

Masterarbeit

Konsistenz und Anwendungen der Vorzeichen-Tiefe-Methode

Jan Krumnack

19. September 2023

Betreuung: Prof. Dr. Christine Müller

Zweitgutachter: JP Dr. Jan Nagel

Fakultät für Mathematik

TU Dortmund

Inhaltsverzeichnis

1	Einleitung	3
2	Grundlegende Motivation der Vorzeichen-Tiefe	6
3	Definitionen, Modell und Tests	8
3.1	Der 3-Vorzeichen-Tiefe-Test und andere Tests	9
4	Bedingungen für Konsistenz	14
4.1	Hinreichende Bedingungen für (C1)	18
5	Konsistenz in Modellen mit deterministischen, äquidistanten Kovariaten	20
5.1	Spezialfall des polynomiellen Modells	27
6	Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell	34
6.1	Möglichkeiten zur (approximativen) Berechnung von $q_\alpha(\Psi_3)$	34
6.2	Möglichkeiten zur (approximativen) Berechnung von $\sup_{\theta \in \Theta_0} \Psi_3(w^N(\theta))$.	36
6.2.1	Gittersuche	36
6.2.2	Datenpunktpaar-Verfahren	38
6.3	Berechnung der 3-Vorzeichen-Tiefe	46
6.4	Vergleich von dem Vorgehen in Pseudocode 2 und 4	48
6.5	Niveau des praktisch implementierten 3-Vorzeichen-Tiefe-Tests	50
6.6	Vergleich von verschiedenen praktischen Tests	52
6.6.1	Robustheit in Hinsicht auf die Fehlerwahrscheinlichkeit 1. Art . .	55
6.6.2	Güte	56
6.6.3	Konsistenz	58

1 Einleitung

In dieser Arbeit beschäftige ich mich mit der Konsistenz und den Anwendungen der Vorzeichen-Tiefe-Methode. Im ersten Teil, welcher hauptsächlich auf den Resultaten von Leckey et al., 2022, basiert, beschäftige ich mich eher mit den theoretischen Resultaten und Beweisen, die ich für die Konsistenzaussagen zum sogenannten *3-Vorzeichen-Tiefe-Test* benötige, während ich im zweiten Teil die Anwendung dieses Tests in der Praxis und dafür in Frage kommende Vorgehensweisen und Methoden behandle.

Die folgende Literaturübersicht basiert auf der Einführung von Leckey et al., 2022.

Literaturübersicht

Rousseeuw und Hubert, 1999, schlugen den Begriff der *Regressionstiefe* als Maß für die Anpassung einer Regressionsfunktion an einen bestimmten Datensatz vor. Dieses Anpassungsmaß basiert auf dem Konzept des *Nonfits*, wonach eine Regressionsfunktion ein Nonfit ist, wenn es eine andere Regressionsfunktion desselben Regressionsmodells gibt, die kleinere Residuen zum Quadrat für alle Datenpunkte liefert. Die Regressionstiefe einer Regressionsfunktion ist dann die Anzahl der Datenpunkte, die entfernt werden müssen, damit die gegebene Regressionsfunktion für die verbleibenden Datenpunkte ein Nonfit ist. Rousseeuw und Hubert, 1999, schlugen auch eine *simpliziale Regressionstiefe* vor, indem sie die Anzahl der Teilmengen der Datenpunkten mit $p + 1$ Elementen zählten, die zu keinem Nonfit führen, wenn $p \in \mathbb{N}$ die Dimension des Regressionsparameters ist. Der Name *simpliziale Regressionstiefe* stammt von der *simplizialen Tiefe* von Liu, 1988, 1990, für die Tiefe eines Lageparameters $\mu \in \mathbb{R}^p$ in einem multivariaten Datensatz der Dimension p . Diese *simpliziale Tiefe* zählt die Anzahl der von $p + 1$ Datenpunkten aufgespannten Simplexe, die den Lageparameter μ enthalten. Die Eigenschaft, dass μ in einem von $p + 1$ Datenpunkten aufgespannten Simplex liegt, ist jedoch gleichbedeutend mit der Eigenschaft, dass Tukeys (Tukey, 1975) *Halbraumtiefe* von μ in Bezug auf diese $p + 1$ Datenpunkte größer als Null ist. Ersetzt man die Halbraumtiefe mit der Regressionstiefe in Lius Definition der simplizialen Tiefe, so erhält man den Begriff der *simplizialen Regressionstiefe*.

Alle Begriffe für simpliziale Tiefen sind U-Statistiken, sodass ihre asymptotische Verteilung abgeleitet werden kann, siehe z.B. Liu, 1990, Dümbgen, 1992, Arcones und Gine, 1993, Müller, 2005, Wellmann et al., 2009, Wellmann und Müller, 2010. Darüber hinaus stellten Rousseeuw und Hubert, 1999, fest, dass sich die Berechnung der simplizialen Regressionstiefe in einigen polynomiellen Modellen auf das Zählen der Teilmengen mit alternierenden Vorzeichen der Residuen reduziert, was die Berechnung erleichtert. Kustosz et al., 2016, stellten hinreichende Bedingungen für lineare und nichtlineare Modelle für diese angenehme Eigenschaft auf. Insbesondere bewiesen sie, dass die Regressionstiefe einer Regressionsfunktion, die durch einen p -dimensionalen Regressionsparameter gegeben ist, durch die relative Anzahl von $p + 1$ -Tupeln von Residuen mit alternierenden Vorzeichen der Residuen gegeben ist. Dies führt zu der Idee, die Tiefe nur über die relative Anzahl von K -Tupeln mit alternierenden Vorzeichen der Residuen zu definieren, wobei K nicht auf $p + 1$ beschränkt ist, d.h. K ist nicht unbedingt durch die Dimension des unbekanntes Regressionsparameters gegeben. Dieser Tiefenbegriff wird

1 Einleitung

K -Vorzeichen-Tiefe oder kurz K -Tiefe genannt.

Auf der Grundlage der K -Vorzeichen-Tiefe definierten Leckey et al., 2023, sogenannte K -Vorzeichen-Tiefe-Tests, bei denen eine Nullhypothese abgelehnt wird, wenn die K -Vorzeichen-Tiefe der Parameter in der Nullhypothese zu klein ist. Dies sind sehr einfache Tests, die robust gegenüber Ausreißern sind, da sie auf Vorzeichen basieren. Wenn der Stichprobenumfang N klein ist, können die kritischen Werte außerdem leicht genau berechnet werden, indem alle 2^N möglichen Vorzeichenkombinationen berücksichtigt werden. Für größere Stichprobenumfänge wird jedoch die asymptotische Verteilung der K -Vorzeichen-Tiefe benötigt, die mit Standardmethoden nicht abgeleitet werden kann. Der Grund dafür ist, dass die K -Vorzeichen-Tiefe keine U-Statistik mehr ist, da die Reihenfolge innerhalb der K -Tupel für die Eigenschaft der alternierenden Vorzeichen relevant ist. Zunächst leiteten Falkenau et al., 2014, die asymptotische Verteilung der K -Vorzeichen-Tiefe für $K = 3$ ab und dann Malcherczyk et al., 2021, für allgemeines K .

Ein weiterer ursprünglicher Nachteil war der Aufwand der Berechnung der Tiefe der K -Vorzeichen-Tiefe, die auch für alle Begriffe der simplizialen Tiefe gilt. Wenn nämlich N die Stichprobengröße ist, hat eine naive Berechnung einen Aufwand der Größenordnung N^K , da $\binom{N}{K}$ Tupel berücksichtigt werden müssen. Die Herleitung der asymptotischen Verteilung in Malcherczyk et al., 2021, führt jedoch zu einer Darstellung der K -Vorzeichen-Tiefe, die in linearer Zeit bzgl. N berechnet werden kann. Darüber hinaus fand Malcherczyk, 2022, eine Blockimplementierung, die viel schneller ist als die Implementierung auf der Grundlage der asymptotischen Darstellung, siehe auch Leckey et al., 2023. Ein dritter ursprünglicher Nachteil war, dass eine Ordnung für die Residuen erforderlich ist. Eine natürliche Ordnung existiert für Zeitreihen und univariate erklärende Variablen, aber nicht für multivariate erklärende Variablen. Für diese Situation untersuchten Horn, 2021b, Horn und Müller, 2020, verschiedene Ordnungsmethoden und stellten fest, dass die Ordnung entlang eines kürzesten Pfades durch die beobachteten erklärenden Variablen zu den effizientesten Tests führt.

Ein offenes Problem war bisher, warum die K -Vorzeichen-Tiefe-Tests für $K \geq 3$ so effizient sind, wie in Kustosz et al., 2016, Falkenau et al., 2014, Leckey et al., 2023, Horn, 2021b, Horn und Müller, 2020, Malcherczyk, 2022, durch Simulationen gezeigt. Dies ist insbesondere deshalb überraschend, weil Leckey et al., 2023, gezeigt haben, dass die K -Vorzeichen-Tiefe-Tests mit $K = 2$, d.h. die 2-Vorzeichen-Tiefe-Tests, äquivalent zu den klassischen Vorzeichentests sind, die nur die Anzahl der positiven oder negativen Residuen zählen, siehe z.B. Gibbons und Chakraborti, 2003. Diese klassischen Vorzeichentests sind nicht in der Lage, bestimmte Alternativen der Nullhypothese abzulehnen, nämlich jene Alternativen, bei denen die erwartete Anzahl der positiven und negativen Vorzeichen der Residuen gleich ist. Infolgedessen sind die klassischen Vorzeichentests bei diesen Alternativen keine konsistenten Tests, d.h. die Trennschärfe dieser Tests bei diesen Alternativen konvergiert bei wachsendem Stichprobenumfang nicht gegen 1. Solche Alternativen treten bereits bei der linearen Regression auf. Leckey et al., 2022, zeigten jedoch, dass die K -Vorzeichen-Tiefe-Tests mit $K = 3$, d.h. 3-Vorzeichen-Tiefe-Tests, konsistente Tests für mehrere relevanten Hypothesen in allgemeinen polynomiellen Modellen sind. Insbesondere zeigten sie, dass K nicht mit der Dimension p des unbekanntem Regressionsparameters verknüpft sein muss, d.h. dass $K = 3$ immer ausreicht.

1 Einleitung

Sie bewiesen dies, indem sie zunächst eine allgemeine Bedingung für die Konsistenz der 3-Vorzeichen-Tiefe-Tests ableiteten.

Zunächst gebe ich in Kapitel 2 eine grundlegende Motivation für die Betrachtung der Vorzeichen-Tiefe, wobei ich dort der Einfachheit halber lediglich ein lineares Regressionsmodell betrachte. Diese Motivation basiert auf einem Dokument, welches ich durch persönliche Kommunikation von Malcherczyk erhalten habe und die Grundlagen und die wesentliche Aufgabenstellung meiner Beschäftigung mit Forschungsarbeiten zu der Vorzeichen-Tiefe im Wintersemester 21/22 darlegt.

In Kapitel 3 definiere ich dann vorzeichenbasierte Tests für ein generelles Regressionsmodell, wobei der 3-Vorzeichen-Tiefe-Test ein Spezialfall solcher Tests ist. Für diesen Test wird die Definition eines zeitstetigen stochastischen Prozesses $\mathcal{W}^N(\theta)$ benötigt, welcher von dem vorgeschlagenen Parameter θ abhängt und lediglich auf den Vorzeichen der zugehörigen Residuen basiert. Die Konvergenzeigenschaften dieses stochastischen Prozesses motivieren dann die Formulierung der vorzeichenbasierten Tests und mit der Beziehung zwischen dem Prozess und der 3-Vorzeichen-Tiefe folgt insbesondere der 3-Vorzeichen-Tiefe-Test. Wie oben schon angedeutet, stammt dieser sehr spezielle Zugang zu dem 3-Vorzeichen-Tiefe-Test über generelle vorzeichenbasierte Tests von Leckey et al., 2022, und er wird benötigt für die späteren Aussagen und Beweise zu der Konsistenz.

In Kapitel 4 leite ich allgemeine Bedingungen für die Konsistenz des 3-Vorzeichen-Tiefe-Tests für das allgemeine Modell und die allgemeine Nullhypothese des Kapitels 3 her. Mein Ansatz ist es, aus dem asymptotischen Verhalten von $\|\mathcal{W}^N(\theta)\|_\infty$ die Konsistenz zu folgern. In Abschnitt 4.1 leite ich dann wiederum hinreichende Bedingungen für das gewünschte asymptotische Verhalten von $\|\mathcal{W}^N(\theta)\|_\infty$ her. Hierbei habe ich die Beweise von Leckey et al., 2022, nachvollzogen, wobei mir einige Unstimmigkeiten aufgefallen sind. So betrachten sie \mathbb{R}^d -wertige Zufallsvariablen als erklärende Variablen, erläutern aber nicht ausreichend, wie genau nun diese Variablen geordnet werden sollen. Um diese Problematik zu umgehen, definiere ich die Variablen als univariate, (fast sicher) natürlich geordnete Zufallsvariablen. Außerdem fiel mir auf, dass bei der Definition vom Ereignis $A_{N,\varepsilon}(\theta)$ Null-Residuen nicht mitbedacht wurden, was daraufhin korrigiert wurde. In der ursprünglichen Definition überprüfte dieses Ereignis, ob Indexbereiche existieren, in denen Residuen dazu neigen, entweder atypisch viele oder wenige positive Vorzeichen zu haben. Die Korrektur besteht darin, dass ich nun nicht an atypisch vielen oder wenigen positiven Vorzeichen interessiert bin, sondern an atypisch vielen positiven oder atypisch vielen negativen Vorzeichen.

In Kapitel 5 behandle ich in einem spezielleren Modell mit deterministischen und äquidistanten Kovariaten weitere Bedingungen für die Konsistenz des 3-Vorzeichen-Tiefe-Tests. Ich betrachte dann in Abschnitt 5.1 das polynomielle Modell als Spezialfall. Es zeigt sich, dass diese weiteren Bedingungen erfüllt sind für bestimmte Nullhypothesen, welche besagen, dass das Modell gegeben ist durch ein Polynom mit einem Grad, der durch eine gegebene obere Schranke begrenzt ist. Leckey et al., 2022, betrachteten unabhängige, identisch verteilte Zufallsvariablen, die einen Träger $[S, T] \subset \mathbb{R}$ und eine stetige Verteilung haben. Deswegen definierten sie die 3-Vorzeichen-Tiefe für die Residu-

en erst nach der Ordnung der univariaten erklärenden Variablen nach ihrer natürlichen Ordnung. Da die Ordnung dann abhängig vom Zufall ist und ich dies als problematisch empfand, definiere ich die erklärenden Variablen als deterministisch und äquidistant. Aufgrund dieser Änderung, musste ich den Beweis von Satz 5.2 selbstständig in relativ großem Ausmaße anpassen.

Bei der Anwendung des 3-Vorzeichen-Tiefe-Tests in Simulationsstudien in Kapitel 6 betrachte ich lediglich ein quadratisches Modell und eine Nullhypothese, welche besagt, dass das Modell durch ein lineares Polynom gegeben ist. Ich erläutere Methoden, die Teststatistik, welche durch ein Supremum gegeben ist, und den kritischen Wert (annähernd) zu berechnen. Die Berechnung des Supremums der Tiefe hat noch keiner in einer größeren Menge untersucht. Dazu habe ich verschiedene Methoden entwickelt. Zunächst ist die sogenannte Gittersuche naheliegend. Diese Methode ist aber im Vergleich zu dem sogenannten Datenpunktpaar-Methode eher unvorteilhaft. Bei dieser Methode betrachte ich Geraden durch Paare von Datenpunkten. Da wir dabei aber stets zwei Null-Residuen erhalten und Tripel mit mindestens einem Null-Residuum nicht bei der 3-Vorzeichen-Tiefe mitgezählt werden, erläutere ich eine alternative Methode, bei der ich für jede im regulären Datenpunktpaar-Verfahren betrachtete Gerade vier leicht verschobene Geraden betrachte, sodass ich Null-Residuen vermeide. Diese Idee stammt von Falkenau, 2016. Da bei dieser alternativen Methode nicht klar ist, in welcher Größenordnung die Geraden verschoben werden sollen, besteht meine wesentliche alternative Methode darin, die Null-Residuen einfach mit den Werten $+1$ oder -1 zu ersetzen, was ausreichend kleinen Verschiebungen der Gerade in der ersten alternativen Methode entspricht. Die zweite alternative Methode vergleichen wir dann mit der regulären Version des Datenpunktpaar-Verfahrens. Eine weitere Fragestellung besteht darin, wie genau dieses Verfahren durchgeführt werden sollte, damit es das geforderte Niveau des 3-Vorzeichen-Tiefe-Tests in der Praxis einhält. Zuletzt vergleichen wir den mit Hilfe des Datenpunktpaar-Verfahrens durchgeführten 3-Vorzeichen-Tiefe-Test mit einem bereits in \mathbb{R} implementierten, robusten und relativ komplexen Test und dem klassischen F -Test in Hinsicht auf Güte und Konsistenz.

2 Grundlegende Motivation der Vorzeichen-Tiefe

Wir betrachten der Einfachheit halber zunächst die Modellgleichung

$$y_n = \theta_0^* + \theta_1^* x_n + e_n, \quad n = 1, \dots, N$$

für den unbekanntem wahren Parameter $(\theta_0^*, \theta_1^*) \in \mathbb{R}^2$. (In den folgenden Abschnitten der Masterarbeit betrachten wir dann allgemeinere Modellgleichungen.) Für die Regressoren (x_1, \dots, x_N) nehmen wir an, dass sie natürlich sortiert sind, d.h.

$$x_1 < \dots < x_N.$$

2 Grundlegende Motivation der Vorzeichen-Tiefe

Die Modellfehler e_1, \dots, e_N sind Realisierungen von Zufallsvariablen E_1, \dots, E_N mit den Eigenschaften

$$E_1, \dots, E_N \text{ sind stochastisch unabhängig,} \quad (2.1)$$

$$\mathbb{P}(E_n > 0) = \mathbb{P}(E_n < 0) = \frac{1}{2} \text{ für alle } n = 1, \dots, N. \quad (2.2)$$

Für jeden Parameter $\theta = (\theta_0, \theta_1) \in \mathbb{R}^2$ definieren wir die Residuen

$$r_n(\theta) := y_n - (\theta_0 + \theta_1 x_n), \quad n = 1, \dots, N.$$

Die Vorzeichen-Tiefe bewertet einen vorgeschlagenen Parameter $\theta \in \mathbb{R}^2$ anhand der Residuen zum vorgeschlagenen Parameter. Die K -Vorzeichen-Tiefe

$$d_K(r_1(\theta), \dots, r_N(\theta)) = \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \dots < n_K \leq N} \left(\prod_{k=1}^K \mathbb{1}\{r_{n_k}(\theta)(-1)^k > 0\} + \prod_{k=1}^K \mathbb{1}\{r_{n_k}(\theta)(-1)^k < 0\} \right)$$

gibt den relativen Anteil der geordneten K -Tupel aus den Residuen an, bei denen die Vorzeichen alternierend sind, d.h. es gibt $K - 1$ Vorzeichenwechsel bzw. eine Struktur der Form $(+ - + \dots)$ oder $(- + - \dots)$. Dabei ist $K \in \mathbb{N} \setminus \{1\}$ ein frei wählbarer Hyper-Parameter.

Je höher der Wert der Tiefe für einen Parameter θ ist, desto passender (bzw. tiefer) wird dieser Parameter verstanden. Dies geht damit einher, dass die Anzahl der Vorzeichenwechsel dann höher ist, was plausibel unter den Annahmen (2.1) und (2.2) ist. Die untere Abbildung 1 illustriert, warum eine zu geringe Anzahl an Vorzeichen-Wechseln gegen die Annahmen und damit gegen den betrachteten Parameter θ spricht.

Die blauen Punkte entsprechen den beobachteten Daten. Links wird das Modell zum wahren Parameter (gestrichelt schwarze Linie links) zu den Daten vorgeschlagen, rechts ein Modell zu einem falschen Parameter (gestrichelt schwarze Linie rechts). Die Residuen entsprechen den roten Strichen, die sich aus dem Abstand von den blauen Punkten und der schwarz gestrichelten Linie ergeben. Die Vorzeichen der Residuen sind oben angegeben.

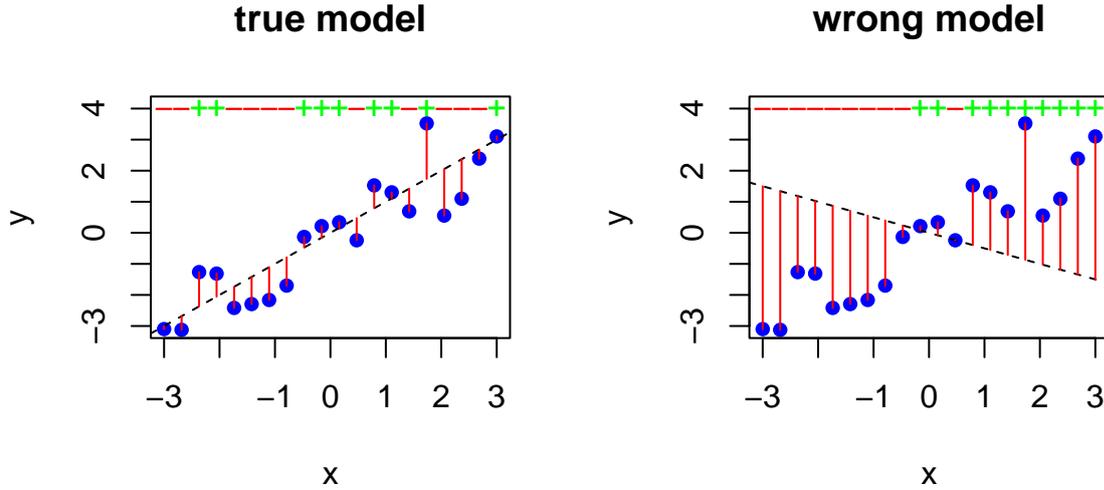


Abbildung 1: Vergleich von zwei Modellen

Rechts sind deutlich weniger Vorzeichenwechsel sichtbar als links. Insbesondere spricht die Struktur der Vorzeichen gegen die Annahme (2.1), da klare Block-Strukturen, d.h. Abhängigkeiten, in den Vorzeichen erkennbar sind. Anhand dieser Abbildung können wir ebenfalls sofort erkennen, dass im Falle $K > 4$ wir sogar eine K -Vorzeichen-Tiefe von Null für den falschen Parameter erhalten.

3 Definitionen, Modell und Tests

In diesem Kapitel führen wir vorzeichenbasierte Tests für ein allgemeines Modell und die Definitionen, die dafür benötigt werden, ein. Ein Spezialfall für einen solchen Test ist der 3-Vorzeichen-Tiefe-Test. Die Konsistenz dieses Tests und die Bedingungen dafür werden wir dann auch in den folgenden Kapiteln studieren.

Notation 3.1. *In einem parametrischen Modell mit einem Parameter $\theta^* \in \Theta$ notieren wir das Wahrscheinlichkeitsmaß und den Erwartungswert mit \mathbb{P}_{θ^*} und \mathbb{E}_{θ^*} . Damit drücken wir aus, dass wir den wahren Modellparameter als θ^* annehmen. Wenn die Verteilung einer Zufallsvariable nicht vom Modellparameter θ^* abhängt, nutzen wir \mathbb{P} und \mathbb{E} stattdessen.*

Mit $\mathbf{1}$ bezeichnen wir die Indikatorfunktion. Dies ist eine $\{0, 1\}$ -wertige Funktion, die gleich eins ist genau dann, wenn die Bedingung in der geschwungenen Klammer erfüllt ist.

Wir bezeichnen eine Funktion $f : [0, 1] \rightarrow \mathbb{R}$ als Càdlàg-Funktion, wenn die linksseitigen Limiten $\lim_{x \uparrow y} f(x)$ existieren für alle $y \in (0, 1]$ und f rechtsstetig ist, d.h. $\lim_{x \downarrow y} f(x) = f(y)$ für alle $y \in [0, 1)$. Die Menge aller Càdlàg-Funktionen mit Definitionsbereich $[0, 1]$ bezeichnen wir als Skorochod-Raum $\mathcal{D}[0, 1]$.

Zunächst definieren wir das sehr allgemeine Modell, das wir betrachten. Das Regressionsmodell hat die Form

$$Y_n = f(X_{n;N}, \theta^*) + E_n, \quad n = 1, \dots, N, \quad (3.1)$$

wobei Y_n die abhängige, reellwertige Zufallsvariable ist, $f : \mathbb{R} \times \mathbb{R}^p \rightarrow \mathbb{R}$ eine reellwertige Funktion, $\theta^* \in \Theta \subset \mathbb{R}^p$, $p \in \mathbb{N}$, der Modellparameter und die Störgrößen E_1, \dots, E_N unabhängige, reellwertige Zufallsvariablen mit $\mathbb{P}(E_n > 0) = 0.5 = \mathbb{P}(E_n < 0)$. Die erklärenden Variablen $X_{1;N}, X_{2;N}, \dots, X_{N;N}$ sind reellwertige Zufallsvariablen für die wir eine natürliche Ordnung

$$X_{1;N} < X_{2;N} < \dots < X_{N;N} \quad \mathbb{P}\text{-f.s.} \quad (3.2)$$

annehmen. In dem Modell (3.1) definieren wir das n -te Residuum abhängig vom Parameter $\theta \in \Theta$ als

$$R_n(\theta) := Y_n - f(X_{n;N}, \theta), \quad n = 1, \dots, N. \quad (3.3)$$

Da für den wahren Parameter θ^* $R_n(\theta^*) = E_n$ gilt, folgt aus den Annahmen für die Störgrößen E_1, \dots, E_N , dass

$$R_1(\theta^*), R_2(\theta^*), \dots, R_N(\theta^*) \text{ sind unabhängig bzgl. } \mathbb{P}_{\theta^*}, \quad (\text{A1})$$

$$\mathbb{P}_{\theta^*}(R_n(\theta^*) > 0) = 0.5 = \mathbb{P}_{\theta^*}(R_n(\theta^*) < 0) \text{ für alle } n = 1, \dots, N. \quad (\text{A2})$$

3.1 Der 3-Vorzeichen-Tiefe-Test und andere Tests

In diesem Abschnitt definieren wir Tests für Hypothesen der Form

$$H_0 : \theta^* \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta^* \in \Theta \setminus \Theta_0, \quad (3.4)$$

wobei $\Theta_0 \subset \Theta$ eine beliebige Teilmenge ist und θ^* der Modellparameter des parametrischen Modells (3.1). Der Einfachheit betrachten wir zunächst den Spezialfall der Einpunkt-Hypothese, wobei $\Theta_0 = \{\theta^0\}$ mit $\theta^0 \in \Theta$, d.h.

$$H_0 : \theta^* = \theta^0 \quad \text{vs.} \quad H_1 : \theta^* \neq \theta^0. \quad (3.5)$$

Mir $\psi(x)$ bezeichnen wir das Vorzeichen von $x \in \mathbb{R}$, d.h.

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

Wenn wir die Hypothese H_0 in (3.5) als wahr annehmen, so sind $\psi(R_1(\theta^0)), \dots, \psi(R_N(\theta^0))$ unabhängig und diskret gleichverteilt mit Träger $\{-1, +1\}$ nach (A1) und (A2). Wir definieren nun $\mathcal{W}^N(\theta) = (\mathcal{W}_t^N(\theta))_{t \in [0,1]}$ mit

$$\mathcal{W}_t^N(\theta) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor tN \rfloor} \psi(R_n(\theta)), \quad t \in [0, 1], \quad (3.6)$$

wobei wir $\mathcal{W}_t^N(\theta) = 0$ für $t < 1/N$ als Konvention setzen. Im Falle $\theta^* = \theta^0$ ist also $\mathcal{W}^N(\theta^0)$ der Pfad einer reskalierten symmetrischen einfachen Irrfahrt. In dem Fall konvergiert $\mathcal{W}^N(\theta^0)$ nach dem Donskerschen Invarianzprinzip in Verteilung gegen eine Standard-Brownsche Bewegung W (eingeschränkt auf den Zeitbereich $[0, 1]$) im Skorochod-Raum $\mathcal{D}[0, 1]$ (Billingsley, 1999, Theorem 14.1). Dies ist die Motivation dafür, dass wir Tests herleiten, die überprüfen inwiefern Eigenschaften von $\mathcal{W}^N(\theta^0)$ sich signifikant von denen der Brownschen Bewegung unterscheiden. Wir betrachten eine stetige Funktion

$$\Psi : \mathcal{D}[0, 1] \rightarrow \mathbb{R},$$

und definieren einen Test für die Einpunkt-Hypothese (3.5) durch

$$\text{Lehne } H_0 \text{ ab, falls } \Psi(\mathcal{W}^N(\theta^0)) < q_\alpha(\Psi), \quad (3.7)$$

wobei $\alpha \in (0, 1)$ das gegebene Signifikanzniveau und $q_\alpha(\Psi)$ das α -Quantil von $\Psi(W)$ mit Standard-Brownscher Bewegung W ist. Auf ähnliche Weise testen wir die allgemeinen Hypothesen (3.4) durch

$$\text{Lehne } H_0 \text{ ab, falls } \sup_{\theta \in \Theta_0} \Psi(\mathcal{W}^N(\theta)) < q_\alpha(\Psi), \quad (3.8)$$

s. Müller, 2005. Dies ist ein asymptotischer Test zum Niveau α , da nach Continuous Mapping Theorem

$$\begin{aligned} \limsup_{N \rightarrow \infty} \mathbb{P}_{\theta^*}(\sup_{\theta \in \Theta_0} \Psi(\mathcal{W}^N(\theta)) < q_\alpha(\Psi)) &\leq \limsup_{N \rightarrow \infty} \mathbb{P}_{\theta^*}(\Psi(\mathcal{W}^N(\theta^*)) \leq q_\alpha(\Psi)) \\ &= \lim_{N \rightarrow \infty} \mathbb{P}_{\theta^*}(\Psi(\mathcal{W}^N(\theta^*)) \leq q_\alpha(\Psi)) \\ &= \mathbb{P}_{\theta^*}(\Psi(W) \leq q_\alpha(\Psi)) \\ &= \alpha \end{aligned} \quad (3.9)$$

für alle $\theta^* \in \Theta_0$.

Bemerkung 3.2. Wir betrachten den Skorochod-Raum $\mathcal{D}[0, 1]$ versehen mit der Skorochod-Metrik

$$d_S(w, v) := \inf_{\lambda \in \Lambda} \max\{\|\lambda - \text{id}\|_\infty, \|w \circ \lambda - v\|_\infty\},$$

wobei Λ die Menge aller stetigen, bijektiven, streng monoton wachsenden Funktionen $\lambda : [0, 1] \rightarrow [0, 1]$ ist, $\text{id} : [0, 1] \rightarrow [0, 1]$, $t \mapsto t$ die Identität auf $[0, 1]$ und $\|\cdot\|_\infty$ die Supremumsnorm auf $[0, 1]$. Für den linksseitigen Limes im Punkt t eines Elements $w = (w_t)_{t \in [0, 1]} \in \mathcal{D}[0, 1]$ schreiben wir $w_{t-} := \lim_{s \uparrow t} w_s$. Es gelten die folgenden Fakten für den Skorochod-Raum $\mathcal{D}[0, 1]$ und dessen Elemente.

(i) Für eine Folge $(w^{(n)})_{n \in \mathbb{N}}$ in $\mathcal{D}[0, 1]$ und $w \in \mathcal{D}[0, 1]$ gilt $w^{(n)} \xrightarrow{n \rightarrow \infty} w$ in $\mathcal{D}[0, 1]$ genau dann, wenn eine Folge $\lambda^{(n)}$ in Λ existiert mit

$$\sup_{t \in [0, 1]} |\lambda^{(n)}(t) - t| \xrightarrow{n \rightarrow \infty} 0 \quad \text{und} \quad \sup_{t \in [0, 1]} |w_{\lambda^{(n)}(t)} - w_t| \xrightarrow{n \rightarrow \infty} 0.$$

- (ii) Die Funktion $\chi : \mathcal{D}[0, 1] \rightarrow \mathbb{R}$, $w \mapsto \inf_{t \in [0, 1]} w_t$ ist wohldefiniert und stetig. Für den Nachweis der Wohldefiniertheit nehmen wir $\chi(w) = -\infty$ für ein $w \in \mathcal{D}[0, 1]$ an. Dann existiert für alle $n \in \mathbb{N}$ ein $t_n \in [0, 1]$ mit $w_{t_n} < -n$. Damit existiert eine Teilfolge $(t_{n_m})_{m \in \mathbb{N}}$ von $(t_n)_{n \in \mathbb{N}}$ und ein $t_0 \in [0, 1]$, sodass $t_{n_m} \uparrow t_0$ (oder $t_{n_m} \downarrow t_0$) für $m \rightarrow \infty$. Damit folgt

$$w_{t_{n_m}} \xrightarrow{m \rightarrow \infty} -\infty,$$

was heißt, dass der linksseitige (oder rechtsseitige) Limes von w in t_0 nicht existiert. Dies führt uns zum Widerspruch. Für den Nachweis der Stetigkeit betrachten wir wieder eine Folge $(w^{(n)})_{n \in \mathbb{N}}$ in $\mathcal{D}[0, 1]$ und $w \in \mathcal{D}[0, 1]$ mit $w^{(n)} \xrightarrow{n \rightarrow \infty} w$ in $\mathcal{D}[0, 1]$. Außerdem sei eine Folge $(\lambda^{(n)})_{n \in \mathbb{N}}$ wie in (i) gegeben. Durch die Anwendung der bekannten allgemeinen Ungleichung $\sup_x (f(x) + g(x)) \leq \sup_x f(x) + \sup_x g(x)$ können wir sehr leicht die allgemeine Ungleichung

$$\left| \inf_x f(x) - \inf_x g(x) \right| \leq \sup_x |f(x) - g(x)|$$

herleiten. Damit folgt

$$\begin{aligned} |\chi(w^{(n)}) - \chi(w)| &= \left| \inf_{t \in [0, 1]} w_t^{(n)} - \inf_{t \in [0, 1]} w_t \right| \\ &= \left| \inf_{t \in [0, 1]} w_t^{(n)} - \inf_{t \in [0, 1]} w_{\lambda^{(n)}(t)} \right| \\ &\leq \sup_{t \in [0, 1]} |w_t^{(n)} - w_{\lambda^{(n)}(t)}| \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Hieraus folgt somit auch die Stetigkeit von

$$\begin{aligned} \mathcal{D}[0, 1] \ni w \mapsto \|w\|_\infty &= \max\left\{-\inf_{t \in [0, 1]} w_t, \sup_{t \in [0, 1]} w_t\right\} \\ &= \max\left\{-\inf_{t \in [0, 1]} w_t, -\inf_{t \in [0, 1]} (-w_t)\right\}. \end{aligned}$$

- (iii) Für alle $w \in \mathcal{D}[0, 1]$ ist die Menge der Sprungstellen $\{t \in (0, 1] \mid |w_t - w_{t-}| > 0\}$ höchstens abzählbar. Dafür nehmen wir an, dass für ein $\varepsilon > 0$ die Menge $S(\varepsilon) := \{t \in (0, 1] \mid |w_t - w_{t-}| > \varepsilon\}$ unendlich ist. Dann existiert also eine Folge $(t_n)_{n \in \mathbb{N}}$ in $S(\varepsilon)$ und ein $t_0 \in [0, 1]$ mit $t_n \uparrow t_0$ (oder $t_n \downarrow t_0$) für $n \rightarrow \infty$ und $|w_{t_n} - w_{t_n-}| > \varepsilon$ für alle $n \in \mathbb{N}$. Dann existieren $t_{n-1} < s_n < t_n$ (oder $t_{n-1} > s_{n-1} > t_n$) für alle $n \in \mathbb{N} \setminus \{1\}$, sodass $|w_{t_n} - w_{s_n}| > \varepsilon/2$ für alle $n \in \mathbb{N} \setminus \{1\}$. Wenn wir dann aber die Folge

$$(\tilde{t}_1, \tilde{t}_2, \dots) := (t_1, s_2, t_2, s_3, \dots) \quad (\text{oder } (\tilde{t}_1, \tilde{t}_2, \dots) := (t_1, s_1, t_2, s_2, \dots))$$

betrachten, so folgt $|w_{\tilde{t}_n} - w_{\tilde{t}_{n+1}}| > \varepsilon/2$ für alle geraden n (oder für alle ungeraden n). Dies bedeutet aber, dass der linksseitige (oder rechtsseitige) Limes in t_0 nicht existiert, was zum Widerspruch führt. Somit ist also

$$\{t \in (0, 1] \mid |w_t - w_{t-}| > 0\} = \bigcup_{n \in \mathbb{N}} \left\{ t \in (0, 1] \mid |w_t - w_{t-}| > \frac{1}{n} \right\}$$

als abzählbare Vereinigung von endlichen Mengen höchstens abzählbar.

(iv) Sei $t \in [0, 1]$. Die Funktion $\pi_t : \mathcal{D}[0, 1] \rightarrow \mathbb{R}$, $w \mapsto w_t$ ist stetig in allen $w \in \mathcal{D}[0, 1]$, die stetig in t sind. Betrachte also ein $w \in \mathcal{D}[0, 1]$, das stetig in t ist, und eine Folge $(w^{(n)})_{n \in \mathbb{N}}$ in $\mathcal{D}[0, 1]$ mit $w^{(n)} \xrightarrow{n \rightarrow \infty} w$ in $\mathcal{D}[0, 1]$. Außerdem sei eine Folge $(\lambda^{(n)})_{n \in \mathbb{N}}$ wie in (i) gegeben. Dann gilt

$$\begin{aligned} |\pi_t(w^{(n)}) - \pi_t(w)| &= |w_t^{(n)} - w_t| \\ &\leq |w_t^{(n)} - w_{\lambda^{(n)}(t)}| + |w_{\lambda^{(n)}(t)} - w_t|, \end{aligned} \quad (3.10)$$

wobei der erste Summand in (3.10) kleiner gleich $\|w^{(n)} - w \circ \lambda^{(n)}\|_\infty \rightarrow 0$ für $n \rightarrow \infty$ ist, und der zweite Summand aufgrund von $\lambda^{(n)}(t) \rightarrow t$ für $n \rightarrow \infty$ und der Stetigkeit von w in t auch gegen 0 konvergiert für $n \rightarrow \infty$. Somit haben wir also die Stetigkeit von π_t in w gezeigt.

Beispiel 3.3. Ein mögliches Beispiel für Ψ ist die Betrachtung der Supremumsnorm von $\mathcal{W}^N(\theta)$. Dafür definieren wir

$$\Psi_\infty(W) := -\|W\|_\infty := -\sup_{t \in [0, 1]} |W_t|$$

für einen Prozess $W = (W_t)_{t \in [0, 1]}$. $\Psi_\infty : \mathcal{D}[0, 1] \rightarrow \mathbb{R}$ ist nach Bemerkung 3.2 (ii) stetig, d.h. nach dem Continuous Mapping Theorem konvergiert $\Psi_\infty(\mathcal{W}^N(\theta^*))$ in Verteilung gegen $\Psi_\infty(W)$ für eine Standard-Brownsche Bewegung W , falls θ^* der wahre Modellparameter ist. Man könnte also die Tests (3.7) und (3.8) mit $\Psi = \Psi_\infty$ definieren.

Definition 3.4 (*K*-Vorzeichen-Tiefe). Die 3-Vorzeichen-Tiefe ist ein Spezialfall der *K*-Vorzeichen-Tiefe, welche für einen Vektor $(r_1, \dots, r_N) \in \mathbb{R}^N$ die relative Häufigkeit an geordneten *K*-Tupel aus dem Vektor angibt, bei denen die Vorzeichen alternierend sind. Für ein $K \in \mathbb{N} \setminus \{1\}$ definieren wir die *K*-Vorzeichen-Tiefe von (r_1, \dots, r_N) durch

$$\begin{aligned} d_K(r_1, \dots, r_N) &:= \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \dots < n_K \leq N} \mathbb{1}\{(r_{n_1}, \dots, r_{n_K}) \in \mathcal{A}_{N, K}\}, \\ \mathcal{A}_{N, K} &:= \{(z_1, \dots, z_K) \in \mathbb{R}^K \mid z_i \cdot z_{i+1} < 0 \text{ für alle } i = 1, \dots, K-1\}. \end{aligned}$$

Bemerkung 3.5. Die obige Darstellung der *K*-Vorzeichen-Tiefe ist in der Tat identisch zu der in Kapitel 2.

Der *K*-Vorzeichen-Tiefe-Test für die Hypothesen (3.4) ist gegeben durch

$$\text{Lehne } H_0 \text{ ab, falls } \sup_{\theta \in \Theta_0} d_K(R_1(\theta), \dots, R_N(\theta)) < q_{K, \alpha},$$

wobei $q_{K, \alpha}$ das α -Quantil von $d_K(R_1(\theta^*), \dots, R_N(\theta^*))$ ist.

Da jedoch diese Darstellung der 3-Vorzeichen-Tiefe relativ kompliziert ist, betrachten wir eine andere Darstellung. Es lässt sich nämlich zeigen, dass die 3-Vorzeichen-Tiefe

bis auf eine lineare Transformation äquivalent zu $\Psi_3(\mathcal{W}^N(\theta))$ ist, wobei $\mathcal{W}^N(\theta)$ in (3.6) gegeben ist und

$$\Psi_3(\mathcal{W}^N(\theta)) := \frac{3}{4} \left(1 - \int_0^1 (\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta))^2 dt \right). \quad (3.11)$$

In Malcherczyk et al., 2021, Theorem 2.3, wurde genauer gezeigt, dass

$$\frac{(N-1)(N-2)}{N} (d_3(R_1(\theta), \dots, R_N(\theta)) - 0.25) = \Psi_3(\mathcal{W}^N(\theta)) \quad \mathbb{P}_\theta\text{-f.s.}, \quad (3.12)$$

falls $R_n(\theta) \neq 0$ \mathbb{P}_θ -f.s. für alle n . Somit ist also der 3-Vorzeichen-Tiefe-Test ein Test der Form (3.8).

$\Psi_3 : \mathcal{D}[0, 1] \rightarrow \mathbb{R}$ ist stetig. Betrachte dafür eine Folge $(w^{(n)})_{n \in \mathbb{N}}$ in $\mathcal{D}[0, 1]$ und $w \in \mathcal{D}[0, 1]$ mit $w^{(n)} \xrightarrow{n \rightarrow \infty} w$ in $\mathcal{D}[0, 1]$. Außerdem sei eine Folge $(\lambda^{(n)})_{n \in \mathbb{N}}$ mit

$$\sup_{t \in [0, 1]} |\lambda^{(n)}(t) - t| \xrightarrow{n \rightarrow \infty} 0 \quad \text{und} \quad \sup_{t \in [0, 1]} |w_{\lambda^{(n)}(t)} - w_t| \xrightarrow{n \rightarrow \infty} 0$$

gegeben. Nach Bemerkung 3.2 (iii) und (iv) existiert eine höchstens abzählbare Menge $E \subset [0, 1]$, sodass $w_t^{(n)} \rightarrow w_t$ für $n \rightarrow \infty$ für alle $t \in [0, 1] \setminus E$, und somit

$$w_t^{(n)} \xrightarrow{n \rightarrow \infty} w_t \quad \text{für fast alle } t \in [0, 1].$$

Da $\lambda(1) = 1$ für alle $\lambda \in \Lambda$, gilt nach Bemerkung 3.2 (i) insbesondere $w_1^{(n)} \rightarrow w_1$ für $n \rightarrow \infty$. Zusammen erhalten wir also

$$(w_1^{(n)} - 2w_t^{(n)})^2 \xrightarrow{n \rightarrow \infty} (w_1 - 2w_t)^2 \quad \text{für fast alle } t \in [0, 1].$$

Nun reicht es aufgrund des Satzes von der majorisierten Konvergenz aus zu zeigen, dass die Funktionenfolge $([0, 1] \ni t \mapsto (w_1^{(n)} - 2w_t^{(n)})^2)_{n \in \mathbb{N}}$ eine integrierbare Majorante besitzt. Es gilt

$$\begin{aligned} (w_1^{(n)} - 2w_t^{(n)})^2 &\leq 9 \|w^{(n)}\|_\infty^2 \\ &\leq 9 (\|w^{(n)} - w \circ \lambda^{(n)}\|_\infty + \underbrace{\|w \circ \lambda^{(n)}\|_\infty}_{=\|w\|_\infty})^2. \end{aligned} \quad (3.13)$$

Der erste Summand in (3.13) ist beschränkt (bzgl. n), da er nach Voraussetzung gegen 0 konvergiert. Der zweite Summand ist endlich, da w eine Càdlàg-Funktion ist und somit auch beschränkt sein muss. (In Bemerkung 3.2 haben wir gezeigt, dass das Infimum einer Càdlàg-Funktion endlich ist. Auf analoge Weise kann man auch die Endlichkeit des Supremums herleiten.) Mit dem Satz von der majorisierten Konvergenz erhalten wir also

$$\lim_{n \rightarrow \infty} \int_0^1 (w_1^{(n)} - 2w_t^{(n)})^2 dt = \int_0^1 \lim_{n \rightarrow \infty} (w_1^{(n)} - 2w_t^{(n)})^2 dt = \int_0^1 (w_1 - 2w_t)^2 dt,$$

woraus nach Darstellung von Ψ_3 die Stetigkeit folgt.

4 Bedingungen für Konsistenz

Bemerkung 3.6. In Malcherczyk et al., 2021 wurde sogar gezeigt, dass für beliebige $K \geq 4$ die K -Vorzeichen-Tiefe asymptotisch äquivalent ist zu $\Psi(\mathcal{W}^N(\theta))$, wobei Ψ eine passende Funktion ist. Falls die Residuen (A1) und (A2) erfüllen für $\theta = \theta^*$, so gilt

$$N \left(d_K(R_1(\theta), \dots, R_N(\theta)) - \left(\frac{1}{2}\right)^{K-1} \right) = \Psi_K(\mathcal{W}^N(\theta)) + o_{\mathbb{P}_\theta}(1),$$

wobei $o_{\mathbb{P}_\theta}(1)$ eine Zufallsvariable ist, die stochastisch gegen 0 konvergiert, und

$$\begin{aligned} \Psi_K(\mathcal{W}^N) = & \mathcal{C}_K - \frac{\langle K \rangle_4}{4} \int_{-0.5}^1 \int_{t \vee 0}^{t+0.5} \left(\frac{1}{2} + t - s\right)^{K-4} (\mathcal{W}_{s \wedge 1}^N - \mathcal{W}_{t \vee 0}^N)^2 ds dt \\ & - \frac{\langle K \rangle_4}{2} \int_{0.5}^1 \int_0^{t-0.5} \left(\frac{1}{2} + s - t\right)^{K-4} \mathcal{W}_s^N (\mathcal{W}_1^N - \mathcal{W}_t^N) ds dt. \end{aligned}$$

Dabei ist $\mathcal{C}_K > 0$ eine Konstante und $\langle K \rangle_4 = K!/(K-4)!$. Aufgrund dieser Darstellung, welche komplizierter ist als Ψ_3 , ist aber nicht klar, ob ähnliche Ergebnisse wie in Kapitel 4 auch für Tests gelten, die für $K \geq 4$ auf $\Psi_K(\mathcal{W}^N(\theta))$ basieren.

4 Bedingungen für Konsistenz

Ein Test φ_N zum Niveau α für die Nullhypothese $H_0 : \theta \in \Theta_0$ heißt konsistent genau dann, wenn

$$\lim_{N \rightarrow \infty} \mathbb{P}_{\theta^*}(\varphi_N \text{ lehnt } H_0 \text{ ab}) = 1 \text{ für alle } \theta^* \notin \Theta_0.$$

In diesem Kapitel leiten wir allgemeine Bedingungen für die Konsistenz des 3-Vorzeichen-Tiefe-Tests her. Ein wichtiger Ansatz für die Folgerung der Konsistenz ist das asymptotische Verhalten von $\|\mathcal{W}^N(\theta)\|_\infty$. Dieser Ansatz wird in dem folgenden Satz verfolgt, welcher unser Hauptresultat ist.

Satz 4.1. Sei Ψ_3 gegeben wie in (3.11) und $\mathcal{W}^N(\theta)$ wie in (3.6). Falls für alle $\theta^* \notin \Theta_0$

$$\mathbb{P}_{\theta^*} \left(\liminf_{N \rightarrow \infty} \frac{\inf_{\theta \in \Theta_0} \|\mathcal{W}^N(\theta)\|_\infty}{\sqrt{N}} > 0 \right) = 1, \quad (\text{C1})$$

dann

$$\mathbb{P}_{\theta^*} \left(\limsup_{N \rightarrow \infty} \sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) = -\infty \right) = 1.$$

Wir können Satz 4.1 beweisen, indem wir zeigen, dass aus (C1) die Divergenz des Integrand in der Definition (3.11) von Ψ_3 gegen unendlich auf einem beliebigen Intervall positiver Länge folgt. Ein wichtiges Resultat für den Beweis von Satz 4.1 ist das folgende Lemma, das wir später beweisen werden.

Lemma 4.2. Wenn (C1) gilt für ein $\theta^* \notin \Theta_0$, so gilt

$$\mathbb{P}_{\theta^*} \left(\lim_{N \rightarrow \infty} \inf_{\theta \in \Theta_0} \sup_{\substack{(a,b) \in [0,1]^2 \\ a < b}} (b-a) \inf_{t \in [a,b]} (\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta))^2 = \infty \right) = 1. \quad (4.1)$$

Beweis von Satz 4.1. Nach Lemma 4.2 können wir annehmen, dass (4.1) für alle $\theta^* \notin \Theta_0$ gilt. Zunächst gilt

$$\begin{aligned} \int_0^1 (\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta))^2 dt &\geq \int_a^b (\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta))^2 dt \\ &\geq (b-a) \inf_{t \in [a,b]} (\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta))^2 \end{aligned}$$

für alle $N \in \mathbb{N}$, $\theta \in \Theta_0$ und $(a,b) \in [0,1]^2$ mit $a < b$. Wenn man also zunächst das Supremum bzgl. der a, b und dann das Infimum bzgl. $\theta \in \Theta_0$ anwendet, erhält man

$$\begin{aligned} &\left\{ \lim_{N \rightarrow \infty} \inf_{\theta \in \Theta_0} \sup_{\substack{(a,b) \in [0,1]^2 \\ a < b}} (b-a) \inf_{t \in [a,b]} (\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta))^2 = \infty \right\} \\ &\subset \left\{ \lim_{N \rightarrow \infty} \inf_{\theta \in \Theta_0} \int_0^1 (\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta))^2 dt = \infty \right\}. \end{aligned}$$

Wegen der Definition (3.11) folgt also

$$\mathbb{P}_{\theta^*} \left(\lim_{N \rightarrow \infty} \sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) = -\infty \right) = 1$$

für alle $\theta^* \notin \Theta_0$. □

Der Beweis von Lemma 4.2 basiert auf der folgenden Lipschitzbedingung von $\mathcal{W}^N(\theta)$.

Lemma 4.3. Für alle $N \in \mathbb{N}$ und $\theta \in \Theta$ gilt

$$|\mathcal{W}_t^N(\theta) - \mathcal{W}_s^N(\theta)| \leq |t-s|\sqrt{N} + \frac{1}{\sqrt{N}} \text{ für alle } s, t \in [0, 1].$$

Beweis. Sei o.B.d.A. $s < t$. Nach Definition ist $(\mathcal{W}_t^N(\theta))_{t \in [0,1]}$ stückweise konstant mit Sprunghöhen $1/\sqrt{N}$, 0 oder $-1/\sqrt{N}$ an den Stellen $t \in \{1/N, 2/N, \dots, N/N = 1\}$ (und sonst keinen Sprüngen). Es lässt sich sehr leicht graphisch herleiten, dass somit die Anzahl an Sprüngen auf Intervall $[s, t]$, $s < t$, höchstens $\lfloor (t-s)N \rfloor + 1$ ist. Somit gilt

$$|\mathcal{W}_t^N(\theta) - \mathcal{W}_s^N(\theta)| \leq (\lfloor (t-s)N \rfloor + 1) \frac{1}{\sqrt{N}} \leq |t-s|\sqrt{N} + \frac{1}{\sqrt{N}}.$$

□

4 Bedingungen für Konsistenz

Korollar 4.4. Sei $\varepsilon \in (0, 8]$, $\theta \in \Theta$ und $N \in \mathbb{N}$ mit $\sqrt{N} \geq 8/\varepsilon$. Wenn $w^N = (w_t^N)_{t \in [0,1]}$ eine Realisierung von $\mathcal{W}^N(\theta)$ mit

$$\|w^N\|_\infty \geq \varepsilon\sqrt{N}$$

ist, so existiert ein Intervall $[a, b] \subset [0, 1]$, sodass $b - a \geq \varepsilon/8$

$$\inf_{t \in [a, b]} |w_t^N| \geq \frac{3}{4} \|w^N\|_\infty.$$

Beweis. Da $w^N = (w_t^N)_{t \in [0,1]}$ stückweise konstant ist, wird das Supremum an einem Punkt T angenommen, d.h. es existiert ein $T \in [0, 1]$, sodass

$$w_T^N = \|w^N\|_\infty \geq \varepsilon\sqrt{N}.$$

Nach Lemma 4.3 gilt

$$|w_T^N - w_s^N| \leq |T - s|\sqrt{N} + \frac{1}{\sqrt{N}} \leq \frac{\varepsilon}{8}\sqrt{N} + \frac{\varepsilon}{8} \leq \frac{\varepsilon}{4}\sqrt{N} \leq \frac{\|w^N\|_\infty}{4}$$

für alle $s \in [0, 1]$ mit $|s - T| \leq \varepsilon/8$. Wenn wir nun $a = \max\{T - \varepsilon/8, 0\}$ und $b = \min\{T + \varepsilon/8, 1\}$ setzen, gilt $|s - T| \leq \varepsilon/8$ für alle $s \in [a, b]$. Aufgrund der umgekehrten Dreiecksungleichung folgt also

$$\begin{aligned} |w_s^N| &\geq |w_T^N| - |w_T^N - w_s^N| \\ &\geq \|w^N\|_\infty - \frac{\|w^N\|_\infty}{4} \\ &= \frac{3}{4} \|w^N\|_\infty \end{aligned}$$

für alle $s \in [a, b]$, und somit auch

$$\inf_{s \in [a, b]} |w_s^N| \geq \frac{3}{4} \|w^N\|_\infty.$$

□

Beweis von Lemma 4.2. Für die Behauptung von Lemma 4.2 genügt es zu zeigen, dass aus dem Ereignis in (C1) das Ereignis in (4.1) folgt. Seien also $w^N(\theta) = (w_t^N(\theta))_{t \in [0,1]}$ Realisierungen von $\mathcal{W}^N(\theta)$ für $\theta \in \Theta_0$ und $N \in \mathbb{N}$ mit

$$\liminf_{N \rightarrow \infty} \frac{\inf_{\theta \in \Theta_0} \|w^N(\theta)\|_\infty}{\sqrt{N}} > 0.$$

Dann existieren $N_0 \in \mathbb{N}$ und $\varepsilon > 0$, sodass

$$\|w^N(\theta)\|_\infty \geq \varepsilon\sqrt{N} \text{ für alle } N \geq N_0 \text{ und } \theta \in \Theta_0. \quad (4.2)$$

4 Bedingungen für Konsistenz

Da wir in der Bedingung (4.2) $N_0 \in \mathbb{N}$ beliebig vergrößern und $\varepsilon > 0$ beliebig verkleinern können, kann $\varepsilon \leq 8$ und $N_0 \geq 8/\varepsilon$ angenommen werden. Nach Korollar 4.4 existiert für alle $N \geq N_0$ ein Intervall $[a_N, b_N]$, sodass

$$b_N - a_N \geq \frac{\varepsilon}{8} \quad \text{und} \quad \inf_{t \in [a_N, b_N]} |w_t^N(\theta)| \geq \frac{3}{4} \|w^N(\theta)\|_\infty. \quad (4.3)$$

Somit folgt mit der umgekehrten Dreiecksungleichung für alle $t \in [a_N, b_N]$

$$\begin{aligned} |w_1^N(\theta) - 2w_t^N(\theta)| &\geq 2|w_t^N(\theta)| - |w_1^N(\theta)| \\ &\stackrel{(4.3)}{\geq} \frac{6}{4} \|w^N(\theta)\|_\infty - \|w^N(\theta)\|_\infty \\ &= \frac{1}{2} \|w^N(\theta)\|_\infty \\ &\stackrel{(4.2)}{\geq} \frac{\varepsilon}{2} \sqrt{N}. \end{aligned}$$

Also folgt für alle $N \geq N_0$

$$\begin{aligned} \sup_{\substack{(a,b) \in [0,1]^2 \\ a < b}} (b-a) \inf_{t \in [a,b]} (w_1^N(\theta) - 2w_t^N(\theta))^2 &\geq (b_N - a_N) \inf_{t \in [a_N, b_N]} (w_1^N(\theta) - 2w_t^N(\theta))^2 \\ &\geq \frac{\varepsilon}{8} \left(\frac{\varepsilon}{2} \sqrt{N} \right)^2 \\ &= \frac{\varepsilon^3}{32} N. \end{aligned}$$

Da die Wahl von $N_0 \in \mathbb{N}$ und $\varepsilon > 0$ nicht von θ abhängt, erhalten wir schließlich

$$\begin{aligned} \liminf_{N \rightarrow \infty} \inf_{\theta \in \Theta_0} \sup_{\substack{(a,b) \in [0,1]^2 \\ a < b}} (b-a) \inf_{t \in [a,b]} (w_1^N(\theta) - 2w_t^N(\theta))^2 &\geq \lim_{N \rightarrow \infty} \frac{\varepsilon^3}{32} N \\ &= \infty. \end{aligned}$$

Somit folgt aus dem Ereignis in (C1) das Ereignis in (4.1), und Lemma 4.2 ist daher bewiesen. \square

Korollar 4.5. *Wenn (C1) gilt für alle $\theta^* \notin \Theta_0$, so ist der 3-Vorzeichen-Tiefe-Test konsistent, d.h.*

$$\lim_{N \rightarrow \infty} \mathbb{P}_{\theta^*} \left(\sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_3) \right) = 1 \quad \text{für alle } \alpha \in (0, 1] \text{ und } \theta^* \notin \Theta_0.$$

4 Bedingungen für Konsistenz

Beweis. Dieses Resultat folgt sofort aus Satz 4.1 und

$$\begin{aligned}
\mathbb{P}_{\theta^*}(\lim_{N \rightarrow \infty} \sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) = -\infty) &= \mathbb{P}_{\theta^*} \left(\bigcap_{L>0} \bigcup_{N_0 \in \mathbb{N}} \bigcap_{N \geq N_0} \sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < -L \right) \\
&\leq \mathbb{P}_{\theta^*} \left(\bigcup_{N_0 \in \mathbb{N}} \bigcap_{N \geq N_0} \sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_3) \right) \\
&= \lim_{N_0 \rightarrow \infty} \mathbb{P}_{\theta^*} \left(\bigcap_{N \geq N_0} \sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_3) \right) \\
&\leq \lim_{N_0 \rightarrow \infty} \mathbb{P}_{\theta^*} \left(\sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^{N_0}(\theta)) < q_\alpha(\Psi_3) \right),
\end{aligned}$$

wobei wir hier die σ -Stetigkeit verwendet haben. □

4.1 Hinreichende Bedingungen für (C1)

Wir können die Gültigkeit der Bedingung (C1) herleiten, indem wir Intervalle $[a, b]$, $a < b$, betrachten, bei denen der absolute Zuwachs $|\mathcal{W}_b^N(\theta) - \mathcal{W}_a^N(\theta)|$ besonders groß ist. Dafür definieren wir

$$\mathcal{N}_{k,l}^+ := \sum_{j=k}^l \mathbb{1}\{R_j(\theta) > 0\}, \quad \mathcal{N}_{k,l}^- := \sum_{j=k}^l \mathbb{1}\{R_j(\theta) < 0\}. \quad (4.4)$$

Dann ist das Ereignis $A_{N,\varepsilon}$ für $\varepsilon > 0$ eine hinreichende Bedingung für große Werte von $\|\mathcal{W}^N(\theta)\|_\infty$, wobei

$$A_{N,\varepsilon}(\theta) := \left\{ \sup_{\substack{k,l \in \{1, \dots, N\} \\ l-k+1 \geq \varepsilon N}} \frac{\max\{\mathcal{N}_{k,l}^+, \mathcal{N}_{k,l}^-\}}{l-k+1} - \frac{1}{2} > \varepsilon \right\}. \quad (4.5)$$

Lemma 4.6. *Sei $\varepsilon > 0$ und $N \in \mathbb{N}$. Falls*

$$\sup_{\substack{k,l \in \{1, \dots, N\} \\ l-k+1 \geq \varepsilon N}} \frac{\max\{\mathcal{N}_{k,l}^+, \mathcal{N}_{k,l}^-\}}{l-k+1} - \frac{1}{2} > \varepsilon,$$

so gilt $\|\mathcal{W}^N(\theta)\|_\infty \geq \varepsilon^2 \sqrt{N}$, d.h. $A_{N,\varepsilon}(\theta) \subset \{\|\mathcal{W}^N(\theta)\|_\infty \geq \varepsilon^2 \sqrt{N}\}$.

Beweis von Lemma 4.6. Es gelte $A_{N,\varepsilon}(\theta)$. Es existieren also $k, l \in \{1, \dots, N\}$ mit $l - k + 1 \geq \varepsilon N$, sodass

$$\frac{\max\{\mathcal{N}_{k,l}^+, \mathcal{N}_{k,l}^-\}}{l-k+1} - \frac{1}{2} > \varepsilon.$$

Wir werden im Folgenden lediglich den Fall

$$\frac{\mathcal{N}_{k,l}^+}{l-k+1} > \frac{1}{2} + \varepsilon \quad (4.6)$$

4 Bedingungen für Konsistenz

betrachten, da der Beweis für den anderen Fall analog folgt. Nun gilt

$$\begin{aligned}\mathcal{W}_{\frac{l}{N}}^N(\theta) - \mathcal{W}_{\frac{k-1}{N}}^N(\theta) &= \frac{1}{\sqrt{N}} \sum_{n=k}^l \psi(R_n(\theta)) \\ &= \frac{1}{\sqrt{N}} (\mathcal{N}_{k,l}^+(\theta) - \mathcal{N}_{k,l}^-(\theta)) \\ &\geq \frac{1}{\sqrt{N}} (\mathcal{N}_{k,l}^+(\theta) - ((l-k+1) - \mathcal{N}_{k,l}^+(\theta))),\end{aligned}$$

wobei die Ungleichung sogar eine Gleichung ist, wenn alle Residuen ungleich Null sind. Somit folgt

$$\begin{aligned}\mathcal{W}_{\frac{l}{N}}^N(\theta) - \mathcal{W}_{\frac{k-1}{N}}^N(\theta) &\geq \frac{1}{\sqrt{N}} (2\mathcal{N}_{k,l}^+(\theta) - (l-k+1)) \\ &\stackrel{(4.6)}{\geq} \frac{1}{\sqrt{N}} \left(2(l-k+1) \left(\frac{1}{2} + \varepsilon \right) - (l-k+1) \right) \\ &= 2 \frac{1}{\sqrt{N}} \varepsilon (l-k+1) \\ &\geq 2\varepsilon^2 \sqrt{N},\end{aligned}$$

da $l-k+1 \geq \varepsilon N$. Wegen der Dreiecksungleichung gilt nun

$$|\mathcal{W}_{\frac{l}{N}}^N(\theta) - \mathcal{W}_{\frac{k-1}{N}}^N(\theta)| \leq 2 \|\mathcal{W}^N(\theta)\|_{\infty},$$

und deswegen

$$\|\mathcal{W}^N(\theta)\|_{\infty} \geq \frac{1}{2} |\mathcal{W}_{\frac{l}{N}}^N(\theta) - \mathcal{W}_{\frac{k-1}{N}}^N(\theta)| \geq \varepsilon^2 \sqrt{N}.$$

Somit haben wir Lemma 4.6 für den Fall (4.6) gezeigt. Für den Fall

$$\frac{\mathcal{N}_{k,l}^-}{l-k+1} > \frac{1}{2} + \varepsilon$$

betrachten wir $\mathcal{W}_{\frac{k-1}{N}}^N(\theta) - \mathcal{W}_{\frac{l}{N}}^N(\theta)$ statt $\mathcal{W}_{\frac{l}{N}}^N(\theta) - \mathcal{W}_{\frac{k-1}{N}}^N(\theta)$ und führen die obigen Abschätzungen auf analoge Weise durch. \square

Korollar 4.7. *Sei $\theta^* \notin \Theta_0$. Falls ein $\varepsilon > 0$ existiert, sodass*

$$\mathbb{P}_{\theta^*} \left(\liminf_{N \rightarrow \infty} \bigcap_{\theta \in \Theta_0} A_{N,\varepsilon}(\theta) \right) = 1, \tag{C2}$$

so gilt (C1). Falls (C2) sogar für alle $\theta^ \notin \Theta_0$ gilt, so ist der 3-Vorzeichen-Tiefe-Test konsistent.*

Beweis. Nach Lemma 4.6 gilt

$$\begin{aligned}
 \liminf_{N \rightarrow \infty} \bigcap_{\theta \in \Theta_0} A_{N,\varepsilon}(\theta) &\subset \liminf_{N \rightarrow \infty} \bigcap_{\theta \in \Theta_0} \{ \|\mathcal{W}^N(\theta)\|_\infty \geq \varepsilon^2 \sqrt{N} \} \\
 &= \liminf_{N \rightarrow \infty} \left\{ \inf_{\theta \in \Theta_0} \frac{\|\mathcal{W}^N(\theta)\|_\infty}{\sqrt{N}} \geq \varepsilon^2 \right\} \\
 &= \left\{ \liminf_{N \rightarrow \infty} \inf_{\theta \in \Theta_0} \frac{\|\mathcal{W}^N(\theta)\|_\infty}{\sqrt{N}} \geq \varepsilon^2 \right\} \\
 &\subset \left\{ \liminf_{N \rightarrow \infty} \inf_{\theta \in \Theta_0} \frac{\|\mathcal{W}^N(\theta)\|_\infty}{\sqrt{N}} > 0 \right\}.
 \end{aligned}$$

Die zweite Aussage aus Korollar 4.7 folgt somit direkt aus Satz 4.1. \square

5 Konsistenz in Modellen mit deterministischen, äquidistanten Kovariaten

In diesem Abschnitt betrachten wir ein Modell der Form $Y_n = f(x_{n;N}, \theta^*) + E_n$, $n = 1, \dots, N$, wobei

$$f(\cdot, \theta) : \mathbb{R} \rightarrow \mathbb{R} \text{ ist stetig für alle } \theta \in \Theta \subset \mathbb{R}^p, \quad (5.1)$$

$$E_1, \dots, E_N \text{ sind reellwertige, unabhängige Zufallsvariablen für alle } N \in \mathbb{N}, \quad (5.2)$$

die Kovariaten $x_{1;N}, x_{2;N}, \dots, x_{N;N}$ sind deterministisch, äquidistant auf $[S, T]$, $S < T$,

$$\text{d.h. } x_{k;N} = S + \frac{k-1}{N-1}(T-S) \text{ für alle } k = 1, \dots, N, \quad (5.3)$$

E_1, \dots, E_N sind identisch verteilt für alle $N \in \mathbb{N}$ mit $\mathbb{P}(E_n > 0) = 0.5 = \mathbb{P}(E_n < 0)$

und E_n hat eine Verteilungsfunktion, die streng monoton wachsend ist

$$\text{auf einem Intervall } [-\delta_0, \delta_0] \text{ für ein } \delta_0 > 0. \quad (5.4)$$

Ein Spezialfalls dieses Modells ist das polynomielle Modell, welches wir in Abschnitt 5.1 betrachten werden. Eine benötigte Definition für unser Hauptresultat in diesem Kapitel beschäftigt sich mit den Vorzeichenwechseln einer Funktion g auf $[S, T]$.

Definition 5.1. Sei $I \in \mathbb{N}$.

(i) $g : [S, T] \rightarrow \mathbb{R}$ hat I Vorzeichenwechsel bei $S < z_1 < z_2 < \dots < z_I < T$, falls g alternierende Vorzeichen auf

$$B_1 := (S, z_1), B_2 := (z_1, z_2), \dots, B_I := (z_{I-1}, z_I), B_{I+1} := (z_I, T)$$

hat und die Funktionswerte auf B_i entweder echt positiv oder echt negativ sind, d.h. entweder $g(x) \cdot (-1)^i > 0$ für alle $i = 1, 2, \dots, I+1$ oder $g(x) \cdot (-1)^i < 0$ für alle $i = 1, 2, \dots, I+1$.

(ii) $g : [S, T] \rightarrow \mathbb{R}$ hat I schwache Vorzeichenwechsel bei $S < z_1 < z_2 < \dots < z_I < T$, falls g schwach alternierende Vorzeichen auf

$$B_1 := (S, z_1), B_2 := (z_1, z_2), \dots, B_I := (z_{I-1}, z_I), B_{I+1} := (z_I, T)$$

hat, d.h. entweder $g(x) \cdot (-1)^i \geq 0$ für alle $i = 1, 2, \dots, I+1$ oder $g(x) \cdot (-1)^i \leq 0$ für alle $i = 1, 2, \dots, I+1$.

Satz 5.2. Sei $\theta^* \notin \Theta_0$. Falls ein Parameter $\theta^0 \in \Theta_0$ existiert, sodass

$$g : [S, T] \rightarrow \mathbb{R} \text{ gegeben durch } g(x) = f(x, \theta^*) - f(x, \theta^0) \\ \text{mindestens } I \text{ Vorzeichenwechsel auf } [S, T] \text{ hat,} \quad (5.5)$$

und für alle $\theta \in \Theta_0$ mit $\theta \neq \theta^0$

$$h : [S, T] \rightarrow \mathbb{R} \text{ gegeben durch } h(x) = f(x, \theta^0) - f(x, \theta) \\ \text{höchstens } I - 1 \text{ schwache Vorzeichenwechsel auf } [S, T] \text{ hat,} \quad (5.6)$$

dann existiert ein $\varepsilon > 0$, sodass (C2) gilt. Falls sogar (5.5) und (5.6) für alle $\theta^* \notin \Theta_0$ gelten, so ist der 3-Vorzeichen-Tiefe-Test konsistent.

Wir benutzen für den Beweis von Satz 5.2 das folgende Lemma.

Lemma 5.3. Falls $g : [S, T] \rightarrow \mathbb{R}$ I Vorzeichenwechsel auf

$$B_1 := (S, z_1), B_2 := (z_1, z_2), \dots, B_I := (z_{I-1}, z_I), B_{I+1} := (z_I, T)$$

hat und $h : [S, T] \rightarrow \mathbb{R}$ höchstens $I - 1$ schwache Vorzeichenwechsel, so existiert ein $i \in \{1, \dots, I+1\}$ mit

$$\text{entweder } g(x) + h(x) \geq g(x) > 0 \text{ für alle } x \in B_i \\ \text{oder } g(x) + h(x) \leq g(x) < 0 \text{ für alle } x \in B_i. \quad (5.7)$$

Beweis. Wir beweisen die Aussage mit vollständiger Induktion bzgl. $I \in \mathbb{N} \setminus \{1\}$.

Induktionsanfang. Für den Induktionsanfang betrachten wir $I = 2$, sodass h also höchstens einen schwachen Vorzeichenwechsel hat. Falls h keinen schwachen Vorzeichenwechsel hat, so ist entweder $h(x) \geq 0$ für alle $x \in [S, T]$ oder $h(x) \leq 0$ für alle $x \in [S, T]$. Somit gilt also (5.7) für entweder $i = 1$ oder $i = 2$. Wenn h einen schwachen Vorzeichenwechsel bei $z_0 \in [S, z_1]$ hat, so hat h keinen Vorzeichenwechsel auf $(z_1, T) = B_2 \cup B_3 \cup \{z_2, T\}$. Da g verschiedene Vorzeichen auf B_2 und B_3 hat, gilt (5.7) für $i = 2$ oder $i = 3$. Wenn h einen schwachen Vorzeichenwechsel bei $z_0 \in [z_2, T]$ hat, verläuft die Argumentation analog. Wenn h einen schwachen Vorzeichenwechsel bei $z_0 \in [z_1, z_2]$ hat, so hat h verschiedene Vorzeichen auf B_1 und B_3 (im schwachen Sinne), während g gleiche Vorzeichen auf B_1 und B_3 hat. Somit gilt auch hier (5.7) für $i = 1$ oder $i = 3$.

Induktionsvoraussetzung. Die Behauptung von Lemma 5.3 gelte für ein $I \in \mathbb{N} \setminus \{1\}$ für alle $S < T$.

Induktionsschritt. Wir wollen nun die Behauptung von Lemma 5.3 für $I + 1$ beweisen. Sei also nun g gegeben mit $I + 1$ Vorzeichenwechseln und h mit höchstens I schwachen Vorzeichenwechseln. g habe also alternierende Vorzeichen auf

$$B_1 := (S, z_1), B_2 := (z_1, z_2), \dots, B_{I+1} := (z_I, z_{I+1}), B_{I+2} := (z_{I+1}, T).$$

Falls h einen schwachen Vorzeichenwechsel bei $z_0 \in (z_{I+1}, T]$ hat, so hat h höchstens $I - 1$ schwache Vorzeichenwechsel auf $[S, z_{I+1}]$. Da g I Vorzeichenwechsel auf $[S, z_{I+1}]$ hat, folgt (5.7) aus der Induktionsvoraussetzung mit $z_{I+1} \hat{=} T$. Falls h einen schwachen Vorzeichenwechsel bei $z_0 \in [S, z_1)$ hat, so folgt (5.7) ebenfalls aus der Induktionsvoraussetzung mit $z_1 \hat{=} S$.

Falls h keinen schwachen Vorzeichenwechsel auf $[S, z_1) \cup (z_{I+1}, T] = B_1 \cup B_{I+1} \cup \{S, T\}$ hat, so hat h also höchstens I schwache Vorzeichenwechsel auf $[z_1, z_{I+1}]$. Wenn h weniger als I schwache Vorzeichenwechsel auf $[z_1, z_{I+1}]$ hat, so folgt (5.7) aus der Induktionsvoraussetzung, indem wir $[S, z_{I+1}] \hat{=} [S, T]$ oder $[z_1, T] \hat{=} [S, T]$ betrachten. Wenn h genau I schwache Vorzeichenwechsel auf $[z_1, z_{I+1}]$ hat, so existieren $\tilde{z}_1, \dots, \tilde{z}_I \in [z_I, z_{I+1}]$, sodass h schwach alternierende Vorzeichen auf

$$C_1 := (S, \tilde{z}_1), C_2 := (\tilde{z}_1, \tilde{z}_2), \dots, C_I := (\tilde{z}_{I-1}, \tilde{z}_I), C_{I+1} := (\tilde{z}_I, T)$$

hat. Da h keine schwachen Vorzeichenwechsel auf $[S, z_1) \cup (z_{I+1}, T] = B_1 \cup B_{I+1} \cup \{S, T\}$ hat, muss $B_1 \subset C_1$ und $B_{I+2} \subset C_{I+1}$ gelten. Wenn I ungerade ist, so sind die Vorzeichen von g auf B_1 und B_{I+2} gleich, während die Vorzeichen (im schwachen Sinne) von h auf C_1 und C_{I+1} unterschiedlich sind. Wenn I gerade ist, so hat mit analoger Begründung g verschiedene Vorzeichen auf B_1 und B_{I+2} , und h hat gleiche Vorzeichen (im schwachen Sinne) auf C_1 und C_{I+1} . In den beiden beschriebenen Fällen gilt (5.7) für $i = 1$ oder $i = I + 2$. \square

Beweis von Satz 5.2. Nach Annahme (5.5) hat g gegeben durch $g(x) = f(x, \theta^*) - f(x, \theta^0)$ alternierende Vorzeichen auf

$$B_1 := (S, z_1), B_2 := (z_1, z_2), \dots, B_I := (z_{I-1}, z_I), B_{I+1} := (z_I, T).$$

Da entweder $g(x) > 0$ auf B_i oder $g(x) < 0$ auf B_i für alle $i = 1, \dots, I + 1$ und g stetig ist aufgrund von (5.1), existieren $\delta \in (0, \delta_0)$, $a_i < b_i$, $i = 1, \dots, I + 1$, mit $[a_i, b_i] \subset B_i$ für alle $i = 1, \dots, I + 1$ und

$$\text{entweder } g(x) > \delta \text{ auf } [a_i, b_i] \text{ oder } g(x) < -\delta \text{ auf } [a_i, b_i] \text{ für alle } i = 1, \dots, I + 1.$$

Nach (5.4) existiert ein $\tilde{\varepsilon} > 0$ mit

$$\mathbb{P}_{\theta^*}(E_n > -\delta) - \frac{1}{2} > \tilde{\varepsilon}, \quad \mathbb{P}_{\theta^*}(E_n < \delta) - \frac{1}{2} > \tilde{\varepsilon}$$

für alle $n = 1, \dots, N$, $N \in \mathbb{N}$. Wir definieren nun

$$N_i^+(\theta) := \sum_{n=1}^N \mathbb{1}\{R_n(\theta) > 0, x_{n,N} \in [a_i, b_i]\},$$

$$N_i^-(\theta) := \sum_{n=1}^N \mathbb{1}\{R_n(\theta) < 0, x_{n,N} \in [a_i, b_i]\}.$$

Zunächst zeigen wir, dass für beliebige $[a, b] \subset [S, T]$, $a < b$,

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{x_{n;N} \in [a, b]\} &\xrightarrow{N \rightarrow \infty} \frac{b-a}{T-S}, \\ \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{E_n > -\delta, x_{n;N} \in [a, b]\} &\xrightarrow{N \rightarrow \infty} \frac{b-a}{T-S} \mathbb{P}_{\theta^*}(E_1 > -\delta) \quad \mathbb{P}_{\theta^*}\text{-f.s.}, \text{ und} \\ \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{E_n < \delta, x_{n;N} \in [a, b]\} &\xrightarrow{N \rightarrow \infty} \frac{b-a}{T-S} \mathbb{P}_{\theta^*}(E_1 < \delta) \quad \mathbb{P}_{\theta^*}\text{-f.s.} \end{aligned}$$

Dafür definieren wir Intervalle $[a_{\subset}, b_{\subset}] \subset [a, b] \subset [a_{\supset}, b_{\supset}]$, dessen Grenzen in einer Form wie in (5.3) gegeben sind, d.h.

$$\begin{aligned} a_{\subset} = a_{\subset}^N &:= S + \left\lceil \frac{a-S}{T-S}(N-1) \right\rceil \frac{1}{N-1}(T-S), \\ a_{\supset} = a_{\supset}^N &:= S + \left\lfloor \frac{a-S}{T-S}(N-1) \right\rfloor \frac{1}{N-1}(T-S), \\ b_{\subset} = b_{\subset}^N &:= S + \left\lceil \frac{b-S}{T-S}(N-1) \right\rceil \frac{1}{N-1}(T-S), \\ b_{\supset} = b_{\supset}^N &:= S + \left\lfloor \frac{b-S}{T-S}(N-1) \right\rfloor \frac{1}{N-1}(T-S). \end{aligned}$$

Die folgende Abbildung 2 veranschaulicht diese Beziehung zwischen $[a, b]$, $[a_{\subset}, b_{\subset}]$ und $[a_{\supset}, b_{\supset}]$.

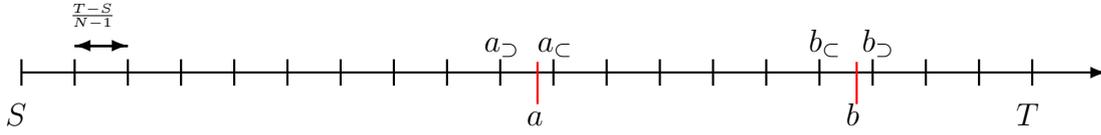


Abbildung 2: Beziehung von $[a, b]$, $[a_{\subset}, b_{\subset}]$ und $[a_{\supset}, b_{\supset}]$

Für den Beweis der Konvergenzen betrachten wir jeweils Abschätzungen nach oben und unten. Dabei können wir $\frac{1}{N} \sum_{n=1}^N \mathbb{1}\{x_{n;N} \in [a, b]\}$ nach oben abschätzen durch

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{x_{n;N} \in [a, b]\} &\leq \frac{1}{N} \sum_{n=1}^N \mathbb{1}\{x_{n;N} \in [a_{\supset}^N, b_{\supset}^N]\} \\ &= \frac{1}{N} \left(\left\lceil \frac{b-S}{T-S}(N-1) \right\rceil - \left\lfloor \frac{a-S}{T-S}(N-1) \right\rfloor + 1 \right) \\ &\xrightarrow{N \rightarrow \infty} \frac{b-a}{T-S}, \end{aligned}$$

und nach oben durch

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a, b]\} &\geq \frac{1}{N} \sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_C^N, b_C^N]\} \\ &= \frac{1}{N} \left(\left\lfloor \frac{b-S}{T-S}(N-1) \right\rfloor - \left\lfloor \frac{a-S}{T-S}(N-1) \right\rfloor + 1 \right) \\ &\xrightarrow{N \rightarrow \infty} \frac{b-a}{T-S}. \end{aligned}$$

Analog zeigen wir die Konvergenz von $\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{E_n > -\delta, x_{n;N} \in [a, b]\}$. Dies können wir nach oben abschätzen durch

$$\begin{aligned} &\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{E_n > -\delta, x_{n;N} \in [a, b]\} \\ &\leq \frac{1}{N} \sum_{n=1}^N \mathbf{1}\{E_n > -\delta, x_{n;N} \in [a_S^N, b_S^N]\} \\ &= \frac{1}{N} \left(\left\lfloor \frac{b-S}{T-S}(N-1) \right\rfloor - \left\lfloor \frac{a-S}{T-S}(N-1) \right\rfloor + 1 \right) \end{aligned} \quad (5.8)$$

$$\cdot \frac{1}{\left\lfloor \frac{b-S}{T-S}(N-1) \right\rfloor - \left\lfloor \frac{a-S}{T-S}(N-1) \right\rfloor + 1} \sum_{n=\left\lfloor \frac{a-S}{T-S}(N-1) \right\rfloor}^{\left\lfloor \frac{b-S}{T-S}(N-1) \right\rfloor} \mathbf{1}\{E_n > -\delta\}. \quad (5.9)$$

Die Konvergenz von dem Term in (5.8) gegen $(b-a)/(T-S)$ haben wir oben schon gezeigt. Die (fast sichere) Konvergenz des Terms in (5.9) gegen $\mathbb{P}_{\theta^*}(E_1 > -\delta)$ folgt wegen

$$\begin{aligned} \left\lfloor \frac{b-S}{T-S}(N-1) \right\rfloor - \left\lfloor \frac{a-S}{T-S}(N-1) \right\rfloor + 1 &\geq \frac{b-S}{T-S}(N-1) - \frac{a-S}{T-S}(N-1) - 1 \\ &= (N-1) \frac{b-a}{T-S} - 1 \\ &\xrightarrow{N \rightarrow \infty} \infty \end{aligned}$$

aus dem starken Gesetz der großen Zahlen und der unabhängigen, identischen Verteilung der E_1, \dots, E_N nach (5.2) und (5.4). Durch eine analoge Abschätzung nach unten erhalten wir also

$$\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{E_n > -\delta, x_{n;N} \in [a, b]\} \xrightarrow{N \rightarrow \infty} \frac{b-a}{T-S} \mathbb{P}_{\theta^*}(E_1 > -\delta) \quad \mathbb{P}_{\theta^*}\text{-f.s.}$$

Natürlich folgt die (fast sichere) Konvergenz von $\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{E_n < \delta, x_{n;N} \in [a, b]\}$ auf gleiche Weise.

Für alle i mit $g(x) > \delta$ auf $[a_i, b_i]$ gilt

$$\begin{aligned}
 N_i^+(\theta^0) &= \sum_{n=1}^N \mathbf{1}\{Y_n - f(x_{n;N}, \theta^0) > 0, x_{n;N} \in [a_i, b_i]\} \\
 &= \sum_{n=1}^N \mathbf{1}\{f(x_{n;N}, \theta^*) + E_n - f(x_{n;N}, \theta^0) > 0, x_{n;N} \in [a_i, b_i]\} \\
 &= \sum_{n=1}^N \mathbf{1}\{g(x_{n;N}) + E_n > 0, x_{n;N} \in [a_i, b_i]\} \\
 &\geq \sum_{n=1}^N \mathbf{1}\{E_n > -\delta, x_{n;N} \in [a_i, b_i]\},
 \end{aligned}$$

und somit folgt für diese i mit den oben gezeigten Konvergenzen

$$\begin{aligned}
 \frac{N_i^+(\theta^0)}{\sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_i, b_i]\}} - \frac{1}{2} &\geq \frac{\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{E_n > -\delta, x_{n;N} \in [a_i, b_i]\}}{\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_i, b_i]\}} - \frac{1}{2} \\
 &\xrightarrow{N \rightarrow \infty} \frac{\frac{b_i - a_i}{T - S} \mathbb{P}_{\theta^*}(E_1 > -\delta)}{\frac{b_i - a_i}{T - S}} - \frac{1}{2} \\
 &= \mathbb{P}_{\theta^*}(E_1 > -\delta) - \frac{1}{2} > \tilde{\varepsilon} \quad \mathbb{P}_{\theta^*}\text{-f.s.}
 \end{aligned}$$

Auf analoge Weise gilt für alle i mit $g(x) < -\delta$ auf $[a_i, b_i]$

$$\begin{aligned}
 N_i^-(\theta^0) &= \mathbf{1}\{g(x_{n;N}) + E_n < 0, x_{n;N} \in [a_i, b_i]\} \\
 &\geq \sum_{n=1}^N \mathbf{1}\{E_n < \delta, x_{n;N} \in [a_i, b_i]\},
 \end{aligned}$$

und

$$\begin{aligned}
 \frac{N_i^-(\theta^0)}{\sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_i, b_i]\}} - \frac{1}{2} &\geq \frac{\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{E_n < \delta, x_{n;N} \in [a_i, b_i]\}}{\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_i, b_i]\}} - \frac{1}{2} \\
 &\xrightarrow{N \rightarrow \infty} \frac{\frac{b_i - a_i}{T - S} \mathbb{P}_{\theta^*}(E_1 < \delta)}{\frac{b_i - a_i}{T - S}} - \frac{1}{2} \\
 &= \mathbb{P}_{\theta^*}(E_1 < \delta) - \frac{1}{2} > \tilde{\varepsilon} \quad \mathbb{P}_{\theta^*}\text{-f.s.}
 \end{aligned}$$

Sei im Folgenden

$$\varepsilon = \min \left\{ \min_{i \in \{1, \dots, I+1\}} \frac{b_i - a_i}{2(T - S)}, \tilde{\varepsilon} \right\}.$$

Nun existiert also ein Ereignis Ω_0 mit $\mathbb{P}_{\theta^*}(\Omega_0) = 1$, sodass für alle $\omega \in \Omega_0$ ein $N_\omega \in \mathbb{N}$

gibt mit

$$\frac{1}{N} \sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_i, b_i]\} > \varepsilon \text{ für alle } i = 1, \dots, I+1,$$

$$\frac{N_i^+(\theta^0)(\omega)}{\sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_i, b_i]\}} - \frac{1}{2} > \varepsilon \text{ für alle } i \text{ mit } g(x) > \delta \text{ auf } [a_i, b_i], \text{ und} \quad (5.10)$$

$$\frac{N_i^-(\theta^0)(\omega)}{\sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_i, b_i]\}} - \frac{1}{2} > \varepsilon \text{ für alle } i \text{ mit } g(x) < -\delta \text{ auf } [a_i, b_i] \quad (5.11)$$

für alle $N \geq N_\omega$.

Betrachte nun ein beliebiges $\theta \in \Theta_0$ mit $\theta \neq \theta^0$. Nach den Voraussetzungen von Satz 5.2 gelten (5.5) und (5.6), und somit existiert nach Lemma 5.3 ein $i^* \in \{1, \dots, I+1\}$, sodass entweder

$$f(x, \theta^*) - f(x, \theta) = g(x) + h(x) \geq g(x) > \delta \text{ auf } [a_{i^*}, b_{i^*}] \subset B_{i^*} \quad (5.12)$$

oder

$$f(x, \theta^*) - f(x, \theta) = g(x) + h(x) \leq g(x) < -\delta \text{ auf } [a_{i^*}, b_{i^*}] \subset B_{i^*}. \quad (5.13)$$

Im Falle (5.12) erhalten wir

$$\begin{aligned} N_{i^*}^+(\theta) &= \sum_{n=1}^N \mathbf{1}\{Y_n - f(x_{n;N}, \theta) > 0, x_{n;N} \in [a_{i^*}, b_{i^*}]\} \\ &= \sum_{n=1}^N \mathbf{1}\{g(x_{n;N}) + h(x_{n;N}) + E_n > 0, x_{n;N} \in [a_{i^*}, b_{i^*}]\} \\ &\geq \sum_{n=1}^N \mathbf{1}\{g(x_{n;N}) + E_n > 0, x_{n;N} \in [a_{i^*}, b_{i^*}]\} \\ &= N_{i^*}^+(\theta^0), \end{aligned}$$

und analog im Fall (5.13)

$$N_{i^*}^-(\theta) \geq N_{i^*}^-(\theta^0).$$

Wenn wir nun also (5.10) und (5.11) benutzen, erhalten wir

$$\sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_{i^*}, b_{i^*}]\} > N\varepsilon, \quad (5.14)$$

$$\frac{\max\{N_{i^*}^+(\theta)(\omega), N_{i^*}^-(\theta)(\omega)\}}{\sum_{n=1}^N \mathbf{1}\{x_{n;N} \in [a_{i^*}, b_{i^*}]\}} - \frac{1}{2} > \varepsilon$$

für alle $\omega \in \Omega_0$, $\theta \in \Theta_0$ und $N \geq N_\omega$. Sei nun $\{\tilde{k}, \tilde{k} + 1, \dots, \tilde{l}\} := \{n \mid x_{n;N} \in [a_{i^*}, b_{i^*}]\}$. Damit gilt nach Definition von $\mathcal{N}_{k,l}^\pm$ und $N_{i^*}^\pm$

$$\begin{aligned} & \sup \left\{ \frac{\max\{\mathcal{N}_{k,l}^+(\theta)(\omega), \mathcal{N}_{k,l}^-(\theta)(\omega)\}}{l - k + 1} - \frac{1}{2} \mid l, k \in \{1, \dots, N\}, l - k + 1 \geq \varepsilon N \right\} \\ & \stackrel{(5.14)}{\geq} \frac{\max\{\mathcal{N}_{\tilde{k},\tilde{l}}^+(\theta)(\omega), \mathcal{N}_{\tilde{k},\tilde{l}}^-(\theta)(\omega)\}}{\tilde{l} - \tilde{k} + 1} - \frac{1}{2} \\ & = \frac{\max\{N_{i^*}^+(\theta)(\omega), N_{i^*}^-(\theta)(\omega)\}}{\sum_{n=1}^N \mathbb{1}\{x_{n;N} \in [a_{i^*}, b_{i^*}]\}} - \frac{1}{2} \\ & > \varepsilon, \end{aligned}$$

bzw. $\omega \in A_{N,\varepsilon}(\theta)$, für alle $\omega \in \Omega_0$, $\theta \in \Theta_0$ und $N \geq N_\omega$. Insgesamt erhalten wir also die Folgerungskette

$$\begin{aligned} \omega \in A_{N,\varepsilon}(\theta) \quad \forall N \geq N_\omega, \theta \in \Theta_0 & \implies \omega \in \bigcap_{\theta \in \Theta_0} A_{N,\varepsilon}(\theta) \quad \forall N \geq N_\omega \\ & \implies \omega \in \liminf_{N \rightarrow \infty} \bigcap_{\theta \in \Theta_0} A_{N,\varepsilon}(\theta), \end{aligned}$$

und somit

$$1 = \mathbb{P}_{\theta^*}(\Omega_0) \leq \mathbb{P}_{\theta^*} \left(\liminf_{N \rightarrow \infty} \bigcap_{\theta \in \Theta_0} A_{N,\varepsilon} \right).$$

Die Aussage von Satz 5.2 folgt schließlich mit Korollar 4.7. \square

5.1 Spezialfall des polynomiellen Modells

Wir zeigen nun, dass (5.5) und (5.6) gelten für bestimmte Hypothesen im polynomiellen Modell. Sei in diesem Abschnitt $f(\cdot, \theta)$ also gegeben durch

$$f(x, \theta) = \sum_{q=0}^p \theta_q x^q \tag{5.15}$$

mit $\theta = (\theta_0, \theta_1, \dots, \theta_p)^\top \in \Theta = \mathbb{R}^{p+1}$ und die Nullhypothese durch

$$H_0 : \theta^* \in \Theta_0^m := \{\theta \in \mathbb{R}^{p+1} \mid 0 = \theta_{p-m} = \theta_{p-m+1} = \dots = \theta_p\} \tag{5.16}$$

für ein $m \in \{0, \dots, p\}$. Mit der Nullhypothese wird also ausgedrückt, dass das Modell gegeben ist durch ein Polynom mit einem Grad, der höchstens $p - m - 1$ ist.

Satz 5.4. *Wenn f gegeben ist wie in (5.15) und H_0 wie in (5.16), so gelten (5.5) und (5.6) (für alle $\theta^* \notin \Theta_0^m$) und somit ist auch der 3-Vorzeichen-Tiefe-Test konsistent für die Nullhypothese H_0 .*

Beweis. Sei $I = p - m$. Falls $\theta^0, \theta \in \Theta_0^m$, so ist $h(x) = f(x, \theta^0) - f(x, \theta)$ ein Polynom mit einem Grad, der höchstens $p - m - 1 = I - 1$ ist. Somit hat h höchstens $p - m - 1 = I - 1$ Nullstellen und somit auch höchstens $p - m - 1 = I - 1$ schwache Vorzeichenwechsel, d.h. (5.6) gilt.

Die Gültigkeit von (5.5) folgt aus dem folgenden Lemma 5.5. □

Lemma 5.5. *Seien $m \in \{0, 1, \dots, p\}$ und $\theta^* \notin \Theta_0^m$ beliebig. Dann existieren $\theta^0 \in \Theta_0^m$ und $S < z_1 < z_2 < \dots < z_I < T$, sodass $g(x) = f(x, \theta^*) - f(x, \theta^0)$ mindestens $I = p - m$ Vorzeichenwechsel bei $S < z_1 < z_2 < \dots < z_I < T$ hat und somit (5.5) gilt.*

Bemerkung 5.6. *Im Falle $m = 0$ ist der Beweis von Lemma 5.5 sehr leicht. Denn dann existiert für beliebige $S < z_1 < z_2 < \dots < z_I < T$ ein Polynom $f(\cdot, \theta^0)$ mit einem Grad, der höchstens $I - 1$ ist, d.h. $\theta^0 \in \Theta_0^m$, sodass $f(x, \theta^0) = f(x, \theta^*)$ für alle $x \in \{z_1, \dots, z_I\}$. Dann ist g gegeben durch $g(x) = f(x, \theta^*) - f(x, \theta^0)$ ein Polynom mit Grad $I = p - m = p$. Da ein Polynom mit Grad I höchstens I Nullstellen hat, so sind z_1, \dots, z_I die einzigen Nullstellen und es muss $g(x) = c \prod_{n=1}^I (x - z_n)$ mit $c \neq 0$ Konstante gelten. Aufgrund von*

$$g'(x) = c \sum_{k=1}^I \prod_{\substack{n=1 \\ n \neq k}}^I (x - z_n) \neq 0 \text{ für alle } x \in \{z_1, \dots, z_I\}$$

existieren Vorzeichenwechsel bei den I Nullstellen z_1, \dots, z_I , sodass g I Vorzeichenwechsel hat und somit (5.5) erfüllt ist.

Der Beweis für den Fall $m \geq 1$ ist aufwendiger. Für $I = p - m$ und beliebigen $S < z_1 < z_2 < \dots < z_I < T$ existiert ein Parameter $\theta(z_1, \dots, z_I) \in \Theta_0^m$ mit $f(x, \theta(z_1, \dots, z_I)) = f(x, \theta^)$ für alle $x \in \{z_1, \dots, z_I\}$. Jedoch besteht die Möglichkeit, dass aufgrund der Wahl der Stützpunkte z_1, \dots, z_I die Ableitungen identisch sind an Stützpunkten, d.h. $f'(x, \theta(z_1, \dots, z_I)) = f'(x, \theta^*)$ für ein $x \in \{z_1, \dots, z_I\}$. An solchen Stellen x könnte es sein, dass g gegeben durch $g(x) = f(x, \theta^*) - f(x, \theta(z_1, \dots, z_I))$ keinen Vorzeichenwechsel hat. Wir werden aber in dem kompletten Beweis von Lemma 5.5 zeigen, dass wir z_1, \dots, z_I so ändern können, damit (5.5) gilt.*

Beweis von Lemma 5.5. Betrachte beliebige $S < z_1 < \dots < z_I < T$. Sei $\theta(z_1, \dots, z_I) \in \Theta_0^m$ der Parameter mit $f(x, \theta(z_1, \dots, z_I)) = f(x, \theta^*)$ für alle $x \in \{z_1, \dots, z_I\}$ und g gegeben durch $g(x) = f(x, \theta^*) - f(x, \theta(z_1, \dots, z_I))$. Um sicherzustellen, dass g mindestens I Vorzeichenwechsel hat auf $[S, T]$, d.h. (5.6) gilt, brauchen wir eine passende Wahl von z_1, \dots, z_I . Wir werden in dem folgenden Beweis zeigen, dass es ausreicht, lediglich die Stützstelle z_I zu ändern, sofern dies nötig ist.

Für $k \in \{1, \dots, I\}$ definieren wir die Newton-Basisfunktionen

$$P_k(x) := \prod_{j=1}^{k-1} (x - z_j) \text{ mit der Konvention } P_1 \equiv 1.$$

Das Interpolationspolynom $Q_{z_1, \dots, z_I}^{\theta^*}$ für $f(\cdot, \theta^*)$ an den Stützstellen z_1, \dots, z_I ist dann gegeben durch

$$Q_{z_1, \dots, z_I}^{\theta^*}(x) := \sum_{k=1}^I [f(z_1, \theta^*), \dots, f(z_k, \theta^*)] P_k(x), \quad (5.17)$$

wobei die Koeffizienten $[f(z_1, \theta^*), \dots, f(z_k, \theta^*)]$ nach dem Schema der dividierten Differenzen (s. z.B. Fröberg, 1969, S. 170, oder Dahlquist und Björck, 1974, S. 278) definiert sind durch

$$\begin{aligned} [f(z_k, \theta^*)] &:= f(z_k, \theta^*), \quad k \in \{1, \dots, I\}, \\ [f(z_j, \theta^*), \dots, f(z_{j+k}, \theta^*)] &:= \frac{[f(z_{j+1}, \theta^*), \dots, f(z_{j+k}, \theta^*)] - [f(z_j, \theta^*), \dots, f(z_{j+k-1}, \theta^*)]}{z_{j+k} - z_j}, \\ &k \in \{1, \dots, I-1\}, \quad j \in \{1, \dots, I-k\}, \end{aligned}$$

sodass dann $Q_{z_1, \dots, z_I}^{\theta^*}(x) = f(x, \theta^*)$ für alle $x \in \{z_1, \dots, z_I\}$. Da wir θ^* und die Stützstellen z_1, \dots, z_{I-1} im Folgenden nicht variieren werden, kürzen wir die Notation ab und schreiben $Q_{z_I} = Q_{z_1, \dots, z_I}^{\theta^*}$. Wir definieren nun für $z_{I-1} < z < T$

$$J(z) := \{i \in \{1, \dots, I\} \mid (\bar{z}_1, \dots, \bar{z}_{I-1}, \bar{z}_I) := (z_1, \dots, z_{I-1}, z) \Rightarrow Q'_{\bar{z}_I}(\bar{z}_i) \neq f'(\bar{z}_i, \theta^*)\}.$$

Im Falle $J(z_I) = \{1, \dots, I\}$ ist die Behauptung von Lemma 5.5 bereits gezeigt, denn dann hat g mindestens I Vorzeichenwechsel auf $[S, T]$.

Betrachte also im Folgenden ein $i_0 \in \{1, \dots, I\}$ mit $Q'_{z_I}(z_{i_0}) = f'(z_{i_0}, \theta^*)$. Ersetzen wir z_I mit \tilde{z}_I , wobei $z_{I-1} < \tilde{z}_I < T$, d.h. die Reihenfolge der Stützstellen ändert sich nicht, so ändern sich nicht die ersten $I-1$ Summanden des Interpolationspolynoms in (5.17). Somit gilt

$$Q_{\tilde{z}_I}(x) - Q_{z_I}(x) = ([f(z_1, \theta^*), \dots, f(\tilde{z}_I, \theta^*)] - [f(z_1, \theta^*), \dots, f(z_I, \theta^*)]) P_I(x), \quad (5.18)$$

und wir erhalten nach Ableiten

$$Q'_{\tilde{z}_I}(x) - Q'_{z_I}(x) = ([f(z_1, \theta^*), \dots, f(\tilde{z}_I, \theta^*)] - [f(z_1, \theta^*), \dots, f(z_I, \theta^*)]) P'_I(x). \quad (5.19)$$

Wegen der Produktregel gilt

$$P'_I(x) = \sum_{k=1}^{I-1} \prod_{\substack{j=1 \\ j \neq k}}^{I-1} (x - z_j),$$

und somit

$$P'_I(\tilde{z}_I) = \sum_{k=1}^{I-1} \prod_{\substack{j=1 \\ j \neq k}}^{I-1} \underbrace{(\tilde{z}_I - z_j)}_{>0} > 0, \quad P'_I(z_i) = \prod_{\substack{j=1 \\ j \neq i}}^{I-1} \underbrace{(z_i - z_j)}_{\neq 0} \neq 0, \quad i = 1, \dots, I-1. \quad (5.20)$$

Betrachte die Hilfsfunktion $\kappa : (z_{I-1}, T) \rightarrow \mathbb{R}$ definiert durch

$$\kappa(x) := [f(z_1, \theta^*), \dots, f(z_{I-1}, \theta^*), f(x, \theta^*)].$$

Man kann aus der iterativen Definition von κ mit vollständiger Induktion bzgl. I nachweisen, dass κ auch durch

$$\kappa(x) = \frac{f(x, \theta^*)}{\prod_{k=1}^{I-1} (x - z_k)} + \sum_{j=1}^{I-1} \frac{f(z_j, \theta^*)}{(z_j - x) \prod_{\substack{k=1 \\ k \neq j}}^{I-1} (z_j - z_k)}, \quad x \in (z_{I-1}, T), \quad (5.21)$$

dargestellt werden kann (s. z.B. Fröberg, 1969, S. 170). Aus dieser Darstellung folgt sofort die Stetigkeit von κ .

Wir werden nun den Beweis in zwei Schritten durchführen. In Schritt (i) zeigen wir, dass es eine Umgebung $B_\delta(z_I) = \{z \in \mathbb{R} \mid |z - z_I| < \delta\} \subset (z_{I-1}, T)$ von z_I gibt, sodass $J(z_I) \subset J(\tilde{z}_I)$ für alle $\tilde{z}_I \in B_\delta(z_I)$, d.h. wenn die Ableitungen an einer Stützstelle unterschiedlich sind, so tun sie dies auch nach einer leichten Verschiebung von z_I . Dann zeigen wir in Schritt (ii), dass in jeder Umgebung von z_I ein \tilde{z}_I existiert mit $Q'_{\tilde{z}_I}(z_{i_0}) \neq f'(z_{i_0}, \theta^*)$, d.h.

$$i_0 \notin J(z_I) \not\subset J(\tilde{z}_I) \ni i_0.$$

(i) *Fall 1.* Der Fall $J(z_I) = \emptyset$ ist trivial.

Fall 2. Betrachte in diesem Fall $J(z_I) \neq \emptyset$.

Definiere nun $C := \min_{i \in J(z_I)} |Q'_{z_I}(z_i) - f'(z_i, \theta^*)| > 0$.

Fall 2.1. In diesem Fall sei $I \notin J(z_I)$.

Aufgrund (5.20) können wir

$$\varepsilon := \min_{i \in J(z_I)} \frac{C}{2|P'_I(z_i)|} > 0$$

definieren. Da κ stetig in z_I ist, existiert ein $\delta > 0$ mit $B_\delta(z_I) \subset (z_{I-1}, T)$, sodass $|\kappa(x) - \kappa(z_I)| < \varepsilon$ für alle $x \in B_\delta(z_I)$. Somit gilt aufgrund (5.19) für alle $i \in J(z_I)$ und $x \in B_\delta(z_I)$

$$\begin{aligned} |Q'_{z_I}(z_i) - Q'_x(z_i)| &= |(\kappa(z_I) - \kappa(x))P'_I(z_i)| \\ &< \varepsilon |P'_I(z_i)| \\ &\leq \frac{C}{2}. \end{aligned}$$

Mit der umgekehrten Dreiecksungleichung folgt also

$$\begin{aligned} |f'(z_i, \theta^*) - Q'_x(z_i)| &= |f'(z_i, \theta^*) - Q'_{z_I}(z_i) + Q'_{z_I}(z_i) - Q'_x(z_i)| \\ &\geq \underbrace{|f'(z_i, \theta^*) - Q'_{z_I}(z_i)|}_{\geq C} - \underbrace{|Q'_{z_I}(z_i) - Q'_x(z_i)|}_{\leq \frac{C}{2}} \\ &\geq \frac{C}{2} \end{aligned}$$

für alle $i \in J(z_I)$ und $x \in B_\delta(z_I)$. Also gilt $J(z_I) \subset J(x)$ für alle $x \in B_\delta(z_I)$.

Fall 2.2. Sei $I \in J(z_I)$ in diesem Fall.

Mit den Argumenten aus Fall 2.1 folgt die Existenz eines $\delta > 0$ mit $B_\delta(z_I) \subset (z_{I-1}, T)$, sodass $J(z_I) \setminus \{I\} \subset J(x)$ für alle $x \in B_\delta(z_I)$. Es genügt also zu zeigen, dass $I \in J(x)$. Indem wir wieder die Stetigkeit von κ in z_I benutzen, erhalten wir ein $\delta > \delta_1 > 0$, sodass

$$|\kappa(x) - \kappa(z_I)| < \frac{C}{4 \sup_{x \in B_\delta(z_I)} P'_I(x)} \text{ für alle } x \in B_{\delta_1}(z_I),$$

und damit auch

$$\begin{aligned} |Q'_{z_I}(x) - Q'_x(x)| &= |(\kappa(z_I) - \kappa(x))P'_I(x)| \\ &< \frac{C}{4 \sup_{x \in B_\delta(z_I)} P'_I(x)} |P'_I(x)| \\ &\leq \frac{C}{4} \end{aligned}$$

für alle $x \in B_{\delta_1}(z_I)$. Außerdem erhalten wir mit der Stetigkeit von $f'(\cdot, \theta^*)$ und Q'_{z_I} ein $\delta_1 > \delta_2 > 0$, sodass

$$|f'(x, \theta^*) - f'(z_I, \theta^*)| < \frac{C}{4}, \quad |Q'_{z_I}(x) - Q'_{z_I}(z_I)| < \frac{C}{4} \text{ für alle } x \in B_{\delta_2}(z_I).$$

Mit der umgekehrten Dreiecksungleichung erhalten wir nun

$$\begin{aligned} &|f'(x, \theta^*) - Q'_x(x)| \\ &= |f'(x, \theta^*) - f'(z_I, \theta^*) + f'(z_I, \theta^*) - Q'_{z_I}(z_I) + Q'_{z_I}(z_I) - Q'_{z_I}(x) \\ &\quad + Q'_{z_I}(x) - Q'_x(x)| \\ &\geq |f'(z_I, \theta^*) - Q'_{z_I}(z_I)| - |f'(x, \theta^*) - f'(z_I, \theta^*)| - |Q'_{z_I}(z_I) - Q'_{z_I}(x)| \\ &\quad - |Q'_{z_I}(x) - Q'_x(x)|. \end{aligned}$$

Aus den vorangegangenen Abschätzungen folgt also

$$\begin{aligned} |f'(x, \theta^*) - Q'_x(x)| &\geq C - \frac{C}{4} - \frac{C}{4} - \frac{C}{4} \\ &= \frac{C}{4} \\ &> 0 \end{aligned}$$

für alle $x \in B_{\delta_2}(z_I)$, und somit $I \in J(x)$ für alle $x \in B_{\delta_2}(z_I)$.

(ii) *Fall 1.* In diesem Fall sei $i_0 \neq I$.

Angenommen, es existiert ein $\delta > 0$ mit $f'(z_{i_0}, \theta^*) = Q'_{z_I}(z_{i_0}) = Q'_x(z_{i_0})$ für alle $x \in$

$B_\delta(z_I)$ (und $B_\delta(z_I) \subset (z_{I-1}, T)$). Hierbei sei anzumerken, dass die erste Gleichung nach Voraussetzung gilt. Da $P'_I(z_{i_0}) \neq 0$, erhalten wir mit (5.19)

$$\kappa(x) = [f(z_1, \theta^*), \dots, f(x, \theta^*)] = [f(z_1, \theta^*), \dots, f(z_I, \theta^*)] = \kappa(z_I)$$

für alle $x \in B_\delta(z_I)$. Nun ist $(z_{I-1}, T) \ni x \mapsto p(x)(\kappa(x) - \kappa(z_I))$ ein Polynom mit

$$p(x)(\kappa(x) - \kappa(z_I)) = 0 \text{ für alle } x \in B_\delta(z_I)$$

und somit auch für alle $x \in (z_{I-1}, T)$, wobei für den Hauptnenner von $\kappa(x) - \kappa(z_I)$

$$\begin{aligned} p(x) &= \left(\prod_{k=1}^{I-1} (x - z_k) \right) \left(\prod_{j=1}^{I-1} \left((z_j - x) \prod_{\substack{k=1 \\ k \neq j}}^{I-1} (z_j - z_k) \right) \right) \\ &\cdot \left(\prod_{k=1}^{I-1} (z_I - z_k) \right) \left(\prod_{j=1}^{I-1} \left((z_j - z_I) \prod_{\substack{k=1 \\ k \neq j}}^{I-1} (z_j - z_k) \right) \right) \neq 0 \end{aligned}$$

für alle $x \in (z_{I-1}, T)$ gilt, d.h. es muss sogar

$$\kappa(x) - \kappa(z_I) = 0 \text{ für alle } x \in (z_{I-1}, T)$$

gelten. Damit gilt

$$\begin{aligned} Q_\xi(x) &= Q_{z_I}(x) \text{ für alle } \xi \in (z_{I-1}, T) \text{ und } x \in [S, T], \text{ und somit insbesondere} \\ f(\xi, \theta^*) &= Q_\xi(\xi) = Q_{z_I}(\xi) \text{ für alle } \xi \in (z_{I-1}, T). \end{aligned}$$

Dies impliziert nun $f(\cdot, \theta^*) = Q_{z_I}$ auf ganz $[S, T]$. Somit ist der Grad von $f(\cdot, \theta^*)$ höchstens $I - 1$, d.h. $\theta^* \in \Theta_0^m$. Da wir aber $\theta^* \notin \Theta_0^m$ in den Voraussetzungen von Lemma 5.5 angenommen haben, ist dies ein Widerspruch. Deshalb existiert also in jeder Umgebung $B_\delta(z_I) \subset (z_{I-1}, T)$ von z_I ein \tilde{z}_I mit $i_0 \in J(\tilde{z}_I)$.

Beachte, dass wir hier $i_0 = I$ ausschließen mussten, da mit $i_0 = I$ das Resultat aus dem Widerspruch wäre, dass aus jeder Umgebung von z_I ein \tilde{z}_I existiert mit $f'(z_I, \theta^*) \neq Q'_{\tilde{z}_I}(z_I)$. Nach Definition von $J(\cdot)$ ist dies aber nicht äquivalent zu $I \in J(x)$.

Fall 2. In diesem Fall sei also nun $i_0 = I$. Wir nehmen hier ähnlich wie zuvor an, dass $Q'_x(x) = f'(x, \theta^*)$ für alle x aus einer Umgebung von z_I . Nach Multiplikation mit $P_I(x)$ erhalten wir für alle x aus dieser Umgebung

$$f'(x, \theta^*)P_I(x) = P_I(x) \left(\kappa(x)P'_I(x) + \underbrace{\sum_{k=1}^{I-1} [f(z_1, \theta^*), \dots, f(z_k, \theta^*)]P'_k(x)}_{=:U(x)} \right),$$

wobei U ein Polynom ist mit einem Grad, der höchstens $I - 3$ ist. Indem wir nun die Darstellung (5.21) von $\kappa(x)$ einsetzen, erhalten wir wieder für alle x aus dieser Umgebung von z_I

$$\begin{aligned}
 f'(x, \theta^*)P_I(x) &= P_I(x)U(x) + P'_I(x)f(x, \theta^*) + P'_I(x) \sum_{j=1}^{I-1} \frac{P_I(x)f(z_j, \theta^*)}{(z_j - x) \prod_{\substack{k=1 \\ k \neq j}}^{I-1} (z_j - z_k)} \\
 &= P_I(x)U(x) + P'_I(x)f(x, \theta^*) + P'_I(x) \sum_{j=1}^{I-1} \frac{f(z_j, \theta^*) \prod_{k=1, k \neq j}^{I-1} (x - z_k)}{(z_j - x) \prod_{k=1, k \neq j}^{I-1} (z_j - z_k)} \\
 &= P_I(x)U(x) + P'_I(x)f(x, \theta^*) - \underbrace{P'_I(x) \sum_{j=1}^{I-1} \left(f(z_j, \theta^*) \prod_{\substack{k=1 \\ k \neq j}}^{I-1} \frac{x - z_k}{z_j - z_k} \right)}_{=:V(x)},
 \end{aligned}$$

wobei V ein Polynom ist mit einem Grad, der höchstens $I - 2$ ist. Da auf beiden Seiten dieser Gleichung nun Polynome stehen, welche übereinstimmen auf einer Umgebung von z_I , müssen diese Polynome identisch sein (auf ganz \mathbb{R}). Um dies ausnutzen zu können, betrachten wir zunächst die Grade der auftretenden Polynome. Sei dafür $\tilde{I} := \deg(f(\cdot, \theta^*)) \geq I$ der Grad vom Polynom $f(\cdot, \theta^*)$. Dann erhalten wir

$$\begin{aligned}
 \deg(x \mapsto f'(x, \theta^*)P_I(x)) &= (\tilde{I} - 1) + (I - 1) = \tilde{I} + I - 2, \\
 \deg(x \mapsto P_I(x)U(x)) &\leq (I - 1) + (I - 3) < \tilde{I} + I - 2, \\
 \deg(x \mapsto P'_I(x)f(x, \theta^*)) &= (I - 2) + \tilde{I} = \tilde{I} + I - 2, \\
 \deg(x \mapsto P'_I(x)V(x)) &\leq (I - 2) + (I - 2) < \tilde{I} + I - 2.
 \end{aligned}$$

Somit ist also $f'(\cdot, \theta^*)P_I - P'_I f(\cdot, \theta^*)$, eine Differenz von zwei Polynomen mit Grad $\tilde{I} + I - 2$, identisch mit einem Polynom, dessen Grad kleiner als $\tilde{I} + I - 2$ ist. Dies bedeutet, dass die Leitkoeffizienten von $f'(\cdot, \theta^*)P_I$ und $P'_I f(\cdot, \theta^*)$ identisch sind. Nun ist der Leitkoeffizient von P_I 1, und somit der Leitkoeffizient von P'_I $I - 1$. Nach den Regeln für Ableitungen von Polynomen ist auch der Leitkoeffizient von $f'(\cdot, \theta^*)$ das \tilde{I} -fache des Leitkoeffizienten von $f(\cdot, \theta^*)$. Wenn wir mit $\text{lc}(p)$ den Leitkoeffizienten eines Polynoms p schreiben, folgt also nun

$$\begin{aligned}
 \tilde{I} \cdot \text{lc}(f(\cdot, \theta^*)) &= (\tilde{I} \cdot \text{lc}(f(\cdot, \theta^*))) \cdot 1 = \text{lc}(x \mapsto f'(x, \theta^*)P_I(x)) \\
 &= \text{lc}(x \mapsto P'_I(x)f(x, \theta^*)) = (I - 1) \cdot \text{lc}(f(\cdot, \theta^*)),
 \end{aligned}$$

d.h. $\tilde{I} = I - 1$, was ein Widerspruch zu $\tilde{I} \geq I$ ist. Somit existiert ein in jeder Umgebung von z_I ein \tilde{z}_I , sodass $i_0 = I \in J(\tilde{z}_I)$.

Nun müssen wir Schritte (i) und (ii) höchstens für alle $i_0 \notin J(z_I)$ iterativ durchführen, wobei z_I in jeder Durchführung des Schritts (ii) (ein wenig) verschoben wird. Dies bedeutet, dass wir die Schritte (i) und (ii) höchstens I -mal durchführen müssen. Danach entspricht der gesuchte Parameter $\theta^0 = \theta(z_1, \dots, z_I)$ den Koeffizienten des Interpolationspolynoms $Q_{z_I} = Q_{z_1, \dots, z_I}^{\theta^*}$. \square

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

In diesem Kapitel wenden wir den 3-Vorzeichen-Tiefe-Test an und vergleichen ihn mit anderen Tests in Bezug auf unter anderem Konsistenz und Güte. Dabei betrachten wir das polynomielle Modell, wie es in Kapitel 5 und insbesondere Abschnitt 5.1 definiert wird, d.h. die Bedingungen (5.1) bis (5.4) gelten und $f(\cdot, \theta)$ ist gegeben durch

$$f(x, \theta) = \sum_{q=0}^p \theta_q x^q,$$

wobei $\theta = (\theta_0, \theta_1, \dots, \theta_p)^\top \in \Theta = \mathbb{R}^{p+1}$.

Im Wesentlichen betrachten wir im Folgenden das quadratische Modell mit $p = 2$, bei dem wir bzgl. der Nullhypothese

$$H_0 : \theta^* \in \Theta_0 := \{\theta \in \mathbb{R}^3 \mid 0 = \theta_2\}$$

testen. Dies bedeutet, dass wir im quadratischen Modell testen, ob wirklich ein quadratisches Modell vorliegt, d.h. $\theta_2^* \neq 0$, und nicht lediglich ein lineares Modell, d.h. $\theta_2^* = 0$.

Um den nach Kapitel 3 durch

$$\text{Lehne } H_0 \text{ ab, falls } \sup_{\theta \in \Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_3)$$

definierten 3-Vorzeichen-Tiefe-Test anwenden zu können, benötigen wir das α -Quantil $q_\alpha(\Psi_3)$ und Methoden, das Supremum $\sup_{\theta \in \Theta_0} \Psi_3(w^N(\theta))$ für Realisierungen $w^N(\theta)$ von $\mathcal{W}^N(\theta)$ für $\theta \in \Theta_0$ approximativ zu berechnen. Dies behandeln wir in den nächsten beiden Abschnitten 6.1 und 6.2.

6.1 Möglichkeiten zur (approximativen) Berechnung von $q_\alpha(\Psi_3)$

$q_\alpha(\Psi_3)$ ist das α -Quantil von $\Psi_3(W)$, wobei W eine Standard-Brownsche Bewegung ist. Dies bedeutet, dass

- (i) $W_0 = 0$ \mathbb{P} -f.s.,
- (ii) die Pfade $t \mapsto W_t$ sind \mathbb{P} -f.s. stetig,
- (iii) $W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ sind stochastisch unabhängig für alle $0 \leq t_1 < t_2 < \dots < t_n$, und
- (iv) $W_t - W_s \sim \mathcal{N}(0, t - s)$ für alle $0 \leq s < t$.

Seien nun $L, M \in \mathbb{N}$. Aufgrund dieser Eigenschaften von W können wir also leicht Werte

$$0 = w_0, w_{\frac{1}{L}}, w_{\frac{2}{L}}, \dots, w_{\frac{L}{L}} = w_1$$

einer Realisierung w von W generieren. Dies können wir z.B. mit der Programmiersprache R tun, s. R Core Team, 2022. Mit diesen Werten können wir dann mit Hilfe einer Riemann-Summe bzgl. der Zerlegung $\{0, 1/L, 2/L, \dots, L/L = 1\}$ von $[0, 1]$ $\Psi_3(w)$ approximieren, d.h.

$$\Psi_3(w) = \frac{3}{4} \left(1 - \int_0^1 (w_1 - 2w_t)^2 dt \right) \approx \frac{3}{4} \left(1 - \sum_{i=1}^L \frac{1}{L} (w_1 - 2w_{\frac{i}{L}})^2 \right).$$

Dieses Vorgehen können wir dann M -mal wiederholen, womit wir M Realisierungen $\Psi_3(w)$ von $\Psi_3(W)$ erhalten. Daraus erhalten wir das empirische α -Quantil dieser M Werte, womit wir für große L, M das theoretische α -Quantil $q_\alpha(\Psi_3)$ näherungsweise bestimmen können.

Eine weitere Möglichkeit, $q_\alpha(\Psi_3)$ (im zweifachen Sinne) zu approximieren, ist die näherungsweise Berechnung des α -Quantils $q_{3,\alpha}$ von $d_3(R_1(\theta^*), \dots, R_N(\theta^*))$ für große N . Wegen der Beziehung zwischen $d_3(R_1(\theta^*), \dots, R_N(\theta^*))$ und $\Psi_3(\mathcal{W}^N(\theta^*))$ in (3.12) und der Erläuterung in Abschnitt 3.1 über die Konvergenz von $\mathcal{W}^N(\theta^*)$ gegen eine Standard-Brownsche Bewegung W sollten wir also mit

$$\frac{(N-1)(N-2)}{N} (q_{3,\alpha} - 0.25) \approx q_\alpha(\Psi_3) \quad (6.1)$$

für große N eine Approximation von $q_\alpha(\Psi_3)$ erhalten. Da $\psi(R_1(\theta^*)), \dots, \psi(R_N(\theta^*))$ unabhängig und diskret gleichverteilt mit Träger $\{-1, +1\}$ sind nach (A1) und (A2), müssten wir für eine exakte Berechnung von $q_{3,\alpha}$ die Vorzeichen-Tiefen $d_3(r_1, \dots, r_N)$ für alle $(r_1, \dots, r_N) \in \{-1, +1\}^N$ berechnen, was die Berechnung von 2^N Vorzeichen-Tiefen bedeutet. Dies ist also für große N nicht empfehlenswert, weshalb wir eher für eine bestimmte (nicht exponentielle) Zahl Realisierungen (r_1, \dots, r_N) von $(R_1(\theta^*), \dots, R_N(\theta^*))$ die Vorzeichen-Tiefe $d_3(r_1, \dots, r_N)$ berechnen würden, woraus wir dann das α -Quantil $q_{3,\alpha}$ näherungsweise berechnen könnten.

Im Folgenden werden wir jedoch in der praktischen Anwendung des 3-Vorzeichen-Tiefe-Tests die approximativen Quantile in Tabelle 1 aus Falkenau, 2016, S. 76, verwenden. Dort wurde eine andere Darstellung des Limes von $\Psi_3(\mathcal{W}^N(\theta^*))$ verwendet, welcher jedoch die gleiche Verteilung wie $\Psi_3(W)$ hat.

α	0.005	0.01	0.025	0.05	0.1	0.2
$q_\alpha(\Psi_3)$	-2.6543216	-2.2403956	-1.6791727	-1.2545411	-0.8270908	-0.4032076

Tabelle 1: Approximative Quantile von $\Psi_3(W)$ (Falkenau, 2016, S. 76)

Mit der R-Funktion `qdepth` aus dem Paket `GSignTest`, s. Horn, 2021a, mit dem Argument `K=3` erhalten wir für $N = 3, 4, \dots, 25$ die exakten Quantile von $d_3(R_1(\theta^*), \dots, R_N(\theta^*))$ und für $N = 26, \dots, 100$ die simulierten, d.h. angenäherten, Quantile. Für Datenmengen $N > 100$ werden die Werte für $N = 100$ gewählt. Wenn wir dann die erhaltenen Werte für $N = 100$ wie in (6.1) normieren, erhalten wir die Quantile in Tabelle 2.

Die Quantile aus Tabelle 1 sind kleiner als die aus Tabelle 2. Somit würden wir bei der Verwendung der Quantile aus Tabelle 2 häufiger die Nullhypothese ablehnen.

Den 3-Vorzeichen-Tiefe-Test werden wir lediglich mit $\alpha = 0.05$ verwenden.

α	0.005	0.01	0.025	0.05	0.1	0.2
$q_\alpha(\Psi_3)$	-2.546364	-2.135796	-1.608048	-1.203048	-0.790524	-0.376296

Tabelle 2: Approximative Quantile von $\Psi_3(W)$ basierend auf der Quantilfunktion `qdepth` (Horn, 2021a)

6.2 Möglichkeiten zur (approximativen) Berechnung von $\sup_{\theta \in \Theta_0} \Psi_3(w^N(\theta))$

Wir betrachten zwei Möglichkeiten zur approximativen Berechnung von

$$\sup_{\theta \in \Theta_0} d_3(r_1(\theta), \dots, r_N(\theta))$$

für Realisierungen $(r_1(\theta), \dots, r_N(\theta))$ von $(R_1(\theta), \dots, R_N(\theta))$ für $\theta \in \Theta_0$. Aufgrund von (3.12) erhalten wir somit auch zwei Möglichkeiten zur approximativen Berechnung von $\sup_{\theta \in \Theta_0} \Psi_3(w^N(\theta))$.

Im Folgenden identifizieren wir die Parametermenge der Nullhypothese $\Theta_0 = \mathbb{R}^2 \times \{0\} \subset \mathbb{R}^3$ mit \mathbb{R}^2 . Dies bedeutet, dass wir ein Parameter $(\theta_0, \theta_1, 0)^\top \in \Theta_0$ als $(\theta_0, \theta_1)^\top \in \mathbb{R}^2$ betrachten.

6.2.1 Gittersuche

Ein naheliegender Ansatz für die approximative Berechnung des Supremums ist die Gitter- bzw. Rastersuche. Dabei legen wir auf einen bestimmten Bereich $[L_0, U_0] \times [L_1, U_1]$ von Θ_0 ein Raster, d.h. ein System aus horizontalen und vertikalen Geraden, wobei benachbarte horizontale (bzw. vertikale) Geraden stets den gleichen Abstand zueinander haben. Für alle Stellen $\theta \in \Theta_0$, an denen sich eine horizontale und vertikale Gerade schneiden, berechnen wir dann $d_3(r_1(\theta), \dots, r_N(\theta))$. Wir geben dann sowohl das θ zurück, für das die zugehörige Vorzeichen-Tiefe am höchsten ist, als auch diese Vorzeichen-Tiefe.

Wenn das Raster fein genug ist, sollte dies aufgrund der Definition von d_3 stets zur exakten Berechnung des Supremums führen. Das Vorgehen ist ebenfalls in Pseudocode 1 beschrieben.

Dieses Verfahren lässt sich auch für das allgemeinere Modell in Abschnitt 5.1 mit Nullhypothese (5.16) verallgemeinern. Dafür betrachten wir stattdessen einen Bereich $[L_0, U_0] \times \dots \times [L_{p-m-1}, U_{p-m-1}]$ und Anzahl an Geraden (A_0, \dots, A_{p-m-1}) . Dann wenden wir das Verfahren auf analoge Weise auf dieses $(p-m)$ -dimensionale Raster an.

Den Bereich $[L_0, U_0] \times [L_1, U_1]$ können wir mit Hilfe einer robusten Regression festlegen. Dafür können wir z.B. die Funktion `r1m` aus dem R-Paket `MASS` benutzen, s. Venables und Ripley, 2002.

Dieses Gittersuche-Verfahren hat jedoch das Problem, dass wir $A_0 \cdot A_1$ Vorzeichen-Tiefen berechnen müssen. Wenn der Bereich $[L_0, U_0] \times [L_1, U_1]$ relativ groß ist und wir somit für ein feines Raster hohe Werte von (A_0, A_1) benötigen, kann dies sehr aufwendig sein. Dabei kann es vorkommen, dass wir mehrere θ in diesem Verfahren betrachten, die die gleichen Residuen-Vorzeichen $(\psi(r_1(\theta)), \dots, \psi(r_N(\theta)))$ erzeugen, und wir somit

Eingabe: Daten $(x_1, y_1), \dots, (x_N, y_N)$, Bereich $[L_0, U_0] \times [L_1, U_1]$, Anzahl an vertikalen und horizontalen Geraden (A_0, A_1) .

Ausgabe: Parameter $\theta \in \Theta_0$ und zugehörige Vorzeichen-Tiefe $d_3(r_1(\theta), \dots, r_N(\theta))$.

- 1 für $i = 1, \dots, A_0$
- 2 für $j = 1, \dots, A_1$
- 3 Sei $\theta^{i,j} = \left(L_0 + \frac{i-1}{A_0-1}(U_0 - L_0), L_1 + \frac{j-1}{A_1-1}(U_1 - L_1) \right)$.
- 4 Berechne Vorzeichen-Tiefe $d^{i,j} = d_3(r_1(\theta^{i,j}), \dots, r_N(\theta^{i,j}))$, wobei $r_n(\theta^{i,j}) = y_n - (\theta_0^{i,j} + \theta_1^{i,j} x_n)$.
- 5 Gebe sowohl ein $\theta^{i,j}$ zurück, sodass $d^{i,j}$ maximal ist, als auch dieses $d^{i,j}$.

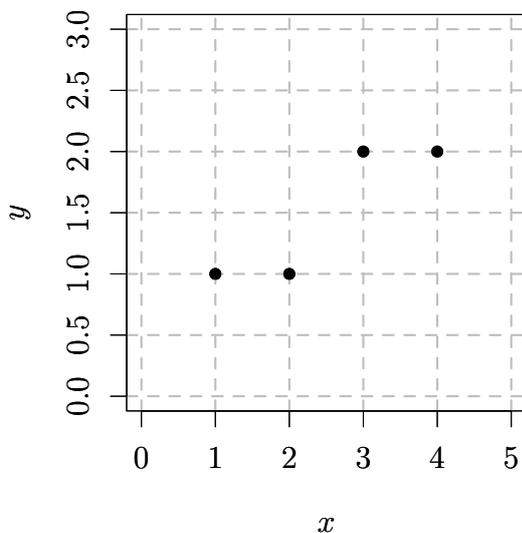
Pseudocode 1: Pseudocode für die Gittersuche

unnötige Berechnungen von Vorzeichen-Tiefen durchführen. Deshalb müsste man sich bei diesem Verfahren merken, für welche Residuen-Vorzeichen wir schon die Vorzeichen-Tiefe berechnet haben. Dafür betrachten wir das folgende Beispiel 6.1.

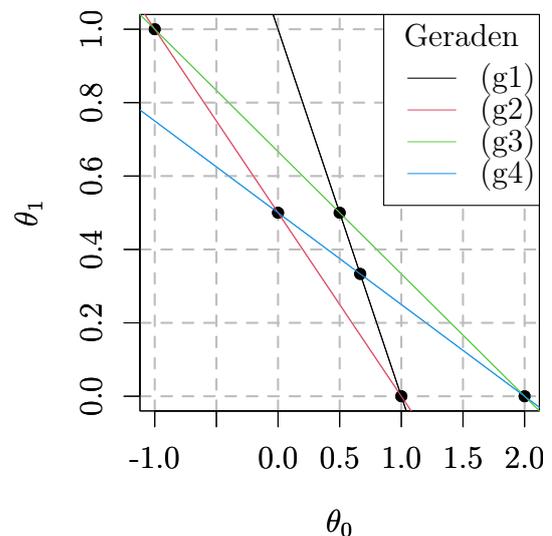
Beispiel 6.1. Gegeben seien die Daten

$$(x_1, y_1) = (1, 1), (x_2, y_2) = (2, 1), (x_3, y_3) = (3, 2), (x_4, y_4) = (4, 2)$$

in Abbildung 3a.



(a) Beispieldaten



(b) Geraden definiert durch (g1) bis (g4)

Abbildung 3: Grafik zu Beispiel 6.1

Wir wollen nun untersuchen, für welche $\theta \in \Theta_0$ ein Residuum $r_n(\theta)$ positiv und für

welche es negativ ist. Dafür betrachten wir

$$y_1 - (\theta_0 + \theta_1 x_1) = r_1(\theta) = 0 \quad \Leftrightarrow \quad \theta_0 + \theta_1 = 1, \quad (\text{g1})$$

$$y_2 - (\theta_0 + \theta_1 x_2) = r_2(\theta) = 0 \quad \Leftrightarrow \quad \theta_0 + 2\theta_1 = 1, \quad (\text{g2})$$

$$y_3 - (\theta_0 + \theta_1 x_3) = r_3(\theta) = 0 \quad \Leftrightarrow \quad \theta_0 + 3\theta_1 = 2, \quad (\text{g3})$$

$$y_4 - (\theta_0 + \theta_1 x_4) = r_4(\theta) = 0 \quad \Leftrightarrow \quad \theta_0 + 4\theta_1 = 2. \quad (\text{g4})$$

Die durch (g1) bis (g4) erhaltenen Geraden und deren Schnittpunkte sind in Abbildung 3b dargestellt. So gilt z.B., dass für alle $\theta \in \Theta_0$ unterhalb der Gerade (g1) das Residuum $r_1(\theta)$ positiv ist, und oberhalb negativ. Gleiches gilt für die anderen Geraden. Wir erhalten durch diese Geraden eine Art Mosaik aus Polyedern, d.h. Mengen der Form $\{x \in \mathbb{R}^n \mid Ax \leq b\}$ für gegebene $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, deren Vereinigung ganz \mathbb{R}^2 ergibt und deren offenen Kerne disjunkt zueinander sind. Die Residuen-Vorzeichen $(\psi(r_1(\theta)), \dots, \psi(r_N(\theta)))$ sind dann auf dem offenen Kern, d.h. dem Inneren, eines solchen Polyeders jeweils konstant, und somit ist dann auch dort die Vorzeichen-Tiefe $d_3(r_1(\theta), \dots, r_N(\theta))$ konstant.

Beispiel 6.1, und insbesondere Abbildung 3b, spricht also eher gegen den Ansatz, für die Suche nach der maximalen Vorzeichen-Tiefe ein Raster aus horizontalen und vertikalen Geraden auf einen Bereich von Θ_0 zu legen.

6.2.2 Datenpunktpaar-Verfahren

Wenn wir für beliebige Daten $(x_1, y_1), \dots, (x_N, y_N)$ so wie in Beispiel 6.1 vorgehen, erhalten wir wieder ein System aus Geraden, welche definiert sind durch $\theta_0 + \theta_1 x_n = y_n$. Bei dem Datenpunktpaar-Verfahren berechnen wir dann für alle Schnittpunkte $\theta \in \Theta_0$ die zugehörige Vorzeichen-Tiefe und geben dann von diesen Vorzeichen-Tiefen die Höchste zusammen mit dem zugehörigen Parameter θ zurück. Eines dieser Schnittpunkte $\theta \in \Theta_0$, welches durch den Schnitt von

$$r_i(\theta) = 0 \text{ und } r_j(\theta) = 0 \quad \Leftrightarrow \quad \theta_0 + \theta_1 x_i = y_i \text{ und } \theta_0 + \theta_1 x_j = y_j \quad (6.2)$$

entsteht, entspricht gerade den Koeffizienten der linearen Funktion durch (x_i, y_i) und (x_j, y_j) . Beachte, dass ein Schnittpunkt von Geraden in (6.2) stets existiert. Dies liegt daran, dass diese Geraden aufgrund von $x_1 < \dots < x_N$ nie parallel sein können. Wenn $x_n \neq 0$, so ist nämlich die Steigung der durch $\theta_0 + \theta_1 x_n = y_n$ definierten Gerade, bzw. der linearen Funktion $\theta_0 \mapsto y_n/x_n - \theta_0/x_n, -1/x_n$. Im Falle $x_n = 0$ erhalten wir eine vertikale Gerade.

Dieses Vorgehen beschreiben wir ebenfalls als Pseudocode.

Wir könnten in Pseudocode 2 anstatt die Koeffizienten $\theta \in \Theta_0$ der linearen Funktion durch (x_i, y_i) und (x_j, y_j) auch für ein gegebenes kleines $\varepsilon > 0$ die Koeffizienten $\theta \in \Theta_0$ der linearen Funktion durch $(x_i, y_i + \delta_i)$ und $(x_j, y_j + \delta_j)$ für alle $\delta_i, \delta_j \in \{+\varepsilon, -\varepsilon\}$ betrachten, wie in Pseudocode 3 beschrieben.

Die Begründung dafür ist, dass wir bei dem Verfahren in Pseudocode 2 stets $\theta \in \Theta_0$ betrachten, für die (mindestens) zwei der zugehörigen Residuen $r_1(\theta), \dots, r_N(\theta)$ Null

Eingabe: Daten $(x_1, y_1), \dots, (x_N, y_N)$.
Ausgabe: Parameter $\theta \in \Theta_0$ und zugehörige Vorzeichen-Tiefe $d_3(r_1(\theta), \dots, r_N(\theta))$.

- 1 **für** $(i, j) \in \{1, \dots, N\}^2$ mit $i < j$
- 2 Sei $\theta^{i,j} = \left(\frac{y_i x_j - x_i y_j}{x_j - x_i}, \frac{y_j - y_i}{x_j - x_i} \right)$.
- 3 Berechne Vorzeichen-Tiefe $d^{i,j} = d_3(r_1(\theta^{i,j}), \dots, r_N(\theta^{i,j}))$, wobei $r_n(\theta^{i,j}) = y_n - (\theta_0^{i,j} + \theta_1^{i,j} x_n)$.
- 4 Gebe sowohl ein $\theta^{i,j}$ zurück, sodass $d^{i,j}$ maximal ist, als auch dieses $d^{i,j}$.

Pseudocode 2: Pseudocode für das Datenpunktpaar-Verfahren

Eingabe: Daten $(x_1, y_1), \dots, (x_N, y_N)$, Shift-Parameter $\varepsilon > 0$.
Ausgabe: Parameter $\theta \in \Theta_0$ und zugehörige Vorzeichen-Tiefe $d_3(r_1(\theta), \dots, r_N(\theta))$.

- 1 **für** $(i, j) \in \{1, \dots, N\}^2$ mit $i < j$
- 2 **für** $(\delta_i, \delta_j) \in \{+\varepsilon, -\varepsilon\}^2$
- 3 Sei $\theta^{i,j,\delta_i,\delta_j} = \left(\frac{(y_i + \delta_i)x_j - x_i(y_j + \delta_j)}{x_j - x_i}, \frac{(y_j + \delta_j) - (y_i + \delta_i)}{x_j - x_i} \right)$.
- 4 Berechne Vorzeichen-Tiefe $d^{i,j,\delta_i,\delta_j} = d_3(r_1(\theta^{i,j,\delta_i,\delta_j}), \dots, r_N(\theta^{i,j,\delta_i,\delta_j}))$, wobei $r_n(\theta^{i,j,\delta_i,\delta_j}) = y_n - (\theta_0^{i,j,\delta_i,\delta_j} + \theta_1^{i,j,\delta_i,\delta_j} x_n)$.
- 5 Gebe sowohl ein $\theta^{i,j,\delta_i,\delta_j}$ zurück, sodass $d^{i,j,\delta_i,\delta_j}$ maximal ist, als auch dieses $d^{i,j,\delta_i,\delta_j}$.

Pseudocode 3: Alternative 1 für das Datenpunktpaar-Verfahren

sind. Nach den Annahmen an unser Modell können wir auch für diese θ annehmen, dass genau zwei Residuen $r_i(\theta)$, $r_j(\theta)$ Null sind. Wenn wir θ nun so wenig verschieben, dass diese beiden Residuen nicht mehr Null sind und sich die Vorzeichen der anderen Residuen nicht ändert, so vergrößern wir die zugehörige Vorzeichen-Tiefe nach Definition. Hierbei würde sich jedoch die Frage stellen, wie wir den Shift-Parameter $\varepsilon > 0$ wählen sollen, um das oben gewünschte Verhalten in allen Durchläufen der Für-Schleife in Zeile 1 von Pseudocode 3 sicherzustellen, d.h.

$$\psi(r_n(\theta^{i,j,\delta_i,\delta_j})) = \psi(r_n(\theta^{i,j,0,0})) \quad \text{für alle } n \in \{1, \dots, N\} \setminus \{i, j\}, \quad (6.3)$$

$$(i, j) \in \{1, \dots, N\} \text{ mit } i < j, (\delta_i, \delta_j) \in \{+\varepsilon, -\varepsilon\}^2,$$

wobei $\theta^{i,j,0,0} = \theta^{i,j}$ mit $\theta^{i,j}$ wie in Pseudocode 2.

Diese Problematik der Wahl des Shift-Parameters $\varepsilon > 0$ führt uns zu einer weiteren Alternative des Datenpunktpaar-Verfahrens, welches in Pseudocode 4 beschrieben wird.

Die Motivation für diese Alternative ist, dass aufgrund der Annahmen an unser Modell ein Shift-Parameter $\varepsilon > 0$ existiert, sodass (6.3) gilt. Dann gilt $\psi(r_i(\theta^{i,j,\delta_i,\delta_j})) = \psi(\delta_i)$ und $\psi(r_j(\theta^{i,j,\delta_i,\delta_j})) = \psi(\delta_j)$ für alle $(i, j) \in \{1, \dots, N\}^2$ mit $i < j$ und $(\delta_i, \delta_j) \in \{+\varepsilon, -\varepsilon\}^2$. Mit diesem ε gibt also das Vorgehen des Pseudocodes 3 die gleiche Vorzeichen-Tiefe zurück wie das Vorgehen des Pseudocodes 4, wobei bei der zweiten Alternative nun natürlich die Wahl des Shift-Parameters ε wegfällt.

Eingabe: Daten $(x_1, y_1), \dots, (x_N, y_N)$.
Ausgabe: Residuen (r_1, \dots, r_N) und zugehörige Vorzeichen-Tiefe $d_3(r_1, \dots, r_N)$,
 Parameter $\theta \in \Theta_0$

- 1 **für** $(i, j) \in \{1, \dots, N\}^2$ mit $i < j$
- 2 Sei $\theta^{i,j} = \left(\frac{y_i x_j - x_i y_j}{x_j - x_i}, \frac{y_j - y_i}{x_j - x_i} \right)$.
- 3 Berechne $r^{i,j} = (r_1(\theta^{i,j}), \dots, r_N(\theta^{i,j}))$, wobei $r_n(\theta^{i,j}) = y_n - (\theta_0^{i,j} + \theta_1^{i,j} x_n)$.
- 4 **für** $(l_i, l_j) \in \{+1, -1\}^2$
- 5 Sei $r^{i,j,l_i,l_j} \in \mathbb{R}^N$ mit

$$r_n^{i,j,l_i,l_j} = \begin{cases} r_n^{i,j}, & \text{falls } n \notin \{i, j\}, \\ l_i, & \text{falls } n = i, \\ l_j, & \text{falls } n = j. \end{cases}$$
- 6 Berechne Vorzeichen-Tiefe $d^{i,j,l_i,l_j} = d_3(r^{i,j,l_i,l_j})$.
- 7 Gebe sowohl ein $\theta^{i,j}$ und r^{i,j,l_i,l_j} zurück, sodass d^{i,j,l_i,l_j} maximal ist, als auch dieses d^{i,j,l_i,l_j} .

Pseudocode 4: Alternative 2 für das Datenpunktpaar-Verfahren

Auch dieses Datenpunktpaar-Verfahren lässt sich für das allgemeinere Modell in Abschnitt 5.1 mit Nullhypothese (5.16) verallgemeinern. Dabei betrachten wir dann eine Für-Schleife bzgl. der Index- $(p-m)$ -Tupel $(i_1, \dots, i_{p-m}) \in \{1, \dots, N\}^{p-m}$ mit $i_1 < \dots < i_{p-m}$. Für ein solches Index- $(p-m)$ -Tupel berechnen wir dann die Koeffizienten $\theta \in \Theta_0^m$ des Interpolationspolynoms bzgl. der Punkte $(x_{i_1}, y_{i_1}), \dots, (x_{i_{p-m}}, y_{i_{p-m}})$, welches einen Grad hat, der höchstens $p-m-1$ ist. Für dieses θ berechnen wir dann die Vorzeichen-Tiefe. Wenn wir den Rechenaufwand für die Berechnung der Vorzeichen-Tiefe und für die Maximum-Bestimmung vernachlässigen, d.h. den Rechenaufwand dafür als konstant annehmen, so ergibt sich bei der Verallgemeinerung von Pseudocode 2 ein Rechenaufwand von $\mathcal{O}\left(\binom{N}{p-m}\right)$ (bzw. $\mathcal{O}(N^{p-m})$, wenn wir p, m nicht als Eingabe und somit als konstant betrachten) und bei der Verallgemeinerung von Pseudocode 3 und 4 $\mathcal{O}(2^{p-m} \binom{N}{p-m})$.

Die Für-Schleife in der Zeile 1 von Pseudocode 2, 3 und 4 durchlaufen wir $\binom{N}{2}$ -mal. Obwohl die Anzahl an Durchläufen somit lediglich quadratisch wächst, kann dies dennoch für große N einen großen Rechenaufwand bedeuten. Die Anzahl an Durchläufen könnten wir also entweder begrenzen, z.B. auf 100 Durchläufe, oder wir durchlaufen für große N stets lediglich einen gewissen Anteil, z.B. 10%, der insgesamt $\binom{N}{2}$ Durchläufe. Wenn wir also weniger als $\binom{N}{2}$ Durchläufe durchführen, benötigen wir eine Möglichkeit, zufällige Indexpaare $(i, j) \in \{1, \dots, N\}^2$ mit $i < j$ auszuwählen. Eine Möglichkeit dafür ist das Ziehen mit Zurücklegen, wobei wir dabei natürlich ein Indexpaar (i, j) mehrmals ziehen könnten. Eine andere Möglichkeit ist das Ziehen ohne Zurücklegen. Dafür ist es bei der Implementierung hilfreich, eine explizite Bijektion zwischen $\{1, \dots, \binom{N}{2}\}$ und $\{(i, j) \in \{1, \dots, N\}^2 \mid i < j\}$ für beliebige N zu nutzen. Eine mögliche Bijektion ist in Tabelle 3 gegeben.

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

1	2	...	$N-1$	N	...	$(N-1) + (N-2)$...	$\binom{N}{2}$
(1,2)	(1,3)	...	(1,N)	(2,3)	...	(2,N)	...	$(N-1, N)$

Tabelle 3: Mögliche Bijektion zwischen $\{1, \dots, \binom{N}{2}\}$ und $\{(i, j) \in \{1, \dots, N\}^2 \mid i < j\}$

Lemma 6.2 (persönliche Kommunikation von Malcherczyk). *Die Funktion*

$$f_N : \left\{ 1, 2, \dots, \binom{N}{2} \right\} \rightarrow \{(i, j) \in \{1, \dots, N\}^2 \mid i < j\},$$

gegeben durch

$$\begin{aligned} \text{GETi}_N(k) &= \left\lfloor \frac{2N+1}{2} - \sqrt{\left(\frac{2N+1}{2}\right)^2 - 2(N+k)} \right\rfloor - 1, \\ (f_N(k))_1 &= \text{GETi}_N(k), \\ (f_N(k))_2 &= (f_N(k))_1 + k - \left(N((f_N(k))_1 - 1) - \frac{((f_N(k))_1 - 1)(f_N(k))_1}{2} \right) \end{aligned}$$

für alle $k \in \{1, \dots, \binom{N}{2}\}$, entspricht der Bijektion in Tabelle 3.

Wenn man diese Funktion in R implementiert, so erhält man z.B. mit $N = 10$ und den Argumenten $k = 1, 2, \dots, \binom{10}{2}$ die folgende Ausgabe, bei der die Spalten den ausgegebenen Indexpaaren $((f_N(k))_1, (f_N(k))_2)$ entsprechen und die Spaltennummer dem Argument k .

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	1	1	1	1	1	1	1	1	1	2	2	2
[2,]	2	3	4	5	6	7	8	9	10	3	4	5
	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]	[,21]	[,22]		
[1,]	2	2	2	2	2	3	3	3	3	3		
[2,]	6	7	8	9	10	4	5	6	7	8		
	[,23]	[,24]	[,25]	[,26]	[,27]	[,28]	[,29]	[,30]	[,31]	[,32]		
[1,]	3	3	4	4	4	4	4	4	5	5		
[2,]	9	10	5	6	7	8	9	10	6	7		
	[,33]	[,34]	[,35]	[,36]	[,37]	[,38]	[,39]	[,40]	[,41]	[,42]		
[1,]	5	5	5	6	6	6	6	7	7	7		
[2,]	8	9	10	7	8	9	10	8	9	10		
	[,43]	[,44]	[,45]									
[1,]	8	8	9									
[2,]	9	10	10									

Beweis von Lemma 6.2. Seien $A_N := \{(i, j) \in \{1, \dots, N\}^2 \mid i < j\}$ und $B_N := \{1, \dots, \binom{N}{2}\}$.

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

1. Zunächst konstruieren wir eine (explizite) Bijektion g von A_N nach B_N wie in Tabelle 3. Dies bedeutet, dass wir wie folgt zuordnen wollen:

$$\begin{array}{rcl}
 (1, 2) & \mapsto & 1, \\
 (1, 3) & \mapsto & 2, \\
 & \vdots & \\
 (1, N) & \mapsto & N - 1, \\
 (2, 3) & \mapsto & N - 1 + 1, \\
 (2, 4) & \mapsto & N - 1 + 2, \\
 & \vdots & \\
 (2, N) & \mapsto & N - 1 + N - 2, \\
 (3, 4) & \mapsto & N - 1 + N - 2 + 1, \\
 & \vdots & \\
 (3, N) & \mapsto & N - 1 + N - 2 + N - 3, \\
 & \vdots & \\
 & \text{usw.} &
 \end{array}$$

Es gilt also

$$g(i, j) = \sum_{m=1}^{i-1} (N - m) + j - i = N(i - 1) - \frac{(i - 1)i}{2} + j - i.$$

Wir zeigen nun, dass $g : A_N \rightarrow B_N$ eine (wohldefinierte) bijektive Abbildung ist.

- (i) Für ein festes $i \in \{1, \dots, N - 1\}$ und $j, \tilde{j} \in \{i + 1, \dots, N\}$ mit $j < \tilde{j}$ gilt $g(i, j) < g(i, \tilde{j})$, da

$$\begin{aligned}
 g(i, j) &= N(i - 1) - \frac{(i - 1)i}{2} + j - i \\
 &< N(i - 1) - \frac{(i - 1)i}{2} + \tilde{j} - i \\
 &= g(i, \tilde{j}).
 \end{aligned}$$

- (ii) Es gilt $g(i, N) < g(i + 1, i + 2)$ für $i \in \{1, \dots, N - 2\}$ und insbesondere $g(i + 1, i + 2) - g(i, N) = 1$, da

$$\begin{aligned}
 g(i + 1, i + 2) - g(i, N) &= Ni - \frac{i(i + 1)}{2} + i + 2 - i - 1 \\
 &\quad - \underbrace{\left(N(i - 1) - \frac{(i - 1)i}{2} + N - i \right)}_{= -Ni + N + \frac{(i - 1)i}{2} - N + i},
 \end{aligned}$$

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

wobei wir diesen Term nun vereinfachen können und somit folgt

$$\begin{aligned} g(i+1, i+2) - g(i, N) &= -i \frac{i+1}{2} + i+1 + i \frac{i-1}{2} \\ &= i \left(\frac{i-1}{2} - \frac{i+1}{2} \right) + i+1 \\ &= -i + i+1 = 1. \end{aligned}$$

(iii) Es gilt $g(N-1, N) = \binom{N}{2}$, da

$$\begin{aligned} g(N-1, N) &= N(N-2) - \frac{(N-2)(N-1)}{2} + N - N + 1 \\ &= N^2 - 2N - \frac{1}{2}N^2 + \frac{3}{2}N - 1 + N - N + 1 \\ &= \frac{1}{2}N^2 - \frac{1}{2}N \\ &= \frac{1}{2}N(N-1) = \binom{N}{2}. \end{aligned}$$

(iv) Wenn wir nun (i) und (ii) sukzessiv anwenden, erhalten wir, dass $g(i, j) < g(\tilde{i}, \tilde{j})$ für alle $i, \tilde{i} \in \{1, \dots, N-1\}$ mit $i < \tilde{i}$ und $j \in \{i+1, \dots, N\}$, $\tilde{j} \in \{\tilde{i}+1, \dots, N\}$. Wegen (iii) ist $\binom{N}{2}$ dann auch der Maximalwert von g .

Somit haben wir nun die Wohldefiniiertheit und Injektivität von g gezeigt. Wegen $|A_N| = |B_N|$ folgt aus der Injektivität auch die Surjektivität, d.h. g ist bijektiv.

2. Da g umkehrbar ist, konstruieren wir nun die Umkehrabbildung g^{-1} . Betrachte hierbei

$$g(i, j) = \underbrace{N(i-1) - \frac{(i-1)i}{2}}_{:=h(i)} + j - i.$$

Die Idee ist nun, dass wir aus dem Funktionswert $h(i)$ bereits das Argument i

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

folgern können. Dies kann wie folgt illustriert werden:

(i, j)	$h(i) = \tilde{z}$	$g(i, j) = z$
$(1, 2) \mapsto$	0	$+ 2 - 1 = 1,$
$(1, 3) \mapsto$	0	$+ 3 - 1 = 2,$
\vdots		
$(1, N) \mapsto$	0	$+ N - 1 = N - 1,$
$(2, 3) \mapsto$	$N - 1$	$+ 3 - 2 = N,$
$(2, 4) \mapsto$	$N - 1$	$+ 4 - 2 = N + 1,$
\vdots		
$(2, N) \mapsto$	$N - 1$	$+ N - 2 = 2N - 3,$
$(3, 4) \mapsto$	$N - 1 + N - 2$	$+ 4 - 3 = 2N - 2,$
\vdots		
$(5, 7) \mapsto$	$N - 1 + N - 2 + N - 3 + N - 4$	$+ 7 - 5 = 4N - 8,$
\vdots		
usw.		

D.h. für $g(i, j) \in \{1, \dots, N - 1\}$ ist $i = 1$, bzw. $(g^{-1})_1(\{1, \dots, N - 1\}) = \{1\}$, $(g^{-1})_1(\{N, \dots, 2N - 3\}) = \{2\}$, usw. Anhand dieser Illustration folgt also, dass der Wert von h bereits die erste Koordinate i bestimmt. Wir zeigen nun, dass die Funktion $h : \{1, \dots, N - 1\} \rightarrow h(\{1, \dots, N - 1\})$ umkehrbar ist, wobei

$$h(\{1, \dots, N - 1\}) = \left\{ 0, N - 1, 2N - 3, 3N - 6, \dots, \binom{N}{2} - 1 \right\} =: C_N.$$

Nun wollen wir die Umkehrvorschrift von h auf C_N finden. Betrachte also die Funktion

$$h : \{1, \dots, N - 1\} \rightarrow h(\{1, \dots, N - 1\}), \quad i \mapsto N(i - 1) - \frac{(i - 1)i}{2}.$$

Nun gilt

$$\begin{aligned} \tilde{z} &= N(i - 1) - \frac{(i - 1)i}{2} \\ &= Ni - N - \frac{1}{2}i^2 + \frac{i}{2} \\ \Leftrightarrow i^2 - i - 2Ni + 2N + 2\tilde{z} &= 0 \\ \Leftrightarrow i_{1/2} &= \frac{1 + 2N}{2} \pm \sqrt{\frac{(1 + 2N)^2}{4} - 2N - 2\tilde{z}}, \end{aligned}$$

wobei wir hier $i = N + \frac{1}{2} - \sqrt{\frac{1}{4} + N + N^2 - 2N - 2\tilde{z}}$ wählen müssen, um Werte in $\{1, \dots, N - 1\}$ zu erhalten. Z.B. erhalten wir für $\tilde{z} = 0 \in C_N$

$$\begin{aligned} i &= N + \frac{1}{2} - \sqrt{\frac{1}{4} - N + N^2} \\ &= N + \frac{1}{2} - \sqrt{\left(N - \frac{1}{2}\right)^2} = 1, \end{aligned}$$

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

für $\tilde{z} = N - 1 \in C_N$

$$\begin{aligned}
 i &= N + \frac{1}{2} - \sqrt{\frac{1}{4} - N + N^2 - 2(N - 1)} \\
 &= N + \frac{1}{2} - \sqrt{\frac{1}{4} - N + N^2 - 2N + 2} \\
 &= N + \frac{1}{2} - \sqrt{N^2 - 3N + \frac{9}{4}} \\
 &= N + \frac{1}{2} - \sqrt{\left(N - \frac{3}{2}\right)^2} \\
 &= \frac{1}{2} + \frac{3}{2} = 2,
 \end{aligned}$$

für $\tilde{z} = \binom{N}{2} - 1 \in C_N$

$$\begin{aligned}
 i &= N + \frac{1}{2} - \sqrt{\frac{1}{4} - N + N^2 - 2\left(\binom{N}{2} - 1\right)} \\
 &= N + \frac{1}{2} - \sqrt{\frac{9}{4}} \\
 &= N - 1,
 \end{aligned}$$

usw. Allgemein gilt

$$\begin{aligned}
 i &= N + \frac{1}{2} - \sqrt{\left(N + \frac{1}{2} - i\right)^2} \\
 &= N + \frac{1}{2} - \sqrt{\left(\left(N - \frac{1}{2}\right) + (1 - i)\right)^2} \\
 &= N + \frac{1}{2} - \sqrt{\left(N - \frac{1}{2}\right)^2 + 2\left(N - \frac{1}{2}\right)(1 - i) + (1 - i)^2} \\
 &= N + \frac{1}{2} - \sqrt{\left(N - \frac{1}{2}\right)^2 - 2\left(N(i - 1) + \frac{1}{2} - \frac{i}{2} - \frac{1}{2} + i - \frac{i^2}{2}\right)} \\
 &= N + \frac{1}{2} - \sqrt{\left(N - \frac{1}{2}\right)^2 - 2h(i)}
 \end{aligned}$$

für $i = 1, \dots, N - 1$. Da wir aber den Definitionsbereich $B_N = \{1, \dots, \binom{N}{2}\}$ anstatt $C_N = \{0, N - 1, 2N - 3, 3N - 6, \dots, \binom{N}{2} - 1\}$ benutzen wollen, müssen wir die Funktion h^{-1} ein wenig abändern. Anstatt

$$h^{-1} : C_N \rightarrow \{1, \dots, N - 1\}, \quad \tilde{z} \mapsto N + \frac{1}{2} - \sqrt{\left(N - \frac{1}{2}\right)^2 - 2\tilde{z}} \quad (6.4)$$

betrachten wir nämlich

$$\begin{aligned} \widetilde{h}^{-1} : B_N &\rightarrow \{1, \dots, N-1\}, z \mapsto \left\lfloor N + \frac{1}{2} - \sqrt{\left(N - \frac{1}{2}\right)^2 - 2z} \right\rfloor - 1 \\ &= \left\lfloor \frac{2N+1}{2} - \sqrt{\left(\frac{2N+1}{2}\right)^2 - 2(N+z)} \right\rfloor - 1. \end{aligned}$$

Diese Änderung liefert das gewünschte Ergebnis, da der obige Term in (6.4) streng monoton wachsend in \tilde{z} ist. Wie wir außerdem aus der obigen Illustration der Rolle von h in der Abbildung g entnehmen können, lässt sich die zweite Koordinate j mit Hilfe der ersten Koordinate i und der Funktion h durch $j = i + (z - h(i))$ herleiten.

Die Umkehrfunktion g^{-1} von g ist als gegeben durch

$$g^{-1} : B_N \rightarrow A_N, z \mapsto \left(\widetilde{h}^{-1}(z), z + \widetilde{h}^{-1}(z) - h(\widetilde{h}^{-1}(z)) \right),$$

was gerade der Funktion f_N in Lemma 6.2 entspricht.

□

Wegen dieser expliziten Bijektion aus Lemma 6.2 reicht es also aus, mit der R-Funktion `sample` mit Argumenten `x=1:choose(N,2)` und `replace=FALSE` eine bestimmte Anzahl an (paarweise verschiedenen) Zahlen aus $\{1, \dots, \binom{N}{2}\}$ zu erhalten, um eine bestimmte Anzahl Indexpaare aus $\{(i, j) \in \{1, \dots, N\} \mid i < j\}$ durch Ziehen ohne Zurücklegen zu gewinnen.

Eine alternativer Ansatz wäre die Benutzung der R-Funktion `combn`. Durch ein Ausführen von `combn(1:N, m=2)` erhalten wir alle Indexpaare $(i, j) \in \{1, \dots, N\}^2$ mit $i < j$ in der gleichen Reihenfolge wie in Tabelle 3. So erhalten wir z.B. mit dem Ausführen von `combn(1:10, m=2)` die gleiche Ausgabe wie die obige. Diese Erzeugung aller Indexpaare ist aber im Vergleich zu der expliziten Bijektion in Lemma 6.2 unvorteilhaft, wenn wir lediglich einen relativen kleinen Anteil an Indexpaaren zufällig auswählen wollen. Für sehr große N benötigt diese Benutzung von `combn` auch eine relativ lange Laufzeit. Wegen dieser Gründe verfolgen wir diesen alternativen Ansatz nicht weiter. Allgemein könnten wir aber natürlich argumentieren, dass es zwischen dem Ziehen mit und ohne Zurücklegen sowieso keinen großen Unterschied gibt, wenn die Stichprobengröße relativ klein im Vergleich zur Größe der Grundgesamtheit ist.

6.3 Berechnung der 3-Vorzeichen-Tiefe

Für die effiziente Berechnung der 3-Vorzeichen-Tiefe benutzen wir die Funktion `calcDepth` mit dem Argument `k=3` aus dem R-Paket `GSignTest`, s. Horn, 2021a. Diese Implementierung beruht auf Malcherczyk, 2022. Hierbei ist zu beachten, dass diese Funktion nicht

übereinstimmt mit der Definition 3.4 der Vorzeichen-Tiefe. $\text{calcDepth}(\text{res}, K=3)$ entfernt nämlich als ersten Schritt die Nullen aus dem Residuenvektor res , und gibt dann lediglich für diesen verkürzten Vektor die korrekte 3-Vorzeichen-Tiefe zurück. Wenn wir also für Residuen (r_1, \dots, r_N) mit M Null-Residuen die korrekte 3-Vorzeichen-Tiefe erhalten möchten, müssen wir die Ausgabe von $\text{calcDepth}(\text{res}, K=3)$ mit dem Faktor $\binom{N-M}{3} / \binom{N}{3}$ korrigieren.

Die Vorzeichen-Tiefe wurde aus der Regressions-Tiefe hergeleitet. Nach Definition der Regressions-Tiefe ist ein Residuen- K -Tupel mit einem Null-Residuum kein Nonfit, d.h. ein Fit, weil es keine Gerade gibt, die überall kleinere Residuen besitzt, s. Rousseeuw und Hubert, 1999. Dies ist die Motivation dafür, eine alternative K -Vorzeichen-Tiefe einzuführen, bei der die Residuen- K -Tupel mit mindestens einem Null-Residuum mitgezählt werden.

Definition 6.3 (Alternative K -Vorzeichen-Tiefe). *Für ein $K \in \mathbb{N} \setminus \{1\}$ definieren wir die alternative K -Vorzeichen-Tiefe von (r_1, \dots, r_N) durch*

$$\begin{aligned} \tilde{d}_K(r_1, \dots, r_N) &:= \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \dots < n_K \leq N} \mathbf{1}\{(r_{n_1}, \dots, r_{n_K}) \in \tilde{\mathcal{A}}_{N,K}\}, \\ \tilde{\mathcal{A}}_{N,K} &:= \{(z_1, \dots, z_K) \in \mathbb{R}^K \mid z_i \cdot z_{i+1} < 0 \text{ für alle } i = 1, \dots, K-1 \\ &\quad \text{oder } z_i = 0 \text{ für ein } i \in \{1, \dots, K-1\}\}. \end{aligned}$$

Die Definitionen von d_K und \tilde{d}_K stimmen fast sicher überein. Wegen $R_n(\theta) \neq 0$ \mathbb{P}_θ -f.s. gilt nämlich

$$\tilde{d}_K(R_1(\theta), \dots, R_N(\theta)) = d_K(R_1(\theta), \dots, R_N(\theta)) \quad \mathbb{P}_\theta\text{-f.s.}$$

Somit gilt (3.12) auch mit \tilde{d}_3 anstelle von d_3 , d.h.

$$\frac{(N-1)(N-2)}{N} (\tilde{d}_3(R_1(\theta), \dots, R_N(\theta)) - 0.25) = \Psi_3(\mathcal{W}^N(\theta)) \quad \mathbb{P}_\theta\text{-f.s.}$$

Diese Beziehung zeigt, dass wir die obigen Verfahren für die Berechnung von

$$\sup_{\theta \in \Theta_0} d_3(r_1(\theta), \dots, r_N(\theta))$$

auch auf analoge Weise mit \tilde{d}_3 anstelle von d_3 durchführen können, um den 3-Vorzeichen-Tiefe-Test anwenden zu können. Es gilt die folgende, leicht nachvollziehbare Beziehung zwischen d_K und \tilde{d}_K .

Lemma 6.4. *Der Residuenvektor (r_1, \dots, r_N) habe M Null-Residuen. Dann gilt*

$$\tilde{d}_K(r_1, \dots, r_N) = \frac{\binom{N}{K} d_K(r_1, \dots, r_N) + \sum_{i=1}^K \binom{M}{i} \binom{N-M}{K-i}}{\binom{N}{K}},$$

und insbesondere für $K = 3$

$$\tilde{d}_3(r_1, \dots, r_N) = \frac{\binom{N}{3} d_3(r_1, \dots, r_N) + \binom{M}{1} \binom{N-M}{2} + \binom{M}{2} \binom{N-M}{1} + \binom{M}{3}}{\binom{N}{3}}.$$

6.4 Vergleich von dem Vorgehen in Pseudocode 2 und 4

In diesem und den folgenden Abschnitten seien die erklärenden Variablen $x_{1;N}, \dots, x_{N;N}$ deterministisch und äquidistant auf $[S, T] = [-3, 3]$, wie in (5.3) erläutert. Die erhaltenen Werte in den Simulationen dieses Abschnittes und der folgenden Abschnitte erhalten wir stets durch 1000 Simulationsdurchläufe.

Wir vergleichen nun das Vorgehen in Pseudocode 2 ($PC2$) und 4 ($PC4$). Um den Rechenaufwand zu verringern, durchlaufen wir in Zeile 1 von Pseudocode 2 und 4 nicht alle Für-Schleifen, sondern wählen mit Hilfe von Lemma 6.2 N Indexpaare durch Ziehen ohne Zurücklegen und rechnen dann lediglich mit diesen N Indexpaaren. Dieses Vorgehen erklärt sich durch die Resultate in Abschnitt 6.5. Hier seien die Störgrößen E_1, \dots, E_N standardnormalverteilte Zufallsvariablen und der wahre Parameter $\theta^* \in \Theta$ sei bekannt, d.h.

$$Y_n = \theta_0^* + \theta_1^* x_{n;N} + \theta_2^* x_{n;N}^2 + E_n, \quad n = 1, \dots, N.$$

In Abbildung 4 betrachten wir für verschiedene Datenmengen N zunächst die durchschnittlichen, von dem entsprechenden Algorithmus ausgegebenen Vorzeichen-Tiefen.

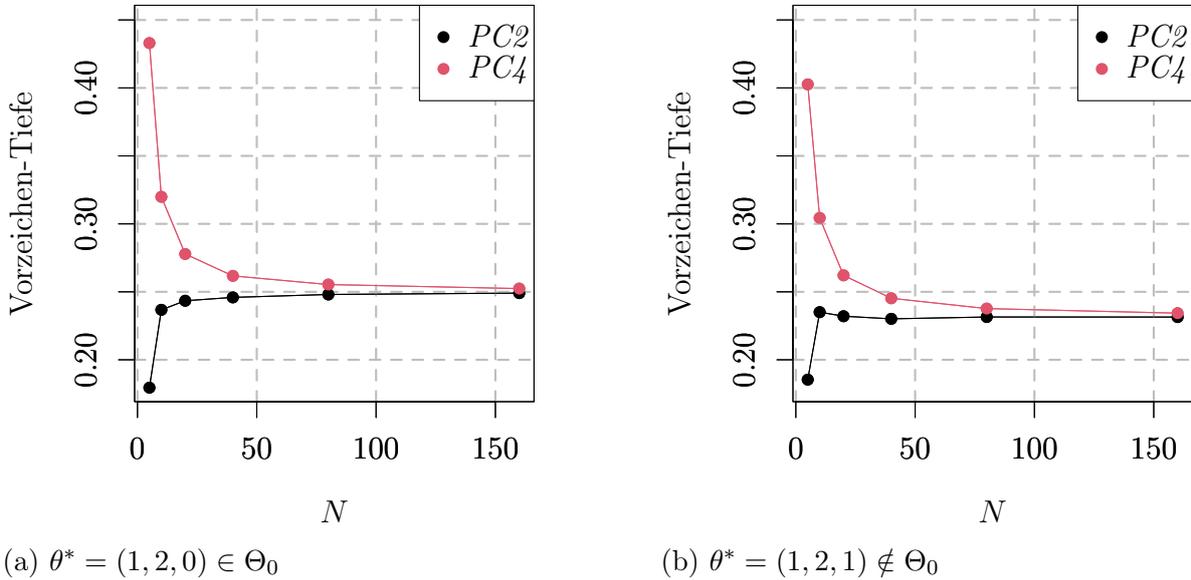


Abbildung 4: Vergleich von dem Vorgehen in Pseudocode 2 ($PC2$) und 4 ($PC4$) in Hinblick auf die ausgegebenen Vorzeichen-Tiefen

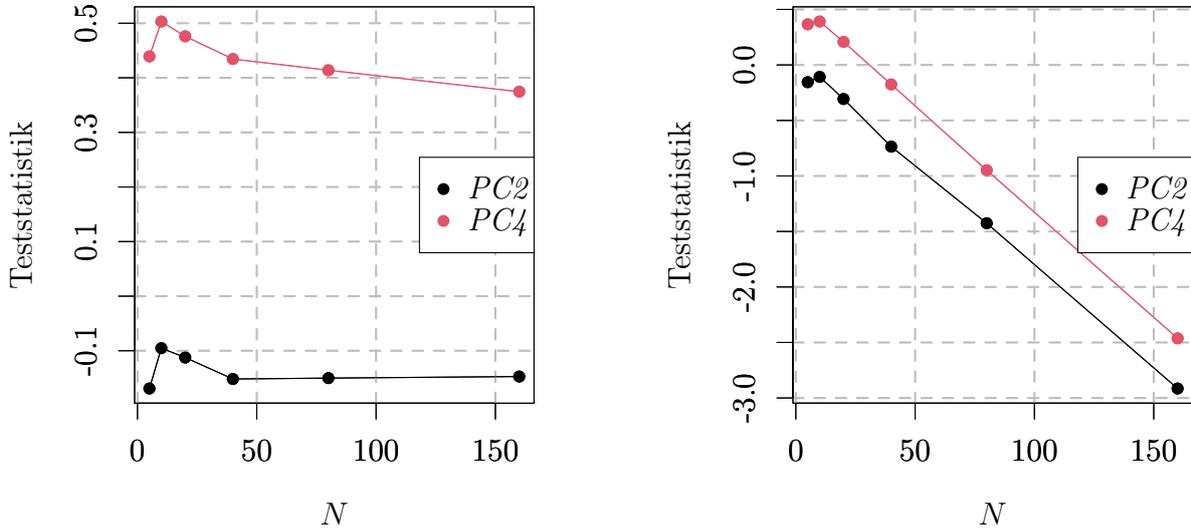
In Abbildung 4a ist der wahre Parameter $\theta^* = (1, 2, 0) \in \Theta_0$ und in 4b $\theta^* = (1, 2, 1) \notin \Theta_0$. Natürlich erhalten wir für das $\theta^* \in \Theta_0$ im Schnitt größere Vorzeichen-Tiefen als für das $\theta^* \notin \Theta_0$. Für alle alle Datenmengen N ist die zugehörige Vorzeichen-Tiefe von $PC4$ größer als die von $PC2$. Dies ist erwartbar, da für einen Residuenvektor mit genau zwei Null-Residuen das Ersetzen dieser Null-Residuen mit Werten aus $\{-1, +1\}$ die zugehörige 3-Vorzeichen-Tiefe nach Definition nicht verringert und höchstens um den

Summanden

$$\frac{\binom{2}{1}\binom{N-2}{2} + \binom{2}{2}\binom{N-2}{1}}{\binom{N}{3}} = \frac{2\binom{N-2}{2} + (N-2)}{\binom{N}{3}} \in O\left(\frac{1}{N}\right) \quad (6.5)$$

erhöht. Für kleine Datenmengen N ist der Unterschied der ausgegebenen Vorzeichen-Tiefen von Pseudocode 2 und 4 relativ groß, während diese (absolute) Differenz für große N beliebig klein wird. Auch dies ist erwartbar, da der Term in (6.5) für $N \rightarrow \infty$ gegen Null konvergiert. Hierbei ist jedoch zu beachten, dass wir in der Teststatistik aufgrund von (3.12) das $(N-1)(N-2)/N$ -fache der Vorzeichen-Tiefe verwenden. Wegen $(N-1)(N-2)/N \in O(N)$ können wir also annehmen, dass wir in der Teststatistik für große N eine Differenz von $O(1)$ zwischen PC_4 und PC_2 erhalten.

Dies motiviert uns dazu, die durchschnittliche Teststatistik in Abbildung 5 zu betrachten. Mit Teststatistik ist dabei $(N-1)(N-2)/N \cdot (d-0.25)$ gemeint, wobei d die Vorzeichen-Tiefe ist, die von PC_2 bzw. PC_4 zurückgegeben wird.



(a) $\theta^* = (1, 2, 0) \in \Theta_0$

(b) $\theta^* = (1, 2, 1) \notin \Theta_0$

Abbildung 5: Vergleich von dem Vorgehen in Pseudocode 2 (PC_2) und 4 (PC_4) in Hinsicht auf die Teststatistik

Mit Abbildung 5 sehen wir nun also unsere Annahme bestätigt, dass wir in Hinsicht auf die Teststatistik eine Differenz zwischen den Werten von PC_2 und PC_4 erkennen können, welche für $N \rightarrow \infty$ nicht gegen Null konvergiert. In Abbildung 5b, wo der wahre Parameter $\theta^* = (1, 2, 1) \notin \Theta_0$ ist, scheint die durchschnittliche Teststatistik monoton fallend für $N \rightarrow \infty$ gegen $-\infty$ zu divergieren. Dies deckt sich mit den theoretischen Resultaten aus Kapitel 4 und 5. So gelten wegen Satz 5.4 die Bedingungen (5.5) und (5.6), woraus wiederum aus Satz 5.2 die Gültigkeit von (C2) folgt. Dies bedeutet, dass aufgrund von Korollar 4.7 auch (C1) gilt. Die Folgerung aus Satz 4.1 können wir dann so interpretieren, dass die Teststatistik des 3-Vorzeichen-Tiefe-Tests für $N \rightarrow \infty$ gegen $-\infty$ divergiert.

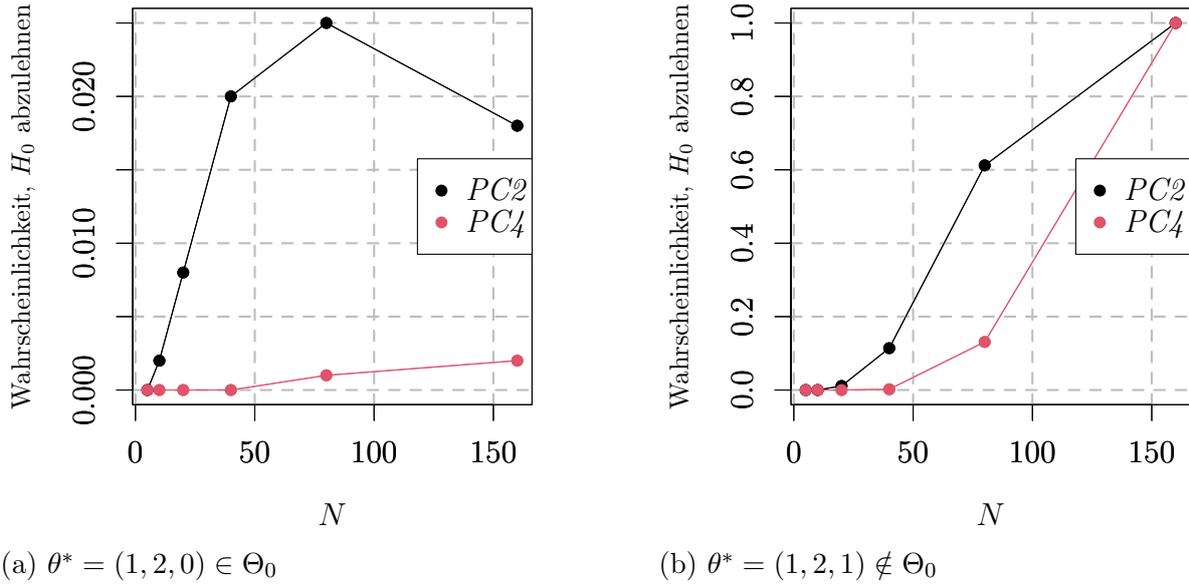


Abbildung 6: Vergleich von dem Vorgehen in Pseudocode 2 (PC_2) und 4 (PC_4) in Hinsicht auf die Wahrscheinlichkeit, H_0 abzulehnen

In Abbildung 6 vergleichen wir nun das Vorgehen in Pseudocode 2 (PC_2) und 4 (PC_4) in Hinsicht auf die Wahrscheinlichkeit, H_0 abzulehnen. Dabei berechnen wir zunächst mit dem Datenpunktpaar-Verfahren des Pseudocodes 2 bzw. 4 annäherungsweise die maximale 3-Vorzeichen-Tiefe d . Wegen (3.12) lehnen wir dann die Nullhypothese $H_0 : \theta_2^* = 0$ genau dann ab, wenn

$$\frac{(N-1)(N-2)}{N}(d-0.25) < -1.2545411,$$

wobei -1.2545411 das 0.05-Quantil aus Tabelle 1 ist. In Abbildung 6a sehen wir zunächst, dass wir bei beiden Vorgehensweisen das Niveau 0.05 einhalten. Dabei ist jedoch die Fehlerwahrscheinlichkeit 1. Art bei PC_4 sehr viel kleiner als bei PC_2 . Dies ist ein negatives Ergebnis für PC_4 , da dies sich negativ auf die Fehlerwahrscheinlichkeit 2. Art im Falle $0 < |\theta_2^*| \ll 1$ auswirkt. So ist die Konvergenzgeschwindigkeit der Trennschärfe gegen 1 von PC_4 langsamer als die von PC_2 in Abbildung 6b, d.h. in Hinsicht auf die Konsistenz ist das Vorgehen PC_2 besser als PC_4 .

Wegen dieser Resultate werden wir im Folgenden nicht weiter das Vorgehen in Pseudocode 4 verwenden, sondern lediglich das in Pseudocode 2.

6.5 Niveau des praktisch implementierten 3-Vorzeichen-Tiefe-Tests

Wir möchten nun den 3-Vorzeichen-Tiefe-Test in der Praxis untersuchen. Dabei berechnen wir mit dem Datenpunktpaar-Verfahren des Pseudocodes 2 annäherungsweise die maximale 3-Vorzeichen-Tiefe d . Wegen (3.12) lehnen wir dann die Nullhypothese

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

$H_0 : \theta_2^* = 0$ genau dann ab, wenn

$$\frac{(N-1)(N-2)}{N}(d-0.25) < -1.2545411,$$

wobei -1.2545411 das 0.05-Quantil aus Tabelle 1 ist. Da wir aber wegen der Begründung in Abschnitt 6.2.2 nicht alle Indexpaare in der Anwendung von Pseudocode 2 betrachten wollen, d.h. nicht alle Für-Schleifen in Zeile 1 durchlaufen wollen, stellt sich nun die Frage, wie viele Indexpaare wir betrachten müssen, damit dieser praktisch implementierte 3-Vorzeichen-Tiefe-Test das Niveau 0.05 einhält.

		Datenmenge N				
		5	14	37	101	275
Anteil bzw. Anzahl betrachteter Indexpaare in der Anwendung von Pseudocode 2	1%	0	0.678	0.483	0.154	0.006
	2%	0	0.497	0.239	0.020	0.000
	3%	0	0.312	0.116	0.004	0.000
	4%	0	0.213	0.065	0.000	0.000
	5%	0	0.140	0.018	0.000	0.000
	$\min\{100, \binom{N}{2}\} \#$	0	0.000	0.000	0.024	0.256
$N \#$	0	0.002	0.012	0.024	0.026	

Tabelle 4: Fehlerwahrscheinlichkeit 1. Art des praktisch implementierten 3-Vorzeichen-Tiefe-Tests

In Tabelle 4 haben wir also nun mit einer Simulation, bei der wir standardnormalverteilte Störgrößen und den wahren Parameter $\theta^* = (1, 2, 0) \in \Theta_0$ verwendet haben, die Fehlerwahrscheinlichkeit 1. Art des praktisch implementierten 3-Vorzeichen-Tiefe-Tests in Abhängigkeit der Datenmenge N und des Anteils bzw. der Anzahl betrachteter Indexpaare in der Anwendung von Pseudocode 2 berechnet. Dabei haben wir die Indexpaare durch Ziehen ohne Zurücklegen wie oben beschrieben ausgewählt. Mit dem Anteil betrachteter Indexpaare ist gemeint, dass wir bei einem Anteil von z.B. 1% eine Anzahl von $\lceil 1\% \cdot \binom{N}{2} \rceil$ Indexpaaren betrachten. Die hier gewählten Datenmengen N_i , $i = 1, \dots, 5$, wurden erhalten durch $N_i = \lfloor 5 \cdot 2.72407^{i-1} \rfloor$, wobei 2.72407 gewählt wurde, da $(750/5)^{1/5} \approx 2.72407$ und somit auch $\lfloor 5 \cdot 2.72407^5 \rfloor = 750$. Es war gedacht, auch die Datenmenge $N = 750$ in der Simulation zu benutzen, jedoch wurde dieser Gedanke aufgrund einer langen Laufzeit verworfen.

Hier können wir sehen, dass das Betrachten eines gewissen (festen) Anteils der Indexpaare nicht vorteilhaft ist. Auch bei einem sehr kleinen Anteil ist die Fehlerwahrscheinlichkeit 1. Art für große Datenmengen N sehr viel kleiner als 0.05, was sich dann negativ auf die Fehlerwahrscheinlichkeit 2. Art im Falle $0 < |\theta_2^*| \ll 1$ auswirkt. Dass wir hier Fehlerwahrscheinlichkeiten 1. Art erhalten, die kleiner als 0.05 sind, deckt sich mit der Ungleichung (3.9). Bei kleinen Datenmengen N führen die kleinen Anteile, welche bei großen Datenmengen noch zu extrem konservativen Testentscheidungen führten, jedoch zu einem Nichteinhalten des Niveaus 0.05.

Wenn wir die Anzahl betrachteter Indexpaare auf 100 begrenzen, steigt die Fehlerwahrscheinlichkeit 1. Art mit wachsendem N , sodass wir bei $N = 275$ schon eine Fehlerwahrscheinlichkeit 1. Art von 0.256 erhalten, was natürlich nicht akzeptabel ist. Wenn wir aber stets N Indexpaare betrachten, erhalten wir bei den hier betrachteten Datenmengen N stets Fehlerwahrscheinlichkeiten 1. Art, die kleiner als 0.05 sind. Wegen dieses Resultats werden wir im Folgenden stets N Indexpaare bei der annäherungsweise Berechnung der maximalen 3-Vorzeichen-Tiefe im praktisch implementierten 3-Vorzeichen-Tiefe-Test betrachten.

Es ist klar, dass der Test bei $N = 5$ nie die Nullhypothese ablehnt, da

$$\frac{(5-1)(5-2)}{5} \cdot (d - 0.25) \geq -\frac{(5-1)(5-2)}{5} \cdot 0.25 = -0.6 \geq -1.2545411$$

für alle $d \in [0, 1]$.

		Datenmenge N				
		5	14	37	101	275
Anteil bzw. Anzahl betrachteter Indexpaare in der Anwendung von Pseudocode 2	1%	0	0.717	0.641	0.938	1
	2%	0	0.470	0.401	0.891	1
	3%	0	0.338	0.256	0.864	1
	4%	0	0.239	0.169	0.814	1
	5%	0	0.173	0.104	0.769	1
	$\min\{100, \binom{N}{2}\} \#$	0	0.000	0.000	0.905	1
$N \#$	0	0.007	0.074	0.881	1	

Tabelle 5: Wahrscheinlichkeit, H_0 abzulehnen, des praktisch implementierten 3-Vorzeichen-Tiefe-Tests

Außerdem betrachten wir nun im Falle, dass der wahre Parameter $\theta^* = (1, 2, 1) \notin \Theta_0$ ist, die Wahrscheinlichkeit, H_0 abzulehnen, des praktisch implementierten 3-Vorzeichen-Tiefe-Tests in Tabelle 5. Es zeigt sich, dass die Trennschärfe desto schneller gegen 1 für $N \rightarrow \infty$ konvergiert, je kleiner der Anteil bzw. die Anzahl betrachteter Indexpaare ist. Obwohl wir in Tabelle 4 bei $N = 275$ und den Anteilen 1%, ..., 5% Fehlerwahrscheinlichkeiten 1. Art beobachten können, die extrem klein sind bzw. sogar Null sind, erreichen wir in den gleichen Zellen in Tabelle 5 mit $\theta^* = (1, 2, 1)$ eine Trennschärfe von 1. Dies ist ein sehr positives Resultat und spricht für unser Verfahren. Hierbei ist aber natürlich zu beachten, dass wir mit einem wahren Parameter $\theta^* = (\theta_0^*, \theta_1^*, \theta_2^*)$ mit $0 < |\theta_2^*| \ll 1$ ein schlechteres Ergebnis erhalten würden, d.h. dann wäre auch ein negativer Einfluss der sehr niedrigen Fehlerwahrscheinlichkeiten 1. Art auf die Fehlerwahrscheinlichkeiten 2. Art deutlicher erkennbar.

6.6 Vergleich von verschiedenen praktischen Tests

Wir vergleichen nun den praktisch implementierten 3-Vorzeichen-Tiefe-Test mit verschiedenen anderen Tests in Hinsicht auf Robustheit, Güte und Konsistenz. Wir betrachten die folgenden Tests.

dp: Bei dem Test, den wir mit *dp* bezeichnen, berechnen wir zunächst mit dem Datenpunktpaar-Verfahren des Pseudocodes 2, wobei wir jedoch die Für-Schleife in Zeile 1 N -mal durchlaufen wie oben beschrieben, annäherungsweise die maximale 3-Vorzeichen-Tiefe d . Wegen (3.12) lehnen wir bei dem Test *dp* dann die Nullhypothese $H_0 : \theta_2^* = 0$ genau dann ab, wenn

$$\frac{(N-1)(N-2)}{N}(d-0.25) < -1.2545411,$$

wobei -1.2545411 das 0.05-Quantil aus Tabelle 1 ist.

dp.alt: Bei dem Test, den wir mit *dp.alt* bezeichnen, gehen wir ähnlich wie bei *dp* vor. Hierbei benutzen wir aber bei Pseudocode 2 nicht die Definition 3.4 der 3-Vorzeichen-Tiefe, sondern die alternative Definition 6.3 der Vorzeichen-Tiefe, d.h. wir ersetzen d_3 durch \tilde{d}_3 .

lm: Bei dem Test, den wir mit *lm* bezeichnen, benutzen wir den F -Test für die Nullhypothese $H_0 : \theta_2^* = 0$. Den zugehörigen p -Wert erhalten wir durch das Ausführen des R-Befehls `anova(lm(y ~ 1+x+I(x^2))) [2, "Pr(>F)"]`. Falls dieser p -Wert kleiner als 0.05 ist, lehnen wir die Nullhypothese ab. Dieser Test ist nicht robust.

lmRob: Bei dem Test, den wir mit *lmRob* bezeichnen, benutzen wir die R-Funktion `lmRob` aus dem Paket `robust`, s. Wang et al., 2022. Mit dieser Funktion können wir eine robuste Regression durchführen, die eine hohe Bruchpunktresistenz und Effizienz hat. Den p -Wert eines Wald-Tests für die Nullhypothese $H_0 : \theta_2^* = 0$, der auf robusten Regressionsschätzungen basiert, erhalten wir durch das Ausführen von `anova(lmRob(y~1+x+I(x^2))) [3, "Pr(F)"]`. Falls dieser p -Wert kleiner als 0.05 ist, lehnen wir die Nullhypothese ab.

Aufgrund der Annahmen an unser Modell können wir annehmen, dass die Residuenvektoren (r_1, \dots, r_N) , die wir im Datenpunktpaar-Verfahren des Pseudocodes 2 betrachten, stets genau zwei Null-Einträge haben. Somit gilt

$$\begin{aligned} & \frac{(N-1)(N-2)}{N} \tilde{d}_3(r_1, \dots, r_N) \\ &= \frac{(N-1)(N-2)}{N} d_3(r_1, \dots, r_N) + \underbrace{\frac{(N-1)(N-2)}{N} \frac{\binom{2}{1} \binom{N-2}{2}}{\binom{N}{3}}}_{\xrightarrow{N \rightarrow \infty} 6} + \underbrace{\frac{(N-1)(N-2)}{N} \frac{\binom{2}{2} \binom{N-2}{1}}{\binom{N}{3}}}_{\xrightarrow{N \rightarrow \infty} 0}. \end{aligned}$$

Dies könnte uns dazu motivieren, als kritischen Wert bei dem Test *dp.alt* $-1.2545411+6$ statt -1.2545411 zu verwenden. Jedoch führt dies in Simulationen mit unserem in R implementierten Datenpunktpaar-Verfahren noch bei $N = 200$ zu extrem hohen Fehlerwahrscheinlichkeiten 1. Art, nämlich über 0.8 bei standardnormalverteilten Störgrößen, sodass diese Idee hier nicht weiter verfolgt wird.

An dieser Stelle sei jedoch auf ein Problem bzw. Fehler unser in R implementierten Funktionen hingewiesen. Wir berechnen durch $r = y - (\theta_0 + \theta_1 x)$ den Residuenvektor,

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

wobei θ_0, θ_1 die entsprechenden Koeffizienten der linearen Funktion durch zwei gegebene Datenpunkte wie in Pseudocode 2 beschrieben sind. Hierbei wurde leider nicht beachtet, dass dies natürlich nicht bedeutet, dass dann in R wegen Rundungsfehlern nicht zwei Einträge des Residuenvektors exakt Null sind. So kann es z.B. sein, dass ein Eintrag, der in der Theorie exakt Null sein sollte, in R ungefähr gleich $1.110223 \cdot 10^{-16}$ ist. Somit stimmen die resultierenden Vorzeichen-Tiefen nicht unbedingt exakt mit der Theorie überein. Genauer gesagt bedeutet dies, dass die resultierenden Vorzeichen-Tiefen größer gleich als die theoretischen Vorzeichen-Tiefen sind und ggf. sogar echt größer sind. Außerdem wurde bei der Implementierung der alternativen Vorzeichen-Tiefe in R als Funktion der Vergleichsoperator `==` verwendet, um Null-Einträge zu identifizieren. Dieser Vergleichsoperator überprüft aber exakte Gleichheit, sodass wir z.B. bei einem Ausführen von `1.110223e-16==0` die Ausgabe `FALSE` erhalten. Dies bedeutet, dass die resultierenden alternativen Vorzeichen-Tiefen kleiner gleich als die theoretischen alternativen Vorzeichen-Tiefen sind, und ggf. sogar echt kleiner sind. Somit könnte dies ein Grund für die oben genannte sehr schlechte Fehlerwahrscheinlichkeit 1. Art sein, wo wir als kritischen Wert bei dem Test `dp.alt -1.2545411 + 6` statt `-1.2545411` verwendet haben.

In Abbildung 7 ist die Problematik anschaulich dargestellt. Wir betrachten die Gerade durch die Datenpunkte (x_2, y_2) und (x_6, y_6) . Lediglich die Datenpunkte (x_1, y_1) und (x_3, y_3) liegen über der Gerade, was theoretisch zu einer 3-Vorzeichen-Tiefe von Null führen sollte. Wegen Rundungsfehler sind aber die Residuen r_2 und r_6 nicht exakt Null, sondern $r_2 \approx 4.440892 \cdot 10^{-16}$ und $r_6 \approx 6.661338 \cdot 10^{-16}$, sodass wir doch eine 3-Vorzeichen-Tiefe von ca. 0.03 erhalten.

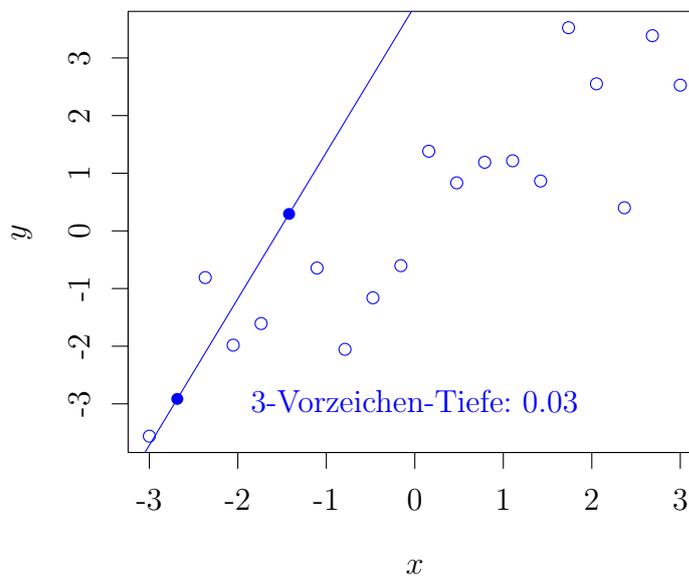


Abbildung 7: Illustration der Auswirkung von Rundungsfehlern

6.6.1 Robustheit in Hinsicht auf die Fehlerwahrscheinlichkeit 1. Art

Wir wollen nun zunächst die Robustheit der verschiedenen obigen Tests in Hinsicht auf die Fehlerwahrscheinlichkeit 1. Art untersuchen und vergleichen.

Dafür sei der wahre Parameter $\theta^* = (1, 1, 0) \in \Theta_0$ und die Störgrößen E_1, \dots, E_N $(1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(0, 100^2)$ -verteilt für ein gegebenes $\varepsilon \in [0, 0.5]$. Damit sind durchschnittlich ein Anteil von ε der Daten Ausreißer. Dabei sagen wir, dass eine Zufallsvariable Z $(1 - \varepsilon)\mathbb{P}^0 + \varepsilon\mathbb{P}^1$ -verteilt ist für gegebene Wahrscheinlichkeitsmaße $\mathbb{P}^0, \mathbb{P}^1$, falls stochastisch unabhängige Zufallsvariablen $X \sim \mathbb{P}^0, Y \sim \mathbb{P}^1, U \sim \mathcal{U}[0, 1]$ existieren, so dass

$$Z = \begin{cases} X, & \text{falls } U < 1 - \varepsilon, \\ Y, & \text{sonst,} \end{cases} \quad \mathbb{P}\text{-f.s.}$$

Hierbei ist $\mathcal{U}[0, 1]$ die stetige Gleichverteilung auf $[0, 1]$.

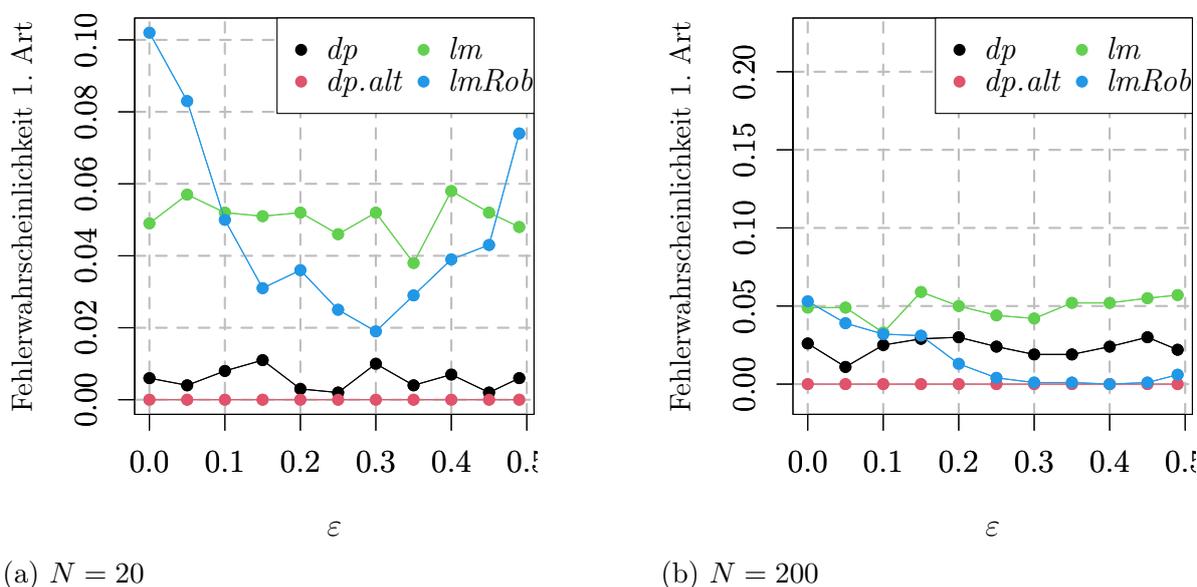


Abbildung 8: Fehlerwahrscheinlichkeit 1. Art in Abhängigkeit von Ausreißer-Anteil ε

Zunächst können wir in Abbildung 8 erkennen, dass die Fehlerwahrscheinlichkeit 1. Art von lm , d.h. von dem F -Test, nicht stark durch die Ausreißer beeinflusst wird und das Niveau 0.05 ungefähr einhält. Die Statistik des F -Tests ist ein Quotient, dessen Nenner die Quadratsumme für die Messfehler und im Zähler die Quadratsumme für den Faktor des quadratischen Terms beinhaltet. Der Einfluss der Ausreißer auf den Nenner und Zähler scheinen sich also auszugleichen. In Abschnitt 6.6.2 zu der Güte sehen wir jedoch, dass Ausreißer durchaus einen Einfluss auf die Fehlerwahrscheinlichkeit 2. Art des F -Tests haben.

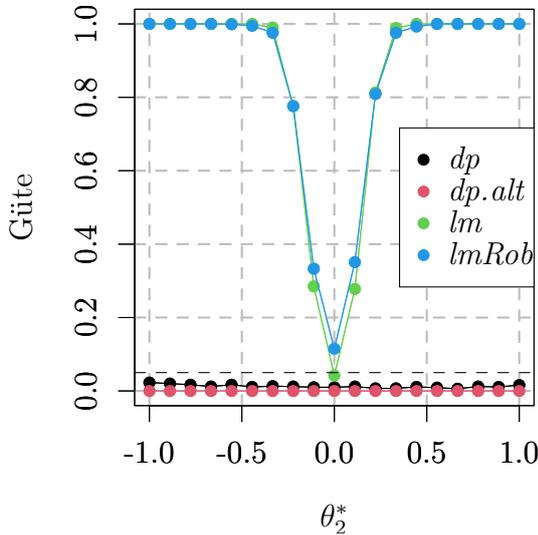
Da die Vorzeichen-Tiefe nur von den Vorzeichen der Residuen abhängt, sehen wir hier somit auch bei den Tests dp und $dp.alt$ keinen Einfluss der Ausreißer auf die Fehlerwahrscheinlichkeit 1. Art. Beide Tests halten hier das Niveau stets ein. Der Test $dp.alt$ lehnt in

dieser Simulation die Nullhypothese nie ab. Wie wir auch in den folgenden Abschnitten sehen werden, zeigt dies, dass $dp.alt$ viel zu konservativ entscheidet.

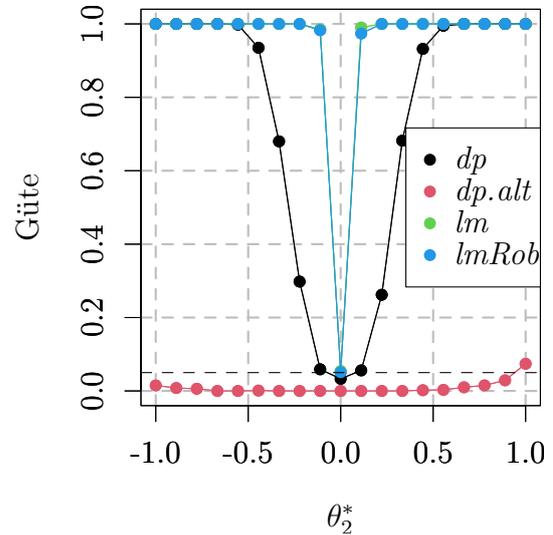
Bei der kleinen Datenmenge $N = 20$ und einem Ausreißer-Anteil $\varepsilon = 0$ hält der Test $lmRob$ das Niveau nicht ein. Mit wachsendem Ausreißer-Anteil sinkt die Fehlerwahrscheinlichkeit 1. Art von $lmRob$ zunächst, wobei sie ab einem Ausreißer-Anteil von ungefähr 0.3 dann wieder steigt. Bei $N = 200$ hält $lmRob$ das Niveau ein, was ein Indiz dafür ist, dass $lmRob$ in der Tat ein asymptotischer Test zum Niveau 0.05 ist. Mit wachsendem Ausreißer-Anteil sinkt dann die Fehlerwahrscheinlichkeit 1. Art von $lmRob$, sodass sie bei einem Ausreißer-Anteil zwischen ungefähr 0.3 und 0.45 sogar fast Null ist. Erst ab einem Ausreißer-Anteil, der sehr nahe von 0.5 ist, steigt sie leicht.

6.6.2 Güte

Nun wollen wir die Gütefunktionen der Tests vergleichen. Dabei variieren wir θ_2^* auf dem Bereich $[-1, 1]$. Die anderen Einträge des Parametervektors θ^* seien stets $(\theta_0^*, \theta_1^*) = (1, 2)$. Wir betrachten zwei verschiedene Verteilungen der Störgrößen, nämlich die Standardnormalverteilung $\mathcal{N}(0, 1)$ und Standard-Cauchy-Verteilung $\text{Cauchy}(0, 1)$. Die Standard-Cauchy-Verteilung ist in manchen Punkten der Standardnormalverteilung ähnlich, d.h. sie ist auch symmetrisch und ihr Modus und Median ist Null, jedoch existiert aufgrund ihrer schweren Ränder ihr Erwartungswert nicht. So ist das 99%-Quantil der Standard-Cauchy-Verteilung ca. 31.82, während das der Standardnormalverteilung lediglich ca. 2.33 ist. Dies führt im Vergleich zur Standardnormalverteilung häufiger zu relativ hohen Realisierungen. Die Verwendung der Standard-Cauchy-Verteilung ist deshalb eine Möglichkeit, um die Robustheit von Schätzern, Tests usw. gegenüber Ausreißern zu beurteilen.

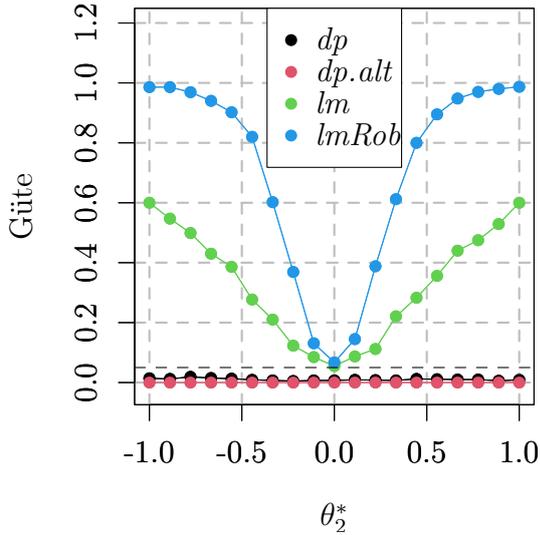


(a) $N = 20, E_n \sim \mathcal{N}(0, 1)$

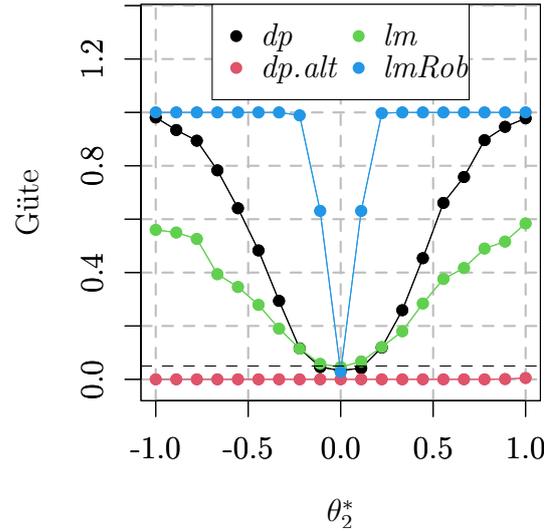


(b) $N = 200, E_n \sim \mathcal{N}(0, 1)$ (Werte von lm und $lmRob$ sind größtenteils identisch)

Abbildung 9: Vergleich der Güte der verschiedenen Tests



(c) $N = 20$, $E_n \sim \text{Cauchy}(0, 1)$



(d) $N = 200$, $E_n \sim \text{Cauchy}(0, 1)$

Abbildung 9: Vergleich der Güte der verschiedenen Tests (fortges.)

Wir erkennen, dass unsere praktisch implementierten 3-Vorzeichen-Tiefe-Tests dp und $dp.alt$ bei kleinen Datenmengen wie hier $N = 20$ viel zu konservativ entscheiden und somit eine sehr schlechte Güte haben. Der Test $dp.alt$ entscheidet in dieser Simulation stets viel zu konservativ, sodass dieser Test in allen vier betrachteten Situationen in Abbildung 9 der Test mit der schlechtesten Güte ist. Bei der relativ großen Datenmenge $N = 200$ entscheidet dp nicht mehr so konservativ. Wegen der hohen Effizienz der Tests lm und $lmRob$ hat dp jedoch bei standardnormalverteilten Störgrößen eine deutlich schlechtere Güte als lm und $lmRob$. Bei den standard-Cauchy-verteilten Störgrößen, welche wir für die Untersuchung der Robustheit in Hinsicht auf die Güte verwenden, hat dp jedoch bei der Datenmenge $N = 200$ eine bessere Güte als lm . Hier ist die Güte von dp aber immer noch schlechter als die von $lmRob$.

Auch die kleine Datenmenge $N = 20$ reicht für lm und $lmRob$ aus, um im Vergleich zu dp in dem hier betrachteten Bereich $\theta_2^* \in [-1, 1]$ eine nennenswerte Güte zu erreichen. Bei standardnormalverteilten Störgrößen in Abbildung 9a und 9b sehen wir eine ungefähr gleich gute Güte von lm und $lmRob$. Bei den Abbildungen 9c und 9d, wo wir in der Situation mit den standard-Cauchy-verteilten Störgrößen sind, können wir die Nichtrobustheit des F -Tests lm erkennen im Gegensatz zu $lmRob$. Wie wir in Abschnitt 6.6.1 schon bemerkt haben, hält $lmRob$ bei $N = 20$ und standardnormalverteilten Störgrößen jedoch nicht das Niveau ein.

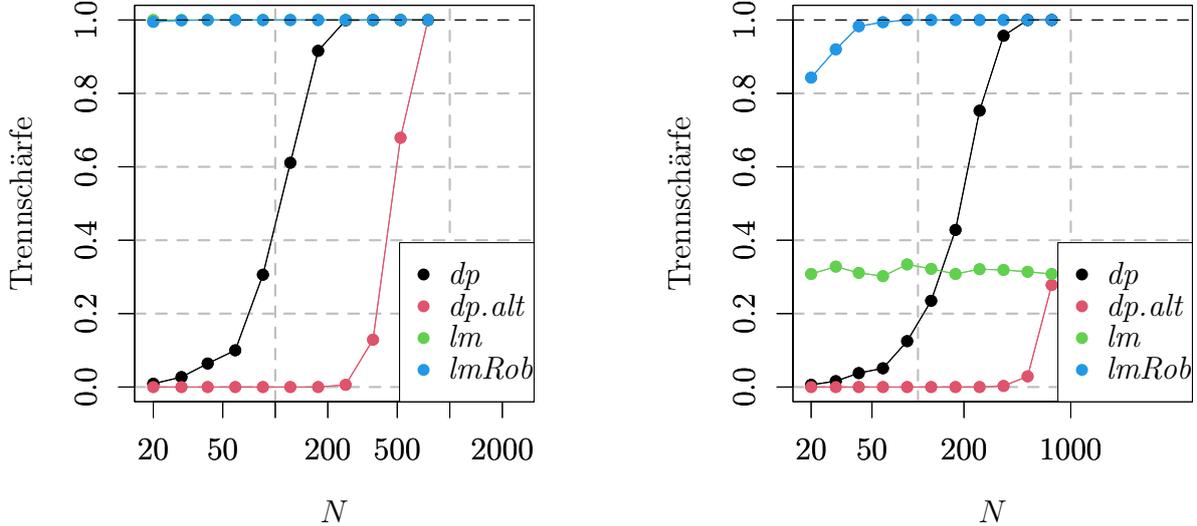
Insgesamt zeigt sich hier die Überlegenheit vom robusten Test $lmRob$ gegenüber den anderen Tests, was sowohl die Effizienz bei standardnormalverteilten Störgrößen als auch die Robustheit gegenüber Ausreißern angeht. Bei hohen Datenmengen wie hier exemplarisch $N = 200$ ist jedoch unser praktisch implementierter 3-Vorzeichen-Tiefe-Test dp trotz seiner Schwächen eine realistische und robuste Alternative.

6.6.3 Konsistenz

Zuletzt vergleichen wir die Konsistenz der Tests. Dabei betrachten wir den wahren Parameter $\theta^* = (1, 2, 0.5) \notin \Theta_0$ und exponentiell wachsende Datenmengen

$$N = 20, 29, 41, 59, 85, 122, 176, 253, 363, 522, 750.$$

Deshalb verwenden wir in Abbildung 10 auch eine logarithmische x -Achse.



(a) $E_n \sim \mathcal{N}(0,1)$ (Werte von lm und $lmRob$ sind größtenteils identisch)

(b) $E_n \sim \text{Cauchy}(0,1)$

Abbildung 10: Vergleich der Konsistenz der verschiedenen Tests

Wir können zunächst in Abbildung 10a mit den standardnormalverteilten Störgrößen die sehr schnelle Konvergenzgeschwindigkeit der Trennschärfen von lm und $lmRob$ gegen 1 erkennen. Schon bei der Datenmenge $N = 20$ sind diese Trennschärfen größer als 0.99. Im Vergleich dazu ist die Konvergenzgeschwindigkeit bei den Tests dp und $dp.alt$ viel langsamer. In dieser Simulation ist dp sehr viel besser als $dp.alt$, da dp schon bei $N = 176$ eine Trennschärfe größer als 0.9 erreicht, während die Trennschärfe von $dp.alt$ erst bei $N = 750$ größer als 0.9 ist. Somit zeigt sich auch hier, dass $dp.alt$ viel zu konservativ entscheidet.

In Abbildung 10b, d.h. der Situation mit standard-Cauchy-verteilten Störgrößen erkennen wir, dass der F -Test lm anscheinend nicht konsistent ist. Hier zeigt sich wieder die Nichtrobustheit des F -Tests. Im Gegensatz dazu ist der robuste Test $lmRob$ weiterhin konsistent mit einer schnellen Konvergenzgeschwindigkeit der Trennschärfe gegen 1. Der robuste, praktisch implementierte 3-Vorzeichen-Tiefe-Test ist auch in dieser Situation konsistent, wobei die Konvergenzgeschwindigkeit der Trennschärfe gegen 1 aber immer noch langsamer ist als die von $lmRob$.

Auch hier bestätigt sich das Fazit von Abschnitt 6.6.2, dass der Test $lmRob$ den anderen Tests überlegen ist. Trotzdem zeigt sich ebenfalls unser praktisch implementierter

6 Anwendung des 3-Vorzeichen-Tiefe-Tests in dem polynomiellen Modell

3-Vorzeichen-Tiefe-Test als realistische Alternative für große Datenmengen, wenn man die Resultate aus den Abschnitten 6.6.2 und 6.6.3 zusammen betrachtet.

Literatur

- Arcones, M. A., & Gine, E. (1993). Limit Theorems for U-Processes. *The Annals of Probability*, 21(3), 1494–1542. Verfügbar 18. September 2023 unter <http://www.jstor.org/stable/2244585>
- Billingsley, P. (1999). *Convergence of Probability Measures* (2. Aufl.). Wiley.
- Dahlquist, G., & Björck, Å. (1974). *Numerical Methods*. Prentice-Hall, Inc.
- Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statistics & Probability Letters*, 14(2), 119–128. [https://doi.org/https://doi.org/10.1016/0167-7152\(92\)90075-G](https://doi.org/https://doi.org/10.1016/0167-7152(92)90075-G)
- Falkenau, C. (2016). *Depth based estimators and tests for autoregressive processes with application on crack growth and oil prices* [Diss., Technische Universität Dortmund].
- Falkenau, C., Leucht, A., & Müller, C. (2014). Tests Based on Simplicial Depth for AR(1) Models With Explosion. *Journal of Time Series Analysis*, 37. <https://doi.org/10.1111/jtsa.12186>
- Fröberg, C.-E. (1969). *Introduction to Numerical Analysis* (2. Aufl.). Addison Wesley Publishing Company.
- Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric Statistical Inference* (4. Aufl.). Marcel Dekker.
- Horn, M., & Müller, C. H. (2020). Tests based on sign depth for multiple regression. SFB Discussion Paper 07/20. <https://doi.org/https://doi.org/10.17877/DE290R-20984>
- Horn, M. (2021b). *Sign depth for parameter tests in multiple regression* [Diss., Technische Universität Dortmund]. <https://doi.org/https://doi.org/10.17877/DE290R-22355>
- Kustosz, C. P., Müller, C. H., & Wendler, M. (2016). Simplified simplicial depth for regression and autoregressive growth processes. *Journal of Statistical Planning and Inference*, 173, 125–146. <https://doi.org/https://doi.org/10.1016/j.jspi.2016.01.005>
- Leckey, K., Jakubzik, M., & Müller, C. (2022). *On the consistency of K-sign depth tests*.
- Leckey, K., Malcherzyk, D., & Horn, M. (2023). Simple powerful robust tests based on sign depth. *Stat Papers*, 64, 857–882. <https://doi.org/https://doi.org/10.1007/s00362-022-01337-5>
- Liu, R. (1988). On a notion of simplicial depth. *Proceedings of the National Academy of Sciences of the United States of America*, 85, 1732–4. <https://doi.org/10.1073/pnas.85.6.1732>
- Liu, R. (1990). On a Notion of Data Depth Based on Random Simplices. *Ann. Stat.*, 18. <https://doi.org/10.1214/aos/1176347507>
- Malcherzyk, D. (2022). *K-sign depth: Asymptotic distribution, efficient computation and applications* [Diss., Technische Universität Dortmund].
- Malcherzyk, D., Leckey, K., & Müller, C. H. (2021). K-sign depth: From asymptotics to efficient implementation. *Journal of Statistical Planning and Inference*, 215, 344–355. <https://doi.org/https://doi.org/10.1016/j.jspi.2021.04.006>

- Müller, C. H. (2005). Depth estimators and tests based on the likelihood principle with application to regression. *Journal of Multivariate Analysis*, 95(1), 153–181. <https://doi.org/https://doi.org/10.1016/j.jmva.2004.06.006>
- Rousseeuw, P. J., & Hubert, M. (1999). Regression Depth. *Journal of the American Statistical Association*, 94(446), 388–402. Verfügbar 3. Juni 2023 unter <http://www.jstor.org/stable/2670155>
- Tukey, J. W. (1975). Mathematics and the Picturing of Data. *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, 2, 523–531.
- Wellmann, R., Harmand, P., & Müller, C. H. (2009). Distribution-free tests for polynomial regression based on simplicial depth. *Journal of Multivariate Analysis*, 100(4), 622–635. <https://doi.org/https://doi.org/10.1016/j.jmva.2008.06.009>
- Wellmann, R., & Müller, C. H. (2010). Tests for multiple regression based on simplicial depth. *Journal of Multivariate Analysis*, 101(4), 824–838. <https://doi.org/https://doi.org/10.1016/j.jmva.2009.12.008>

Software und zugehörige Literatur

- Horn, M. (2021a). *GSignTest: Robust Tests for Regression-Parameters via Sign Depth* [R package version 1.0.8]. <https://github.com/melaniehorn/GSignTest>
- Posit team. (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC. Boston, MA. <http://www.posit.co/>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Sharpsteen, C., & Bracken, C. (2023). *tikzDevice: R Graphics Output in LaTeX Format* [R package version 0.12.4]. <https://CRAN.R-project.org/package=tikzDevice>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4. Aufl.) [ISBN 0-387-95457-0]. Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>
- Wang, J., Zamar, R., Marazzi, A., Yohai, V., Salibian-Barrera, M., Maronna, R., Zivot, E., Rocke, D., Martin, D., Maechler, M., & Konis, K. (2022). *robust: Port of the S+ "Robust Library"* [R package version 0.7-1]. <https://CRAN.R-project.org/package=robust>

Eidesstattliche Versicherung

(Affidavit)

Krumnack, Jan

Name, Vorname
(surname, first name)

209947

Matrikelnummer
(student ID number)

Bachelorarbeit
(Bachelor's thesis)

Masterarbeit
(Master's thesis)

Titel
(Title)

Konsistenz und Anwendungen der Vorzeichen-Tiefe-Methode

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem oben genannten Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present thesis with the above-mentioned title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution before.

Dortmund, 19.09.2023

Ort, Datum
(place, date)

Krumnack

Unterschrift
(signature)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the Chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, Section 63 (5) North Rhine-Westphalia Higher Education Act (*Hochschulgesetz, HG*).

The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund University will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification.*

Dortmund, 19.09.2023

Ort, Datum
(place, date)

Krumnack

Unterschrift
(signature)

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.