\begin{aligned}
\frac{\partial Q(\theta, z)}{\partial \theta_1} &= -2y_{n-1}^{\theta_2} r_n(\theta, y), \\
\frac{\partial Q(\theta, z)}{\partial \theta_2} &= -2\theta_1 y_{n-1}^{\theta_2} log(y_{n-1}) r_n(\theta, y), \\
\frac{\partial Q(\theta, z)}{\partial \theta_3} &= -2r_n(\theta, y).
\end{aligned}$$

Hence, tangential depth is

$$d_T^{ARc}(\theta, z) = \min_{\|u\|=1, u\in\mathbb{R}^3} \# \left\{ n | (u_1, u_2, u_3) \cdot (y_{n-1}^{\theta_2}, \theta_1 y_{n-1}^{\theta_2} log(y_{n-1}), 1)^\top \cdot r_n(\theta, z) \geq 0 \right\}. \tag{52}$$

If we denote

$$w_u(y)r_n(\theta, y) = u^\top \nu(y) r_n(\theta, y)$$

with $\nu(y) = (y_{n-1}^{\theta_2}, \theta_1 y_{n-1}^{\theta_2} log(y_{n-1}), 1)^\top$, we have to analyse the roots of $w_u(y)$ to allow a simplification of the full simplicial depth to sign changes of residuals by Theorem 21. In particular, the roots of

$$\begin{aligned} w_u(x) &= u_1 x^{\theta_2} + u_2 \theta_1 x^{\theta_2} log(x) + u_3 \\ &= (u_1 + u_2 \theta_1 log(x)) x^{\theta_2} + u_3 \end{aligned} \tag{53}$$

have to be checked. This can be done as in Kustosz et al. (2016b). Therefore we apply the following Lemma.

Lemma 102. *If $\theta_1 \neq 0, \theta_2 \neq 0$, then $w_u : [0, \infty) \to \mathbb{R}$ given by (53) has the following properties:*
a) w_u has exactly one extremum at $x = exp\left(-\frac{1}{\theta_2} - \frac{u_2}{u_3 \theta_1}\right)$ for all $u = (u_1, u_2, u_3)^\top \in \mathbb{R}^3$ with $u_3 \neq 0$.
b) For all $0 < \xi_1 < \xi_2$, there exists a vector $u_+ \in \mathbb{R}^3$ with $w_{u_+}(\xi_1) = w_{u_+}(\xi_2) = 0$ and $w_{u_+}(x) > 0$ for all $x \in (\xi_1, \xi_2)$ and a vector $u_- \in \mathbb{R}^3$ with $w_{u_-}(\xi_1) = w_{u_-}(\xi_2) = 0$ and $w_{u_-}(x) < 0$ for all $x \in (\xi_1, \xi_2)$.

Proof. The proof is given in the Examples section of Kustosz et al. (2016b). □

Lemma 102 proves that the conditions of Theorem 21 are satisfied. This simplifies the simplicial depth statistic.

Corollary 103. *For model (51) and $y_n > y_{n-1} \forall n$ simplicial depth can be simplified to*

$$\begin{aligned} &d_S(\theta, z_*) = \\ &\frac{1}{\binom{N}{4}} \sum_{1 \leq n_1 < n_2 < \ldots < n_4 \leq N} \Bigg(\mathbb{1}\{r_{n_1}(\theta, z) > 0, r_{n_2}(\theta, z) < 0, r_{n_3}(\theta, z) > 0, r_{n_4}(\theta, z) < 0\} \\ &\qquad + \mathbb{1}\{r_{n_1}(\theta, z) < 0, r_{n_2}(\theta, z) > 0, r_{n_3}(\theta, z) < 0, r_{n_4}(\theta, z) > 0\} \\ &\qquad + 1 - \prod_{k=1}^{4} \mathbb{1}\{r_{n_k}(\theta, z) \neq 0\} \Bigg), \end{aligned}$$

if $\theta_1, \theta_2 \neq 0$.

Proof. See section 5.4. in Kustosz et al. (2016b). □

This corollary allows us to apply the full simplicial depth and the simplified depth notions to the extended model for $K = 3$. Since we know the limit distributions for our statistics in this case, all proposed methods based on the simplified depth notions can be applied. The consistency can be simply checked as in the example section presented in Kustosz et al. (2016b). An important restriction in of the evaluation of the method is that parameters leading to stationary processes with sign changes have to be excluded. This, in particular, means that for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, the parameter range has to be restricted to $\Theta_0 \cup \Theta_1 = \Theta_a$, whereby Θ_a is the set of all parameters which guarantee explosive or at least non-negative processes. Since at least the selection of θ_3 thereby directly depends on the error distribution, Θ_a cannot be determined in general, if we assume that the error distribution is unknown. The resulting parameters and confidence sets should be checked carefully to exclude such solutions. A simple possibility to restrict the parameters is to set $\Theta_a = (0, \infty) \times (0, \infty) \times (0, \infty)$. Note that this excludes possible parameters, when for example $\theta_3 < 0$ is overcompensated by θ_1, θ_2 large enough for $y_0 > 0$ and appropriate errors. Here, we set $\Theta_a = (0, \infty) \times (0, \infty) \times (a, \infty)$ for simulations, allowing $a < 0$, as reasonable space for θ and advise to check the results for violations of the growth assumption. The main problem of this model for application is the influence of θ_2. If θ_2 is close to zero the effect of the autoregressive parameter reduces to a constant. This follows from the definition of the residuals

$$r_n(\theta, y) = y_n - y_{n-1} - \theta_1 y_{n-1}^{\theta_2} - \theta_3.$$

Applied to data for which the differences alternate around a constant value, we get a positive value of depth if

$$r_n((\theta_1, 0, \theta_3)^\top, y) = y_n - y_{n-1} - \theta_1 - \theta_3$$

alternate. Obviously, this solution is not identifiable for (θ_1, θ_3). Unfortunately, this happens for the Maurer and Heeke (2010) data. Hence, the resulting depth based confidence sets are unbounded, if the full model is applied.
To illustrate this problem, we evaluate d_S^1 and d_S^2. In the application, we set $a = -1$ and calculate depth on a grid defined by $[0, 1] \times [0.1, 2.5] \times [-1, 1]$ with steps of 0.01 in each interval. This defines 4.892.541 grid points on which we evaluate the simplified depth statistics. Another necessary modification is caused by the inhomogeneity of the data. Due to the low variance of the time series and due to inhomogeneous parameters, the resulting confidence intervals have to be computed on very high

levels to contain any parameters. To avoid such problems, we add random noise to the data to introduce slightly more variance. This overrides small changes in the parameters and thereby connects phases of slightly differing parameters. Due to the increase of the general depth level, then reliable and non-trivial confidence sets can be constructed at reasonable levels. But however, parameters in the confidence sets for the full model lead to unit root processes. This means that no information on the actual growth can be won.
This major problem is depicted in Figure 86.

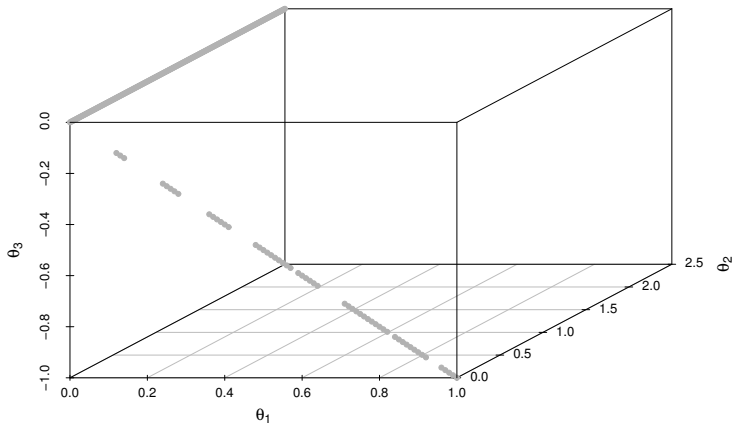


Fig. 86: $1 - 1 \cdot 10^{-14}$ parameter confidence set for TR01 series based on d_S^1 and the three parameter model.

Here, we see the $(1 - 1 \cdot 10^{-14})$ confidence set for the TR01 series based on all observations beginning at index 431 and ending at index 2517. Obviously, the resulting parameters either satisfy $\theta_1 = 0$ or $\theta_2 = 0$, what results in pure white noise processes. For lower levels no parameters are included in the confidence set.
A simple noise generation by addition of $\mathcal{N}(0, 0.001)$ errors leads to larger confidence sets. Now, we can evaluate the 5% level. However, the unit root processes are included in the confidence sets, but now also non-trivial solutions can be found. In particular, the depth maximising parameters on the considered grid now are located on a curve and are non-trivial, see Figure 87.
By additional analysis, we see that the resulting parameters in the 95% confidence set are results from robust fits with respect to the residuals, what can be seen in Figure 88. Here, subsequent observations y_n against y_{n-1} are depicted. Further, the fit via the autoregressive relation based on the parameters with maximal depth (red lines) and from parameters in the 95% confidence set (grey lines) are shown. All fits from the confidence interval are not affected by the outliers. The green line shows

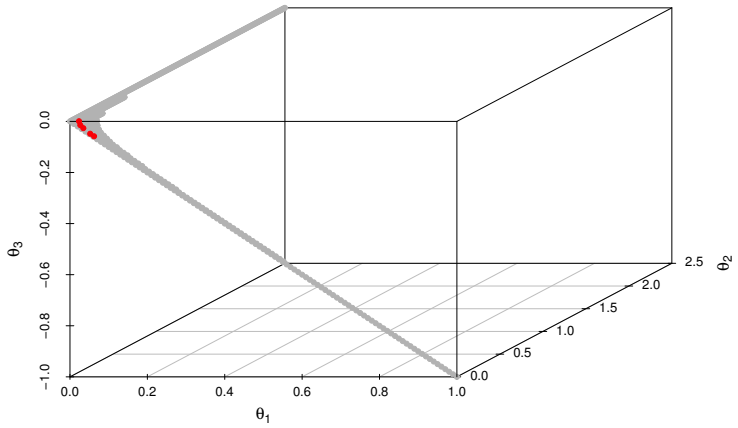


Fig. 87: 95% parameter confidence set for TR01 series with noise based on d_S^1 and the three parameter model.

the fit of a least squares estimator for the three parameter model which is more affected by the jumps in the process. Similar results can be observed for TR02 and TR03 and both considered statistics. The results are presented in the Appendix A. In case of d_S^2 the intervals are slightly larger and the influence by outliers seems to be more pronounced.

Summarising, we see that the usage of a unknown θ_2 leads to identification problems, since a parameter value close to zero implies variable solutions for (θ_1, θ_3) with constant and positive depth. Unfortunately, the calculated confidence sets include a continuous transition to this parameter region. Hence, the resulting confidence sets are unbounded and imply results which are random walks with drift expressed in a solution which is not unique. These random walks then do not include any information about the process, except for the drift level. The remaining information is transferred to the unknown errors with median zero.

Hence, we cannot apply the full model to the Maurer and Heeke (2010) data.

Application Based on a Two Parameter Model

To resolve this identification problem, we reduce the model. To allow an influence of θ_2, we include a constant in the new model, but assume it to be fixed and known. This reduces the model to a variation defined by

$$Y_n = Y_{n-1} + \theta_1 Y_{n-1}^c + \theta_3 + E_n,$$

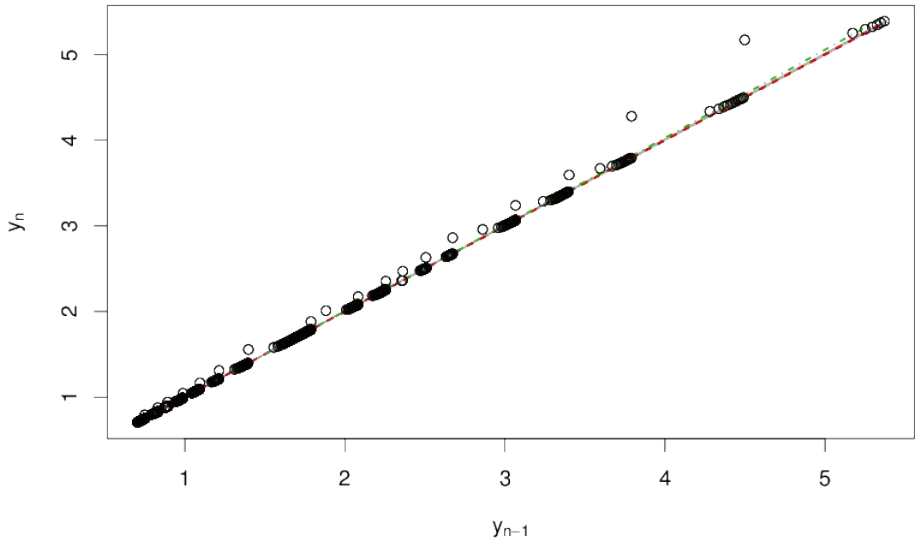


Fig. 88: Fits from the 95% parameter confidence set for TR01 series with noise based on d_S^1 and the three parameter model. The unfilled dots are the true observations (y_{n-1}, y_n). The parameters which maximise depth define the red dashed lines. The green line is a fit defined by the OLS estimate for the three parameter model.

under the same assumptions, as used for the two parameter autoregressive models in Section 2.3. By usage of the quality function

$$Q(\theta, z) = (y_n - y_{n-1} - \theta_1 y_{n-1}^c - \theta_3)^2 =: r_n(\theta, y)^2$$

and its derivatives

$$\frac{\partial Q(\theta, z)}{\partial \theta_1} = -2y_{n-1}^c r_n(\theta, y), \frac{\partial Q(\theta, z)}{\partial \theta_1} = -2r_n(\theta, y),$$

it directly follows that for $y_n > 0$ for all n simplicial depth, based on tangential depth, reduces to the two parameter simplicial depth

$$d_S^{ARcr}(\theta, y)$$
$$= \frac{1}{\binom{N}{3}} \sum_{1 \le n_1 < n_2 < n_3 \le N} \mathbb{1}_{\{r_{n_1}(\theta,y)>0, r_{n_2}(\theta,y)<0, r_{n_3}(\theta,y)>0\}} + \mathbb{1}_{\{r_{n_1}(\theta,y)<0, r_{n_2}(\theta,y)>0, r_{n_3}(\theta,y)<0\}},$$

if we again neglect the terms resulting from zero residuals. Hence, the limit theorems for the full and simplified depth in the two parameter case can be applied, if the residuals are calculated based on the appropriate model equation.

In the final application, we assume that $c = 1$ holds and analyse the data in the linear AR(1) setting with and without intercept, to allow prediction based on the algorithm proposed in Section 4.8.

6.2 Heuristic Derivation of a S-N Curve

Main Research Questions

Our modelling approach can be applied to analyse questions considering the available crack growth time series. Thereby, we assume that the Paris-Erdogan equation holds and that the processes can be described by our models. We want to answer the following questions.

Q1 Are the parameters constant over time?

Q2 How are the parameters changing over the experiments due to the non identical experimental settings?

Q3 Is it possible to predict future values of the processes and is it possible to continue the censored processes?

Q4 Can we construct a S-N curve relating lifetime and stress?

Summary of the Analysis Approach

We use the results from Sections 2 and 4 to propose a heuristic derivation of a S-N curve for prestressed concrete. On the way, we answer the first questions raised above. Note that the results cannot be interpreted as reliable derivation of this curve, since the number of available experiments is very low, but nevertheless the method can be applied when more data is available to deliver reliable results.
Based on our recent arguments, we base the derivation on the linear processes, since they allow the analysis of change points and prediction for the observed series. The censored series will be used to approximate the time of failure under low loading by the available methods.
The general idea is structured as follows.
First, we analyse the available data with respect to potential parameter changes in our model. Therefore, we apply the linear one and two parameter models. Under the assumption that the processes are homogeneous between these changes, we then analyse critical crack widths dependent on the actual loading in each experiment. Thereby, we will be able to detect in which phases the censored experiments are at

the censoring times. A simple model then allows us to derive critical crack widths for the phase changes of the censored series. To allow prediction for the censored series, we propose simple models to extrapolate the jump heights and the jump frequencies.
Finally, we apply the prediction methods based on simplicial depth to generate prediction intervals for the uncensored series and a modified version which relies on the extrapolated parameters for the censored series.
This will define a Wöhler curve with point-wise prediction intervals.

Application of the Methods

In Figure 89, we see the five observed series. The series TR01, TR02 and TR03 are not censored, while TR04 and TR05 are. Thereby, we declare a series as censored, if it does not reach a crack width of 3 mm which we define a total failure here. To analyse the data on an equal scale, the step widths are adjusted to be identical for all series. Therefore, we have 570 observations for TR01, 70 for TR02, 1241 for TR03, 929 for TR04 and 1994 for TR05 now. Since at the beginning of each experiment

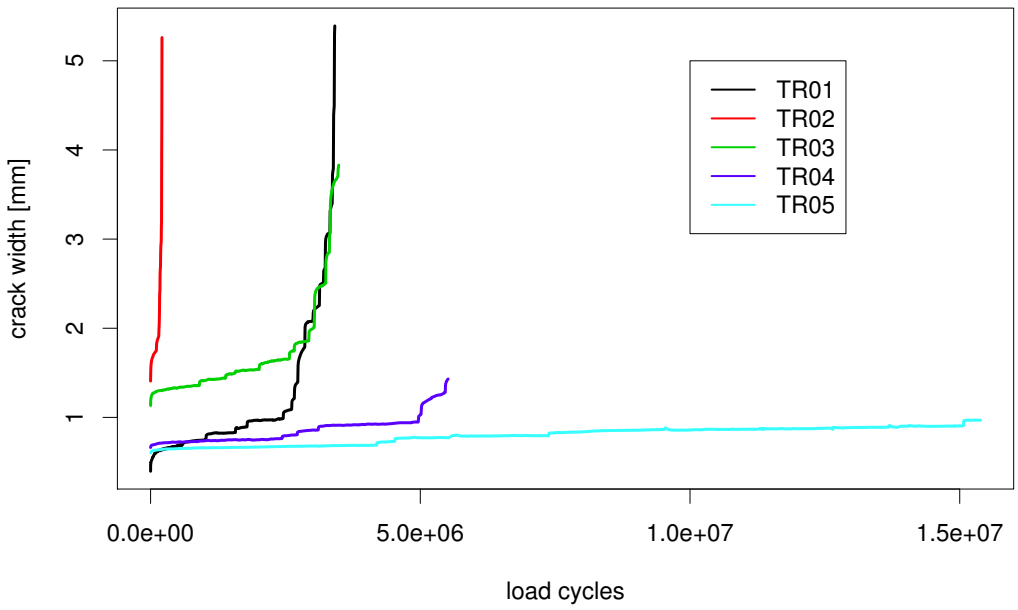


Fig. 89: Maurer and Heeke experiments.

the crack growth is unstable we distinguish a first initialisation phase which needs to be detected by visual inspection and is given by the engineers and a supercritical

phase where the crack explodes and which we want to detect by our algorithms. To do this, we apply the one and two dimensional change point detection methods to the series TR01, TR02 and TR03. The results are depicted in Figure 90. Since the end of the first phases, the initialisation, is too close to the beginning of the series, we have to define it by hand for TR01 and TR02. In case of TR03 it is auto-detected. The second phase is set to the first points, where all of the methods detect changes consistently, since the results are quite unstable. Thereby, we get the results given in Table 23.

Experiment	Change Point 1	Change Point 2
TR01	31	484
TR02	10	55
TR03	51	944

Tab. 23: Change points in full experiments.

The crack widths at the change point observation indexes are the critical values for the phase changes in our simulations. To get critical values for TR04 and TR05, we define a simple model by

$$cw_{cp1,TR_i} = \alpha_1 \cdot \Delta\sigma_{TR_i} + \alpha_2 \cdot \sigma_{max,TR_i} + E_n,$$

whereby cw_{cp1,TR_i} is the crack width at the first change point and σ_{max,TR_i} and $\Delta\sigma_{TR_i}$ are the respective maximal loads and load ranges for the experiments $TR0i, i \in \{1,2,3,4,5\}$. We get a reasonable fit by the linear model, as shown in Figure 91. A similar model for the second change point can be defined by

$$cw_{cp2} = \alpha_1 \cdot \Delta\sigma_{TR_i} + \alpha_2 \cdot cw_{cp2,TR_i} + E_n$$

with a fit shown in Figure 92. Applied to the loads of TR04 and TR05 we get the values for the critical crack widths as presented in Table 24.

Experiment	Width at Change Point 1	Width at Change Point 2
TR04	0.863	1.501
TR05	0.111	1.101

Tab. 24: Change points in censored experiments.

Hence, we can assume that both censored experiments are still in the first phase, but beyond initialisation, since TR04 is censored at a crack width of 1.433 mm and TR05 is censored a width of 0.972 mm. In principal, prediction intervals for the full series already can be constructed based on the available methods and information.

But however, for TR04 and TR05, we neither can bootstrap from residuals in the second phase, since it is not observed, nor we can estimate the parameters in this phase. Hence, we use a heuristic extrapolation based on the results for TR01, TR02 and TR03 series.
Comparing the empirical simplicial depth shapes in the one parameter case for the three series, restricted on the second and third phase, in Figure 93 we see that the main difference between the phases is a shift of the maximal depth. Further, depth gets a bit wider. Here, this widening will be neglected for simplicity. The shift can be modelled by a linear model. We define it by

$$max2dS_{TR_i} = \alpha_1 \cdot max1dS_{TR_i} + \alpha_0 + E_n.$$

Since we can calculate the simplicial depth shape for TR04 and TR05 in the first phase, an extrapolated empirical depth can be defined by shifting the phase one results as suggested by the linear model. The fit is shown in Figure 94.

We also have to model the jumps. To allow a non-parametric usage of the errors, we assume that in each phase the occurrence of jumps is triggered by a homogeneous Poisson process. Such a process is defined as follows.

Definition 104. *A stochastic process $(N_t)_{t\geq 0}$ is called **Poisson process** with intensity parameter $\lambda \in \mathbb{R}_+$, if*
(i) $\mathbb{P}(N_0 = 0) = 1$
(ii) $(N_{t_i} - N_{t_j})_{i,j\in\mathbb{N}}, 1 \leq j < i \leq N$ *independent for each index set* $0 < t_1 < ... < t_N$
(iii) $N_t - N_s \sim Pois(\lambda(t-s))$.

For more details on Poisson processes see Ross (1996, pp. 59).
In our application, we assume that the parameters change when we switch the phase. By a linear model given by

$$\lambda_{2,TR_i} = \alpha_1 \lambda_{1,TR_i} + \alpha_0 + E_n$$

the parameters of the Poisson process can be extrapolated based on observed jumps in the first phase. Again, the fit is quite reliable, when the model is applied to TR01, TR02 and TR03, see Figure 95. We now want to use the residuals from the first phase for a bootstrap procedure in the second phase. Therefore, we estimate a scaling parameter which is defined by the change in the mean of the increments between two phases. This allows us to rescale the jump heights in the simulations of

the censored series. The model for the scaling parameters is defined on a log scale by

$$scale_{2,TR_i} = \alpha_1 ln(scale_{1,TR_i}) + \alpha_0 + E_n.$$

The resulting fit is depicted in Figure 96. We then inflate the bootstrapping residuals for the censored series by a multiplication with this scaling parameter. Applied to TR04 and TR05, we get scaling and jump parameters for the second phases.
With these parameters, we can predict the processes TR04 and TR05. By the simple prediction method, we can also generate prediction intervals for TR01, TR02 and TR03.
To calculate prediction intervals at a 95% level, based on a three step procedure, we adjust the nominal levels to $\alpha_n = 0.01695$. For the completely observed series, we simulate each process starting at a known value up to the critical value for the first phase and then starting from the beginning of the second phase up to the critical value of 3. This gives us two independent empirical distributions for the arrival at the second phase and the arrival at the point, where the test concrete beam fails. These distributions are based on $(1 - \alpha)$ prediction intervals. By drawing from the intervals independently and combining the times we get an empirical interval for the time of failure. Therefore, for each interval 1000 Paths are generated. Then, for each combination, 100.000 Samples are constructed.
The results, measured in load cycles are presented in Table 25.

Experiment	Lower Bound	Upper Bound	Median	True Value
TR01	363	917	604	599
TR02	42	90	66	75
TR03	819	1536	1182	1227

Tab. 25: 95 % prediction intervals for the fully observed series.

The empirical densities for prediction are presented in Figure 97. Since we use a bootstrapping scheme to generate processes for prediction, we also can analyse the empirical distributions defined by the arrival times at the critical crack sizes. We see that the predictive densities are lightly skewed but show a similar shape for TR01, TR02 and TR03. To allow prediction for TR04 and TR05 we use the extrapolation parameters to rescale the residuals and to generate the jumps by simulated processes. In Figure 98, we see some of the simulated paths. These resulting arrival densities are given in Figure 99. While the prediction for TR04 resembles the results for TR01 to TR03, the results for TR05 show heavier tails. This is caused

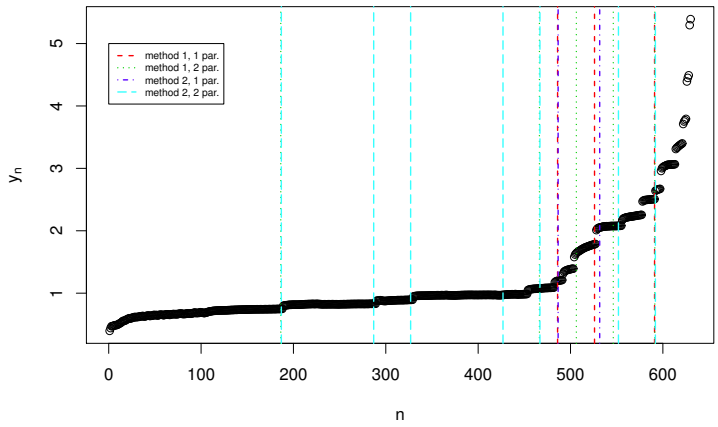
by a higher probability of very slow growth. Nevertheless, we still have explosive processes here, so that a time of failure and a prediction interval can be computed. The results are given in Table 26.

Experiment	Lower Bound	Upper Bound	Median
TR04	767	1520	1133
TR05	3903	22702	7334

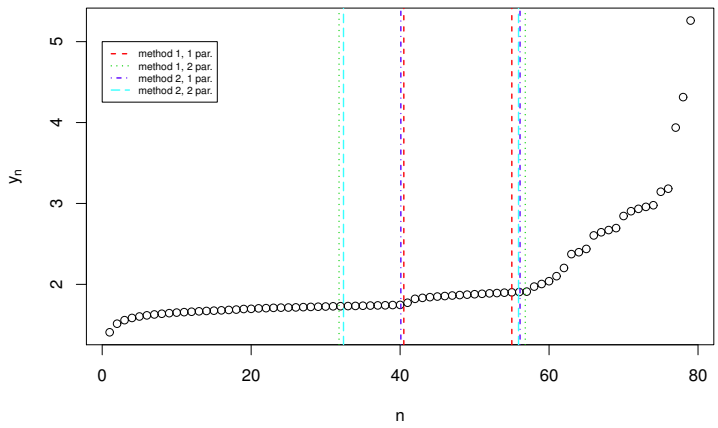
Tab. 26: 95 % prediction intervals for the censored series.

Summarising, we now have prediction intervals for all series and can construct the Wöhler curve by plotting the load cycles on a logarithmic scale with base 10 against the load amplitude $\Delta\sigma$. The curve is presented in Figure 100. The resulting S-N curve has some interesting implications. The results for the completely observed experiments can be connected by a line, as expected. The median failure times for TR01 and TR03, which were performed under similar load amplitudes, but different maximal load levels, are comparable. The only difference is that the prediction intervals for TR03 are smaller, what is caused by the higher applied maximal and minimal loads. The forecast interval for TR04 is located on a line continuing the connection of the observed experiments. This induces that TR04 is following the same fatigue characteristics as TR01 to TR03. The result for TR05 is more surprising. Here, the connection to TR04 is a line with a lower slope parameter. This implies that the fatigue characteristics lead to a significantly slower degradation than in the remaining experiments. The upper bounds for TR05 are located at over 8.000.000 load cycles, what coincides with a time of 277 days. However, the data does not allow a valid prognosis for lower loads, since we use data from the first phase to extrapolate the unobserved parts of the crack growth series.
We now can give answers to the main questions.
By the change point detection methods and the analysis of the depth shapes, we know that the parameters are not constant over time in the individual experiments. Further, by the depth shapes and the jump frequencies, we can claim that the parameters are changing, depending on the applied loads. Hence, they differ due to non identical experimental settings. By bootstrapping and extrapolated parameters, it is possible to continue the processes in a consistent way. Thereby, a S-N curve can be constructed.
The central result is that with respect to the available data, we cannot verify the existence of a long term reliability bound for prestressed concrete, but observe a decline in the fatigue characteristics under low loading, leading to higher reliability than expected by a linear relation. This probably extends to an asymptote for

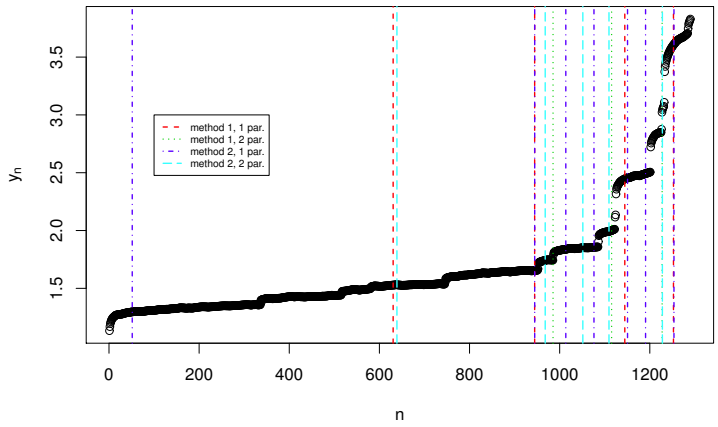
$\Delta\sigma > 0$, when lower loadings are applied. To verify this assumptions experiments under lower loads remain necessary. A possible application of our result could be a proposal for designs of experiments with upper bounds based on our confidence intervals.



(a) TR01



(b) TR02



(c) TR03

Fig. 90: Change points in full series.

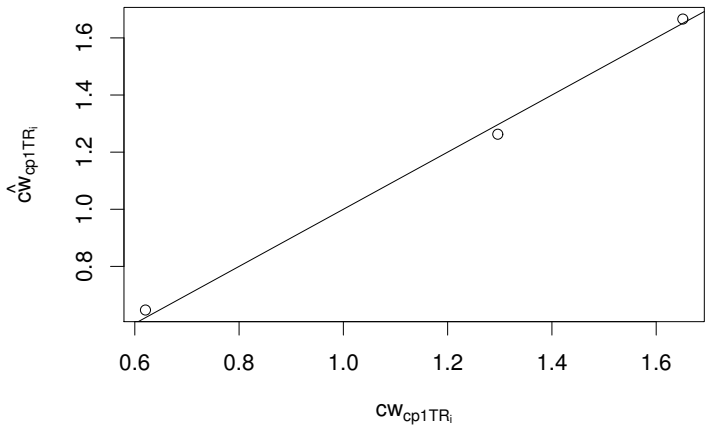


Fig. 91: Model for first change point. The black line starts at 0 and has slope 1.

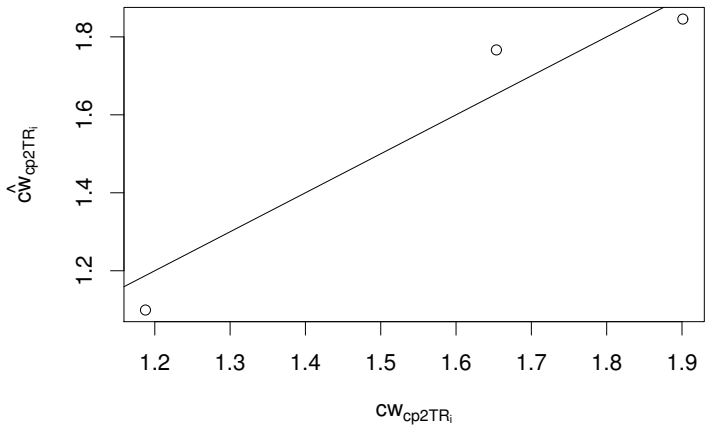


Fig. 92: Model for second change point. The black line starts at 0 and has slope 1.

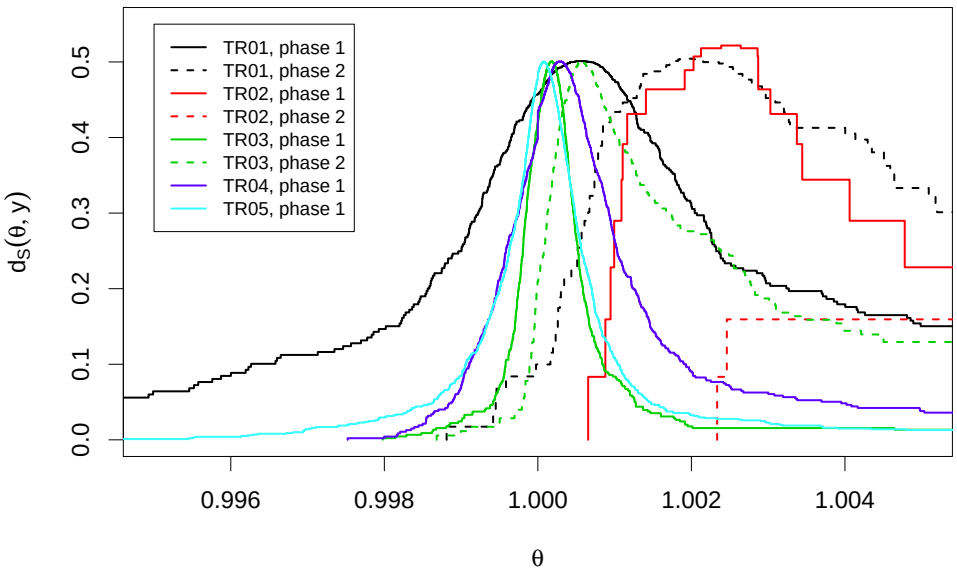


Fig. 93: Empirical depth for the Maurer and Heeke series divided by phases. The solid lines are the curves for the first phase. The dashed lines represent the depth in the second phase.

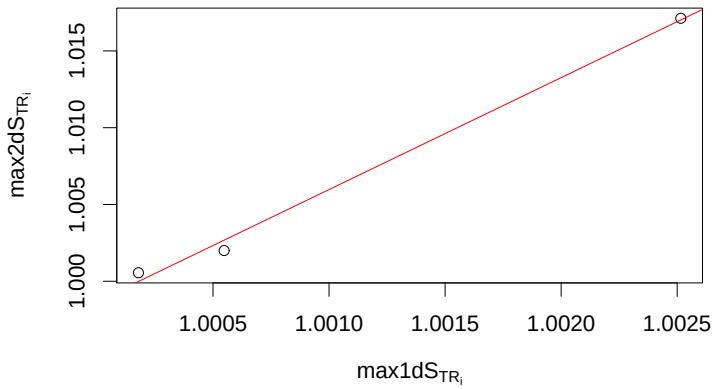


Fig. 94: Model for maximal depth in the second phase. The red line represents the fitted model.

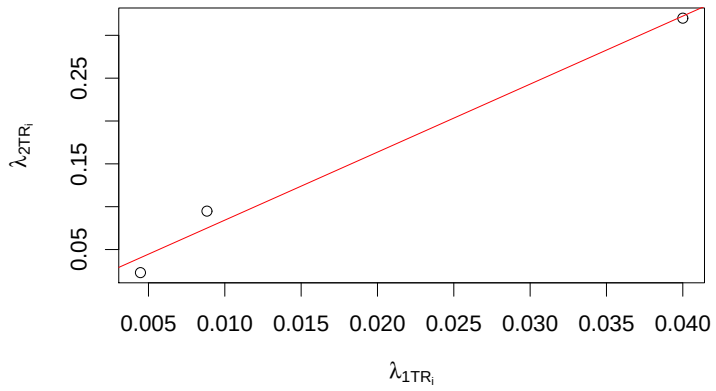


Fig. 95: Model for jump intensity in the second phase. The red line represents the fitted model.

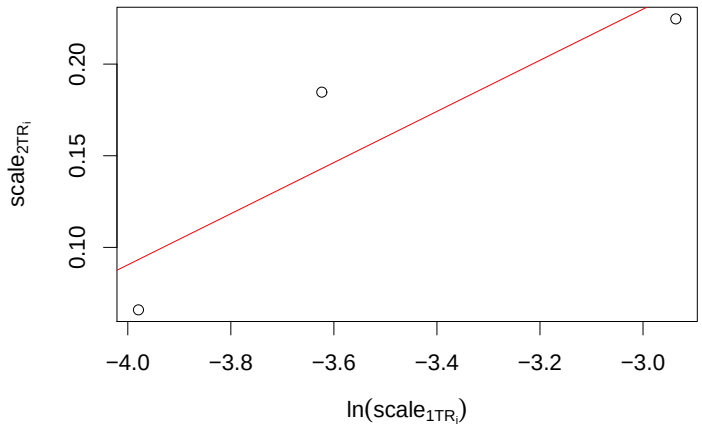
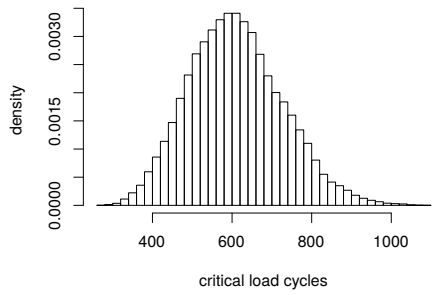
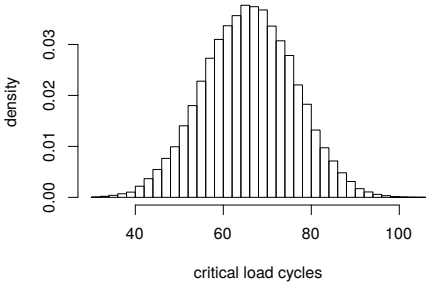


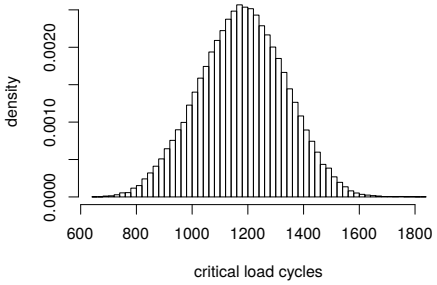
Fig. 96: Model for jump hight scaling in the second phase. The red line represents the fitted model.



(a) TR01

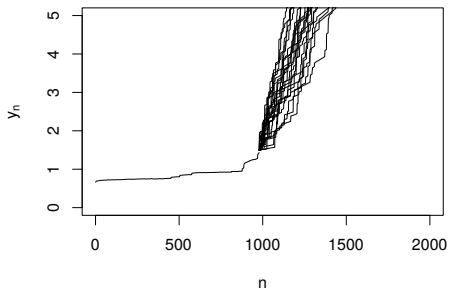


(b) TR02

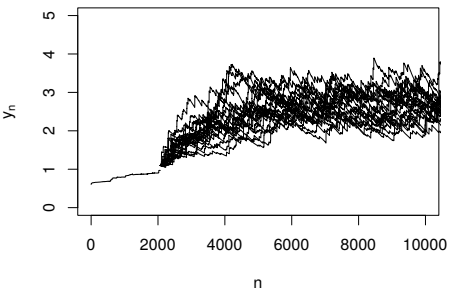


(c) TR03

Fig. 97: Empirical densities to predict failure of the fully observed series.

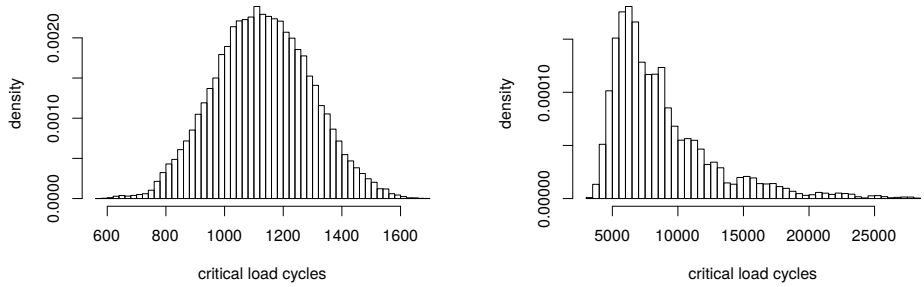


(a) TR04



(b) TR05

Fig. 98: Example continuations of TR04 and TR05.



(a) TR04 (b) TR05

Fig. 99: Empirical densities to predict failure of the censored series.

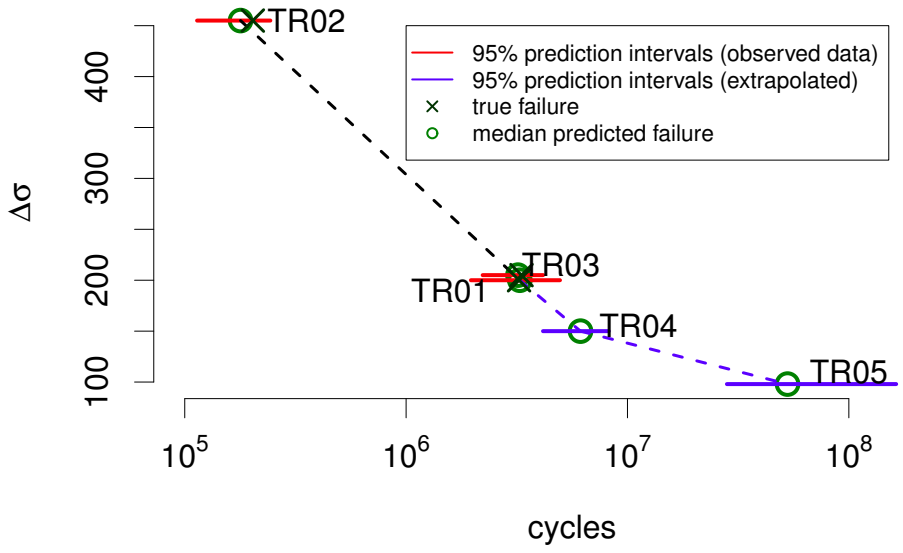


Fig. 100: S-N curve for the prestressed concrete series.

6.3 Change Points in Oil Prices

In this section, we apply our model to detect parameter changes in an oil price series.

Modelling Assumption

Starting with the fact that oil is a non-regenerative resource, the price of oil can be modelled by rules introduced by Hotelling (1931). These rules imply that the mean price follows an exponential model with a positive recovery rate. Simple calculations yield that a discretely observed oil price process can be modelled by an autoregressive process. Due to the positive recovery rates, this process should be explosive. Hence, we are able to apply our tests and estimators to detect changes in the parameter of such an explosive oil price process robustly.
The main argument in Hotellings' theory on the price path of non-regenerative resources is that the owner of a good, in this case oil, has to decide how to sell the good which he as in a storage to maximise earnings over time. If he sells the good to fast, he will be out of it in times of high demand and low supply, and hence cannot profit of high prices. If he sells his good to slow, he can run into unsold amounts of the good when the price reaches a boundary, where a substitute comes into the market.
A central assumption is that the owner of the good aims to maximise the future profit. Thereby, he has to take the interest rate δ into account. One unit of the good is discounted to its present value by $e^{-\delta t}$. The theory of Hotelling shows that under a quite simple model the optimal price path is given by

$$\frac{dp_t}{dt} = c \cdot p_t,$$

what can be translated to a discrete model via

$$p_{t+1} = \theta_1 p_t + \theta_3 + E_{t+1}$$

for the oil price.

Application to Brent Oil Prices

This allows us to apply our methods for linear autoregressive processes to oil prices, if we assume that we have a growth process. Note that theoretically this assumption should hold, if the rule is true. Empirically, the oil prices also show periods of

decrease, what contradicts this assumption. But as we at least can assume a strictly positive process with exponential long term growth, we apply our phase change detection method here, to analyse the price path for structural changes. We already pointed out that close to the unit root slightly decreasing processes can also be treated by our models.
The application of the change point detection for a linear AR(1) model with intercept is depicted in Figure 101. Both methods were applied to daily spot prices in Euro for European Brent Oil available for all trading days from 20th of May 1987 to the 1st of June 2015. Hence, we apply our methods on a series of 7166 daily prices. Since no prices are available on non trading days, the AR(1) process was defined by indices on the trading days only. To apply the methods, we selected a bandwidth of 30 days and evaluated the prices on each trading day. Change point clusters below the length of 50 % of the largest cluster were excluded to achieve robust results. For many of the changes it is difficult to relate them with crucial dates, since the

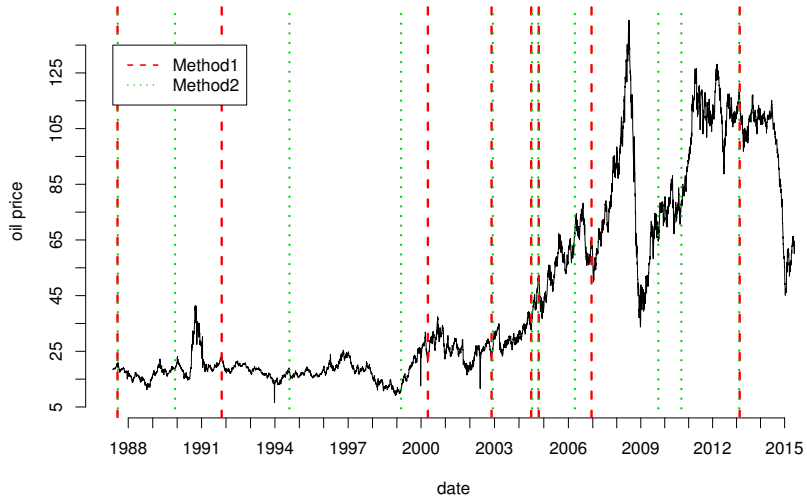


Fig. 101: Change points in the Brent Oil Prices detected by simplicial depth statistics for linear AR(1) processes with intercept. Source: U.S. Energy Information Administration (Dec 2015).

oil market does not always react suddenly to political tensions or specific events. In addition, the reasons for oil price changes are of different kinds, as political tensions, nature events, demand and supply changes and many more. Nevertheless, the changes often appear in phases which can be related to some events. The figure shows that the detected change points in fact divide the process with respect to falling and rising phases quite well. This indicates that with 30 day windows the

phases could have been also detected by our methods. We now give some reasons for the detected changes.
The first jointly detected change is in July 1987. Here, the price behaviour in fact changes due to the end of the role of Saudi Arabia as swing producer of oil supply. The next joint change point is detected in 2002. The change of the process can be explained by a change form a stagnation due to the 9/11 Attacks to an explosive behaviour caused by speculations. The next two changes appear in the ascending process in 2004 and identify a phase of low spare capacity. The last change 2013 can be explained by an significant increase in the production capacity of the US oil supply and a general stabilisation of the oil market.
The remaining points were just detected by one of the proposed methods. Thereby the second method detects changes in the slope parameter, as 1999 or 1994, quite well. These changes are affected by political tensions in Iraq and the end of the Asian financial crisis. Further, the changes of the price dynamics in the Sub-Prime Crisis after 2009 are detected by the second method. The first method also detects the most prominent of these changes but at slightly different points in time. In general, the behaviour of the prices show significant changes in the autoregressive relationship switching from explosive to random walk or even decreasing behaviour remarkably often. Due to the good performance close to the unit root, our change point detection might be quite interesting for application in finance, since prices often show such mildly explosive behaviour, if for example non-regenerative goods are considered.

7 Summary and Outlook

We introduced simplicial depth based statistics to allow inference for specific autoregressive growth processes. This thesis solves some questions for (non-linear) autoregressive processes of order one. As a result, we proposed a formal framework to analyse crack growth in prestressed concrete via autoregressive models. Fitting to this framework, we extended ideas from regression depth to autoregressive processes to allow statistical inference. This included the definition of estimators, tests and confidence regions for the discussed models. In particular, our methods were also outlier robust and distribution free, with respect to the model errors and can be extended to more general growth models, as for example polynomial regression or the Michelis-Menten model. By appropriate simplifications, we also proposed variations of the test statistics to allow fast, but less efficient evaluations for large samples. In the thesis we have shown that the methods are not necessarily uniformly best solutions but work very reliable under the conditions set by our experimental data. This, in particular, means that the estimators lose efficiency under standard assumptions like normally distributed errors and are not as efficient as standard robust estimators in symmetric contamination cases, but are superior, if we consider asymmetric contaminated errors or atypical skewed error distributions. Thereby, we also have shown that the introduced methods are reliable in case of linear and non-linear autoregressive processes of order one for errors satisfying $med(E_n) = 0$ and in addition to this are distribution free, with respect to the exact error distributions allowing a wide range of applications.

The simulations of the methods have shown that depth methods can be quite useful to achieve reliable results when reasonable doubts on standard assumptions, like symmetric or centred normally distributed errors, exist. In such cases, we pay the price of a slightly higher uncertainty and of computation time but the results are robust with respect to outliers and asymmetry.

Beyond the theoretical achievements introduced in this theses, we also presented solutions to compute the proposed methods. Thereby, we proposed vectorised and parallel calculation schemes for the resulting test statistics. Further, we analysed the properties of simplicial depth for autoregressive processes, to allow a data driven selection of candidates. Based on these candidates, the calculation of estimates could be improved by using data driven optimisation, instead of user defined grids. Additionally, these candidates allowed an efficient calculation of confidence sets, since they define the boundaries of these sets in a data driven way. Finally, we also could construct change point detection and prediction methods, based on our implemen-

tations which allowed us to apply the methods to fundamental research questions in engineering.
In case of crack growth in prestressed concrete, assumptions like centred symmetric errors are violated in a natural way. Hence, our method can be assumed to be more reliable than an application of standard methods. Due to our implementation of estimators, tests and prediction methods, we were able to propose a way which estimates a S-N curve for censored experiments. The resulting curve clearly deviates from a linear relation of load and lifetime on a log scale for the experiment under low loading as actually assumed by standard hypothesis in construction engineering. This indicates that the fatigue dynamics are about to change for load below of $100N/m$. However, based on the low number of replications and experiments, more reliable results are just possible, if we generate more data under repeated conditions to validate the methods and preliminary results.
The application on oil prices illustrates an application on data from non-negative autoregressive processes close to the unit-root. This makes the methods an interesting choice for a wider range of real world applications, including financial products and economic examples where such processes appear quite often.
Beside of these results, the thesis also raises some further research questions.
A central further step beyond the thesis is the derivation of the asymptotic distribution of the simplicial depth statistic for a parameter dimension larger than two. In this case the limit of

$$\sum_{n_1<\ldots<n_{k+1}} \left(\prod_{j=1}^{k} \mathbb{1}_{\{(-1)^j E_{n_j}(\theta,y)>0)\}} + \prod_{j=1}^{k} \mathbb{1}_{\{(-1)^{j-1} E_{n_j}(\theta,y)<0)\}} \right)$$

has to be derived. As in the two parameter case, the main task will be an asymptotic symmetrisation of the test statistic, to make the application of the continuous mapping theorem available. So far, for the case with $\theta \in \mathbb{R}^d, d > 2$ at least approximations of the quantiles can be simulated by using the median zero assumption on the residuals.
Even with the limits at hand, models with high parameter dimension lead to additional problems. The complexity of the calculation increases in the parameter dimension, since larger groups of residuals have to be compared. This restricts the usage of matrix based methods in the implementation and hence makes it necessary to discuss implementations with a more efficient memory allocation. Therefore, it will be useful to revise the presented functions and to implement them in, for example, `C` to allow a faster computation.

Since this thesis was aimed to introduce methods to be applied to a specific problem in engineering, we also neglected a deeper analysis of the robustness of the proposed methods. Of course, the definition of the test statistic and the simulations indicate that the method should be robust to innovation outliers, and to a certain extend to additive outliers also, but so far no theoretical results on influence functions or breakdown points are available. This should be addressed in future work.
The proposed extensions for prediction and phase change detection also leave some interesting open questions. In case of prediction, we already have shown that the method appears to deliver valid prediction intervals. Here, a theoretical result on the bootstrap validity is desirable. In case of the change point detection, a method based on CUSUM depth statistics or a deeper understanding of joint confidence regions could improve the results and allow valid change point tests.
The error distribution also leaves an issue which could be addressed theoretically in future work. We assume to observe an error process which induces $med(E_n) = 0$. Further, we allow skewed error distributions. Hence, it is very likely that the continuous time error process can be modelled by some Lévy-Process with restricted parameters. Thereby, the restriction has to consider the median zero assumption and the growth conditions. It would be desirable, if in combination of the intercept of the model, the Lévy-Triple of potential processes could be specified.
Summarising, this work presents a consistent package for estimation, testing, parameter inference, prediction and change point detection of growth processes under mild conditions on the underlying error distribution. It includes theoretical results on the statistics and empirical verifications for the derived prediction and change point detection algorithms. Additionally, all methods are implemented in `R` and hence available for further usage. Due to the general results, an application to several other models and problems is possible.

References

Agostinelli, C. and Romanazzi, M. (2011). Local depth. *Journal of Statistical Planning and Inference*, 141:817–830.

Anderson, T. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics*, 30:676–687.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18:1–22.

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., and Delignette-Muller, M.-L. (2015). A toolbox for nonlinear regression in R: The package nlstools. *Journal of Statistical Software*, 66(5):1 – 21.

Beale, E. (1960). Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society*, 22:41–88.

Bischl, B., Lang, M., Mersmann, O., Rahnenführer, J., and Weihs, C. (2015). BatchJobs and BatchExperiments: Abstraction mechanisms for using R in batch environments. *Journal of Statistical Software*, 64(11):1 – 25.

Burr, M. A., Rafalin, E., and Souvaine, D. L. (2004). Simplicial depth: An improved definition, analysis, and efficiency for the finite sample case. In *Proceedings of the 16th Canadian Conference on Computationan Geometry (CCCG'04)*, pages 136–139.

Chiquet, J., Limnios, N., and Eid, M. (2009). Piecewise deterministic markov processes applied to fatigue crack growth modelling. *Journal of Statistical Planning and Inference*, 139:1657–1667.

Conover, W. J. (1971). *Practical Nonparametric Statistics*. John Wiley & Sons.

Denby, L. and Martin, R. D. (1979). Robust estimation of the first-order autoregressive parameter. *Journal of the American Statistical Association*, 74(365):140–146.

Denecke, L. and Müller, C. H. (2011). Robust estimators and tests for copulas based on likelihood depth. *Computational Statistics and Data Analysis*, 55:2724–2738.

Denecke, L. and Müller, C. H. (2012). Consistency and robustness of tests and estimators based on depth. *Journal of Statistical Planning and Inference*, 142:2501–2517.

Denecke, L. and Müller, C. H. (2014a). Consistency of the likelihood depth estimator for the correlation coefficient. *Statistical Papers*, 55:3–13.

Denecke, L. and Müller, C. H. (2014b). New robust tests for the parameters of the weibull distribution for complete and censored data. *Metrika*, 77(5):585–607.

Edelsbrunner, H., Kirkpatrick, D. G., and Seidel, R. (1983). On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559.

Efron, B. (1979). Bootstrap methods: another look at the jacknife. *Annals of Statistics*, 9:1218–1228.

Hoeffding, W. (1948). A class of statistics with asymasymptotic normal distribution. *Annals of Mathematical Statistics*, 19(3):293–325.

Hotelling, H. (1931). The economics of exhaustible resources. *The Journal of Political Economy*, 39(2):137–175.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics*, 1(5):799 – 821.

Huggins, R. (1989). The sign test for stochastic processes. *Australian Journal of Statistics*, 31:153–165.

Jae, K. H. (2014). *BootPR: Bootstrap Prediction Intervals and Bias-Corrected Forecasting*. R package version 0.60.

Kaufmann, H. and Kruse, R. (2013). Bias-corrected estimation in potentially mildly explosive autoregressive models. CREATES Research Paper 2013 - 10.

Kilian, L. (1998). Estimators confidence intervals for impulse response functions. *The Review of Economics and Statistics*, 80:218–230.

Kim, J. (2001). Bootstrap-after-bootstrap prediction intervals for autoregressive models. *Journal of Business and Economic Statistics*, 19:117–128.

Kim, J. H. (2003). Forecasting autoregressive time series with bias-corrected parameter estimators. *International Journal of Forecasting*, 19:493–502.

Kloeden, P. E. and Platen, E. (1992). *Numerical solution of sochastic differential equations*. Springer, Berlin, Heidelberg.

Knight, K. (1998). Limlimit distributions for l1 regression estimators under general conditions. *The Annals of Statistics*, 26(2):755–770.

Kustosz, C., Leucht, A., and Müller, C. H. (2016a). Tests based on simplicial depth for ar(1) models with explosion. *Journal of Time Series Analysis*. In press.

Kustosz, C. and Müller, C. H. (2014). Analysis of crack growth with robust distribution-free estimators and tests for nonstationary autoregressive processes. *Statistical Papers*, 55:125–140.

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

Lin, L. and Chen, M. (2006). Robust estimating equation based on statistical depth. *Statistical Papers*, 47:263–278.

Lin, Y., Wu, W., and Yang, J. (1985). Stochastic modeling of fatigue crack propagation. In *Probabilistic Methods in the Mechanics of Solids and Structures*. Springer, Berlin, Heidelberg.

Liu, R. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414.

Liu, R., Parelius, J., and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858.

Luenberger, D. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons.

Mahalanobis, P. (1936). On the generalized distance in statistics. In *Proceedings of the National Institute of Science of India*, volume 2, pages 49–55.

Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, 11(3/4):173–220.

Maronna, R., Martin, R. D., and Yohai, V. (2006). *Robust statistics*. John Wiley & Sons.

Maurer, R. and Heeke, G. (2010). Ermüdungsfestigkeit von spannbeton aus einer älteren spannbetonbrücke. Technical report, TU Dortmund University.

Mizera, I. (2002). On depth and deep points: a calculus. *The Annals of Statistics*, 30:1681–1736.

Muggeo, V. M. R. (2003). Estimating regression models with unkown break-points. *Statistics in Medicine*, 22:3055–3071.

Muggeo, V. M. R. (2008). segmented: an r package to fit regression models with broken-line relationships. *R News*, 8(1):20 – 25.

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

Novomestky, F. (2012). *matrixcalc: Collection of functions for matrix calculations*. R package version 1.0-3.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, 1:327–332.

Paindaveine, D. and Van Bever, G. (2013). From depth to local depth: A focus on centrality. *Journal of the American Statistical Association*, 108(503):1105–1119.

Paris, P. and Erdogan, F. (1963). A critical analysis of crack propagation laws. *Journal of Basic Engineering*, 85(1):528–534.

Pateiro-Lopez, B. and Rodriguez-Casal, A. (2015). *alphahull: Generalization of the Convex Hull of a Sample of Points in the Plane*. R package version 2.0.

RCoreTeam (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ritz, C. and Streibig, J. C. (2008). *Nonlinear regression with R*. Springer, New York.

Ross, S. M. (1996). *Stochastic Processes*. John Wiley & Sons.

Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., and Maechler, M. (2015). *robustbase: Basic Robust Statistics*. R package version 0.92-3.

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

Shaman, P. and Stine, R. A. (1988). The bias of autoregressive coefficient estimatiors. *Journal of the American Statistical Association*, 83:842–848.

Snaidy, P., Sieniawska, R., and Zurowski, S. (1998). Application of filtered poisson process in modeling fatigue cracks propagation. *Zeitschrift für Angewandte Mathematik und Mechanik*, 78(3):1083–1084.

Sobczyk, K. (1979). On the stochastic model for fatigue crack propagation. *Bulletin of the Polish Academy of Sciences, Technical Sciences*, 27(5/6).

Sobczyk, K. (1987). Stochastic models for fatigue damage of materials. *Advances in Applied Probability*, 19(3):652–673.

Spencer, B., Tang, J., and Artley, M. (1989). Stochastic approach to modeling fatigue crack growth. *American Institute of Aeronautics and Astronautics Journal*, 27(11):1628–1635.

Stine, R. A. and Shaman, P. (1989). A fixed point characterization for bias of autoregressive estimators. *The Annals of Statistics*, 17:1275–1284.

Struyf, A. and Rousseeuw, P. (1999). Halfspace depth and regression depth characterize the empirical distribution. *Journal of Multivariate Analysis*, 69:135 – 153.

Stute, W. and Gründer, B. (1993). Nonparametric prediction intervals for explosive ar(1)-processes. *Journal of Nonparametric Statistics*, 2(2):155–167.

Thombs, L. A. and Schuchany, W. R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, 85:486–492.

Tierney, L., Rossini, A. J., Li, N., and Sevcikova, H. (2013). *snow: Simple Network of Workstations*. R package version 0.3-13.

Tukey, J. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver 1974*, pages 523–531.

van der Vaart, A. W. (2007). *Asymptotic Statistics*. Cambridge University Press.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York.

Wang, X. and Yu, J. (2013). Limit theory for an explosive autoregressive process. Working Paper, No 08-2013. Singapore Management University, School of Economics.

Wellmann, R., Harmand, P., and Müller, C. H. (2009). Distribution-free tests for polynomial regression based on simplicial depth. *Journal of Multivariate Analysis*, 100:622–635.

Wellmann, R. and Müller, C. H. (2010). Depth notions for orthogonal regression. *Journal of Multivariate Analysis*, 101:2358–2371.

Witting, H. and Müller-Funk, U. (1995). *Mathematische Statistik II*. Teubner, Stuttgart.

Wu, W., Ni, C., and Liou, H. (2001). Random outcome and stochastic analysis of some fatigue crack growth data. *Journal of Mechanics*, 17(2):61–68.

Yang, J. and Manning, S. (1990). Stochastic crack growth analysis methodologies for metallic structures. *Engineering Fracture Mechanics*, 37(5):1105–1124.

Yu, H. (2002). Rmpi: Parallel statistical computing in r. *R News*, 2(2):10 – 14.

Zeileis, A., Kleiber, C., Krämer, W., and Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis*, 44:109–123.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002). strucchange: An r package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2):1 – 38.

Zio, E. and Zoia, A. (2009). Parameter identification in degradation modeling by reversible-jump markov chain monte carlo. *IEEE Transactions on reliability*, 58(1):123–131.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Annals of Statistics*, 28(2):461–482.

A Additional Simulations

In this Section we present the results of additional simulations which examine the performance of our methods in cases for which the methods are not verified by theoretical results or which extend the results presented in the main thesis.

A.1 Estimation

The first simulations show the estimators for the linear AR(1) model without intercept, when the underlying autoregression parameter implies a stationary or a unit root process.
We begin with the stationary case. In Figures 102 and 103 the estimator comparisons are presented in boxplots. In Table 27 the comparison of mean squared errors is presented.

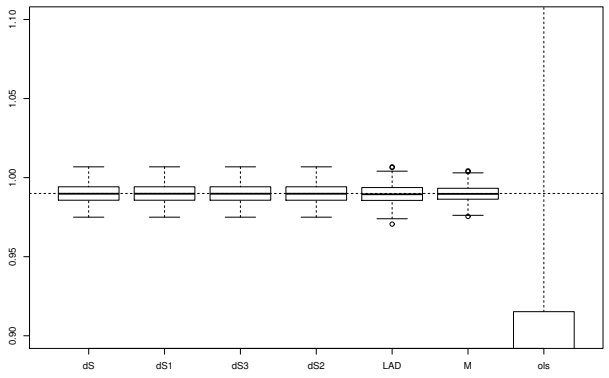


Fig. 102: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\kappa_1 = 0.99$ based on $\mathcal{N}(0, 0.1)$ errors.

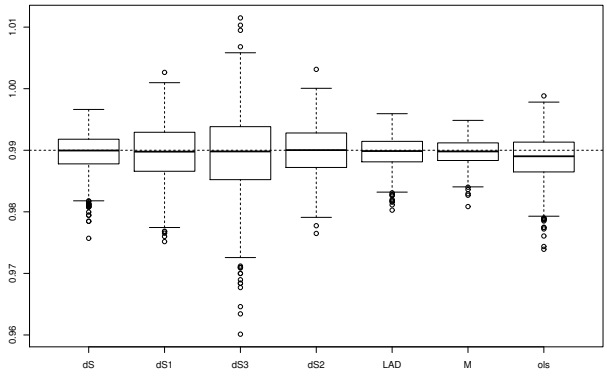


Fig. 103: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\kappa_1 = 0.99$ based on $\mathcal{N}(0, 0.1)$ errors.

N	d_S	d_S^1	d_S^2	d_S^3	LAD	M	OLS
10	$3.28 \cdot 10^{-5}$	$3.28 \cdot 10^{-5}$	$3.28 \cdot 10^{-5}$	$3.28 \cdot 10^{-5}$	$43.26 \cdot 10^{-5}$	$2.27 \cdot 10^{-5}$	$9.42 \cdot 10^{-2}$
250	$1.00 \cdot 10^{-5}$	$2.00 \cdot 10^{-5}$	$4.68 \cdot 10^{-5}$	$1.60 \cdot 10^{-5}$	$6.31 \cdot 10^{-6}$	$4.13 \cdot 10^{-6}$	$1.44 \cdot 10^{-5}$

Tab. 27: MSE for linear AR processes without intercept with parameter $\kappa_1 = 0.99$ based on $\mathcal{N}(0, 0.1)$ errors.

The results are quite similar to the explosive case. In the small sample the robust estimators perform very well, while the OLS estimator suffers a rather large small sample bias. For the larger sample size the improvement of the robust estimators is low, while the OLS clearly improves towards the true value. In general the M-estimator shows the best performance while our robust alternatives are slightly worse.

For the unit-root process we get the results presented in Figures 104 and 105.

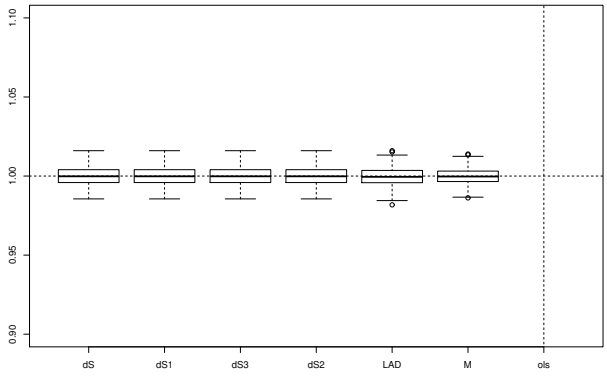


Fig. 104: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\kappa_1 = 1$ based on $\mathcal{N}(0, 0.1)$ errors.

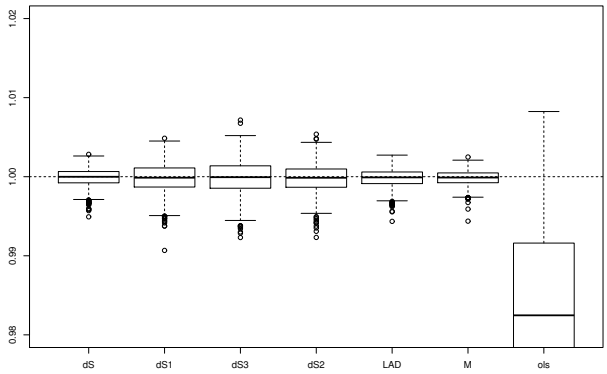


Fig. 105: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\kappa_1 = 1$ based on $\mathcal{N}(0, 0.1)$ errors.

In the unit root case the M-estimator performs best. The relation between all other estimators is qualitatively the same as without contamination. The only remarkable effect is that the OLS estimator now suffers a very large bias which is just weakly corrected by increasing sample sizes.
The comparison of the estimators for contaminated normal errors in stationary or unit root processes is presented in Figures 106, 107 and 110, 111. For the Fréchet distributed errors the results can be found in Figures 108, 109 and 112, 113.

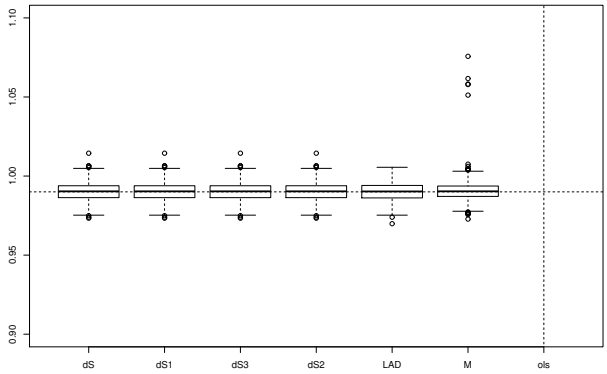


Fig. 106: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\kappa_1 = 0.99$ based on $\mathcal{CN}(0, 0.1)$ errors.

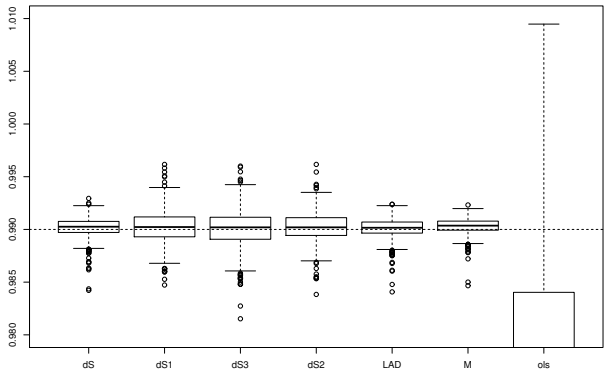


Fig. 107: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\kappa_1 = 0.99$ based on $\mathcal{CN}(0, 0.1)$ errors.

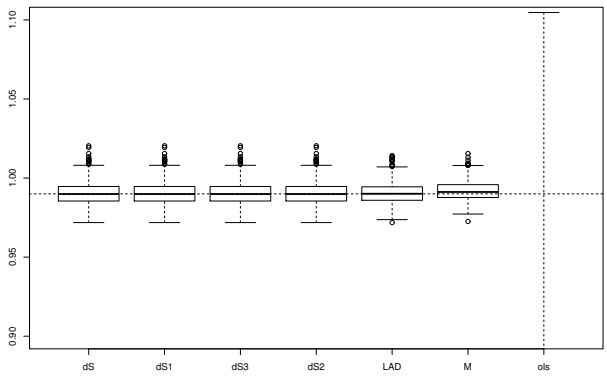


Fig. 108: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\kappa_1 = 0.99$ based on $\mathcal{F}(1.928, -2, 10)$ errors.

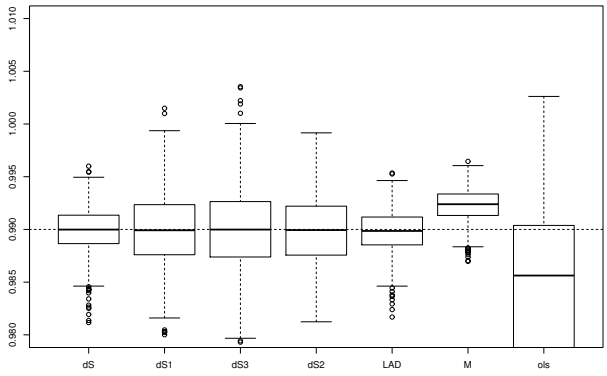


Fig. 109: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\kappa_1 = 0.99$ based on $\mathcal{F}(1.928, -2, 10)$ errors.

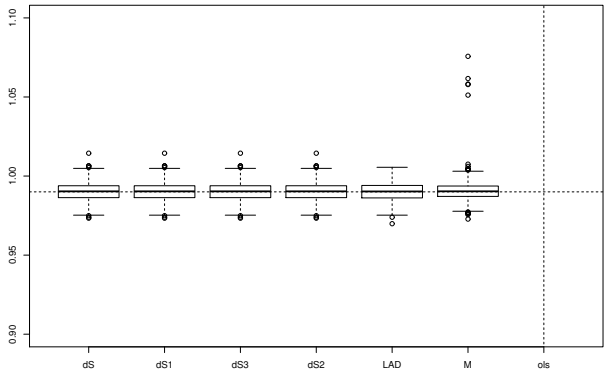


Fig. 110: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\kappa_1 = 1$ based on $\mathcal{CN}(0, 0.1)$ errors.

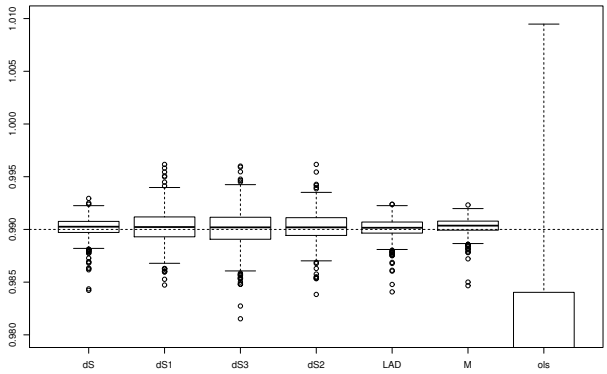


Fig. 111: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\kappa_1 = 1$ based on $\mathcal{CN}(0, 0.1)$ errors.

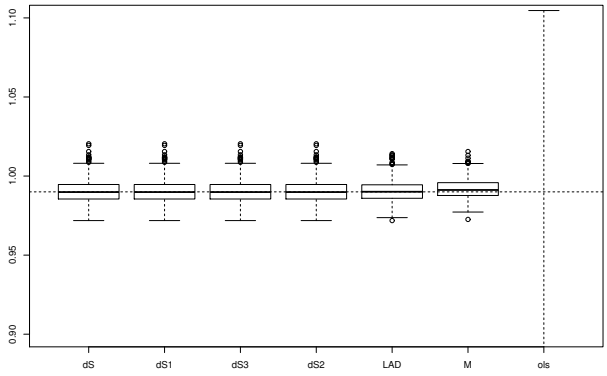


Fig. 112: Estimator performance for linear AR processes without intercept with $N = 10$ observations and parameter $\kappa_1 = 1$ based on $\mathcal{F}(1.928, -2, 10)$ errors.

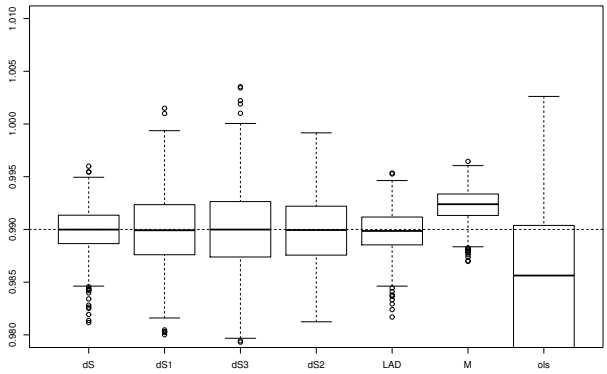


Fig. 113: Estimator performance for linear AR processes without intercept with $N = 250$ observations and parameter $\kappa_1 = 1$ based on $\mathcal{F}(1.928, -2, 10)$ errors.

The simulations indicate that even if we were not able to prove the consistency of the estimators in the unit root and stationary case, the proposed estimators work quite well in these situations.

A.2 Tests

The following figures show projections of the power functions in the two-parameter linear AR(1) model. The results are discussed in Section 5.

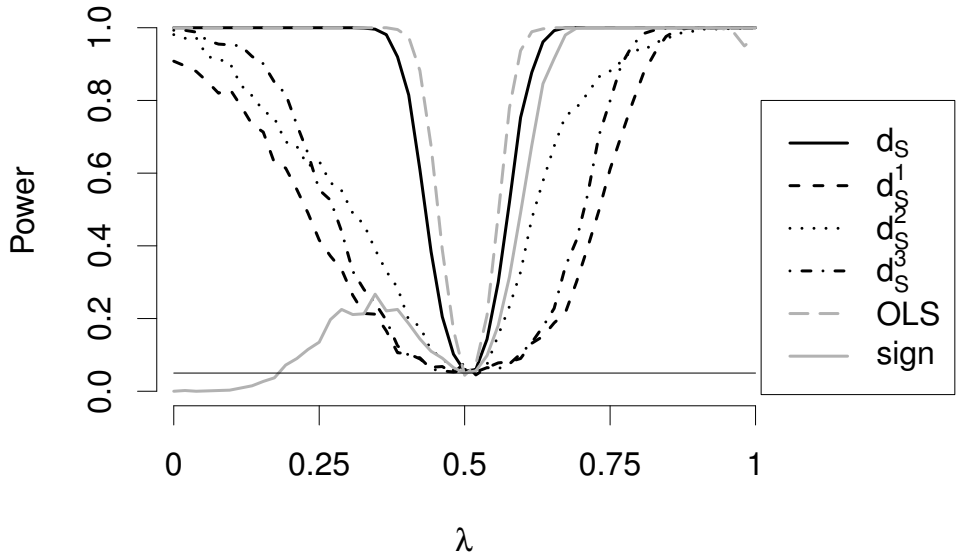


Fig. 114: Power evaluated along $\theta = (0.725, 0.9995)^\top + \lambda(-1.05, 0.021)^\top$ for normally distributed errors.

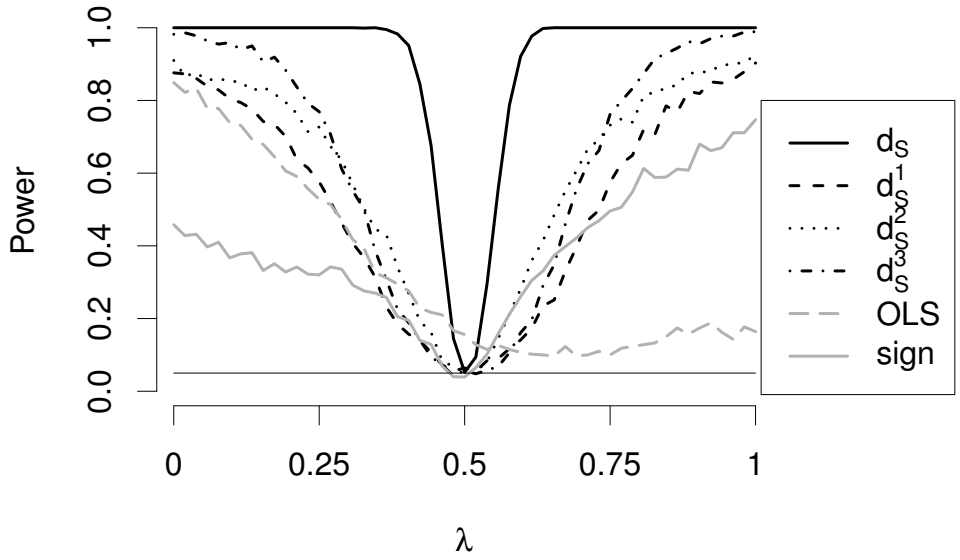


Fig. 115: Power evaluated along $\theta = (0.725, 0.9995)^\top + \lambda(-1.05, 0.021)^\top$ for errors with contaminated normal distribution.

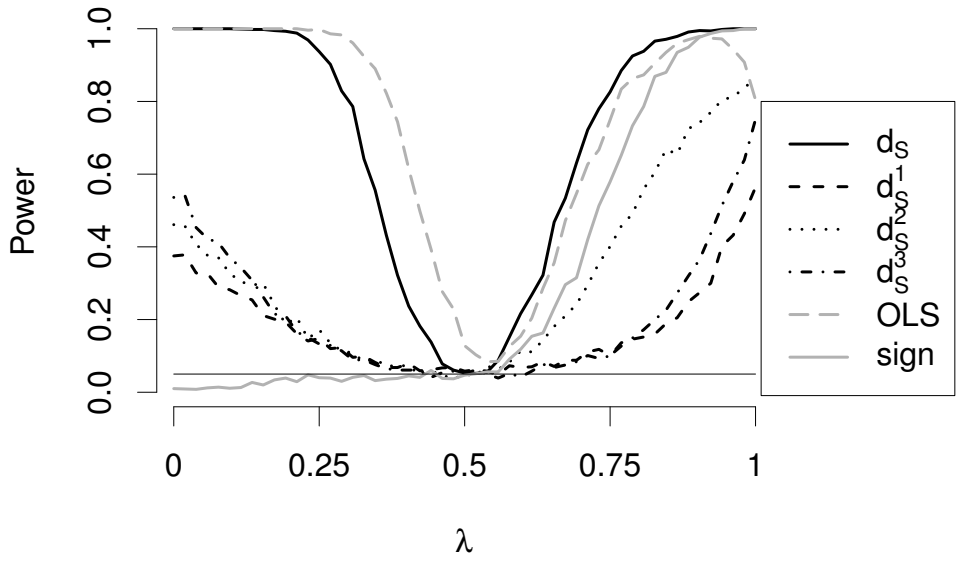


Fig. 116: Power evaluated along $\theta = (0.725, 0.9995)^\top + \lambda(-1.05, 0.021)^\top$ for errors with Fréchet distribution.

A.3 Applications

The next figures show the confidence sets for the three parameter model based on the d_S^1 and d_S^2 statistic for TR01, TR02 and TR03.

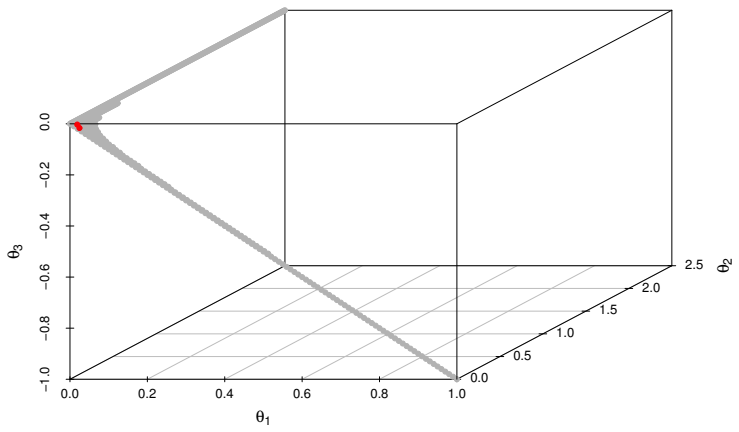


Fig. 117: 95% parameter confidence set for TR01 series with noise based on d_S^2 and the three parameter model. The red dots maximise the depth.

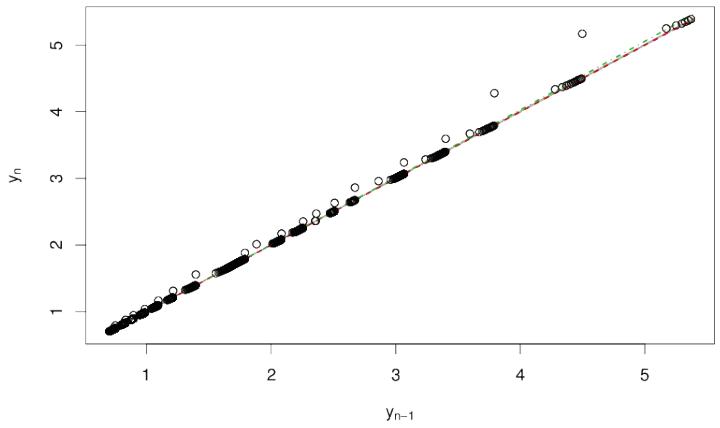


Fig. 118: Fits from the 95% parameter confidence set for TR01 series with noise based on d_S^1 and the three parameter model. The unfilled dots represent the data by (y_{n-1}, y_n). The parameters which maximise depth define the red fitted lines. The green line is a fit defined by the OLS estimate for the three parameter model.

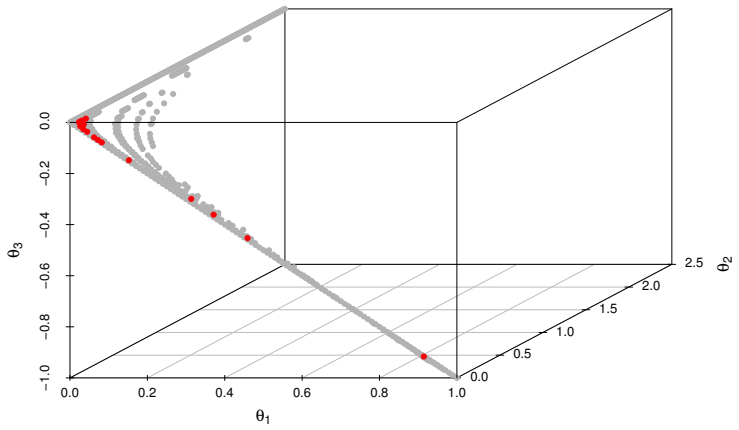


Fig. 119: 95% parameter confidence set for TR02 series with noise based on d_S^1 and the three parameter model. The red dots maximise the depth.

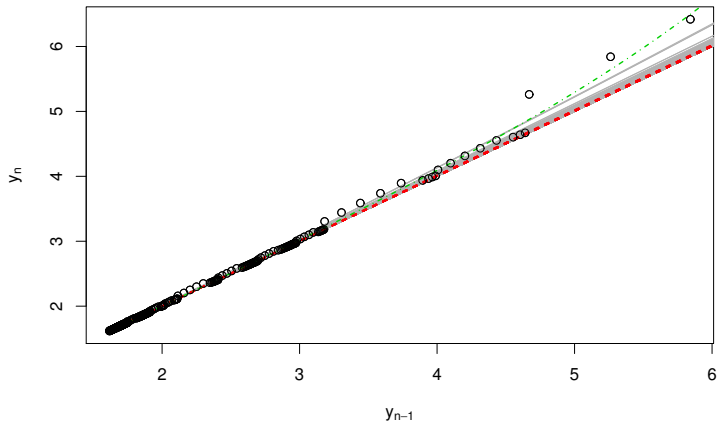


Fig. 120: Fits from the 95% parameter confidence set for TR02 series with noise based on d_S^1 and the three parameter. The unfilled dots represent the data by (y_{n-1}, y_n). The parameters which maximise depth define the red fitted lines. The green line is a fit defined by the OLS estimate for the three parameter model.

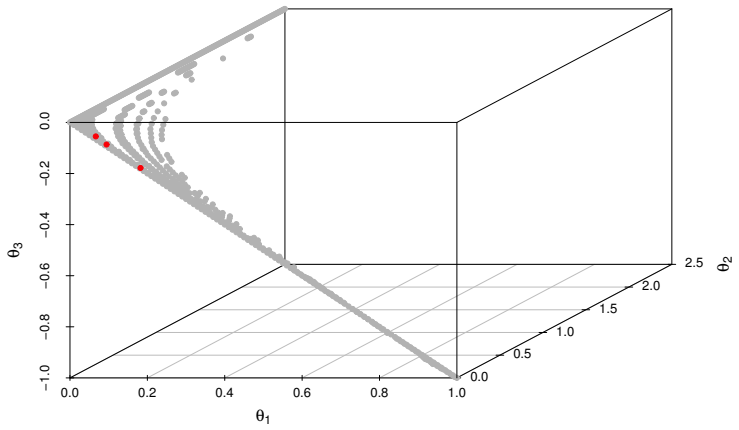


Fig. 121: 95% parameter confidence set for TR02 series with noise based on d_S^2 and the three parameter model. The red dots maximise the depth.

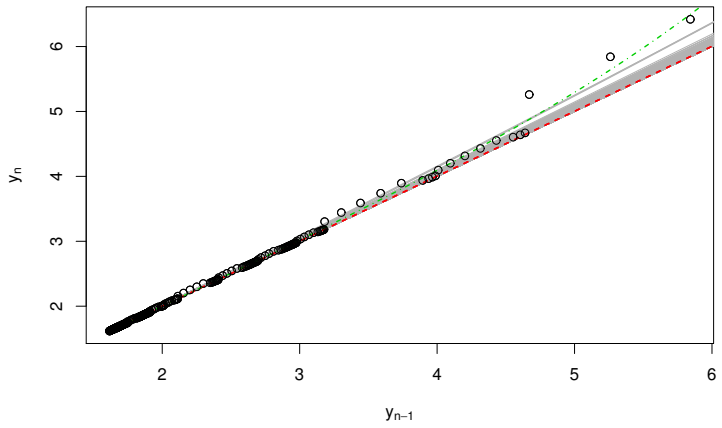


Fig. 122: Fits from the 95% parameter confidence set for TR02 series with noise based on d_S^2 and the three parameter model. The unfilled dots represent the data by (y_{n-1}, y_n). The parameters which maximise depth define the red fitted lines. The green line is a fit defined by the OLS estimate for the three parameter model.

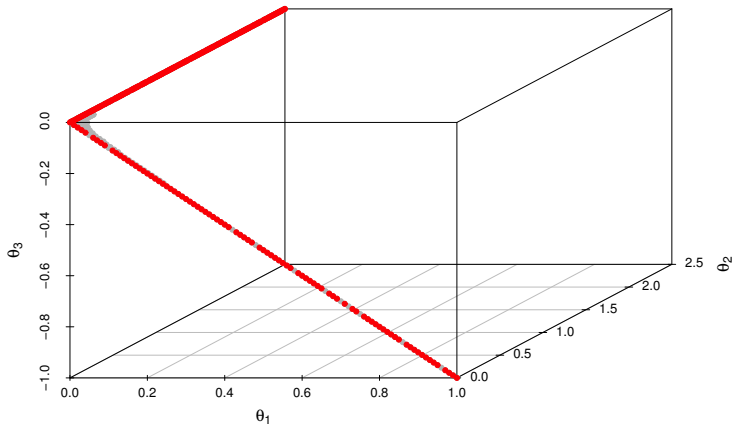


Fig. 123: 95% parameter confidence set for TR02 series with noise based on d_S^1 and the three parameter model. The red dots maximise the depth.

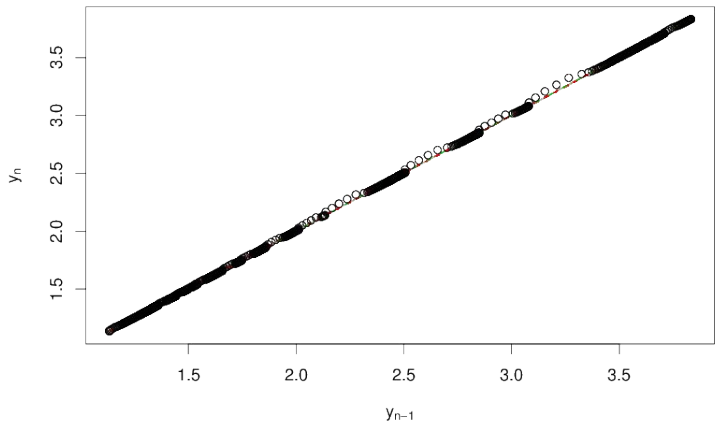


Fig. 124: Fits from the 95% parameter confidence set for TR03 series with noise based on d_S^1 and the three parameter model. The unfilled dots represent the data by (y_{n-1}, y_n). The parameters which maximise depth define the red fitted lines. The green line is a fit defined by the OLS estimate for the three parameter model.

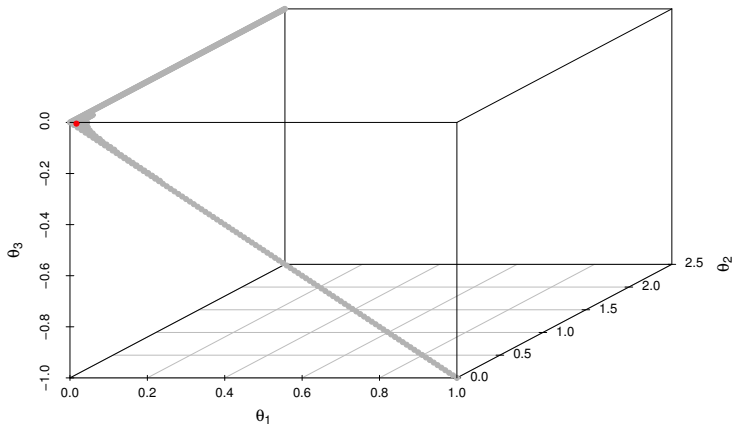


Fig. 125: 95% parameter confidence set for TR02 series with noise based on d_S^2 and the three parameter model. The red dots maximise the depth.

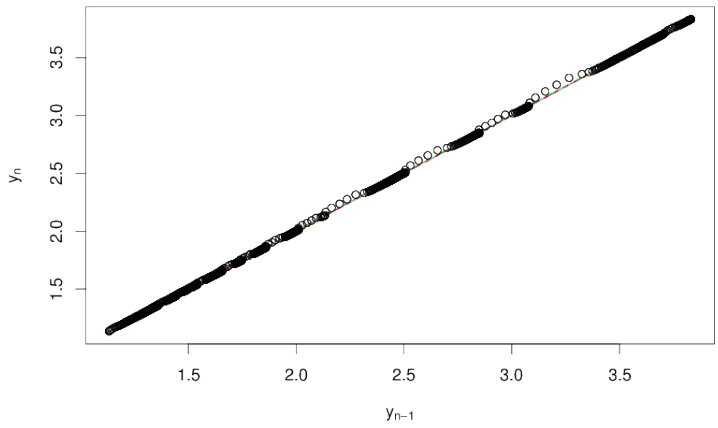


Fig. 126: Fits from the 95% parameter confidence set for TR03 series with noise based on d_S^2 and the three parameter model. The unfilled dots represent the data by (y_{n-1}, y_n). The parameters which maximise depth define the red fitted lines. The green line is a fit defined by the OLS estimate for the three parameter model.

A.4 Proofs

In this section, the proofs of miscellaneous subsidiary statements are given.

Proof of Lemma 67. Denote $A = (A_1, A_2), B = (B_1, B_2)$ and $C = (C_1, C_2)$. Then the line from A to B can be written as

$$L_{AB} : P = (P_1, P_2) = A + \lambda \cdot (B - A), \lambda \in \mathbb{R}.$$

For $\lambda = \frac{1}{2}$ we are at the middle of L_{AB}. Denote this point by $M_{AB} := A/2 + B/2$. Then the straight from M_{AB} to C is defined by

$$L_{M_{AB}C} : (1/2)(B + A) + \mu[C - (1/2)(B + A)], \mu \in \mathbb{R}.$$

Analogously one can calculate

$$L_{M_{AC}B} : (1/2)(C + A) + \nu[B - (1/2)(C + A)], \nu \in \mathbb{R}$$

and

$$L_{M_{CB}A}(1/2)(C + B) + \tau[A - (1/2)(C + B)], \tau \in \mathbb{R}.$$

The lines $L_{M_{AB}C}$ and $L_{M_{AC}B}$ intersect at $\mu = \nu = \tau = \frac{1}{3}$, since

$$\begin{aligned} &(1/2)(B + A) + (1/3)[C - (1/2)(B + A)] \\ =&B/2 + A/2 + C/3 - B/6 - A/6 \\ =&(1/3) \cdot (A + B + C), \end{aligned}$$

$$\begin{aligned} &(1/2)(C + A) + (1/3)[B - (1/2)(C + A)] \\ =&C/2 + A/2 + B/3 - C/6 - A/6 \\ =&(1/3) \cdot (A + B + C). \end{aligned}$$

and

$$\begin{aligned} &(1/2)(C + B) + (1/3)[A - (1/2)(C + B)] \\ =&C/2 + B/2 + A/3 - C/6 - B/6 \\ =&(1/3) \cdot (A + B + C). \end{aligned}$$

This coincides with the centroid S. Considering the last combination $L_{M_{CBA}}$ delivers the same result. Since the straights from $L_{M_{ABC}}, L_{M_{ACB}}$ and $L_{M_{CBA}}$ are inside $conv(A, B, C)$ for $\mu, \nu, \tau \in (0, 1)$, we conclude that $S \in \overline{conv(A, B, C)}$ holds. □

B The rexpar Package

In this part of the Appendix we shortly go through the functionality of supplementary `rexpar` package which is available for `R` in a `github` repository and includes the main functions introduced in this thesis.
The description is divided into central functions, important helper functions and auxiliary functions. For details on the arguments of the functions the documentation of the `rexpar` package can be used. The package can be downloaded at `https://github.com/ChrisKust/rexpar.git` in the recent version. Here, we describe the main function, used in this thesis.

B.1 Main Functions

We start with the main test statistics.

`dS_lin1(theta,y,mod)`
This function calculates the full simplicial depth for the autoregressive process without intercept for a fixed AR parameter `theta` and an observed process `y`. The `mod` argument allows to apply the modified weighted version for stationary processes.

`dS_lin2(theta,res,y,ncores,model,cpow)`
This function calculates the full simplicial depth for a parameter `theta` and an observed process `y` for a specified type of `model`. Alternatively, one can also directly plug-in residuals without a model choice by the argument `res`. At the current stage, we have implemented the linear autoregressive model with intercept and the nonlinear two parameter autoregressive model. The `ncores` argument allows the usage of multiple cores to calculate the test statistic. In case of the two parameter model with fixed power, `cpow` can be used to set a fixed power parameter.

`dS1_lin2(theta,y,model)`
This function calculates the simplified simplicial depth, defined by subsequent nonoverlapping blocks of residuals (d_S^1 in this thesis) for each of the three models and a given parameter `theta` and on observation vector `y`. The model thereby is specified in the argument `model`.

`dS2_lin2(theta,y,model, res)`
This function calculates the simplified simplicial depth defined by the central observation and increasing/decreasing indices for the remaining residuals (d_S^3 in this thesis) for each of the three models and a given parameter `theta` and on observation vector `y`. The model thereby is specified in the argument `model`. Alternatively, one can also directly plug-in residuals without a model choice by the argument `res`

`dS3_lin2(theta,y,model,res)`
This function calculates the simplified simplicial depth defined by the groups of subsequent residuals with overlapping (d_S^2 in this thesis) for each of the three models and a given parameter `theta` and on observation vector `y`. The model thereby is specified in the argument `model`. Alternatively, one can also directly plug-in residuals without a model choice by the argument `res`

Now we turn to the tests.

`dS_lin1_test(thetaN,alpha,y,mod)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case on a linear autoregressive process without intercept. The `mod` argument allows a weighting of the residuals for stationary processes.

`dS_lin2_test(thetaN,alpha,y,ncores)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a linear autoregressive process with intercept. Thereby the `ncores` argument can be used to perform a multicore computation.

`dS_nlin_test(thetaN,alpha,y,ncores)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a nonlinear autoregressive process with two parameters. Thereby the `ncores` argument can be used to perform a multicore computation.

`dS1_lin1_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a linear autoregressive process without intercept based on the simplified notion d_S^1. It allows the usage of the approximate normal limit distribution or an exact calculation of the quantiles of the distribution under H_0 by the argument `exact`.

`dS2_lin1_test(thetaN,alpha,y,exact)`

This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a linear autoregressive process without intercept based on the simplified notion d_S^3. It allows the usage of the approximate normal limit distribution or an exact calculation of the quantiles of the distribution under H_0 by the argument `exact`.

`dS3_lin1_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a linear autoregressive process without intercept based on the simplified notion d_S^2. It allows the usage of the approximate normal limit distribution or an exact calculation of the quantiles of the distribution under H_0 by the argument `exact`.

`dS1_lin2_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a linear autoregressive process with intercept based on the simplified notion d_S^1. The argument `exact` allows the usage of the exact distribution of d_S^1 to calculate the critical values.

`dS2_lin2_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a linear autoregressive process with intercept based on the simplified notion d_S^3. The argument `exact` allows the usage of the exact distribution of d_S^3 to calculate the critical values.

`dS3_lin2_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a linear autoregressive process with intercept based on the simplified notion d_S^2. The argument `exact` allows the usage of the exact distribution of d_S^2 to calculate the critical values.

`dS1_nlin_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a nonlinear autoregressive process based on the simplified notion d_S^1. The argument `exact` allows the usage of the exact distribution of d_S^1 to calculate the critical values.

`dS2_nlin_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a nonlinear autoregressive process based on the

simplified notion d_S^3. The argument `exact` allows the usage of the exact distribution of d_S^3 to calculate the critical values.

`dS3_nlin_test(thetaN,alpha,y,exact)`
This function implements the test for $H_0 : \theta =$ `thetaN` with level `alpha` applied to an observation vector `y` in case of a nonlinear autoregressive process based on the simplified notion d_S^2. The argument `exact` allows the usage of the exact distribution of d_S^2 to calculate the critical values.

The estimators are implemented in the following functions.

`est_lin1(y,maxit,candy,acc,plots,eps,unique,notion)`
This function estimates the parameter of a linear autoregressive process without intercept based on one of the proposed depth statistics, specified by `notion` for an observed process `y`. Thereby several optimisation parameters can be set. The `maxit` argument specifies the maximal number of iterations. By the `candy` argument the user can specify, if all candidates are considered by the calculation of the edges of the residual simplexes or if inner points should be considered. The `acc` argument defines an accuracy parameter to determine the size of search regions in the iterative procedure. The `plots` option allows to turn on plots of the evaluations in the iterative procedure. The `eps` parameter specifies a search step, if inner points of the candidate simplexes are evaluated. The `unique` argument reduces the result of the estimator to one value, instead of a complete set in the maximising region by selecting a central observation in the set.

`est_lin2(y,maxit,candy,candy_eps,perc,acc,plots,`
`normtype,pv,wgt,unique,notion,opt_rude)`
This function estimates the parameter of a linear autoregressive process with intercept based on one of the proposed depth statistics, specified by `notion` for an observed process `y`. Thereby several optimisation parameters can be set. The `maxit` argument specifies the maximal number of iterations. By the `candy` argument the user can specify, if all candidates are considered by the calculation of the edges of the residual simplexes or if inner points should be considered. The `candy_eps` option allows to define an alternative way to explore the candidate region based on small steps away from the edges. The `prec` parameter defines a precision value which is used as soft stopping criterion in the algorithm. The `acc` argument defines an accuracy parameter to determine the size of search regions in the iterative procedure. The `plots` option allows to turn on plots of the evaluations in the iterative procedure. The parameters `nortype`,`pv` and `wgt` are used to define the search regions in

the iterative algorithm. The `unique` argument reduces the result of the estimator to one value, instead of a complete set in the maximising region by selecting a central observation in the set. The `opt_rude` option uses the candidate parameters just for a starting guess and then tries to optimise depth by a Nelder-Mead algorithm. This is not reliable for the linear model but useful in the nonlinear case.

`est_nlin1(y,maxit,candy,perc,acc,plots,`
`normtype,pv,wgt,unique,notion,opt_rude)`
This function estimates the parameter of a non-linear autoregressive process based on one of the proposed depth statistics, specified by `notion` for an observed process `y`. Thereby several optimisation parameters can be set. The `maxit` argument specifies the maximal number of iterations. By the `candy` argument the user can specify, if all candidates are considered by the calculation of the edges of the residual simplexes or if inner points should be considered. The `prec` parameter defines a precision value which is used as soft stopping criterion in the algorithm. The `acc` argument defines an accuracy parameter to determine the size of search regions in the iterative procedure. The `plots` option allows to turn on plots of the evaluations in the iterative procedure. The parameters `nortype`,`pv` and `wgt` are used to define the search regions in the iterative algorithm. The `unique` argument reduces the result of the estimator to one value, instead of a complete set in the maximising region by selecting a central observation in the set.The `opt_rude` option uses the candidate parameters just for a starting guess and then tries to optimise depth by a Nelder-Mead algorithm. This is not reliable for the linear model but useful in the non-linear case.

The confidence set construction is also implemented in three functions.

`lin1_CI(y,level,plots,notion,eps)`
The function calculates confidence intervals for the linear autoregressive process without intercept given by an observation vector `y` for a desired `level`. The `plots` option produces a figure which shows the evaluated candidates and the resulting regions. The desired depth statistic can be set by `notion`. The `eps` argument defines an ϵ for the candidate selection to shift the candidates from the roots of the residuals.

`lin2_CI(y,level,plots,notion,ncoresC,mid)`
The function calculates confidence regions for the linear autoregressive process with intercept given by an observation vector `y` for a desired `level`. The `plots` option produces a figure which shows the evaluated candidates and the resulting regions.

The desired depth statistic can be set by `notion`. The `ncoresC` option allows the usage of a multicore computation. The `mid` argument allows a calculation of candidates set by the centre points of the candidate simplexes, instead of a slight shifting from the edges by default.

`nlin_CI(y,level,plots,notion,ncoresC,addPar,spar)`
The function calculates confidence regions for the non-linear autoregressive process given by an observation vector `y` for a desired `level`. The `plots` option produces a figure which shows the evaluated candidates and the resulting regions. The desired depth statistic can be set by `notion`. The `ncoresC` option allows the usage of a multicore computation. The `addpar` argument allows to add the resulting confidence set to an existing plot. Further the `spar` value sets the parameter which defines the smoothness of the resulting Alpha shape which approximates the confidence set.

The changepoint and prediction methods are not fully implemented in the package, but for some cases they are available.

`changepoints_lin2(y,level,bw,sw,plots,method,ncoresCP,mincper,mincp)`
This function implements the change point detection for a linear autoregressive model with two parameters. Thereby the observation vector `y` is used. The confidence sets applied are calculated on a preset `level` for each segment defined by a bandwidth `bw` and in steps with step-width `sw`. The `plots` option shows the resulting change points if desired. The `method` option switches between the classification via non intersecting confidence regions and the estimator based approach. By `ncoresCP` a parallel computation can be performed. The parameters `mincper` and `mincp` are tuning parameters which allow to specify a minimal expected number of change-points and a minimal percentage of the largest cluster of changepoints to define a parameter change. Another function, `changepoints_lin2_cl` replaces the argument `ncoresCP` by an argument `cluster` and allows to apply the method on a predefined cluster for parallel computation also. This is in particular useful, if the method is applied on a high performance cluster.

`changepoints_lin1(y,level,bw,sw,plots,method,ncoresCP,mincper,mincp)`
This function implements the change point detection for the linear model without intercept. The arguments are as in the `changepoints_lin2(...)` function. Here a cluster version is not available.

`predict_lin1(y,CritLen,CritTime,NSim,alpha,restrict)`
This function implements a bootstrapping based prediction procedure based on

an observed linear autoregressive process without intercept `y` for a critical value `CritLen` or a specific point in time `CritTime` based on `NSim` simulated processes. Further prediction intervals with a specified level `alpha` are constructed. The `restrict` option allows the usage of the truncated empirical depth shape to simulate the parameters.

B.2 Interesting Helper Functions

This section briefly describes helper functions which are used in the package for various reasons. Here we skip the detailed description of all arguments. The full function descriptions can be found in the documentation of the `R` package.

We begin with some functions which are useful to generate the limit distribution of the full simplicial depth with two dimensional parameter.

`simulateGP(g,sigma)`
This function simulates the two-dimensional Gaussian process which is the limit distribution of the two parameter simplicial depth statistics. Thereby a mean vector `g` and a covariance matrix `sigma` have to be specified appropriately. The covariance matrix can be calculated by the `sigmaMat(t,nclust)` function. This function needs a vector of indices `t` and can be performed on multiple cores by the `nclust` option. The mean vector can be computed by `muVec(t,y)`. Again `t` is the time index set and `y` is a scaling parameter.

Once the Gaussian process is simulated the limit distribution can be calculated by the following function. `LimitApprox(g,Y)`
This function approximates the integral given in Kustosz et al. (2016a) on a given grid `t` and an observed bivariate Gaussian process `Y`.

The quantiles of a simulations as presented in Kustosz et al. (2016a) is stored in the `SimQuants` matrix.
We also have got functions to simulate processes based on the distributions, used in this thesis.

`RandomARMod_lin2(nobs,intercept,arp,start,cont,sd)`
This function simulates a linear autoregressive process of length `nobs` with intercept parameter `intercept` and autoregression parameter `arp` starting at `start` with an error distribution given by `cont`. The parameter `sd` sets the standard deviation in case of normal errors. For other error distribution, it is neglected.

`RandomARMod_nlin1(nobs,arp,power,start,cont,sd)`
This function simulates a nonlinear autoregressive process of length `nobs` with power parameter `power` and autoregression parameter `arp` starting at `start` with an error distribution given by `cont`. The parameter `sd` sets the standard deviation in case of normal errors. For other error distribution, it is neglected.

The package also includes functions to calculate the residuals based on the proposed models.

`resARMod_lin2(theta,dat)`
This function calculates the residuals of a linear autoregressive process with intercept based on an observed process given by `dat` and a parameter `theta`.

`resARMod_nlin1(theta,dat)`
This function calculates the residuals of a nonlinear autoregressive process based on an observed process given by `dat` and a parameter `theta`.

B.3 Internal Functions and Auxilliary Functions

The following list summarises some remaining functions which are subroutines of the main functions or auxiliary functions used for examples or further work.
The first functions implement some results to apply OLS test statistics for explosive autoregressive processes with intercept.

- `ols_expl(y)` (estimation)
- `ols_ts(y,thetaT)` (test statistic)
- `ols_test(theta0,y,alpha)` (resulting test)
- `power_ols(thetas,N,R,sv,cont,theta0,alpha)` (power simulation study)

The next set of functions allows implements the candidate parameters sets for the empirical depth evaluation for the different model equations.

- `lin1_theta(dat)` (canditate edges for linear model without intercept)
- `lin1_theta_eps(dat,eps)` (inner points of candidate intervals for linear model without intercept)
- `lin2_theta_f(dat)` (candidate edges for linear model with intercept)

- `nlin1_theta_f(dat)` (candidate intersections for non-linear model)
- `Tri_Eps(y,perc,eps)` (inner points in candidate simplices for the linear model with intercept based on small steps away from edges)
- `Tri_Eps_dist(y,perc,eps)`(inner points in candidate simplices for the linear model with intercept based on small steps away from edges with check)
- `Tri_Mid(y,perc,candy)` (inner points in candidate simplices for the linear model with intercept based on center points of simplices)
- `Tri_Mid_n1(y,perc,candy)` (inner points in candidate simplices for the non-linear model based on center points of approximated simplices)
- `convex_hull_intersect(points1,points2,alpha,y1,y2,notion,plots)` (convex hull calculation to connect results from candidate evaluation to confidence region)
- `convex_hull_plot(x,y,col)` (function to plot convex hull based on points `x,y`)
- `straight_intersect(v1,v2,v3,v4,plots)` (calculation of the intersections from straights)
- `eps_ind(Mat,iX,eps)` (internal function to calculate points in simplices)
- `eps_ind_dist(v1,Mat,eps)` (internal function to calculate points in simplices)
- `mind(Mat,iX)` (function to calculate center of a simplex)

To implement the prediction method we make use of the `draw_from_depth(depthI,testvec,lower,upper)` function which allows us to draw from an empirical depth shape with restriction on a lower and upper bound.
The next set of functions is necessary to simulate the Gaussian limit process. In implements some of the functions which are necessary for the covariance matrix.

- `find1(x,pars)`
- `find2(x,pars)`
- `intfun2(t2,t1)`

The next two functions are used to speed up the computation of the complete simplicial depth by matrix operations.

- `inv_tri(Matrix)`
- `oner(resid)`

The `follow_ups(dat,mincper,steps,mincp)` function collects some calculations used in the change point detection procedure.
The `Ele_Norm(Cvec,center,pv,nortype,wgt)` function evaluates the distance of a point to another point with respect to a specified norm. It is used in the estimation procedure to speed up computation.
Finally the `dS_lin2_loop(theta,y)` function calculates full simplicial depth for the linear model with intercept by a simple loop implementation. It is used for runtime comparisons.
The complete package is available via
`https://github.com/ChrisKust/rexpar.git`.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Dissertation ist bisher keiner anderen Fakultät vorgelegt worden. Ich erkläre, dass ich bisher kein Promotionsverfahren erfolglos beendet habe und dass keine Aberkennung eines bereits erworbenen Doktorgrades vorliegt.

Christoph Falkenau