

Technische Universität Dortmund
Fakultät Statistik

Bachelorarbeit

Vorzeichen-Varianz-Tests im Vergleich mit anderen Varianz-Tests

Erstgutachter: Prof. Dr. Christine Müller
Zweitgutachter: Dr. Mirko Jakubzik

Verfasser: Marius Timm
Matrikelnummer: 235500
Adresse: Zum Bergfried 12, 51147 Köln
Datum: 7. April 2025

Inhaltsverzeichnis

1	Einleitung	3
2	Problemstellung	4
3	Statistische Methoden	4
3.1	K-Vorzeichen-Tiefe-Test	5
3.2	χ^2 -Test für die Varianz	8
3.3	Test basierend auf dem Q-Schätzer	9
4	Simulationsdesign	15
5	Auswertung der Simulationsergebnisse	16
5.1	Nicht-kontaminierte Daten	16
5.2	Kontaminierte Daten	18
6	Diskussion	24
7	Fazit	31
	Anhang	32
	Literatur	35

1 Einleitung

In der statistischen Analyse ist das Testen der Varianz σ^2 in vielen Anwendungsbereichen von großer Bedeutung. Der Parameter steht für die Streuung bzw. Homogenität der Daten, weshalb sich aus ihm wertvolle Informationen ableiten lassen. Es existieren bereits etablierte klassische Methoden zum Testen auf den Parameter σ^2 , die sich besonders gut für homogene Stichproben eignen. In der Praxis treten jedoch auch Unregelmäßigkeiten in Form von untypisch großen oder kleinen Werten auf, die sogenannten Ausreißer. Durch die restriktiven Annahmen über die Verteilung in klassischen Ansätzen können daher die Ergebnisse mit Ausreißern in der Stichprobe verzerrt sein. Aus diesem Grund sind Testverfahren erforderlich, die auch in Anwesenheit von Ausreißern zuverlässige Ergebnisse liefern. Ein vielversprechender Ansatz hierfür ist der K-Vorzeichen-Tiefe-Test. In Leckey et al. (2023) wird diese Erweiterung des klassischen Vorzeichen-Tests vorgestellt, die ebenso simpel ist, jedoch eine wesentlich höhere Teststärke aufweist. Eine Eigenschaft dieses Testverfahrens ist, dass keine Annahmen über die Verteilung erforderlich sind, wodurch es in verschiedenen Situationen anwendbar ist. In dieser Untersuchung wird daher geprüft, ob sich der K-Vorzeichen-Tiefe-Test als geeignetes Verfahren zum robusten Testen der Varianz erweist. Dabei wird analysiert, wie sich die Teststärke verschiedener Variationen des K-Vorzeichen-Tiefe-Tests in unterschiedlichen Szenarien verhält. Zum Vergleich werden zwei weitere Varianztests herangezogen: Einer davon stellt einen klassischen Ansatz dar, während der andere die Eigenschaften teilt, robust gegenüber Ausreißern zu sein und keine Annahmen über die Verteilung zu machen. Für dieses Vorgehen werden im Rahmen einer Simulation Zufallszahlen erzeugt, die eine Stichprobe bilden und auf die Testverfahren angewendet werden. In die Stichproben werden systematisch Ausreißer eingefügt, die sich je nach Szenario in Größe und Umfang unterscheiden. Die Ergebnisse werden anhand von Güteplots veranschaulicht, die zeigen, ob das Signifikanzniveau eingehalten wird und wie hoch die Teststärke ist. In den Ergebnissen der Simulation wurde deutlich, wie unterschiedlich die Tests auf Ausreißer reagieren und dass bestimmte Variationen des K-Vorzeichen-Tiefe-Tests besonders gut abschneiden.

Diese Arbeit gliedert sich in 6 weitere Abschnitte. Im folgenden Abschnitt wird das Modell und die Testhypothesen kurz erläutert. Anschließend werden die Testverfahren im dritten Abschnitt vorgestellt und definiert. Im vierten Abschnitt werden die Szenarien

und der Aufbau der Simulation erklärt. Die Simulationsergebnisse der Tests werden dann im fünften Abschnitt vorgestellt. Daraufgehend werden im sechsten Abschnitt die vorgestellten Ergebnisse interpretiert und im letzten Abschnitt werden die wichtigsten Ergebnisse zusammengefasst und ein Ausblick für weitere Untersuchungen gegeben.

2 Problemstellung

Seien $X_1, \dots, X_N$ unabhängig und identisch verteilte Zufallsvariablen mit dem Erwartungswert μ und der Varianz σ^2 . Das Ziel dieser Untersuchung ist das robuste Testen der folgenden Hypothesen:

$$H_0 : \sigma = \sigma_0 \text{ vs. } H_1 : \sigma \neq \sigma_0.$$

Ein Test gilt für diese Untersuchung als robust, wenn er gute Ergebnisse für $X_1, \dots, X_N \sim \mathcal{N}(\mu, \sigma^2)$ liefert und wenn die Ergebnisse nicht deutlich schlechter werden, wenn sich Ausreißer unter den Realisationen der Zufallsvariablen $X_1, \dots, X_N$ befinden. Gute Ergebnisse bedeuten in diesem Kontext, dass die Testentscheidungen bei Abweichungen von der Nullhypothese richtig getroffen werden und das Signifikanzniveau α unter der Nullhypothese nicht überschritten wird. Ohne Einschränkung der Allgemeinheit nehmen wir an, dass $\mu = 0$ und $\sigma_0 = 1$. Das Signifikanzniveau der Tests wird auf $\alpha = 0.05$ festgelegt.

Im folgenden wird die Abkürzung „u.i.v.“ für „unabhängig identisch verteilt“ verwendet. Außerdem wird auch der Begriff „Kontamination“ (Verunreinigung) für Stichproben mit Ausreißern verwendet.

3 Statistische Methoden

In diesem Abschnitt werden die für die Untersuchung verwendeten Testverfahren vorgestellt. Für den K-Vorzeichen-Tiefe-Test werden zwei Methoden zur Berechnung der Tiefe sowie die Skalierung der Ausreißer definiert. Abschließend folgt die Definition des Tests. Der darauffolgende χ^2 -Test stellt den klassischen Varianztest dar. Im Unterabschnitt zum Test, der auf dem Q-Schätzer basiert, wird zunächst der Q-Schätzer vorgestellt, auf dessen Grundlage anschließend ein statistischer Test konstruiert wird.

3.1 K-Vorzeichen-Tiefe-Test

Der K-Vorzeichen-Tiefe-Test ist ein nicht-parametrischer und robuster Test, welcher lediglich auf den Vorzeichen der Residuen beruht. Grundlage des Tests ist, wie der Name schon sagt, die so genannte K-Vorzeichen-Tiefe, in dieser Arbeit auch abgekürzt mit K-Tiefe. Die K-Vorzeichen-Tiefe ist der relative Anteil von den K-Tupeln der Residuen R_n , bei denen ein Vorzeichenwechsel stattfindet, wobei $K \in \mathbb{N} \setminus \{1\}$ ein frei wählbarer Parameter ist. Tests basierend auf der K-Tiefe beruhen auf der Voraussetzung, dass die Wahrscheinlichkeit für positive bzw. negative Vorzeichen der Residuen $R_n(\theta)$, mit $n = 1, \dots, N$ und dem Parameter $\theta \in \Theta \subset \mathbb{R}^p, p \in \mathbb{N}$, genau $\frac{1}{2}$ entspricht (Leckey et al., 2023), d.h. wenn

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

gilt. Ferner bedeutet (1), eine ungefähr gleiche Anzahl an positiven und negativen Vorzeichen der von θ abhängigen Residuen, welche zu Vorzeichenwechseln in den K-Tupeln führt. Die K-Tiefe $d_K(r_1, \dots, r_N)$ mit den Realisationen $r_1, \dots, r_N$ von R_n , ist für $K \geq 2$ wie folgt definiert (Leckey et al., 2023):

$$d_K^F(r_1, \dots, r_N) := \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < n_2 < \dots < n_K \leq N} \left(\prod_{k=1}^K \mathbb{1}_{(0, \infty)}\{(-1)^k r_{n_k}\} + \prod_{k=1}^K \mathbb{1}_{(-\infty, 0)}\{(-1)^k r_{n_k}\} \right). \quad (2)$$

Das Zeichen $\mathbb{1}$, steht für die Indikatorfunktion, die bei $\mathbb{1}_A\{x\} = \mathbb{1}\{x \in A\}$ gleich 1 ist, wenn x in dem Intervall A liegt, andernfalls ist sie gleich 0. Man spricht von der vollen K-Vorzeichen-Tiefe, wenn sie die Form wie in (2) hat. Eine vereinfachte Version der K-Vorzeichen-Tiefe, ist nach Kustojs et al. (2016) für $K \geq 2$, folgendermaßen definiert:

$$d_K^S(r_1, \dots, r_N) := \frac{1}{N - K + 1} \sum_{n=1}^{N-K+1} \left(\prod_{k=1}^K \mathbb{1}_{(0, \infty)}\{(-1)^k r_{n+k-1}\} + \prod_{k=1}^K \mathbb{1}_{(-\infty, 0)}\{(-1)^k r_{n+k-1}\} \right). \quad (3)$$

Im Gegensatz zu der vollen K-Tiefe, werden bei der vereinfachten K-Tiefe $N - K + 1$ Tupel betrachtet, anstatt $\binom{N}{K}$ Tupel, wodurch sie sich schneller berechnen lässt. Trotz

größerem Rechenaufwand, ist die Güte der Tests, die auf der vollständigen K-Tiefe basieren, meist besser als die der vereinfachten K-Tiefe (Kustozs et al., 2016). Die R-Funktion `componentKDepth` aus dem Paket `KDepth` von Nagy (2024), bietet einen effizienten Algorithmus zur Bestimmung der vollen und vereinfachten K-Tiefe.

Bevor der K-Vorzeichen-Tiefe-Test für die Hypothese $H_0 : \sigma = \sigma_0$ definiert wird, werden die skalierten Residuen definiert, damit die Voraussetzung (1) für unseres Modell erfüllt ist. Die Idee um dies zu gewährleisten ist eine Skalierung der Residuen durch den Median einer Verteilung. Dazu gelten folgende Überlegungen (Müller, 2024): Sei $X_n \sim \mathcal{N}(\mu, \sigma^2)$ u.i.v., dann ist $\frac{X_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ und wegen der Unabhängigkeit von $X_1, \dots, X_N$, gilt

$$\frac{(X_n - \mu)^2}{\sigma^2} \sim \chi_1^2, \quad (4)$$

$$\frac{(X_n - \mu)^2}{\sigma^2} + \frac{(X_m - \mu)^2}{\sigma^2} \sim \chi_2^2 \text{ mit } n \neq m, \quad (5)$$

wobei χ_k^2 für die zentrale χ^2 -Verteilung mit k Freiheitsgraden steht. Der Median einer χ_1^2 -Verteilung, kann mithilfe von R mit dem Funktionsaufruf `qchisq(0.5, 1)` ausgegeben werden. Der Median von den Zufallsvariablen der Form wie in (4), ist somit gegeben durch

$$\text{med} \left(\frac{(X_n - \mu)^2}{\sigma^2} \right) = 0.4549364.$$

Der exakte Median einer Summer zweier χ_k^2 -verteilten Zufallsvariablen, lässt sich durch die Beziehung der χ^2 -Verteilung zur Gammaverteilung herleiten. Denn $X \sim \chi_k^2$ impliziert, dass $X \sim \mathcal{G}(\alpha, \theta)$ mit $\alpha = \frac{k}{2}$ und $\theta = 2$, wobei $\mathcal{G}(\alpha, \theta)$ für die Gammaverteilung mit dem Formparameter α und dem Skalenparameter θ steht. Für $k = 2$ und $\alpha = 1$ ist eine Gammaverteilung dieser Form äquivalent zu einer Exponentialverteilung mit dem Skalenparameter θ , d.h. $X \sim \mathcal{E}(\theta)$. Der Median von $X \sim \mathcal{E}(2)$ ist gegeben, durch

$$1 - e^{-\frac{x}{2}} = \frac{1}{2} \Leftrightarrow \frac{1}{2} = e^{-\frac{x}{2}} \Leftrightarrow 2 = e^{\frac{x}{2}} \Leftrightarrow \ln(2) = \frac{x}{2} \Leftrightarrow x = 2 \ln(2).$$

Für die Summe von zwei χ_k^2 -verteilten Zufallsvariablen bedeutet das

$$\text{med} \left(\frac{(X_n - \mu)^2}{\sigma^2} + \frac{(X_m - \mu)^2}{\sigma^2} \right) = 2 \ln(2) = 1.386294 \text{ für } n \neq m.$$

Mit den Medianen erhält man die skalierten Residuen

$$Y_n^1 := \frac{(X_n - \mu)^2}{\sigma_0^2} - 0.4549364 \text{ für } n = 1, \dots, N, \text{ und} \quad (6)$$

$$Y_n^2 := \frac{(X_n - \mu)^2}{\sigma_0^2} + \frac{(X_{N-n+1} - \mu)^2}{\sigma_0^2} - 2 \ln(2) \text{ für } n = 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor. \quad (7)$$

Die hochgestellten Zahlen bei Y_n^1 bzw. Y_n^2 dienen hier einzig zur Nummerierung. Y_n^2 kann auch mit aufeinanderfolgenden Paaren definiert werden, was hier jedoch irrelevant ist, da die Zufallsvariablen unabhängig und identisch verteilt sind. Wir haben unter H_0 für alle $n = 1, \dots, N$ bzw. $n = 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor$,

$$\begin{aligned} P_{\sigma_0}(Y_n^1 > 0) &= \frac{1}{2} = P_{\sigma_0}(Y_n^1 < 0), \\ P_{\sigma_0}(Y_n^2 > 0) &= \frac{1}{2} = P_{\sigma_0}(Y_n^2 < 0), \end{aligned} \quad (8)$$

wodurch die Voraussetzung (1) für $R_n = Y_n^1$ bzw. $R_n = Y_n^2$ erfüllt ist.

Korollar 3.1 (siehe Malcherzyk et al. (2021), Leckey et al. (2023)). *Der K -Vorzeichen-Tiefe-Test, gegeben durch*

Lehne $H_0 : \sigma = \sigma_0$ ab, falls

$$d_K^J(r_1, \dots, r_N) < q_K^J(\alpha),$$

wobei $q_K^J(\alpha)$ für das α -Quantil von $d_K^J(R_1, \dots, R_N)$ steht, mit $R_1, \dots, R_N$, welche (1) erfüllen und $J = F, S$, ist ein Test zum Niveau α .

Die 5%-Quantile der vollen und vereinfachten 2- und 3-Vorzeichen-Tiefe, können aus Tabelle 1 entnommen werden.

Tabelle 1: Simulierte 5%-Quantile der K-Tiefen für $K = 1, 2$ bei 10^6 Simulationsdurchläufen.

Tiefe	5%-Quantil	
	N = 30	N = 100
d_2^S	0.3448276	0.414141
d_3^S	0.1071429	0.163265
d_2^F	0.4597701	0.484848
d_3^F	0.2068966	0.237612

3.2 χ^2 -Test für die Varianz

Der χ^2 -Test ist ein klassischer Test zur Untersuchung der Nullhypothese $H_0 : \sigma^2 = \sigma_0^2$ für normalverteilte Stichproben. Den Namen hat der Test wegen seiner Teststatistik, welche auf der χ^2 -Verteilung basiert. Sei $X_1, \dots, X_N$ u.i.v. mit $X_i \sim \mathcal{N}(\mu, \sigma)$, $i = 1, \dots, N$, dann ist die Teststatistik definiert durch

$$T = \frac{(N-1)S^2}{\sigma_0^2} = \frac{1}{\sigma_0^2} \sum_{i=1}^N (X_i - \bar{X})^2 \quad (9)$$

mit der empirischen Varianz S^2 und dem arithmetischen Mittel $\bar{X}$ von $X_1, \dots, X_N$ (Bamberg et al., 2021). Des weiteren lässt sich mit der Definition der χ^2 -Verteilung (siehe Mosler (2006)) zeigen, dass T einer χ^2 -Verteilung folgt mit $N-1$ Freiheitsgraden, also $T \sim \chi_{N-1}^2$.

Korollar 3.2 (siehe Bamberg et al. (2021)) *Sei $X_1, \dots, X_N$ u.i.v. mit $X_i \sim \mathcal{N}(\mu, \sigma)$, $i = 1, \dots, N$, dann ist der χ^2 -Test, gegeben durch*

$$\begin{aligned} &\text{Lehne } H_0 : \sigma^2 = \sigma_0^2 \text{ ab, falls} \\ &T \in \left(0, \chi_{\frac{\alpha}{2}; N-1}^2\right) \cup \left(\chi_{1-\frac{\alpha}{2}; N-1}^2, \infty\right), \end{aligned}$$

mit T so wie in (9) und $\chi_{\frac{\alpha}{2}; N-1}^2$ bzw. $\chi_{1-\frac{\alpha}{2}; N-1}^2$ als dem $\frac{\alpha}{2}$ - bzw. $(1 - \frac{\alpha}{2})$ -Quantil der χ^2 -Verteilung mit $N-1$ Freiheitsgraden, ein Test zum Niveau α .

Zusätzlich ist noch anzumerken, dass der χ^2 -Test bei nicht normalverteilten Zufallsvariablen selbst bei einem hohen Stichprobenumfang generell nicht anwendbar ist (Bamberg et al., 2021).

3.3 Test basierend auf dem Q-Schätzer

Der Q-Schätzer ist ein Schätzverfahren um die Varianz oder andere Parameter einer Verteilung zu schätzen. Grundlage dieses Schätzers sind Quantile der empirischen Verteilung der Daten. Die empirische Verteilung setzt sich aus den absoluten Abweichungen aller Datenpunkte zusammen, die nicht den gleichen Wert besitzen. In Rousseeuw (1993) wird gezeigt, dass der Schätzer eine robuste Methode ist mit einem Bruchpunkt¹ (engl. breakdown point) von 50%. Ebenso wird auch erwähnt, dass der Nachteil dieses Schätzverfahrens die hohe Rechenlaufzeit ist. Im Folgenden wird der Q-Schätzer definiert und der dazugehörige Test hergeleitet. Die Konstruktion des Tests entstammt aus Müller (2024). Der Test basierend auf dem Q-Schätzer wird im weiteren Verlauf dieser Arbeit auch als Q-Schätzer-Test bezeichnet.

Der Q-Schätzer ist folgendermaßen definiert

$$\sigma_Q := \sigma_Q(x) = \frac{1}{4} - \text{Quantil von } \{|x_n - x_m|; x_n \neq x_m\} = \inf \left\{ y; H_N(y) \geq \frac{1}{4} \right\} = H_N^{-1} \left(\frac{1}{4} \right) \quad (10)$$

mit

$$H_N(y) := H_{N,x}(y) := \frac{1}{N(N-1)} \sum_{n=1}^N \sum_{m=1, m \neq n}^N \mathbf{1}_{(-\infty, y]}(|x_n - x_m|). \quad (11)$$

Nach Hoeffding (1948), ist $H_N(y)$ eine U-Statistik für die Realisierungen $x = (x_1, \dots, x_N)$, wobei $y \in \mathbb{R}$ ein fester Wert ist. Deshalb ist nach dem Satz von Hoeffding (siehe Serfling (1980), S. 192) die Verteilung einer U-Statistik und somit die von $H_N(y)$ bekannt.

Lemma 3.1 Wenn $X_1, \dots, X_N$ unabhängig identisch verteilte Zufallsvariablen sind, $s \in \mathbb{N}$ und für

$$H_N(y) := \frac{1}{\binom{N}{s}} \sum_{1 \leq n_1 < n_2 < \dots < n_s \leq N} U_y(X_{n_1}, \dots, X_{n_s})$$

¹Verhältnis von Ausreißern im Datensatz, bei dem Schätzer verzerrt wird (Müller, 2001).

mit

$$U_y(x_1, \dots, x_s) = \mathbb{1}_{(-\infty, y]} h(x_1, \dots, x_s),$$

eine Funktion $h : \mathbb{R}^s \rightarrow \mathbb{R}$ existiert, die

$$h(x_{\pi_1}, \dots, x_{\pi_s}) = h(x_1, \dots, x_s)$$

für jede Permutation $\pi : \{1, \dots, s\} \rightarrow \{1, \dots, s\}$ erfüllt, dann ist U_y eine symmetrische Kernfunktion und somit konvergiert

$$\sqrt{N}(H_N(y) - H_F(y)) \longrightarrow \mathcal{N}(0, s^2 v_y^2)$$

in Verteilung mit

$$\begin{aligned} H_F(y) &:= E(U_y(X_1, \dots, X_s)) = P(h(X_1, \dots, X_s) \leq y), \\ v_y^2 &:= \text{var}(\psi(X_1)) \end{aligned}$$

wobei $\psi(x_1) = E(U_y(X_1, \dots, X_s) | X_1 = x_1) = P(h(X_1, \dots, X_s) \leq y | X_1 = x_1)$.

Korollar 3.3 Unter den Voraussetzungen von Lemma 3.1 und wenn $H'_F(H_F^{-1}) > 0$ gilt für $p \in (0, 1)$, dann gilt

$$\sqrt{N}(H_N^{-1}(p) - H_F^{-1}(p)) \longrightarrow \mathcal{N}\left(0, \frac{s^2 v^2}{(H'_F(H_F^{-1}(p)))^2}\right)$$

wobei $v^2 := \text{var}(\psi(X_1))$ mit

$$\psi(x_1) := P(h(X_1, \dots, X_s) \leq H_F^{-1}(p) | X_1 = x_1).$$

Beweis. Der Beweis folgt analog zum Beweis von Satz 1 in Ghosh (1971). Siehe Müller (2001) für weitere komplizierte Q-Schätzer. $\square$

Satz 3.1 *Der Q-Schätzer hat die folgende asymptotische Verteilung*

$$\sqrt{N} \left(\frac{1}{\sqrt{2}\Phi^{-1}\left(\frac{5}{8}\right)} \hat{\sigma}_Q(X) - \sigma \right) \longrightarrow \mathcal{N}(0, V^2)$$

wobei

$$V^2 := \left(\frac{1}{\Phi^{-1}\left(\frac{5}{8}\right)} \right)^2 \frac{v^2 \sigma^2}{\left(\phi\left(\Phi^{-1}\left(\frac{5}{8}\right)\right) \right)^2}$$

mit

$$v^2 := \text{var} \left(\Phi \left(\frac{X_1}{\sigma} + \sqrt{2}\Phi^{-1}\left(\frac{5}{8}\right) \right) - \Phi \left(\frac{X_1}{\sigma} - \sqrt{2}\Phi^{-1}\left(\frac{5}{8}\right) \right) \right)$$

und Φ als die kumulative Verteilungsfunktion der Standardnormalverteilung und ϕ als ihre Ableitung.

Beweis. Zuerst beachte man, dass der Q-Schätzer oder genauer genommen die Funktion $H_N(y)$ eine U-Statistik ist mit $s = 2$, $h(x_1, x_2) = |x_1 - x_2|$, und

$$U_y(x_1, x_2) = \mathbb{1}_{(-\infty, y]}(|x_1 - x_2|).$$

Für $X_1 - X_2 \sim \mathcal{N}(0, 2\sigma^2)$, gilt

$$\begin{aligned} H_F(y) &= E(U_y(X_1, X_2)) = P(|X_1 - X_2| \leq y) = P(-y \leq |X_1 - X_2| \leq y) \\ &= P(-y \leq X_1 - X_2 \leq y) = P\left(-\frac{y}{\sqrt{2}\sigma} \leq \frac{X_1 - X_2}{\sqrt{2}\sigma} \leq \frac{y}{\sqrt{2}\sigma}\right) = 2\Phi\left(\frac{y}{\sqrt{2}\sigma}\right) - 1 \end{aligned}$$

und

$$H'_F(y) = 2\phi\left(\frac{y}{\sqrt{2}\sigma}\right) \frac{1}{\sqrt{2}\sigma} = \sqrt{2}\phi\left(\frac{y}{\sqrt{2}\sigma}\right) \frac{1}{\sigma}.$$

Für $y = H_F^{-1}\left(\frac{1}{4}\right)$ bedeutet das

$$\begin{aligned} H_F(y) = \frac{1}{4} &\iff 2\Phi\left(\frac{y}{\sqrt{2}\sigma}\right) - 1 = \frac{1}{4} \iff 2\Phi\left(\frac{y}{\sqrt{2}\sigma}\right) = \frac{5}{4} \\ &\iff \Phi\left(\frac{y}{\sqrt{2}\sigma}\right) = \frac{5}{8} \iff \frac{y}{\sqrt{2}\sigma} = \Phi^{-1}\left(\frac{5}{8}\right) \iff H_F^{-1}\left(\frac{1}{4}\right) = y = \sqrt{2}\sigma\Phi^{-1}\left(\frac{5}{8}\right) \end{aligned}$$

und

$$H'_F \left(H_F^{-1} \left(\frac{1}{4} \right) \right) = \sqrt{2} \phi \left(\frac{\sqrt{2} \sigma \Phi^{-1} \left(\frac{5}{8} \right)}{\sqrt{2} \sigma} \right) \frac{1}{\sigma} = \sqrt{2} \phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right) \frac{1}{\sigma}.$$

Ferner gilt

$$\begin{aligned} \psi(x_1) &:= P \left(h(X_1, X_2) \leq H_F^{-1} \left(\frac{1}{4} \right) \mid X_1 = x_1 \right) = P(|X_1 - X_2| \leq y \mid X_1 = x_1) \\ &= P(|x_1 - X_2| \leq y) = P(-y \leq X_2 - x_1 \leq y) = P(x_1 - y \leq X_2 \leq x_1 + y) \\ &= P \left(\frac{x_1 - y}{\sigma} \leq \frac{X_2}{\sigma} \leq \frac{x_1 + y}{\sigma} \right) = \Phi \left(\frac{x_1 + y}{\sigma} \right) - \Phi \left(\frac{x_1 - y}{\sigma} \right) \\ &= \Phi \left(\frac{x_1 + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right)}{\sigma} \right) - \Phi \left(\frac{x_1 - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right)}{\sigma} \right) \\ &= \Phi \left(\frac{x_1}{\sigma} + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) - \Phi \left(\frac{x_1}{\sigma} - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) \end{aligned}$$

so dass

$$v^2 := \text{var} \left(\Phi \left(\frac{X_1}{\sigma} + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) - \Phi \left(\frac{X_1}{\sigma} - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) \right).$$

Mit Korollar 3.3 erhält man

$$\sqrt{N} \left(H_N^{-1} \left(\frac{1}{4} \right) - H_F^{-1} \left(\frac{1}{4} \right) \right) = \sqrt{N} \left(\hat{\sigma}_Q(X) - \sqrt{2} \sigma \Phi^{-1} \left(\frac{5}{8} \right) \right) \longrightarrow \mathcal{N}(0, V_0^2)$$

mit

$$V_0^2 = \frac{s^2 v^2}{\left(H'_F \left(H_F^{-1} \left(\frac{1}{4} \right) \right) \right)^2} = \frac{4v^2}{\left(\sqrt{2} \phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right) \frac{1}{\sigma} \right)^2} = \frac{2v^2 \sigma^2}{\left(\phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right) \right)^2}$$

woraus folgt, dass

$$\sqrt{N} \left(\frac{1}{\sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right)} \hat{\sigma}_Q(X) - \sigma \right) \longrightarrow \mathcal{N}(0, V^2)$$

mit

$$V^2 = \left(\frac{1}{\sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right)} \right)^2 V_0^2 = \left(\frac{1}{\Phi^{-1} \left(\frac{5}{8} \right)} \right)^2 \frac{v^2 \sigma^2}{\left(\phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right) \right)^2}. \quad \square$$

Korollar 3.4 *Der Q-Schätzer erfüllt*

$$\sqrt{N} \left(\frac{\phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right) \hat{\sigma}_Q(X)}{\sqrt{2}v} - \frac{\Phi^{-1} \frac{5}{8} \phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right)}{v} \right) \rightarrow \mathcal{N}(0, 1)$$

mit

$$v^2 = \text{var} \left(\Phi \left(\frac{X_1}{\sigma} + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) - \Phi \left(\frac{X_1}{\sigma} - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) \right).$$

Der Wert von v^2 kann ermittelt werden, indem man M standardnormalverteilte Zufallsvariablen $z_1, \dots, z_M$ simuliert und die empirische Varianz von $w_1, \dots, w_M$ gegeben durch

$$w_m = \left(\Phi \left(z_m + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) - \Phi \left(z_m - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) \right)$$

für $m = 1, \dots, M$ bestimmt. Für $M = 100.000.000$ ergab sich $v^2 = 0.008891178$. Ein genauerer Wert für v^2 kann durch numerische Integration des folgenden Ausdrucks bestimmt werden:

$$v^2 = \int \left(\Phi \left(z_m + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) - \Phi \left(z - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) \right)^2 \phi(z) dz - \left(\frac{1}{4} \right)^2$$

unter der Berücksichtigung, dass

$$\begin{aligned} & E \left(\Phi \left(\frac{X_1}{\sigma} + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) - \Phi \left(\frac{X_1}{\sigma} - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) \right) \\ &= E(\psi(X_1)) = E \left(P \left(|X_1 - X_2| \leq H_F^{-1} \left(\frac{1}{4} \right) | X_1 \right) \right) \\ &= P \left(|X_1 - X_2| \leq H_F^{-1} \left(\frac{1}{4} \right) \right) = H_F \left(H_F^{-1} \left(\frac{1}{4} \right) \right) = \frac{1}{4}. \end{aligned}$$

gilt. Der Q-Schätzer kann mit der R-Funktion `Qn` aus dem Paket `robustbase` von Maechler et al. (2024) bestimmt werden. In Abbildung 1 sind die Ergebnisse dargestellt, um zu überprüfen, ob

$$\sqrt{N} \left(\frac{\phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right) \hat{\sigma}_Q(X)}{\sqrt{2}v} - \frac{\Phi^{-1} \left(\frac{5}{8} \right) \phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right)}{v} \right) \quad (12)$$

annäherungsweise standardnormalverteilt ist. Die Ergebnisse zeigen, dass die Verteilungsannahme für die Teststatistik des Q-Schätzer-Tests, gerechtfertigt ist.

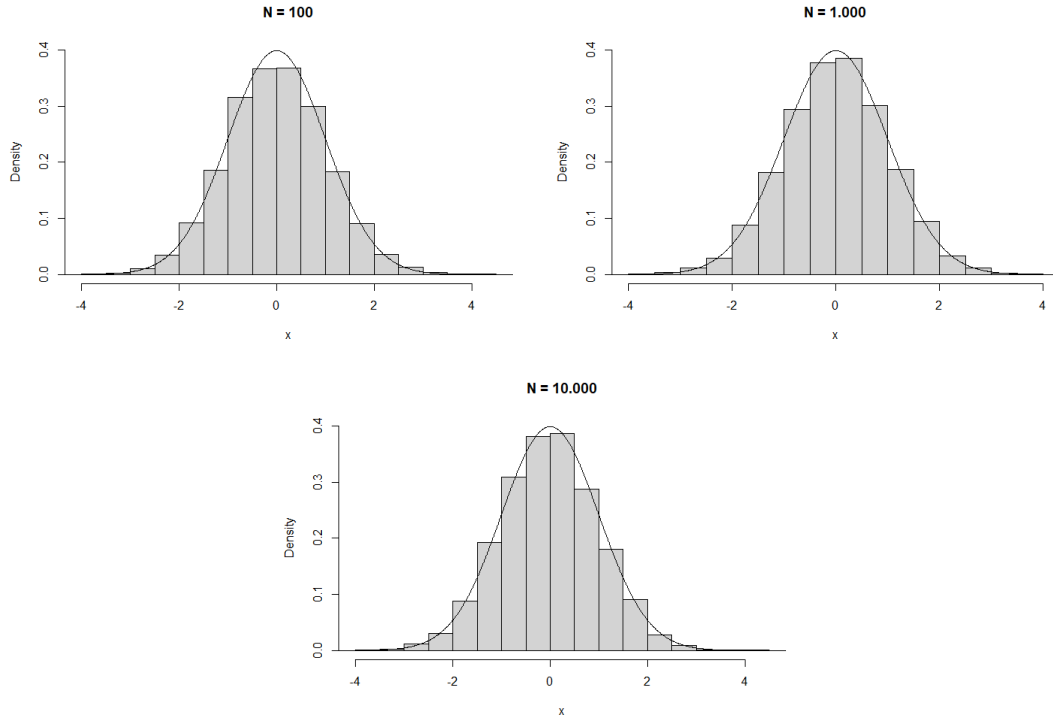


Abbildung 1: Simulation der standardisierten Teststatistik basierend auf dem Q -Schätzer (12) für $N = 100, 1.000, 10.000$ mit jeweils 10.000 Simulationsdurchläufen.

Korollar 3.5 *Der Test, der auf dem Q -Schätzer basiert, gegeben durch*
Lehne $H_0 : \sigma = \sigma_0$ ab, falls

$$\left| \sqrt{N} \left(\frac{\phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right) \hat{\sigma}_Q(X)}{\sqrt{2}v \sigma_0} - \frac{\Phi^{-1} \left(\frac{5}{8} \right) \phi \left(\Phi^{-1} \left(\frac{5}{8} \right) \right)}{v} \right) \right| > q_{N(0,1)} \left(1 - \frac{\alpha}{2} \right),$$

wobei $q_{N(0,1)}$ das α -Quantil der Standardnormalverteilung ist und

$$v^2 = \text{var} \left(\Phi \left(Z + \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) - \Phi \left(Z - \sqrt{2} \Phi^{-1} \left(\frac{5}{8} \right) \right) \right) \text{ mit } Z \sim \mathcal{N}(0, 1),$$

ist ein asymptotischer Test zum Niveau α für $H_0 : \sigma = \sigma_0$.

4 Simulationsdesign

Zum Vergleich der vorangegangenen Tests, soll durch eine Simulation kenntlich gemacht werden, wie sich die Güte der Tests in verschiedenen Szenarien verhält. Dabei dient die relative Ablehnungsrate der Nullhypothese $H_0 : \sigma = \sigma_0$ mit $\sigma_0 = 1$ aus 10.000 Simulationsdurchläufen als die empirische Güte. In jedem Simulationsdurchlauf werden N normalverteilte Zufallszahlen erzeugt, aus denen die jeweilige Teststatistik berechnet und eine Testentscheidung getroffen wird. Für den Q-Schätzer wird in der Voreinstellung von **Qn** mit dem Faktor $\frac{1}{\sqrt{2\Phi^{-1}(\frac{5}{8})}}$ multipliziert, um Konsistenz bei der Normalverteilung zu erreichen, welcher aber hier ausgeglichen wird. Für die K-Tiefe-Tests werden sowohl die volle als auch die vereinfachte K-Vorzeichen-Tiefe mit jeweils $K = 2$ und $K = 3$ verwendet. Zusammen mit dem χ^2 -Test und dem Q-Schätzer-Test ergibt dies insgesamt sechs Testentscheidungen pro Simulation. Da $\mu = 0$ fest ist, bleibt die Varianz bzw. Standardabweichung als der einzige variable Parameter für die Erzeugung der Zufallszahlen. Für die Auswertung wurden die folgenden Werte für die Standardabweichung festgelegt:

$$\sigma_s \in \{0.1, 0.25, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, \\ 1, 1.05, 1.1, 1.15, 1.2, 1.25, 1.3, 1.35, 1.4, 1.45, 1.5, 1.75, 2, 2.5, 3, 4\}.$$

Da auch die Robustheit gegenüber Ausreißern untersucht werden soll, wurden Szenarien erstellt, in denen mit bestimmter Wahrscheinlichkeit Ausreißer in die Stichprobe eingefügt werden. Hierfür wird unter anderem der Begriff „Kontamination“ der Daten verwendet. Auf die Verteilung der Zufallszahlen X_n , $n = 1, \dots, N$ hat das je nach Szenario den nachfolgenden Effekt

- 1) ohne Kontamination: $X_n \sim \mathcal{N}(0, \sigma_s)$
- 2) mit Kontamination: $X_n \sim (1 - \epsilon)\mathcal{N}(0, \sigma_s) + \frac{\epsilon}{2}\tau + \frac{\epsilon}{2}\tau^{-1}$
 - a) $\epsilon = 0.1, \tau = 10$
 - b) $\epsilon = 0.1, \tau = 10$
 - c) $\epsilon = 0.2, \tau = 100$
 - d) $\epsilon = 0.2, \tau = 100$

In den Szenarien mit Ausreißern, gibt ϵ die Wahrscheinlichkeit für einen Ausreißer an und τ den Wert des Ausreißers. Der Term $\frac{\epsilon}{2}\tau^{-1}$ wurde hinzugefügt, um die Bedingung über die Vorzeichenverteilung der Residuen (siehe Gleichung 1) nicht zu verletzen. Für die Anzahl der zu erzeugenden Zufallszahlen, wurden die Werte $N = 30, 60, 100, 200$ gewählt. Im Fall $N = 30$ bzw. $N = 100$ werden die Residuen wie in Y_n^1 (6) skaliert, während sie im Fall $N = 60$ bzw. $N = 200$ wie in Y_n^2 (7) skaliert werden.

Durch die Wahl von N und den fünf Szenarien ergeben sich insgesamt 20 Kombinationen. Für jede dieser Kombinationen wird mit jedem Wert von σ_s jeweils 10.000-mal simuliert. Die Simulation sowie die rechnerische und grafische Auswertung erfolgen vollständig in R (R Core Team, 2024).

5 Auswertung der Simulationsergebnisse

In diesem Kapitel werden die Ergebnisse der Simulationen anhand von Güteplots vorgestellt. Für jede der 20 Kombinationen von N und den fünf Szenarien wurde jeweils ein Güteplot erstellt, in dem die relative Ablehnungsrate auf der y-Achse und die Werte für σ auf der x-Achse für jeden Test abgetragen sind. Je Szenario werden für $N \in \{30, 60, 100, 200\}$ jeweils vier Plots in einer 2×2 -Anordnung gegenübergestellt.

5.1 Nicht-kontaminierte Daten

Zunächst werden in Abbildung 2 die Güteplots für Szenario 1 gezeigt, um zu veranschaulichen, wie die Tests ohne Ausreißer performen. Es zeigt sich, dass die relative Ablehnungsrate für $\sigma > 1$ bei $N = 100$ und $N = 200$ schneller ansteigt als bei $N = 30$ und $N = 60$. Für $\sigma < 1$ zeigt sich ein ähnliches Verhalten, mit dem Unterschied, dass die Ablehnungsrate für diese Werte von σ sinkt. Dies deutet darauf hin, dass die Teststärke mit zunehmenden N steigt. Auch zeigt sich, dass die K-Tiefe-Tests mit $R_n = Y_n^2$ eine höhere Teststärke aufweisen als die mit $R_n = Y_n^1$. Die vereinfachten K-Tiefe-Tests erweisen sich in diesem Szenario als die Tests mit der geringsten Teststärke, wobei der Test mit $K = 2$ bessere Ergebnisse aufweist als der mit $K = 3$. Gleiches gilt für die vollen K-Tiefe-Tests, die jedoch eine deutlich höhere Teststärke aufweisen als die vereinfachten

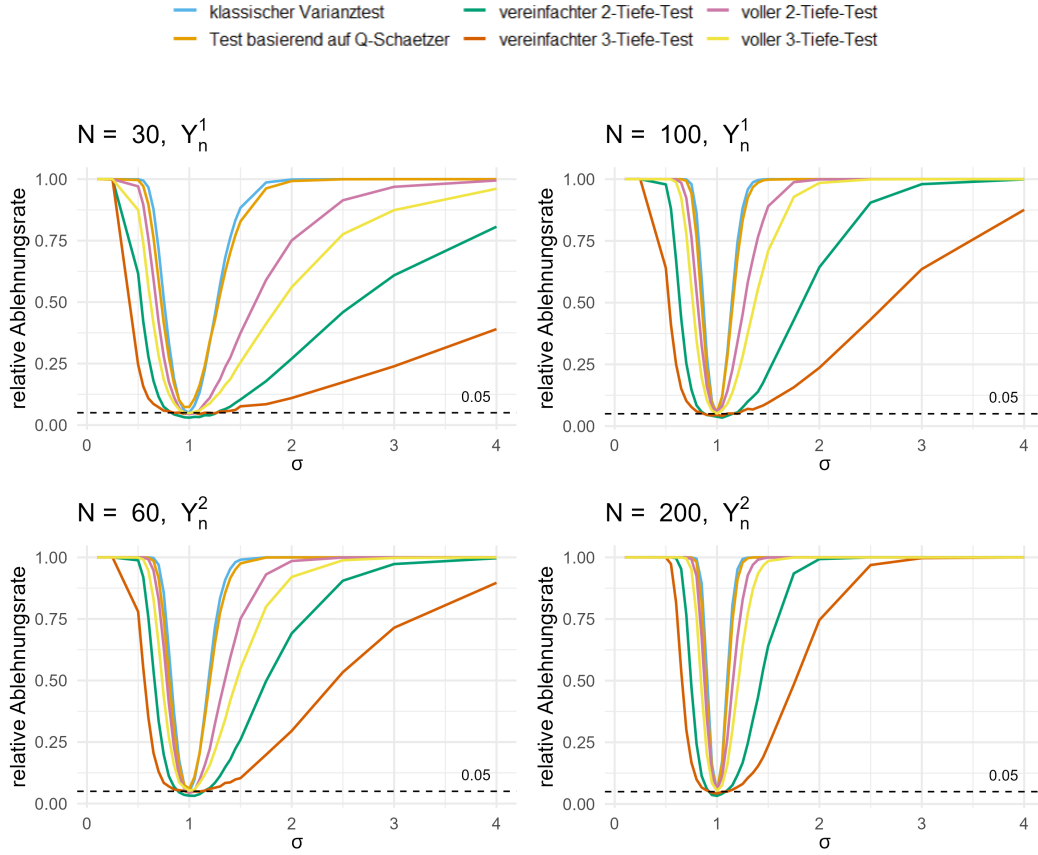


Abbildung 2: Güteplots der Tests für Szenario 1 mit Stichprobengröße $N \in \{30, 60, 100, 200\}$ und dem skalierten Residuen Y_n^1 bzw. Y_n^2 mit $n = 1, \dots, N$ für die K-Tiefe-Tests.

Tests. Um zu überprüfen, ob das Signifikanzniveau der Tests eingehalten wird, zeigt Tabelle 2 im Anhang die relative Ablehnungsrate unter der Nullhypothese. Im Szenario ohne Ausreißer hält der Test basierend auf dem Q-Schätzer das Signifikanzniveau für alle N nicht ein – ganz im Gegensatz zu den vereinfachten K-Tiefe-Tests, die es für alle N einhalten. Allerdings ist die Teststärke der vereinfachten K-Tiefe-Tests deutlich geringer. Bei den vollen K-Tiefe-Tests wird das Signifikanzniveau insbesondere bei höheren N leicht überschritten. Der klassische Varianztest ist in diesem Szenario der beste Test, da er sowohl eine relativ hohe Teststärke aufweist als auch das Signifikanzniveau weitgehend einhält.

5.2 Kontaminierte Daten

Nun folgen die Szenarien mit Ausreißern, beginnend mit Szenario 2a). Hier haben die Zufallszahlen folgende Verteilung:

$$X_n \sim 0.9 \cdot \mathcal{N}(0, \sigma_s) + 0.05 \cdot \tau + 0.05 \cdot \tau^{-1} \text{ mit } \tau = 10$$

für alle $n = 1, \dots, N$. Die Simulationsergebnisse hierzu sind in Abbildung 3 zu sehen. An dieser Stelle sei vorweggenommen, dass der klassische Varianztest in allen Szenarien mit Ausreißern eine deutlich zu hohe Ablehnungsrate an der Stelle $\sigma = 1$ aufweist. Dies unterstreicht nochmals, dass dieser Test für Stichproben mit Ausreißern ungeeignet ist. Daher wird im weiteren Verlauf dieses Abschnitts nicht mehr auf ihn eingegangen. Für den Q-Schätzer-Test zeigt sich, dass dieser zu streng bewertet, wenn sich Ausreißer in der Stichprobe befinden. Zudem verschiebt sich das Minimum der Gütefunktion entlang der x-Achse nach links. Abgesehen von dieser Verschiebung nimmt die Teststärke des Q-Schätzer-Tests mit steigendem N zu. Die Ergebnisse der K-Tiefe-Tests in diesem Szenario ähneln denen aus Szenario 1). Die Unterschiede bestehen darin, dass die Teststärke bei dieser Art von Ausreißern geringer ist und sich das Minimum der Gütefunktion – ähnlich wie beim Q-Schätzer-Test, jedoch nur bei den Residuen $R_n = Y_n^2$ - nach links verschoben hat. Zudem ist die relative Ablehnungsrate des vereinfachten 3-Tiefe-Tests für $\sigma < 1$ bei $N = 60$ deutlich geringer als bei den anderen Tests sowie den übrigen Werten für N . Das Signifikanzniveau wird - bis auf den vollen 2-Tiefe-Test mit $N = 100$ - bei allen K-Tiefe-Tests mit dem Residuen $R_n = Y_n^1$ eingehalten. Das Minimum der Gütefunktion von dem vollen K-Tiefe-Test mit den Residuen $R_n = Y_n^2$ liegt nur knapp über 0.05, weshalb diese Tests das Signifikanzniveau ohne die Linksverschiebung nur leicht überschreiten würden. Der vollen 2-Tiefe-Test mit $N = 60$ würde das Niveau ohne Linksverschiebung sogar einhalten.

Als nächstes betrachten wir das Szenario 2b). In diesem Szenario ist die Wahrscheinlichkeit für Ausreißer gleich wie in 2a), wobei der Wert der Ausreißer von $\tau = 10$ auf $\tau = 100$ erhöht wird. Abbildung 4 zeigt die Ergebnisse dieser Anpassung. Die Ergebnisse zeigen, dass die Erhöhung des Ausreißerwertes die Güte der Tests kaum beeinflusst. Der einzige nennenswerte Unterschied besteht darin, dass das Signifikanzniveau des vollen 3-Tiefe-Tests für $N = 30$ anders wie im Szenario mit kleineren Ausreißern knapp nicht eingehalten wird.

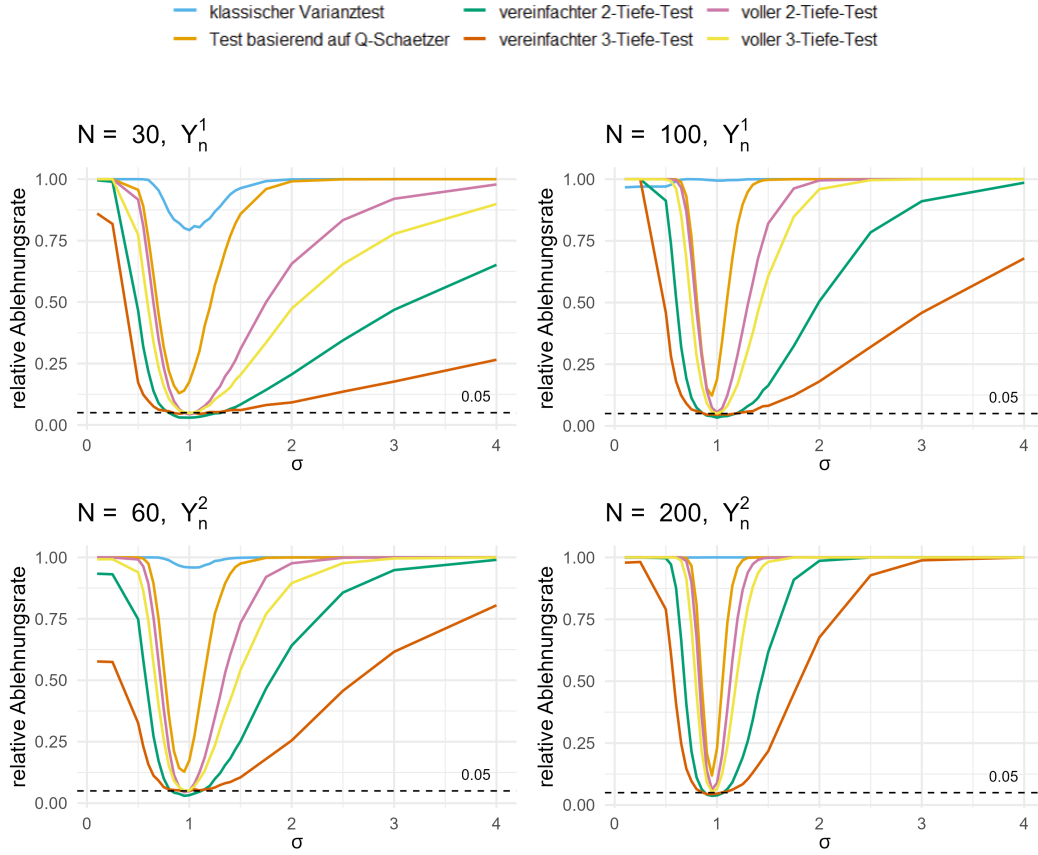


Abbildung 3: Güteplots der Tests für Szenario 2a) mit Stichprobengröße $N \in \{30, 60, 100, 200\}$ und dem skalierten Residuen Y_n^1 bzw. Y_n^2 mit $n = 1, \dots, N$ für die K-Tiefe-Tests.

Im folgenden Szenario wird die Wahrscheinlichkeit für Ausreißer erhöht, so dass

$$X_n \sim 0.8 \cdot \mathcal{N}(0, \sigma_s) + 0.1 \cdot \tau + 0.1 \cdot \tau^{-1} \text{ mit } \tau = 10$$

für alle $n = 1, \dots, N$ gilt. In Abbildung 5 ist zu erkennen, dass der Q-Schätzer bei einer höheren Anzahl an Ausreißern noch strenger ablehnt. Die Linksverschiebung der Gütefunktion bleibt auch in diesem Szenario bestehen. Die K-Tiefe-Tests zeigen bei einer höheren Anzahl an Ausreißern eine geringere Teststärke. Auffällig ist, dass die Güte für den vereinfachten 3-Tiefe-Test für $\sigma < 1$ bei $N = 60$ hier äußerst niedrig ausfällt. Auch bei den anderen K-Tiefe-Tests zeigt sich für $\sigma < 1$ bei $N = 60$ eine deutlich geringere Teststärke. Abgesehen davon ähneln die Ergebnisse der K-Tiefe-Tests

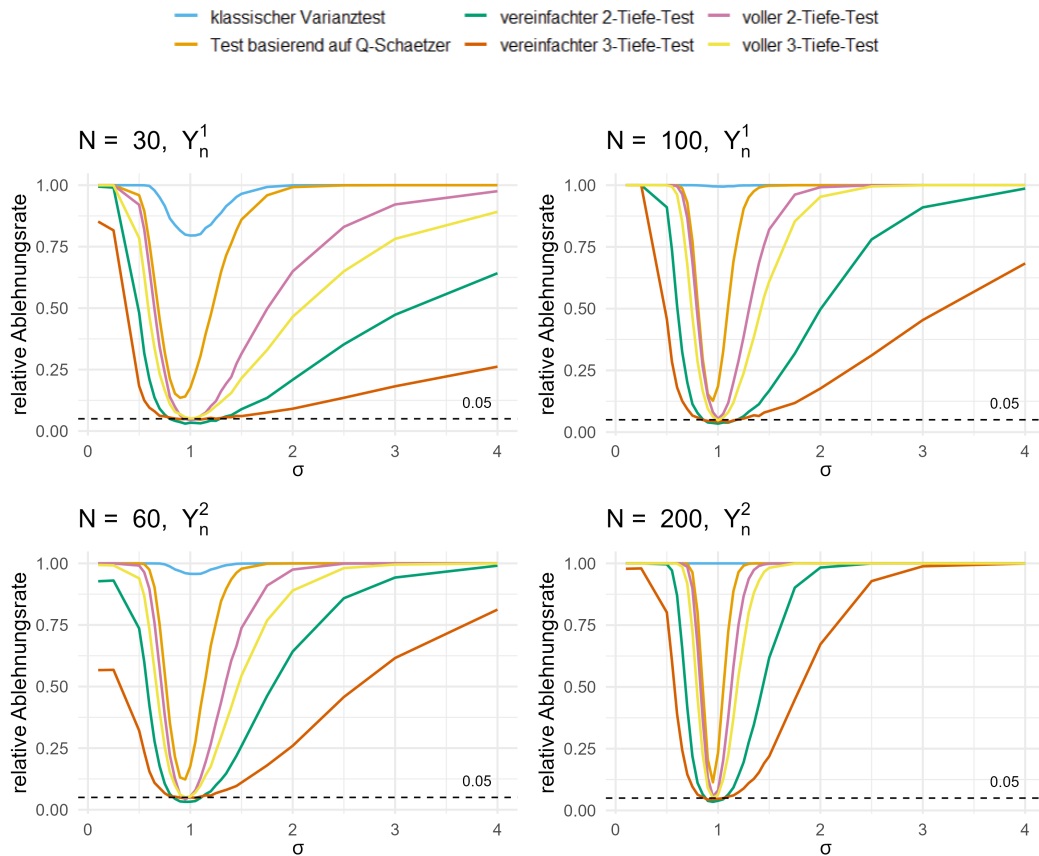


Abbildung 4: Güteplots der Tests für Szenario 2b) mit Stichprobengröße $N \in \{30, 60, 100, 200\}$ und dem skalierten Residuen Y_n^1 bzw. Y_n^2 mit $n = 1, \dots, N$ für die K-Tiefe-Tests.

für $N = 30$ und $N = 100$ denen aus Szenario 2b). Zudem halten die vollen K-Tiefe-Tests das Signifikanzniveau bei $\sigma = 1$ ein. Der volle 2-Tiefe-Test hat auch in diesem Szenario die höchste Teststärke, während der vereinfachte 3-Tiefe-Test die geringste Teststärke aufweist. Vergleicht man die Gütefunktion der K-Tiefe-Tests für $N = 60$ und $N = 200$ mit dem vorherigen Szenario, dann stellt man fest, dass die Gütefunktion hier etwas weiter nach Links verschoben ist.

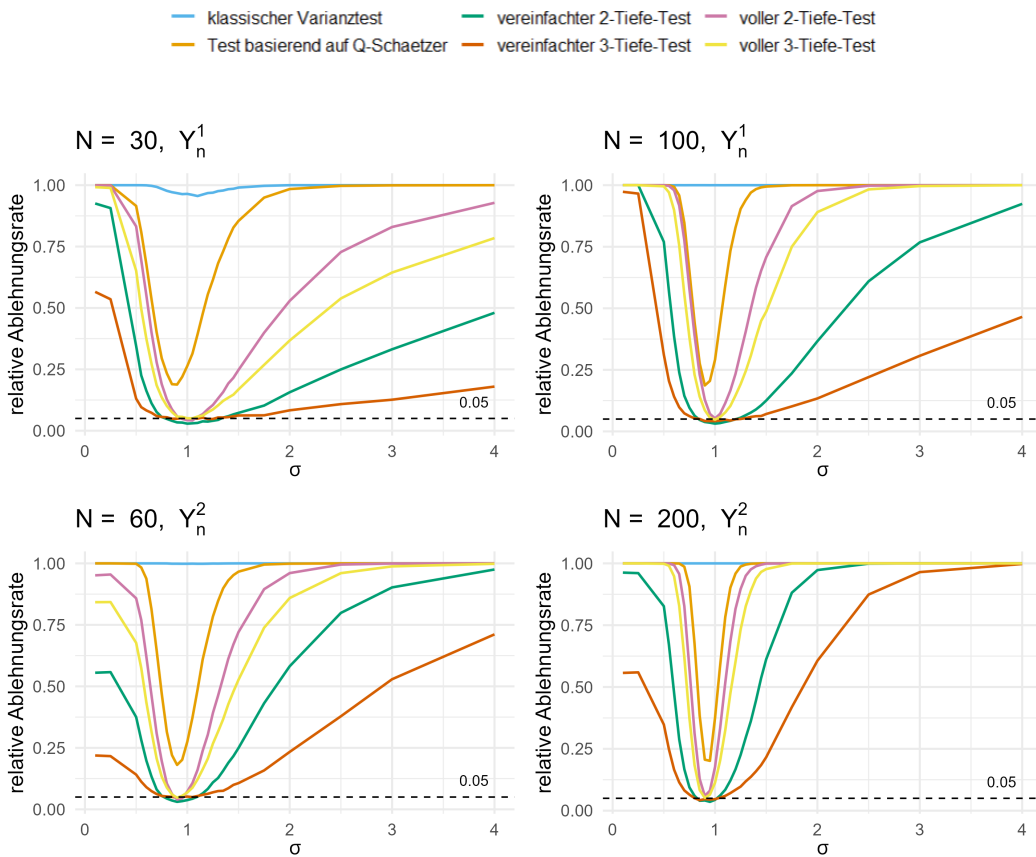


Abbildung 5: Güteplots der Tests für Szenario 2c) mit Stichprobengröße $N \in \{30, 60, 100, 200\}$ und dem skalierten Residuen Y_n^1 bzw. Y_n^2 mit $n = 1, \dots, N$ für die K-Tiefe-Tests.

Im letzten Szenario 2d) wird bei gleichbleibender Ausreißerwahrscheinlichkeit der Wert eines Ausreißers von $\tau = 10$ auf $\tau = 100$ erhöht. Wie bereits im Vergleich zwischen den Szenarien 2a) und 2b), lässt sich in Abbildung 6 kaum ein Unterschied zwischen Szenario 2d) und 2c) erkennen.

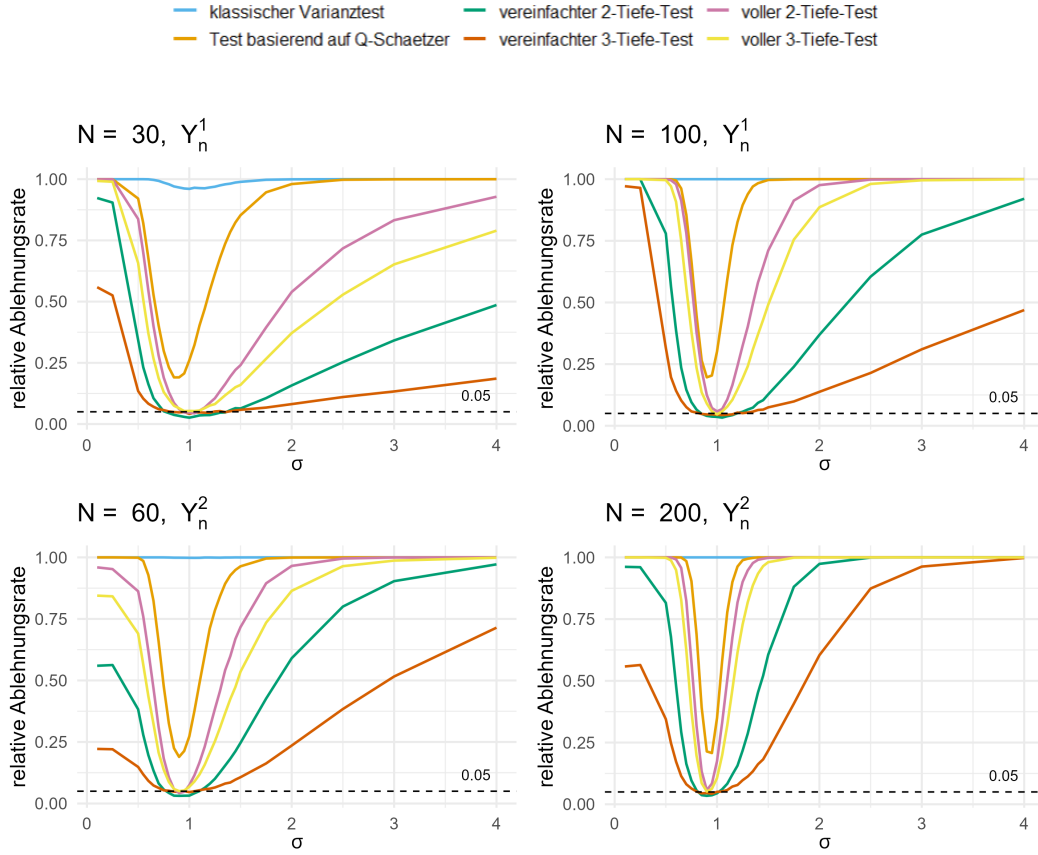


Abbildung 6: Güteplots der Tests für Szenario 2d) mit Stichprobengröße $N \in \{30, 60, 100, 200\}$ und dem skalierten Residuen Y_n^1 bzw. Y_n^2 mit $n = 1, \dots, N$ für die K-Tiefe-Tests.

Die Ergebnisse zeigen, dass die Gütefunktion der K-Tiefe-Tests unter dem Einfluss von Ausreißern nach links verschoben ist, wenn die Residuen wie bei Y_n^2 definiert werden. Um die Ursache für dieser Linksverschiebung zu untersuchen, wird an dieser Stelle geprüft ob die Annahme (1) der K-Tiefe verletzt wird. Dazu werden $N = 200$ normalverteilte Zufallszahlen mit der Verteilung wie in 2d) simuliert, d.h. $\epsilon = 0.2$ und $\tau = 100$. Aus den N Zufallszahlen werden die Residuen nach Y_n^2 bestimmt mit $n = 1, \dots, \lfloor \frac{N}{2} \rfloor$. Auf die skalierten Residuen wird anschließend die Signumfunktion angewendet, um aus der Differenz die Verteilung der Vorzeichen zu bestimmen. Die Vorzeichendifferenz ergibt sich also aus:

$$\sum_{n=1}^{\lfloor \frac{N}{2} \rfloor} \text{sig}(Y_n^2)$$

mit der Signumfunktion

$$\text{sgn}(x) = \begin{cases} 1, & \text{falls } x > 0 \\ 0, & \text{falls } x = 0 \\ -1, & \text{falls } x < 0 \end{cases}$$

für alle $x \in \mathbb{R}$. Zum Vergleich wird das gleiche Vorgehen für Y_n^1 mit $N = 100$ angewandt, da die Gütefunktion der K-Tiefe-Tests keine Linksverschiebung bei dieser Art von skalierten Residuen aufweist. Das Ergebnis in Abbildung 7 zeigt, dass sich deutlich mehr positive Vorzeichen unter den Werten für Y_n^2 befinden. Damit die Annahme (1) nicht verletzt wird, müssen sich die Häufigkeiten der Vorzeichendifferenzen um die null verteilen, so wie bei der oberen Grafik. In Abbildung 8 ist die gleiche Simulation für das Szenario 2b) zu sehen. Vergleicht man die Ergebnisse von Abbildung 7 und Abbildung 8, zeigt sich dass der Anteil der positiven Vorzeichen wächst, je mehr Ausreißer in der Stichprobe sind.

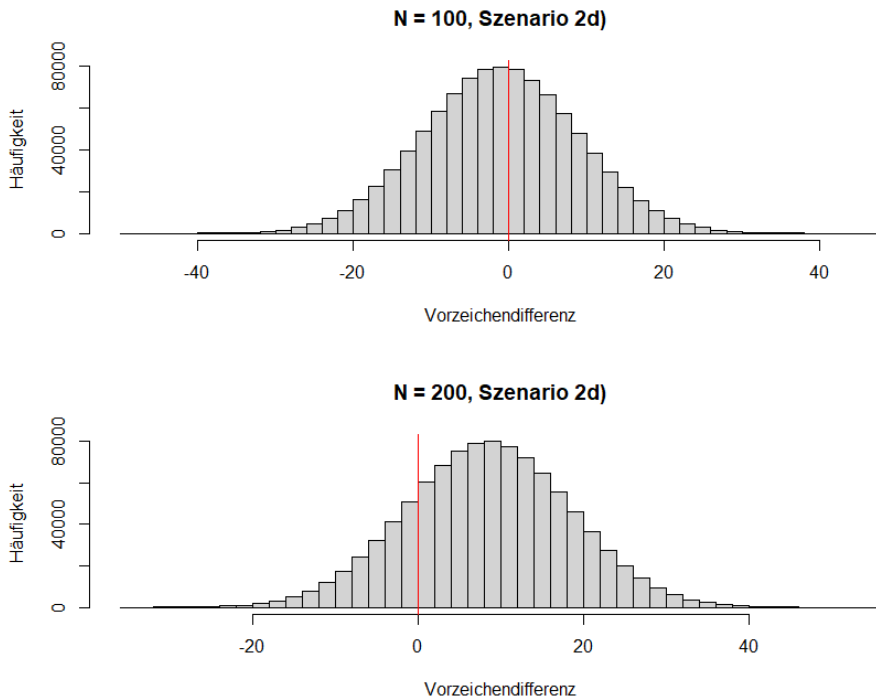


Abbildung 7: Simulation der Vorzeichenverteilung für Y_n^1 und Y_n^2 mit Zufallszahlen nach Szenario 2d) und 10^6 Simulationsdurchläufen.

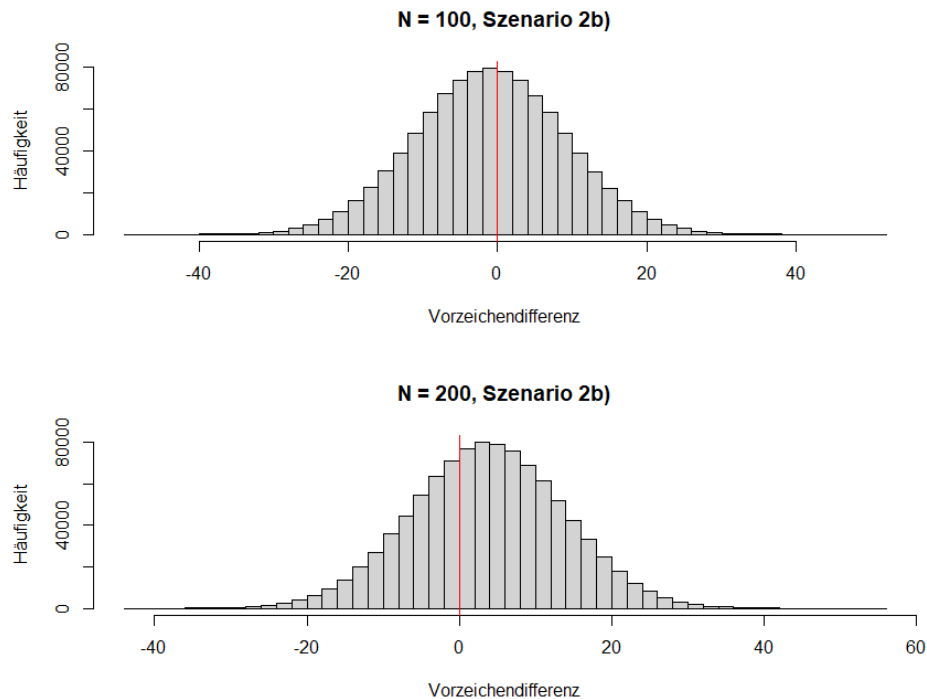


Abbildung 8: Simulation der Vorzeichenverteilung für Y_n^1 und Y_n^2 mit Zufallszahlen nach Szenario 2b) und 10^6 Simulationsdurchläufen.

6 Diskussion

In der vorliegenden Arbeit wurden verschiedene Tests zur Schätzung der Varianz mit und ohne Ausreißern untersucht. Die Ergebnisse zeigen, dass sich die Tests - vor allem mit Ausreißern - hinsichtlich ihrer Güte voneinander unterscheiden. Zusammenfassend zeigen die Ergebnisse, dass die K-Vorzeichen-Tiefe-Tests im Vergleich zu den anderen Tests robust gegenüber Ausreißern sind und dass der volle 2-Tiefe-Test in allen Szenarien die zuverlässigsten Ergebnisse liefert. Für die Interpretation sollte berücksichtigt werden, dass die Ergebnisse auf Stichproben mit Ausreißern und einem Ausgleichsterm basieren, was in der Praxis nicht zwangsläufig der Fall ist.

Der klassische Varianztest hat die größte Teststärke unter den Tests, wenn sich keine Ausreißer in der Stichprobe befinden. Die größte Schwäche dieses Tests zeigt sich in seiner Anfälligkeit gegenüber Ausreißern. Die Resultate belegen, dass der Test bei Vorhandensein

von Ausreißern keine zuverlässigen Ergebnisse liefert. Dennoch dienen die Ergebnisse dieses Tests im Szenario ohne Ausreißer als optimaler Vergleichswert für die anderen Tests.

Ähnliche Ergebnisse zum klassischen Varianztest sind auch für den Test mit dem Q-Schätzer zu beobachten. In dem Szenario ohne Ausreißer ist die Teststärke des Q-Schätzer-Tests fast identisch zu der des klassischen Varianztests. Ein klarer Unterschied bei den Ergebnissen zeigt sich bei den Szenarien mit Ausreißern. Denn der Q-Schätzer-Test reagiert bei weitem nicht so empfindlich auf Ausreißer wie der klassische Varianztest. Dennoch ist eine Linksverschiebung der Gütefunktion zu beobachten, und zudem hält der Test das Signifikanzniveau in Anwesenheit von Ausreißern nicht ein. Die verschiedenen Konfigurationen in den Szenarien veranschaulichen, wie der Q-Schätzer-Test auf Ausreißer reagiert. Es zeigt sich, dass die Testentscheidungen bei größeren Ausreißerwerten strenger ausfallen, was der Grund dafür ist, dass der Test das Signifikanzniveau nicht einhält. Des weiteren zeigt sich, dass die Linksverschiebung der Gütefunktion umso größer ist, desto mehr Ausreißer sich in der Stichprobe befinden. Die Ursache für die Linksverschiebung der Gütekurve kommt dadurch, dass der Q-Schätzer auf empirischen Quantilen basiert. Durch das Einfügen von Ausreißern wird die empirische Verteilung $H_N(y)$ nach rechts verzerrt, wodurch das p-Quantil $H_N^{-1}(p)$ zu groß geschätzt wird. Da die Schätzung für den Q-Schätzer $\sigma_Q = H_N^{-1}\left(\frac{1}{4}\right)$ Bestandteil der Teststatistik ist, wirken sich somit Ausreißer unmittelbar auf die Testentscheidung aus. Ein weiteres Ergebnis von dem Q-Schätzer-Test ist, dass die Kurve der Gütefunktion umso stärker staucht, je größer die Stichprobe ist. Das bedeutet, dass der Test mit einer größeren Stichprobe Abweichungen von der Nullhypothese besser erkennen kann.

Die K-Vorzeichen-Tiefe-Tests zeigen im Vergleich zu den anderen Tests in jedem Szenario moderate Ergebnisse. Dabei hängen die Testergebnisse maßgeblich von der Wahl von K sowie der Verwendung der vollständigen oder vereinfachten Tiefe ab. Jedoch weisen die K-Tiefe-Test in jedem Szenario das gleiche Muster auf hinsichtlich der Teststärke. Darunter zeigt der vollen 2-Tiefe-Test die besten Ergebnisse und der vereinfachte 3-Tiefe-Test die schlechtesten. Dies weist darauf hin, dass die 2-Tiefe für dieses Testproblem die bessere Wahl von K ist. Grund dafür ist wahrscheinlich, dass beim K-Tiefe-Test durch Y_n^1 bzw. Y_n^2 die ursprüngliche Fragestellung in ein Lokationsproblem überführt wird. Für Lokationsprobleme eignet sich der klassische Vorzeichen-Test, welcher äquivalent zum Vorzeichen-Tiefe-Test mit $K = 2$ ist (Leckey et al., 2023). Aufgrund dieser Ergebnisse wurde der gleiche Simulationsvorgang für die Vorzeichen-Tiefe mit $K = 4$ durchgeführt.

Im Anhang sind in den Güteplots von Abbildung 10, Abbildung 11 und Abbildung 12 die Ergebnisse des vollen 2-Vorzeichen-Tiefe-Tests mit denen des vollen 4-Vorzeichen-Tiefe-Tests aufgeführt. Es wurden nur die Szenarien 1, 2a) und 2d) ausgewählt, da hier die Konfiguration bezüglich der Ausreißer am unterschiedlichsten sind. An den Ergebnissen zeigt sich, dass der volle 4-Vorzeichen-Tiefe-Test auch gute Ergebnisse erzielt, jedoch immer noch etwas schlechter als der volle 2-Vorzeichen-Tiefe-Test. Die geringen Unterschiede zwischen der 3- und 4-Vorzeichen-Tiefe zu der 2-Vorzeichen-Tiefe könnten auch darauf zurückzuführen sein, dass die Anordnung der Stichprobe bei der 3- und 4-Vorzeichen-Tiefe eine Rolle spielt, was bei $K = 2$ nicht der Fall ist. Demnach kann es vorkommen, dass Testentscheidungen allein durch die Anordnung der Stichprobe unterschiedlich ausfallen.

Wie bereits bei der Vorstellung der K-Tiefe-Tests erwähnt, haben die vollen K-Tiefe-Tests auch hier eine höhere Teststärke. Da bei der vollen K-Tiefe $\binom{N}{K}$ Tupel und bei der vereinfachten $N - K + 1$ Tupel in die Bestimmung der Tiefe eingehen, werden dadurch deutlich mehr Informationen aus der Stichprobe bei der vollen K-Tiefe entnommen was zu besseren Testentscheidungen führt. Trotz der unterschiedlichen Teststärken halten alle K-Tiefe-Tests das Signifikanzniveau weitgehend ein, was ihre Robustheit unterstreicht. Die Schwächen der K-Tiefe-Tests liegen einerseits in der hohen Rechenlaufzeit zur Bestimmung der Tiefe (Leckey et al., 2023) und andererseits in den Auswirkungen der Verletzung der Voraussetzung (1) auf die Ergebnisse. Ein Blick auf die Ergebnisse von Abbildung 8 und Abbildung 7 mit den jeweiligen Güteplots zeigt, dass die Skalierung der Residuen nach Y_n^2 bei Kontamination zur Verletzung der Voraussetzung führt, was in einer Linksverschiebung der Gütefunktion resultiert. Die Linksverschiebung war bei einer ersten Simulation ohne den Term τ^{-1} noch größer, weshalb dieser in die Verteilungsvorschrift bei Kontamination mit aufgenommen wurde. Diese Änderung hat die Verschiebung jedoch nicht beseitigt sondern nur verringert. Bezüglich dazu folgende theoretische Überlegungen: Sei $\epsilon = 0.2$ und τ so, dass

$$\tau^2 > 2 \ln(2) \text{ bzw.} \tag{13}$$

$$\frac{1}{\tau^2} < 2 \ln(2) \text{ und } \frac{2}{\tau^2} < 2 \ln(2) \tag{14}$$

gilt. Aufgrund der stochastischen Unabhängigkeit von X_n , ergeben sich folgende Wahrscheinlichkeiten für die Kontamination bzw. Nicht-Kontamination von X_n :

1. $P(\text{„}X_n \text{ und } X_{N-n+1} \text{ sind nicht kontaminiert“}) = 0.8^2$
2. $P(\text{„}X_n \text{ oder } X_{N-n+1} \text{ ist kontaminiert mit } \tau\text{“}) = 0.8 \cdot 0.1 \cdot 2$
3. $P(\text{„}X_n \text{ oder } X_{N-n+1} \text{ ist kontaminiert mit } \tau^{-1}\text{“}) = 0.8 \cdot 0.1 \cdot 2$
4. $P(\text{„}X_n \text{ und } X_{N-n+1} \text{ sind kontaminiert mit } \tau\text{“}) = 0.1^2$
5. $P(\text{„}X_n \text{ und } X_{N-n+1} \text{ sind kontaminiert mit } \tau^{-1}\text{“}) = 0.1^2$
6. $P(\text{„eine von } X_n \text{ und } X_{N-n+1} \text{ ist kontaminiert mit } \tau \text{ und die andere mit } \tau^{-1}\text{“}) = 0.1^2 \cdot 2.$

Für die Wahrscheinlichkeit eines positiven Vorzeichens von Y_n^2 folgt

1. Fall: $0.8^2 \cdot P(X_n^2 + X_{N-n+1}^2 - 2 \ln(2) > 0) = 0.8^2 \cdot 0.5 = 0.32$
2. Fall: $0.8 \cdot 0.1 \cdot 2 \cdot P(X_n^2 + X_{N-n+1}^2 - 2 \ln(2) > 0) \stackrel{(13)}{=} 0.8 \cdot 0.1 \cdot 2 \cdot 1 = 0.16$
3. Fall: $0.8 \cdot 0.1 \cdot 2 \cdot P(X_n^2 + X_{N-n+1}^2 - 2 \ln(2) > 0) \stackrel{(14)}{=} 0.8 \cdot 0.1 \cdot 2 \cdot \delta = 0.16\delta$
4. Fall: $0.1^2 \cdot P(X_n^2 + X_{N-n+1}^2 - 2 \ln(2) > 0) = 0.1^2 \cdot P(2 \cdot \tau^2 - 2 \ln(2) > 0) \stackrel{(13)}{=} 0.1^2 \cdot 1 = 0.01$
5. Fall: $0.1^2 \cdot P(X_n^2 + X_{N-n+1}^2 - 2 \ln(2) > 0) = 0.1^2 \cdot P(2 \cdot \tau^{-2} - 2 \ln(2) > 0) \stackrel{(14)}{=} 0.1^2 \cdot 0 = 0$
6. Fall: $0.1^2 \cdot 2 \cdot P(X_n^2 + X_{N-n+1}^2 - 2 \ln(2) > 0) \stackrel{(13)}{=} 0.1^2 \cdot 2 \cdot 1 = 0.02$

wobei $\delta > 0$. Bildet man die Summe dieser Wahrscheinlichkeiten erhält man

$$0.32 + 0.16 + 0.16\delta + 0.01 + 0 + 0.02 > 0.48 + 0.03 = 0.51 > 0.5,$$

was bedeutet, dass die Wahrscheinlichkeit für ein positives Vorzeichen von Y_n^2 höher ist. Die Linksverschiebung der Gütefunktion ergibt sich wie folgt:

Gilt $P(X_n^2 + X_{N-n+1}^2 - 2\ln(2) > 0) > 0.5$ für $X_n \sim 0.8 \cdot X_{n,1} + 0.1 \cdot \tau + 0.1 \cdot \tau^{-1}$

mit $X_{n,1} \sim \mathcal{N}(0, 1)$ so gilt für $X_{n,\sigma} \sim \mathcal{N}(0, \sigma)$ mit $\sigma < 1$:

$$\begin{aligned} \frac{1}{\sigma} X_{n,\sigma} \sim \mathcal{N}(0, 1) &\Rightarrow P(X_{n,\sigma}^2 + X_{N-n+1,\sigma}^2 - 2\ln(2) > 0) \\ &= P\left(\frac{1}{\sigma} X_{n,\sigma}^2 + \frac{1}{\sigma} X_{N-n+1,\sigma}^2 - \frac{1}{\sigma} 2\ln(2) > 0\right) \\ &= P\left(X_{n,1}^2 + X_{N-n+1,1}^2 - \frac{1}{\sigma} 2\ln(2) > 0\right) \\ &< P(X_{n,1}^2 + X_{N-n+1,1}^2 - 2\ln(2) > 0) \end{aligned}$$

Damit wird die Wahrscheinlichkeit $P(X_n^2 + X_{N-n+1}^2 - 2\ln(2) > 0)$ mit

$X_n \sim 0.8 \cdot X_{n,\sigma} + 0.1 \cdot \tau + 0.1 \cdot \tau^{-1}$ für $\sigma < 1$ kleiner und es existiert ein $\sigma < 1$ mit $P(X_n^2 + X_{N-n+1}^2 - 2\ln(2) > 0) = 0.5$.

Des weiteren sind auch Unterschiede bei der Teststärke zwischen Y_n^1 und Y_n^2 in den Ergebnissen ohne Kontamination festzustellen. Für die Erklärung dieser Unterschiede folgende Annahmen: Sei $X_n \sim \mathcal{N}(0, \sigma)$ u.i.v., $\sigma > 0$, $Z_n \sim \mathcal{N}(0, 1)$ u.i.v. und c der Median der χ^2 -Verteilung mit einem Freiheitsgrad. Es gilt

$$\begin{aligned} \frac{X_n - \mu}{\sigma} = \frac{X_n}{\sigma} = Z_n \sim \mathcal{N}(0, 1) &\Rightarrow X_n = \sigma Z_n \text{ bzw.} \\ X_n + X_m = \sigma Z_n + \sigma Z_m &\text{ mit } n \neq m. \end{aligned}$$

Ferner gilt

$$\begin{aligned} X_n^2 &= \sigma^2 Z_n^2 \text{ und } Z_n^2 \sim \chi_1^2 \text{ sowie} \\ X_n^2 + X_m^2 &= \sigma^2 Z_n^2 + \sigma^2 Z_m^2 = \sigma^2 (Z_n^2 + Z_m^2) \\ \text{und } Z_n^2 + Z_m^2 &\sim \chi_2^2 \text{ mit } n \neq m. \end{aligned}$$

Für die Wahrscheinlichkeit eines negativen Vorzeichens bei Y_n^1 erhält man

$$\begin{aligned} P(X_n^2 - c < 0) &= P(\sigma^2 Z_n^2 < c) \\ &= P(Z_n^2 < \frac{c}{\sigma^2}) \\ &= F_{\chi_1^2} \left(\frac{c}{\sigma^2} \right) \end{aligned}$$

und für Y_n^2

$$\begin{aligned} P(X_n^2 + X_m^2 - 2 \cdot \ln(2) < 0) &= P(X_n^2 + X_m^2 < 2 \cdot \ln(2)) \\ &= P(\sigma^2(Z_n^2 + Z_m^2) < 2 \cdot \ln(2)) \\ &= P(Z_n^2 + Z_m^2 < \frac{2 \cdot \ln(2)}{\sigma^2}) \\ &= F_{\chi_2^2} \left(\frac{2 \cdot \ln(2)}{\sigma^2} \right). \end{aligned}$$

Unter der Nullhypothese ist der Anteil der negativen Vorzeichen für Y_n^1 und Y_n^2 genau 0.5, d.h. $F_{\chi_1^2}(c) = F_{\chi_2^2}(2 \cdot \ln(2)) = 0.5$. In Abbildung 9 sind die Verteilungsfunktionen von $F_{\chi_1^2} \left(\frac{c}{\sigma^2} \right)$ und $F_{\chi_2^2} \left(\frac{2 \cdot \ln(2)}{\sigma^2} \right)$ in Abhängigkeit von σ zu sehen, welche äquivalent zur negativen Vorzeichenverteilung von Y_n^1 bzw. Y_n^2 sind. An dieser Abbildung ist zu erkennen, dass die Wahrscheinlichkeit bei $F_{\chi_2^2} \left(\frac{2 \cdot \ln(2)}{\sigma^2} \right)$ für $\sigma \in [0.5, 1.5]$ schneller fällt als bei $F_{\chi_1^2} \left(\frac{c}{\sigma^2} \right)$. Somit geraten die Anteile der Vorzeichen bei Abweichungen der Nullhypothese schneller in ein Ungleichgewicht, wodurch eine bessere Trennschärfe hervorgeht.

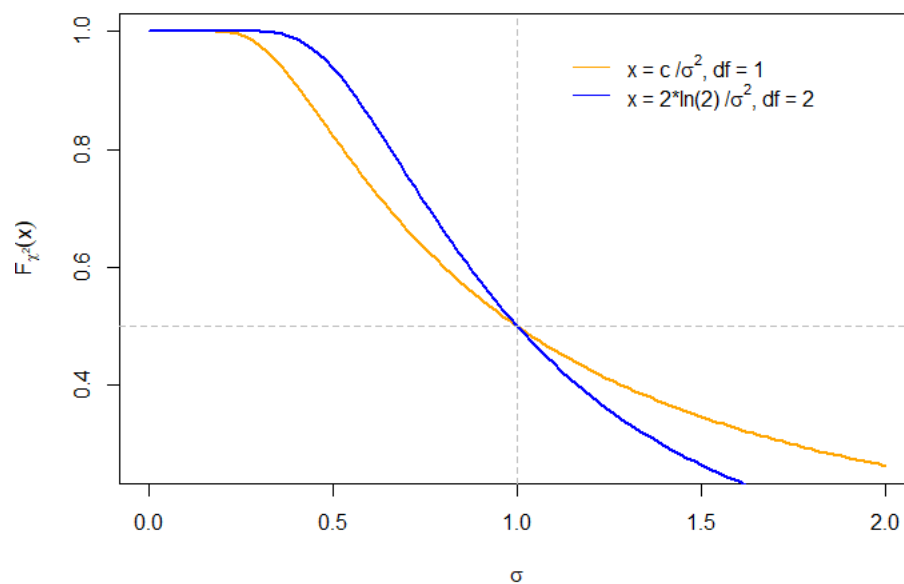


Abbildung 9: Verteilungsfunktion zweier unterschiedlichen χ^2 -Verteilungen mit $(1, 0.5)$ als dem Schnittpunkt der gestrichelten Linien.

7 Fazit

Die zugrunde liegenden Ergebnisse liefern wertvolle Einblicke zu den Eigenschaften der ausgewählten Testverfahren. Zum Testen der Nullhypothese $H_0 : \sigma = \sigma_0$ erweisen sich die K-Vorzeichen-Tiefe-Tests selbst bei Kontamination der Daten als gute Wahl. Insbesondere gilt dies für den vollen 2-Vorzeichen-Tiefe-Test, welcher in allen Szenarien Robustheit und gute Ergebnisse gezeigt hat. Der Q-Schätzer-Test hat besonders gute Ergebnisse bei nicht-kontaminierten Daten geliefert, welche sich aber für Art und Umfang der Kontamination immer weiter verschlechtert haben. Bei größeren Stichproben sind die Ergebnisse des Q-Schätzer-Tests im Vergleich deutlich besser. Darüber hinaus könnte untersucht werden, ob das Signifikanzniveau bei hinreichend großen Stichproben vom Q-Schätzer-Test eingehalten wird. In Hinsicht auf den K-Vorzeichen-Tiefe-Test könnte mit dem Vergleich von weiteren Testverfahren überprüft werden, ob der 2-Vorzeichen-Tiefe-Test auch der gleichmäßig beste Test ist. Jedoch hat sich nach eingehender Recherche gezeigt, dass es nur eine begrenzte Anzahl von Tests gibt, mit denen die gleiche Nullhypothese sinnvoll durchgeführt werden kann. Auch die Skalierung der Residuen haben eine wichtige Bedeutung für die Ergebnisse der Tests. Für $R_n = Y_n^2$ wurden die Ergebnisse nämlich etwas verzerrt. Der Vorteil bei Y_n^2 gegenüber Y_n^1 ist jedoch, dass zur Berechnung der Vorzeichen-Tiefe doppelt so viele Daten für die gleiche Rechenlaufzeit verwendet werden können. Zusätzlich sind die Ergebnisse auch zuverlässiger, da mehrere Daten zur Verfügung stehen, wie man im nicht-kontaminierten Fall sehen kann. Bezüglich des Simulationsdesigns liegt eine mögliche Optimierung in der Art und Weise, wie Ausreißer eingefügt werden. Das Auftreten von Ausreißern in den Stichproben sollte möglichst realitätsnah simuliert werden, um die Aussagekraft der Ergebnisse zu erhalten.

Anhang

Test	Stichprobengröße N			
	30	60	100	200
klassischer Varianztest	0.0504	0.0477	0.0473	0.0502
Test basierend auf Q-Schätzer	0.0739	0.0603	0.0573	0.0528
vereinfachter 2-Tiefe-Test	0.0301	0.0323	0.0381	0.0330
vereinfachter 3-Tiefe-Test	0.0464	0.0437	0.0455	0.0424
vollständiger 2-Tiefe-Test	0.0442	0.0429	0.0610	0.0577
vollständiger 3-Tiefe-Test	0.0484	0.0527	0.0510	0.0536

Tabelle 2: Relative Ablehnungsrate der Tests nach 10.000 Simulationen für Szenario 1) an der Stelle $\sigma = 1 = \sigma_0$ mit Signifikanzniveau $\alpha = 5\%$.

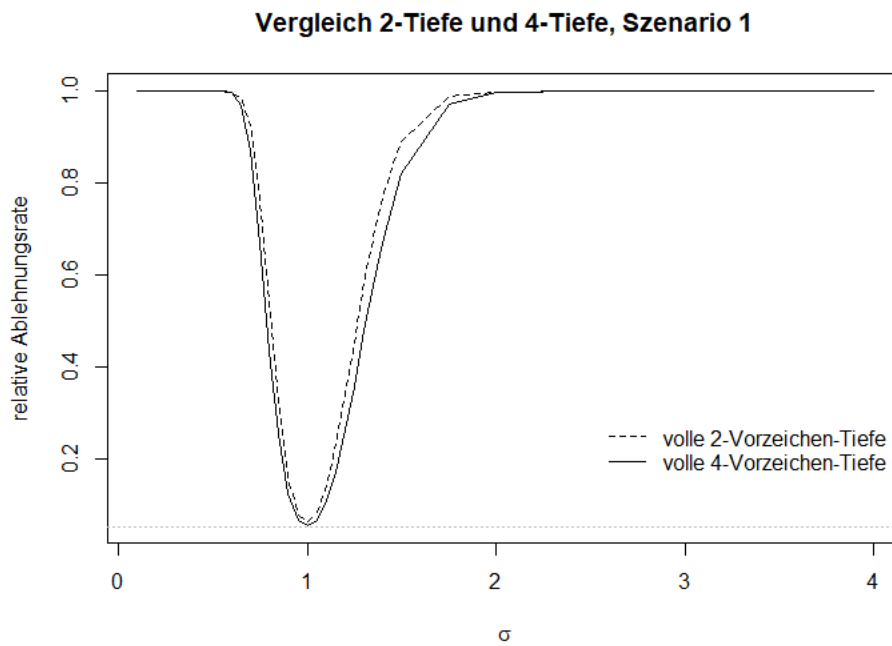


Abbildung 10: Vergleich des vollen K-Vorzeichen-Tiefe-Tests mit $K = 2$ und $K = 4$ in Szenario 1 mit $N = 100$ und $q_4^F(0.05) = 0.05154639$.

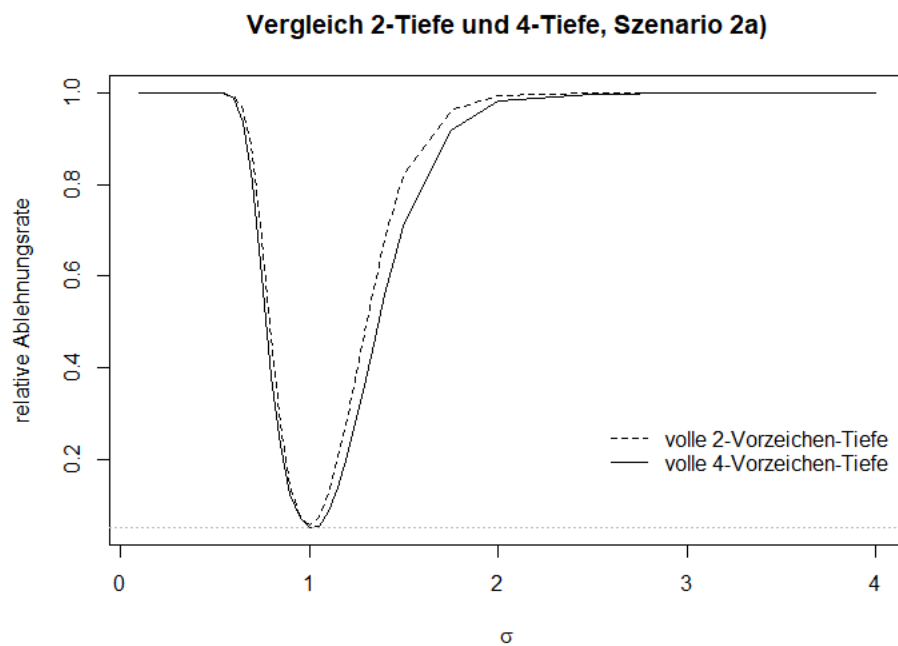


Abbildung 11: Vergleich des vollen K-Vorzeichen-Tiefe-Tests mit $K = 2$ und $K = 4$ in Szenario 2d) mit $N = 100$ und $q_4^F(0.05) = 0.05154639$.

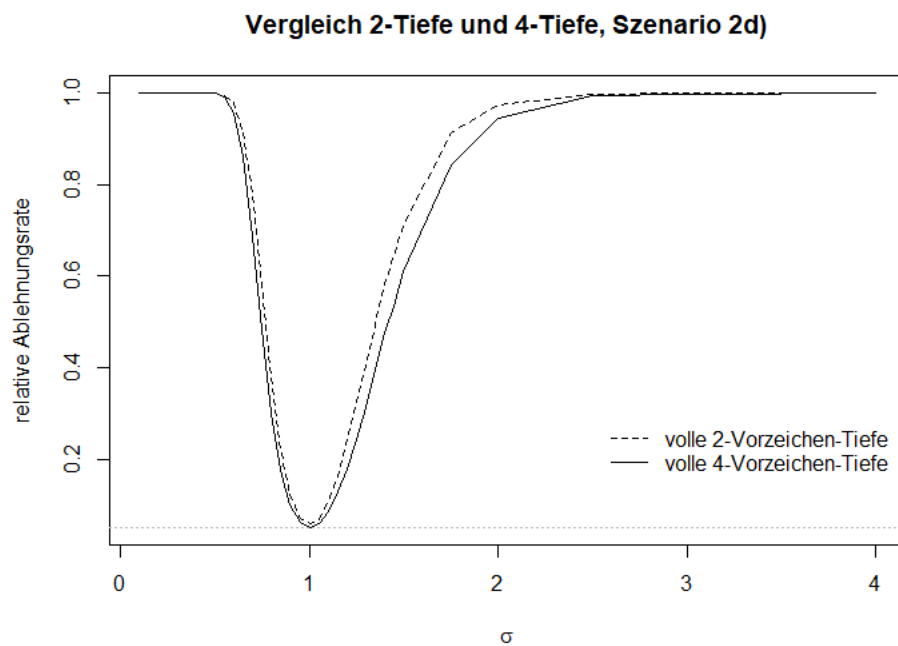


Abbildung 12: Vergleich des vollen K-Vorzeichen-Tiefe-Tests mit $K = 2$ und $K = 4$ in Szenario 2d) mit $N = 100$ und $q_4^F(0.05) = 0.05154639$.

Literatur

- G. Bamberg, F. Baur, and M. Krapp. *Statistik*. Oldenbourg Verlag, 2021.
- J.K. Ghosh. A new proof of the bahadur representation of quantiles and an application. *The Annals of Mathematical Statistics*, 42:1957–1961, 1971.
- Wassily Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19, 1948.
- C. P. Kustozs, C. H. Müller, and M. Wendler. Simplified simplicial depth for regression and autoregressive growth processes. *Journal of Statistical Planning and Inference*, 173:125–146, 2016.
- K. Leckey, D. Malcherzyk, M. Horn, and C. H. Müller. Simple powerful robust tests based on sign depth. *Statistical Papers*, 64:857–882, 2023.
- M. Maechler, P. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, E. L. T. Conceicao, and M. Anna di Palma. *robustbase: Basic Robust Statistics*. R package version 0.99-4-1, 2024.
- D. Malcherzyk, K. Leckey, and C. H. Müller. K-sign depth: From asymptotics to efficient implementation. *Journal of Statistical Planning and Inference*, 215:344–355, 2021.
- F. Mosler, K. und Schmid. *Wahrscheinlichkeitsrechnung und schließende Statistik*. Springer-Verlag, 2006.
- C.H. Müller. Robust scale tests, 2024. Unveröffentlichte Projektskizze, Technische Universität Dortmund.
- S. Müller, C.H. und Uhlig. Estimation of variance components with high breakdown point and high efficiency. *Biometrika*, 88:353–366, 2001.
- S. Nagy. *KDepth: K-depths for one-dimensional and bivariate data*. R package version 0.1.5, 2024. URL <https://github.com/NagyStanislav/Kdepth>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL <https://www.R-project.org/>.

C. Rousseeuw, P.J. und Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88:1273–1283, 1993.

Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, 1980.

Eidesstattliche Versicherung

(Affidavit)

Timm, Marius

Name, Vorname
(surname, first name)

235500

Matrikelnummer
(student ID number)

☒ Bachelorarbeit
(Bachelor's thesis)

☐ Masterarbeit
(Master's thesis)

Titel
(Title)

Vorzeichen-Varianz-Tests im Vergleich mit anderen
Varianz-Tests

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.

Köln, 06.04.2025

Ort, Datum
(place, date)

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:*

Köln, 06.04.2025

Ort, Datum
(place, date)

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.