

New characterisation-based tests for symmetry

CJL van Heerden

 **orcid.org 0000-0003-4007-8622**

Thesis accepted in fulfilment of the requirements for the degree
Doctor of Philosophy in Science with Statistics at the
North-West University

Promoter: Prof C Pretorius

Graduation August 2021

22675752

New characterisation-based tests for symmetry

Carl J. L. van Heerden
North-West University
Potchefstroom

March 2021

Acknowledgements

I wish to express my sincere appreciation and gratitude to the following individuals.

Our Heavenly Father, thank You for blessing me with the ability and opportunity to receive an excellent education. To my supervisor, Prof Charl Pretorius, thank you for sharing your passion for and knowledge of mathematics and statistics with me. Without you this study would not have been possible. A final word of thanks goes to the South African Statistical Association and the National Research Foundation of South Africa for generous financial support.

Disclaimer. *This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Number 114630). The findings and conclusions expressed in this thesis are that of the author alone, and the National Research Foundation accepts no liability whatsoever in this regard.*

Abstract

We propose new tests for symmetry of a distribution around an *unspecified* centre. The first set of tests is based on a lesser-known characterisation involving symmetric integration of the distribution function around the estimated centre. The tests are straightforward to implement and, unlike many of the existing tests, do not require the practitioner to make a choice of a tuning parameter. This is an important consideration from a practical perspective, as an optimal choice of the tuning parameter usually depends on the unknown underlying distribution and it is known that the performance of tests can be quite sensitive, in terms of both size and power, to the choice of the tuning parameter. Numerical simulations indicate that the new tests have favourable finite-sample properties in the sense that they are level preserving and exhibit competitive power. The asymptotic null distribution of the Kolmogorov–Smirnov type test statistic is derived in the multiple linear regression setup where the goal is to test for symmetry of the error distribution, and the test is shown to be asymptotically consistent against general alternatives.

The second set of tests is based on the idea of substituting smooth kernel-type distribution function estimators for the nonsmooth empirical counterparts appearing in classical Kolmogorov–Smirnov and Cramér–von Mises tests for symmetry. We demonstrate theoretically that, in finite samples, this leads to tests which potentially have significantly higher power. Numerical studies support this and also show that these tests have empirical size close to the nominal level. Even though kernel-type distribution function estimators have been studied extensively in the literature, its benefits in goodness-of-fit testing have only recently been recognised (see, e.g., Racine and van Keilegom, 2020).

Recently, Allison and Pretorius (2017) proposed two new tests for symmetry around a *specified* centre. Although these tests were shown to compete very well with other modern tests for symmetry in a Monte Carlo study, the authors did not investigate the asymptotic properties of their tests. In this study we derive the asymptotic null distribution of one of these tests in the context of testing whether the observations have a distribution symmetric around an unknown centre.

The study is concluded with an application of the new tests to real-world data.

Keywords: Bootstrap test; characterisation of symmetry; goodness-of-fit test; kernel smoothing; linear regression model; test for symmetry.

Contents

1	Introduction	8
1.1	Objectives	10
1.2	Outline	10
1.3	Frequently-used notation	11
2	Literature study	13
2.1	Tests based on the distribution function	14
2.2	Tests based on the characteristic function	16
2.3	Other tests	25
2.4	Additional tests that did not form part of this study	29
2.5	Regression estimators	30
2.6	Bootstrap hypothesis testing	33
2.7	Some symmetric and skew probability distributions	37
3	New tests based on the integrated empirical process	42
3.1	The case of independent observations	44
3.2	Error terms in linear regression models	53
3.3	Numerical study	59
4	Tests based on smooth distribution function estimation	76
4.1	Linear regression	78
4.2	Theoretical analysis	84
4.3	Numerical study	88
5	Some asymptotic results for existing tests for symmetry	103
5.1	Case of a specified centre of symmetry	104
5.2	Case of an unknown median	112
6	Empirical study	120
7	Conclusions	125

A Some basic results on kernel smoothing	126
A.1 Proof of Lemma 4.1	127
Bibliography	129

Chapter 1

Introduction

Many authors have stressed that symmetry plays an important role in nonparametric statistics as well as in structured models where a standard assumption is that the errors are symmetrically distributed. For example, Bickel (1982) showed that under the assumption of symmetry, slope parameters in a linear regression model can be estimated adaptively to attain the same efficiency as that of the maximum-likelihood estimators if the true distribution of the errors were known (also see Hillebrand and Müller, 2006). In GARCH models, symmetry of the errors is necessary for obtaining consistent quasi-maximum-likelihood estimates of model parameters (see Newey and Steigerwald, 1997). As a result, a multitude of tests have been developed to test this hypothesis, and testing for symmetry remains an area of active research (see, e.g., Hušková and Meintanis, 2012; Ngatchou-Wandji and Harel, 2013; Nikitin and Ahsanullah, 2015; Milošević and Obradović, 2016, 2019; Gaigall, 2020; Ivanović, Milošević and Obradović, 2020). We propose new tests for symmetry of univariate distributions, and study the theoretical and finite-sample properties of these tests in the case where the centre of symmetry is unspecified. The new tests are shown to exhibit favourable finite-sample and asymptotic properties and are very competitive alternatives to existing tests. It is shown that the tests have similar level-preserving and power properties as some of the modern powerful tests for symmetry.

The first new tests are based on a lesser-known property of symmetric distributions which was pointed out by Silver and Costa (1997). We demonstrate that this property in fact characterises symmetric distributions, and can therefore be used to construct tests which are consistent against general alternatives (nonsymmetric distributions). Numerical studies using these tests yield favourable results, which show that they are competitive alternatives to existing tests for symmetry. We mention specifically that the new tests are level preserving and also have power comparable to popular tests based on the characteristic function (such as

the test of Hušková and Meintanis, 2012), which have been shown to typically have higher power than classical tests (e.g. Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling). An attractive feature of these new tests is that they do not require the statistician to make a choice of a tuning parameter, unlike many of the modern tests (such as the existing tests based on the characteristic function).

The second set of new tests is based on the idea of kernel smoothing the distribution function estimates used in classical tests for symmetry. This idea, which has gained popularity recently (see, e.g., Pérez-Veiga and Pardo-Fernández, 2019; Racine and van Keilegom, 2020), has been shown to result in substantial power gains. The idea of kernel smoothing distribution function estimators is an established concept and is well studied (see Nadaraya, 1964; Reiss, 1981; Bowman, Hall and Prvan, 1998), but its benefits in goodness-of-fit testing are only now being recognised. The tests are straightforward to construct and involve substituting kernel distribution function estimators for the usual empirical distribution functions.

Apart from proposing and refining these new tests, we derive their asymptotic null distributions to confirm that the tests are well behaved for increasing sample sizes. We also theoretically confirm that the tests are able to detect deviations from symmetry by showing that the tests are asymptotically consistent against general alternatives. With application to regression modelling in mind, these asymptotic properties are derived in the context of testing whether the disturbance in a parametric regression model is distributed symmetrically around zero. As a special case this includes testing whether independent and identically distributed random variables have a distribution which is symmetric around an unspecified centre.

Recently, Allison and Pretorius (2017) proposed two new tests for symmetry around a known median. Although these tests were shown to compete very well with modern tests for symmetry (specifically the tests by Hettmansperger, McKean and Sheather, 2002; Neumeyer, Dette and Nagel, 2005; Hušková and Meintanis, 2012; Meintanis, Allison and Santana, 2016) in a Monte Carlo study, Allison and Pretorius (2017) did not investigate the asymptotic properties of their tests. In this study we investigate the asymptotic behaviour of one of these tests when used to test whether the observations were generated by a distribution symmetric around an unknown median.

The theoretical work in this study is complemented with numerical simulation studies to compare the finite-sample performance of the newly-developed tests with that of some of the established tests for symmetry found in the literature. The simulation studies are carried out in the general context of linear regression models.

Future avenues of research include extending the asymptotic results to the case of testing for symmetry of the errors appearing in nonparametric regression mod-

els, or of errors appearing in time series models allowing a Wold representation (which include ARMA models as a special case). Another possibility is to generalise the testing procedures to multidimensional settings to test for radial or elliptical symmetry.

1.1 Objectives

The main objectives of this study can be summarised as follows:

- Provide an overview of some of the existing tests for symmetry of univariate distributions.
- Propose a new characterisation of symmetry of a univariate distribution and use this characterisation to construct simple new tests. Derive the asymptotic null distributions of the new tests and show that the tests are consistent against general alternatives.
- Investigate the finite-sample and asymptotic effects of smoothing the empirical distribution function estimators in classical Kolmogorov–Smirnov and Cramér–von Mises tests for univariate symmetry.
- Study the asymptotic properties of a test for symmetry recently proposed by Allison and Pretorius (2017). Specifically establish consistency of the test as well as derive the limit null distribution, which was not pursued by the original authors.
- Use Monte Carlo simulation to assess the finite-sample behaviour of all newly-proposed tests and compare the behaviour of these tests to that of existing tests for symmetry.
- Illustrate the application of the new tests to real-world data.

1.2 Outline

The thesis is structured as follows. In Chapter 2 we provide a literature review of some of the most established tests for symmetry of a univariate distribution and discuss some of the properties associated with each of them. We discuss bootstrap hypothesis testing and specifically how the bootstrap can be applied when testing for symmetry. In Chapter 3 we propose new tests for symmetry of a univariate distribution based on a lesser-known characterisation of symmetry by Silver and Costa (1997). We study asymptotic properties of the new tests and show that they have favourable finite-sample level and power properties. In Chapter 4 we

investigate the finite-sample and asymptotic effects of kernel smoothing in classical Kolmogorov–Smirnov and Cramér–von Mises tests for symmetry of a univariate distribution. We provide a theoretical justification of the increase in power observed in the numerical studies. In Chapter 5 we study the asymptotic properties of a test for symmetry proposed by Allison and Pretorius (2017), and Chapter 6 concludes the thesis with an empirical application of the newly-proposed tests to real-world data.

1.3 Frequently-used notation

For easy reference we provide the following list of notation occurring throughout the text.

- $\mathbb{R}^d$ is the set of all d -dimensional real numbers; $\mathbb{R} := \mathbb{R}^1$.
- $\mathbb{N} = \{1, 2, \dots\}$ is the set of natural numbers.
- $a \wedge b := \min(a, b)$ and $a \vee b := \max(a, b)$ for any $a, b \in \mathbb{R}$.
- $x_i, i = 1, 2, \dots, d$, denotes the i th element of $x \in \mathbb{R}^d$, i.e.

$$x = (x_1, x_2, \dots, x_d).$$

We say that x is a d -vector.

- $\|\cdot\|$ denotes the Euclidean norm of a vector. That is, if $x \in \mathbb{R}^d$, then

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}.$$

- For any $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer less than or equal to x and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x .
- $\xrightarrow{d}$ and $\xrightarrow{P}$ denote convergence *in distribution* and *in probability*, respectively.

To avoid ambiguity, we explicitly define the asymptotic notation $O(\cdot)$, $o(\cdot)$, $O_P(\cdot)$ and $o_P(\cdot)$, which will be used throughout to describe error terms in asymptotic expressions. An insightful discussion of these definitions, as well as equivalent alternative definitions, is given by Janson (2009). Here, the sequences $\{a_n\}$ and $\{b_n\}$ denote sequences of real numbers, with $a_n > 0$ for all n .

Definition 1.1. We say that $b_n = o(a_n)$ as $n \rightarrow \infty$ if for every $\varepsilon > 0$ there exists an integer n_ε such that $a_n^{-1}|b_n| \leq \varepsilon$ for all $n \geq n_\varepsilon$, that is,

$$b_n = o(a_n) \iff \lim_{n \rightarrow \infty} \frac{b_n}{a_n} = 0.$$

Definition 1.2. We say that $b_n = O(a_n)$ as $n \rightarrow \infty$ if there exist constants C and n_0 such that $a_n^{-1}|b_n| \leq C$ for all $n \geq n_0$, that is,

$$b_n = O(a_n) \iff \limsup_{n \rightarrow \infty} \frac{|b_n|}{a_n} < \infty.$$

Probabilistic counterparts of the above two definitions will also be needed.

Definition 1.3. Let $\{X_n\}$ be any sequence of random variables. We say that $X_n = o_P(a_n)$ as $n \rightarrow \infty$ if for every $\varepsilon > 0$ there exists an integer n_ε such that

$$P\left(\frac{|X_n|}{a_n} > \varepsilon\right) < \varepsilon \quad \text{for all } n \geq n_\varepsilon.$$

The above definition is a generalisation of the usual definition of convergence in probability and we shall sometimes use the equivalent notation

$$X_n \xrightarrow{P} 0$$

as $n \rightarrow \infty$ instead of

$$X_n = o_P(1).$$

By using the standard definition of a limit, one sees that the statement in Definition 1.3 can be written as follows:

$$\forall \varepsilon, \delta > 0 \exists n_{\varepsilon, \delta} > 0 : P(|X_n - X| > \delta) < \varepsilon \quad \forall n \geq n_{\varepsilon, \delta}.$$

The use of a single quantity ε in Definition 1.3 to bound the quantity $|X_n - X|$ as well as control the asymptotic deviation of $P(|X_n - X| > \varepsilon)$ from zero is purely for notational convenience. One must realise, however, that ε plays two distinct roles, and that there need not be any relationship between these two roles. The point is, convergence in probability means that the probability of $|X_n - X|$ being greater than *any arbitrary (but fixed) positive number* tends to zero.

Definition 1.4. Let $\{X_n\}$ be any sequence of random variables. We say that $X_n = O_P(a_n)$ as $n \rightarrow \infty$ if for every $\varepsilon > 0$ there exist a constant C_ε and an integer n_ε such that $P(a_n^{-1}|X_n| > C_\varepsilon) < \varepsilon$ for all $n \geq n_\varepsilon$, that is,

$$X_n = O_P(a_n) \iff \lim_{C \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left(\frac{|X_n|}{a_n} > C\right) = 0.$$

Chapter 2

Literature study

This study concerns the problem of testing whether a sequence of independent and identically distributed random variables $Y_1, \dots, Y_n$ have a distribution which is symmetric around an *unknown* median η . More formally, denoting the distribution function of the Y_i by F , we would like to test the hypothesis

$$H_0 : F(\eta + t) + F(\eta - t) = 1 \quad \text{for all } t \geq 0, \quad (2.1)$$

against general alternatives.

The assumption of symmetry plays an important role in many statistical procedures and modelling approaches. This includes the case of the classical linear regression model where it is known that symmetry of the error distribution is a sufficient condition for the existence of asymptotically efficient adaptive estimators of the slope parameter (see Bickel, 1982). In GARCH models, symmetry of the errors is necessary for obtaining consistent quasi-maximum-likelihood estimates of model parameters (see Newey and Steigerwald, 1997). There exist many other examples; see Ivanović et al. (2020) for a discussion. Antille, Kersting and Zucchini (1982) also mention several other reasons that the assumption of symmetry plays an important role. For example, if the distribution is asymmetric, then there is no longer one natural measure of location which might have implications in the interpretation of results. They also mention that some popular robust techniques are based on the assumption of symmetry.

An intuitive way of testing for symmetry around an unknown centre is to estimate the centre of the data and then applying a test for symmetry around zero by relocating the data using the estimated centre.

For clarity, we now introduce some additional notation which will be used throughout the remainder of the thesis. The observations $Y_1, \dots, Y_n$ have the representation

$$Y_i = \eta + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.2)$$

where $\varepsilon_1, \dots, \varepsilon_n$ are *unobserved* i.i.d. random variables with distribution function F_ε (say). The hypothesis in (2.1) is true if and only if F_ε is the distribution function of a random variable that is symmetric around zero. That is, the original problem is equivalent to testing the hypothesis

$$H_0 : F_\varepsilon(t) + F_\varepsilon(-t) = 1 \quad \text{for all } t \geq 0, \quad (2.3)$$

against general alternatives.

The model in (2.2) can be regarded as a location model, or an intercept-only model, where the intercept η needs to be estimated by an appropriate estimator, say $\hat{\eta}_n$. Since the ε_i are unobservable, we base all tests for the hypothesis in (2.3) on the residuals

$$\hat{\varepsilon}_i = Y_i - \hat{\eta}_n, \quad i = 1, \dots, n, \quad (2.4)$$

which are simply location-shifted versions of the original observations.

In this chapter we will describe some of the existing tests for symmetry in terms of the location (intercept-only) model in (2.2). That is, we will assume that the $Y_1, \dots, Y_n$ are centred around a fixed η which needs to be estimated. In later chapters, however, we shall generalise this to the case where the $Y_1, \dots, Y_n$ are generated by a linear regression model. In such cases it is intuitive and customary to use base tests on the regression residuals. As many of the tests in the literature were introduced and studied in the special case of (2.2), we first restrict our attention to the location model only.

2.1 Tests based on the distribution function

2.1.1 Kolmogorov–Smirnov and Cramér–von Mises tests

The classical Kolmogorov–Smirnov and Cramér–von Mises tests for symmetry are probably the most extensively studied tests for symmetry in the literature. The earliest work considers the problem of testing for symmetry around a *specified* η . Works include those by Smirnov (1947), Butler et al. (1969), Orlov (1972), Rothman and Woodroffe (1972), Srinivasan and Godio (1974), Hill and Rao (1977) and Koziol (1980). Interestingly, all these authors found that the Kolmogorov–Smirnov and Cramér–von Mises tests for symmetry are typically distribution free. This is very convenient from a practical perspective, as critical values can be tabulated in advance and directly used by practitioners regardless of which distribution generated the data.

In the case where the centre of symmetry η is *unknown* and needs to be estimated, these tests typically lose their distribution-free character. For the classical Kolmogorov–Smirnov and Cramér–von Mises tests, this was first pointed out by

Koziol (1983). Similar findings were also reported for certain generalisations of the Cramér–von Mises tests (Aki, 1987).

We now formally describe the above-mentioned classical tests that may be used for testing the hypothesis in (2.3). Define the empirical symmetry process

$$\widehat{S}_n(t) = \widehat{F}_n(t) + \widehat{F}_n(-t-) - 1, \quad t \in \mathbb{R},$$

where $\widehat{F}_n$ is the empirical distribution function based on the residuals defined in (2.4), i.e. $\widehat{F}_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}(\widehat{\varepsilon}_i \leq t)$. Throughout the remainder of the thesis, the notation $\widehat{F}_n(-t-)$ denotes the left limit of the function $\widehat{F}_n$ at the point $-t$, i.e.

$$\widehat{F}_n(-t-) = \lim_{\delta \downarrow 0} \widehat{F}_n(-t - \delta).$$

Classical Kolmogorov–Smirnov and Cramér–von Mises statistics for testing the hypothesis in (2.3) are

$$\begin{aligned} \widehat{U}_n &= \sup_{t \geq 0} |\widehat{S}_n(t)| = \sqrt{n} \sup_{t \geq 0} |\widehat{F}_n(t) + \widehat{F}_n(-t-) - 1|, \\ \widehat{W}_n &= \int_0^\infty \widehat{S}_n^2(t) d\widehat{Q}_n(t) = n \int_0^\infty [\widehat{F}_n(t) + \widehat{F}_n(-t-) - 1]^2 d\widehat{Q}_n(t), \end{aligned} \quad (2.5)$$

respectively, where $\widehat{Q}_n(t) = \widehat{F}_n(t) - \widehat{F}_n(-t-)$. In light of (2.3), one can expect that $\widehat{U}_n$ and $\widehat{W}_n$ will typically be close to zero under the null hypothesis of symmetry, and large under alternatives. A test for symmetry therefore rejects for large values of these statistics.

The classical Kolmogorov–Smirnov and Cramér–von Mises tests were studied more recently in the case of testing whether the errors in a linear regression model (Neumeyer et al., 2005) or a nonparametric regression model (Neumeyer and Dette, 2007) are symmetric around zero. In these settings the tests are not distribution free and depend on the unknown location parameter η and the regression estimators in a complicated way. Also see Koul (2002).

2.1.2 A test by Boos (1982)

For the case where η is unspecified, Boos (1982) proposed a test for symmetry based on a minimisation of the Cramér–von Mises distance characterisation of the Hodges–Lehmann estimator.

The test statistic proposed by Boos (1982) is a suitably-scaled minimisation of the distance

$$d_n(\eta) = \int_{-\infty}^{\infty} [F_n(\eta - x) + F_n(\eta + x) - 1]^2 dx,$$

where $F_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}(Y_i \leq t)$ is the empirical distribution function based on the observations $Y_1, \dots, Y_n$. Boos (1982) showed that this statistic can equivalently be

expressed in closed form as

$$d_n(\eta) = \frac{-2}{n^2} \sum_{1 \leq i < j \leq n} |Y_i - Y_j| + \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left| \frac{Y_i + Y_j}{2} - \eta \right|,$$

and that $d_n(\eta)$ is minimised by

$$\hat{\eta}_{\text{HL}} = \text{median} \left\{ \frac{Y_i + Y_j}{2}, 1 \leq i, j \leq n \right\}.$$

That is, $\hat{\eta}_{\text{HL}}$ is the median of all the pairwise means of all possible two-element subsets of the sample. This is a variant of the well-known Hodges–Lehmann estimator (Hodges and Lehmann, 1963).

As a test for symmetry, Boos (1982) then suggested using a test based on the statistic

$$B = n \left[\frac{\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n |Y_i + Y_j - 2\hat{\eta}_{\text{HL}}|}{\sum_{i < j} |Y_i - Y_j|} - 1 \right]. \quad (2.6)$$

This statistic is obtained by multiplying $n^2(n-1)^{-1}d_n(\hat{\eta}_{\text{HL}})$ by Gini's mean difference

$$\binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} |Y_i - Y_j|.$$

The null hypothesis of symmetry is rejected for large values of B . Critical values for this test statistic were computed by Boos (1982) using Monte Carlo simulation with the observations Y_i generated from a logistic distribution. He found that these values did not differ greatly from the asymptotic critical values.

2.2 Tests based on the characteristic function

A few modern tests for symmetry, which have proven to be quite powerful under alternatives, make use of the empirical characteristic function. Recall that the characteristic function of a random variable X is defined as

$$\varphi(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX)) + i \mathbb{E}(\sin(tX)) =: C_\varphi(t) + iS_\varphi(t),$$

where $C_\varphi(\cdot)$ is the *real* part and $S_\varphi(\cdot)$ the *imaginary* part of $\varphi(\cdot)$. The characteristic function has several properties which make it very attractive from a theoretical point of view. In particular, we mention that the characteristic function of a random variable X always exists (unlike, e.g., the moment generating function) and is uniformly continuous and bounded.

Based on a set of data $X_1, \dots, X_n$ generated by the characteristic function φ , a natural estimator of φ is the empirical characteristic function φ_n defined as

$$\varphi_n(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j} = C_{\varphi,n}(t) + iS_{\varphi,n}(t), \quad (2.7)$$

where $C_{\varphi,n}(t) = n^{-1} \sum_{j=1}^n \cos(tX_j)$ and $S_{\varphi,n}(t) = n^{-1} \sum_{j=1}^n \sin(tX_j)$.

The use of the empirical characteristic function in goodness-of-fit testing has been shown to have several advantages over tests based on the empirical distribution function. Specifically, procedures based on the empirical characteristic function

1. typically work under weaker assumptions,
2. are often more convenient from a computational point of view in the sense that test statistics can typically be written in closed form, and
3. are easily extendable to the multivariate context, which is not always the case for methods based on the empirical distribution function.

For a thorough review and developments in the application of the empirical characteristic function in goodness-of-fit testing procedures, see Csörgő (1984), Hušková and Meintanis (2008a), Hušková and Meintanis (2008b) and Meintanis (2016).

We now discuss several aspects of goodness-of-fit tests employing the empirical characteristic function. Although some of the ideas mentioned are applicable to goodness-of-fit tests in general, we present these ideas in the specific context of testing for symmetry around an unknown centre η .

2.2.1 A test by Hušková and Meintanis (2012)

A well-known characterisation of symmetry states that the distribution of a random variable X is symmetric around the origin if and only if the imaginary part of the characteristic function of X is zero. That is, if we denote the characteristic function of a random variable X by φ , then

$$X \text{ is symmetric around zero} \iff \operatorname{Im}(\varphi(t)) = 0, \forall t \in \mathbb{R}. \quad (2.8)$$

This characterisation was first used to develop a test for symmetry by Feuerverger and Mureika (1977) and later also by Koutrouvelis (1985), Ghosh and Ruymgaart (1992), Neuhaus and Zhu (1998), Henze, Klar and Meintanis (2003) and Hušková and Meintanis (2012) in several different contexts.

An intuitive choice for a test criterion is to calculate the distance of the imaginary part of the empirical characteristic function from zero, i.e.

$$H = n \int_{-\infty}^{\infty} S_{\varphi,n}^2(t) w(t) dt, \quad (2.9)$$

where, as in (2.7), $S_{\varphi,n}(\cdot)$ denotes the imaginary part of the empirical characteristic function and $w(\cdot)$ is an appropriate nonnegative weight function which is symmetric around zero. This test rejects for large values of the test statistic H .

The introduction of a weight function is quite common in tests based on the empirical characteristic function and is necessary to ensure existence of the integral appearing in H . Important considerations for the choice of weight function are computational simplicity as well as level and power properties of the resulting tests. For a discussion on the role of the weight function, see Hušková and Meintanis (2007).

Observe that the statistic H can be written as

$$\begin{aligned}
H &= n \int_{-\infty}^{\infty} \left[\frac{1}{n} \sum_{j=1}^n \sin(tX_j) \right]^2 w(t) dt \\
&= \frac{1}{n} \int_{-\infty}^{\infty} \sum_{j=1}^n \sum_{k=1}^n \sin(tX_j) \sin(tX_k) w(t) dt \\
&= \frac{1}{2n} \int_{-\infty}^{\infty} \sum_{j=1}^n \sum_{k=1}^n [\cos(tX_j - tX_k) - \cos(tX_j + tX_k)] w(t) dt \\
&= \frac{1}{2n} \sum_{j=1}^n \sum_{k=1}^n [I_w(X_j - X_k) - I_w(X_j + X_k)],
\end{aligned}$$

where

$$I_w(z) = \int_{-\infty}^{\infty} \cos(tz) w(t) dt.$$

Calculation of the test statistic boils down to evaluating the integral in I_w . The weight function w may be chosen in a way that facilitates integration and we will now show that, if chosen correctly, the integral in I_w may be computed without numerical quadrature. First note that, if R is a random variable with density w , then the real part of the characteristic function of R (evaluated at the point z) is given by

$$\operatorname{Re} \left[\mathbb{E} \left(e^{-izR} \right) \right] = \mathbb{E}(\cos(zR)) = \int_{-\infty}^{\infty} \cos(tz) w(t) dt = I_w(z).$$

If we now choose w within the family of spherical distributions, we have by the property stated in (2.8) that $I_w(z)$ is a multiple of the characteristic function of a random variable with density w .

A popular choice for the weight function is

$$w(t) = w_{1,a}(t) = e^{-at^2}, \quad a > 0.$$

Making use of the fact that the characteristic function (evaluated at z) corresponding to a $N(0, \sigma^2)$ distribution is given by $e^{-\frac{1}{2}\sigma^2 z^2}$, this choice of w leads to

$$\begin{aligned}
I_{w_{1,a}}(z) &= \int_{-\infty}^{\infty} \cos(tz) e^{-at^2} dt \\
&= \sqrt{\frac{\pi}{a}} \int_{-\infty}^{\infty} \cos(tz) \sqrt{\frac{a}{\pi}} e^{-at^2} dt \\
&= \sqrt{\frac{\pi}{a}} e^{-z^2/(4a)}.
\end{aligned} \tag{2.10}$$

Another popular choice for w is

$$w(t) = w_{2,a}(t) = e^{-a|t|}, \quad a > 0,$$

which is a multiple of the Laplace density function. In this case,

$$\begin{aligned} I_{w_{2,a}}(z) &= \int_{-\infty}^{\infty} \cos(tz) e^{-a|t|} dt \\ &= \frac{2}{a} \int_{-\infty}^{\infty} \cos(tz) \frac{a}{2} e^{-a|t|} dt \\ &= \frac{2a}{a^2 + z^2}, \end{aligned} \tag{2.11}$$

where we made use of the fact that the characteristic function (evaluated at z) of a Laplace distribution with mean zero and scale parameter σ is given by $(1 + \sigma^2 z^2)^{-1}$.

In fact, for any spherical density w_a there exists a function $\Psi_a(\cdot)$ such that $I_{w_a}(x) = \Psi_a(|x|)$. The function $\Psi(\cdot)$ is called the *characteristic kernel* or *characteristic generator* and its functional form depends on the subfamily of spherical distributions. Subfamilies of spherical distributions with simple kernels are the class of *spherical stable distributions* (see Nolan, 2013) with

$$\Psi_a(u) = e^{-au^\gamma}, \quad a > 0, 0 < \gamma \leq 2,$$

and the class of *generalised multivariate Laplace distributions* (see Kozubowski, Podgórski and Rychlik, 2013) with

$$\Psi_a(u) = (1 + u^2/a)^{-\gamma}, \quad a > 0, \gamma > 0.$$

Hušková and Meintanis (2012) derived the limit null distribution of H for the case of a homoskedastic linear regression model as well as for a nonparametric heteroskedastic regression model. Since the asymptotic null distribution depends on the error distribution F_ε , it does not provide a good approximation for the critical values of the test. To ensure a consistent test, Hušková and Meintanis (2012) followed Neumeyer et al. (2005) by using a symmetric wild bootstrap procedure to approximate the critical values (as described in Section 2.6.2 of this thesis).

2.2.2 A test by Allison and Pretorius (2017) based on a characterisation by Ahsanullah

Allison and Pretorius (2017) assumed that the distribution function F has a *known* median which they assumed, without loss of generality, to be zero and considered the following characterisation of a symmetric distribution by Ahsanullah (1992): *Let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ denote the order statistics of a random sample $X_1, \dots, X_n$ from a continuous distribution function F . Then $|X_{(1)}|$ and $|X_{(n)}|$ are identically distributed if and only if F is symmetric with respect to zero.*

Allison and Pretorius (2017) proposed a test for symmetry based on this characterisation by considering the difference between the empirical characteristic functions of $|X_{(1)}|$ and $|X_{(n)}|$. In contrast, Nikitin and Ahsanullah (2015) also proposed a test based on this characterisation, but instead used the empirical distribution function. The test of Allison and Pretorius (2017) is based on the process

$$\begin{aligned}
V_n(t) &= \mathbf{E} \left(e^{it|X_{(n)}|} \right) - \mathbf{E} \left(e^{it|X_{(1)}|} \right) \\
&= n \int_0^\infty e^{itx} [F^{n-1}(x)f(x) + F^{n-1}(-x)f(-x)] dx \\
&\quad - n \int_0^\infty e^{itx} [(1-F(x))^{n-1}f(x) + (1-F(-x))^{n-1}f(-x)] dx \\
&= n \int_{-\infty}^\infty e^{it|x|} [F^{n-1}(x) - (1-F(x))^{n-1}] dF(x).
\end{aligned} \tag{2.12}$$

From the characterisation of Ahsanullah (1992) stated above, it follows that F is symmetric around zero if and only if $V_n(t) = 0$ for all t .

In the same spirit as Hušková and Meintanis (2012), Allison and Pretorius (2017) used (2.12) and proposed the test statistic

$$A_a^{(1)} = n \int_{-\infty}^\infty |\widehat{V}_n(t)|^2 w_a(t) dt,$$

where $\widehat{V}_n(t)$ is an empirical counterpart of $V_n(t)$ and $w_a(t)$ is an appropriate weight function. Allison and Pretorius (2017) estimated $V_n(t)$ by substituting the (scaled) empirical function $\widetilde{F}_n(t) = (n-1)^{-1} \sum_{k=1}^n \mathbf{1}(X_{(k)} < t)$. They obtained

$$\begin{aligned}
\widehat{V}_n(t) &= n \int_{-\infty}^\infty e^{it|x|} [\widetilde{F}_n^{n-1}(x) - (1 - \widetilde{F}_n(x))^{n-1}] dF_n(x) \\
&= n \sum_{k=1}^n e^{it|X_{(k)}|} [\widetilde{F}_n^{n-1}(X_{(k)}) - (1 - \widetilde{F}_n(X_{(k)}))^{n-1}] \cdot \frac{1}{n} \\
&= \sum_{k=1}^n e^{it|X_{(k)}|} \left[\left(\frac{k-1}{n-1} \right)^{n-1} - \left(1 - \frac{k-1}{n-1} \right)^{n-1} \right] \\
&= \sum_{k=1}^n \left[\left(\frac{k-1}{n-1} \right)^{n-1} - \left(\frac{n-k}{n-1} \right)^{n-1} \right] [\cos(t|X_{(k)}|) + i \sin(t|X_{(k)}|)] \\
&=: \sum_{k=1}^n u_{k,n} [\cos(t|X_{(k)}|) + i \sin(t|X_{(k)}|)].
\end{aligned}$$

Observe that $|u_{k,n}|$ is large for values of k near the endpoints of the allowable range $\{1, 2, \dots, n\}$ and small for values near the centre of this range. Intuitively, this means that the weight associated with the smallest observations and largest observations in the sample will be larger than that of intermediate observations.

$A_a^{(1)}$ can be written as follows:

$$\begin{aligned}
A_a^{(1)} &= n \int_{-\infty}^{\infty} |\widehat{V}_n(t)|^2 w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \left| \sum_{k=1}^n u_{k,n} [\cos(t|X_{(k)}|) + i \sin(t|X_{(k)}|)] \right|^2 w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \left[\left(\sum_{k=1}^n u_{k,n} \cos(t|X_{(k)}|) \right)^2 + \left(\sum_{k=1}^n u_{k,n} \sin(t|X_{(k)}|) \right)^2 \right] w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \left[\sum_{j=1}^n \sum_{k=1}^n u_{j,n} u_{k,n} \cos(t|X_{(j)}|) \cos(t|X_{(k)}|) \right. \\
&\quad \left. + \sum_{j=1}^n \sum_{k=1}^n u_{j,n} u_{k,n} \sin(t|X_{(j)}|) \sin(t|X_{(k)}|) \right] w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \sum_{j=1}^n \sum_{k=1}^n u_{j,n} u_{k,n} \cos(t|X_{(j)}| - t|X_{(k)}|) w_a(t) dt \\
&= n \sum_{j=1}^n \sum_{k=1}^n u_{j,n} u_{k,n} \int_{-\infty}^{\infty} \cos(t|X_{(j)}| - t|X_{(k)}|) w_a(t) dt \\
&= n \sum_{j=1}^n \sum_{k=1}^n u_{j,n} u_{k,n} I_{w_a}(|X_{(j)}| - |X_{(k)}|),
\end{aligned}$$

where, as before, $I_{w_a}(z) = \int_{-\infty}^{\infty} \cos(tz) w_a(t) dt$. Here we have made use of the fact that $\cos a \cos b + \sin a \sin b = \cos(a - b)$, and the fact that

$$\int_{-\infty}^{\infty} \cos(tX_{(j)}) \sin(t|X_{(k)}|) w(t) dt = 0.$$

Hence, for the weight function $w_a(t) = e^{-a|t|}$, we have from (2.11) that

$$A_a^{(1)} = n \sum_{j=1}^n \sum_{k=1}^n u_{j,n} u_{k,n} \frac{2a}{a^2 + (|X_{(j)}| - |X_{(k)}|)^2}, \quad (2.13)$$

which is in closed form and easily calculable. Analogously, for the weight function $w_a(t) = e^{-at^2}$, we have from (2.10) that

$$A_a^{(1)} = n \sqrt{\frac{\pi}{a}} \sum_{j=1}^n \sum_{k=1}^n u_{j,n} u_{k,n} e^{-\frac{1}{4a}(|X_{(j)}| - |X_{(k)}|)^2}. \quad (2.14)$$

2.2.3 A test by Allison and Pretorius (2017) based on a characterisation by Baringhaus and Henze

We now move on to a second test proposed by Allison and Pretorius (2017). Their test is based on the following characterisation of a symmetric distribution by Baringhaus and Henze (1992): *Let X_1 and X_2 be i.i.d. random variables having a continuous distribution function F . F is symmetric around zero if and only if $|X_1|$ and $|\max(X_1, X_2)|$ have the same distribution.*

The test of Allison and Pretorius (2017) is based on the difference between the empirical characteristic functions of $|X_1|$ and $|\max(X_1, X_2)|$. Baringhaus and Henze (1992) developed Kolmogorov–Smirnov and Cramér–von Mises type tests for symmetry based on this same characterisation, but by making use of the empirical distribution functions of $|X_1|$ and $|\max(X_1, X_2)|$.

Let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ denote the order statistics of the random sample $X_1, \dots, X_n$. Define the empirical characteristic function of $|X_1|$ by

$$\chi_n(t) = \frac{1}{n} \sum_{k=1}^n e^{it|X_{(k)}|},$$

and the empirical characteristic function of the $\binom{n}{2}$ random variables $|\max(X_j, X_k)|$, $1 \leq j < k \leq n$, by

$$\phi_n(t) = \binom{n}{2}^{-1} \sum_{1 \leq j < k \leq n} e^{it|\max(X_j, X_k)|}.$$

Using this notation, the test of Allison and Pretorius (2017) is based on the test statistic

$$A_a^{(2)} = n \int_{-\infty}^{\infty} |\chi_n(t) - \phi_n(t)|^2 w_a(t) dt, \quad (2.15)$$

where $w_a(t)$ is an appropriate weight function. If $X_1, \dots, X_n$ is a random sample from a symmetric distribution, the difference between $\chi_n(t)$ and $\phi_n(t)$ should be small according to the characterisation of Baringhaus and Henze (1992). Therefore, a test for symmetry rejects for large values of $A_a^{(2)}$.

We now show that $A_a^{(2)}$ can be written in a form that is suitable for practical implementation. First observe that

$$\begin{aligned} \phi_n(t) &= \frac{2}{n(n-1)} \sum_{1 \leq j < k \leq n} e^{it|X_{(j)} \vee X_{(k)}|} \\ &= \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{k=j+1}^n e^{it|X_{(k)}|} \\ &= \frac{2}{n(n-1)} \sum_{k=2}^n \sum_{j=1}^{k-1} e^{it|X_{(k)}|} \\ &= \frac{2}{n(n-1)} \sum_{k=1}^n (k-1) e^{it|X_{(k)}|}. \end{aligned}$$

$A_a^{(2)}$ can then be written as

$$\begin{aligned}
A_a^{(2)} &= n \int_{-\infty}^{\infty} |\chi_n(t) - \phi_n(t)|^2 w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \left| \frac{1}{n} \sum_{k=1}^n e^{it|X_{(k)}|} - \left(\frac{n}{2}\right)^{-1} \sum_{k=1}^n (k-1) e^{it|X_{(k)}|} \right|^2 w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \left| \frac{1}{n} \sum_{k=1}^n [\cos(t|X_{(k)}|) + i \sin(t|X_{(k)}|)] \right. \\
&\quad \left. - \left(\frac{n}{2}\right)^{-1} \sum_{k=1}^n (k-1) [\cos(t|X_{(k)}|) + i \sin(t|X_{(k)}|)] \right|^2 w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \left\{ \left[\frac{1}{n} \sum_{k=1}^n \cos(t|X_{(k)}|) - \left(\frac{n}{2}\right)^{-1} \sum_{k=1}^n (k-1) \cos(t|X_{(k)}|) \right]^2 \right. \\
&\quad \left. + \left[\frac{1}{n} \sum_{k=1}^n \sin(t|X_{(k)}|) - \left(\frac{n}{2}\right)^{-1} \sum_{k=1}^n (k-1) \sin(t|X_{(k)}|) \right]^2 \right\} w_a(t) dt \\
&= n \int_{-\infty}^{\infty} \left\{ \left[\sum_{k=1}^n \frac{n-2k+1}{n(n-1)} \cos(t|X_{(k)}|) \right]^2 + \left[\sum_{k=1}^n \frac{n-2k+1}{n(n-1)} \sin(t|X_{(k)}|) \right]^2 \right\} w_a(t) dt \\
&= \frac{1}{n(n-1)^2} \int_{-\infty}^{\infty} \sum_{j=1}^n \sum_{k=1}^n (n-2j+1)(n-2k+1) [\cos(t|X_{(j)}|) \cos(t|X_{(k)}|) \\
&\quad + \sin(t|X_{(j)}|) \sin(t|X_{(k)}|)] w_a(t) dt \\
&= \frac{1}{n(n-1)^2} \sum_{j=1}^n \sum_{k=1}^n (n-2j+1)(n-2k+1) \int_{-\infty}^{\infty} \cos(t|X_{(j)}| - t|X_{(k)}|) w_a(t) dt \\
&= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n c_{j,n} c_{k,n} I_{w_a}(|X_{(j)}| - |X_{(k)}|),
\end{aligned}$$

where $c_{j,n} = (n-2j+1)/(n-1)$ and, as before, $I_{w_a}(z) = \int_{-\infty}^{\infty} \cos(tz) w_a(t) dt$. Using the weight function $w_a(t) = e^{-a|t|}$, the test statistic becomes

$$A_a^{(2)} = \frac{2a}{n} \sum_{j=1}^n \sum_{k=1}^n \frac{c_{j,n} c_{k,n}}{a^2 + (|X_{(j)}| - |X_{(k)}|)^2}. \quad (2.16)$$

Using the weight function $w_a(t) = e^{-at^2}$, we get

$$A_a^{(2)} = \frac{1}{n} \sqrt{\frac{\pi}{a}} \sum_{j=1}^n \sum_{k=1}^n c_{j,n} c_{k,n} e^{-\frac{1}{4a}(|X_{(j)}| - |X_{(k)}|)^2}. \quad (2.17)$$

Allison and Pretorius (2017) showed that this test is quite competitive in finite-sample settings. However, the authors did not investigate the asymptotic properties of the test. In Chapter 5 we show that this test is consistent against general alternatives, and we also derive the null distribution in the case where the centre is unspecified.

2.2.4 A probability weighted empirical characteristic function test by Meintanis et al. (2016)

To alleviate the difficult problem of choosing the weight function $w_a(t)$ appearing in L_2 -type procedures, Meintanis, Swanepoel and Allison (2014) proposed a statistically meaningful weight function based on the empirical distribution function. Instead of the usual empirical characteristic function, they suggested using the so-called probability weighted characteristic function defined as

$$\chi(t; a) := E\left(W(X; at)e^{itX}\right) = \int_{-\infty}^{\infty} W(x; at)e^{itx} dF(x),$$

for some $a > 0$, where the probability weight is given by

$$W(x; \beta) = [F(x)(1 - F(x))]^{|\beta|}, \quad \beta \in \mathbb{R}, \quad x \in \mathbb{R}.$$

As its empirical counterpart, they used

$$\chi_n(t; a) := \frac{1}{n} \sum_{j=1}^n \widehat{W}(X_j; at)e^{itX_j}, \quad t \in \mathbb{R},$$

where the estimated probability weight is defined by

$$\widehat{W}(x; \beta) = [F_n(x)(1 - F_n(x))]^{|\beta|}, \quad \beta \in \mathbb{R}, \quad x \in \mathbb{R}.$$

Here, F_n denotes the usual empirical distribution function. Note that $\chi_n(t; a)$ can also be written in trigonometric form as

$$\chi_n(t; a) = O_n(t; a) + iP_n(t; a),$$

where $O_n := n^{-1} \sum_{j=1}^n \widehat{W}(X_j; at) \cos(tX_j)$ denotes the real part and $P_n := n^{-1} \sum_{j=1}^n \widehat{W}(X_j; at) \sin(tX_j)$ denotes the imaginary part of the probability weighted empirical characteristic function.

Among some other properties of the probability weighted characteristic function which relate to the corresponding properties of the classical characteristic function, Meintanis et al. (2014) proved that if F is symmetric around zero, then $\chi(t; a)$ is real. Hence, similar to that of Hušková and Meintanis (2012), Meintanis et al. (2016) suggested using a test based on the test statistic

$$PW_a = \int_{-\infty}^{\infty} |P_n(t; a)|^2 dt.$$

Writing $\widehat{W}(x; \beta) = (\widehat{w}(x))^{|\beta|}$, with $\widehat{w}(x) = F_n(x)(1 - F_n(x))$, and $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ for the order statistics corresponding to the random sample $X_1, \dots, X_n$, Meintanis et al. (2016) showed that PW_a is equivalent to

$$PW_a = \frac{1}{2n^2} \sum_{j=1}^{n-1} \sum_{k=1}^{n-1} [s(X_{(j)} - X_{(k)}; \widehat{w}_{jk}) - s(X_{(j)} + X_{(k)}; \widehat{w}_{jk})],$$

where $\widehat{w}_{jk} = \widehat{w}(X_{(j)})\widehat{w}(X_{(k)})$, and

$$s(y; b) := s_a(y; b) = \int_{-\infty}^{\infty} b^{|t|} \cos(ty) dt = \frac{-2a \log b}{y^2 + a^2 \log^2 b}, \quad 0 < b < 1.$$

The null hypothesis of symmetry of F is rejected for large values of PW_a . The exact critical values for PW_a cannot be obtained, but can be approximated using a model-based or wild bootstrap procedure (see Section 2.6).

2.3 Other tests

2.3.1 The triples test

Davis and Quade (1978) and Randles, Fligner, Policello and Wolfe (1980) independently proposed an asymptotically distribution-free test for symmetry around an unknown median called the triples test. The test is defined as follows.

For three distinct integers i, j, k such that $1 \leq i, j, k \leq n$, define

$$f^*(Y_i, Y_j, Y_k) = \frac{1}{3} \left[\text{sign}(Y_i + Y_j - 2Y_k) + \text{sign}(Y_i + Y_k - 2Y_j) + \text{sign}(Y_j + Y_k - 2Y_i) \right].$$

We say the triple of observations (Y_i, Y_j, Y_k) forms a *right triple* if $f^*(Y_i, Y_j, Y_k) = \frac{1}{3}$, a *left triple* if $f^*(Y_i, Y_j, Y_k) = -\frac{1}{3}$, and whenever $f^*(Y_i, Y_j, Y_k) = 0$, the triple is neither left nor right.

The test statistic proposed by Randles et al. (1980) is based on the U -statistic

$$\begin{aligned} \widehat{\zeta} &= \binom{n}{3}^{-1} \sum_{i < j < k} f^*(Y_i, Y_j, Y_k) \\ &= \left[3 \binom{n}{3} \right]^{-1} \cdot [(\text{number of right triples}) - (\text{number of left triples})]. \end{aligned}$$

This U -statistic is an estimator of

$$E(\widehat{\zeta}) = P(Y_1 + Y_2 - 2Y_3 > 0) - P(Y_1 + Y_2 - 2Y_3 < 0).$$

The triples test is based on the idea that the distribution of $Y_1 + Y_2 - 2Y_3$ is the same as that of $-Y_1 - Y_2 + 2Y_3$ if the underlying distribution of the Y_i is symmetric. The statistic $\widehat{\zeta}$ will therefore be close to zero under the null hypothesis. The test of Randles et al. (1980) rejects the null hypothesis of symmetry for large absolute values of $\widehat{\zeta}$.

Randles et al. (1980) proposed using the normalised test statistic

$$DQ = \frac{\sqrt{n}\widehat{\zeta}}{\widehat{\sigma}_D}, \quad (2.18)$$

where

$$\hat{\sigma}_D = n \binom{n}{3}^{-1} \sum_{c=1}^3 \binom{3}{c} \binom{n-3}{3-c} \hat{\gamma}_c,$$

with

$$\begin{aligned} \hat{\gamma}_1 &= \frac{1}{n} \sum_{i=1}^n (\hat{f}_1^*(Y_i) - \hat{\zeta})^2, \\ \hat{\gamma}_2 &= \binom{n}{2}^{-1} \sum_{j < k} (\hat{f}_2^*(Y_j, Y_k) - \hat{\zeta})^2, \\ \hat{\gamma}_3 &= \frac{1}{9} - \hat{\zeta}^2, \\ \hat{f}_1^*(Y_i) &= \binom{n-1}{2}^{-1} \sum_{\substack{j < k \\ j \neq i \neq k}} f^*(Y_i, Y_j, Y_k), \\ \hat{f}_2^*(Y_j, Y_k) &= \frac{1}{n-2} \sum_{\substack{i=1 \\ i \neq j \neq k \\ i \neq k}} f^*(Y_i, Y_j, Y_k). \end{aligned}$$

The authors showed that, under the null hypothesis of symmetry, the statistic DQ has an asymptotic standard normal distribution, regardless of the underlying distribution of the observations (i.e. is distribution free).

In the numerical study we find that this test is quite powerful under alternatives and has empirical size very close to the nominal level. However, Randles et al. (1980) noted that this test is only consistent against asymmetric alternatives for which the median and the mean are not equal. Nevertheless, they remarked that this class of distributions for which $E(\hat{\zeta}) = 0$ is small.

2.3.2 The trimmed Wilcoxon test (Antille et al., 1982)

The original untrimmed Wilcoxon test for symmetry was introduced by Gupta (1967). Antille et al. (1982) proposed a trimmed version of the test which resulted in improved stability of level under the null hypothesis and higher efficiency under alternatives.

The trimmed Wilcoxon test is defined as follows. Let $\hat{\eta}_n$ be an estimator of the centre η and let R_i be the rank of $|Y_i - \hat{\eta}_n|$ among $|Y_1 - \hat{\eta}_n|, \dots, |Y_n - \hat{\eta}_n|$. Define

$$TW_\lambda = \frac{1}{\sqrt{n}} \sum_{i=1}^n G_\lambda \left(\frac{R_i}{2(n+1)} \right) \text{sign}(Y_i - \hat{\eta}_n), \quad (2.19)$$

where $G_\lambda(x) = \min(x, \frac{1}{2} - \lambda)$, with $0 \leq x \leq \frac{1}{2}$ and λ a parameter such that $0 \leq \lambda \leq \frac{1}{2}$, and

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0. \end{cases}$$

The test rejects the null hypothesis of symmetry for large absolute values of TW_λ . For $\lambda = 0$ this corresponds exactly to the test statistic of Gupta (1967).

Antille et al. (1982) stated that, under the null hypothesis of symmetry, the test statistic TW_λ is asymptotically normal with zero mean, and variance

$$\sigma^2(\lambda) = \frac{1}{4}(1 - 2\lambda)^2 \left(\frac{1}{12} + \frac{\lambda}{3} - \lambda^2 \right) + \frac{1}{4} \left[A - \left(\frac{1}{2} - 2\lambda^2 \right) \right]^2,$$

where

$$A = (f(\mu))^{-1} \int_{\lambda}^{1-\lambda} f(F^{-1}(u)) du,$$

with f and F denoting the density and distribution functions of the Y_i , respectively. Antille et al. (1982) noted that to ensure robustness of the size and stability of the power of the test, the variance must be bounded from above and below, respectively. They found that TW_λ performed best in this regard among three tests considered.

Clearly, the test of Antille et al. (1982) is not asymptotically distribution free and depends on the underlying distribution in a complicated way. Antille et al. (1982) simulated the critical values of this test using Monte Carlo samples from a normal distribution. They concluded from their simulation study that TW_λ was best used in the case of heavy-tailed distributions, where one should use $\lambda = 0$ or $\lambda = \frac{1}{6}$. In other cases they proposed using Finch's (1977) test based on gaps with $\lambda = 0.05$.

2.3.3 A test by Cabilio and Masaro (1996)

For a population with mean μ , variance σ^2 and median η , Yule and Kendall (1911) defined a measure of skewness as

$$\frac{\mu - \eta}{\sigma}.$$

A sample version of this measure, also discussed in Gastwirth (1971), is given by

$$\frac{\bar{X} - \hat{\eta}}{S}, \tag{2.20}$$

where $\bar{X}$, $\hat{\eta}$, and S^2 denote the sample mean, median, and variance, respectively.

Cabilio and Masaro (1996) proposed a test for symmetry based on a suitably normalised version of (2.20). Specifically, they proposed a test which rejects for large absolute values of

$$CM = \frac{\sqrt{n}(\bar{X} - \hat{\eta})}{S}. \tag{2.21}$$

Cabilio and Masaro (1996) obtained the critical values for this test by means of Monte Carlo simulation by sampling observations from the standard normal distribution. This test, however, is not distribution free and we will use the bootstrap (see Section 2.6) to approximate the critical values of the test.

2.3.4 A test by Nikitin and Ahsanullah (2015)

Based on the characterisation of symmetry by Ahsanullah (1992) stated in Section 2.2.2, Nikitin and Ahsanullah (2015) proposed a test for symmetry around zero based on the statistic

$$NA_1 = \int_{-\infty}^{\infty} (G_n(t) - H_n(t)) dQ_n(t),$$

where $G_n(t)$, $H_n(t)$ and $Q_n(t)$ are defined as

$$\begin{aligned} G_n(t) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}(|\min(X_i, X_j)| < t), \\ H_n(t) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{1}(|\max(X_i, X_j)| < t), \\ Q_n(t) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}(|X_i| < t), \end{aligned}$$

for $t \in \mathbb{R}$. It was shown that this test statistic is distribution free and that it can be written in closed, calculable form as

$$NA_1 = \frac{1}{n^3} \sum_{1 \leq i, j, k \leq n} [\mathbf{1}(|\min(X_i, X_j)| < |X_k|) - \mathbf{1}(|\max(X_i, X_j)| < |X_k|)].$$

This test statistic is two sided and hence rejects for large absolute values of NA . Nikitin and Ahsanullah (2015) showed that NA is distribution free in the case of testing for symmetry around a specified median.

Nikitin and Ahsanullah (2015) also proposed an L_2 -type norm test statistic of the form

$$NA_2 = \int_{-\infty}^{\infty} (G_n(t) - H_n(t))^2 dQ_n(t),$$

which, after some algebra, can be written in calculable form as

$$NA_2 = \frac{4}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n (n-2i+1)(n-2j+1) \mathbf{1}(\max(|X_{(i)}|, |X_{(j)}|) < |X_{(k)}|).$$

A third possible test statistic based on the same characterisation was proposed and studied by Allison and Pretorius (2017). The test statistic is given by

$$NA_3 = \int_{-\infty}^{\infty} (F_n^{\max}(t) - F_n^{\min}(t))^2 dQ_n(t),$$

where

$$F_n^{\max}(t) = (F_n(t))^n - (F_n(-t))^n,$$

and

$$F_n^{\min}(t) = (1 - F_n(-t))^n - (1 - F_n(t))^n,$$

with $F_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}(X_i \leq x)$ being the empirical distribution function of $X_1, \dots, X_n$.

In the numerical study of Allison and Pretorius (2017) it was found that tests based on NA_1 and NA_3 do not consistently exhibit power comparable to existing tests. The test based on NA_2 seems to have more favourable finite-sample power properties in comparison with the classical Kolmogorov–Smirnov test, but does not compare well with more modern tests.

2.3.5 A test by Zardasht (2018)

Zardasht (2018) proposed a bootstrap test for symmetry based on sample quantiles. Assuming that F is continuous and strictly increasing, Zardasht (2018) showed that $F^{-1}(1-p) + F^{-1}(p) - 2F^{-1}(0.5) = 0$ under the null hypothesis of symmetry and hence the absolute value of this difference can be used as a measure of deviation of F from symmetry. The test is constructed as follows.

Let $X_1, \dots, X_n$ be a random sample from an unknown continuous distribution F and let $\hat{x}_p = \inf\{x \in \mathbb{R} : F_n(x) \geq p\}$, where F_n is the empirical distribution function based on the sample $X_1, \dots, X_n$. That is, $\hat{x}_p$ is the p -level empirical percentile of the sample $X_1, \dots, X_n$. The test statistic proposed by Zardasht (2018) is then

$$Z = \max_{0 < p < 0.5} |\hat{x}_{1-p} + \hat{x}_p - 2\hat{x}_{0.5}|, \quad (2.22)$$

where large values of Z lead to rejection of the null hypothesis of symmetry. Since this test is nonparametric but not distribution free, Zardasht (2018) used the bootstrap to obtain approximate critical values of this test.

2.4 Additional tests that did not form part of this study

Some tests that did not form part of this study, but which are listed here for completeness, include the following:

- A rank-based test by Bhattacharya, Gastwirth and Wright (1982).
- Tests based on the distribution function by Baringhaus and Henze (1992) (Kolmogorov–Smirnov and Cramér–von Mises type).
- A sign test by Gastwirth (1971) and the hybrid test by Modarres and Gastwirth (1998) which is a percentile-modified two-sample Wilcoxon test.
- An asymptotically nonparametric test by Gupta (1967) based on a sample skewness coefficient and a test by Staudte (2014) based on a variance stabilising transformation of a skewness coefficient.

- Gaps tests by Finch (1977) and Antille et al. (1982).
- Tests based on a runs statistic by McWilliams (1990) and Modarres and Gastwirth (1996).
- A test by Nabeya (1987) based on Aki's (1987) nonparametric test which uses a Kolmogorov–Smirnov type distance.

2.5 Regression estimators

Our main theoretical results of Chapters 3 and 4 are derived for the general case where the goal is to test whether the distribution of the error in a linear regression model is symmetric around zero. Existing tests for symmetry have been studied in this context under several different sets of assumptions on the regressors and the errors (see, e.g., Koul, 2002; Neumeyer et al., 2005; Hušková and Meintanis, 2012). In this section we give an exposition of typical assumptions in the linear regression setup, which will also partly serve as motivation for the assumptions used in Chapters 3 and 4.

We first define some notation. Let $(Y_1; x_1), \dots, (Y_n; x_n)$ be a random sample generated by the model

$$Y_i = x_i^\top \beta + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.23)$$

where $\beta \in \mathbb{R}^p$ is an unknown parameter vector, $x_i = (x_{i1}, \dots, x_{ip})^\top \in \mathbb{R}^p$ are fixed (nonrandom) regressors, and $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d., zero-mean random errors with uniformly continuous density $f > 0$. Denote the $n \times p$ design matrix, which we assume to have rank $p \leq n$, by

$$\mathbf{X}_n = \begin{bmatrix} x_1^\top \\ x_2^\top \\ \vdots \\ x_n^\top \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & & \ddots & \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}.$$

Also define

$$\mathbf{D}_n = \mathbf{X}_n^\top \mathbf{X}_n = \sum_{k=1}^n x_k x_k^\top.$$

In coming chapters, our main assumptions on the regressors are the following:

(R1a) $\max_{1 \leq i \leq n, 1 \leq v \leq p} |x_{iv}| = o(n^{1/2}).$

(R1b) $n^{-1} \mathbf{D}_n \rightarrow \mathbf{D}$ for some positive definite matrix $\mathbf{D}$.

Assumptions (R1a) and (R1b), along with $E(\varepsilon_i^2) < \infty$, are satisfied by reasonable designs and are also sufficient for asymptotic normality of the ordinary least-squares

estimators of β (see van der Vaart, 2000, p. 21). Hušková and Meintanis (2012) studied their symmetry test under Assumptions (R1a) and (R1b), among others.

An alternative basic condition is the so-called Noether condition used by Koul (2002) and Neumeyer et al. (2005):

$$(R3) \max_{i=1,\dots,n} \mathbf{x}_i^\top (\mathbf{X}_n^\top \mathbf{X}_n)^{-1} \mathbf{x}_i = o(1).$$

It is clear that if Assumptions (R1a) and (R1b) are satisfied, then Assumption (R3) is also satisfied. Many useful results derived by Koul (2002), such as asymptotic representations of the regression parameter estimators, are based on Assumption (R3).

To help motivate our set of conditions used in Chapters 3 and 4, it is instructive to consider the situation where the parameter vector is estimated using an M -estimator (Huber, 1981). This is a general class of estimators which includes the ordinary least-squares estimators and maximum-likelihood estimators as special cases. Let ψ be a nondecreasing function of bounded total variation satisfying

$$E(\psi(\varepsilon_1)) = \int \psi(t) dF(t) = 0, \quad E(\psi^2(\varepsilon_1)) = \int \psi^2(t) dF(t) < \infty \quad \text{and} \quad \int f(t) d\psi(t) > 0, \quad (2.24)$$

where F is the distribution function of the regression errors. The function ψ is often referred to as the *score function*. The M -estimator $\hat{\beta}_n$ of β is then defined as a solution to the system of equations

$$\sum_{i=1}^n \mathbf{x}_i \psi(Y_i - \mathbf{x}_i^\top \hat{\beta}_n) = 0.$$

Under Assumption (R3) and the moment assumptions in (2.24), we see from (4.2.7) of Koul (2002, p. 103) that the M -estimator $\hat{\beta}_n$ has the asymptotic representation

$$\mathbf{D}_n^{1/2} (\hat{\beta}_n - \beta) = d_{\psi,f} \mathbf{D}_n^{-1/2} \sum_{k=1}^n \mathbf{x}_k \psi(\varepsilon_k) + o_P(1), \quad (2.25)$$

as $n \rightarrow \infty$, where

$$d_{\psi,f} = \frac{1}{\int f(t) d\psi(t)}.$$

Using Assumption (R1b) and multiplying both sides of (2.25) by $\mathbf{D}_n^{1/2}$, we get

$$\hat{\beta}_n - \beta = d_{\psi,f} \mathbf{D}_n^{-1} \sum_{k=1}^n \mathbf{x}_k \psi(\varepsilon_k) + o_P\left(\frac{1}{\sqrt{n}}\right). \quad (2.26)$$

Now, using this expansion, consider the quantity

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta) &= d_{\psi,f} \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{k=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \mathbf{x}_k \psi(\varepsilon_k) + o_P(1) \\ &= d_{\psi,f} \frac{1}{\sqrt{n}} \sum_{k=1}^n b_k \psi(\varepsilon_k) + o_P(1), \end{aligned} \quad (2.27)$$

as $n \rightarrow \infty$, where

$$b_k = \sum_{i=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \mathbf{x}_k.$$

Clearly, the asymptotic behaviour of the random variable $n^{-1} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta)$ depends on the design matrix $\mathbf{X}_n$ through the b_k .

Suppose now that the regression model has an intercept term, i.e. we may take $x_{i1} = 1$ for all $i = 1, \dots, n$. It can then be shown (see below) that

$$\frac{1}{n} \sum_{k=1}^n b_k = \frac{1}{n} \sum_{k=1}^n b_k^2 = 1 \quad (2.28)$$

for all $n \geq p$. Now, using (2.27), write

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta) = d_{\psi,f} \frac{1}{\sqrt{n}} \sum_{k=1}^n \psi(\varepsilon_k) + R_n + o_P(1), \quad (2.29)$$

where

$$R_n = d_{\psi,f} \frac{1}{\sqrt{n}} \sum_{k=1}^n (b_k - 1) \psi(\varepsilon_k).$$

Since $E(\psi(\varepsilon_1)) = 0$, we also have $E(R_n) = 0$. Moreover, by independence of the ε_k ,

$$E \left[\frac{1}{\sqrt{n}} \sum_{k=1}^n (b_k - 1) \psi(\varepsilon_k) \right]^2 = \frac{1}{n} \sum_{k=1}^n (b_k - 1)^2 E[\psi^2(\varepsilon_1)] = 0.$$

This means that R_n is degenerate at zero and (2.29) becomes

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta) = d_{\psi,f} \frac{1}{\sqrt{n}} \sum_{k=1}^n \psi(\varepsilon_k) + o_P(1). \quad (2.30)$$

From this expansion one sees that the limit behaviour of $n^{-1} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta)$ does not depend on the regressors. We have shown that this expansion is typically valid in the setting of a linear model with an intercept term, but there are many other general situations where such an expansion is valid (see Hušková and Meintanis, 2007). Because of this, we base our asymptotic analyses in the expansion in Chapters 3 and 4 on estimators allowing the representation in (2.30).

Proof of (2.28)

Assume that the linear model in (2.23) has an intercept term. Specifically we assume that the first entry of $\mathbf{x}_i$, $i = 1, \dots, n$, is equal to 1. Define

$$\tilde{\mathbf{x}}_i = (1, x_{i2}, x_{i3}, \dots, x_{ip}), \quad i = 1, \dots, n.$$

Then we may express $\mathbf{D}_n$ as

$$\mathbf{D}_n = \mathbf{X}_n^\top \mathbf{X}_n = \begin{bmatrix} n & \sum_i \tilde{\mathbf{x}}_i^\top \\ \sum_i \tilde{\mathbf{x}}_i & \sum_i \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^\top \end{bmatrix}.$$

Using A.9.1 on p. 466 of Seber and Lee (2012), we get that

$$\mathbf{D}_n^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{1}{n^2} \sum_i \tilde{\mathbf{x}}_i^\top \mathbf{A}_n^{-1} \sum_j \tilde{\mathbf{x}}_j & -\frac{1}{n} \sum_i \tilde{\mathbf{x}}_i^\top \mathbf{A}_n^{-1} \\ -\frac{1}{n} \mathbf{A}_n^{-1} \sum_i \tilde{\mathbf{x}}_i & \mathbf{A}_n^{-1} \end{bmatrix},$$

where

$$\mathbf{A}_n = \sum_{i=1}^n \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^\top - \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{x}}_i \sum_{j=1}^n \tilde{\mathbf{x}}_j^\top.$$

Straightforward matrix multiplication shows that

$$\sum_k \mathbf{x}_k^\top \mathbf{D}_n^{-1} = \left[n, \sum_k \tilde{\mathbf{x}}_k^\top \right] \mathbf{D}_n^{-1} = \left[1, -\sum_i \tilde{\mathbf{x}}_i^\top \mathbf{A}_n^{-1} + \sum_k \tilde{\mathbf{x}}_k^\top \mathbf{A}_n^{-1} \right] = (1, 0)$$

and

$$\sum_k \mathbf{x}_k^\top \mathbf{D}_n^{-1} \sum_\ell \mathbf{x}_\ell = (1, 0) \begin{bmatrix} n \\ \sum_\ell \tilde{\mathbf{x}}_\ell \end{bmatrix} = n.$$

Hence,

$$\frac{1}{n} \sum_{k=1}^n b_k = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \mathbf{x}_k = 1.$$

The result in (2.28) now follows easily by noting that

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n b_k^2 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \mathbf{x}_k \mathbf{x}_j^\top \mathbf{D}_n^{-1} \mathbf{x}_k \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \mathbf{x}_k \mathbf{x}_k^\top \mathbf{D}_n^{-1} \mathbf{x}_j \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \left(\sum_{k=1}^n \mathbf{x}_k \mathbf{x}_k^\top \right) \mathbf{D}_n^{-1} \mathbf{x}_j \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \mathbf{D}_n \mathbf{D}_n^{-1} \mathbf{x}_j \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{x}_i^\top \mathbf{D}_n^{-1} \mathbf{x}_j \\ &= \frac{1}{n} \sum_{k=1}^n b_k. \end{aligned}$$

2.6 Bootstrap hypothesis testing

The test statistics used in goodness-of-fit tests for symmetry around an unknown median are typically not distribution free and have limit null distributions that typically depend on the unknown underlying distribution that generated the observations, as well as the estimator used to estimate the unknown location parameter (see, e.g., Koziol, 1980; Koul, 2002). In some cases (e.g. Neumeyer et al., 2005; Hušková and Meintanis, 2012) these limit null distributions may depend on the

mentioned unknowns in a complicated way and may be difficult to approximate directly.

An alternative approach is to use the bootstrap to approximate the finite-sample critical values of the tests. We shall specifically consider two suitable bootstrap approaches:

- (i) a model-based bootstrap (Section 2.6.1) appropriate under the assumption that the model errors are identically distributed, and
- (ii) a wild bootstrap approach (Section 2.6.2) which works under more general conditions (e.g. if the model errors are heteroskedastic).

Another bootstrap procedure developed by Freedman et al. (1981) and improved by Mammen (1993) is the so-called *pairs bootstrap* which involves directly resampling with replacement from the paired data. It has been shown by Horowitz (2001) that inference based on the pairs bootstrap is not always accurate. Although the ideas proposed by Mammen (1993) and Flachaire (1999) can highly improve the accuracy of this procedure, Flachaire (1999) found that significant size distortions still remain. For these reasons, and the fact that the two above-mentioned approaches are popular approaches when testing for symmetry, we do not pursue the use of the pairs bootstrap in this study.

We consider the general case where we observe $Y_1, \dots, Y_n$ generated by the linear model

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.31)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown regression parameters, $\mathbf{x}_i \in \mathbb{R}^p$ is a vector of regressors, and $\varepsilon_1, \dots, \varepsilon_n$ are independent zero-mean random variables with distribution function F . Seeing that the errors in the above model are unobservable, we base the test on the residuals

$$\hat{\varepsilon}_i = Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n,$$

where $\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}_n(Y_1, \dots, Y_n, \mathbf{x}_1, \dots, \mathbf{x}_n)$ is an estimator of $\boldsymbol{\beta}$. At this stage we do not specify a specific choice for $\hat{\boldsymbol{\beta}}_n$.

Suppose that we would like to apply a test for symmetry which rejects for large values of the test statistic

$$T_n = T(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n).$$

To obtain a test with a Type I error probability of α , we need to find a critical value c_α satisfying

$$\mathbb{P}_{H_0}(T_n > c_\alpha) = \alpha, \quad (2.32)$$

where $\mathbb{P}_{H_0}$ denotes the probability measure under the null hypothesis.

In order to apply the bootstrap correctly, we need to calculate a bootstrap version of the statistic T_n while enforcing the null hypothesis during resampling. This means that the bootstrap errors ε_i^* , $i = 1, \dots, n$, need to be resampled in such a way that the null hypothesis of symmetry around zero holds. In the following two subsections we will discuss separately the two approaches mentioned in (i) and (ii) above.

A general bootstrap scheme for estimating the critical value c_α is the following:

1. Obtain ε_i^* , $i = 1, \dots, n$, according to approach (i) or (ii).
2. Calculate the bootstrap responses Y_i^* as

$$Y_i^* = \mathbf{x}_i^\top \hat{\beta}_n + \varepsilon_i^*, \quad i = 1, \dots, n.$$

3. Let $\hat{\beta}_n^* = \beta_n(Y_1^*, \dots, Y_n^*, \mathbf{x}_1, \dots, \mathbf{x}_n)$ be the bootstrap analogue of $\hat{\beta}_n$, i.e. $\hat{\beta}_n^*$ is the same function of $(Y_1^*, \dots, Y_n^*, \mathbf{x}_1, \dots, \mathbf{x}_n)$ as $\hat{\beta}_n$ is of $(Y_1, \dots, Y_n, \mathbf{x}_1, \dots, \mathbf{x}_n)$. The bootstrap residuals are then given by

$$\hat{\varepsilon}_i^* = Y_i^* - \mathbf{x}_i^\top \hat{\beta}_n^*.$$

4. Based on these residuals, we calculate the bootstrap analogue of T_n as

$$T_n^* = T(\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*).$$

A bootstrap estimator of the critical value c_α defined in (2.32) is given by $\hat{c}_\alpha$ satisfying

$$P_{H_0}^*(T_n^* > \hat{c}_\alpha) = \alpha,$$

where $P_{H_0}^*$ denotes the bootstrap probability measure under the null hypothesis. Implementation of the above bootstrap procedure typically involves repeating steps 1–4 above a large number of times (say B) to obtain B replications of the bootstrap statistic T_n^* , say $T_n^*(1), \dots, T_n^*(B)$. The estimated critical value $\hat{c}_\alpha$ may then be approximated by finding the $(1 - \alpha)$ -level empirical quantile of $T_n^*(1), \dots, T_n^*(B)$.

2.6.1 Model-based bootstrap

In this approach we assume that the model errors ε_i are independent and identically distributed. An example of a situation where such an assumption is reasonable, is in the case where we want to test whether i.i.d. observations are symmetric around an *unknown* median β . This is a special case of the model in (2.31) where there is only an intercept term and no regressors, i.e. $p = 1$ and $x_i = 1$ for all $i = 1, \dots, n$ so that the model in (2.31) reduces to the mean model $Y_i = \beta + \varepsilon_i$.

In order to enforce the null hypothesis of symmetry around zero, we sample the bootstrap errors ε_i^* , $i = 1, \dots, n$, with replacement from the set of symmetrised observations

$$\widehat{\varepsilon}'_1, \dots, \widehat{\varepsilon}'_n, -\widehat{\varepsilon}'_1, \dots, -\widehat{\varepsilon}'_n,$$

where $\widehat{\varepsilon}'_i$, $i = 1, \dots, n$, are the centred residuals defined by

$$\widehat{\varepsilon}'_i = \widehat{\varepsilon}_i - \frac{1}{n} \sum_{j=1}^n \widehat{\varepsilon}_j. \quad (2.33)$$

Clearly, the bootstrap distribution of ε_i^* is given by

$$P^*(\varepsilon_i^* \leq t) = \frac{1}{2n} \sum_{i=1}^n [\mathbf{1}(\widehat{\varepsilon}'_i \leq t) + \mathbf{1}(-\widehat{\varepsilon}'_i \leq t)] = \frac{1}{2} [\widehat{F}_n(t) + 1 - \widehat{F}_n(-t-)], \quad (2.34)$$

where $\widehat{F}_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}(\widehat{\varepsilon}'_i \leq t)$. It follows readily that $P^*(\varepsilon_i^* \leq t) = 1 - P^*(\varepsilon_i^* \leq -t-)$, so that the bootstrap distribution of ε_i^* is truly symmetric around zero.

We note here that the distribution function in (2.34) coincides with the non-parametric maximum-likelihood estimator of F under the assumption that F is symmetric around zero (Modarres, 2002). This estimator, also referred to as the symmetrised estimator, was also studied earlier in the context of testing for symmetry of errors in an intercept-only model by Schuster (1973); also see Schuster and Barker (1987). Theoretical justification of the resampling procedure used in the latter two references was given by Arcones and Giné (1991).

2.6.2 Wild bootstrap

In some cases the assumption of an i.i.d. error structure as assumed in Section 2.6.1 is unreasonable, for example in the case of a heteroskedastic error structure. Following a suggestion of Beran (1986) and Wu (1986), Liu (1988) proposed the so-called *wild bootstrap* and established its asymptotic validity under certain conditions. Probably the most popular version of the wild bootstrap is that proposed by Mammen (1993), which involves resampling in such a way that the first three moments (i.e. also skewness) of the bootstrap distribution coincides with those of the true underlying distribution. However, in our setting we need to resample while enforcing the null hypothesis of symmetry.

The bootstrap procedure we describe here corresponds to the bootstrap procedures used by other authors in similar settings, including Neumeyer et al. (2005) and Hušková and Meintanis (2012). The procedure involves randomly multiplying the centred residuals by -1 or $+1$ (with equal probability) instead of resampling directly from the residuals. Apart from being asymptotically justified by Mammen (1993), extensive numerical studies by Flachaire (2005) and Davidson and

Flachaire (2008) showed the good finite-sample properties of this version of the wild bootstrap, including the case where the regression errors are heteroskedastic.

The idea is simple, and to generate the bootstrap errors ε_i^* , $i = 1, \dots, n$, we proceed as follows. First, independently from all other random variables, generate n independent observations $\xi_1, \dots, \xi_m$ from the two-point distribution

$$\xi_i = \begin{cases} +1 & \text{with probability } \frac{1}{2}, \\ -1 & \text{with probability } \frac{1}{2}. \end{cases}$$

The bootstrap errors are then given by

$$\varepsilon_i^* = \xi_i \hat{\varepsilon}_i', \quad i = 1, \dots, n$$

where $\hat{\varepsilon}_i'$, $i = 1, \dots, n$, are the centred residuals defined in (2.33).

2.7 Some symmetric and skew probability distributions

In this study we complement the theoretical investigation with numerical studies which will show that the newly-proposed tests have desirable finite-sample properties. To evaluate both size and power properties, we consider the distributions listed and defined in Table 2.1, where the top section of the table contains symmetric distributions and the bottom section some skew distributions. For the skew distributions, the parameter ϑ controls the degree of skewness. The coefficients reported in the last column are Pearson's moment coefficient of skewness. For any random variable X with density function f , this coefficient is defined as the third standardised moment

$$\mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \int_{-\infty}^{\infty} \left(\frac{x - \mu}{\sigma} \right)^3 f(x) dx,$$

where

$$\mu = \int_{-\infty}^{\infty} x f(x) dx \quad \text{and} \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Figures 2.1 and 2.2 show the densities of the considered symmetric and skew distributions, but shifted and scaled to have zero mean and unit variance. That is, for any distribution f defined in Table 2.1, the density f_s shown in Figure 2.1 or 2.2 is given by

$$f_s(x) = \sigma f(\sigma x + \mu),$$

with the mean μ and variance σ^2 as defined above.

To sample observations from the normal, Student t, uniform, logistic and χ^2 distributions, we used the built-in R functions `rnorm`, `rt`, `runif`, `rlogis` and `rchisq` functions, respectively. For the remaining distributions one can obtain samples as follows:

Table 2.1: List of considered distributions with their corresponding density functions, range and skewness coefficients.

Distribution	Notation	Density function	Parameter range	Support	Skewness coefficient
Standard normal [†]	N	$f(x) = \phi(x)$		$x \in \mathbb{R}$	0
Student's t	t_ϑ	$f(x) = \frac{\Gamma(\frac{\vartheta+1}{2})}{\sqrt{\vartheta\pi}\Gamma(\frac{\vartheta}{2})} \left(1 + \frac{x^2}{\vartheta}\right)^{-(\vartheta+1)/2}$	$\vartheta > 0$	$x \in \mathbb{R}$	0
Uniform(0,1)	U	$f(x) = 1$		$x \in (0, 1)$	0
Logistic	L	$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$		$x \in \mathbb{R}$	0
Normal mixture	NM $_\vartheta$	$f(x) = \frac{1}{2}\phi(x) + \frac{1}{2\sqrt{\vartheta}}\phi\left(\frac{x}{\sqrt{\vartheta}}\right)$	$\vartheta > 0$	$x \in \mathbb{R}$	0
Asymmetric Laplace	AL $_\vartheta$	$f(x) = \begin{cases} \frac{\vartheta}{1+\vartheta^2}e^{-\vartheta x} & \text{if } x \geq 0, \\ \frac{\vartheta}{1+\vartheta^2}e^{x/\vartheta} & \text{if } x < 0. \end{cases}$	$\vartheta > 0$	$x \in \mathbb{R}$	$\frac{2(1-\vartheta^6)}{(1+\vartheta^4)^{3/2}}$
Chi-square	χ^2_ϑ	$f(x) = \frac{x^{\vartheta/2-1}e^{-x/2}}{2^{\vartheta/2}\Gamma(\vartheta/2)}$	$\vartheta \in \mathbb{N}$	$x > 0$	$2\sqrt{\frac{2}{\vartheta}}$
Skew normal [†]	SN $_\vartheta$	$f(x) = 2\phi(x)\Phi(\vartheta x)$	$\vartheta \in \mathbb{R}$	$x \in \mathbb{R}$	$\frac{\sqrt{2}(4-\pi)\vartheta^3}{(\pi(1+\vartheta^2)-2\vartheta^2)^{3/2}}$
Tukey's g	TU $_\vartheta$	$f(x) = \begin{cases} \phi\left(\frac{\ln(\vartheta x+1)}{\vartheta}\right)\left(\frac{1}{\vartheta x+1}\right) & \text{if } \vartheta \neq 0, \\ \phi(x) & \text{if } \vartheta = 0. \end{cases}$	$\vartheta \in \mathbb{R}$	$\begin{cases} x > -\frac{1}{\vartheta} & \text{if } \vartheta \neq 0, \\ x \in \mathbb{R} & \text{if } \vartheta = 0. \end{cases}$	$\sqrt{3e^{2\vartheta^2} + e^{3\vartheta^2} - 4}$

[†] The functions ϕ and Φ respectively denote the standard normal density and distribution functions, i.e. $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $\Phi(x) = \int_{-\infty}^x \phi(t)dt$.

- *normal mixture* distribution — sample from a $N(0, 1)$ or a $N(0, \vartheta^2)$ distribution with equal probability;
- if Z_1 and Z_2 are independent $N(0, 1)$ random variables, then the random variable

$$\frac{\vartheta|Z_1| + Z_2}{\sqrt{1 + \vartheta^2}}$$

has a *skew normal* distribution with parameter ϑ ;

- if Z has a $N(0, 1)$ distribution, then the random variable

$$\frac{1}{\vartheta} \left(\exp^{\vartheta Z_1} - 1 \right)$$

has *Tukey's g* distribution with parameter ϑ ;

- if V_1 and V_2 are independent standard exponentially distributed random variables, then the random variable

$$\frac{V_1}{\vartheta} - \vartheta V_2$$

has an *asymmetric Laplace* distribution with parameter ϑ .

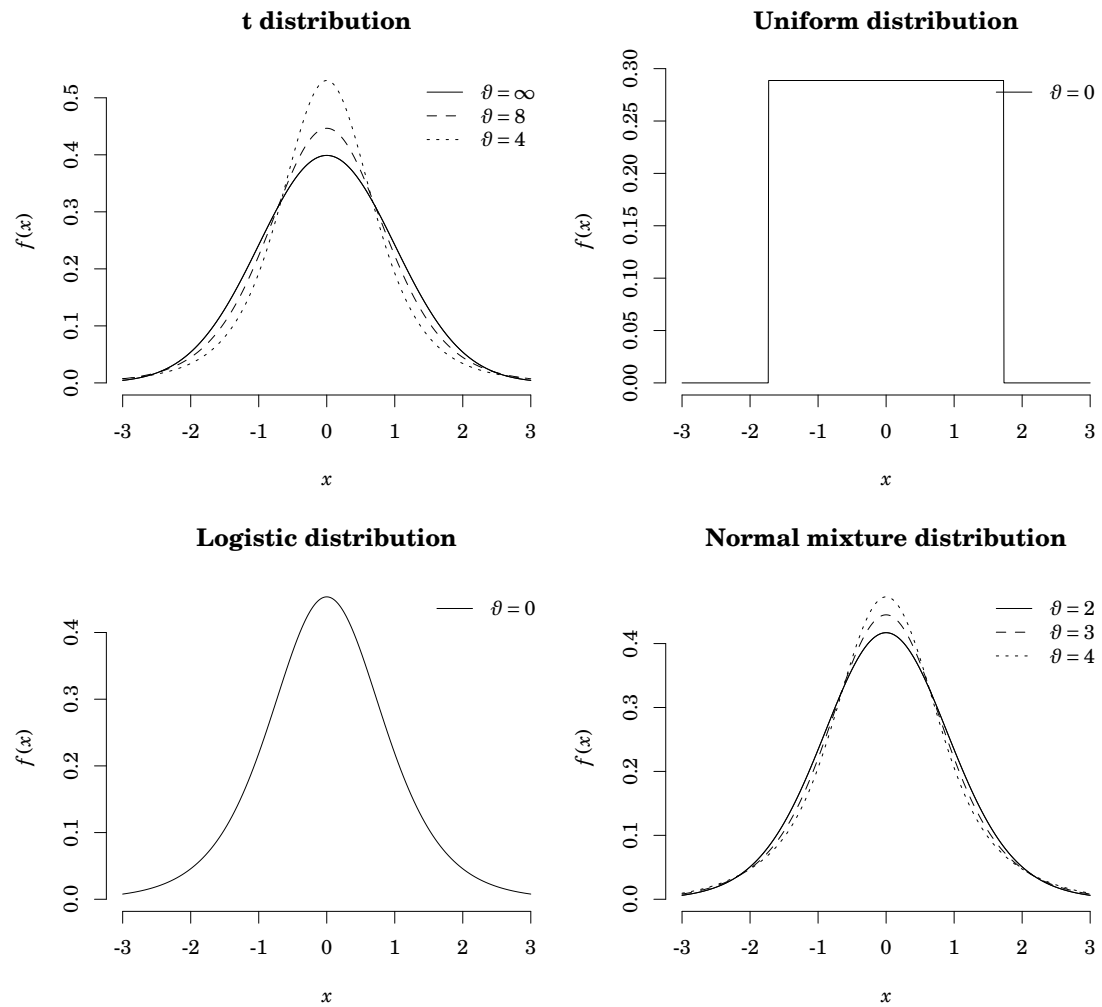


Figure 2.1: Densities of the considered symmetric distributions with various choices of the parameter ϑ . Note that the plots shown here are the densities of *standardised* versions of the distributions defined in Table 2.1, i.e. scaled and shifted to have mean 0 and variance 1.

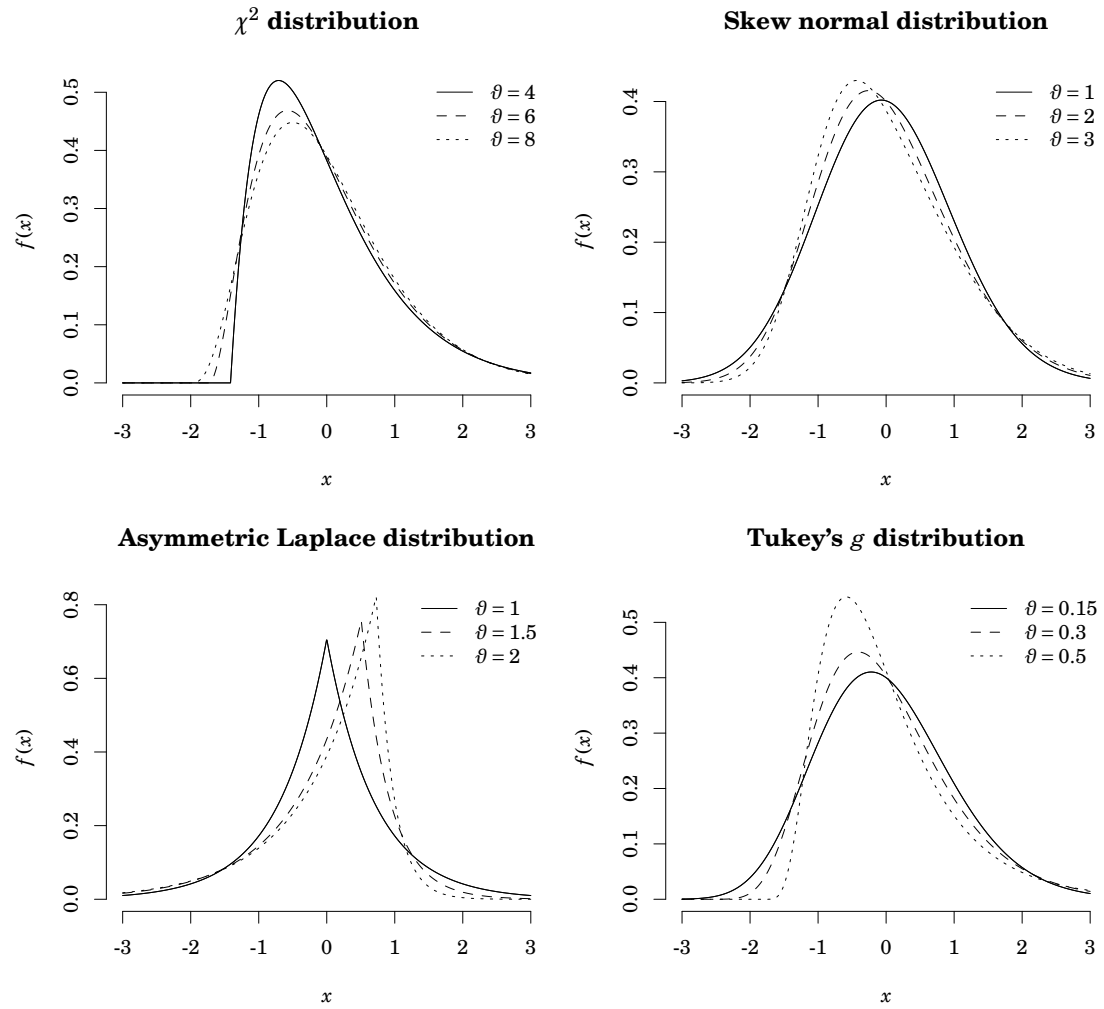


Figure 2.2: Densities of the considered skew distributions with various choices of the parameter θ . Note that the plots shown here are the densities of *standardised* versions of the distributions defined in Table 2.1, i.e. scaled and shifted to have mean 0 and variance 1.

Chapter 3

New tests based on the integrated empirical process

Denoting the median of the continuous distribution function F by η , we would like to test the hypothesis

$$H_0 : F(\eta + t) + F(\eta - t) = 1 \quad \text{for all } t \geq 0, \quad (3.1)$$

against general alternatives. Our first set of new tests is based on an interesting property of symmetric distributions pointed out by Silver and Costa (1997), which involves symmetric integration of the distribution function around its median. They showed that, under H_0 , the relation $\int_{-t}^t F(\eta + x) dx = t$ holds for all $t \geq 0$. Indeed, by the fundamental theorem of calculus it follows directly that the property of symmetry stated in (3.1) holds *if and only if*

$$V(t) := \int_{-t}^t \left[F(\eta + x) - \frac{1}{2} \right] dx = 0 \quad \text{for all } t \geq 0, \quad (3.2)$$

assuming that F is continuous. Although this result seems obvious, it does not seem to be widely known and such a restatement of (3.1) allows the introduction of goodness-of-fit tests with certain attractive properties. We mention that this characterisation was also pointed out by Malan (2017, p. 52), but the development of goodness-of-fit tests for symmetry was not investigated.

Recently, Neumeyer et al. (2005) revisited the classical Kolmogorov–Smirnov and the Cramér–von Mises tests and studied their theoretical properties in the general context of testing whether the error distribution in a linear regression model is symmetric. In the same context, as well as in the case where the parametric form of the regression function is unknown, Hušková and Meintanis (2012) proposed and studied a test based on the empirical characteristic function. Their test is based on the well-known fact that the imaginary component of the characteristic function of

a random variable is zero if and only if the random variable has a distribution symmetric around zero (cf. Feuerverger and Mureika, 1977; Koutrouvelis, 1985; Ghosh and Ruymgaart, 1992). These more modern tests have proven to be quite powerful against asymmetric alternatives (also see Henze et al., 2003). A drawback of these tests, and many other existing tests (e.g. Fan and Gencay, 1995; Ahmad and Li, 1997), is that they require the user to make a choice of a tuning parameter. Numerical results from these papers suggested that the performance of these tests could be very sensitive to the choice of such parameters (this is also noted by Neumeyer et al., 2005), which complicates application of these tests in practical settings. Moreover, choosing these parameters in an optimal, data-dependent way is by no means a straightforward task and remains an active area of research (see, e.g., Tenreiro, 2019).

In view of this, the new tests proposed in this chapter are attractive from a practical perspective in that they require no choice of any unknown parameters. In addition, from our numerical study the newly-proposed tests seem to compare favourably with even the most modern tests, both in that the new tests preserve the nominal level quite well and are competitive in terms of power.

We finally mention that our tests are related to tests based on the so-called *integrated empirical process* originally proposed by Henze and Nikitin (2000); also see Klar (2001); Henze and Nikitin (2003). In these papers, the authors defined the integrated empirical distribution function as the Lebesgue–Stieltjes integral of the empirical distribution function with respect to the hypothesised distribution function (e.g. the exponential or normal distribution function). However, our tests differ in that the underlying distribution under the null hypothesis is unspecified and one needs to resort to integrating with respect to some other measure, such as the empirical distribution function.

The remainder of this chapter is structured as follows. In Section 3.1 we introduce the new tests and derive expressions which are suitable for easy implementation. In that section we also study the large-sample behaviour of the new tests in the case where we would like to test whether the observations were generated independently by a common distribution with unknown median. Specifically, we derive their limit null distributions and also formalise consistency against nonsymmetric alternatives. The theoretical results are then generalised in Section 3.2 to the case where the objective is to test whether the linear regression error may be assumed to have a symmetric distribution. In Section 3.3 we show that the newly-proposed tests have favourable finite-sample properties and compare favourably against existing tests.

3.1 The case of independent observations

Suppose we have at our disposal a random sample $Y_1, \dots, Y_n$ of realisations of a random variable Y . In light of the discussion in the introduction, we define an empirical version of the integral in (3.2) by

$$V_n(t) = \sqrt{n} \int_{-t}^t \left[F_n(\hat{\eta}_n + x) - \frac{1}{2} \right] dx, \quad (3.3)$$

where F_n and $\hat{\eta}_n$ are appropriate estimators of F and η , respectively. Under the hypothesis in (3.1) it is expected that $V_n(t)$ will be close to zero for all $t \geq 0$. This suggests using a test that rejects for large values of the Kolmogorov–Smirnov and Cramér–von Mises type statistics

$$\hat{U}_n^{(I)} = \sup_{t \geq 0} |V_n(t)| \quad \text{and} \quad \hat{W}_n^{(I)} = \int_0^\infty V_n^2(x) d\hat{Q}_n(x), \quad (3.4)$$

where $\hat{Q}_n(x) = n^{-1} \sum_{i=1}^n \mathbf{1}(|Y_i - \hat{\eta}_n| \leq x)$.

As an estimator of F we use the usual empirical distribution function

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Y_i \leq x).$$

In this case the process $V_n(\cdot)$ may be expressed as

$$V_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-t}^t \left[\mathbf{1}(Y_i - \hat{\eta}_n \leq x) - \frac{1}{2} \right] dx = \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\hat{\eta}_n - Y_i) (|\hat{\eta}_n - Y_i| \wedge t). \quad (3.5)$$

To see this, first observe that, if $Y_i - \hat{\eta}_n \leq -t$, then

$$\mathbf{1}(Y_i - \hat{\eta}_n \leq x) - \frac{1}{2} = \frac{1}{2}, \quad \forall x \in [-t, t],$$

so that

$$\int_{-t}^t \left[\mathbf{1}(Y_i - \hat{\eta}_n \leq x) - \frac{1}{2} \right] dx = t \quad \text{whenever} \quad Y_i - \hat{\eta}_n \leq -t.$$

Similarly,

$$\int_{-t}^t \left[\mathbf{1}(Y_i - \hat{\eta}_n \leq x) - \frac{1}{2} \right] dx = -t \quad \text{whenever} \quad Y_i - \hat{\eta}_n > t,$$

and

$$\int_{-t}^t \left[\mathbf{1}(Y_i - \hat{\eta}_n \leq x) - \frac{1}{2} \right] dx = \hat{\eta}_n - Y_i \quad \text{whenever} \quad -t < Y_i - \hat{\eta}_n \leq t.$$

Hence,

$$\begin{aligned} \int_{-t}^t \left[\mathbf{1}(Y_i - \hat{\eta}_n \leq x) - \frac{1}{2} \right] dx &= \begin{cases} t & \text{if } Y_i - \hat{\eta}_n \leq -t, \\ \hat{\eta}_n - Y_i & \text{if } -t < Y_i - \hat{\eta}_n \leq t, \\ -t & \text{if } Y_i - \hat{\eta}_n > t. \end{cases} \\ &= \text{sign}(\hat{\eta}_n - Y_i) (|\hat{\eta}_n - Y_i| \wedge t), \end{aligned}$$

from which (3.5) follows.

We note here that the process V_n has certain desirable properties. Firstly, since $V_n(t)$ is defined as an integral, it is continuous for all values of $t \geq 0$. In fact, it can easily be shown that V_n is uniformly continuous, which facilitates theoretical analysis. Secondly, $|V_n(t)|$ is invariant under reflecting all the data at the sample median, i.e. the value of $|V_n(t)|$ remains unaffected when replacing $Y_1, \dots, Y_n$ with $-Y_1, \dots, -Y_n$. This property is desirable when testing for symmetry and therefore symmetry tests are usually constructed with this in mind (see, e.g., Rothman and Woodroffe, 1972).

3.1.1 Technical assumptions

To study the asymptotic properties of our tests, we require the following mild assumption on the distribution of the observations:

- (A1) $Y_1, \dots, Y_n$ are independent and identically distributed with continuous, positive density f satisfying $\sup_{x \in \mathbb{R}} f(x) < \infty$. Furthermore, $E(Y_1^2) < \infty$.

Since the true median η is assumed unknown, it needs to be estimated using a suitable estimator. For our theoretical results we require that the estimator has the following property:

- (A2) The estimator $\hat{\eta}_n$ admits the asymptotic representation

$$\sqrt{n}(\hat{\eta}_n - \eta) = \frac{d_\eta}{\sqrt{n}} \sum_{i=1}^n \psi(Y_i - \eta) + o_P(1), \quad (3.6)$$

as $n \rightarrow \infty$, where d_η is some constant and ψ is a measurable antisymmetric function satisfying $E[\psi(Y_1 - \eta)] = 0$ and $E[\psi^2(Y_1 - \eta)] < \infty$.

Assumption (A2) allows for a range of estimators of the location parameter η , including the sample mean and the sample median. Below we briefly discuss three examples of estimators admitting the representation in (3.6).

Example 3.1. If $\hat{\eta}_n$ the sample *mean*, then (3.6) holds with

$$d_\eta = 1 \quad \text{and} \quad \psi(x) = x.$$

Example 3.2. Suppose $\hat{\eta}_n$ is chosen as the sample *median*. From Serfling (1980) we have that

$$\sqrt{n}(\hat{\eta}_n - \eta) = \frac{1}{f(\eta)} \sqrt{n} \left(\frac{1}{2} - F_n(\eta) \right) + o_P(1) = \frac{1}{f(\eta)\sqrt{n}} \sum_{i=1}^n \left(\frac{1}{2} - \mathbf{1}(Y_i \leq \eta) \right) + o_P(1),$$

provided that $f(\eta) > 0$; also see Ghosh (1971). The sample median therefore has the representation in (3.6) with

$$d_\eta = \frac{1}{f(\eta)} \quad \text{and} \quad \psi(x) = \frac{1}{2} - \mathbf{1}(x \leq 0).$$

Example 3.3. The α -trimmed mean of a distribution function F is defined as

$$\frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x dF(x), \quad 0 \leq \alpha < \frac{1}{2}.$$

Suppose we estimate the unknown centre η using the plug-in estimator

$$\hat{\eta}_n = \frac{1}{1-2\alpha} \int_{F_n^{-1}(\alpha)}^{F_n^{-1}(1-\alpha)} x dF_n(x), \quad 0 \leq \alpha < \frac{1}{2}.$$

It is known (see, e.g., Milošević and Obradović, 2019, p. 48) that this estimator has the representation in (3.6) with

$$d_\eta = \frac{1}{1-2\alpha} \quad \text{and} \quad \psi(x) = \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} [F(t) - \mathbf{1}(x + \eta \leq t)] dt.$$

3.1.2 Limit null distribution

The following lemma shows that the process V_n is asymptotically equivalent to a properly-scaled sum of independent random processes. This allows application of a functional central limit theorem to obtain the asymptotic distribution of V_n .

Lemma 3.1. *Under Assumptions (A1) and (A2) the process V_n admits the asymptotic representation $V_n(t) = \tilde{V}_n(t) + o_P(1)$, uniformly in $t \geq 0$, where*

$$\tilde{V}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\text{sign}(-\varepsilon_i)(|\varepsilon_i| \wedge t) + d_\eta \psi(\varepsilon_i) Q(t)], \quad (3.7)$$

with $\varepsilon_i = Y_i - \eta$ and $Q(t) = P(|\varepsilon_1| \leq t)$.

Proof. We start from (3.3) and, for all $t \geq 0$, write $V_n(t) = A_n(t) + B_n(t) + C_n(t)$, where

$$\begin{aligned} A_n(t) &= \sqrt{n} \int_{-t}^t \left[F_n(\eta + x) - \frac{1}{2} \right] dx, \\ B_n(t) &= \sqrt{n} \int_{-t}^t [F(\hat{\eta}_n + x) - F(\eta + x)] dx, \\ C_n(t) &= \sqrt{n} \int_{-t}^t [F_n(\hat{\eta}_n + x) - F(\hat{\eta}_n + x)] dx - \sqrt{n} \int_{-t}^t [F_n(\eta + x) - F(\eta + x)] dx. \end{aligned}$$

Define the function $g(t) = \int_0^t F(s) ds$. We may then write

$$\int_{-t}^t F(s+h) ds = \int_{-t+h}^{t+h} F(s) ds = \int_0^{t+h} F(s) ds + \int_{-t+h}^0 F(s) ds = g(t+h) - g(-t+h).$$

Note that, since F is everywhere continuous by Assumption (A1), we have by the fundamental theorem of calculus that $g'(t) = F(t)$. By the mean value theorem, there exists a $\theta \in (0, 1)$ such that

$$g(t+h) = g(t) + h g'(t + \theta h),$$

that is,

$$\int_0^{t+h} F(x)dx = \int_0^t F(x)dx + hF(t) + h[F(t+h) - F(t)].$$

Now, using the fact that $f = F'$ is bounded, another application of the mean value theorem shows that

$$\sup_{t \in \mathbb{R}} |F(t+h) - F(t)| \leq |h| \sup_{t \in \mathbb{R}} f(t) \leq |h| K_f,$$

where $K_f = \sup_{t \in \mathbb{R}} f(t)$. Hence,

$$g(t+h) = \int_0^{t+h} F(x)dx = \int_0^t F(x)dx + hF(t) + O(h^2),$$

and therefore also

$$\int_{-t}^t F(s+h)ds = g(t+h) - g(-t+h) = \int_{-t}^t F(x)dx + h(F(t) - F(-t)) + O(h^2), \quad (3.8)$$

uniformly in $t \geq 0$, as $h \rightarrow 0$. Using this result, we get that

$$B_n(t) = \sqrt{n} \int_{-t}^t [F(s + \hat{\eta}_n) - F(s)] ds = \sqrt{n} \hat{\eta}_n (F(t) - F(-t)) + O_P(\sqrt{n} \hat{\eta}_n^2),$$

uniformly in $t \geq 0$. By Assumption (A2) we know that $\sqrt{n} \hat{\eta}_n = O_P(1)$, so that the error term in the last display is properly negligible. Using the asymptotic representation in Assumption (A2), we get

$$B_n(t) = \sqrt{n} \int_{-t}^t [F(s + \hat{\eta}_n) - F(s)] ds = (F(t) - F(-t)) \frac{d_\eta}{\sqrt{n}} \sum_{i=1}^n \psi(Y_i - \eta) + o_P(1),$$

uniformly in $t \geq 0$.

We now show that the term $C_n(t)$ is asymptotically negligible. First observe that

$$\begin{aligned} C_n(t) &= \sqrt{n} \int_{\hat{\eta}_n - t}^{\hat{\eta}_n + t} (F_n(x) - F(x)) dx - \sqrt{n} \int_{\eta - t}^{\eta + t} (F_n(x) - F(x)) dx \\ &= \sqrt{n} \int_{\eta + t}^{\hat{\eta}_n + t} (F_n(x) - F(x)) dx + \sqrt{n} \int_{\eta - t}^{\hat{\eta}_n - t} (F_n(x) - F(x)) dx. \end{aligned}$$

By Assumption (A2) we know that $\sqrt{n}(\hat{\eta}_n - \eta) = O_P(1)$, which, in conjunction with the Glivenko–Cantelli lemma (Kallenberg, 2002, p. 75), yields

$$\sup_{t \geq 0} \left| \sqrt{n} \int_{\eta + t}^{\hat{\eta}_n + t} (F_n(x) - F(x)) dx \right| \leq \sqrt{n} |\hat{\eta}_n - \eta| \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = o_P(1).$$

By the same arguments,

$$\sup_{t \geq 0} \left| \sqrt{n} \int_{\eta - t}^{\hat{\eta}_n - t} (F_n(x) - F(x)) dx \right| = o_P(1),$$

from which we obtain $\sup_t |C_n(t)| = o_P(1)$.

The representation of V_n given in (3.7) now follows directly by noting that, as before,

$$A_n(t) = \sqrt{n} \int_{-t}^t \left[F_n(\eta + x) - \frac{1}{2} \right] dx = \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\eta - Y_i) (|\eta - Y_i| \wedge t). \quad \square$$

The fact that V_n is asymptotically equivalent to a properly-normalised sum of independent random processes may then be used to establish the following result.

Theorem 3.1. *Under H_0 and Assumptions (A1) and (A2), the process V_n converges in distribution to a zero-mean Gaussian process G with covariance structure*

$$\begin{aligned} \mathbb{E}(G(s)G(t)) = & 2 \int_0^\infty (s \wedge u)(t \wedge u) dF_\eta(u) + 2d_\eta^2 Q(s)Q(t) \int_0^\infty \psi^2(u) dF_\eta(u) \\ & - 2d_\eta Q(t) \int_0^\infty (s \wedge u) \psi(u) dF_\eta(u) - 2d_\eta Q(s) \int_0^\infty (t \wedge u) \psi(u) dF_\eta(u), \end{aligned} \quad (3.9)$$

for $0 \leq s \leq t$, where $F_\eta(t) = P(Y_1 - \eta \leq t)$ and $Q(t) = P(|Y_1 - \eta| \leq t) = 2F_\eta(t) - 1$.

Proof. Let $\varepsilon_i = Y_i - \eta$, $i = 1, \dots, n$. Observe that, under the null hypothesis, $\mathbb{E}(V_n(t)) = 0$ for all $t \geq 0$. Define

$$A_i(t) = \text{sign}(-\varepsilon_i)(t \wedge |\varepsilon_i|), \quad B_i(t) = d_\eta \psi(\varepsilon_i) Q(t), \quad i = 1, \dots, n,$$

so that we may write

$$\tilde{V}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (A_i(t) + B_i(t)).$$

Noting that $\mathbb{E}(A_i(t)) = \mathbb{E}(B_i(t)) = 0$ and that the ε_i are independent, we have that

$$\mathbb{E}(\tilde{V}_n(s)\tilde{V}_n(t)) = \mathbb{E}(A_1(s)A_1(t)) + \mathbb{E}(B_1(s)B_1(t)) + \mathbb{E}(A_1(s)B_1(t)) + \mathbb{E}(B_1(s)A_1(t)).$$

Direct calculation shows that

$$\begin{aligned}
\mathbf{E}(A_1(s)A_1(t)) &= \mathbf{E}[(s \wedge |\varepsilon_1|)(t \wedge |\varepsilon_1|)] \\
&= \int_0^\infty (s \wedge x)(t \wedge x) dQ(x) \\
&= 2 \int_0^\infty (s \wedge x)(t \wedge x) dF_\eta(x), \\
\mathbf{E}(B_1(s)B_1(t)) &= d_\eta^2 Q(s)Q(t) \mathbf{E}(\psi^2(\varepsilon_1)), \\
\mathbf{E}(A_1(s)B_1(t)) &= d_\eta Q(t) \mathbf{E}[\text{sign}(-\varepsilon_1)(s \wedge |\varepsilon_1|)\psi(\varepsilon_1)] \\
&= d_\eta Q(t) \int_{-\infty}^\infty \text{sign}(-x)(s \wedge |x|)\psi(x) dF_\eta(x) \\
&= d_\eta Q(t) \left[- \int_0^\infty (s \wedge x)\psi(x) dF_\eta(x) + \int_{-\infty}^0 (s \wedge (-x))\psi(x) dF_\eta(x) \right] \\
&= d_\eta Q(t) \left[- \int_0^\infty (s \wedge x)\psi(x) dF_\eta(x) - \int_0^\infty (s \wedge x)\psi(-x) dF_\eta(-x) \right] \\
&= -2d_\eta Q(t) \int_0^\infty (s \wedge x)\psi(x) dF_\eta(x),
\end{aligned}$$

where we have made use of the fact that $\psi(-x) = -\psi(x)$ and, under H_0 , $F_\eta(-x) = 1 - F_\eta(x)$. Similarly,

$$\mathbf{E}(B_1(s)A_1(t)) = -2d_\eta Q(s) \int_0^\infty (t \wedge x) dF_\eta(x).$$

Combining all these results we obtain the covariance structure in the theorem.

To study the large-sample behaviour of $\tilde{V}_n$, we note that, since f is strictly positive, the behaviour of $\tilde{V}_n(t)$, $t \geq 0$, is equivalent to that of $W_n(s)$, $s \in [\frac{1}{2}, 1]$, defined through a continuous quantile transform as

$$\begin{aligned}
W_n(s) &= \tilde{V}_n(F_\eta^{-1}(s)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\text{sign}(-\varepsilon_i) \left(|\varepsilon_i| \wedge F_\eta^{-1}(s) \right) + d_\eta \psi(\varepsilon_i) (2s - 1) \right] \\
&=: \frac{1}{\sqrt{n}} \sum_{i=1}^n T_i(s), \quad s \in [\tfrac{1}{2}, 1].
\end{aligned}$$

Since W_n is a sum of independent zero-mean random processes, it follows by the multivariate central limit theorem that the finite-dimensional distributions of W_n are asymptotically normal with covariance structure given in (3.9).

It now only remains to verify that the process W_n is tight in $C[\frac{1}{2}, 1]$, the space of continuous functions on $[\frac{1}{2}, 1]$. Since W_n is a sum of independent, zero-mean random elements, we may apply Rosenthal's inequality (see, e.g., Petrov, 1995, p. 59) to obtain

$$\begin{aligned}
\mathbf{E}(|W_n(t) - W_n(s)|^2) &\leq \frac{c_1}{n} \sum_{i=1}^n \mathbf{E} \left\{ \left[\left(|\varepsilon_i| \wedge F_\eta^{-1}(t) \right) - \left(|\varepsilon_i| \wedge F_\eta^{-1}(s) \right) \right]^2 \right\} \\
&\quad + \frac{c_2}{n} d_\eta^2 (t - s)^2 \sum_{i=1}^n \mathbf{E}(\psi^2(\varepsilon_i)).
\end{aligned}$$

It is not difficult to see that

$$\mathbb{E} \left\{ \left[\left(|\varepsilon_1| \wedge F_\eta^{-1}(t) \right) - \left(|\varepsilon_1| \wedge F_\eta^{-1}(s) \right) \right]^2 \right\} \leq \left[F_\eta^{-1}(t) - F_\eta^{-1}(s) \right]^2.$$

This, together with the fact that $\mathbb{E}(|\psi(\varepsilon_1)|^2) < \infty$, shows that there exist constant $c_1, c_3 > 0$ such that

$$\mathbb{E}(|W_n(t) - W_n(s)|^2) \leq c_1 \left| F_\eta^{-1}(t) - F_\eta^{-1}(s) \right|^2 + c_3 |t - s|^2.$$

By Theorem 12.3 of Billingsley (1968, p. 130) it follows that the process W_n is tight in $C[\frac{1}{2}, 1]$. $\square$

Theorem 3.1 may be used to obtain the limit null distribution of the statistic $\hat{U}_n^{(I)}$ in (3.4). By the continuous mapping theorem,

$$\hat{U}_n^{(I)} = \sup_{t \geq 0} |V_n(t)| \xrightarrow{d} \sup_{t \geq 0} |G(t)|.$$

Although this result may be used to approximate the asymptotic critical values of the test based on $\hat{U}_n^{(I)}$, it is clear that (3.9) depends on unknown quantities that need to be estimated. Furthermore, depending on the underlying distribution, convergence of the finite-sample critical values to that of the asymptotic distribution could be quite slow.

To illustrate how the underlying distribution F and the choice of the estimator $\hat{\eta}_n$ of η affects the null distribution of $\hat{U}_n^{(I)}$, we conducted a small Monte Carlo study. We generated a random sample of $n \in \{10, 20, 50, 100, 500\}$ observations from one of three choices of F : a standard normal distribution, a uniform $[-1, 1]$ distribution, or a t_4 distribution. The median η was assumed to be unknown and was estimated using either the sample mean, the sample median, or the 10%-trimmed sample mean. This was repeated independently 10 000 times for each configuration to obtain the approximate finite-sample quantiles shown in Tables 3.1 to 3.3. Clearly, both the underlying distribution F and the choice of $\hat{\eta}_n$ have an influence on the finite-sample null distribution of $\hat{U}_n^{(I)}$.

It is of interest to compare these finite-sample critical values with the asymptotic critical values implied by Theorem 3.1. For each configuration, we sampled 10 000 independent sample paths of a zero-mean Gaussian process G with covariance structure given in (3.9) to approximate quantiles of $\sup_{t \geq 0} |G(t)|$. The approximate theoretical quantiles obtained are given in the right-most columns of Tables 3.1 to 3.3. For the quantities F , d_η and ψ appearing in (3.9) we used the true values corresponding to each configuration. The values of d_η and ψ for each of the considered estimators are given in Examples 3.1 to 3.3.

Table 3.1: A selection of finite-sample and asymptotic upper-level quantiles of $\hat{U}_n^{(I)}$ in the case where η is estimated using the *sample mean*.

F	Level	Finite-sample quantiles					Asymptotic
		$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 500$	
$N(0, 1)$	90%	0.3921	0.4064	0.4140	0.4085	0.4064	0.4063
	95%	0.4640	0.4755	0.4824	0.4728	0.4782	0.4790
	99%	0.6191	0.6120	0.6297	0.6073	0.6178	0.6158
t_4	90%	0.7357	0.8189	0.8588	0.8778	0.9058	0.9093
	95%	0.9742	1.0369	1.0630	1.0592	1.0786	1.0646
	99%	1.7720	1.7445	1.6205	1.5114	1.4253	1.3691
$U[-1, 1]$	90%	0.2304	0.2425	0.2478	0.2509	0.2541	0.2516
	95%	0.2683	0.2817	0.2906	0.2968	0.2946	0.2944
	99%	0.3504	0.3671	0.3728	0.3785	0.3841	0.3869

Table 3.2: A selection of finite-sample and asymptotic upper-level quantiles of $\hat{U}_n^{(I)}$ in the case where η is estimated using the *sample median*.

F	Level	Finite-sample quantiles					Asymptotic
		$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 500$	
$N(0, 1)$	90%	1.0272	1.1234	1.2046	1.2230	1.2432	1.2519
	95%	1.2355	1.3475	1.4454	1.4585	1.4695	1.4891
	99%	1.6402	1.7719	1.8835	1.9224	1.9419	1.9757
t_4	90%	1.4638	1.5881	1.6656	1.6702	1.7439	1.7409
	95%	1.8317	1.9531	1.9860	1.9959	2.0460	2.0449
	99%	2.8618	2.7324	2.7181	2.6433	2.6520	2.6523
$U[-1, 1]$	90%	0.6114	0.7440	0.8297	0.8716	0.9114	0.9504
	95%	0.7039	0.8683	0.9742	1.0246	1.0931	1.1340
	99%	0.8814	1.0770	1.2756	1.3510	1.4286	1.5019

Table 3.3: A selection of finite-sample and asymptotic upper-level quantiles of $\hat{U}_n^{(I)}$ in the case where η is estimated using the *10%-trimmed sample mean*.

F	Level	Finite-sample quantiles					Asymptotic
		$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 500$	
$N(0, 1)$	90%	0.3986	0.3949	0.4023	0.4077	0.4136	0.4113
	95%	0.4882	0.4692	0.4762	0.4796	0.4858	0.4827
	99%	0.6545	0.6277	0.6218	0.6277	0.6398	0.6273
t_4	90%	0.8014	0.8657	0.9634	0.9744	1.0009	1.0072
	95%	1.0746	1.1469	1.2194	1.1922	1.1971	1.1966
	99%	2.0161	2.0400	1.8191	1.6792	1.5529	1.5206
$U[-1, 1]$	90%	0.1769	0.1833	0.1890	0.1942	0.2014	0.2130
	95%	0.2037	0.2092	0.2142	0.2213	0.2318	0.2468
	99%	0.2568	0.2692	0.2702	0.2722	0.2956	0.3049

To avoid estimation of these unknown quantities in application of the tests, one can make use of a bootstrap approach (see Section 2.6) as we have done in our numerical studies.

Using Theorem 3.1 to obtain the null distribution of the Cramér–von Mises type tests based on $\widehat{W}_n^{(I)}$ is more delicate and we do not pursue it here. We expect, however, since $\sup_{t \in \mathbb{R}} |\widehat{Q}_n(t) - Q(t)| \rightarrow 0$, that the following convergence result will hold:

$$\widehat{W}_n^{(I)} = \int_0^\infty V_n^2(t) d\widehat{Q}_n(t) \xrightarrow{d} \int_0^\infty G^2(t) dQ(t),$$

as $n \rightarrow \infty$.

3.1.3 Consistency

To establish consistency of the tests against general alternatives, we require the following result.

Lemma 3.2. *Under Assumptions (A1) and (A2) it holds for each fixed $t \geq 0$ that*

$$n^{-1/2}V_n(t) \xrightarrow{P} V(t).$$

Proof. The proof is straightforward and we just note the following. For any fixed $t \geq 0$ we have by Lemma 3.1 that

$$\frac{1}{\sqrt{n}}V_n(t) = V(t) + \int_{-t}^t [F_n(\eta + x) - F(\eta + x)] dx + \frac{1}{\sqrt{n}}B_n(t) + o_P(1).$$

The Glivenko–Cantelli lemma implies that $\int_{-t}^t [F_n(\eta + x) - F(\eta + x)] dx = o_P(1)$ for each fixed $t \geq 0$. Also, since F is continuous and bounded,

$$\begin{aligned} B_n(t) &= \sqrt{n} \int_{-t}^t [F(\widehat{\eta}_n + x) - F(\eta + x)] dx \\ &= \sqrt{n} \int_{(\widehat{\eta}_n - \eta) - t}^{(\widehat{\eta}_n - \eta) + t} F_\eta(x) dx - \sqrt{n} \int_{-t}^t F_\eta(x) dx \\ &= \sqrt{n} \int_t^{(\widehat{\eta}_n - \eta) + t} F_\eta(x) dx + \sqrt{n} \int_{(\widehat{\eta}_n - \eta) - t}^{-t} F_\eta(x) dx. \end{aligned}$$

Hence,

$$\sup_{t \geq 0} |B_n(t)| \leq 2\sqrt{n} |\widehat{\eta}_n - \eta| = O_P(1). \quad \square$$

If H_0 does not hold, it is clear from (3.2) that there exists a $t^* > 0$ such that $V(t^*) \neq 0$. Consequently, under alternatives, the quantity $n^{-1/2}V_n(t^*)$ converges in probability to a nonzero constant by Lemma 3.2. Therefore, applying the continuous mapping theorem, one immediately sees that

$$\sup_{t \geq 0} |n^{-1/2}V_n(t)| \geq n^{-1/2}|V_n(t^*)| \xrightarrow{P} |V(t^*)| > 0,$$

so that $\widehat{U}_n^{(I)} = \sup_{t \geq 0} |V_n(t)| \xrightarrow{P} \infty$ under alternatives. Hence the test based on $\widehat{U}_n^{(I)}$ is consistent.

3.2 Error terms in linear regression models

We now generalise the results of the preceding section to the case where the objective is to test whether the distribution of error terms appearing in a regression model is symmetric around zero. We assume that $(Y; \mathbf{x})$ is an observation generated by the multivariate linear model

$$Y = \mathbf{x}^\top \boldsymbol{\beta} + \varepsilon, \quad (3.10)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown regression parameters and the regressors $\mathbf{x} \in \mathbb{R}^p$ are assumed fixed (i.e. nonrandom). The null hypothesis of interest is

$$H_0 : F_\varepsilon(t) + F_\varepsilon(-t) = 1 \quad \text{for all } t \geq 0,$$

where F_ε denotes the distribution function of ε .

Suppose $(Y_1; \mathbf{x}_1), \dots, (Y_n; \mathbf{x}_n)$ is a random sample generated by the model in (3.10). Since the errors $\varepsilon_1, \dots, \varepsilon_n$ are unobserved, we base our test on the residuals

$$\hat{\varepsilon}_i = Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_n = \varepsilon_i - \mathbf{x}_i^\top (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}), \quad i = 1, 2, \dots, n,$$

where $\hat{\boldsymbol{\beta}}_n$ is an appropriate estimator for $\boldsymbol{\beta}$. In this case an appropriate version of the process $V_n(t)$ in (3.3) is

$$V_n^\varepsilon(t) = \sqrt{n} \int_{-t}^t \left(F_n^{\hat{\varepsilon}}(s) - \frac{1}{2} \right) ds = \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(-\hat{\varepsilon}_i) (|\hat{\varepsilon}_i| \wedge t), \quad t \geq 0,$$

where

$$F_n^{\hat{\varepsilon}}(s) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{\varepsilon}_i \leq s)$$

is the empirical distribution function constructed from the residuals.

3.2.1 Technical assumptions

Our theoretical results rely on a few mild assumptions which are typical in related literature. We separately formulate the requirements on the model errors, the regression design (where $\mathbf{X}_n$ denotes the $n \times p$ design matrix), and the estimator of $\boldsymbol{\beta}$:

- (B1) $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. with continuous, positive density f satisfying $\sup_{x \in \mathbb{R}} f(x) < \infty$. Furthermore, $E(\varepsilon_1^2) < \infty$.
- (B2) $\max_{1 \leq i \leq n, 1 \leq v \leq p} |x_{iv}| = o(n^{1/2})$, where x_{iv} is the v th element of the vector $\mathbf{x}_i$, and $n^{-1} \mathbf{X}_n^\top \mathbf{X}_n \rightarrow \mathbf{C}$ for some positive definite matrix $\mathbf{C}$.

(B3) The estimator $\hat{\beta}_n$ satisfies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta) = \frac{d_\beta}{\sqrt{n}} \sum_{i=1}^n \psi_\beta(\varepsilon_i) + o_P(1),$$

as $n \rightarrow \infty$, where d_β is some constant and ψ_β is a measurable antisymmetric function satisfying $E(\psi_\beta(\varepsilon_1)) = 0$ and $E(\psi_\beta^2(\varepsilon_1)) < \infty$.

Assumption (B2) on the design matrix is quite standard in the regression literature and covers most reasonable designs (cf. Koul, 2002; Hušková and Meintanis, 2012). Assumption (B3) allows for a wide array of choices of $\hat{\beta}_n$ as estimator of β and typically includes ordinary least-squares estimators (when the model has an intercept) as well as the more general class of M-estimators. For a more detailed discussion on regression assumptions and their implications in the context of goodness-of-fit tests for the error distribution in linear regression models, see Hušková and Meintanis (2007, pp. 369–372). Also see our discussion in Section 2.5.

3.2.2 Asymptotic null distribution

In order to derive the limit null distributions of the tests based on V_n^ε , the first step is to find an asymptotic representation of V_n^ε as a sum of independent random processes. Specifically, we require a result such as the one stated in Lemma 3.1, but for the linear regression case.

Lemma 3.3. *Under Assumptions (B1)–(B3), the process V_n^ε admits the representation*

$$V_n^\varepsilon(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\text{sign}(-\varepsilon_i)(|\varepsilon_i| \wedge t) + d_\beta \psi_\beta(\varepsilon_i) Q_\varepsilon(t)] + o_P(1),$$

uniformly in $t \geq 0$, where $Q_\varepsilon(t) = P(|\varepsilon_1| \leq t) = F_\varepsilon(t) - F_\varepsilon(-t)$.

Proof. Let $F_n^\varepsilon(s) = n^{-1} \sum_{i=1}^n \mathbf{1}(\varepsilon_i \leq s)$ and write $V_n^\varepsilon(t) = A_n^\varepsilon(t) + B_n^\varepsilon(t) + C_n^\varepsilon(t)$, where

$$\begin{aligned} A_n^\varepsilon(t) &= \sqrt{n} \int_{-t}^t \left(F_n^\varepsilon(s) - \frac{1}{2} \right) ds, \\ B_n^\varepsilon(t) &= \sqrt{n} \int_{-t}^t \left[\frac{1}{n} \sum_{i=1}^n F_\varepsilon(s + \mathbf{x}_i^\top (\hat{\beta}_n - \beta)) - F_\varepsilon(s) \right] ds, \\ C_n^\varepsilon(t) &= \sqrt{n} \int_{-t}^t \left[\frac{1}{n} \sum_{i=1}^n \mathbf{1}(\varepsilon_i \leq s + \mathbf{x}_i^\top (\hat{\beta}_n - \beta)) - \frac{1}{n} \sum_{i=1}^n F_\varepsilon(s + \mathbf{x}_i^\top (\hat{\beta}_n - \beta)) \right] ds \\ &\quad - \sqrt{n} \int_{-t}^t (F_n^\varepsilon(s) - F_\varepsilon(s)) ds. \end{aligned}$$

Notice that, by Assumption (B2),

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n [x_i^\top (\hat{\beta}_n - \beta)]^2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{\beta}_n - \beta)^\top x_i x_i^\top (\hat{\beta}_n - \beta) \\
&= \frac{1}{\sqrt{n}} (\hat{\beta}_n - \beta)^\top (\mathbf{X}_n^\top \mathbf{X}_n) (\hat{\beta}_n - \beta) \\
&= O_P(n^{-1/2}).
\end{aligned} \tag{3.11}$$

Therefore, from the expansion in (3.8) and the representation in Assumption (B3),

$$\begin{aligned}
B_n^\varepsilon(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta) (F_\varepsilon(t) - F_\varepsilon(-t)) + O_P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n [x_i^\top (\hat{\beta}_n - \beta)]^2\right) \\
&= Q_\varepsilon(t) \frac{1}{\sqrt{n}} \sum_{i=1}^n d_\beta \psi_\beta(\varepsilon_i) + o_P(1),
\end{aligned}$$

uniformly in $t \geq 0$. It now only remains to show that $C_n^\varepsilon(\cdot)$ is negligible.

Observe that

$$\begin{aligned}
C_n^\varepsilon(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-t}^t [\mathbf{1}(\varepsilon_i \leq s + x_i^\top (\hat{\beta}_n - \beta)) - F_\varepsilon(s + x_i^\top (\hat{\beta}_n - \beta))] ds \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-t}^t [\mathbf{1}(\varepsilon_i \leq s) - F_\varepsilon(s)] ds \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-t+x_i^\top (\hat{\beta}_n - \beta)}^{t+x_i^\top (\hat{\beta}_n - \beta)} [\mathbf{1}(\varepsilon_i \leq s) - F_\varepsilon(s)] ds - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-t}^t [\mathbf{1}(\varepsilon_i \leq s) - F_\varepsilon(s)] ds \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_t^{t+x_i^\top (\hat{\beta}_n - \beta)} [\mathbf{1}(\varepsilon_i \leq s) - F_\varepsilon(s)] ds \\
&\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-t}^{-t+x_i^\top (\hat{\beta}_n - \beta)} [\mathbf{1}(\varepsilon_i \leq s) - F_\varepsilon(s)] ds.
\end{aligned}$$

We shall only treat the first term since the second term can be treated in exactly the same way. Write

$$\begin{aligned}
&\int_t^{t+x_i^\top (\hat{\beta}_n - \beta)} \mathbf{1}(\varepsilon_i \leq s) ds \\
&= \mathbf{1}(\varepsilon_i \leq t) \int_t^{t+x_i^\top (\hat{\beta}_n - \beta)} ds + \int_t^{t+x_i^\top (\hat{\beta}_n - \beta)} [\mathbf{1}(\varepsilon_i \leq s) - \mathbf{1}(\varepsilon_i \leq t)] ds \\
&= x_i^\top (\hat{\beta}_n - \beta) \mathbf{1}(\varepsilon_i \leq t) + \int_t^{t+x_i^\top (\hat{\beta}_n - \beta)} [\mathbf{1}(\varepsilon_i \leq s) - \mathbf{1}(\varepsilon_i \leq t)] ds \\
&=: D_{i,n}^{(1)}(t) + E_{i,n}^{(1)}(t).
\end{aligned}$$

Also, by Taylor expansion,

$$\begin{aligned}
\int_t^{t+x_i^\top (\hat{\beta}_n - \beta)} F_\varepsilon(s) ds &= x_i^\top (\hat{\beta}_n - \beta) F_\varepsilon(t) + [x_i^\top (\hat{\beta}_n - \beta)]^2 f_\varepsilon\left(t + x_i^\top (\tilde{\beta}^{(2)} - \beta)\right) \\
&=: D_{i,n}^{(2)}(t) + E_{i,n}^{(2)}(t),
\end{aligned}$$

for some $\tilde{\beta}^{(2)}$ such that $\|\tilde{\beta}^{(2)} - \beta\| \leq \|\hat{\beta}_n - \beta\|$. Using these representations, we see that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n (D_{i,n}^{(1)}(t) - D_{i,n}^{(2)}(t)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta) [\mathbf{1}(\varepsilon_i \leq t) - F_\varepsilon(t)] \\ &= (\hat{\beta}_n - \beta)^\top \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i [\mathbf{1}(\varepsilon_i \leq t) - F_\varepsilon(t)] \\ &= \sum_{v=1}^p (\hat{\beta}_{nv} - \beta_v) \left(\frac{1}{n} \sum_{i=1}^n x_{iv}^2 \right)^{1/2} \sum_{i=1}^n d_{iv} [\mathbf{1}(\varepsilon_i \leq t) - F_\varepsilon(t)], \end{aligned}$$

where $d_{iv} = x_{iv}(\sum_{i=1}^n x_{iv}^2)^{-1/2}$, and $\hat{\beta}_{nv}$ and β_v denote the v th elements of $\hat{\beta}_n$ and β , respectively. By Theorem 2.2.1 of Koul (2002) one can easily verify that, under Assumption (B2), the process $\sum_{i=1}^n d_{iv} [\mathbf{1}(\varepsilon_i \leq t) - F_\varepsilon(t)]$ converges weakly to a Gaussian process and is bounded in probability. Furthermore, by Assumptions (B2) and (B3) we have $n^{-1} \sum_{i=1}^n x_{iv}^2 = o(n)$ and $\hat{\beta}_n - \beta = O_P(n^{-1/2})$, so that

$$\sup_{t \geq 0} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (D_{i,n}^{(1)}(t) - D_{i,n}^{(2)}(t)) \right| = o_P(1).$$

To treat the term $n^{-1/2} \sum_{i=1}^n E_{i,n}^{(1)}(t)$, observe that, by the Cauchy–Schwarz inequality,

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n E_{i,n}^{(1)}(t) \right| &\leq \frac{2}{\sqrt{n}} \sum_{i=1}^n |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| \mathbf{1}(|\varepsilon_i - t| \leq |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|) \\ &\leq \frac{2}{\sqrt{n}} \left\{ \sum_{i=1}^n [\mathbf{x}_i^\top (\hat{\beta}_n - \beta)]^2 \right\}^{1/2} \left\{ \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|) \right\}^{1/2} \\ &= 2 \left\{ \sum_{i=1}^n [\mathbf{x}_i^\top (\hat{\beta}_n - \beta)]^2 \right\}^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|) \right\}^{1/2} \\ &= O_P(1) \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|) \right\}^{1/2}. \end{aligned}$$

Now, fix any $\delta > 0$ and observe that

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|) \\ &\leq \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left(|\varepsilon_i - t| \leq \max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| \right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left(|\varepsilon_i - t| \leq \max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|, \max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| \leq \delta \right) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \mathbf{1} \left(|\varepsilon_i - t| \leq \max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|, \max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| > \delta \right) \\ &\leq \frac{1}{n} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq \delta) + \mathbf{1} \left(\max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| > \delta \right). \end{aligned}$$

By the Glivenko–Cantelli lemma,

$$\sup_{t \geq 0} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq \delta) - \mathbf{P}(|\varepsilon_1 - t| \leq \delta) \right| \xrightarrow{\mathbf{P}} 0.$$

Furthermore, since $\max_i |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| = o_{\mathbf{P}}(1)$, there exists an $n_\delta > 0$ such that

$$\mathbf{P} \left(\mathbf{1} \left(\max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| > \delta \right) > 0 \right) = \mathbf{P} \left(\max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| > \delta \right) < \delta,$$

for all $n \geq n_\delta$, which means that also $\mathbf{1}(\max_i |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| > \delta) \xrightarrow{\mathbf{P}} 0$. Therefore,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|) \leq \mathbf{P}(|\varepsilon_1 - t| \leq \delta) + o_{\mathbf{P}}(1),$$

uniformly in $t \geq 0$. As the relation in the last display holds for any $\delta > 0$, we may choose δ arbitrarily small. Also, by the uniform continuity of F_ε , we know that

$$\sup_{t \geq 0} \mathbf{P}(|\varepsilon_1 - t| \leq \delta) = \sup_{t \geq 0} [F_\varepsilon(t + \delta) - F_\varepsilon(t - \delta)] \rightarrow 0$$

as $\delta \rightarrow 0$. Hence,

$$\sup_{t \geq 0} \frac{1}{n} \sum_{i=1}^n \mathbf{1}(|\varepsilon_i - t| \leq |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)|) = o_{\mathbf{P}}(1).$$

Finally, from (3.11) and the boundedness of f_ε , and setting $K_{f_\varepsilon} = \sup_{t \geq 0} f_\varepsilon(t)$,

$$\begin{aligned} \sup_{t \geq 0} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n E_{i,n}^{(2)}(t) \right| &\leq \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbf{x}_i^\top (\hat{\beta}_n - \beta)]^2 \sup_{t \geq 0} \left| f_\varepsilon \left(t + \mathbf{x}_i^\top (\hat{\beta}_n^{(2)} - \beta) \right) \right| \\ &\leq K_{f_\varepsilon} \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbf{x}_i^\top (\hat{\beta}_n - \beta)]^2 \\ &= O_{\mathbf{P}} \left(n^{-1/2} \right). \end{aligned}$$

By the same arguments one may treat the remaining terms in $C_n^\varepsilon(t)$ to show that $\sup_{t \geq 0} |C_n^\varepsilon(t)| = o_{\mathbf{P}}(1)$. $\square$

Lemma 3.3 shows that, under the stated assumptions, the process V_n^ε is asymptotically equivalent to a properly-scaled sum of independent random processes. Following exactly the same steps as in the proof of Theorem 3.1, we get the following result.

Theorem 3.2. *Under H_0 and Assumptions (B1)–(B3), the process V_n^ε converges in distribution to a zero-mean Gaussian process G with covariance structure*

$$\begin{aligned} \mathbf{E}(G(s)G(t)) &= 2 \int_0^\infty (s \wedge u)(t \wedge u) dF_\varepsilon(u) + 2d_\beta^2 Q(s)Q(t) \int_0^\infty \psi^2(u) dF_\varepsilon(u) \\ &\quad - 2d_\beta Q(t) \int_0^\infty (s \wedge u) \psi(u) dF_\varepsilon(u) - 2d_\beta Q(s) \int_0^\infty (t \wedge u) \psi(u) dF_\varepsilon(u), \end{aligned} \tag{3.12}$$

$0 \leq s \leq t$, where $Q(t) = \mathbf{P}(|\varepsilon_1| \leq t) = 2F_\varepsilon(t) - 1$.

We note here that the limit null distribution G given in Theorem 3.2 depends on the unknown error distribution F_ε and the chosen estimator of β (which determines ψ_β and d_β). This is not surprising and is in fact in agreement with findings by other authors. We specifically mention the work of Koul (2002), who asserted that tests for symmetry based on the empirical symmetry process are in general not distribution free. Observe, however, that the covariance structure in (3.12) does not depend on the regressors under the stated assumptions. It is possible to relax Assumption (A3) to include a wider range of model configurations (such as a model without intercept) and estimators, but typically in these more general situations the covariance structure in (3.12) may depend on the regressors.

3.2.3 Consistency

In the case where the objective is to test whether the regression errors have a symmetric distribution, we may establish consistency of the tests using similar arguments to those used in the last paragraph of Section 3.1.3. To be able to do this, we require a result analogous to Lemma 3.2, but tailored for the linear regression case. This is stated and proved in the following lemma.

Lemma 3.4. *Under Assumptions (B1)–(B3) it holds for each fixed $t \geq 0$ that $n^{-1/2}V_n^\varepsilon(t) \xrightarrow{P} V_\varepsilon(t)$ as $n \rightarrow \infty$.*

Proof. Looking at the proof of Lemma 3.3, we see that the negligibility of the term $C_n^\varepsilon(t)$ does not rely on the assumption of symmetry, and therefore also goes through under alternatives. Therefore, for any fixed $t \geq 0$,

$$\frac{1}{\sqrt{n}}V_n^\varepsilon(t) = V_\varepsilon(t) + \int_{-t}^t (F_n^\varepsilon(s) - F_\varepsilon(s)) ds + \frac{1}{\sqrt{n}}B_n^\varepsilon(t) + o_P(1) = V_\varepsilon(t) + \frac{1}{\sqrt{n}}B_n^\varepsilon(t) + o_P(1),$$

where we made use of the fact that

$$\left| \int_{-t}^t (F_n^\varepsilon(s) - F_\varepsilon(s)) ds \right| \leq 2t \sup_{s \in \mathbb{R}} |F_n^\varepsilon(s) - F_\varepsilon(s)| = o_P(1),$$

by the Glivenko–Cantelli lemma.

To show that the term $n^{-1/2}B_n(t)$ is negligible, first observe that

$$\begin{aligned} \sup_s |F_\varepsilon(s + \mathbf{x}_i^\top (\hat{\beta}_n - \beta)) - F_\varepsilon(s)| \\ \leq \sup \left\{ |F_\varepsilon(u) - F_\varepsilon(v)| : |u - v| \leq \max_{i=1, \dots, n} |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| \right\} \xrightarrow{P} 0 \end{aligned}$$

by the uniform continuity of F_ε and the assumption that $\max_i |\mathbf{x}_i^\top (\hat{\beta}_n - \beta)| \xrightarrow{P} 0$. Therefore it holds for each fixed $t \geq 0$ that

$$\frac{1}{\sqrt{n}}|B_n(t)| \leq 2t o_P(1) = o_P(1),$$

which completes the proof. $\square$

3.3 Numerical study

To complement the theoretical results derived in this chapter, we now provide some results from numerical simulations which shed some light on the finite-sample behaviour of the newly-proposed tests.

3.3.1 Case I: Intercept-only model (no regressors)

We first compare the finite-sample performance of the *integrated* versions of the Kolmogorov–Smirnov and Cramér–von Mises tests $\hat{U}_n^{(I)}$ and $\hat{W}_n^{(I)}$ proposed in this chapter with that of the classical (nonintegrated) Kolmogorov–Smirnov and Cramér–von Mises tests $\hat{U}_n$ and $\hat{W}_n$ defined in (2.5).

Observations $Y_1, \dots, Y_n$ were randomly sampled from either the skew normal distribution, Tukey’s g distribution or the χ^2 distribution (all defined in Table 2.1) for several choices of the parameter ϑ . The true centre of the distribution was assumed to be unknown and estimated using the sample mean $\hat{\eta}_n$. The tests were then applied to the generated data and it was recorded whether the tests rejected the null hypothesis of symmetry. This procedure was repeated 1 000 times (independently), where in each iteration we used the model-based bootstrap (with 1 000 replications in each Monte Carlo iteration) described in Section 2.6.1 to obtain critical values of the tests.

Power curves (at significance levels $\alpha \in \{0.01, 0.05, 0.1\}$) for the case of the skew normal distribution are shown in Figure 3.1. The solid lines correspond to the integrated tests, whereas the dashed lines correspond to the classical tests. For the skew normal distribution, a choice of $\vartheta = 0$ corresponds to the null hypothesis of symmetry, whereas larger values of ϑ are associated with higher degrees of asymmetry; cf. the skewness coefficient stated in Table 2.1. Clearly, in the cases considered, the integrated versions of the tests achieve the considered nominal levels (indicated as ticks on the vertical axis at 1%, 5% and 10%) when $\vartheta = 0$, with no indication of being oversized. Also notice that the classical, nonintegrated tests are conservative in the sense that, for each chosen nominal level, the percentage of rejections under the null hypothesis is lower than the nominal level.

As the value of ϑ is increased, there is an increase in power of all the considered tests, which is certainly a desirable property. It is clear that the integrated tests have significantly higher power than the classical counterparts under all the alternatives shown in Figure 3.1. For example, when $\vartheta = 3$ and the sample size is $n = 75$, the power (at a nominal level of 5%) of the integrated Kolmogorov–Smirnov test is approximately 55% whereas the classical Kolmogorov–Smirnov test has power of approximately 25%. This is a relative increase in power of around 120%.

Another feature of the new integrated tests is that, under alternatives, the power seems to increase as the sample size is increased. This is in agreement with the consistency result stated in Lemma 3.2 and the subsequent discussion.

The power curves shown in Figures 3.2 and 3.3, which correspond to the case where observations are sampled from Tukey's g distribution and the χ^2 distribution, respectively, show similar features as pointed out above for the case of the skew normal distribution. There is slight indication that the tests are oversized for sample size $n = 75$ in the case of Tukey's g distribution, but this quickly disappears with increasing the sample size.

3.3.2 Case II: Linear regression

We now study the behaviour of the new tests in the context of testing whether the distribution of errors in a linear regression can be assumed symmetric around zero.

Our simulation setup is similar to that of Neumeyer et al. (2005), who considered the linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

with $\beta_0 = 0$ and $\beta_1 = \beta_2 = 1$, and the x_i being fixed regressors (these are sampled randomly from a uniform(0,1) distribution once and kept fixed over all Monte Carlo iterations). The errors were simulated from the distributions in Table 2.1 and, if necessary, were adjusted to have zero mean and unit variance. See Section 2.7 for more detail on how to sample from these distributions.

Seeing that the errors are unobservable, we base all tests on the residuals defined as

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2, \quad i = 1, 2, \dots, n,$$

where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are the ordinary least-squares estimators of the regression parameters β_0 , β_1 and β_2 , respectively (see Section 2.5). On the basis of $\{\hat{\varepsilon}_i\}_{i=1}^n$ the test statistic is computed, say $T = T(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)$.

As the limit distribution stated in Theorem 3.2 is complicated and depends on several unknowns which need to be estimated, we turn to the bootstrap for approximation of the critical values of the test. We use the following wild bootstrap algorithm proposed and described in Section 2.6.2, but we restate it here for this specific situation:

1. Generate independent random variables $\{\xi_i\}_{i=1}^n$ from the two-point distribution

$$P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}.$$

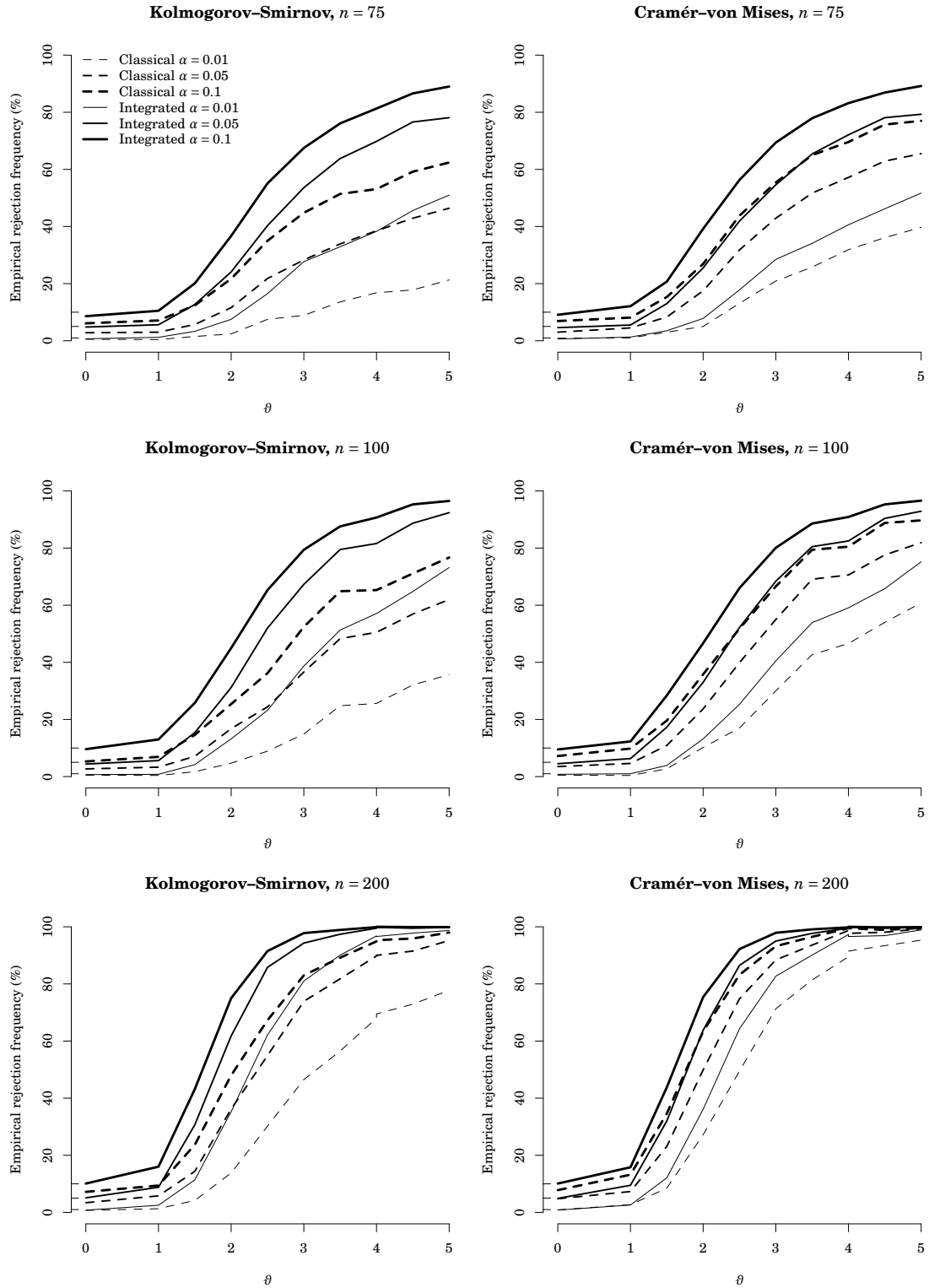


Figure 3.1: Power curves for the classical and integrated Kolmogorov-Smirnov and Cramér-von Mises tests for the case where observations were sampled from a skew normal distribution and where there are no regressors.

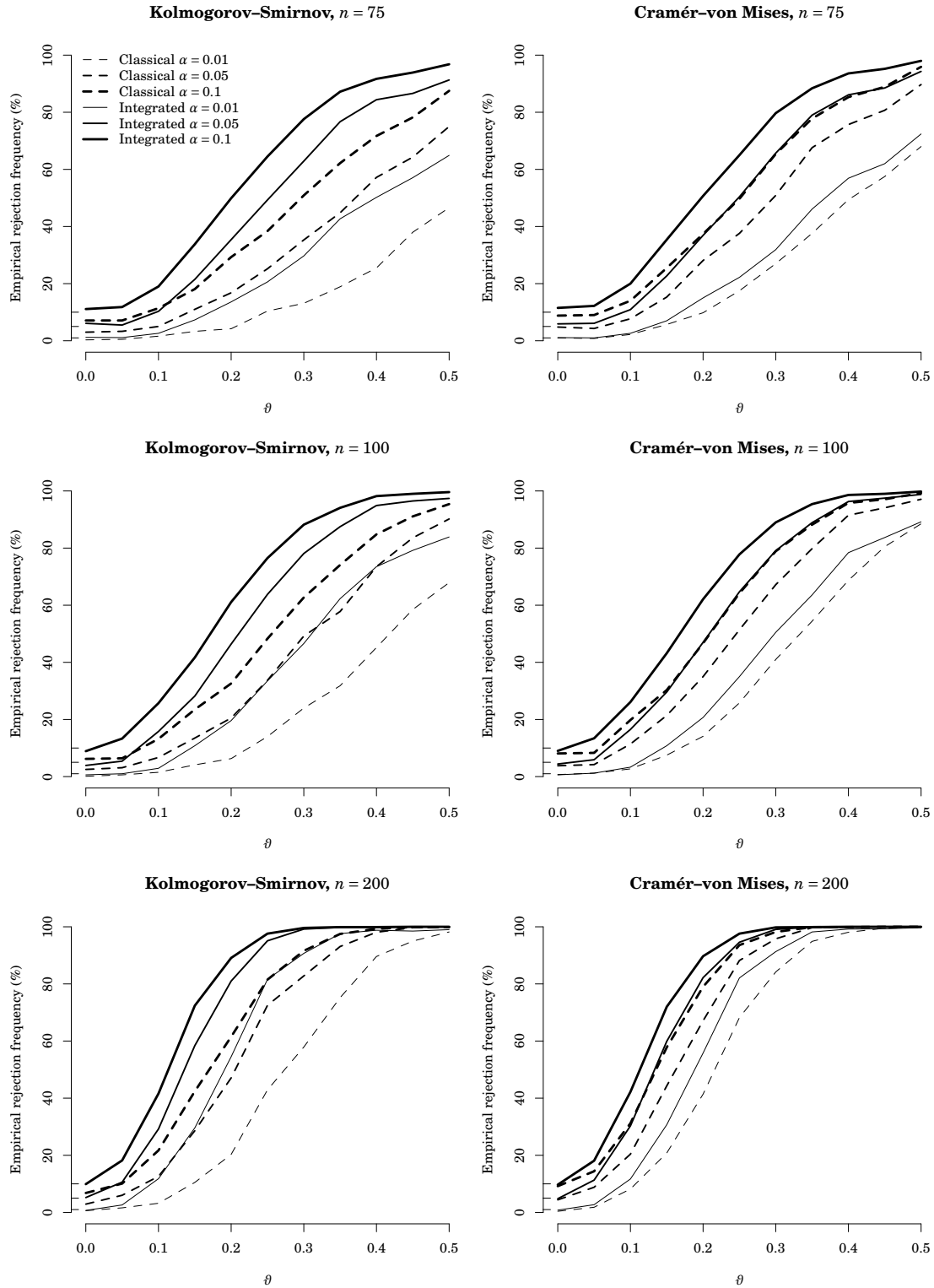


Figure 3.2: Power curves for the classical and integrated Kolmogorov–Smirnov and Cramér–von Mises tests for the case where observations were sampled from Tukey’s g distribution and where there are no regressors.

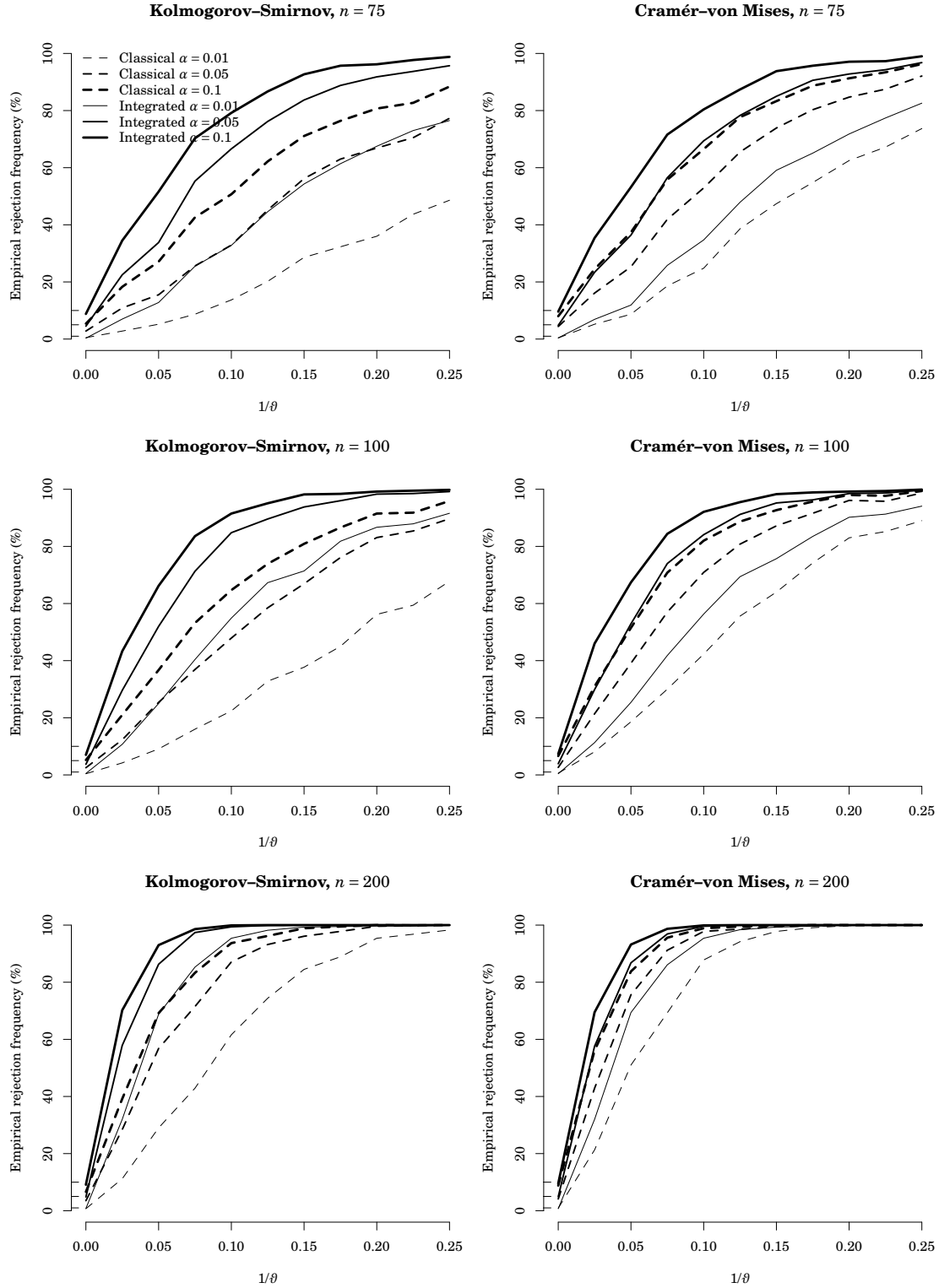


Figure 3.3: Power curves for the classical and integrated Kolmogorov-Smirnov and Cramér-von Mises tests for the case where observations were sampled from the χ^2 distribution and where there are no regressors.

2. Compute the bootstrap responses

$$Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 + \xi_i \hat{\varepsilon}_i, \quad i = 1, 2, \dots, n.$$

3. Calculate the estimators $\hat{\beta}_0^*$, $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ from the bootstrap data $\{Y_i^*, x_i, x_i^2\}_{i=1}^n$ using least-squares estimation.

4. Based on the bootstrap residuals

$$\hat{\varepsilon}_i^* = Y_i^* - \hat{\beta}_0^* - \hat{\beta}_1^* x_i - \hat{\beta}_2^* x_i^2,$$

compute a bootstrap analogue of the test statistic, namely $T^* = T(\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*)$.

Repeat steps 1–4 a large number of times, say B , to obtain $\{T_b^*\}_{b=1}^B$. Reject the null hypothesis at a prescribed significance level α if $T > T_{(\lfloor B(1-\alpha) \rfloor)}^*$, where $T_{(1)}^* \leq T_{(2)}^* \leq \dots \leq T_{(B)}^*$ denote the order statistics associated with the replications T_b^* . In all simulations we chose $B = 1000$ bootstrap replications to approximate the critical value.

As in the previous subsection, we first compare the finite-sample performance of the *integrated* versions of the Kolmogorov–Smirnov and Cramér–von Mises tests $\hat{U}_n^{(I)}$ and $\hat{W}_n^{(I)}$ proposed in this chapter with that of the classical (nonintegrated) counterparts.

We constructed power curves for the case where the regression errors were sampled from either a skew normal distribution, Tukey’s g distribution, or the χ^2 distribution. The resulting power curves are shown in Figures 3.4–3.6. Comparing these plots with the power curves of the intercept-only case shown in Figures 3.1–3.3, we see very similar behaviour. Apart from good approximation of the nominal size, the integrated versions of the tests yield significantly higher power than their classical nonintegrated counterparts.

We now move on to a more extensive numerical study in which we compare the new tests with the following existing tests for symmetry:

- classical Kolmogorov–Smirnov (denoted by $\hat{U}_n$) and Cramér–von Mises ($\hat{W}_n$) tests described in Section 2.1.1;
- the test of Boos (1982), denoted by B , described in Section 2.1.2;
- a test by Cabilio and Masaro (1996) (CM) described in Section 2.3.3;
- a test based on triples (DQ) proposed independently by Davis and Quade (1978) and Randles et al. (1980) (see Section 2.3.1);
- a nonparametric test Z based on sample quantiles by Zardasht (2018) (see Section 2.3.5);

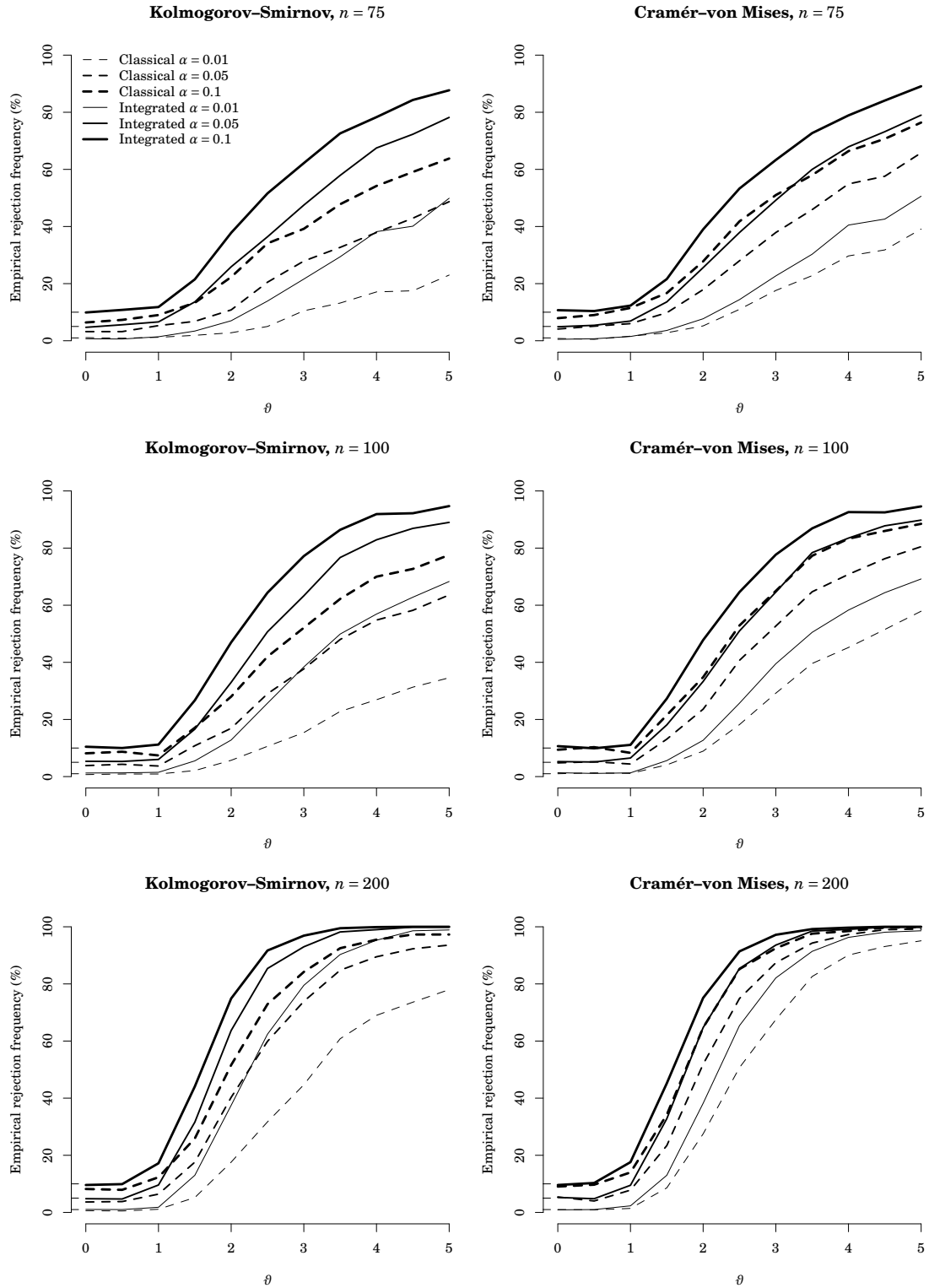


Figure 3.4: Power curves for the classical and integrated Kolmogorov–Smirnov and Cramér–von Mises tests for the case where the regression errors were sampled from a skew normal distribution.

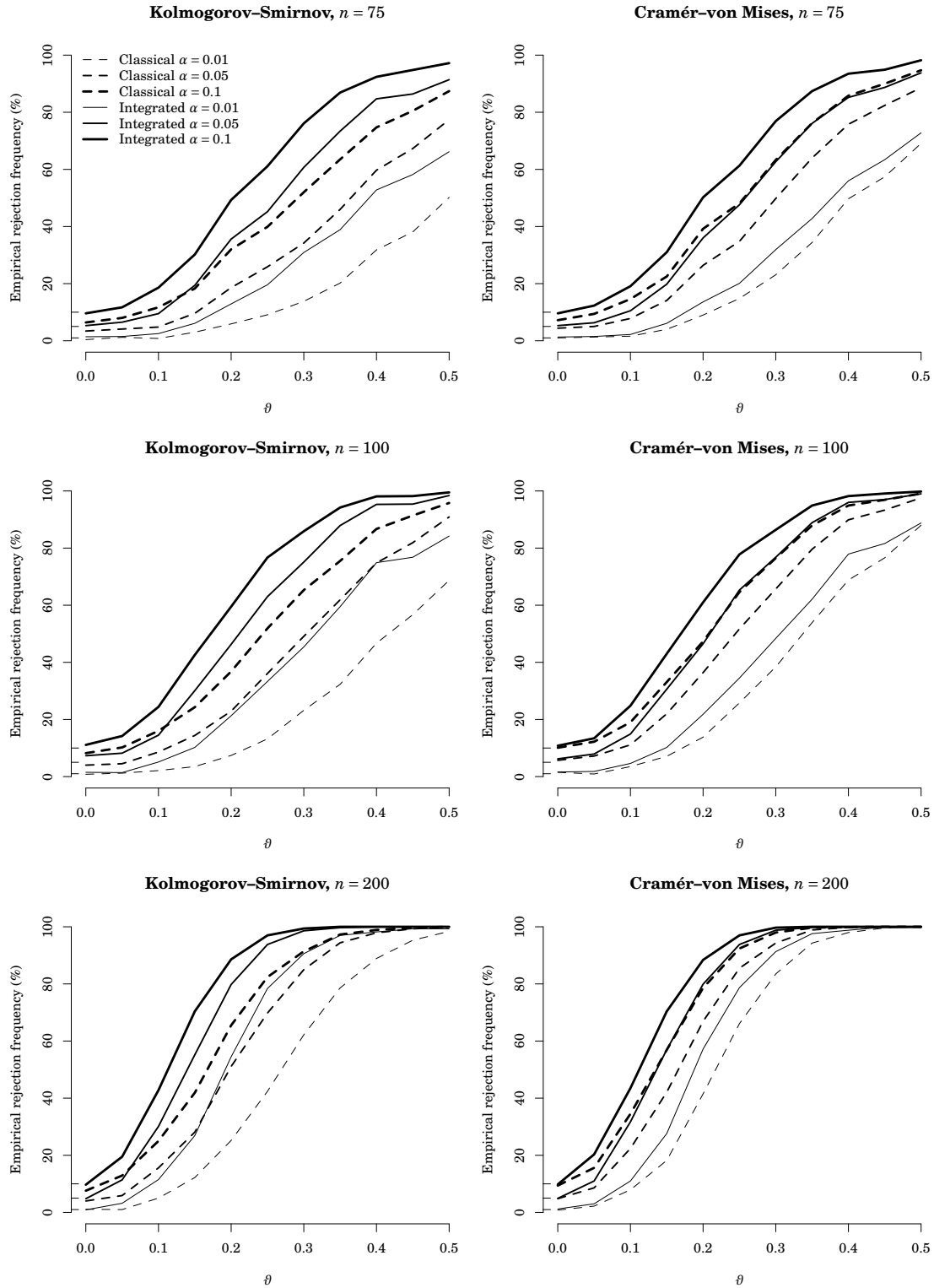


Figure 3.5: Power curves for the classical and integrated Kolmogorov–Smirnov and Cramér–von Mises tests for the case where the regression errors were sampled from Tukey’s g distribution.

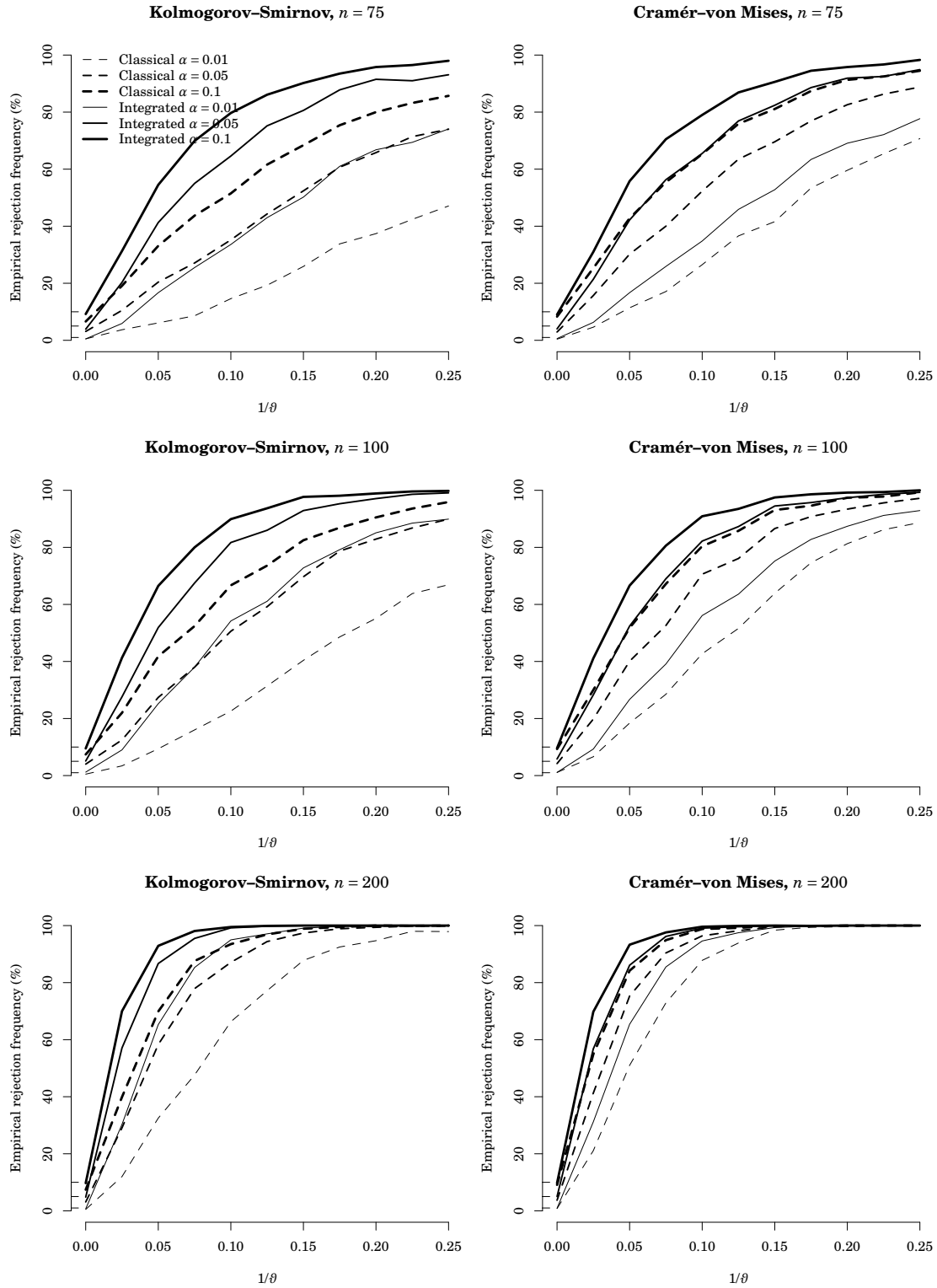


Figure 3.6: Power curves for the classical and integrated Kolmogorov–Smirnov and Cramér–von Mises tests for the case where the regression errors were sampled from the χ^2 distribution.

- the trimmed Wilcoxon test based on signed ranks (Antille et al., 1982) with tuning parameter λ , denoted by TW_λ (see Section 2.3.2);
- the test of Hušková and Meintanis (2012) based on the empirical characteristic function (see Section 2.2.1), where we use the weight function $w_a(t) = e^{-a|t|}$ (statistic labelled $H_a^{(1)}$) and the weight function $w_a(t) = e^{-at^2}$ (statistic labelled $H_a^{(2)}$); and
- two tests of Allison and Pretorius (2017) described in Sections 2.2.2 and 2.2.3 based on the empirical characteristic function and using weight function $w_a(t) = e^{-a|t|}$, denoted by $A_a^{(1)}$ and $A_a^{(2)}$ for the statistics defined in (2.13) and (2.16), respectively.

The critical values for all tests were obtained using the wild bootstrap procedure described earlier in this section, where we again use $B = 1000$ bootstrap replications to approximate the critical values.

Tables 3.4–3.9 present the rejection percentages of all tests (for significance levels $\alpha = 0.05, 0.1$) based on 2000 Monte Carlo iterations. The results show several desirable qualities of the newly-proposed tests. Firstly, both new tests based on $\hat{U}_n^{(I)}$ and $\hat{W}_n^{(I)}$ seem to respect the nominal level reasonably well, except in a few cases where the errors originate from a heavy-tailed distribution. Nevertheless, in these cases similar or more severe size distortions can be observed for several of the existing tests. Secondly, under alternatives the power of both tests seem to increase as the sample size is increased, which is in agreement with the consistency result given earlier. In addition, the power of both tests seem to be higher for distributions with higher skewness coefficients (e.g. compare the results for the χ_8^2 , χ_6^2 and χ_4^2 distributions, which have skewness coefficients of 1.00, 1.15 and 1.41, respectively).

In terms of power the new tests compare favourably with the considered existing tests. Notice that in many cases the new tests significantly outperform the classical Kolmogorov–Smirnov and Cramér–von Mises tests, while achieving power comparable and often even higher than the more modern tests, such as those based on the characteristic function.

Finally we wish to reiterate that our tests do not rely on any tuning parameters like many of the modern tests. This, together with the simplicity of calculating the statistics $\hat{U}_n^{(I)}$ and $\hat{W}_n^{(I)}$ in (3.4), makes the new tests straightforward to implement. It is clear from the results that the choice of such parameters could have a significant effect on the finite-sample performance of tests, both in terms of size and power. Moreover, choosing these tuning parameters in an ‘optimal’ data-driven way is by no means an easy task, and remains an open question in the literature (see,

e.g., Tenreiro, 2009). In light of this, and keeping in mind its good finite-sample size and power properties, the newly-proposed tests seem to be attractive alternative tests from a practical perspective.

Avenues for future research

Using the integrated empirical distribution function in goodness-of-fit tests for symmetry clearly has finite-sample advantages in terms of power. As ideas for future research we mention the following:

- generalising the proposed tests to the multivariate case to test for radial or spherical symmetry;
- studying the limit behaviour (under the null and alternatives) in the case where the goal is to test for symmetry of errors in a nonparametric regression model, as was done by Neumeyer and Dette (2007) and Hušková and Meintanis (2012);
- deriving the local approximate Bahadur efficiencies of the new tests as was done for several other tests by Milošević and Obradović (2019).

Table 3.4: Rejection percentages for symmetric error distributions (at a significance level of $\alpha = 5\%$).

Dist.	ϑ	n	$\hat{U}_n$	$\hat{W}_n$	B	CM	DQ	Z	TW_0	$TW_{.05}$	$TW_{.17}$	$H_{0.5}^{(1)}$	$H_1^{(1)}$	$H_{1.5}^{(1)}$	$H_{0.5}^{(2)}$	$H_1^{(2)}$	$H_{1.5}^{(2)}$	$A_{0.5}^{(1)}$	$A_1^{(1)}$	$A_{1.5}^{(1)}$	$A_{0.5}^{(2)}$	$A_1^{(2)}$	$A_{1.5}^{(2)}$	$\hat{U}_n^{(I)}$	$\hat{W}_n^{(I)}$
t_ϑ	∞	50	5.2	7.3	7.0	5.2	6.1	4.6	5.2	5.3	6.0	7.3	7.6	7.4	7.3	7.2	7.7	4.6	4.1	3.8	7.6	7.3	7.7	7.1	7.0
		75	5.0	5.2	5.4	5.4	5.2	3.6	4.8	5.3	5.0	5.6	5.9	5.8	5.8	5.7	5.9	4.4	4.0	3.7	6.4	6.2	6.2	6.0	5.7
		100	4.2	5.5	5.4	4.8	6.0	4.0	5.7	5.6	5.5	5.8	5.9	6.0	6.2	6.4	6.4	3.8	3.1	3.0	6.2	6.1	6.2	5.8	6.3
	8	50	4.6	6.0	5.9	4.4	5.5	5.4	6.6	6.6	6.2	6.0	6.5	6.7	6.7	7.3	7.8	6.0	6.0	6.2	6.8	6.8	7.3	6.6	6.8
		75	5.1	7.0	6.8	4.8	6.8	6.2	7.7	7.7	7.4	6.4	7.0	7.3	7.6	8.0	8.2	5.4	5.6	5.4	7.0	7.0	7.1	7.0	7.5
		100	5.0	5.4	5.3	4.4	5.1	5.2	6.1	6.2	5.2	5.7	5.4	5.3	5.4	6.0	6.5	4.8	5.3	5.3	5.2	5.8	6.3	5.7	5.4
	4	50	5.5	7.7	7.3	4.8	6.3	8.0	8.7	8.8	7.7	7.4	7.6	8.0	8.5	9.6	10.3	6.0	6.7	7.6	7.0	7.7	8.4	8.9	8.0
		75	5.6	6.6	6.0	5.0	5.8	8.2	8.3	8.4	7.4	5.4	6.3	7.2	7.5	8.2	8.7	5.3	6.0	6.2	6.5	6.7	7.2	7.6	7.2
		100	5.8	7.2	7.2	5.0	6.0	6.8	8.1	7.8	7.4	7.0	7.0	7.2	7.0	7.3	8.2	4.8	5.3	5.3	6.1	6.3	6.7	7.4	7.2
	U	50	4.7	7.0	6.8	7.3	7.1	5.4	2.3	2.6	5.3	7.8	8.2	7.9	8.2	7.7	7.7	5.7	4.0	3.6	10.4	9.2	8.2	8.2	8.2
		75	5.9	6.8	7.3	7.8	6.4	5.3	2.2	2.4	4.3	7.9	8.0	7.7	7.6	7.4	7.2	5.0	3.6	3.2	8.5	7.8	7.6	7.5	7.2
		100	5.2	6.4	6.6	5.5	6.3	3.9	2.0	2.7	4.7	6.7	6.6	7.0	7.0	6.5	6.4	5.4	3.8	3.5	8.6	7.8	7.3	6.8	6.6
	L	50	5.4	6.8	6.3	5.5	6.2	6.0	6.8	7.0	7.1	6.7	7.2	7.3	7.6	8.2	8.3	5.1	5.0	5.0	6.2	6.8	6.9	7.5	7.3
		75	6.2	6.5	6.3	4.8	6.2	5.2	7.2	7.2	7.0	7.0	6.9	7.0	7.2	7.2	7.3	4.9	4.7	4.8	5.7	6.1	6.3	7.0	6.8
		100	5.6	6.2	6.4	4.0	6.0	4.9	6.6	6.4	6.9	6.2	6.8	6.6	6.3	6.6	6.8	5.0	4.7	4.9	6.2	6.0	6.0	6.4	6.3
NM_ϑ	2	50	4.7	6.1	6.0	5.2	5.7	4.7	5.0	5.6	5.4	5.8	6.2	6.3	6.4	6.8	6.6	5.0	4.2	4.1	7.2	6.9	6.5	6.4	6.3
		75	4.8	5.7	6.2	4.3	5.8	5.0	6.2	6.3	6.6	6.0	6.5	6.5	6.2	6.8	6.8	4.5	4.3	4.0	6.2	6.3	6.8	6.2	6.5
		100	4.8	5.4	5.7	4.6	5.7	4.2	6.0	5.8	5.6	5.9	5.4	5.8	6.0	6.3	6.4	3.6	3.9	3.8	5.6	5.8	6.2	6.2	5.8
	3	50	4.7	7.2	6.7	5.0	6.4	5.1	6.0	6.3	6.7	7.0	7.4	7.6	7.7	7.7	7.9	5.3	4.8	4.6	7.2	7.2	7.3	7.6	7.4
		75	5.2	6.2	5.7	4.8	5.3	5.2	6.4	6.8	6.6	5.6	6.0	6.2	6.2	6.8	7.0	5.3	4.8	4.6	5.6	6.0	6.9	6.3	7.0
		100	5.0	5.9	5.5	5.2	5.7	4.5	6.4	6.2	6.2	5.5	5.4	6.0	6.2	6.6	6.8	4.2	4.3	4.2	5.4	5.4	5.8	6.5	6.0
	4	50	5.3	7.2	7.2	5.1	6.2	5.4	7.0	7.3	7.3	7.2	7.8	7.8	7.8	8.2	8.7	5.8	5.2	4.8	7.2	7.8	7.7	8.1	7.7
		75	5.9	6.7	6.4	5.0	6.0	5.3	7.6	8.0	7.2	5.8	6.6	6.8	7.0	7.3	7.8	5.6	5.0	5.2	6.2	6.2	7.0	7.4	7.4
		100	5.2	6.1	5.8	5.3	6.0	4.6	6.8	6.6	6.8	5.5	5.8	6.3	6.4	7.0	6.9	4.6	4.3	4.0	5.3	5.2	5.7	6.5	6.2

Table 3.5: Rejection percentages for symmetric error distributions (at a significance level of $\alpha = 10\%$).

Dist.	ϑ	n	$\hat{U}_n$	$\hat{W}_n$	B	CM	DQ	Z	TW_0	$TW_{.05}$	$TW_{.17}$	$H_{0.5}^{(1)}$	$H_1^{(1)}$	$H_{1.5}^{(1)}$	$H_{0.5}^{(2)}$	$H_1^{(2)}$	$H_{1.5}^{(2)}$	$A_{0.5}^{(1)}$	$A_1^{(1)}$	$A_{1.5}^{(1)}$	$A_{0.5}^{(2)}$	$A_1^{(2)}$	$A_{1.5}^{(2)}$	$\hat{U}_n^{(I)}$	$\hat{W}_n^{(I)}$
t_ϑ	0	50	9.7	11.8	11.8	10.0	11.8	10.2	10.6	11.1	11.2	12.5	12.6	12.6	12.8	13.4	13.4	10.1	9.9	9.4	12.8	13.6	13.2	12.7	12.6
		75	8.8	10.8	10.6	10.2	10.2	10.1	9.9	10.3	10.3	11.2	10.8	11.1	10.8	11.9	12.2	9.1	8.8	8.8	11.9	11.8	11.8	11.2	10.8
		100	8.1	10.5	10.9	9.5	10.3	8.9	10.2	10.5	10.1	10.9	11.1	11.0	10.9	11.7	11.6	8.8	8.3	7.9	11.5	11.8	12.1	10.7	10.5
	8	50	9.5	11.7	11.2	8.9	11.7	11.8	11.7	11.8	11.7	12.0	11.8	12.8	12.7	13.8	14.4	11.8	11.3	11.7	12.2	12.3	12.3	12.4	12.8
		75	10.4	12.4	11.8	10.8	12.0	11.0	12.3	12.4	12.6	11.9	12.7	12.6	13.1	13.2	13.7	10.6	10.4	10.9	12.1	12.2	12.5	12.8	12.7
		100	9.6	11.2	10.4	9.2	9.9	10.8	11.0	11.4	11.6	10.4	10.8	11.1	10.8	11.4	11.7	10.4	10.3	10.4	9.8	11.2	11.8	11.1	11.1
	4	50	9.8	12.3	12.3	10.8	10.8	15.0	14.3	14.1	13.0	12.7	12.9	13.8	13.8	15.1	16.5	11.8	12.8	13.2	12.8	13.6	14.5	14.4	13.8
		75	10.7	11.4	11.1	10.1	10.6	13.3	13.2	13.2	12.8	10.7	11.5	12.1	12.5	13.6	14.1	10.8	11.3	12.6	11.9	12.1	12.6	12.6	12.5
		100	10.5	12.0	11.3	10.2	10.3	12.8	12.6	12.6	12.4	11.3	11.3	11.9	12.2	13.0	13.2	9.8	11.2	11.9	10.9	11.5	11.7	12.3	11.8
	U	50	9.8	13.1	13.1	12.9	12.5	10.5	5.7	6.7	9.9	13.8	14.3	14.2	14.2	13.7	13.6	10.2	8.6	7.9	16.0	15.0	14.4	14.1	13.2
		75	10.5	12.8	11.8	13.1	11.6	10.3	5.5	6.2	9.4	13.1	13.1	12.9	12.6	12.2	11.9	10.3	8.1	7.5	14.2	12.5	12.2	12.7	12.4
		100	9.4	12.3	12.4	11.1	11.8	8.4	5.2	6.1	10.0	12.5	12.4	12.7	12.6	12.6	12.6	10.3	8.3	7.8	14.6	13.4	12.9	12.7	12.5
	L	50	10.7	11.8	11.8	11.0	11.5	11.6	12.4	12.9	12.4	11.8	12.6	12.9	12.8	13.8	14.7	10.6	11.2	11.2	10.7	11.6	12.3	13.4	13.6
		75	11.6	13.3	12.4	10.4	11.3	11.2	12.3	12.8	11.8	12.8	12.4	12.2	12.2	12.4	12.9	10.3	10.7	10.4	11.4	12.3	12.9	11.9	12.7
		100	9.9	11.4	11.2	9.2	11.7	10.8	11.8	11.9	11.8	11.2	11.7	12.3	12.2	12.3	12.3	9.2	9.4	9.9	11.3	11.8	11.8	12.1	12.1
NM_ϑ	2	50	9.4	11.8	11.3	10.2	10.8	10.4	10.4	10.4	11.4	11.7	12.5	12.6	12.2	12.6	12.8	10.1	9.7	9.0	13.2	13.4	13.0	12.7	12.2
		75	10.2	10.9	10.8	9.9	11.2	10.6	11.0	11.2	10.8	10.8	11.2	12.2	12.2	11.7	11.7	9.2	9.1	9.1	10.6	11.3	11.9	11.7	11.6
		100	9.6	10.2	10.3	10.1	10.2	9.7	11.6	11.4	11.3	10.4	10.5	11.0	11.2	11.4	11.9	8.9	8.8	8.8	10.9	10.2	10.6	11.4	11.2
	3	50	10.1	12.0	11.8	9.9	11.4	10.2	11.4	11.8	11.9	12.2	12.5	12.7	12.5	13.7	14.1	11.8	10.9	11.3	12.6	13.3	13.6	13.6	12.7
		75	9.8	11.5	11.2	9.6	11.2	11.2	11.6	11.8	11.5	11.2	11.6	11.7	11.6	12.1	12.2	10.3	9.7	10.0	11.2	11.3	11.4	11.9	11.7
		100	9.7	10.6	10.4	9.9	10.2	9.7	11.2	11.6	11.2	10.7	10.3	10.8	11.0	11.8	12.2	9.2	8.9	9.2	11.4	10.9	11.1	11.6	11.2
	4	50	10.2	12.2	12.3	10.3	11.3	11.3	12.2	12.4	12.4	12.6	13.2	12.9	13.5	14.6	14.9	11.1	10.9	11.0	12.7	13.6	14.2	13.6	13.1
		75	10.7	11.8	11.4	10.2	10.8	11.1	11.9	12.4	11.6	11.2	11.6	11.7	11.9	12.3	12.8	10.4	10.1	10.1	11.4	12.1	11.9	12.2	12.1
		100	9.9	10.8	10.5	10.0	10.2	9.8	12.0	11.8	11.4	10.9	10.2	10.7	11.1	11.7	11.6	9.4	9.1	9.4	10.8	10.8	11.3	11.2	11.1

Table 3.6: Rejection percentages for asymmetric error distributions (at a significance level of $\alpha = 5\%$).

Dist.	ϑ	n	$\hat{U}_n$	$\hat{W}_n$	B	CM	DQ	Z	TW_0	$TW_{.05}$	$TW_{.17}$	$H_{0.5}^{(1)}$	$H_1^{(1)}$	$H_{1.5}^{(1)}$	$H_{0.5}^{(2)}$	$H_1^{(2)}$	$H_{1.5}^{(2)}$	$A_{0.5}^{(1)}$	$A_1^{(1)}$	$A_{1.5}^{(1)}$	$A_{0.5}^{(2)}$	$A_1^{(2)}$	$A_{1.5}^{(2)}$	$\hat{U}_n^{(I)}$	$\hat{W}_n^{(I)}$
χ_ϑ^2	4	50	54.8	76.4	80.2	54.6	82.3	80.6	76.3	79.0	81.5	74.4	82.6	85.5	86.4	87.9	88.4	72.6	81.8	84.0	69.6	77.3	80.8	86.4	85.3
		75	82.3	93.0	95.2	73.7	96.8	96.6	94.7	96.2	96.7	92.1	96.4	97.8	98.0	98.7	98.7	91.0	95.6	96.3	90.2	93.3	95.0	97.7	97.8
		100	90.6	97.9	98.9	84.2	99.1	99.4	99.3	99.4	99.3	97.7	99.1	99.6	99.8	99.8	99.8	97.2	99.1	99.4	97.2	98.3	98.6	99.5	99.5
	6	50	41.4	59.4	64.3	40.0	67.7	66.0	60.9	63.9	65.2	57.9	66.8	70.8	72.8	75.2	76.7	58.4	66.0	68.2	55.0	62.9	68.2	71.9	71.4
		75	63.8	80.7	85.8	56.6	87.7	85.3	87.4	89.4	88.3	78.7	87.2	89.6	90.3	92.0	92.2	79.8	86.1	87.1	78.2	85.1	88.6	90.4	90.1
		100	79.5	92.8	95.0	71.7	95.8	96.0	96.4	97.2	95.9	90.6	95.2	96.7	97.4	98.3	98.5	91.8	96.2	96.8	91.0	94.4	96.2	97.1	97.1
	8	50	30.9	46.7	50.6	29.4	54.9	51.3	49.2	51.5	51.8	43.8	53.2	57.6	59.7	63.0	64.2	44.6	51.4	52.9	43.6	51.9	56.5	59.8	58.5
		75	51.9	69.3	74.8	45.6	76.9	77.3	78.3	80.3	77.6	66.0	76.2	80.2	81.6	84.7	85.5	71.1	77.2	79.0	67.0	75.2	79.6	81.3	80.7
		100	66.0	82.4	87.1	58.8	89.1	89.7	90.4	92.0	89.3	79.6	88.0	90.7	92.1	93.7	94.0	82.9	88.6	89.6	80.7	87.0	89.8	91.7	91.3
	10	50	25.6	38.8	42.5	25.0	46.8	42.5	42.8	44.9	45.2	36.7	44.9	49.4	51.6	54.4	55.2	38.1	43.3	43.9	37.0	43.3	46.9	50.5	50.1
		100	55.6	72.7	78.7	49.2	81.4	81.8	84.5	85.8	82.4	67.7	79.2	83.3	84.9	87.7	89.5	73.2	80.2	81.2	70.3	79.7	84.0	85.2	84.3
SN_ϑ	1	50	4.3	6.0	6.3	4.9	6.7	4.5	5.1	5.2	5.8	6.5	7.1	7.5	7.2	7.7	7.4	4.9	4.3	4.0	6.0	6.5	6.8	7.1	7.2
		75	5.8	6.8	6.9	5.4	6.6	5.8	6.4	7.0	7.0	7.1	7.2	7.3	7.3	7.8	7.8	5.6	5.0	4.5	7.0	6.9	7.3	7.4	7.3
		100	6.0	7.3	7.7	5.7	8.6	5.9	8.4	8.8	8.6	6.8	7.7	8.3	9.0	9.2	9.3	5.2	5.2	5.0	6.7	7.2	8.2	8.8	8.9
	2	50	10.7	15.9	16.0	11.4	17.9	14.4	15.9	17.0	17.4	14.6	17.6	19.8	20.1	21.6	22.0	12.6	13.2	12.9	13.6	15.8	18.0	20.2	19.6
		75	16.7	23.6	24.8	14.2	26.8	22.2	25.8	26.7	27.0	20.6	25.2	28.1	28.9	30.5	30.8	17.7	19.0	18.9	19.9	23.5	26.0	29.0	28.6
		100	21.1	28.2	30.7	18.2	33.6	29.6	34.6	35.1	34.0	25.0	31.4	34.3	35.4	38.0	39.0	23.0	25.6	25.6	25.2	31.4	34.0	35.8	35.5
	3	50	20.4	30.2	34.2	20.2	37.0	31.1	31.8	34.0	34.4	29.4	35.4	38.8	40.3	42.4	43.7	24.4	26.7	27.6	25.9	33.0	37.4	39.9	39.5
		75	34.4	45.8	48.9	28.4	53.7	45.0	51.8	53.6	52.5	42.0	50.7	54.5	56.4	59.9	61.0	36.0	41.5	42.3	39.5	48.0	52.6	55.8	55.9
		100	43.6	58.6	62.2	41.3	65.6	60.3	65.7	66.6	65.0	54.5	63.1	66.4	67.8	72.2	73.2	46.6	52.4	53.6	51.4	60.1	65.2	68.4	67.7

Table 3.7: Rejection percentages for asymmetric error distributions (at a significance level of $\alpha = 5\%$).

Dist.	ϑ	n	$\hat{U}_n$	$\hat{W}_n$	B	CM	DQ	Z	TW_0	$TW_{.05}$	$TW_{.17}$	$H_{0.5}^{(1)}$	$H_1^{(1)}$	$H_{1.5}^{(1)}$	$H_{0.5}^{(2)}$	$H_1^{(2)}$	$H_{1.5}^{(2)}$	$A_{0.5}^{(1)}$	$A_1^{(1)}$	$A_{1.5}^{(1)}$	$A_{0.5}^{(2)}$	$A_1^{(2)}$	$A_{1.5}^{(2)}$	$\hat{U}_n^{(I)}$	$\hat{W}_n^{(I)}$
AL_ϑ	1	50	7.4	10.2	8.6	5.0	6.6	7.9	11.0	11.2	10.2	8.6	8.8	9.0	9.5	10.7	11.3	5.9	6.6	6.8	7.4	8.8	9.6	9.4	9.1
		75	7.5	8.1	6.9	5.0	5.8	7.8	8.6	8.2	8.2	7.0	6.8	7.0	7.0	7.4	8.1	6.0	6.3	6.4	5.5	6.0	6.5	7.0	6.9
		100	7.5	8.2	7.4	5.7	6.1	7.0	8.1	7.8	8.2	7.4	7.3	7.1	6.9	7.2	7.4	5.8	6.3	6.6	5.7	6.5	6.7	7.4	6.8
	1.5	50	51.5	61.9	61.2	44.4	58.4	52.8	59.7	59.6	61.1	58.4	62.7	63.1	63.0	62.1	62.6	26.4	35.1	38.2	37.8	41.2	46.2	61.5	64.2
		75	72.7	81.1	79.6	66.0	77.9	66.2	77.5	77.3	79.6	77.5	80.2	79.6	78.0	76.6	76.3	31.2	42.4	47.4	54.2	58.3	61.2	77.8	80.4
		100	86.7	91.4	90.9	81.0	89.0	77.2	88.6	88.3	89.6	89.6	90.2	90.1	89.6	87.6	86.6	38.4	49.8	55.2	68.8	71.0	73.8	88.6	90.6
	2	50	77.1	87.3	88.2	71.4	88.6	83.9	84.7	85.2	88.2	86.4	89.7	90.7	91.0	91.6	91.8	54.9	69.0	73.7	68.3	73.6	78.1	90.7	90.8
		75	93.4	97.7	98.0	88.4	98.2	93.4	97.0	97.0	98.3	97.5	98.2	98.5	98.3	98.2	98.2	68.0	80.7	85.2	88.6	90.2	92.0	98.3	98.5
		100	98.0	99.6	99.7	96.2	99.7	97.7	99.3	99.4	99.7	99.4	99.7	99.8	99.7	99.7	99.6	76.6	88.0	91.3	96.3	96.8	97.1	99.7	99.8
TU_ϑ	0.15	50	11.2	15.1	16.0	11.0	17.0	13.6	15.2	15.7	16.1	14.4	17.0	18.4	18.8	19.7	20.0	12.4	13.2	13.2	14.3	16.1	17.3	18.9	18.9
		75	13.9	18.7	19.9	12.0	22.1	19.3	21.9	22.8	22.4	17.0	20.7	23.4	24.1	26.7	27.8	14.7	16.6	17.2	17.0	20.4	23.0	24.8	24.2
		100	16.8	23.2	25.2	14.3	27.5	24.5	29.5	30.4	28.1	20.2	25.1	27.8	29.3	33.1	34.4	20.1	21.8	23.0	21.0	26.1	28.6	29.3	29.7
	0.3	50	27.4	39.1	43.6	26.6	46.6	43.7	43.4	45.5	45.5	36.9	46.0	49.9	51.5	54.6	56.2	33.3	39.1	41.8	34.4	41.2	46.4	51.0	50.3
		75	40.8	55.5	60.3	35.4	62.7	60.3	64.8	66.6	63.2	50.8	62.1	66.2	68.3	71.3	73.3	49.5	57.9	60.4	49.9	58.4	63.8	68.6	67.0
		100	53.0	69.2	73.7	45.7	77.2	76.6	80.2	81.1	78.0	64.0	74.8	80.0	81.0	84.6	85.4	62.8	71.2	73.6	63.4	72.3	77.7	80.5	80.2
	0.5	50	57.8	76.8	80.0	55.9	80.7	79.1	78.2	80.0	81.4	75.8	82.5	85.0	85.9	87.5	88.0	65.2	74.6	78.6	64.7	71.2	74.4	85.0	84.5
		75	80.3	92.5	94.2	69.9	94.6	94.0	95.2	96.0	95.2	91.2	94.9	96.2	96.8	97.3	97.5	85.2	92.6	94.1	86.8	90.2	92.2	96.3	96.3
		100	92.3	98.0	98.7	86.4	98.8	99.0	99.1	99.1	99.1	97.7	98.8	99.2	99.4	99.6	99.7	94.0	98.1	98.8	95.3	96.8	97.3	99.4	99.2

Table 3.8: Rejection percentages for asymmetric error distributions (at a significance level of $\alpha = 10\%$).

Dist.	ϑ	n	$\hat{U}_n$	$\hat{W}_n$	B	CM	DQ	Z	TW_0	$TW_{.05}$	$TW_{.17}$	$H_{0.5}^{(1)}$	$H_1^{(1)}$	$H_{1.5}^{(1)}$	$H_{0.5}^{(2)}$	$H_1^{(2)}$	$H_{1.5}^{(2)}$	$A_{0.5}^{(1)}$	$A_1^{(1)}$	$A_{1.5}^{(1)}$	$A_{0.5}^{(2)}$	$A_1^{(2)}$	$A_{1.5}^{(2)}$	$\hat{U}_n^{(I)}$	$\hat{W}_n^{(I)}$
χ_ϑ^2	4	50	70.2	83.8	87.3	66.6	89.6	89.6	86.3	87.8	88.8	83.1	89.3	91.8	92.6	93.7	93.8	83.0	89.7	91.2	79.0	85.1	88.0	92.0	92.0
		75	90.9	96.8	98.0	84.4	98.8	98.9	98.2	98.6	98.5	96.1	98.6	99.1	99.3	99.4	99.4	95.4	97.5	98.1	94.4	96.6	97.7	99.1	99.2
		100	95.8	99.2	99.6	91.3	99.7	99.8	99.6	99.8	99.8	99.1	99.8	99.8	99.8	99.9	99.9	99.0	99.6	99.8	98.7	99.1	99.2	99.8	99.8
	6	50	55.6	70.6	74.8	52.8	78.8	79.2	74.7	78.0	77.2	68.9	77.3	81.0	82.8	85.2	86.1	71.8	79.7	81.0	67.2	75.2	79.6	82.3	81.4
		75	75.8	89.0	91.4	68.1	92.8	93.3	94.0	94.5	93.3	87.8	92.1	94.0	94.7	95.7	96.0	88.4	92.9	93.6	87.2	91.4	93.1	94.6	94.0
		100	89.0	96.3	97.3	82.0	98.2	98.4	98.8	99.1	98.4	95.3	97.8	98.8	99.0	99.3	99.4	96.1	98.6	99.0	94.9	97.0	98.2	98.9	98.8
	8	50	43.1	58.5	62.3	41.5	67.0	66.9	65.4	68.6	66.5	56.0	64.8	70.2	72.6	75.8	77.4	60.1	67.0	69.2	56.9	64.5	68.7	71.8	70.8
		75	65.7	80.2	83.7	59.0	85.9	86.4	86.9	88.3	86.4	77.3	84.7	87.8	88.7	91.1	91.7	81.3	86.2	87.6	76.9	84.4	87.0	88.9	88.0
		100	77.8	89.4	92.8	71.0	93.8	94.9	95.9	96.4	94.8	88.4	93.5	94.9	95.3	96.8	97.0	91.0	94.4	95.0	88.3	92.3	94.0	95.6	95.6
	10	50	36.3	51.1	55.5	37.6	59.4	58.4	56.8	58.8	57.8	48.8	57.6	61.6	63.4	66.7	67.8	51.9	57.4	59.2	48.4	55.4	60.8	63.1	62.1
		100	69.7	83.0	86.7	61.6	89.4	90.5	91.7	92.3	89.8	79.7	87.8	90.9	91.8	94.0	94.6	83.2	88.6	89.6	81.4	87.7	91.1	92.0	91.8
SN_ϑ	1	50	8.7	11.1	11.7	10.7	11.2	11.3	10.5	10.8	11.2	12.3	12.6	12.7	12.9	13.2	13.7	10.8	10.0	9.7	12.2	12.4	12.8	12.8	12.9
		75	10.8	12.8	12.7	10.9	13.3	11.7	12.9	13.3	13.2	12.7	12.9	13.2	13.8	14.3	14.4	10.7	10.7	10.6	12.3	12.9	13.1	13.6	13.5
		100	11.2	13.1	13.4	10.7	13.4	12.1	14.4	14.7	13.9	12.6	13.7	14.2	14.6	16.0	16.3	11.7	11.7	11.6	12.5	13.6	14.7	14.2	14.2
	2	50	17.4	24.9	25.8	19.0	28.6	25.5	25.4	26.4	27.8	23.6	27.5	29.8	30.2	31.2	32.1	21.3	23.0	23.4	22.0	25.8	28.0	30.8	30.9
		75	26.0	33.1	34.9	24.0	37.5	34.8	39.5	40.4	39.0	31.1	34.9	38.5	39.6	42.2	43.0	27.6	30.7	31.2	29.2	34.8	38.0	40.0	39.9
		100	31.7	40.2	42.5	29.7	46.1	43.0	47.2	47.8	45.5	36.8	43.9	46.9	47.9	51.3	52.0	34.1	37.7	38.5	36.0	41.5	44.9	48.1	48.5
	3	50	30.0	42.5	45.5	31.5	49.5	45.0	45.5	46.8	47.4	40.0	47.7	52.0	52.5	54.5	55.5	35.5	39.9	41.2	38.1	44.9	49.4	52.1	52.2
		75	47.8	58.0	60.9	40.9	65.6	61.6	65.5	67.3	66.0	53.9	62.4	67.1	68.6	71.6	73.2	51.3	56.3	57.4	52.3	59.8	65.4	68.2	68.7
		100	57.9	69.9	73.3	52.9	77.4	73.0	78.0	79.1	76.4	67.0	73.9	78.1	79.4	81.9	82.8	60.1	66.2	67.6	64.3	72.2	76.3	78.8	79.8

Table 3.9: Rejection percentages for asymmetric error distributions (at a significance level of $\alpha = 10\%$).

Dist.	ϑ	n	$\hat{U}_n$	$\hat{W}_n$	B	CM	DQ	Z	TW_0	$TW_{.05}$	$TW_{.17}$	$H_{0.5}^{(1)}$	$H_1^{(1)}$	$H_{1.5}^{(1)}$	$H_{0.5}^{(2)}$	$H_1^{(2)}$	$H_{1.5}^{(2)}$	$A_{0.5}^{(1)}$	$A_1^{(1)}$	$A_{1.5}^{(1)}$	$A_{0.5}^{(2)}$	$A_1^{(2)}$	$A_{1.5}^{(2)}$	$\hat{U}_n^{(I)}$	$\hat{W}_n^{(I)}$
AL $_{\vartheta}$	1	50	13.6	15.8	14.1	10.7	12.3	13.4	17.4	17.2	16.1	14.2	14.8	15.3	15.2	16.2	17.0	11.1	12.2	12.6	13.9	13.7	14.6	15.3	14.7
		75	12.4	13.4	11.6	10.6	10.8	13.0	14.1	14.1	12.8	12.5	11.8	11.9	11.8	12.9	13.3	10.6	11.2	12.1	11.2	11.1	11.0	12.2	11.9
		100	11.8	12.6	11.4	9.9	10.4	11.9	12.8	12.5	12.4	11.9	11.6	11.6	11.6	11.8	11.9	10.3	10.2	10.8	10.9	10.8	10.9	12.1	11.8
	1.5	50	65.4	71.0	70.3	59.4	69.1	64.2	69.2	69.3	71.2	68.7	71.1	71.4	71.2	71.4	71.5	36.4	45.0	49.4	48.5	52.9	56.8	70.7	72.2
		75	81.6	87.2	86.6	76.2	86.2	77.0	85.0	84.8	87.4	85.7	87.2	86.6	85.8	84.2	83.2	42.7	53.2	57.7	66.5	69.2	72.0	85.2	87.0
		100	92.0	94.2	94.2	88.9	93.8	85.2	93.0	92.8	93.6	93.8	94.3	93.6	93.3	92.2	91.2	50.7	61.5	65.7	79.3	81.9	83.0	92.7	94.2
	2	50	86.0	92.3	93.2	81.7	93.9	90.8	91.5	91.6	93.5	92.0	94.3	95.2	95.0	95.6	95.7	67.3	78.7	83.3	78.9	82.2	84.2	94.7	95.4
		75	96.8	98.8	99.2	92.9	99.2	97.5	98.7	98.7	99.4	98.7	99.3	99.3	99.3	99.2	99.2	76.8	87.2	91.4	93.7	94.5	95.4	99.2	99.3
		100	99.0	99.8	99.8	98.0	100.0	98.9	99.7	99.7	99.9	99.8	99.9	100.0	100.0	100.0	99.8	84.6	92.0	94.5	98.1	98.0	98.2	100.0	100.0
TU $_{\vartheta}$	0.15	50	17.1	24.0	25.1	19.6	27.0	24.4	25.6	26.1	26.6	23.1	26.7	28.2	28.8	31.3	31.6	20.7	22.4	22.4	21.9	25.0	26.1	29.4	29.0
		75	22.4	28.3	30.6	20.5	33.2	31.4	34.2	35.2	33.4	26.0	30.8	34.4	35.5	38.9	40.1	26.0	29.0	29.6	25.9	30.6	34.0	34.3	35.9
		100	25.8	33.8	35.6	23.0	38.8	37.0	41.4	42.5	39.0	30.8	36.3	39.2	40.5	44.2	45.4	30.6	33.5	34.4	31.0	36.1	39.5	41.1	40.5
	0.3	50	38.5	52.7	56.2	39.0	58.7	59.0	58.5	60.0	59.0	49.8	58.4	62.5	64.2	67.1	68.3	46.8	54.2	56.0	46.0	53.9	59.1	63.8	63.4
		75	55.0	67.9	72.4	47.2	74.6	75.0	75.9	77.6	75.2	64.4	73.2	77.9	79.3	82.2	83.2	64.2	71.6	73.9	62.2	70.5	75.2	79.2	78.8
		100	67.0	79.9	83.3	57.9	85.3	86.8	88.3	88.9	86.3	76.2	84.0	87.2	88.8	91.0	91.6	76.7	83.4	84.5	74.2	81.6	85.8	89.0	87.6
	0.5	50	71.3	84.7	86.3	68.3	88.2	88.7	86.3	87.4	88.1	83.1	89.0	91.0	91.8	92.6	93.0	76.5	84.7	87.1	74.8	80.4	84.2	90.6	91.0
		75	89.0	95.6	96.9	80.6	97.3	97.4	97.7	98.2	97.5	95.0	97.3	98.0	98.5	98.8	98.8	91.7	96.3	97.2	91.6	94.2	95.1	98.2	98.0
		100	96.7	99.2	99.4	92.6	99.6	99.8	99.6	99.7	99.7	98.8	99.6	99.8	99.8	99.8	99.8	97.2	99.3	99.5	97.6	98.4	98.6	99.8	99.8

Chapter 4

Tests based on smooth distribution function estimation

Kernel distribution function estimation has been extensively studied in the literature, starting with early references such as Nadaraya (1964), Winter (1973) and Yamato (1973). Reiss (1981) showed that the relative deficiency of the nonsmooth empirical distribution function with respect to an appropriately chosen kernel-type distribution function estimator tends to infinity as $n \rightarrow \infty$ (also see Falk, 1983). It was only recently, however, that the benefits of using kernel-smoothed *distribution* function estimators in goodness-of-fit tests started drawing attention. It was pointed out by Racine and van Keilegom (2020), who developed a smooth test for independence in a mixed-data nonparametric regression model context, that substantial power gains can be achieved by substituting kernel-smoothed distribution function estimators for the nonsmooth empirical distribution functions appearing in classical Kolmogorov–Smirnov and Cramér–von Mises tests.

In this chapter we investigate the effects of using kernel distribution function estimators when testing for symmetry. We show in a numerical study that the smooth tests not only respect the nominal level under the null hypothesis, but could also yield significantly higher power under alternatives. Apart from obtaining the null distribution of the test statistics and establishing consistency of the tests, we also provide a theoretical justification for the observed power gains.

We now give an overview of the theoretical concepts that underlie the main ideas of this chapter. Consider the Kolmogorov–Smirnov and Cramér–von Mises type functionals

$$U_n(F) = \sup_{t \geq 0} |S_n(t; F)|, \quad W_n(F) = \int_0^\infty S_n^2(t; F) dQ(t),$$

where S_n denotes the symmetry process defined as

$$S_n(t; F) = \sqrt{n}(F(t) + F(-t-) - 1)$$

and $Q(t) = Q(t; F) := P(|\varepsilon_1| \leq t) = F(t) - F(-t-)$. Clearly, if F is the distribution function of a random variable that is symmetric around zero, one has $S_n(t; F) = 0$ for all t and all n , and consequently also $U_n(F) = W_n(F) = 0$.

Suppose we observe a random sample $\varepsilon_1, \dots, \varepsilon_n$ generated from F . As sample versions of $U_n(F)$ and $W_n(F)$ one might consider the statistics

$$U_n(F_n) = \sup_{t \geq 0} |S_n(t; F_n)| = \sqrt{n} \sup_{t \geq 0} |F_n(t) + F_n(-t-) - 1|,$$

$$W_n(F_n) = \int_0^\infty S_n^2(t; F_n) dQ_n(t) = n \int_0^\infty (F_n(t) + F_n(-t-) - 1)^2 dQ_n(t),$$

where F_n is simply the empirical distribution function

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\varepsilon_i \leq t), \quad t \in \mathbb{R},$$

and $Q_n(t) = Q_n(t; F_n) = F_n(t) - F_n(-t-)$. Tests for symmetry based on $U_n(F_n)$ and $W_n(F_n)$ have been studied extensively in the literature in various contexts; see, e.g., Rothman and Woodroffe (1972); Baringhaus and Henze (1992); Neumeyer et al. (2005). One can expect that, typically, $U_n(F_n)$ and $W_n(F_n)$ will be close to zero under the null hypothesis of symmetry, and large under alternatives. Intuitively these statistics should then be suitable for use in testing for symmetry.

In this chapter we consider an alternative approach that involves using a kernel-smoothed estimator of the distribution function F instead of the nonsmooth empirical distribution function F_n . To this end, define the function

$$F_n^h(t) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{t - \varepsilon_i}{h_n}\right) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^t \frac{1}{h_n} k\left(\frac{s - \varepsilon_i}{h_n}\right) ds, \quad t \in \mathbb{R}, \quad (4.1)$$

where $h_n \in \mathbb{R}^+$ is a bandwidth and k is a symmetric kernel with corresponding distribution function K , i.e. $k = K'$.

It is well known in the literature that, under appropriate conditions, F_n^h has the asymptotic representation

$$F_n^h(t) = F_n(t) + \frac{1}{2} \kappa_2 f'(t) h_n^2 + o_P(h_n^2), \quad (4.2)$$

as $n \rightarrow \infty$, where $\kappa_2 = \int z^2 k(z) dz$. From this it follows that

$$S_n(t; F_n^h) = S_n(t; F_n) + \frac{1}{2} \kappa_2 (f'(t) + f'(-t)) \sqrt{n} h_n^2 + o_P(1). \quad (4.3)$$

Clearly, whenever f is a density symmetric around zero, one has that $f'(t) = -f'(-t)$ and hence the bias term disappears completely so that

$$S_n(t; F_n^h) = S_n(t; F_n) + o_P(1).$$

This suggests that symmetry tests based on the unsmoothed process $S_n(t; F_n)$ and the smoothed process $S_n(t; F_n^h)$ will typically have the same behaviour under the null hypothesis.

4.1 Linear regression

We consider the situation where the objective is to test whether the distribution of error terms appearing in a regression model is symmetric around zero. We assume that $(Y; \mathbf{x})$ is an observation generated by the multivariate linear model

$$Y = \mathbf{x}^\top \boldsymbol{\beta} + \varepsilon, \quad (4.4)$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown regression parameters and the regressors $\mathbf{x} \in \mathbb{R}^p$ are assumed fixed (i.e. nonrandom). The null hypothesis of interest is

$$H_0 : F(t) + F(-t) = 1 \quad \text{for all } t \geq 0,$$

where F denotes the distribution function of ε .

Suppose $(Y_1; \mathbf{x}_1), \dots, (Y_n; \mathbf{x}_n)$ is a random sample generated by the model in (4.4). Since the errors $\varepsilon_1, \dots, \varepsilon_n$ are unobserved, we base our test on the residuals

$$\hat{\varepsilon}_i = \varepsilon_i - \mathbf{x}_i^\top (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}), \quad i = 1, 2, \dots, n,$$

where $\hat{\boldsymbol{\beta}}_n$ is an appropriate estimator for $\boldsymbol{\beta}$. Based on these residuals, we define the kernel-smoothed distribution function estimator of F by

$$\hat{F}_n^h(t) = \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^t k\left(\frac{s - \hat{\varepsilon}_i}{h_n}\right) ds = \frac{1}{n} \sum_{i=1}^n K\left(\frac{t - \hat{\varepsilon}_i}{h_n}\right), \quad \forall t \in \mathbb{R}, \quad (4.5)$$

where $h_n \in \mathbb{R}^+$ is a bandwidth and k is a symmetric kernel with corresponding distribution function K , i.e. $k = K'$. Now, analogous to the test criteria $U_n(F_n)$ and $W_n(F_n)$ defined earlier, we introduce the smooth test criteria

$$\begin{aligned} U_n(\hat{F}_n^h) &= \sqrt{n} \sup_{t \geq 0} \left| \hat{F}_n^h(t) + \hat{F}_n^h(-t) - 1 \right|, \\ W_n(\hat{F}_n^h) &= n \int_0^\infty \left(\hat{F}_n^h(t) + \hat{F}_n^h(-t) - 1 \right)^2 d\hat{Q}_n^h(t), \end{aligned} \quad (4.6)$$

where $\hat{Q}_n^h(t) = Q_n(t; \hat{F}_n^h) = \hat{F}_n^h(t) - \hat{F}_n^h(-t)$. As with the classical tests, a goodness-of-fit test for symmetry of the errors around zero rejects for large values of $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$.

4.1.1 Asymptotic behaviour of the test statistics

Our asymptotic analysis is based on the following formal assumptions on the distribution of the model errors:

(E1) $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed with continuous, positive density f .

(E2) f' is Hölder continuous, i.e. there exist constants $C, \alpha > 0$ such that

$$|f'(s) - f'(t)| \leq C|s - t|^\alpha \quad \text{for all } s, t \in \mathbb{R}.$$

The assumptions above are quite mild and are typical in the related literature (see, e.g., Koul, 2002; Racine and van Keilegom, 2020). We note that Condition (E2) implies that f' is continuous on $\mathbb{R}$.

As a choice for the kernel K , we allow absolutely continuous distribution functions with corresponding density k satisfying the following assumption:

(K1) k is a density function that is symmetric around zero and $k(u) = 0$ whenever $|u| > a$ for some $a > 0$. Furthermore, k is q times continuously differentiable with $k^{(j)}(a) = k^{(j)}(-a) = 0$ for all $j = 0, \dots, q$.

Assumption (K1) defines a class of continuously differentiable kernels of bounded support. We note that such kernels have finite moments. Specifically, letting

$$\kappa_j = \int z^j k(z) dz = \int z^j dK(z),$$

it is straightforward to verify that

$$\kappa_j = 0 \text{ whenever } j \text{ odd,}$$

and

$$\kappa_j < \infty \text{ whenever } j \text{ even.}$$

A popular example of such a kernel is the Epanechnikov kernel (see Li and Racine, 2007, p. 35) defined by

$$k(u) = \begin{cases} \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}u^2\right) & \text{if } u^2 < 5, \\ 0 & \text{otherwise.} \end{cases} \quad (4.7)$$

Other popular kernels satisfying Assumption (K1) are given in Table 4.1, along with their corresponding density and distribution functions.

Under these conditions, one can verify that the expansion in (4.2) is valid. This is formalised in the following lemma.

Lemma 4.1. *Suppose that Conditions (E1), (E2) and (K1) hold. If $h_n = cn^{-\gamma}$ for some $c > 0$ and some $\gamma \in (0, 1)$, then the expansion in (4.2) is valid uniformly in $t \in \mathbb{R}$.*

This result is not new and is stated here merely for completeness. We also provide a proof in Section A.1 of the Appendix, which illustrates how results like these rely on the modulus of continuity of the empirical process (see Stute, 1982).

Table 4.1: Three well-known kernels of bounded support satisfying Assumption (K1).

Kernel	$K(u)$	$k(u)$	Support
Epanechnikov	$\frac{3}{4\sqrt{5}} \left(\frac{2\sqrt{5}}{3} + u - \frac{u^3}{15} \right)$	$\frac{3}{4\sqrt{5}} \left(1 - \frac{u^2}{5} \right)$	$u^2 < 5$
Biweight	$\frac{15}{16} \left(\frac{8}{15} + u - \frac{2}{3}u^3 + \frac{1}{5}u^5 \right)$	$\frac{15}{16} (1 - u^2)^2$	$ u < 1$
Triweight	$\frac{35}{32} \left(\frac{16}{35} + u - u^3 + \frac{3}{5}u^5 - \frac{1}{7}u^7 \right)$	$\frac{35}{32} (1 - u^2)^3$	$ u < 1$

We are now ready to study the limit behaviour of the process $\hat{F}_n^h(t)$ based on the observed residuals. The following assumptions are required.

(R1) $\max_{1 \leq i \leq n, 1 \leq v \leq p} |x_{iv}| = o(n^{1/2})$, where x_{iv} is the v th element of the vector x_i , and $n^{-1} \mathbf{X}_n^\top \mathbf{X}_n \rightarrow \mathbf{C}$ for some positive definite matrix $\mathbf{C}$.

(R2) The estimator $\hat{\beta}_n$ satisfies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta) = \frac{d_\beta}{\sqrt{n}} \sum_{i=1}^n \psi_\beta(\varepsilon_i) + o_P(1),$$

as $n \rightarrow \infty$, where d_β is some constant and ψ_β is a measurable antisymmetric function satisfying $E(\psi_\beta(\varepsilon_1)) = 0$ and $E(\psi_\beta^2(\varepsilon_1)) < \infty$.

As discussed in Section 2.5, these assumptions are fairly standard in the regression literature. Assumption (R2) is general in the sense that we do not specify any specific type of estimator, but include several important special cases as was pointed out in Section 2.5.

We may now state and prove the following result.

Theorem 4.1. *Suppose that Assumptions (E1), (E2), (R1) and (R2) hold, and that the bandwidth sequence satisfies $\sqrt{n}h_n^2 = O(1)$. Then the expansion*

$$\hat{F}_n^h(t) = F_n(t) + f(t) \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^\top (\hat{\beta}_n - \beta) + \frac{1}{2} \kappa_2 f'(t) h_n^2 + o_P(n^{-1/2})$$

is valid uniformly in $t \in \mathbb{R}$, where $\kappa_2 = \int z^2 k(z) dz$.

Proof. Using integration by parts, first observe that

$$\begin{aligned}
\widehat{F}_n^h(t) &= \frac{1}{n} \sum_{i=1}^n K\left(\frac{t - \widehat{\varepsilon}_i}{h_n}\right) \\
&= \int_{-\infty}^{\infty} K\left(\frac{t-x}{h_n}\right) d\widehat{F}_n(x) \\
&= \left[K\left(\frac{t-x}{h_n}\right) \widehat{F}_n(x) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \widehat{F}_n(x) dK\left(\frac{t-x}{h_n}\right) \\
&= \frac{1}{h_n} \int_{-\infty}^{\infty} \widehat{F}_n(x) k\left(\frac{t-x}{h_n}\right) dx \\
&= \int_{-\infty}^{\infty} \widehat{F}_n(t - h_n z) dK(z).
\end{aligned}$$

By Corollary 2.3.5 of Koul (2002, p. 63), we may write

$$\widehat{F}_n(t) = F_n(t) + f(t) \frac{1}{n} \sum_{i=1}^n x_i^\top (\widehat{\beta}_n - \beta) + o_P(n^{-1/2}),$$

uniformly in $t \in \mathbb{R}$, provided that Assumptions (R1) and (E1) are satisfied. Therefore, since $\int_{-\infty}^{\infty} k(z) dz = 1$,

$$\widehat{F}_n^h(t) = \int_{-\infty}^{\infty} F_n(t - h_n z) k(z) dz + \frac{1}{n} \sum_{i=1}^n x_i^\top (\widehat{\beta}_n - \beta) \int_{-\infty}^{\infty} f(t - h_n z) k(z) dz + o_P(n^{-1/2}),$$

uniformly in $t \in \mathbb{R}$, or

$$\begin{aligned}
\widehat{F}_n^h(t) &= F_n^h(t) + \frac{1}{n} \sum_{i=1}^n x_i^\top (\widehat{\beta}_n - \beta) \int_{-\infty}^{\infty} f(t - h_n z) k(z) dz + o_P(n^{-1/2}) \\
&= F_n^h(t) + \frac{1}{n} \sum_{i=1}^n x_i^\top (\widehat{\beta}_n - \beta) \left(f(t) + \frac{1}{2} h_n^2 \kappa_2 f''(t) \right) + O(n^{-1}) + o_P(n^{-1/2}) \\
&= F_n^h(t) + f(t) \frac{1}{n} \sum_{i=1}^n x_i^\top (\widehat{\beta}_n - \beta) + o_P(n^{-1/2}) \\
&= F_n(t) + f(t) \frac{1}{n} \sum_{i=1}^n x_i^\top (\widehat{\beta}_n - \beta) + \frac{1}{2} h_n^2 \kappa_2 f'(t) + o_P(n^{-1/2}),
\end{aligned}$$

where the last step follows from (4.2). $\square$

From Theorem 4.1 we obtain one of the main results of this chapter, which has an important implication for the limit null distributions of our tests. Observe that, under the assumptions of Theorem 4.1, the empirical symmetry process has the representation

$$\begin{aligned}
S_n(t; \widehat{F}_n^h) &= S_n(t; F_n) + (f(t) + f(-t)) \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^\top (\widehat{\beta}_n - \beta) \\
&\quad + \frac{1}{2} \kappa_2 (f'(t) + f'(-t)) \sqrt{n} h_n^2 + o_P(1). \quad (4.8)
\end{aligned}$$

Compare this to the expansion of the nonsmooth symmetry process $S_n(t; \hat{F}_n)$ given at the top of p. 711 of Neumeyer et al. (2005). Note that the expression on p. 711 of Neumeyer et al. (2005) contains a typographical error (missing square root in the denominator of the penultimate term), and for that reason we reproduce the correct expression here:

$$S_n(t; \hat{F}_n) = S_n(t; F_n) + (f(t) + f(-t)) \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta) + o_P(1). \quad (4.9)$$

Comparing (4.8) and (4.9), we see that the smoothed version in (4.8) contains an extra bias term.

Under symmetry of the error distribution, one has the relation $f'(-t) = -f'(t)$. This means that, under H_0 ,

$$S_n(t; \hat{F}_n^h) = S_n(t; F_n) + 2f(t) \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta) + o_P(1), \quad (4.10)$$

uniformly in $t \in \mathbb{R}$. Therefore, under symmetry of the error distribution, the bias term $\frac{1}{2} \kappa_2(f'(t) + f'(-t)) h_n^2$ disappears and hence has no effect on the asymptotic behaviour of $S_n(\hat{F}_n^h)$. Comparing (4.9) and (4.10), we see that, under H_0 ,

$$\sup_{t \geq 0} |S_n(t; \hat{F}_n^h) - S_n(t; \hat{F}_n)| \xrightarrow{P} 0. \quad (4.11)$$

This implies that, under H_0 , smoothing of the empirical distribution function typically has no effect on the limit behaviour of the classical Kolmogorov–Smirnov and Cramér–von Mises tests.

4.1.2 Consistency

We now show that, under the stated assumptions, the smoothed tests based on $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$ are consistent against general alternatives. First note that, under alternatives, we have from (4.8) that

$$\begin{aligned} \frac{1}{\sqrt{n}} S_n(t; \hat{F}_n^h) &= \frac{1}{\sqrt{n}} S_n(t; F_n) + (f(t) + f(-t)) \frac{1}{n} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta) \\ &\quad + \frac{1}{2} \kappa_2(f'(t) + f'(-t)) h_n^2 + o_P(n^{-1/2}). \end{aligned}$$

Since it is assumed that $h_n^2 = O(n^{-1/2})$, we get that

$$\begin{aligned} \frac{1}{\sqrt{n}} S_n(t; \hat{F}_n^h) &= \frac{1}{\sqrt{n}} S_n(t; F_n) + (f(t) + f(-t)) \frac{1}{n} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta) + O_P(n^{-1/2}) \\ &= \frac{1}{\sqrt{n}} S_n(t; \hat{F}_n) + o_P(1), \end{aligned}$$

uniformly in $t \in \mathbb{R}$, provided that $\kappa_2 < \infty$ and $\sup_{t \in \mathbb{R}} |f'(t)| < \infty$.

By the Glivenko–Cantelli lemma,

$$\sup_{t \in \mathbb{R}} \left| \frac{1}{\sqrt{n}} S_n(t; \hat{F}_n) - \frac{1}{\sqrt{n}} S_n(t; F) \right| \xrightarrow{P} 0.$$

Looking at Assumption (R2), we see that $n^{-1/2} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta)$ is asymptotically normally distributed by the central limit theorem. This means that

$$\frac{1}{n} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta) = O_P(n^{-1/2})$$

and, together with the fact that $\sup_{t \in \mathbb{R}} |f(t)| < \infty$, we have

$$\sup_{t \in \mathbb{R}} \left| (f(t) + f(-t)) \frac{1}{n} \sum_{i=1}^n x_i^\top (\hat{\beta}_n - \beta) \right| \leq 2 \sup_{t \in \mathbb{R}} |f(t)| O_P\left(\frac{1}{\sqrt{n}}\right) \xrightarrow{P} 0.$$

Therefore,

$$\frac{1}{\sqrt{n}} S_n(t; \hat{F}_n^h) = \frac{1}{\sqrt{n}} S_n(t; F) + o_P(1) = F(t) + F(-t) - 1 + o_P(1), \quad (4.12)$$

uniformly in $t \in \mathbb{R}$.

If H_0 is not true, then we know from the definition of symmetry around zero that there exists at least one $t^* \in \mathbb{R}$ such that $F(t^*) + F(-t^*) - 1 = K^* \neq 0$. Hence, under alternatives,

$$\begin{aligned} U_n(\hat{F}_n^h) &= \sup_{t \in \mathbb{R}} |S_n(t; \hat{F}_n^h)| \\ &= \sqrt{n} \sup_{t \in \mathbb{R}} |F(t) + F(-t) - 1| + o_P(1) \\ &\geq \sqrt{n} |F(t^*) + F(-t^*) - 1| + o_P(1) \\ &= \sqrt{n} K^* + o_P(1) \\ &\xrightarrow{P} \infty, \end{aligned}$$

so that the test based on $U_n(\hat{F}_n^h)$ is consistent.

To see that the Cramér–von Mises type test based on $W_n(\hat{F}_n^h)$ is consistent, first note that

$$\begin{aligned} \frac{1}{n} W_n(\hat{F}_n^h) &= \frac{1}{n} \int_0^\infty S_n^2(t; \hat{F}_n^h) d\hat{Q}_n^h(t) \\ &= \frac{1}{n} \int_0^\infty S_n^2(t; F) d\hat{Q}_n^h(t) + \frac{1}{n} \int_0^\infty \left(S_n(t; \hat{F}_n^h) - S_n(t; F) \right)^2 d\hat{Q}_n^h(t) \\ &\quad + \frac{1}{n} \int_0^\infty S_n(t; F) \left(S_n(t; \hat{F}_n^h) - S_n(t; F) \right) d\hat{Q}_n^h(t) \\ &= \frac{1}{n} \int_0^\infty S_n^2(t; F) d\hat{Q}_n^h(t) + o_P(1), \end{aligned}$$

where the last step follows from the uniform convergence in (4.12), the fact that $n^{-1/2} S_n(t; F) \leq 1$ for all $n \geq 1$ and all $t \in \mathbb{R}$, and $\int_0^\infty d\hat{Q}_n^h(t) = 1$. Now, since

$\sup_{t \in \mathbb{R}} |\widehat{Q}_n^h(t) - Q(t)| \xrightarrow{P} 0$ and $n^{-1/2}S_n(\cdot; F)$ is bounded and continuous, we have by the Helly–Bray theorem (van der Vaart, 2000, p. 6) that

$$\begin{aligned} \frac{1}{n}W_n(\widehat{F}_n^h) &= \frac{1}{n} \int_0^\infty S_n^2(t; F) dQ(t) + o_P(1) \\ &= \int_0^\infty (F(t) + F(-t) - 1)^2 dQ(t) + o_P(1). \end{aligned}$$

As the integrand on the right is nonnegative, the integral will be zero if and only if $F(t) + F(-t) - 1 = 0$ for all $t \geq 0$, i.e. under the null hypothesis. Therefore, under alternatives, we have that $W_n(\widehat{F}_n^h) \xrightarrow{P} \infty$ as $n \rightarrow \infty$ and the test is consistent.

Neumeyer et al. (2005, see p. 703) argued that, under the stated assumptions, the classical nonsmooth tests based on $U_n(\widehat{F}_n)$ and $W_n(\widehat{F}_n)$ are consistent against contiguous alternatives converging to the null hypothesis at a rate of $n^{-1/2}$. Since the expansions in (4.8) and (4.9) are asymptotically equivalent up to the bias term, we cannot directly use the result of Neumeyer et al. (2005) to claim that the smooth tests are also consistent against contiguous alternatives. It is expected however, that a similar result will hold under appropriate conditions on the bandwidth.

4.2 Theoretical analysis

Racine and van Keilegom (2020) claim that the improvement in power seen in the smoothed versions of the Kolmogorov–Smirnov and Cramér–von Mises tests can be ascribed to a reduction in the (integrated) variance of the test criterion. In this section we show that, also in our setting, the smoothed symmetry process does indeed have lower (integrated) variance than the nonsmooth counterpart.

To simplify our calculations and to facilitate exposition, we consider only the case where we test whether a random sample $\varepsilon_1, \dots, \varepsilon_n$ has a distribution symmetric around a *known* median η , which we assume to be zero without any loss of generality. For the case where η is unknown, one would require an asymptotic representation such as the one given in Proposition 3.1 of Zhao, Gijbels and van Keilegom (2020).

We restrict our attention to contiguous alternatives and consider the case where $\varepsilon_1, \dots, \varepsilon_n$ is a random sample from the density

$$g_n(t) = \left(1 + \frac{\nu}{\sqrt{n}}u(t)\right)f_0(t), \quad (4.13)$$

where f_0 is a density symmetric around 0, $\nu \neq 0$ and $u(\cdot)$ is a measurable function with the following properties:

$$\int u(t)f_0(t)dt = 0 \quad \text{and} \quad 0 < \int u^2(t)f_0(t)dt = 0 < \infty.$$

To facilitate notation, we shall denote the nonsmooth and smoothed estimator of the symmetry process by

$$\mathbb{S}_n(t) := S_n(t; F_n) = \sqrt{n}(F_n(t) + F_n(-t-) - 1) \quad (4.14)$$

and

$$\mathbb{S}_n^h(t) := S_n(t; F_n^h) = \sqrt{n}(F_n^h(t) + F_n^h(-t-) - 1), \quad (4.15)$$

respectively. Here, F_n denotes the empirical distribution function constructed from the ε_i and, as before, F_n^h denotes the smoothed counterpart given by

$$F_n^h(t) = \int_{-\infty}^{\infty} F_n(t - hz)k(z)dz = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \mathbf{1}(\varepsilon_i \leq t - hz)k(z)dz.$$

The processes in (4.14) and (4.15) both estimate the true symmetry process $S_n(t; G_n) = \sqrt{n}(G_n(t) + G_n(-1) - 1)$, where G_n is the distribution function corresponding to the density defined in (4.13). The mean integrated square error of $\mathbb{S}_n$ is therefore given by

$$\text{MISE}(\mathbb{S}_n) = \int_{-\infty}^{\infty} \mathbb{E}(S_n(t; F_n) - S_n(t; G_n))^2 dt = \int_{-\infty}^{\infty} \text{Var}(\mathbb{S}_n(t)) dt,$$

where we have made use of the fact the $S_n(t; F_n)$ is an unbiased estimator of $S_n(t; G_n)$. On the other hand, the smoothed process $\mathbb{S}_n^h(t)$ is a biased estimator of $S_n(t; G_n)$ and the mean integrated square error is

$$\text{MISE}(\mathbb{S}_n^h) = \int_{-\infty}^{\infty} \mathbb{E}(S_n(t; F_n^h) - S_n(t; G_n))^2 dt = \int_{-\infty}^{\infty} \text{Var}(\mathbb{S}_n^h(t)) dt + n \int_{-\infty}^{\infty} b_n^2(t) dt, \quad (4.16)$$

where

$$\sqrt{n}b_n(t) = \mathbb{E}(\mathbb{S}_n^h(t)) - S_n(t; G_n). \quad (4.17)$$

The following theorem shows that the mean integrated square error of the smoothed symmetry process $\mathbb{S}_n^h$ is smaller than that of the nonsmooth process $\mathbb{S}_n$ which arises mainly due to a reduction in variance. This leads to power gains that are evident in the simulations reported in Section 4.3.

Theorem 4.2. *Suppose that Conditions (E1), (E2) and (K1) hold. If the bandwidth satisfies $\sqrt{n}h_n^2 = O(1)$, then*

$$\text{MISE}(\mathbb{S}_n^h) = \text{MISE}(\mathbb{S}_n) - 4h_n \Psi(K) + O(h_n^2),$$

where $\Psi(K) = \int uK(u)dK(u) > 0$.

Proof. Direct calculation (cf. Neumeyer et al., 2005, Theorem 3.1) shows that

$$\begin{aligned} \text{MSE}(\mathbb{S}_n(t)) &= \text{Var}(\mathbb{S}_n(t)) \\ &= G_n(t)(1 - G_n(t)) + G_n(-t)(1 - G_n(-t)) + 2[G_n(t \wedge (-t)) - G_n(t)G_n(-t)]. \end{aligned}$$

Hence, the mean integrated square error of $\mathbb{S}_n$ is given by

$$\begin{aligned}\text{MISE}(\mathbb{S}_n) &= \int_{-\infty}^{\infty} \text{Var}(\mathbb{S}_n(t)) dt \\ &= \int_{-\infty}^{\infty} \{G_n(t)(1 - G_n(t)) + G_n(-t)(1 - G_n(-t)) + 2[G_n(t \wedge (-t)) - G_n(t)G_n(-t)]\} dt \\ &= \int_0^{\infty} [G_n(t)(1 - G_n(t)) + G_n(-t)(1 - G_n(-t)) + 2G_n(-t)(1 - G_n(t))] dt.\end{aligned}$$

We now turn to the smoothed empirical symmetry process defined in (4.15), which can conveniently be decomposed as follows:

$$\begin{aligned}\mathbb{S}_n^h(t) &= \sqrt{n} \int_{-\infty}^{\infty} [F_n(t - hz) + F_n(-t - hz) - G_n(t - hz) - G_n(-t - hz)] k(z) dz + \sqrt{n} b_n(t) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_{-\infty}^{\infty} [\mathbf{1}(\varepsilon_i \leq t - hz) + \mathbf{1}(\varepsilon_i \leq -t - hz) - G_n(t - hz) - G_n(-t - hz)] k(z) dz \\ &\quad + \sqrt{n} b_n(t),\end{aligned}$$

where $b_n(t)$ appearing in the bias term is defined in (4.17). Observe that $b_n(t)$ is equal to

$$\begin{aligned}b_n(t) &= \int_{-\infty}^{\infty} [G_n(t - hz) + G_n(-t - hz)] k(z) dz - (G_n(t) + G_n(-t)) \\ &= \frac{1}{2} \kappa_2 (g'_n(t) + g'_n(-t)) h_n^2 + O(h_n^3).\end{aligned}$$

Now, since f_0 in (4.13) is symmetric around zero, we have that $f_0(t) = f_0(-t)$ and therefore also

$$g_n(t) - g_n(-t) = \frac{v}{\sqrt{n}} (u(t)f_0(t) - u(-t)f_0(-t)) = \frac{v}{\sqrt{n}} (u(t) - u(-t))f_0(t),$$

so that

$$g'_n(t) + g'_n(-t) = \frac{v}{\sqrt{n}} \frac{d}{dt} (u(t) - u(-t)) f_0(t).$$

Under the integrability condition $\int [\frac{d}{dt} u(t) f_0(t)]^2 dt < \infty$,

$$n \int_{-\infty}^{\infty} b_n^2(t) dt \sim \frac{1}{4} \kappa_2^2 v^2 h_n^4 = O(h_n^4).$$

Hence, the mean integrated square error of $\mathbb{S}_n^h$ in (4.16) becomes

$$\text{MISE}(\mathbb{S}_n^h) = \int_{-\infty}^{\infty} \text{Var}(\mathbb{S}_n^h(t)) dt + O(h_n^4).$$

We now derive an asymptotic expansion for the integrated variance term in the last display. Define $H(t) = \mathbf{1}(\varepsilon_1 \leq t) - G_n(t)$. Since the ε_i are independent and

identically distributed,

$$\begin{aligned}
& \text{Var}(\mathbb{S}_n^h(t)) \\
&= \text{Var}\left(\int_{-\infty}^{\infty} [\mathbf{1}(\varepsilon_1 \leq t - hz) + \mathbf{1}(\varepsilon_1 \leq -t - hz) - G_n(t - hz) - G_n(-t - hz)]k(z)dz\right) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Cov}(H(t - hz_1) + H(-t - hz_1), H(t - hz_2) + H(-t - hz_2))k(z_1)k(z_2)dz_1dz_2.
\end{aligned} \tag{4.18}$$

Straightforward calculations show that

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Cov}[H(t - hz_1), H(t - hz_2)]k(z_1)k(z_2)dz_1dz_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E}[H(t - hz_1)H(t - hz_2)]k(z_1)k(z_2)dz_1dz_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{E}[(\mathbf{1}(\varepsilon_1 \leq t - hz_1) - G_n(t - hz_1))(\mathbf{1}(\varepsilon_1 \leq t - hz_2) - G_n(t - hz_2))] \\
&\quad \times k(z_1)k(z_2)dz_1dz_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_n((t - hz_1) \wedge (t - hz_2)) - G_n(t - hz_1)G_n(t - hz_2)k(z_1)k(z_2)dz_1dz_2.
\end{aligned} \tag{4.19}$$

The first term in (4.19) becomes

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_n((t - hz_1) \wedge (t - hz_2))k(z_1)k(z_2)dz_1dz_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z_2} G_n(t - hz_2)k(z_1)k(z_2)dz_1dz_2 + \int_{-\infty}^{\infty} \int_{z_2}^{\infty} G_n(t - hz_1)k(z_1)k(z_2)dz_1dz_2 \\
&= G_n(t) \int_{-\infty}^{\infty} \int_{-\infty}^{z_2} k(z_1)k(z_2)dz_1dz_2 - hg_n(t) \int_{-\infty}^{\infty} \int_{-\infty}^{z_2} z_2k(z_1)k(z_2)dz_1dz_2 \\
&\quad + G_n(t) \int_{-\infty}^{\infty} \int_{z_2}^{\infty} k(z_1)k(z_2)dz_1dz_2 - hg_n(t) \int_{-\infty}^{\infty} \int_{z_2}^{\infty} z_1k(z_1)k(z_2)dz_1dz_2 + g'_n(t)O(h^2) \\
&= G_n(t) - hg_n(t) \int_{-\infty}^{\infty} \int_{-\infty}^{z_2} z_2k(z_1)k(z_2)dz_1dz_2 - hg_n(t) \int_{-\infty}^{\infty} \int_{z_2}^{\infty} z_1k(z_1)k(z_2)dz_1dz_2 \\
&\quad + g'_n(t)O(h^2).
\end{aligned}$$

Now, noting that

$$\Psi(k) = \int_{-\infty}^{\infty} \int_{-\infty}^{z_2} z_2k(z_1)k(z_2)dz_1dz_2 = \int_{-\infty}^{\infty} \int_{z_2}^{\infty} z_1k(z_1)k(z_2)dz_1dz_2,$$

we get

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_n((t - hz_1) \wedge (t - hz_2))k(z_1)k(z_2)dz_1dz_2 = G_n(t) - 2hg_n(t)\Psi(k) + g'_n(t)O(h^2). \tag{4.20}$$

Also,

$$\begin{aligned}
\int_{-\infty}^{\infty} G_n(t - hz_1)k(z_1)dz_1 &= G_n(t) + hg_n(t) \int_{-\infty}^{\infty} z_1k(z_1)dz_1 + g'_n(t)O(h^2) \\
&= G_n(t) + g'_n(t)O(h^2).
\end{aligned} \tag{4.21}$$

Substituting (4.20) and (4.21) in (4.19), we get that

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Cov}[H(t-hz_1), H(t-hz_2)] k(z_1) k(z_2) dz_1 dz_2 \\ = G_n(t)(1-G_n(t)) - 2hg_n(t)\Psi(k) + g'_n(t)O(h^2). \end{aligned} \quad (4.22)$$

In a similar fashion one can show that

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Cov}[H(-t-hz_1), H(-t-hz_2)] k(z_1) k(z_2) dz_1 dz_2 \\ = G_n(-t)(1-G_n(-t)) - 2hg_n(-t)\Psi(k) + g'_n(-t)O(h^2) \end{aligned} \quad (4.23)$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Cov}[H(t-hz_1), H(-t-hz_2)] k(z_1) k(z_2) dz_1 dz_2 \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Cov}[H(-t-hz_1), H(t-hz_2)] k(z_1) k(z_2) dz_1 dz_2 \\ = G_n(t \wedge (-t)) - G_n(t)G_n(-t) + g'_n(t)O(h^2). \end{aligned} \quad (4.24)$$

Therefore, using (4.22), (4.23) and (4.24), we have from (4.18) that

$$\text{Var}(\mathbb{S}_n^h(t)) = \text{Var}(\mathbb{S}_n(t)) - 2h_n(g_n(t) + g_n(-t))\Psi(k) + g'_n(t)O(h^2).$$

Integrating, we get that

$$\int_{-\infty}^{\infty} \text{Var}(\mathbb{S}_n^h(t)) dt = \int_{-\infty}^{\infty} \text{Var}(\mathbb{S}_n(t)) dt - 4h_n\Psi(k) + O(h_n^2). \quad \square$$

4.3 Numerical study

The theoretical results of Section 4.2 suggest that the smoothed symmetry process typically has a lower mean integrated square error than its nonsmooth counterpart, which should lead to an increase in power of the smooth tests due to more accurate approximation of the true underlying symmetry process. We now present the results from a numerical study that show that these power gains are quite significant in finite samples. An additional feature is that these power gains are obtained when using an optimal cross-validation bandwidth, which is attractive from a practical perspective in that the practitioner does not carry the burden of having to choose an appropriate bandwidth.

We first consider the case where $Y_1, \dots, Y_n$ is a random sample from a distribution with unknown median which needs to be estimated. In Section 4.3.2 we move on to the more general setting of linear regression where the goal is to test for symmetry of the unobserved regression errors.

In all simulations we used 1 000 independent Monte Carlo simulations.

4.3.1 Calculation of the test statistics

The test criteria in (4.6) are not in closed form and do not allow direct numerical evaluation in applications. For this reason we employ discretised versions of the test criteria which we now describe.

Let $0 = t_0 < t_1 < \dots < t_N < \infty$ be N appropriately chosen evaluation points. We approximate the test criteria as follows:

$$U_n(\hat{F}_n^h) = \sup_{t \geq 0} |S_n(t; \hat{F}_n^h)| \approx \max_{j=1, \dots, N} |S_n(t_j; \hat{F}_n^h)|,$$

and

$$\begin{aligned} W_n(\hat{F}_n^h) &= \int_0^\infty S_n^2(t; \hat{F}_n^h) d\hat{Q}_n^h(t) = \int_0^\infty S_n^2(t; \hat{F}_n^h) \hat{q}_n^h(t) dt \\ &\approx \frac{1}{N} \sum_{j=1}^N S_n^2(t_j; \hat{F}_n^h) \hat{q}_n^h(t_j) \Delta_j, \end{aligned}$$

where $\hat{q}_n^h(t) = \frac{d}{dt} \hat{Q}_n^h(t) = \hat{f}_n^h(t) + \hat{f}_n^h(-t)$ and $\Delta_j = t_j - t_{j-1}$. Recall that, since $\hat{F}_n^h(t)$ is absolutely continuous, the estimated kernel density estimator $\hat{f}_n^h$ has the closed-form expression

$$\hat{f}_n^h(t) = \frac{1}{nh_n} \sum_{i=1}^n k\left(\frac{t - \hat{\varepsilon}_i}{h_n}\right), \quad t \in \mathbb{R},$$

with $k = K'$.

In the simulations we chose $N = 500$ equally-spaced evaluation points and simply took $t_N = \frac{3}{2} \max_{j=1, \dots, n} |\hat{\varepsilon}_j|$. We found that choosing larger values of t_N does not have a noticeable effect on the numerical results. Alternatively one could make use of the following argument to choose the upper bound t_N . Since $F(t) = P(|\varepsilon_1| \leq t)$ is a true distribution function, we can find for any $\delta > 0$ an $t_N = t_N(\delta)$ such that $P(|\varepsilon_1| \leq t_N(\delta)) = F(t_N(\delta)) - F(-t_N(\delta)) > 1 - \delta$. The quantile $t_N(\delta)$ could then be estimated by the corresponding empirical quantile $\hat{t}_N(\delta)$ satisfying $\hat{F}_n^h(\hat{t}_N(\delta)) - \hat{F}_n^h(-\hat{t}_N(\delta)) > 1 - \delta$. We did not pursue this idea further.

4.3.2 Case I: Intercept-only model (no regressors)

We first consider the case of the mean (intercept-only) model where there are no regressors. Formally, we observe $Y_1, \dots, Y_n$ generated by the model

$$Y_i = \eta + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed zero-mean random variables and η is an unknown location parameter to be estimated. As an estimator of η we simply use $n^{-1} \sum_{i=1}^n Y_i$. Critical values of all tests were obtained using the

model-based bootstrap described in Section 2.6.1 using $B = 1000$ bootstrap replications.

The statistics in (4.6) require a choice of the kernel distribution function K and the smoothing parameter h_n . Throughout, we used the Epanechnikov kernel as defined in (4.7). As the optimal bandwidth in kernel-type distribution function estimation is of the form $h_n = cn^{-1/3}$ (see, e.g., Li, Li and Racine, 2017), we simply took this bandwidth choice with $c = 1$ for all experiments in this section.

Figures 4.1 to 4.3 show plots which compare the power curves of the smooth test (solid lines) with those of the nonsmooth tests (dashed lines) when the observations were sampled from a skew normal distribution, Tukey's g distribution and the asymmetric Laplace distribution, respectively. Overall it seems that the smooth tests respect the nominal level of the test reasonably well for all three considered significance levels $\alpha = 0.01, 0.05, 0.1$. Notice that in many cases the classical nonsmooth Kolmogorov–Smirnov test seems to be rather conservative (i.e. undersized), whereas the smooth Kolmogorov–Smirnov test has empirical size closer to the nominal level in many of the considered cases.

Under the alternatives (larger values of the parameter ϑ) all tests show a clear increase in power in most cases, which is certainly a desirable property for any goodness-of-fit test. Comparing the solid curves (smooth tests) to the dashed curves (nonsmooth tests), it is evident that there is a drastic improvement in power of the Kolmogorov–Smirnov test in almost all cases considered. Take for example the skew normal distribution with parameter $\vartheta = 5$. At the 5% level of significance for a sample size of $n = 75$, the smooth Kolmogorov–Smirnov test has power of around 60%, whereas the nonsmooth counterpart has power of only 40% meaning a 50% improvement in power.

The power gains for the Cramér–von Mises test are also clear, however to a slightly lesser degree than for the Kolmogorov–Smirnov test. For the example described in the previous paragraph, the improvement seen here is only around 12% (from about 58% to just over 65%). Nevertheless, this improvement is still significant. Interestingly, looking at Figure 4.3, one sees that there is almost no benefit in using the smoothed Cramér–von Mises test over its classical nonsmooth counterpart in the case of the asymmetric Laplace distributions. This shows that significant improvement in power is not available uniformly over all distributions. We point out that even in this case, the smooth tests have empirical size close to that of the classical tests.

We highlight another feature of the smooth tests which is also clear from the plots. In Figures 4.1 to 4.3 it can be seen that the power gains diminish as the sample size is increased. This is not surprising, since, as shown in Theorem 4.2, the

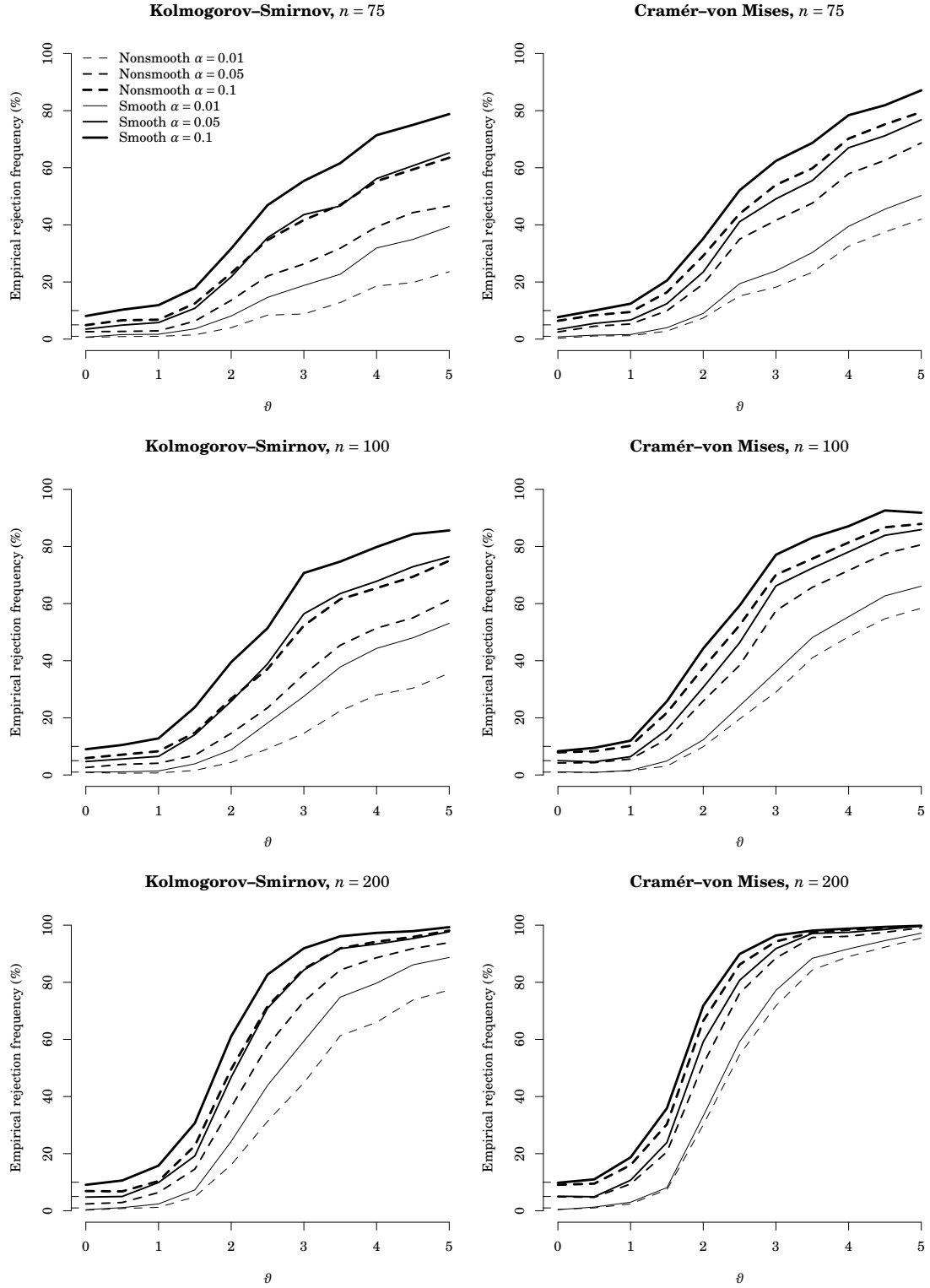


Figure 4.1: Power curves for the skew normal distribution in the case where there are no regressors.

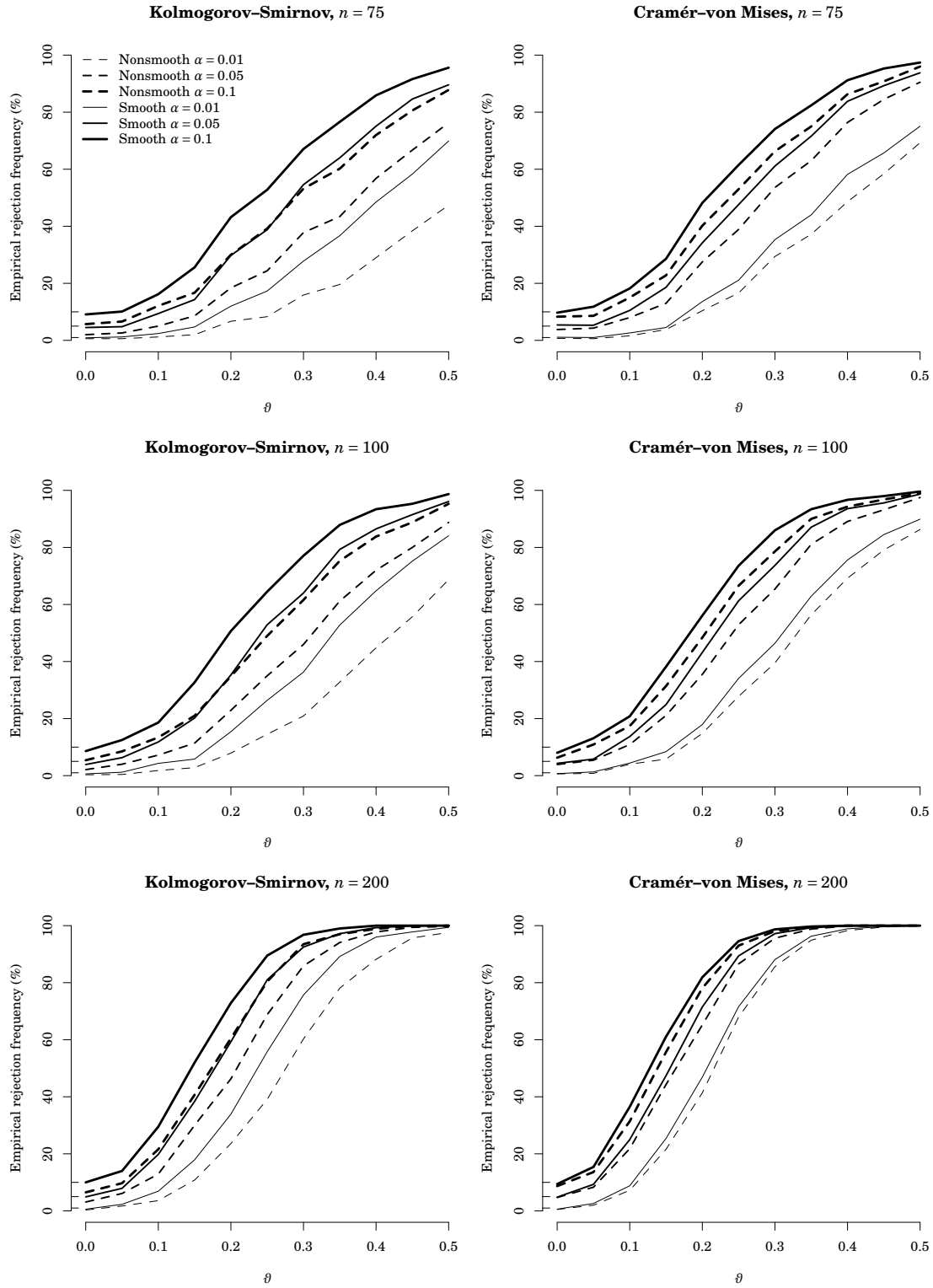


Figure 4.2: Power curves for Tukey's g distribution in the case where there are no regressors.

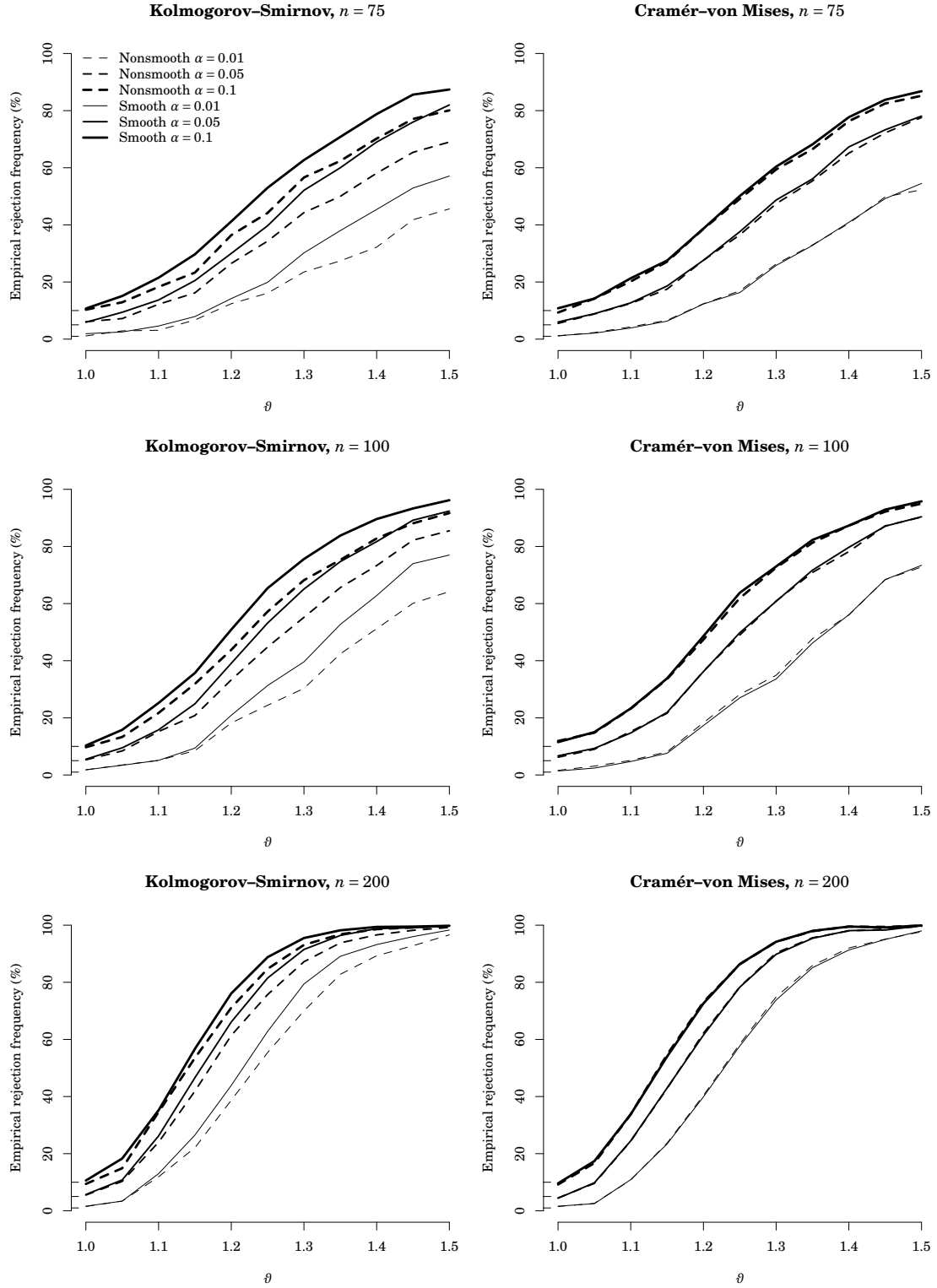


Figure 4.3: Power curves for the asymmetric Laplace distribution in the case where there are no regressors.

mean integrated square errors of the nonsmooth and smooth symmetry processes are equivalent up to asymptotically negligible error terms (under appropriate conditions).

4.3.3 Case II: Linear regression

Consider again the case where the goal is to test whether the errors appearing in a linear regression model are symmetric around zero. As we are interested in comparing the effect of smoothing on the behaviour of the classical tests, we compare the performance of the nonsmooth Kolmogorov–Smirnov and Cramér–von Mises tests based on (2.5) with their smooth counterparts based on the test criteria (4.6).

As in Section 3.3.2, we sample observations from the model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

with $\beta_0 = 0$ and $\beta_1 = \beta_2 = 1$, and the x_i being fixed regressors (these are sampled randomly from a uniform(0,1) distribution once and kept fixed over all Monte Carlo iterations). The errors were simulated from the distributions in Table 2.1 and, if necessary, were adjusted to have zero mean and unit variance. See Section 2.7 for more detail on how to sample from these distributions.

Seeing that the errors are unobservable, we base all tests on the residuals defined as

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2, \quad i = 1, 2, \dots, n,$$

where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are the ordinary least-squares estimators of the regression parameters β_0 , β_1 and β_2 , respectively (see Section 2.5). On the basis of $\{\hat{\varepsilon}_i\}_{i=1}^n$ the test statistic is computed. Critical values of the tests were obtained using the wild bootstrap procedure set out in Section 3.3.2. All bootstrap approximations were based on $B = 1000$ bootstrap replications.

For the kernel K appearing in (4.6), we again chose the Epanechnikov kernel defined in (4.7). As before, we chose a fixed bandwidth of $h_n = cn^{-1/3}$.

Table 4.2 contains the rejection percentages of the nonsmooth and smooth Kolmogorov–Smirnov and Cramér–von Mises tests in the case where the errors were sampled from a symmetric distribution. In most cases we see that the empirical size of the smoothed Kolmogorov–Smirnov test (based on $U_n(\hat{F}_n^h)$) is slightly higher than that of the nonsmooth counterpart (based on $U_n(\hat{F}_n)$). This increase in size is more prominent for heavy-tailed distributions, where the smooth Kolmogorov–Smirnov tests seem to be slightly oversized for smaller samples. On the other hand, the smooth Cramér–von Mises test (based on $W_n(\hat{F}_n^h)$) has empirical size comparable to the nonsmooth counterpart (based on $W_n(\hat{F}_n)$).

Tables 4.3 and 4.4 show the rejection percentages of the tests under alternatives, i.e. when the errors are sampled from an asymmetric distribution. In agreement with the power curves shown in the previous section, we see that the smooth Kolmogorov–Smirnov test yields significantly higher power than its classical non-smooth counterpart. In some cases the improvement in power is around 50%, e.g. in the case of the χ^2_{10} distribution where a power increase from 25% to 37.1% is effected by smoothing (at a significance level of 5% and for sample size $n = 50$).

Similar to the observation made from the power curves in the previous section, the power gains achieved by using the smoothed Cramér–von Mises test instead of the classical nonsmooth counterpart are less prominent. Nevertheless, in most cases there is an improvement in power when using the smoothed Cramér–von Mises test.

It is clear from the numerical results that, in finite-sample settings, one can use kernel-smoothed distribution function estimators to obtain level-preserving tests which are more powerful than the classical nonsmooth counterparts. Up to this point, little thought has been put into the choice of the bandwidth parameter h_n and the kernel K . In the next section we investigate the performance of the tests when using an optimal, data-driven bandwidth choice.

4.3.4 Cross-validation bandwidths

To choose the bandwidth parameter h_n appearing in (4.5), we employ leave-one-out cross-validation as set out in Bowman et al. (1998). Specifically, we seek an $h_n > 0$ that minimises the function

$$\text{CV}(h_n) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \left[\mathbf{1}(\hat{\varepsilon}_i \leq s) - \hat{F}_{n(i)}^{h_n}(s) \right]^2 ds, \quad (4.25)$$

where $\hat{F}_{n(i)}^{h_n}$ denotes the leave-one-out smoothed kernel distribution function estimator based on the sample with the i th observation removed, i.e.

$$\hat{F}_{n(i)}^{h_n}(s) = \frac{1}{n-1} \sum_{j \neq i} K\left(\frac{s - \hat{\varepsilon}_j}{h_n}\right).$$

Observe that the function in (4.25) requires numerical integration. Instead, we follow Li et al. (2017) and use a computationally more efficient discretised cross-validation function. Specifically, we choose

$$h_n = \text{argmin}_{h>0} \widehat{\text{CV}}(h), \quad (4.26)$$

where

$$\widehat{\text{CV}}(h_n) = \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N \left[\mathbf{1}(\hat{\varepsilon}_i \leq s_j) - \hat{F}_{n(i)}^{h_n}(s_j) \right]^2. \quad (4.27)$$

Here, $s_1 < \dots < s_N$ are N appropriately chosen evaluation points.

Li et al. (2017) show that, under quite mild assumptions, the bandwidths obtained by means of (4.26) and (4.27) are asymptotically equivalent to nonstochastic optimal smoothing parameters. More specifically, bandwidths obtained in this way are asymptotically equivalent to bandwidths of the form $cn^{-1/3}$ for some constant $c > 0$. Bandwidths of this form satisfy $\sqrt{n}h_n^2 = O(1)$, which is a requirement for the theorems of this chapter to hold.

We repeated the Monte Carlo experiments described at the beginning of Section 4.3.3, but this time choosing smoothing bandwidths using the cross-validation procedure described above. Tables 4.5 to 4.7 show the rejection percentages obtained in these experiments.

Looking at the rejection percentages in Table 4.5, we see that, even when using cross-validation bandwidths, the smooth tests respect the nominal level reasonably well. In addition, under alternatives the power (shown in Tables 4.6 and 4.7) of the smooth tests is comparable to the rejection percentages obtained using fixed bandwidths (Tables 4.3 and 4.4). The difference observed could be attributed to Monte Carlo error and more precise conclusions regarding the comparison cannot be drawn from the results shown. Nevertheless, the results are reassuring and suggest that reasonable size and high power can be achieved using automatic bandwidth selection procedures. This is important from a practical perspective, as it takes the burden of choosing bandwidths away from the practitioner.

Avenues for future research

We have shown that kernel distribution function smoothing provides some advantages in terms of power of goodness-of-fit tests for symmetry. Natural extensions of the ideas presented and ideas for possible future research are the following:

- studying the effect of the choice of the kernel distribution function K and the bandwidth parameter h_n on the finite-sample results;
- extending the theoretical results to testing whether the distribution of the errors in a nonparametric regression model is symmetric around zero (as was done by Neumeyer and Dette, 2007; Hušková and Meintanis, 2012), while allowing for both continuous and discrete regressors as in Racine and van Keilegom (2020);
- deriving the local approximate Bahadur efficiencies of the smoothed tests as was done for several other tests by Milošević and Obradović (2019).

Table 4.2: Rejection percentages of the nonsmooth tests based on $U_n(\hat{F}_n)$ and $W_n(\hat{F}_n)$ and the smooth counterparts $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$ using *fixed* bandwidths ($h_n = n^{-1/3}$) where the regression errors were sampled from symmetric distributions. α denotes the significance level of the test.

Dist.	ϑ	n	$\alpha = 0.05$				$\alpha = 0.1$			
			$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$	$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$
N	0	50	4.4	5.8	5.6	5.4	8.9	11.1	10.7	10.9
		75	5.0	6.0	5.6	5.8	9.1	10.3	10.3	9.9
		100	4.8	5.6	5.2	5.0	9.1	10.4	10.6	10.3
t_ϑ	8	50	4.0	5.8	5.3	6.0	8.9	11.3	10.7	12.1
		75	4.3	5.8	5.4	5.8	8.9	11.0	10.2	10.9
		100	5.0	5.6	5.3	5.0	9.1	10.5	9.9	10.5
	4	50	6.6	8.1	7.9	8.6	11.2	13.7	14.3	14.6
		75	6.2	7.0	7.4	8.0	11.2	12.8	13.2	13.5
		100	5.6	6.4	6.4	6.2	10.2	11.7	11.9	12.0
	U	50	5.2	8.0	7.8	8.2	10.4	12.9	12.6	13.1
		75	5.2	6.2	6.0	5.8	9.3	11.4	11.3	11.1
		100	4.2	6.6	6.6	6.9	9.3	11.1	12.3	12.2
L	0	50	5.4	6.6	6.4	6.5	9.5	11.8	11.1	12.0
		75	5.1	6.0	6.0	6.0	9.2	10.7	11.1	10.8
		100	3.8	4.3	5.0	4.1	8.4	8.9	10.0	10.2
NM $_\vartheta$	2	50	5.0	6.8	6.5	6.7	10.3	11.4	11.2	11.6
		75	5.0	6.7	5.8	6.5	9.4	11.8	10.9	11.7
		100	5.0	5.5	5.7	5.5	9.8	11.6	11.6	10.9
	3	50	5.0	6.8	7.0	6.8	9.4	12.2	11.4	12.3
		75	5.8	6.6	6.3	7.0	10.2	11.9	11.8	12.2
		100	5.2	5.7	5.7	6.0	10.2	11.7	11.6	11.5
	4	50	5.3	7.0	7.0	8.1	10.5	12.6	12.8	13.1
		75	6.3	7.2	6.8	7.0	11.1	12.0	12.3	12.2
		100	5.3	6.2	6.1	6.2	10.1	11.1	12.1	11.7

Table 4.3: Rejection percentages of the nonsmooth tests based on $U_n(\hat{F}_n)$ and $W_n(\hat{F}_n)$ and the smooth counterparts $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$ using *fixed* bandwidths ($h_n = n^{-1/3}$) where the regression errors were sampled from asymmetric distributions. α denotes the significance level of the test.

Dist.	ϑ	n	$\alpha = 0.05$				$\alpha = 0.1$			
			$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$	$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$
χ^2_ϑ	4	50	56.3	74.9	75.0	81.7	70.5	82.8	85.4	90.3
		75	82.6	90.3	92.5	94.6	90.0	94.7	96.4	97.5
		100	90.1	95.6	98.0	98.5	95.2	98.1	99.2	99.7
	6	50	41.8	59.1	58.4	65.1	53.8	68.2	69.2	74.5
		75	65.2	76.1	79.7	85.5	76.8	87.2	88.4	91.3
		100	77.6	87.6	92.5	95.2	87.8	93.9	96.4	97.5
	8	50	33.6	49.1	49.7	54.4	45.6	59.7	60.2	65.8
		75	54.1	66.2	71.1	76.2	69.0	77.9	81.8	85.8
		100	69.1	76.8	85.0	88.9	79.2	88.4	92.4	94.4
	10	50	25.0	37.1	37.3	42.1	37.0	50.8	48.2	56.7
		75	42.3	52.6	57.0	63.9	55.7	66.2	69.8	75.4
		100	55.2	68.3	74.6	79.6	69.7	80.0	84.3	87.9
SN $_\vartheta$	1	50	5.6	7.6	6.7	8.2	9.7	11.8	12.1	12.1
		75	4.6	6.2	5.9	6.4	8.9	11.9	11.9	13.1
		100	6.1	7.6	7.0	8.0	10.0	12.3	13.0	13.4
	2	50	10.9	16.2	15.3	17.5	18.6	25.8	23.5	27.3
		75	16.8	21.3	21.6	23.9	26.5	33.2	32.7	35.9
		100	22.3	29.1	30.3	32.0	33.2	39.3	40.5	44.3
	3	50	22.2	32.8	32.4	36.6	31.5	44.1	43.2	47.9
		75	33.0	43.4	45.6	49.7	45.8	55.8	58.6	63.2
		100	43.2	54.1	58.2	62.6	59.3	68.1	71.4	75.9

Table 4.4: Rejection percentages of the nonsmooth tests based on $U_n(\hat{F}_n)$ and $W_n(\hat{F}_n)$ and the smooth counterparts $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$ using *fixed* bandwidths ($h_n = n^{-1/3}$) where the regression errors were sampled from asymmetric distributions. α denotes the significance level of the test.

Dist.	ϑ	n	$\alpha = 0.05$				$\alpha = 0.1$			
			$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$	$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$
AL $_{\vartheta}$	1.00	50	6.3	7.5	8.4	7.9	11.9	12.9	13.5	13.0
		75	8.7	9.0	10.3	9.2	13.6	14.9	15.6	15.1
		100	7.1	6.9	7.0	7.0	11.9	11.8	13.0	11.6
	1.50	50	53.3	64.3	63.3	65.6	65.6	74.0	73.2	74.7
		75	71.5	80.8	79.2	79.8	81.7	87.5	85.9	87.9
		100	85.4	90.8	90.6	90.7	91.1	94.5	94.6	95.2
	2.00	50	76.8	89.1	87.5	91.2	87.3	93.8	93.4	95.2
		75	93.4	97.2	97.5	98.0	96.4	98.8	98.8	99.3
		100	99.1	99.9	100.0	100.0	100.0	100.0	100.0	100.0
TU $_{\vartheta}$	0.15	50	8.9	14.0	13.4	15.2	15.4	21.4	21.5	22.7
		75	12.2	18.2	19.0	21.2	21.1	27.5	27.7	30.5
		100	15.4	22.6	24.7	26.5	26.8	34.7	35.4	38.7
	0.30	50	25.2	38.7	38.5	45.5	37.0	52.6	52.3	57.5
		75	39.3	50.4	53.9	60.3	52.0	64.1	66.0	72.6
		100	55.8	65.4	70.3	75.8	67.1	78.0	82.1	85.9
	0.50	50	54.9	74.3	73.9	80.6	67.3	84.4	83.0	88.0
		75	81.0	90.5	92.5	94.8	90.4	95.8	95.5	97.1
		100	92.0	96.0	97.6	99.1	96.2	98.9	99.2	99.6

Table 4.5: Rejection percentages of the nonsmooth tests based on $U_n(\hat{F}_n)$ and $W_n(\hat{F}_n)$ and the smooth counterparts $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$ using *cross-validation* bandwidths where the regression errors were sampled from symmetric distributions. α denotes the significance level of the test.

Dist.	ϑ	n	$\alpha = 0.05$				$\alpha = 0.1$			
			$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$	$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$
N	0	50	4.4	5.3	5.4	4.9	8.8	10.3	10.1	10.0
		75	4.4	5.1	5.3	4.5	8.0	9.6	10.0	9.6
		100	5.8	6.4	6.5	6.3	10.6	10.9	10.9	11.0
t_ϑ	8	50	4.7	6.8	7.1	6.9	10.2	12.6	12.7	13.2
		75	4.4	5.4	5.4	5.0	9.1	10.3	10.6	10.8
		100	5.3	5.6	5.9	5.7	9.1	10.7	11.4	10.9
	4	50	5.6	8.3	8.2	8.6	11.5	14.9	14.1	14.4
		75	5.5	6.5	6.3	6.4	10.2	11.7	11.9	12.3
		100	5.8	7.3	6.9	7.4	11.6	12.5	12.2	12.7
	U	50	4.2	4.7	6.0	5.1	8.9	10.1	11.2	10.5
		75	4.2	6.4	7.4	7.0	9.0	10.5	12.4	11.1
		100	4.6	5.5	5.8	5.2	8.6	11.7	11.9	11.5
L	0	50	4.4	5.5	6.7	5.9	9.3	10.9	11.7	11.5
		75	4.6	5.5	5.3	5.7	7.8	9.4	9.6	9.8
		100	3.9	4.2	3.7	4.0	8.0	8.9	8.6	8.6
NM $_\vartheta$	2	50	5.0	5.7	6.3	5.3	9.2	11.2	11.8	11.4
		75	3.9	5.0	4.5	4.9	8.3	9.7	9.0	10.0
		100	4.7	5.2	6.0	5.2	8.9	10.0	10.1	10.5
	3	50	5.3	6.3	6.3	6.3	9.1	11.9	11.9	12.4
		75	4.5	5.5	5.6	6.1	7.6	10.0	10.1	10.1
		100	5.4	6.9	6.7	6.1	9.4	12.0	11.7	11.9
	4	50	5.9	8.0	7.9	8.5	12.1	14.5	15.1	14.6
		75	5.5	5.9	5.5	5.0	10.4	11.0	11.3	11.2
		100	5.4	5.9	6.3	6.2	8.8	11.5	11.4	11.2

Table 4.6: Rejection percentages of the nonsmooth tests based on $U_n(\hat{F}_n)$ and $W_n(\hat{F}_n)$ and the smooth counterparts $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$ using *cross-validation* bandwidths where the regression errors were sampled from asymmetric distributions. α denotes the significance level of the test.

Dist.	ϑ	n	$\alpha = 0.05$				$\alpha = 0.1$			
			$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$	$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$
χ^2_ϑ	4	50	57.8	74.7	77.7	79.1	70.5	85.1	85.0	88.4
		75	80.8	89.4	91.6	93.3	88.8	95.0	95.5	96.6
		100	92.0	97.0	98.4	98.9	96.6	98.9	99.1	99.6
	6	50	40.1	55.8	57.8	60.4	53.2	68.2	69.9	72.3
		75	66.2	78.8	80.8	84.7	78.2	88.0	87.9	92.0
		100	78.9	89.7	92.0	94.1	88.0	94.8	96.0	96.9
	8	50	31.0	45.6	47.1	50.5	44.3	60.6	59.7	65.5
		75	54.1	65.7	67.7	71.2	66.1	78.3	79.4	83.8
		100	67.8	80.5	82.9	86.9	79.8	89.5	90.7	92.9
	10	50	26.3	37.8	38.9	41.7	38.6	52.6	51.7	56.1
		75	43.8	55.2	59.0	62.7	56.5	70.6	71.5	76.0
		100	56.3	71.4	73.5	77.9	70.7	81.6	83.4	87.4
SN_ϑ	1	50	4.9	6.5	6.9	6.4	9.8	10.0	10.7	10.2
		75	5.8	5.8	5.9	5.5	9.9	11.9	12.3	11.6
		100	5.4	7.2	6.7	7.3	9.9	13.7	13.0	13.4
	2	50	10.6	15.1	16.1	16.7	18.6	24.1	24.8	25.5
		75	16.3	22.2	22.5	23.9	24.6	32.1	32.2	36.2
		100	21.8	28.3	29.4	32.2	33.8	42.5	43.3	46.4
	3	50	16.6	26.1	28.0	29.0	27.6	40.5	40.4	42.7
		75	30.6	42.9	43.6	46.4	44.2	56.5	57.4	61.3
		100	42.7	54.6	57.2	60.5	56.0	67.8	68.8	71.9

Table 4.7: Rejection percentages of the nonsmooth tests based on $U_n(\hat{F}_n)$ and $W_n(\hat{F}_n)$ and the smooth counterparts $U_n(\hat{F}_n^h)$ and $W_n(\hat{F}_n^h)$ using *cross-validation* bandwidths where the regression errors were sampled from asymmetric distributions. α denotes the significance level of the test.

Dist.	ϑ	n	$\alpha = 0.05$				$\alpha = 0.1$			
			$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$	$U_n(\hat{F}_n)$	$U_n(\hat{F}_n^h)$	$W_n(\hat{F}_n)$	$W_n(\hat{F}_n^h)$
AL $_{\vartheta}$	1.00	50	8.7	8.8	9.5	10.0	14.2	15.3	15.9	15.7
		75	8.9	9.7	9.8	10.3	14.7	15.2	15.6	15.8
		100	6.8	7.4	6.9	7.3	11.9	12.8	13.2	13.5
	1.50	50	51.9	59.4	59.5	61.5	63.5	69.9	69.0	69.8
		75	74.4	80.1	80.7	81.8	82.2	87.2	86.9	87.3
		100	85.3	90.7	90.8	90.8	91.4	94.3	94.7	94.6
	2.00	50	76.5	87.4	88.0	89.8	87.3	94.2	94.0	95.2
		75	94.6	97.8	98.4	98.4	97.1	99.3	99.6	99.7
		100	99.2	99.9	100.0	100.0	99.7	100.0	100.0	100.0
TU $_{\vartheta}$	0.15	50	8.9	13.3	15.5	14.1	17.7	22.9	23.7	24.3
		75	13.7	17.1	17.6	18.8	22.5	28.0	29.3	30.5
		100	17.3	23.2	25.2	26.7	27.1	35.2	36.7	39.4
	0.30	50	25.0	34.7	36.5	38.0	34.7	48.3	47.4	52.6
		75	41.3	53.6	54.6	58.6	55.7	65.9	65.9	71.9
		100	55.5	67.7	70.5	74.5	68.8	79.6	81.3	85.9
	0.50	50	58.6	74.3	75.6	78.2	72.0	83.9	85.0	87.0
		75	81.6	89.7	92.6	93.7	89.8	95.1	95.5	96.9
		100	93.4	98.2	98.9	99.4	97.1	99.5	99.7	99.8

Chapter 5

Some asymptotic results for existing tests for symmetry

In this chapter we consider a test recently proposed by Allison and Pretorius (2017). Their test uses the empirical characteristic function as main tool and is based on the following characterisation pointed out by Baringhaus and Henze (1992).

Theorem 5.1. *Let X_1 and X_2 be independent random variables with distribution F . The distribution F is symmetric around zero if and only if the random variables $|X_1|$ and $|X_1 \vee X_2|$ have the same probability distribution.*

This characterisation implies that the characteristic functions of $|X_1|$ and $|X_1 \vee X_2|$ will coincide if and only if the distribution of X_1 (and X_2) is symmetric around zero. From this it follows that a suitable criterion for evaluating departures from symmetry is

$$\int_{-\infty}^{\infty} |\chi(t) - \phi(t)|^2 w(t) dt, \quad (5.1)$$

where $w(\cdot)$ is an appropriate weight function and

$$\chi(t) = \mathbf{E} \left(e^{it|X_1|} \right), \quad \phi(t) = \mathbf{E} \left(e^{it|X_1 \vee X_2|} \right).$$

Clearly, since the characteristic function of a random variable completely and uniquely specifies the distribution of the random variable (see, e.g., Billingsley, 1995, Theorem 26.2), the quantity in (5.1) will be zero if and only if X_1 has a distribution which is symmetric around zero.

The above discussion motivates the test proposed by Allison and Pretorius (2017), which we now describe. Given a random sample of observations from a distribution F , say $X_1, \dots, X_n$, one may test the null hypothesis of symmetry using the test statistic

$$A_n = n \int_{-\infty}^{\infty} |\chi_n(t) - \phi_n(t)|^2 w(t) dt, \quad (5.2)$$

where $\chi_n(\cdot)$ and $\phi_n(\cdot)$ are empirical counterparts of $\chi(\cdot)$ and $\phi(\cdot)$, respectively. Specifically, Allison and Pretorius (2017) used the estimators

$$\chi_n(t) = \frac{1}{n} \sum_{k=1}^n e^{it|X_k|} \quad \text{and} \quad \phi_n(t) = \frac{2}{n(n-1)} \sum_{1 \leq j < k \leq n} e^{it|X_j \vee X_k|}. \quad (5.3)$$

Note that the statistic A_n corresponds to the statistic $A_a^{(2)}$ defined in (2.15). In this chapter we use the notation A_n to emphasise the dependence of the statistic on n .

In order to facilitate practical implementation of the test, $\phi_n(t)$ can be rewritten as follows (see Section 2.2.3) using the order statistics $X_{(1)} < \dots < X_{(n)}$ obtained by sorting $X_1, \dots, X_n$ in ascending order:

$$\begin{aligned} \phi_n(t) &= \frac{2}{n(n-1)} \sum_{1 \leq j < k \leq n} e^{it|X_{(j)} \vee X_{(k)}|} \\ &= \frac{2}{n(n-1)} \sum_{k=1}^n (k-1) e^{it|X_{(k)}|}. \end{aligned}$$

This allows us to write

$$\chi_n(t) - \phi_n(t) = \frac{1}{n} \sum_{k=1}^n c_{kn} e^{it|X_{(k)}|} = \frac{1}{n} \sum_{k=1}^n c_{nk} \cos(tX_{(k)}) + i \frac{1}{n} \sum_{k=1}^n c_{nk} \sin(t|X_{(k)}|),$$

where $c_{kn} = (n - 2k + 1)/(n - 1)$. The statistic A_n then becomes

$$\begin{aligned} A_n &= n \int_{-\infty}^{\infty} |\chi_n(t) - \phi_n(t)|^2 w(t) dt \\ &= \int_{-\infty}^{\infty} \left[\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n c_{nk} \cos(tX_{(k)}) \right)^2 + \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n c_{nk} \sin(t|X_{(k)}|) \right)^2 \right] w(t) dt \\ &= \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n c_{nk} \cos(tX_{(k)}) + \frac{1}{\sqrt{n}} \sum_{k=1}^n c_{nk} \sin(t|X_{(k)}|) \right)^2 w(t) dt, \end{aligned}$$

where the last step follows from the fact that

$$\int_{-\infty}^{\infty} \cos(tX_{(j)}) \sin(t|X_{(k)}|) w(t) dt = 0.$$

In the arguments above we constructed a test for symmetry around zero. The test may readily be generalised to testing for symmetry around any *known* median. Suppose one would like to test whether the distribution that generated a random sample, say $Y_1, \dots, Y_n$, is symmetric around η , then the appropriate statistic would be as defined in (5.2) and (5.3) with $X_k = Y_k - \eta$.

5.1 Case of a specified centre of symmetry

We now study the asymptotic behaviour of A_n where the objective is to test for symmetry around a specified median η . As discussed above, we will take $\eta = 0$ throughout without any loss of generality. Our analysis is based on the following assumptions:

(S1) $X_1, \dots, X_n$ are independent and identically distributed and $E(X_1^2) < \infty$.

(S2) The weight function $w(\cdot)$ is positive and even, that is,

$$w(t) > 0, \quad w(-t) = w(t), \quad \forall t \in \mathbb{R}.$$

These assumptions are typical in the literature for tests using the empirical characteristic function as main tool. Compare the assumptions used in, for example, Hušková and Meintanis (2007) and Hušková and Meintanis (2012).

5.1.1 Asymptotic null distribution

To aid further analysis, we first provide an asymptotic representation of A_n as a sum of independent random processes. Let $C_\chi(t)$ and $S_\chi(t)$ denote the real and imaginary parts of $\chi_n(t)$, i.e.

$$C_\chi(t) = \operatorname{Re}(\chi_n(t)) = \frac{1}{n} \sum_{k=1}^n \cos(tX_k),$$

$$S_\chi(t) = \operatorname{Im}(\chi_n(t)) = \frac{1}{n} \sum_{k=1}^n \sin(t|X_k|).$$

Likewise, define

$$C_\phi(t) = \operatorname{Re}(\phi_n(t)) = \frac{2}{n(n-1)} \sum_{1 \leq j < k \leq n} \cos(t(X_j \vee X_k))$$

$$S_\phi(t) = \operatorname{Im}(\phi_n(t)) = \frac{2}{n(n-1)} \sum_{1 \leq j < k \leq n} \sin(t|X_j \vee X_k|).$$

Using this notation, we may write

$$\begin{aligned} |\chi_n(t) - \phi_n(t)|^2 &= (C_\chi(t) - C_\phi(t))^2 + (S_\chi(t) - S_\phi(t))^2 \\ &= (C_\chi(t) - C_\phi(t) + S_\chi(t) - S_\phi(t))^2 - 2(C_\chi(t) - C_\phi(t))(S_\chi(t) - S_\phi(t)). \end{aligned}$$

Now, since w is assumed to be an even function, one has

$$\int_{-\infty}^{\infty} \cos(tk_1) \sin(tk_2) w(t) dt = 0, \quad \forall k_1, k_2 \in \mathbb{R},$$

and hence it follows that

$$\begin{aligned} \int_{-\infty}^{\infty} C_\chi(t) S_\chi(t) w(t) dt &= \int_{-\infty}^{\infty} C_\chi(t) S_\phi(t) w(t) dt \\ &= \int_{-\infty}^{\infty} C_\phi(t) S_\chi(t) w(t) dt = \int_{-\infty}^{\infty} C_\phi(t) S_\phi(t) w(t) dt = 0. \end{aligned}$$

This shows that the test statistic may be expressed as

$$\begin{aligned} A_n &= n \int_{-\infty}^{\infty} |\chi_n(t) - \phi_n(t)|^2 w(t) dt \\ &= n \int_{-\infty}^{\infty} (C_\chi(t) - C_\phi(t) + S_\chi(t) - S_\phi(t))^2 w(t) dt \\ &=: \int_{-\infty}^{\infty} W_n^2(t) w(t) dt. \end{aligned} \tag{5.4}$$

This representation of A_n allows us to study its limit distribution in the setting of the Hilbert space of measurable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ endowed with the norm $\|f\| = (\int_{-\infty}^{\infty} f(t)^2 w(t) dt)^{1/2}$. We may therefore write the statistic as

$$A_n = \|W_n\|^2. \quad (5.5)$$

Notice that W_n is a real-valued function.

The function W_n defined in (5.4) above can be expressed as

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n g(t|X_k|) - \frac{1}{\sqrt{n(n-1)}} \sum_{j \neq k} g(t|X_j \vee X_k|), \quad (5.6)$$

where, for all $u \in \mathbb{R}$, we define the trigonometric function

$$g(u) = \cos(u) + \sin(u). \quad (5.7)$$

Observe that, in view of Theorem 5.1, one has under H_0 that $E(g(t|X_1|)) = E(g(t|X_1 \vee X_2|))$ and hence also $E(W_n(t)) = 0$ for all $t \in \mathbb{R}$. However, W_n is not a sum of i.i.d. random processes. Therefore, in order to study the asymptotic behaviour of W_n under the null hypothesis, we consider the Hájek projection (see, e.g., van der Vaart, 2000, p. 157)

$$\widetilde{W}_n(t) = \sum_{\ell=1}^n E(W_n(t)|X_\ell) = \frac{1}{\sqrt{n}} \sum_{k=1}^n (g(t|X_k|) + \Psi(t) - 2\Psi(t|X_k)), \quad (5.8)$$

where, letting X denote an independent copy of X_1 , we define

$$\begin{aligned} \Psi(t) &= E(g(t|X|)), \\ \Psi(t|X_k) &= E(g(t|X_k \vee X|)|X_k). \end{aligned}$$

Clearly, under H_0 , the statistic $\widetilde{W}_n$ is a sum of independent zero-mean random processes and will prove to be a convenient proxy for studying the large-sample behaviour of W_n .

The asymptotic null distribution of the test statistic A_n is stated in the following theorem.

Theorem 5.2. *Under H_0 , Assumptions (S1)–(S2) and the condition $\int_{-\infty}^{\infty} t^2 w(t) dt < \infty$, there exists a zero-mean Gaussian process $\mathcal{W} = \{\mathcal{W}(t); t \in \mathbb{R}\}$ such that*

$$A_n \xrightarrow{d} \|\mathcal{W}\|^2,$$

as $n \rightarrow \infty$. Moreover, the covariance structure of $\mathcal{W}$ is given by

$$\begin{aligned} \text{Cov}(\mathcal{W}(s), \mathcal{W}(t)) &= \text{Cov}(g(s|X_1|), g(t|X_1|)) + 4E(\Psi(s|X_1)\Psi(t|X_1)) \\ &\quad - 2E(g(s|X_1|)\Psi(t|X_1)) - 2E(g(t|X_1|)\Psi(s|X_1)). \end{aligned}$$

Proof. Our proof relies on showing that the asymptotic behaviour of W_n and $\widetilde{W}_n$ are equivalent in the sense that

$$\left| \|W_n\| - \|\widetilde{W}_n\| \right| \xrightarrow{\mathbb{P}} 0. \quad (5.9)$$

If this is true, then one has

$$\begin{aligned} A_n &= \|W_n\|^2 = \left[\|\widetilde{W}_n\| - (\|W_n\| - \|\widetilde{W}_n\|) \right]^2 \\ &= \|\widetilde{W}_n\|^2 + (\|W_n\| - \|\widetilde{W}_n\|)^2 - 2\|\widetilde{W}_n\|(\|W_n\| - \|\widetilde{W}_n\|) \\ &= \|\widetilde{W}_n\|^2 + o_{\mathbb{P}}(1), \end{aligned}$$

as $n \rightarrow \infty$, provided that $\|\widetilde{W}_n\| = O_{\mathbb{P}}(1)$. The boundedness of $\|\widetilde{W}_n\|^2$ follows readily from the integrability of the weight function w and the boundedness of g . Specifically, given any $\beta \geq 1$, one has for all $u \in \mathbb{R}$ that

$$|g(u)|^\beta = |\cos(u) + \sin(u)|^\beta \leq 2^{\beta-1} |\cos(u)|^\beta + 2^{\beta-1} |\sin(u)|^\beta \leq 2^\beta.$$

Therefore, since the right-hand side in the last display does not depend on u , we have

$$\sup_{u \in \mathbb{R}} |g(u)|^\beta \leq 2^\beta, \quad \beta \geq 1. \quad (5.10)$$

We now move on to show that the random processes W_n and $\widetilde{W}_n$ are asymptotically equivalent. Choose any $t \in \mathbb{R}$ and write

$$\mathbb{E} \left[(W_n(t) - \widetilde{W}_n(t))^2 \right] = \mathbb{E} (B_n(t) + C_n(t) + D_n(t))^2,$$

where

$$\begin{aligned} B_n(t) &= -\frac{1}{\sqrt{n(n-1)}} \sum_{j \neq k} g(t | X_j \vee X_k |), \\ C_n(t) &= \frac{2}{\sqrt{n}} \sum_{k=1}^n \Psi(t | X_k), \\ D_n(t) &= -\sqrt{n} \Psi(t). \end{aligned}$$

To facilitate exposition, we make use of the following shorthand notation:

$$\begin{aligned} \sum_{(i,j)} a_{ij} &\equiv \sum_{i=1}^n \sum_{j \neq i}^n a_{ij}, & \sum_{(i,j,k)} &\equiv \sum_{i=1}^n \sum_{j \neq i}^n \sum_{\substack{k \neq i \\ k \neq j}}^n, & \sum_{(i,j,k,\ell)} &\equiv \sum_{i=1}^n \sum_{j \neq i}^n \sum_{\substack{k \neq i \\ k \neq j}}^n \sum_{\substack{\ell \neq i \\ \ell \neq j \\ \ell \neq k}}^n. \end{aligned}$$

We first treat the quadratic terms, starting with

$$\begin{aligned}
\mathbb{E}(B_n^2(t)) &= \frac{1}{n(n-1)^2} \sum_{(j,k)} \sum_{(l,m)} \mathbb{E}[g(|X_j \vee X_k|, t) g(|X_l \vee X_m|, t)] \\
&= \frac{1}{n(n-1)^2} \sum_{(j,k,l,m)} \mathbb{E}[g(|X_j \vee X_k|, t)] \mathbb{E}[g(|X_l \vee X_m|, t)] \\
&\quad + \frac{2}{n(n-1)^2} \sum_{(j,k)} \mathbb{E}[g^2(|X_j \vee X_k|, t)] \\
&\quad + \frac{4}{n(n-1)^2} \sum_{(j,k,l)} \mathbb{E}[g(|X_j \vee X_k|, t) g(|X_j \vee X_l|, t)] \\
&= \frac{(n-2)(n-3)}{n-1} \{\mathbb{E}[g(|X_1 \vee X_2|, t)]\}^2 + \frac{2}{n-1} \mathbb{E}[g^2(|X_1 \vee X_2|, t)] \\
&\quad + \frac{4(n-2)}{n-1} \mathbb{E}[g(|X_1 \vee X_2|, t) g(|X_1 \vee X_3|, t)] \\
&= (n-4) \{\mathbb{E}[g(|X_1 \vee X_2|, t)]\}^2 + 4 \mathbb{E}[g(|X_1 \vee X_2|, t) g(|X_1 \vee X_3|, t)] + O(n^{-1}).
\end{aligned}$$

Notice that, by virtue of (5.10), all expectations encountered in the above (and the remaining) calculations are finite. This is necessary to ensure that the error term is truly of order $O(n^{-1})$. Now, since the X_k are mutually independent and identically distributed, we get by the tower law that

$$\begin{aligned}
\mathbb{E}[g(|X_1 \vee X_2|, t) g(|X_1 \vee X_3|, t)] &= \mathbb{E}\{\mathbb{E}[g(|X_1 \vee X_2|, t) g(|X_1 \vee X_3|, t) | X_1]\} \\
&= \mathbb{E}\{\mathbb{E}[g(|X_1 \vee X_2|, t) | X_1] \mathbb{E}[g(|X_1 \vee X_3|, t) | X_1]\} \\
&= \mathbb{E}\left\{\mathbb{E}[g(|X_1 \vee X_2|, t) | X_1]\right\}^2,
\end{aligned}$$

so that

$$\mathbb{E}(B_n^2(t)) = (n-4)(\mathbb{E}[g(|X_1|, t)])^2 + 4 \mathbb{E}\left\{\mathbb{E}[g(|X_1 \vee X_2|, t) | X_1]\right\}^2 + O(n^{-1}).$$

The second quadratic term is

$$\begin{aligned}
\mathbb{E}(C_n^2(t)) &= \frac{4}{n} \sum_j \sum_k \mathbb{E}\{\mathbb{E}[g(|X_j \vee Y|, t) | X_j] \mathbb{E}[g(|X_k \vee Y|, t) | X_k]\} \\
&= \frac{4}{n} \sum_j \mathbb{E}\left\{\mathbb{E}[g(|X_j \vee Y|, t) | X_j]\right\}^2 + \frac{4}{n} n(n-1) (\mathbb{E}[g(|X_1 \vee X_2|, t)])^2 \\
&= 4 \mathbb{E}\left\{\mathbb{E}[g(|X_1 \vee X_2|, t) | X_1]\right\}^2 + 4(n-1) (\mathbb{E}[g(|X_1|, t)])^2.
\end{aligned}$$

We now move on to the cross terms:

$$\begin{aligned}
\mathbb{E}(B_n(t)C_n(t)) &= \frac{-2}{n(n-1)} \sum_{(j,k)} \sum_l \mathbb{E}\{g(|X_j \vee X_k|, t) \mathbb{E}[g(|X_l \vee Y|, t)|X_l]\} \\
&= \frac{-4}{n(n-1)} \sum_{(j,k)} \mathbb{E}\{g(|X_j \vee X_k|, t) \mathbb{E}[g(|X_j \vee Y|, t)|X_j]\} \\
&\quad - \frac{2}{n(n-1)} \sum_{(j,k,l)} (\mathbb{E}[g(|X_1 \vee X_2|, t)])^2 \\
&= -4 \mathbb{E}\{g(|X_1 \vee X_2|, t) \mathbb{E}[g(|X_1 \vee Y|, t)|X_1]\} - 2(n-2) \{\mathbb{E}[g(|X_1|, t)]\}^2 \\
&= -4 \mathbb{E}\left\{\mathbb{E}[g(|X_1 \vee X_2|, t)|X_1]\right\}^2 - 2(n-2) \{\mathbb{E}[g(|X_1|, t)]\}^2, \\
\mathbb{E}(B_n(t)D_n(t)) &= n \{\mathbb{E}[g(|X_1|, t)]\}^2, \\
\mathbb{E}(C_n(t)D_n(t)) &= -2n \{\mathbb{E}[g(|X_1|, t)]\}^2.
\end{aligned}$$

Finally, combining terms, we get

$$\sup_{t \in \mathbb{R}} \mathbb{E}\left(|W_n(t) - \widetilde{W}_n(t)|^2\right) = O(n^{-1}).$$

Since by assumption $\int_{-\infty}^{\infty} w(t)dt < \infty$, one has

$$\begin{aligned}
\mathbb{E}(\|W_n - \widetilde{W}_n\|^2) &= \mathbb{E}\left(\int_{-\infty}^{\infty} |W_n(t) - \widetilde{W}_n(t)|^2 w(t)dt\right) \\
&= \int_{-\infty}^{\infty} \mathbb{E}\left(|W_n(t) - \widetilde{W}_n(t)|^2\right) w(t)dt \\
&\leq \int_{-\infty}^{\infty} \sup_{t \in \mathbb{R}} \mathbb{E}\left(|W_n(t) - \widetilde{W}_n(t)|^2\right) w(t)dt \\
&\rightarrow 0.
\end{aligned}$$

It now follows by the Markov inequality that, for any $\varepsilon > 0$,

$$\mathbb{P}(\|W_n - \widetilde{W}_n\|^2 > \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}(\|W_n - \widetilde{W}_n\|^2) \rightarrow 0,$$

which implies that $\|W_n - \widetilde{W}_n\| \xrightarrow{\mathbb{P}} 0$. Therefore, by the triangle inequality,

$$|\|W_n\| - \|\widetilde{W}_n\|| \leq \|W_n - \widetilde{W}_n\| \xrightarrow{\mathbb{P}} 0,$$

which proves (5.9).

Noting that for each $t \in \mathbb{R}$, the statistic $\widetilde{W}_n(t)$ is a sum of i.i.d. random variables, and that $\mathbb{E}(\widetilde{W}_1) = 0$ and $\mathbb{E}(\|\widetilde{W}_1\|^2) < \infty$, we obtain immediately that $\widetilde{W}_n(\cdot) \xrightarrow{d} \mathcal{W}(\cdot)$ by the Hilbert space central limit theorem for i.i.d. random elements (see van der Vaart and Wellner, 1996, Section 1.8). Consequently, by the continuous mapping theorem, $\|\widetilde{W}_n\|^2 \xrightarrow{d} \|\mathcal{W}\|^2$ and assertion (5.5) follows immediately by (5.9). $\square$

5.1.2 Consistency

We now show that the test proposed by Allison and Pretorius (2017) is consistent against general alternatives. Our argument relies on the convergence result

$$\frac{1}{n}A_n = \frac{1}{n} \int_{\mathbb{R}} W_n^2(t)w(t)dt \xrightarrow{P} \int_{\mathbb{R}} |\chi(t) - \phi(t)|^2 w(t)dt, \quad (5.11)$$

which holds under the null hypothesis as well as general alternatives as will formally be shown in the proof of Theorem 5.3 below. By Theorem 5.1, the characteristic functions χ and ϕ will coincide if and only if the null hypothesis of symmetry is true. Consequently, since the weight function w is positive, we have that

$$H_0 \text{ is true} \iff \int_{\mathbb{R}} |\chi(t) - \phi(t)|^2 w(t)dt = 0.$$

Therefore, under alternatives, one has that $\int_{\mathbb{R}} |\chi(t) - \phi(t)|^2 w(t)dt > 0$ and also

$$A_n = n \int_{\mathbb{R}} |\chi(t) - \phi(t)|^2 w(t)dt + o_P(n) \xrightarrow{P} \infty,$$

as $n \rightarrow \infty$, which confirms that the test is consistent.

Theorem 5.3. *Under Assumptions (S1) and (S2), the statement in (5.11) holds true.*

Proof. In this proof, we make use of the representation of W_n as shown in (5.6). In what follows, D denotes an arbitrary compact subset of $\mathbb{R}$. Now, firstly, since g is bounded and continuous, we have by the uniform law of large numbers given by Jennrich (1969, Theorem 2) that

$$\sup_{t \in D} \left| \frac{1}{n} \sum_{k=1}^n g(t|X_k|) - E(g(t|X_1|)) \right| \xrightarrow{P} 0. \quad (5.12)$$

Furthermore, using well-known results on U -statistics (see, e.g., Theorem A in Section 5.4 of Serfling, 1980) and the fact that $E[|g(t|X_1 \vee X_2|)|] < \infty$, we have for each $t \in \mathbb{R}$ that

$$n^{-1/2}B_n(t) = \frac{1}{n(n-1)} \sum_{j \neq k} g(t|X_j \vee X_k|) \xrightarrow{P} E[g(t|X_1 \vee X_2|)]. \quad (5.13)$$

By Theorem 1 of Andrews (1992), this convergence is uniform on the compact set D , provided that $n^{-1/2}B_n(t)$ is stochastically equicontinuous on D , i.e. if for every $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that

$$\limsup_{n \rightarrow \infty} P \left(n^{-1/2} \sup_{t \in D} \sup_{|s-t| < \delta_\varepsilon} |B_n(s) - B_n(t)| > \varepsilon \right) < \varepsilon. \quad (5.14)$$

We now show that this condition indeed holds. Fix any $\varepsilon > 0$ and choose

$$\delta_\varepsilon = \frac{\varepsilon^2}{2E(|X_1 \vee X_2|)}.$$

The quantity in the last display is positive and finite since $0 < \mathbb{E}(|X_1 \vee X_2|) < \infty$. Observe that

$$\begin{aligned} n^{-1/2} \sup_{t \in D} \sup_{|s-t| < \delta_\varepsilon} |B_n(s) - B_n(t)| &\leq \sup_{t \in D} \sup_{|s-t| < \delta_\varepsilon} \frac{1}{n(n-1)} \sum_{j \neq k} |g(s|X_j \vee X_k) - g(t|X_j \vee X_k)| \\ &\leq \sup_{t \in D} \sup_{|s-t| < \delta_\varepsilon} |s-t| \frac{1}{n(n-1)} \sum_{j \neq k} |X_j \vee X_k| \\ &\leq 2\delta_\varepsilon \frac{1}{n(n-1)} \sum_{j \neq k} |X_j \vee X_k|, \end{aligned}$$

where we made use of the inequalities $|\sin(b) - \sin(a)| \leq |b - a|$ and $|\cos(b) - \cos(a)| \leq |b - a|$. By Chebyshev's inequality, this implies

$$\begin{aligned} \mathbb{P} \left(n^{-1/2} \sup_{t \in D} \sup_{|s-t| < \delta_\varepsilon} |B_n(s) - B_n(t)| > \varepsilon \right) &\leq \mathbb{P} \left(2\delta_\varepsilon \frac{1}{n(n-1)} \sum_{j \neq k} |X_j \vee X_k| > \varepsilon \right) \\ &< \frac{2\delta_\varepsilon}{\varepsilon} \mathbb{E}(|X_1 \vee X_2|) = \varepsilon, \end{aligned}$$

for all $n \geq 1$. Hence, the condition in (5.14) is satisfied and (5.13) holds uniformly in $t \in D$. From (5.12) and (5.13) now follows that

$$\sup_{t \in D} \left| \frac{1}{\sqrt{n}} W_n(t) - W(t) \right| \xrightarrow{\mathbb{P}} 0,$$

where

$$W(t) = \mathbb{E}[g(t|X_1)] - \mathbb{E}[g(t|X_1 \vee X_2)].$$

Therefore, for any compact subset D of $\mathbb{R}$,

$$\frac{1}{n} \int_D W_n^2(t) w(t) dt \xrightarrow{\mathbb{P}} \int_D W^2(t) w(t) dt. \quad (5.15)$$

To complete the proof, we show that the compact set D in the last display may be replaced by the entire real line. Firstly, since w is integrable, there exists for every $\varepsilon > 0$ a compact set $D_\varepsilon \subset \mathbb{R}$ such that

$$\int_{\mathbb{R} \setminus D_\varepsilon} w(t) dt < \frac{1}{16} \varepsilon^2.$$

Using (5.10) we know that $|n^{-1/2} W_n(t)| \leq 4$, and therefore

$$\left| \frac{1}{n} \int_{\mathbb{R} \setminus D_\varepsilon} W_n^2(t) w(t) dt \right| \leq \int_{\mathbb{R} \setminus D_\varepsilon} \left| \frac{1}{n} W_n^2(t) \right| w(t) dt \leq 16 \int_{\mathbb{R} \setminus D_\varepsilon} w(t) dt < \varepsilon.$$

Similarly, $|\int_{\mathbb{R} \setminus D_\varepsilon} W^2(t) w(t) dt| < \varepsilon$, whence we have

$$\begin{aligned} &\left| \frac{1}{n} \int_{\mathbb{R}} W_n^2(t) w(t) dt - \int_{\mathbb{R}} W^2(t) w(t) dt \right| \\ &\leq \left| \frac{1}{n} \int_{D_\varepsilon} W_n^2(t) w(t) dt - \int_{D_\varepsilon} W^2(t) w(t) dt \right| + \left| \frac{1}{n} \int_{\mathbb{R} \setminus D_\varepsilon} W_n^2(t) w(t) dt \right| + \left| \int_{\mathbb{R} \setminus D_\varepsilon} W^2(t) w(t) dt \right| \\ &\leq \left| \frac{1}{n} \int_{D_\varepsilon} W_n^2(t) w(t) dt - \int_{D_\varepsilon} W^2(t) w(t) dt \right| + 2\varepsilon. \end{aligned}$$

This, together with (5.15), shows that

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n} \int_{\mathbb{R}} W_n^2(t)w(t)dt - \int_{\mathbb{R}} W_n^2(t)w(t)dt\right| > 3\varepsilon\right) \\ \leq \mathbb{P}\left(\left|\frac{1}{n} \int_{D_\varepsilon} W_n^2(t)w(t)dt - \int_{D_\varepsilon} W_n^2(t)w(t)dt\right| > \varepsilon\right) < \varepsilon, \end{aligned}$$

for large enough n . Hence, as $n \rightarrow \infty$,

$$\frac{1}{n} \int_{\mathbb{R}} W_n^2(t)w(t)dt \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} W^2(t)w(t)dt. \quad (5.16)$$

Following similar arguments as in Section 5.1.1, we see that the integral $\int_{\mathbb{R}} |\chi(t) - \phi(t)|^2 w(t)dt$ has the representation

$$\int_{\mathbb{R}} |\chi(t) - \phi(t)|^2 w(t)dt = \int_{\mathbb{R}} W^2(t)w(t)dt.$$

The statements in (5.11) and (5.16) are therefore equivalent. $\square$

5.2 Case of an unknown median

We now consider the case where the true median η is unknown and needs to be estimated by an appropriate estimator, which we denote by $\hat{\eta}_n$. A suitable version of the test statistic A_n is then

$$\hat{A}_n = \|\widehat{W}_n\|^2,$$

where, analogous to (5.6), we define

$$\widehat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n g(t|X_k - \hat{\eta}_n|) - \frac{1}{\sqrt{n}(n-1)} \sum_{j \neq k} g(t|(X_j - \hat{\eta}_n) \vee (X_k - \hat{\eta}_n)|). \quad (5.17)$$

As in previous chapters, the estimator of η should be well behaved in the following sense:

(S3) The estimator $\hat{\eta}_n$ admits the asymptotic representation

$$\sqrt{n}(\hat{\eta}_n - \eta) = \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k - \eta) + o_{\mathbb{P}}(1),$$

as $n \rightarrow \infty$, where d_η is some constant and ψ is a measurable odd function satisfying $\mathbb{E}[\psi(X_1 - \eta)] = 0$ and $\mathbb{E}[\psi^2(X_1 - \eta)] < \infty$.

To simplify the expressions that follow, we shall assume without loss of generality that $\eta = 0$.

The first objective now is to find an asymptotic expansion of $\widehat{W}_n(t)$ which will illuminate how estimation of η influences the behaviour $W_n(t)$.

Lemma 5.1. *Suppose that Assumptions (S1)–(S3) and the condition $\int_{-\infty}^{\infty} t^4 w(t) dt < \infty$ hold. Then, for any compact subset D of $\mathbb{R}$, the process $\widehat{W}_n(t)$ has the representation*

$$\widehat{W}_n(t) = W_n(t) + tm_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) + t^2 o_P(1), \quad (5.18)$$

uniformly in $t \in D$,

$$\begin{aligned} m_F(t) &= \mathbb{E} [\text{sign}(X_1 \vee X_2) g(-t|X_1 \vee X_2|)] \\ &= 4 \int_0^\infty g(-tx) F(x) dF(x) - 2 \int_0^\infty g(-tx) dF(x). \end{aligned}$$

Proof. Choose any compact $D \subset \mathbb{R}$. Consider the continuous function

$$h_x(\delta) = \sin(t|x - \delta|) = \begin{cases} \sin(t(x - \delta)) & \text{if } x - \delta \geq 0, \\ -\sin(t(x - \delta)) & \text{if } x - \delta < 0, \end{cases}$$

which is everywhere differentiable except at $\delta = x$. In fact,

$$h'_x(\delta) = -t \text{sign}(x - \delta) \cos(t(x - \delta)), \quad \forall \delta \in \mathbb{R} \setminus \{x\}.$$

Now, whenever $|x| > |\delta|$, we have by the mean value theorem that there exists a $\theta \in (0, 1)$ such that

$$\begin{aligned} h_x(\delta) &= h_x(0) + \delta h'_x(\theta\delta) \\ &= \sin(t|x|) - t\delta \text{sign}(x) \cos(tx) + \delta (h'_x(\theta\delta) - h'_x(0)). \end{aligned}$$

Using the inequality $|\cos(b) - \cos(a)| \leq |b - a|$, we have

$$h_x(\delta) = \sin(t|x - \delta|) = \sin(t|x|) - t\delta \text{sign}(x) \cos(tx) + t^2 O(\delta^2), \quad (5.19)$$

uniformly in $|x| > |\delta|$ and $t \in D$. In the situation where x lies between 0 and δ , there is a point between 0 and δ where the function h_x is nondifferentiable which violates the conditions of the classical mean value theorem. Moreover, in order to proceed we require that the error bound be uniform in x (i.e. independent of x). In view of these complications, we treat the cases $|x| > |\delta|$ and $|x| \leq |\delta|$ separately in what follows.

Fix $\varepsilon > 0$. Since $\widehat{\eta}_n = O_P(n^{-1/2})$, there exists a $C_\varepsilon > 0$ and an $n_1 = n_1(\varepsilon)$ such that

$$\mathbb{P} \left(|\widehat{\eta}_n| > \frac{C_\varepsilon}{\sqrt{n}} \right) < \frac{\varepsilon}{2} \quad (5.20)$$

for all $n \geq n_1$. Define the event

$$A_{n,k} = \left\{ |X_k| > \frac{C_\varepsilon}{\sqrt{n}} \right\}$$

and consider the decomposition

$$T_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t|X_k - \hat{\eta}_n|) \mathbf{1}(A_{n,k}) + \frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t|X_k - \hat{\eta}_n|) \mathbf{1}(A_{n,k}^c) =: T_n^A(t) + T_n^{A^c}(t). \quad (5.21)$$

We shall now show that

$$T_n^A(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t|X_k|) - t\hat{\eta}_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \text{sign}(X_k) \cos(tX_k) + t^2 o_P(1), \quad (5.22)$$

$$T_n^{A^c}(t) = |t| o_P(1), \quad (5.23)$$

uniformly in $t \in D$.

Using (5.19) and $\hat{\eta}_n^2 = O_P(n^{-1})$, we may expand $T_n^A(t)$ as follows:

$$T_n^A(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t|X_k|) \mathbf{1}(A_{n,k}) - t\hat{\eta}_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \text{sign}(X_k) \cos(tX_k) \mathbf{1}(A_{n,k}) + t^2 o_P(1). \quad (5.24)$$

Clearly,

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t|X_k|) \mathbf{1}(A_{n,k}^c) \right| &\leq \frac{1}{\sqrt{n}} \sum_{k=1}^n |tX_k| \mathbf{1}(A_{n,k}^c) \\ &\leq |t| C_\varepsilon \frac{1}{n} \sum_{k=1}^n \mathbf{1}(A_{n,k}^c) \\ &= |t| C_\varepsilon P(A_{n,1}^c) + |t| o_P(1) \\ &= |t| o_P(1), \end{aligned} \quad (5.25)$$

uniformly in $t \in D$, where we made use of the weak law of large numbers and the continuity of F . More specifically, since F is continuous with median $\eta = 0$ (i.e. $F(0) = \frac{1}{2}$),

$$P(A_{n,1}^c) = P\left(|X_1| \leq \frac{C_\varepsilon}{\sqrt{n}}\right) = 2F\left(\frac{C_\varepsilon}{\sqrt{n}}\right) - 1 \rightarrow 0 \quad (5.26)$$

as $n \rightarrow \infty$. By similar arguments,

$$\left| t\hat{\eta}_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \text{sign}(X_k) \cos(tX_k) \mathbf{1}(A_{n,k}^c) \right| \leq |t\sqrt{n}\hat{\eta}_n| \frac{1}{n} \sum_{k=1}^n \mathbf{1}(A_{n,k}^c) = |t| o_P(1), \quad (5.27)$$

uniformly in $t \in D$. Here we made use of the assumption $\sqrt{n}\hat{\eta}_n = o_P(1)$ and the fact that the cosine function is bounded. In light of (5.25) and (5.27), we see that the representations in (5.22) and (5.24) are asymptotically equivalent.

We now consider the term $T_n^{A^c}(t)$. By the law of total probability and (5.20),

$$\begin{aligned} P\left(|T_n^{A^c}(t)| > \varepsilon\right) &= P\left(|T_n^{A^c}(t)| > \varepsilon, |\hat{\eta}_n| \leq \frac{C_\varepsilon}{\sqrt{n}}\right) + P\left(|T_n^{A^c}(t)| > \varepsilon, |\hat{\eta}_n| > \frac{C_\varepsilon}{\sqrt{n}}\right) \\ &\leq P\left(|T_n^{A^c}(t)| > \varepsilon, |\hat{\eta}_n| \leq \frac{C_\varepsilon}{\sqrt{n}}\right) + \frac{\varepsilon}{2}, \end{aligned} \quad (5.28)$$

for all $n \geq n_1$. Now, whenever $|\hat{\eta}_n| \leq C_\varepsilon n^{-1/2}$, it also holds that

$$\begin{aligned} \left| T_n^{A^c}(t) \right| &= \left| \frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(t|X_k - \hat{\eta}_n|) \mathbf{1}(A_{n,k}^c) \right| \\ &\leq \frac{1}{\sqrt{n}} \sum_{k=1}^n |t| |X_k - \hat{\eta}_n| \mathbf{1}(A_{n,k}^c) \\ &\leq 2|t|C_\varepsilon \frac{1}{n} \sum_{k=1}^n \mathbf{1}(A_{n,k}^c). \end{aligned}$$

From this inequality, and by making use of the Chebyshev inequality, it follows that

$$\begin{aligned} \mathbb{P}\left(\left|T_n^{A^c}(t)\right| > \varepsilon, |\hat{\eta}_n| \leq \frac{C_\varepsilon}{\sqrt{n}}\right) &\leq \mathbb{P}\left(2|t|C_\varepsilon \frac{1}{n} \sum_{k=1}^n \mathbf{1}(A_{n,k}^c) > \varepsilon, |\hat{\eta}_n| \leq \frac{C_\varepsilon}{\sqrt{n}}\right) \\ &\leq \mathbb{P}\left(2|t|C_\varepsilon \frac{1}{n} \sum_{k=1}^n \mathbf{1}(A_{n,k}^c) > \varepsilon\right) \\ &\leq 2|t|C_\varepsilon \frac{1}{\varepsilon} \mathbb{P}(A_{n,1}^c). \end{aligned}$$

The result in (5.26) implies that, for any arbitrary positive constant and in particular also for $\delta_\varepsilon = \varepsilon(4|t|C_\varepsilon)^{-1}$, there exists an $n_2 = n_2(\varepsilon)$ such that $\mathbb{P}(A_{n,1}^c) < \delta_\varepsilon$ for all $n \geq n_2$. Therefore, for all $n \geq n_3 = n_3(\varepsilon) := \max(n_1, n_2)$, we have

$$\mathbb{P}\left(\left|T_n^{A^c}(t)\right| > \varepsilon, |\hat{\eta}_n| \leq \frac{C_\varepsilon}{\sqrt{n}}\right) \leq \frac{\varepsilon}{2}.$$

This, in conjunction with (5.28), confirms that there exists an n_3 such that

$$\mathbb{P}\left(\left|T_n^{A^c}(t)\right| > \varepsilon\right) < \varepsilon$$

for all $n \geq n_3$. This verifies (5.23).

Again using Taylor expansion, but with none of the technical difficulties encountered above, one sees that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{k=1}^n \cos(t|X_k - \hat{\eta}_n|) &= \frac{1}{\sqrt{n}} \sum_{k=1}^n \cos(tX_k) + t\hat{\eta}_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \sin(tX_k) + t^2 o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^n \cos(t|X_k|) + t\hat{\eta}_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \text{sign}(X_k) \sin(t|X_k|) + t^2 o_P(1), \end{aligned}$$

which, together with (5.21)–(5.23), shows that

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n g(t|X_k - \hat{\eta}_n|) = \frac{1}{\sqrt{n}} \sum_{k=1}^n g(t|X_k|) - t\hat{\eta}_n \frac{1}{\sqrt{n}} \sum_{k=1}^n \text{sign}(X_k) g(-t|X_k|) + t^2 O_P(n^{-1/2}).$$

Observing that $g(t|(X_j - \hat{\eta}_n) \vee (X_k - \hat{\eta}_n|)) = g(t|(X_j \vee X_k) - \hat{\eta}_n|)$, we have by exactly the same arguments that

$$\begin{aligned} \frac{1}{\sqrt{n(n-1)}} \sum_{j \neq k} g(t|(X_j - \hat{\eta}_n) \vee (X_k - \hat{\eta}_n|)) &= \frac{1}{\sqrt{n(n-1)}} \sum_{j \neq k} \sum_{k=1}^n g(t|X_j \vee X_k|) \\ &\quad - t\hat{\eta}_n \frac{1}{\sqrt{n(n-1)}} \sum_{j \neq k} \sum_{k=1}^n \text{sign}(X_k) g(-t|X_j \vee X_k|) + t^2 O_P(n^{-1/2}). \end{aligned}$$

Therefore, under the stated assumptions, the process $\widehat{W}_n(t)$ has the representation

$$\widehat{W}_n(t) = W_n(t) - t\sqrt{n}\widehat{\eta}_n \left(\bar{W}_n^{(1)}(t) - \bar{W}_n^{(2)}(t) \right) + t^2 O_P \left(n^{-1/2} \right),$$

where we define

$$\begin{aligned} \widehat{W}_n^{(1)}(t) &= \frac{1}{n} \sum_{k=1}^n \text{sign}(X_k) g(-t|X_k|), \\ \widehat{W}_n^{(2)}(t) &= \frac{1}{n(n-1)} \sum_{j \neq k} \text{sign}(X_j \vee X_k) g(-t|X_j \vee X_k|). \end{aligned}$$

Under H_0 , the random variables $\text{sign}(X_k)$ and $|X_k|$ are independent and hence, making use of the fact that the trigonometric functions are bounded,

$$\mathbb{E} \left(\bar{W}_n^{(1)}(t) \right) = \mathbb{E} \left[\text{sign}(X_1) g(-t|X_1|) \right] = \mathbb{E} \left[\text{sign}(X_1) \right] \mathbb{E} \left[g(-t|X_1|) \right] = 0,$$

for all $t \in \mathbb{R}$. Furthermore, straightforward calculations show that

$$\mathbb{E} \left(\bar{W}_n^{(2)}(t) \right) = 4 \int_0^\infty g(-tx) F(x) dF(x) - 2 \int_0^\infty g(-tx) dF(x) =: m_F(t).$$

Now, by making use of Andrews' (1992) stochastic equicontinuity condition as in the proof of Theorem 5.3, one can show that

$$\sup_{t \in D} \left| \bar{W}_n^{(1)}(t) \right| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t \in D} \left| \bar{W}_n^{(2)}(t) - m_F(t) \right| \xrightarrow{P} 0.$$

We have now established that the process $\widehat{W}_n(t)$ has the representation

$$\widehat{W}_n(t) = W_n(t) + t m_F(t) \sqrt{n} \widehat{\eta}_n + t^2 O_P \left(n^{-1/2} \right),$$

uniformly in $t \in D$. Substituting the expansion in Assumption (S3) for $\sqrt{n}\widehat{\eta}_n$ in the above expression yields (5.18). $\square$

Having at our disposal the representation of $\widehat{W}_n$ as a sum of independent random processes, we may use a standard central limit theorem to find the asymptotic null distribution of $\widehat{W}_n$.

Theorem 5.4. *Under H_0 , Assumptions (S1)–(S3) and the condition $\int_{-\infty}^\infty t^4 w(t) dt < \infty$, there exists a zero-mean Gaussian process $\mathcal{W} = \{\mathcal{W}(t); t \in \mathbb{R}\}$ such that*

$$A_n \xrightarrow{d} \|\mathcal{W}\|^2,$$

as $n \rightarrow \infty$.

Proof. Choose any compact set $D \subset \mathbb{R}$. By (5.18) and the assumption that $\int_{-\infty}^{\infty} t^4 w(t) dt < \infty$, we have

$$\int_D \widehat{W}_n^2(t) w(t) dt = \int_D \left(W_n(t) + t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right)^2 w(t) dt + o_P(1). \quad (5.29)$$

To see why the cross-term appearing in the derivation of the above expression is negligible, first note that

$$\begin{aligned} & \int_D \left| W_n(t) + t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right| t^2 w(t) dt \\ & \leq \int_D W_n^2(t) t^2 w(t) dt + \left| \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right| \int_D |m_F(t) t^3| w(t) dt \\ & = O_P(1), \end{aligned}$$

where we have made use of the fact that W_n and m_F are bounded, $n^{-1/2} \sum_{k=1}^n \psi(X_k) = O_P(1)$ by the central limit theorem, and $\int_{-\infty}^{\infty} t^4 w(t) dt < \infty$. Again using (5.18), we get that

$$\begin{aligned} & \left| \int_D \left(W_n(t) + t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right) \left(\widehat{W}_n(t) - W_n(t) - t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right) w(t) dt \right| \\ & \leq \int_D \left| W_n(t) + t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right| t^2 w(t) dt o_P(1) = o_P(1). \end{aligned}$$

Now, using the same reasoning as above along with the result in (5.9), we find that (5.29) becomes

$$\begin{aligned} \int_D \widehat{W}_n^2(t) w(t) dt &= \int_D \left(\widetilde{W}_n(t) + t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right)^2 w(t) dt + o_P(1) \\ &=: \int_D \check{W}_n^2(t) w(t) dt + o_P(1). \end{aligned} \quad (5.30)$$

We will now use Theorem 22 of Ibragimov and Has'minskii (1981) to show that there exists a Gaussian process $\mathcal{W}(\cdot)$ such that

$$\int_D \check{W}_n^2(t) w(t) dt \xrightarrow{d} \int_D \mathcal{W}^2(t) w(t) dt. \quad (5.31)$$

We only need to verify Conditions (1)–(3) of Theorem 22:

(1) Observe that the first condition is satisfied, since

$$\begin{aligned} \mathbb{E} \int_D \check{W}_n^2(t) w(t) dt &= \mathbb{E} \int_D \left(\widetilde{W}_n(t) + t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right)^2 w(t) dt \\ &\leq 2 \mathbb{E} \int_D \widetilde{W}_n^2(t) w(t) dt + \frac{2d_\eta^2}{n} \mathbb{E} \left(\sum_{k=1}^n \psi(X_k) \right)^2 \int_D t^2 m_F^2(t) w(t) dt \\ &< \infty, \end{aligned}$$

where we made use of the boundedness of $\widetilde{W}_n$ and m_F , the integrability condition on w , and Rosenthal's inequality (see, e.g., Petrov, 1995, p. 59) to confirm that $\mathbb{E}[(\sum_{k=1}^n \psi(X_k))^2] = O(n)$.

- (2) Fix any $s, t \in D$. By the mean value theorem and the boundedness of the sine and cosine functions,

$$|g(t|X_k|) - g(s|X_k|)| \leq |X_k||t - s|,$$

with g as in (5.7). Therefore, recalling the definition of $\widetilde{W}_n$ in (5.8), we get by Rosenthal's inequality that

$$\begin{aligned} & \mathbb{E} \left[(\widetilde{W}_n(t) - \widetilde{W}_n(s))^2 \right] \\ & \leq C_1 \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[(g(t|X_k|) + \Psi(t) - 2\Psi(t|X_k|) - g(s|X_k|) - \Psi(s) + 2\Psi(s|X_k|))^2 \right] \\ & \leq C_2(t-s)^2 \mathbb{E}(X_k^2) \\ & \leq C_3(t-s)^2, \end{aligned}$$

for some $C_1, C_2, C_3 > 0$. Therefore, by the Cauchy–Schwarz inequality and the boundedness of $\widetilde{W}_n$,

$$\begin{aligned} \mathbb{E}(|\widetilde{W}_n^2(t) - \widetilde{W}_n^2(s)|) & \leq \sqrt{\mathbb{E}(|\widetilde{W}_n(t) - \widetilde{W}_n(s)|^2) \mathbb{E}(|\widetilde{W}_n(t) + \widetilde{W}_n(s)|^2)} \\ & \leq C_4 \sqrt{\mathbb{E}(|\widetilde{W}_n(t) - \widetilde{W}_n(s)|^2)} \\ & \leq C_5|t - s|, \end{aligned}$$

for some $C_4, C_5 > 0$. Also, since $\mathbb{E}[(n^{-1/2} \sum_{k=1}^n \psi(X_k))^2] = O(1)$ as pointed out above,

$$\mathbb{E} \left[\left| \left(t m_F(t) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right)^2 - \left(s m_F(s) \frac{d_\eta}{\sqrt{n}} \sum_{k=1}^n \psi(X_k) \right)^2 \right| \right] \leq C_6|t - s|,$$

for some $C_6 > 0$, where we made use of the fact that $|m_F(t) - m_F(s)| \leq K|t - s|$ for some $K > 0$. Hence,

$$\mathbb{E} \left[(\check{W}_n(t) - \check{W}_n(s))^2 \right] \leq C_7|t - s|,$$

for some $C_7 > 0$, which verifies Condition (2) of Theorem 22 of Ibragimov and Has'minskii (1981).

- (3) Since $\check{W}_n$ is a suitably-normalised sum of independent zero-mean random processes with finite covariance, it follows directly by the multivariate central limit theorem that, for any arbitrarily chosen set of fixed points $-\infty < t_1 < t_2 < \dots < t_k < \infty$, the distribution of $(\check{W}_n(t_1), \dots, \check{W}_n(t_k))$ converges weakly to a multivariate normal distribution, say $(\mathcal{W}(t_1), \dots, \mathcal{W}(t_k))$.

This confirms the convergence in (5.31).

By Slutsky's lemma, it now follows from (5.30) that

$$\int_D \widehat{W}_n^2(t)w(t)dt \xrightarrow{d} \int_D \mathcal{W}^2(t)w(t)dt.$$

Fix a $\nu > 0$. Since $w(\cdot)$ is integrable and $\widehat{W}_n^2(t)$ is bounded, there exist a compact set $F_\nu \in \mathbb{R}$ and a constant C_ν such that

$$\mathbb{E} \left(\int_{\mathbb{R} \setminus F_\nu} \widehat{W}_n^2(t)w(t)dt \right) \leq C_\nu \int_{\mathbb{R} \setminus F_\nu} w(t)dt < \nu, \quad \text{for all } n.$$

Analogously, we can choose $F_\nu \in \mathbb{R}$ and C_ν such that also

$$\mathbb{E} \left(\int_{\mathbb{R} \setminus F_\nu} \mathcal{W}^2(t)w(t)dt \right) \leq C_\nu \int_{\mathbb{R} \setminus F_\nu} w(t)dt < \nu, \quad \text{for all } n.$$

This proves the theorem. □

Chapter 6

Empirical study

In this chapter we present a short application of the newly-proposed tests for symmetry to real data.

As a practical application we consider data from a study by the US National Institute of Standards and Technology regarding the calibration of ultrasonic field measurements of depths of defects in the Alaskan pipeline relative to more accurate laboratory measurements. The data that support the findings of this study were obtained from the NIST/SEMATECH e-Handbook of Statistical Methods, which is openly available at <https://www.itl.nist.gov/div898/handbook/pmd/section6/pmd62.htm> (National Institute of Standards and Technology, 2012). In the study, the investigators considered the simple linear regression model

$$Y = \beta_0 + \beta_1 x + \varepsilon,$$

where the response variable Y is the field measurement and the regressor x the laboratory measurement. A scatter plot of the data, a scatter plot of the residuals obtained from a least-squares fit, and a kernel density estimate (using Silverman's normal reference rule as the choice of the bandwidth) of the error density are shown in the left-hand column of Figure 6.1.

The least-squares estimates of β_0 and β_1 are shown in the first column of Table 6.1, along with the p -value of the Shapiro–Wilk (SW) test for normality of the regression errors. The p -values of some of the tests for symmetry considered in this thesis are also shown. $\hat{U}_n^{(I)}$ and $\hat{W}_n^{(I)}$ denote the tests introduced in Chapter 3 based on the integrated empirical distribution function. $\hat{U}_n^h$ and $\hat{W}_n^h$ denote the tests of Chapter 4 based on the kernel distribution function estimation. For the remainder of the tests, we used the same notation as defined in Section 3.3.

Clearly, at the 5% significance level it is doubtful whether the errors can be assumed to have a normal distribution. Therefore, the classical linear regression

assumption of normality of the errors is violated. However, all the symmetry tests, including the new tests based on $\widehat{U}_n^{(I)}$, $\widehat{W}_n^{(I)}$, $\widehat{U}_n^h$ and $\widehat{W}_n^h$, do not reject the null hypothesis of symmetry of the errors.

As was mentioned in Chapter 1, symmetry of the regression errors can be advantageous in estimation and inference. For example, if it is known that the errors have a distribution symmetric around zero, adaptive estimators of the regression coefficients can attain the same efficiency as the maximum-likelihood estimators if the true distribution of the errors were known (Bickel, 1982). Therefore, based on the results in Table 6.1, it is not unreasonable to assume that the true regression errors have a symmetric distribution in this application, which allows one to apply adaptive estimators (such as those constructed by Newey, 1988) to efficiently estimate the regression parameters. As mentioned by Fan and Gencay (1995), this motivates the need for consistent tests of symmetry in linear regression models in order to verify the assumption of symmetry before applying the adaptive estimators.

In the original study the investigators pointed out that the residuals exhibit some degree of heteroskedasticity. This violates one of the main assumptions on which classical inferential procedures are based. For example, the estimated standard errors (used in standard tests for significance of regression parameters) will be unreliable in this application as they rely on the assumption of homoskedasticity. As a remedy the investigators proposed a logarithmic transformation of both the response and regressor, i.e. they considered the model

$$\ln Y = \beta_0 + \beta_1 \ln x + \varepsilon.$$

This transformation was shown to eliminate the problem of heteroskedasticity. However, looking at the p -values of the symmetry tests applied to the after-transformation residuals (right-hand side of Table 6.1) one sees that this transformation introduces skewness in the residuals, which is a violation of classical linear regression assumptions. The conclusion of skewness is supported by the density plots in Figure 6.1, where some degree of skewness is evident in the right-hand plot.

This application illustrates that transforming data to address one violation of regression assumptions (here, heteroskedasticity of the errors) could introduce violation of another assumption (nonnormality of the errors). In this particular application, a solution could be to use the original (untransformed) data and proceed under the assumption of symmetry of the errors, which would not be unreasonable given the results from the tests for symmetry. Due to heteroskedasticity of the errors, one no longer has the guarantee provided by the Gauss–Markov theorem that

the least-squares estimators of β_0 and β_1 will be the most efficient unbiased estimators. Nevertheless, as was mentioned above, assuming symmetry of the regression errors allows one to use adaptive estimators to overcome the issue of efficiency. Alternatively, one could use heteroskedasticity-consistent estimators of the standard errors of the estimators (e.g., the well-known estimators proposed by Andrews, 1991) which would allow one to proceed with classical inferential procedures.

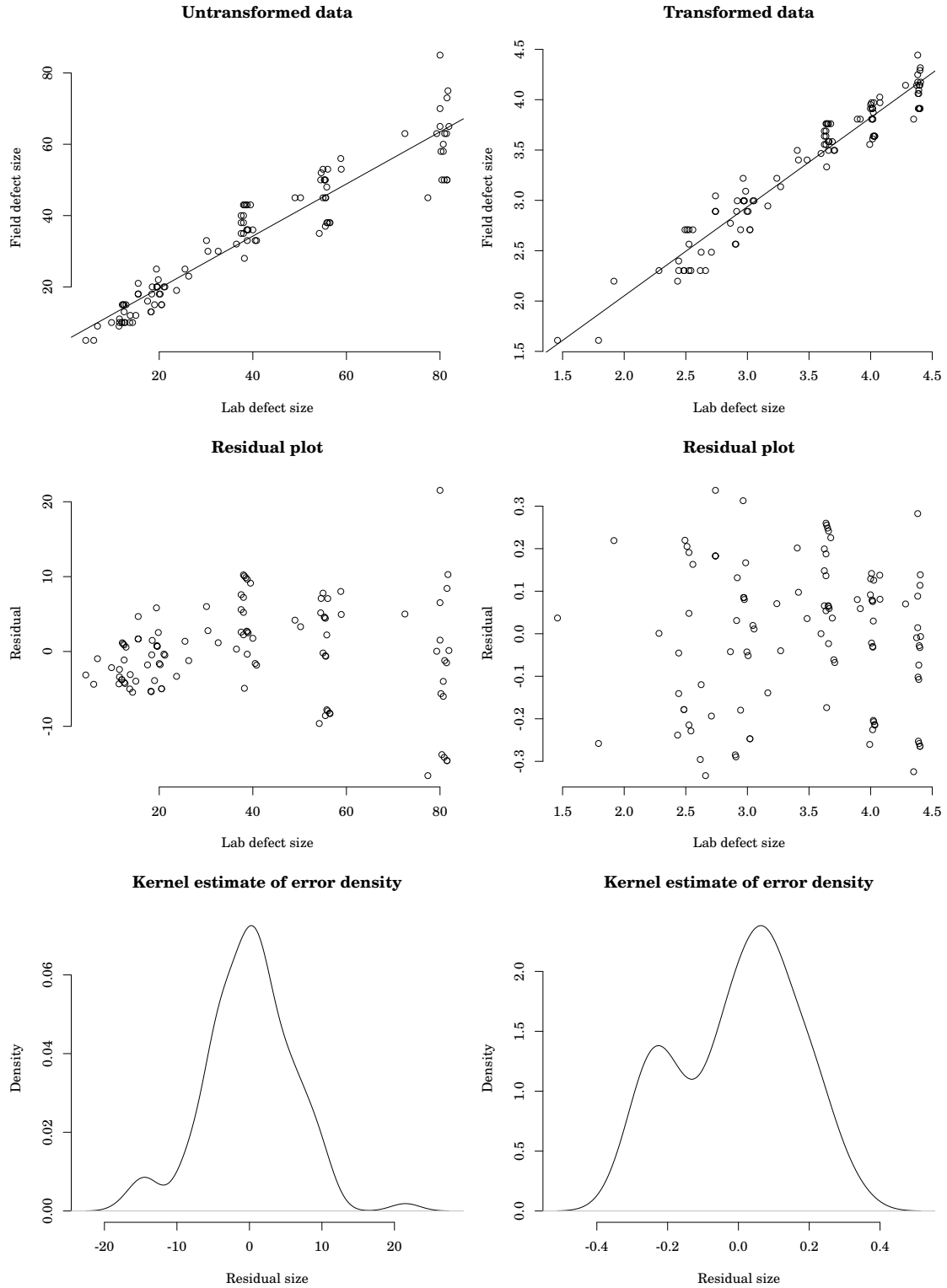


Figure 6.1: Scatter plots of the untransformed (left) and transformed (right) data and residuals, along with plots of the kernel density estimates calculated from the residuals.

Table 6.1: Model parameter estimates for the ultrasonic calibration data along with p -values of tests for normality and symmetry based on the regression residuals. All p -values were obtained using the wild bootstrap scheme described in Section 2.6.2 with 2000 bootstrap replications. For the new kernel-based, smoothed tests of Chapter 4, the Epanechnikov kernel defined in (4.7) was used, with the bandwidth selected using the cross-validation procedure described in Section 4.3.4.

		Untransformed data	ln-ln transformation
Parameter estimates	β_0	4.994	0.281
	β_1	0.731	0.885
Test for normality	SW	0.048	0.004
New tests for symmetry	$\hat{U}_n^{(I)}$	0.530	0.007
	$\widehat{W}_n^{(I)}$	0.617	0.011
	$\hat{U}_n^h$	0.476	0.006
	$\widehat{W}_n^h$	0.644	0.005
Existing tests for symmetry	$\hat{U}_n$	0.769	0.009
	$\widehat{W}_n$	0.697	0.002
	B	0.348	0.004
	CM	0.751	0.011
	DQ	0.591	0.014
	Z	0.573	0.048
	TW_0	0.362	0.035
	$TW_{.05}$	0.332	0.050
	$TW_{.17}$	0.556	0.020
	$H_{0.5}^{(1)}$	0.359	0.016
	$H_1^{(1)}$	0.327	0.044
	$H_{1.5}^{(1)}$	0.317	0.057
	$H_{0.5}^{(2)}$	0.310	0.070
	$H_1^{(2)}$	0.312	0.074
	$H_{1.5}^{(2)}$	0.312	0.076

Chapter 7

Conclusions

In this study we proposed new characterisation-based tests for symmetry of a univariate distribution around an unspecified centre. All newly-proposed tests were shown to be consistent against general asymmetric alternatives and the null distributions of the test statistics were derived. The theoretical results were established in the general setting where the problem is to test for symmetry of the errors appearing in a linear regression model. Numerical studies presented in the thesis showed that the newly-proposed tests are competitive alternatives to existing tests in terms of both level and power.

We investigated the effect of using kernel-smoothed distribution function estimators over the classical empirical distribution function in classical goodness-of-fit tests for symmetry. It was found that significant power improvements seen by other authors in other contexts (e.g., Racine and van Keilegom, 2020) are also possible when using kernel smoothing in tests for symmetry.

In Chapter 5 we revisited a test recently proposed by Allison and Pretorius (2017) and showed that this test is indeed consistent against general alternatives. The null distribution of the test was also derived.

Appendix A

Some basic results on kernel smoothing

Lemma A.1. *Suppose $\{h_n\}$ is a sequence of positive real numbers satisfying*

$$(i) h_n \rightarrow 0, \quad (ii) nh_n \rightarrow \infty, \quad (iii) \ln h_n^{-1} \rightarrow o(nh_n), \quad (iv) \ln h_n^{-1} / \ln \ln n \rightarrow \infty$$

as $n \rightarrow \infty$. Let $\beta_n(t) = \sqrt{n}(F_n(t) - F(t))$ for all $t \in \mathbb{R}$ and all $n \in \mathbb{N}$. Then, for any kernel k satisfying Assumption (K1), we have

$$\sup_{t \in \mathbb{R}} \left| \int_{-\infty}^{\infty} [\beta_n(t - h_n z) - \beta_n(t)] k(z) dz \right| = O_P \left(\sqrt{h_n \ln h_n^{-1}} \right),$$

as $n \rightarrow \infty$.

Proof. Set $a_n = ah_n$, where a is the constant defined in Assumption (C1). It follows from (2.13) of Stute (1982) that there exist an $n_0 > 0$ and a positive constant $d < \infty$ such that

$$\begin{aligned} |\beta_n(t - h_n z) - \beta_n(t)| &\leq \sup_{|u-v| \leq |z|h_n} |\beta_n(u) - \beta_n(v)| \\ &\leq \sup_{|u-v| \leq a_n} |\beta_n(u) - \beta_n(v)| \\ &\leq d \sqrt{a_n \ln a_n^{-1}} \end{aligned}$$

almost surely for all $n \geq n_0$ and for all $|z| \leq a$. Notice that the right-hand side neither depends on z nor t . This, together with the assumption that K has bounded support and that $\int k(z) dz = 1$, completes the proof. $\square$

The next corollary shows that bandwidths of the form $h_n = cn^{-\gamma}$, for some constants $c > 0$ and $0 < \gamma < 1$, satisfy properties (i)–(iv) required by Lemma A.1. To simplify the mathematical expressions in the main text, we assume that h_n is of this form.

Corollary A.1. *Let Assumption (K1) hold and define the bandwidth sequence $\{h_n\}$ by $h_n = cn^{-\gamma}$, for some constants $c > 0$ and $0 < \gamma < 1$. Then, as $n \rightarrow \infty$,*

$$\sup_{t \in \mathbb{R}} \left| \int_{-\infty}^{\infty} [\beta_n(t - h_n z) - \beta_n(t)] k(z) dz \right| = o_P(1). \quad (\text{A.1})$$

Proof. Without loss of generality we may take $c = 1$. We then have

- (i) $h_n \rightarrow 0$;
- (ii) $nh_n \sim n^{1-\gamma} \rightarrow \infty$;
- (iii) $\frac{\ln h_n^{-1}}{nh_n} \sim \frac{\ln n^\gamma}{n^{1-\gamma}} = \frac{\gamma \ln n}{n^{1-\gamma}} \sim \frac{n^{-1}}{n^{-\gamma}} \rightarrow 0$;
- (iv) $\frac{\ln h_n^{-1}}{\ln \ln n} \sim \frac{\ln n^\gamma}{\ln \ln n} = \frac{\gamma \ln n}{\ln \ln n} \sim \frac{1/n}{(\ln n)^{-1}(1/n)} = \ln n \rightarrow \infty$.

Hence, all the requirements of Lemma A.1 are satisfied. $\square$

A.1 Proof of Lemma 4.1

Using integration by parts, observe that

$$\begin{aligned} F_n^h(t) &= \int_{-\infty}^{\infty} K\left(\frac{t-x}{h_n}\right) dF_n(x) \\ &= \left[K\left(\frac{t-x}{h_n}\right) F_n(x) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F_n(x) dK\left(\frac{t-x}{h_n}\right) \\ &= \frac{1}{h_n} \int_{-\infty}^{\infty} F_n(x) k\left(\frac{t-x}{h_n}\right) dx \\ &= \int_{-\infty}^{\infty} F_n(t - h_n z) dK(z). \end{aligned}$$

Therefore we may write

$$F_n^h(t) = \int_{-\infty}^{\infty} [F_n(t - h_n z) - F(t - h_n z)] dK(z) + \int_{-\infty}^{\infty} F(t - h_n z) dK(z). \quad (\text{A.2})$$

Now F is twice differentiable. Therefore, by Taylor expansion, there exists a $\theta_t \in (0, 1)$ such that

$$F(t - h_n z) = F(t) - h_n z f(t) + \frac{1}{2} (h_n z)^2 f'(t - \theta_t h_n z), \quad \text{for each } t \in \mathbb{R}.$$

Furthermore, f' is Hölder continuous, which means that

$$\sup_{t \in \mathbb{R}} |f'(t - \theta_t h_n z) - f'(t)| \leq \sup_{t \in \mathbb{R}} |\theta_t h_n z|^\alpha < |h_n z|^\alpha,$$

for some $\alpha > 0$, so that we may write

$$F(t - h_n z) = F(t) - h_n z f(t) + \frac{1}{2} (h_n z)^2 f'(t) + O((h_n z)^{2+\alpha}),$$

uniformly in $t \in \mathbb{R}$. Since k is a symmetric density with bounded support, we have that $\int z dK(z) = 0$ and $\mu_j(K) = \int z^j dK(z) < \infty$ for all $j > 0$, so that

$$\begin{aligned} \int F(t - h_n z) dK(z) &= F(t) - h_n f(t) \int z dK(z) + \frac{1}{2} h_n^2 f'(t) \int z^2 dK(z) + O(h_n^{2+\alpha}) \\ &= F(t) + \frac{1}{2} \mu_2(K) f'(t) h_n^2 + o(h_n^2), \end{aligned}$$

uniformly in $t \in \mathbb{R}$. Combining this result with (A.2) yields

$$F_n^h(t) = F(t) + \int_{-\infty}^{\infty} [F_n(t - h_n z) - F(t - h_n z)] k(z) dz + \frac{1}{2} \mu_2(K) f'(t) h_n^2 + o(h_n^2).$$

Intuitively, if h_n is very small, one would expect the integral on the right-hand side to behave approximately like $F_n(t) - F(t)$. Indeed, using the notation $\beta_n(t) = \sqrt{n}(F_n(t) - F(t))$, we have by Corollary A.1 that

$$\begin{aligned} \sqrt{n} F_n^h(t) &= \sqrt{n} F_n(t) + \int_{-\infty}^{\infty} [\beta_n(t - h_n z) - \beta_n(t)] k(z) dz + \frac{1}{2} \mu_2(K) f'(t) \sqrt{n} h_n^2 + o(\sqrt{n} h_n^2) \\ &= \sqrt{n} F_n(t) + \frac{1}{2} \mu_2(K) f'(t) \sqrt{n} h_n^2 + o_P(1), \end{aligned}$$

uniformly in $t \in \mathbb{R}$, which completes the proof. □

Bibliography

- Ahmad, I. A. and Li, Q. (1997). Testing symmetry of an unknown density function by kernel method, *Journal of Nonparametric Statistics* **7**: 279–293.
- Ahsanullah, M. (1992). On some characteristic property of symmetric distributions, *Pakistan Journal of Statistics* **8**: 19–22.
- Aki, S. (1987). On nonparametric tests for symmetry, *Annals of the Institute of Statistical Mathematics* **39**: 457–472.
- Allison, J. S. and Pretorius, C. (2017). A Monte Carlo evaluation of the performance of two new tests for symmetry, *Computational Statistics* **32**: 1323–1338.
- Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* pp. 817–858.
- Andrews, D. W. K. (1992). Generic uniform convergence, *Econometric Theory* **8**: 241–257.
- Antille, A., Kersting, G. and Zucchini, W. (1982). Testing symmetry, *Journal of the American Statistical Association* **77**: 639–646.
- Arcones, M. A. and Giné, E. (1991). Some bootstrap tests of symmetry for univariate continuous distributions, *The Annals of Statistics* **19**: 1496–1511.
- Baringhaus, L. and Henze, N. (1992). A characterization of and new consistent tests for symmetry, *Communications in Statistics – Theory and Methods* **21**: 1555–1566.
- Beran, R. (1986). Discussion: Jackknife, bootstrap and other resampling methods in regression analysis, *The Annals of Statistics* **14**: 1295–1298.
- Bhattacharya, P. K., Gastwirth, J. L. and Wright, A. L. (1982). Two modified Wilcoxon tests for symmetry about an unknown location parameter, *Biometrika* **69**: 377–382.
- Bickel, P. J. (1982). On adaptive estimation, *The Annals of Statistics* **10**: 647–671.

- Billingsley, P. (1968). *Convergence of Probability Measures*, 2nd edn, New York: John Wiley & Sons.
- Billingsley, P. (1995). *Probability and Measure*, 3rd edn, New York: John Wiley & Sons.
- Boos, D. D. (1982). A test for asymmetry associated with the Hodges–Lehmann estimator, *Journal of the American Statistical Association* **77**: 647–651.
- Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions, *Biometrika* **85**: 799–808.
- Butler, C. C. et al. (1969). A test for symmetry using the sample distribution function, *The Annals of Mathematical Statistics* **40**: 2209–2210.
- Cabilio, P. and Masaro, J. (1996). A simple test of symmetry about an unknown median, *The Canadian Journal of Statistics* **24**: 349–361.
- Csörgő, S. (1984). Testing by the empirical characteristic function: A survey, *Asymptotic Statistics*, Vol. 2, Elsevier, pp. 45–56.
- Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last, *Journal of Econometrics* **146**: 162–169.
- Davis, C. E. and Quade, D. (1978). U -statistics for skewness or symmetry, *Communications in Statistics – Theory and Methods* **7**: 413–418.
- Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statistica Neerlandica* **37**: 73–83.
- Fan, Y. and Gencay, R. (1995). A consistent nonparametric test of symmetry in linear regression models, *Journal of the American Statistical Association* **90**: 551–557.
- Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications, *The Annals of Statistics* **5**: 88–97.
- Finch, S. J. (1977). Robust univariate test of symmetry, *Journal of the American Statistical Association* **72**: 387–392.
- Flachaire, E. (1999). A better way to bootstrap pairs, *Economics Letters* **64**: 257–262.
- Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap, *Computational Statistics and Data Analysis* **49**: 361–376.

- Freedman, D. A. et al. (1981). Bootstrapping regression models, *The Annals of Statistics* **9**: 1218–1228.
- Gaigall, D. (2020). Rothman–Woodrooffe symmetry test statistic revisited, *Computational Statistics and Data Analysis* **142**: 106837.
- Gastwirth, J. L. (1971). On the sign test for symmetry, *Journal of the American Statistical Association* **66**: 821–823.
- Ghosh, J. K. (1971). A new proof of the bahadur representation of quantiles and an application, *The Annals of Mathematical Statistics* pp. 1957–1961.
- Ghosh, S. and Ruymgaart, F. H. (1992). Applications of empirical characteristic functions in some multivariate problems, *The Canadian Journal of Statistics* **20**: 429–440.
- Gupta, M. K. (1967). An asymptotically nonparametric test of symmetry, *The Annals of Mathematical Statistics* **38**: 849–866.
- Henze, N., Klar, B. and Meintanis, S. G. (2003). Invariant tests for symmetry about an unspecified point based on the empirical characteristic function, *Journal of Multivariate Analysis* **87**: 275–297.
- Henze, N. and Nikitin, Y. Y. (2000). A new approach to goodness-of-fit testing based on the integrated empirical process, *Journal of Nonparametric Statistics* **12**: 391–416.
- Henze, N. and Nikitin, Y. Y. (2003). Two-sample tests based on the integrated empirical process, *Communications in Statistics – Theory and Methods* **32**: 1767–1788.
- Hettmansperger, T. P., McKean, J. W. and Sheather, S. J. (2002). Finite sample performance of tests for symmetry of the errors in a linear model, *Journal of Statistical Computation and Simulation* **72**: 863–879.
- Hill, D. L. and Rao, P. V. (1977). Tests of symmetry based on Cramér–von Mises statistics, *Biometrika* **64**: 489–494.
- Hillebrand, M. and Müller, C. H. (2006). On consistency of redescending M-kernel smoothers, *Metrika* **63**: 71–90.
- Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location based on rank tests, *The Annals of Mathematical Statistics* **34**: 598–611.

- Horowitz, J. L. (2001). The bootstrap, *Handbook of Econometrics*, Vol. 5, Elsevier, pp. 3159–3228.
- Huber, P. J. (1981). *Robust Statistics*, New York: John Wiley & Sons.
- Hušková, M. and Meintanis, S. G. (2007). Omnibus tests for the error distribution in the linear regression model, *Statistics* **41**: 363–376.
- Hušková, M. and Meintanis, S. G. (2008a). Testing procedures based on the empirical characteristic functions I: Goodness-of-fit, testing for symmetry and independence, *Tatra Mountains Mathematical Publications* **39**: 225–233.
- Hušková, M. and Meintanis, S. G. (2008b). Testing procedures based on the empirical characteristic functions II: k-sample problem, change point problem, *Tatra Mountains Mathematical Publications* **39**: 235–243.
- Hušková, M. and Meintanis, S. G. (2012). Tests for symmetric error distribution in linear and nonparametric regression models, *Communications in Statistics – Simulation and Computation* **41**: 833–851.
- Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*, New York: Springer.
- Ivanović, B., Milošević, B. and Obradović, M. (2020). Comparison of symmetry tests against some skew-symmetric alternatives in i.i.d. and non-i.i.d. setting, *Computational Statistics and Data Analysis* **151**: 106991.
- Janson, S. (2009). Probability asymptotics: notes on notation. Accessed 05/11/2014. **URL:** <http://arxiv.org/abs/1108.3924>
- Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators, *The Annals of Mathematical Statistics* **40**: 633–643.
- Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd edn, New York: Springer.
- Klar, B. (2001). Goodness-of-fit tests for the exponential and the normal distribution based on the integrated distribution function, *Annals of the Institute of Statistical Mathematics* **53**: 338–353.
- Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edn, New York: Springer.

- Koutrouvelis, I. A. (1985). Distribution-free procedures for location and symmetry inference problems based on the empirical characteristic function, *Scandinavian Journal of Statistics* **12**: 257–269.
- Koziol, J. A. (1980). On a Cramér–von Mises-type statistic for testing symmetry, *Journal of the American Statistical Association* **75**: 161–167.
- Koziol, J. A. (1983). Tests for symmetry about an unknown value based on the empirical distribution function, *Communications in Statistics – Theory and Methods* **12**: 2823–2846.
- Kozubowski, T. J., Podgórski, K. and Rychlik, I. (2013). Multivariate generalized Laplace distribution and related random fields, *Journal of Multivariate Analysis* **113**: 59–72.
- Li, C., Li, H. and Racine, J. S. (2017). Cross-validated mixed-datatype bandwidth selection for nonparametric cumulative distribution/survivor functions, *Econometric Reviews* **36**: 970–987.
- Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.
- Liu, R. Y. (1988). Bootstrap procedures under some non-iid models, *The Annals of Statistics* **16**: 1696–1708.
- Malan, R. D. (2017). *An evaluation of various tests for symmetry*, Master’s thesis, North-West University, Potchefstroom.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models, *The Annals of Statistics* **21**: 255–285.
- McWilliams, T. P. (1990). A distribution-free test for symmetry based on a runs statistic, *Journal of the American Statistical Association* **85**: 1130–1133.
- Meintanis, S. G. (2016). A review of testing procedures based on the empirical characteristic function, *South African Statistical Journal* **50**: 1–14.
- Meintanis, S. G., Allison, J. S. and Santana, L. (2016). Goodness-of-fit tests for semiparametric and parametric hypotheses based on the probability weighted empirical characteristic function, *Statistical Papers* **57**: 957–976.
- Meintanis, S. G., Swanepoel, J. W. H. and Allison, J. S. (2014). The probability weighted characteristic function and goodness-of-fit testing, *Journal of Statistical Planning and Inference* **146**: 122–132.

- Milošević, B. and Obradović, M. (2016). Characterization based symmetry tests and their asymptotic efficiencies, *Statistics and Probability Letters* **119**: 155–162.
- Milošević, B. and Obradović, M. (2019). Comparison of efficiencies of some symmetry tests around an unknown centre, *Statistics* **53**: 43–57.
- Modarres, R. (2002). Efficient nonparametric estimation of a distribution function, *Computational Statistics and Data Analysis* **39**: 75–95.
- Modarres, R. and Gastwirth, J. L. (1996). A modified runs test for symmetry, *Statistics and Probability Letters* **31**: 107–112.
- Modarres, R. and Gastwirth, J. L. (1998). Hybrid test for the hypothesis of symmetry, *Journal of Applied Statistics* **25**: 777–783.
- Nabeya, S. (1987). On Aki's nonparametric test for symmetry, *Annals of the Institute of Statistical Mathematics* **39**: 473–482.
- Nadaraya, E. A. (1964). Some new estimates for distribution functions, *Theory of Probability and Its Applications* **9**: 497–500.
- National Institute of Standards and Technology (2012). NIST/SEMATECH e-Handbook of Statistical Methods.
URL: <https://www.itl.nist.gov/div898/handbook/pmd/section6/pmd62.htm>
- Neuhaus, G. and Zhu, L. X. (1998). Permutation tests for reflected symmetry, *Journal of Multivariate Analysis* **67**: 129–153.
- Neumeyer, N. and Dette, H. (2007). Testing for symmetric error distribution in nonparametric regression models, *Statistica Sinica* **17**: 775–795.
- Neumeyer, N., Dette, H. and Nagel, E. R. (2005). A note on testing symmetry of the error distribution in linear regression models, *Journal of Nonparametric Statistics* **17**: 697–715.
- Newey, W. K. (1988). Adaptive estimation of regression models via moment restrictions, *Journal of Econometrics* **38**: 301–339.
- Newey, W. K. and Steigerwald, D. G. (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models, *Econometrica: Journal of the Econometric Society* **65**: 587–599.
- Ngatchou-Wandji, J. and Harel, M. (2013). A Cramér–von Mises test for symmetry of the error distribution in asymptotically stationary stochastic models, *Statistical Inference for Stochastic Processes* **16**: 207–236.

- Nikitin, Y. Y. and Ahsanullah, M. (2015). New U -empirical tests of symmetry based on extremal order statistics, and their efficiencies, in M. Hallin, D. Mason, D. Pfeifer and J. Steinebach (eds), *Mathematical Statistics and Limit Theorems. Festschrift in Honour of Paul Deheuvels*, Springer, pp. 231–248.
- Nolan, J. P. (2013). Multivariate elliptically contoured stable distributions: theory and estimation, *Computational Statistics* **28**: 2067–2089.
- Orlov, A. I. (1972). On testing the symmetry of distributions, *Theory of Probability and Its Applications* **17**: 357–361.
- Pérez-Veiga, N. and Pardo-Fernández, J. C. (2019). The effect of smoothing on tests based on regression residuals. Poster presented at the XoveTIC2019 meeting in Galicia, September, 2019.
- Petrov, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford University Press.
- Racine, J. S. and van Keilegom, I. (2020). A smooth nonparametric, multivariate, mixed-data location-scale test, *Journal of Business and Economic Statistics* **38**: 784–795.
- Randles, R. H., Fligner, M. A., Policello, G. E. and Wolfe, D. A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry, *Journal of the American Statistical Association* **75**: 168–172.
- Reiss, R.-D. (1981). Nonparametric estimation of smooth distribution functions, *Scandinavian Journal of Statistics* **8**: 116–119.
- Rothman, E. D. and Woodroffe, M. (1972). A Cramér–von Mises type statistic for testing symmetry, *The Annals of Mathematical Statistics* **43**: 2035–2038.
- Schuster, E. F. (1973). On the goodness-of-fit problem for continuous symmetric distributions, *Journal of the American Statistical Association* **68**: 713–715.
- Schuster, E. F. and Barker, R. C. (1987). Using the bootstrap in testing symmetry versus asymmetry, *Communications in Statistics – Simulation and Computation* **16**: 69–84.
- Seber, G. A. F. and Lee, A. J. (2012). *Linear Regression Analysis*, New York: John Wiley & Sons.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.

- Silver, E. A. and Costa, D. (1997). A property of symmetric distributions and a related order statistic result, *The American Statistician* **51**: 32–33.
- Smirnov, N. V. (1947). Sur un critère de symétrie de la loi de distribution d'une variable aléatoire, *Compte Rendu (Doklady) de l'Academic des Sciences de l' URSS*, Vol. 56, pp. 11–14.
- Srinivasan, R. and Godio, L. (1974). A Cramér–von Mises type statistic for testing symmetry, *Biometrika* **61**: 196–198.
- Staudte, R. G. (2014). Inference for quantile measures of skewness, *TEST* **23**: 751–768.
- Stute, W. (1982). The oscillation behavior of empirical processes, *The Annals of Probability* **10**: 86–107.
- Tenreiro, C. (2009). On the choice of the smoothing parameter for the BHEP goodness-of-fit test, *Computational Statistics and Data Analysis* **53**: 1038–1053.
- Tenreiro, C. (2019). On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests, *Journal of Statistical Computation and Simulation* **89**: 1780–1797.
- van der Vaart, A. W. (2000). *Asymptotic Statistics*, Vol. 3, Cambridge University Press.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, New York: Springer.
- Winter, B. (1973). Strong uniform consistency of integrals of density estimators, *Canadian Journal of Statistics* **1**: 247–253.
- Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *The Annals of Statistics* **14**: 1261–1295.
- Yamato, H. (1973). Uniform convergence of an estimator of a distribution function.
- Yule, G. U. and Kendall, M. G. (1911). *An Introduction to the Theory of Statistics*, 13th edn, London: Griffin.
- Zardasht, V. (2018). A bootstrap test for symmetry based on quantiles, *Hacettepe Journal of Mathematics and Statistics* **47**: 1061–1069.
- Zhao, Y., Gijbels, I. and van Keilegom, I. (2020). Inference for semiparametric gaussian copula model adjusted for linear regression using residual ranks, *Bernoulli* **26**: 2815–2846.