Birkbeck Working Papers in Economics & Finance

ISSN 1745-8587

Department of Economics, Mathematics and Statistics

BWPEF 1806

Bootstrap-Assisted Tests of Symmetry for Dependent Data

Zacharias Psaradakis
Birkbeck, University of London

Márian Vávra
Birkbeck Centre for Applied Macroeconomics, University of London

July 2018

▪ Birkbeck, University of London ▪ Malet Street ▪ London ▪ WC1E 7HX ▪

Bootstrap-Assisted Tests of Symmetry for Dependent Data

Zacharias Psaradakis

Department of Economics, Mathematics and Statistics

Birkbeck, University of London

E-mail: z.psaradakis@bbk.ac.uk

Marián Vávra

Research Department, National Bank of Slovakia

Birkbeck Centre for Applied Macroeconomics, University of London

E-mail: marian.vavra@nbs.sk

July 11, 2018

Abstract

The paper considers the problem of testing for symmetry (about an unknown centre) of the marginal distribution of a strictly stationary and weakly dependent stochastic process. The possibility of using the autoregressive sieve bootstrap and stationary bootstrap procedures to obtain critical values and P-values for symmetry tests is explored. Bootstrap-assisted tests for symmetry are straightforward to implement and require no prior estimation of asymptotic variances. The small-sample properties of a wide variety of tests are investigated using Monte Carlo experiments. A bootstrap-assisted version of the triples test is found to have the best overall performance.

Keywords: Autoregressive sieve bootstrap; Stationary bootstrap; Symmetry; Weak dependence.

1 Introduction

The problem of testing for symmetry of a probability distribution about a specified or unspecified centre has attracted considerable attention in the literature. In view of the fact that many nonparametric and robust statistical procedures rely on the assumption of symmetry, this is not perhaps surprising. Symmetry, or lack of it, is also important in terms of the definition and estimation of location since the centre of symmetry of a distribution is its only natural location parameter. Some well-known problems, such as, for instance, testing for lack of treatment effect through paired comparisons or testing for time-reversibility of a stochastic process, may be reduced to a test for symmetry. In the context of statistical model building, a test for symmetry is a useful specification test since departures from symmetry would imply that certain families of parametric models (e.g., models for stochastic processes that admit a linear representation with respect to independent and symmetrically distributed noise) should not be considered as candidate models. Tests for symmetry may also be useful in evaluating the validity of different hypotheses and models to the extent that the latter rely on or imply distributional symmetry (as is the case, for example, with many of the option and asset pricing models, rational expectations models, and dynamic stochastic general equilibrium models found in the economics and finance literature).

The present paper focuses on the problem of testing for symmetry when the centre is unspecified, as is often the case in applied settings involving real-world data. Unlike most of the voluminous work on this subject, which deals with independent, identically distributed (i.i.d.) observations, we consider the more general case of dependent observations from strictly stationary stochastic processes. Available tests for symmetry of the one-dimensional marginal distribution of dependent data include tests based on moment conditions (Bai and Ng (2005), Psaradakis (2016)), distribution distance measures (Psaradakis (2003), Racine and Maasoumi (2007), Maasoumi and Racine (2009)), the characteristic function (Chen et al. (2000)), and order statistics (Psaradakis and Vávra (2015)). In our analysis, we investigate the properties of a wide variety of tests, most of which have been developed in the context of testing for symmetry of i.i.d. data. When there are deviations from the assumption of independence, such tests cannot, of course, be expected to have the correct level unless the structure of dependence of the observations is taken into account in the construction of critical regions for the tests. To complicate matters further, when the centre of symmetry is unspecified, many of the available tests are based on test statistics the large-sample null distributions of which depend on the unknown marginal distribution of the data and/or other unknown quantities, even under i.i.d. conditions.

We argue that a convenient way of overcoming these difficulties is to use bootstrap approximations to the null sampling distributions of the test statistics of interest. In the presence of dependence, the key issue is how to implement the bootstrap in a way which ensures that bootstrap samples replicate, as close as possible, the dependence structure of the observed data. To do so, we adopt symmetrized versions of two well-known bootstrap procedures for dependent data, namely the autoregressive sieve bootstrap (ARSB) and the stationary bootstrap (STB), introduced by Kreiss (1992) and Politis and Romano (1994), respectively. Under the assumption that the underlying stochastic process admits an infinite-order autoregressive representation, the ARSB approximates the covariance structure of the data by an autoregressive sieve, that is, a sequence of autoregressive models the order of which increases slowly with the sample size. These finite-parameter models are then used to generate bootstrap data by resampling from the estimated residuals. Without assuming a parametric structure for the underlying stochastic process, the STB preserves the short-distance dependence structure of the data by resampling overlapping blocks of consecutive observations, with the block length being random and the average block length growing slowly with the sample size. Unlike other bootstrap methods based on block resampling, the STB produces bootstrap data which retain the stationarity property of the original data.

We explore how these bootstrap procedures may be used to obtain P-values and/or critical values for tests of symmetry (about an unknown centre) of the one-dimensional marginal distribution of strictly stationary and weakly dependent stochastic processes. We consider nineteen different tests, and investigate their level and power properties across six (linear and nonlinear) data-generating processes and eight (symmetric and asymmetric) noise distributions. Our simulation results show that bootstrap procedures provide a practical and efficient way of robustifying symmetry tests to deviations from the assumption of independence, without prior estimation of the asymptotic variance of the relevant test statistics, and deliver tests that achieve good level and power properties even with a relatively small number of bootstrap replications.

The remainder of the paper is organized as follows. Section 2 provides an overview of the symmetry tests of interest. Section 3 discusses how the ARSB and STB may be used to implement tests for symmetry in the presence of serial dependence. Section 4 examines the small-sample properties of bootstrap-assisted tests for symmetry by means of Monte Carlo simulations. Section 5 summarizes and concludes.

2 Tests for Symmetry

2.1 Problem Formulation

Consider a strictly stationary, real-valued, discrete-time stochastic process $\mathcal{X} = \{X_t\}_{t=-\infty}^{\infty}$ with mean $\mu = \mathbb{E}(X_t)$ and finite, positive variance $\sigma^2 = \mathbb{E}[(X_t - \mu)^2]$. It is assumed that $\mathcal{X}$ is weakly dependent in the sense that its autocovariance sequence $\gamma_m = \mathbb{E}[(X_t - \mu)(X_{t+m} - \mu)]$, $m = 0, \pm 1, \ldots$, is absolutely summable. Given an observable segment $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ of the process, the aim is to test the null hypothesis that the one-dimensional marginal distribution of $\mathcal{X}$ is symmetric about μ, that is,

$$X_t - \mu \overset{\mathfrak{D}}{=} \mu - X_t, \tag{1}$$

where the symbol $\overset{\mathfrak{D}}{=}$ denotes equality in distribution. Under (1), $\mathbb{P}(X_t \geq \mu) = \mathbb{P}(X_t \leq \mu) \geq \frac{1}{2}$.

2.2 Tests Based on Measures of Skewness

A classical test for symmetry is based on the empirical third standardized cumulant of X_t. Letting $Z_t = (X_t - \bar{X})/\hat{\sigma}$, where $\bar{X} = n^{-1}\sum_{t=1}^{n} X_t$ and $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}(X_t - \bar{X})^2$, the test rejects for large absolute values of the statistic

$$\mathcal{S}_1 = n^{-1/2}\sum_{t=1}^{n} Z_t^3 \tag{2}$$

(see, e..g., Gupta (1967), Bai and Ng (2005)).

Psaradakis and Vávra (2015) considered a test based on a measure of skeweness that involves a linear combination of quantiles. Letting $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics of $\mathcal{X}_n$, the test rejects for large absolute values of the statistic

$$\mathcal{S}_2 = \sqrt{n}\boldsymbol{\delta}_k'\xi_k, \tag{3}$$

where $\xi_k = (X_{(\lceil nq_1 \rceil)}, \dots, X_{(\lceil nq_k \rceil)}, X_{(\lceil n(1-q_1) \rceil)}, \dots, X_{(\lceil n(1-q_k) \rceil)}, X_{(\lceil n/2 \rceil)})'$ for some positive integer k and constants $0 < q_1 < \cdots < q_k < \frac{1}{2}$, $\boldsymbol{\delta}_k$ is a $(2k+1) \times 1$ vector whose last component is -2 and all other components are $1/k$, and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. As in Psaradakis and Vávra (2015), the number of quantiles k is selected by minimizing the quantity $\log(\boldsymbol{\delta}_k' \hat{\boldsymbol{\Omega}}_k \boldsymbol{\delta}_k) + (k/n) \log n$ in the range $1 \leq k \leq \lfloor \sqrt{n} \rfloor$, where $\hat{\boldsymbol{\Omega}}_k$ is a consistent estimator of the asymptotic covariance matrix of $\sqrt{n}\xi_k$ and $\lfloor x \rfloor$ denotes the largest integer less than or equal to x; the selected quantiles are taken to be evenly spaced over the range $0.05 \leq q_1 < \cdots < q_k < 0.5$. The covariance estimator $\hat{\boldsymbol{\Omega}}_k$ is constructed in the manner described in Psaradakis and Vávra (2015, p. 590), except that the 'plug-in' procedure of Andrews (1991) is used to obtain a bandwidth estimate.

Another class of tests is based on the empirical analogue of measures of skewness which involve the difference between the mean and a median of X_t. Putting $\ddot{X} = X_{(\lceil n/2 \rceil)}$, the test of Cabilio and Masaro (1996) rejects for large absolute values of the statistic

$$\mathcal{S}_3 = (n/\hat{\sigma}^2)^{1/2}(\bar{X} - \ddot{X}), \tag{4}$$

which is a (normalized) empirical analogue of Yule's measure of skewness. Miao et al. (2006) modified the test by using the mean deviation about the median as a measure of dispersion. Their test statistic can be written as

$$\mathcal{S}_4 = n^{3/2}(2/\pi)^{1/2} \left(\sum_{t=1}^{n} |X_t - \ddot{X}| \right)^{-1} (\bar{X} - \ddot{X}), \tag{5}$$

and may be regarded as a robustified empirical analogue of Yule's measure of skewness (see also Ekström and Jammalamadaka (2012)). Finally, the test considered by Mira (1999) rejects for large absolute values of the statistic

$$\mathcal{S}_5 = \sqrt{n}(\bar{X} - \ddot{X}), \tag{6}$$

which is the (normalized) sample analogue of Bonferroni's measure of skewness.

The test statistics (2)–(6) have Gaussian asymptotic null distributions with zero mean, but their asymptotic variances depend on the unknown distribution of X_t (except that of $\mathcal{S}_1$) and on other unknown quantities. Holgersson (2007), Zheng and Gastwirth (2010), and Lyubchich et al. (2016) investigated how suitable bootstrap procedures may be used to approximate the distributions of $\mathcal{S}_1$, $\mathcal{S}_3$, $\mathcal{S}_4$ and $\mathcal{S}_5$ for independent data.

2.3 Tests based on Distribution, Density, and Characteristic Functions

Boos (1982) considered a test based on the Cramér–von Mises distance between the empirical distribution function of $\mathcal{X}_n$ and its symmetrization with respect to the Hodges–Lehmann location estimator. The test rejects for large values of the statistic

$$\mathcal{S}_6 = n\left\{\left(2\sum\sum_{1\le t<s\le n}|X_t - X_s|\right)^{-1}\left(\sum_{t=1}^{n}\sum_{s=1}^{n}|X_t + X_s - 2\ddot{W}|\right) - 1\right\}, \tag{7}$$

where $\ddot{W}$ is the median of the pairwise averages $W_{ts} = \frac{1}{2}\{X_{(t)} + X_{(s)}\}$, $1 \le t \le s \le n$. The asymptotic null distribution of $\mathcal{S}_6$, obtained by Boos (1982) under i.i.d. conditions, depends on the distribution of X_t.

The test proposed by Schuster and Barker (1987) is based on the Kolmogorov distance between the empirical distribution function of $\mathcal{X}_n$ and its symmetrization with respect to the Schuster–Narvarte estimator of the centre of symmetry. The test rejects for large values of the statistic

$$\mathcal{S}_7 = n^{-1/2}\min\{\ell : M_1(\ell) \le M_2(\ell), 0 \le \ell \le n-1\}, \tag{8}$$

where

$$M_1(\ell) = \max\left\{W_{ts} : 1 \le t \le \lfloor (n-\ell+1)/2\rfloor, s = n-\ell+1-t\right\},$$

$$M_2(\ell) = \min\left\{W_{ts} : \ell+1 \le t \le \lfloor (n+\ell+1)/2\rfloor, s = n+\ell+1-t\right\}.$$

Since $\mathcal{S}_7$ is not asymptotically distribution-free, Schuster and Barker (1987) used a bootstrap approximation to its null sampling distribution assuming i.i.d. conditions (see also Arcones and Giné (1991)). An ARSB-based version of the test appropriate for weakly dependent data was investigated in Psaradakis (2003).

Racine and Maasoumi (2007) and Maasoumi and Racine (2009) considered a test based on an empirical analogue of the squared Hellinger–Matusita distance between the distributions of X_t and $2\mu - X_t$. The test rejects for large values of the statistic

$$\mathcal{S}_8 = \frac{1}{2}\int_{-\infty}^{\infty}\left\{\hat{f}_1^{1/2}(x) - \hat{f}_2^{1/2}(x)\right\}^2 \mathrm{d}x, \tag{9}$$

where $\hat{f}_1$ and $\hat{f}_2$ are nonparametric kernel estimators of the density functions of X_t and $2\bar{X} - X_t$, respectively. Maasoumi and Racine (2009) approximated the null distribution of $\mathcal{S}_8$ by means of a STB procedure to take dependence into account. In our implementation of the test, kernel density estimates are obtained using a standard Gaussian kernel and the normal reference bandwidth given in Silverman (1986, p. 47).

Exploiting the fact that the characteristic function of $X_t - \mu$ is real-valued under (1), Henze et al. (2003) proposed a test based on the imaginary part of the empirical characteristic function of standardized data. The test rejects for large values of the test statistic

$$\mathcal{S}_9 = \frac{(\pi/c)^{1/2}}{2n} \sum_{t=1}^{n} \sum_{s=1}^{n} \left\{ \exp\left(-\frac{1}{4c}(Z_t - Z_s)^2\right) - \exp\left(-\frac{1}{4c}(Z_t + Z_s)^2\right) \right\}, \tag{10}$$

where $c > 0$ is a pre-specified constant (we set $c = 1$). Since the asymptotic null distribution of $\mathcal{S}_9$ depends on the distribution of X_t, the authors considered a permutation procedure for implementing the test under i.i.d. conditions.

The family of tests proposed by Chen et al. (2000) also exploits the fact that the imaginary part of the characteristic function of a symmetric random variable equals zero. The test considered here rejects for large absolute values of the statistic

$$\mathcal{S}_{10} = n^{-1/2} \sum_{t=1}^{n} \left(\frac{Z_t}{1 + Z_t^2} \right). \tag{11}$$

Under general conditions that allow for weak dependence, the asymptotic null distribution of $\mathcal{S}_{10}$ is Gaussian with zero mean but variance which depends on unknown quantities (see Chen (2001)).

Premaratne and Bera (2005) developed a score test for symmetry within the family of Pearson Type IV distributions. The test rejects for large absolute values of the statistic

$$\mathcal{S}_{11} = n^{-1/2} \sum_{t=1}^{n} \arctan Z_t, \tag{12}$$

which is asymptotically normal with zero mean (but not asymptotically pivotal) under the null hypothesis.

We note that tests based on statistics such as (2), (11) and (12) (as well as (13) and (15) below) may also be viewed as belonging to a general class of tests which exploit the fact that the expectation (if it exists) of an odd function of $X_t - \mu$ equals zero when (1) is true. Psaradakis (2016) considered ARSB-based versions of such tests which are valid under general serial dependence conditions that allow for autocovariance sequences that may or may not be absolutely summable.

2.4 Tests Based on Signs, Ranks, and Spacings

There is a wide variety of tests based on statistics which involve signs, ranks, or spacings. Such tests essentially compare the behaviour of observations that lie below the centre of symmetry with that of observations that lie above.

The sign test of Gastwirth (1971) rejects for large absolute values of the statistic

$$\mathcal{S}_{12} = n^{-1/2} \sum_{t=1}^{n} \{\mathbb{I}(X_t - \bar{X} \le 0) - \tfrac{1}{2}\}, \tag{13}$$

where $\mathbb{I}(A)$ stands for an indicator that equals 1 when condition A is true and 0 otherwise.

The weighted sign test considered in Antille and Kersting (1977) and Antille et al. (1982) rejects for large absolute values of the statistic

$$\mathcal{S}_{13} = n^{-1/2} \sum_{t=1}^{\lfloor (n-1)/2 \rfloor} J(t/n)\{\mathbb{I}(V_t - V_{n-t} \le 0) - \tfrac{1}{2}\}, \tag{14}$$

where $V_t = X_{(t+1)} - X_{(t)}$, $1 \le t \le n-1$, are the spacings of the order statistics of $\mathcal{X}_n$ and $J(x) = \mathbb{I}(0.05 \le x \le 0.5)$. The modification of the test investigated by Ekström and Jammalamadaka (2007) rejects for large absolute values of the statistic

$$\mathcal{S}_{14} = n^{-1/2} \sum_{t=1}^{\lfloor (n-1)/2 \rfloor} \{\mathbb{I}(V_t - V_{n-t} \le 0) - \tfrac{1}{2}\}. \tag{15}$$

Another test based on spacings is the variant of the test of Finch (1977) considered in Antille et al. (1982). It rejects for large absolute values of the statistic

$$\mathcal{S}_{15} = n^{-1/2} \sum_{t=1}^{\lfloor (n-1)/2 \rfloor} J\left(\frac{t}{n+1}\right)\left(\frac{V_t - V_{n-t}}{V_t + V_{n-t}}\right), \tag{16}$$

with $J(x) = \mathbb{I}(0.05 \le x \le 0.5)$.

The Wilcoxon-type test of Antille and Kersting (1977) rejects for large absolute values of the statistic

$$\mathcal{S}_{16} = n^{-3/2} \sum\sum_{1 \le t < s \le n} \{\mathbb{I}(X_t + X_s - 2\ddot{X} \le 0) - \tfrac{1}{2}\}. \tag{17}$$

The trimmed Wilcoxon signed-rank test of Antille et al. (1982) rejects for large absolute values of the statistic

$$\mathcal{S}_{17} = n^{-1/2} \sum_{t=1}^{n} G_a\left(\frac{R_t}{2n+2}\right) \operatorname{sgn}(X_t - \ddot{X}), \tag{18}$$

where $G_a(x) = \min\{x, \frac{1}{2} - a\}$ for $0 \leq x \leq \frac{1}{2}$ and $0 \leq a \leq \frac{1}{2}$, $R_t = \sum_{s=1}^{n} \mathbb{I}(|X_s - \ddot{X}| \leq |X_t - \ddot{X}|)$ is the rank of $|X_t - \ddot{X}|$, and $\text{sgn}(x) = \mathbb{I}(x > 0) - \mathbb{I}(x < 0)$ is the signum function. We set $a = 0$ in our implementation of the test (for this values of a, $\mathcal{S}_{17}$ is equal to a statistic originally investigated by Gupta (1967)).

For some fixed $0 < d < 1$, the modified Wilcoxon test of Bhattacharya et al. (1982) focuses on the upper and lower $\lfloor d/2 \rfloor$ fraction of the data, and rejects for large absolute values of the statistic

$$\mathcal{S}_{18} = N^{-2} \sum_{t=1}^{N} \sum_{s=1}^{N} \tfrac{1}{2} \left\{ 1 - \text{sgn}(X_{(N+1)} - X_{(N+1-t)} - X_{(n-N+s)} + X_{(n-N)}) \right\}, \qquad (19)$$

where $N = \lfloor nd/2 \rfloor$. We set $d = 1/4$ in our implementation of the test.

Finally, the triples test of Davis and Quade (1978) and Randles et al. (1980) rejects for large absolute values of the statistic

$$\mathcal{S}_{19} = \sqrt{n} \binom{n}{3}^{-1} \sum\sum\sum_{1 \leq t < s < r \leq n} g(X_t, X_s, X_r), \qquad (20)$$

where

$$g(X_t, X_s, X_r) = \tfrac{1}{3} \left[\text{sgn}(X_t + X_s - 2X_r) + \text{sgn}(X_t + X_r - 2X_s) + \text{sgn}(X_s + X_r - 2X_t) \right].$$

The asymptotic null distributions of these statistics (under i.i.d. conditions) can be found in the references previously cited. The statistics (13)–(20) are all asymptotically normal with zero mean under symmetry, but their asymptotic variances, except that of $\mathcal{S}_{19}$, depend on the unknown distribution of X_t; the asymptotic variance of $\mathcal{S}_{19}$ is given by the asymptotic variance of the third-order U-statistic based on the kernel g.

3 Bootstrap Approximations

As indicated already, testing for symmetry about an unknown centre is complicated by the fact that, even when $\mathcal{X}_n$ is a sample of i.i.d. observations, the asymptotic null distributions of the test statistics (2)–(20) depend on the unknown distribution of X_t or on other unknown quantities. Correlation among observations complicates matters further still because the asymptotic variances of the test statistics generally depend on the correlation structure of $\mathcal{X}$ too. For test statistics which are asymptotically normal under the null, one could in principle overcome this problem, and obtain tests with asymptotically correct levels, through 'Studentization', that is, by dividing each test statistic by a consistent estimator (that does not depend on the distribution of X_t) of its asymptotic standard deviation. However, for the majority of the test statistics under consideration here, the variance of their asymptotic null distribution is currently available only under i.i.d. conditions. As a result, Studentization with consistent estimators of these asymptotic variances will not produce asymptotically correct tests if there are deviations from the i.i.d. assumption. Even in cases where asymptotic variances are estimated by means of suitable autocorrelation-robust methods (e.g., Bai and Ng (2005), Psaradakis and Vávra (2015)) or data-resampling methods (e.g., Chen et al. (2000), Chen (2001)), conventional large-sample approximations to the null distributions of the relevant test statistics may not necessarily be accurate for the relatively small sample sizes that are relevant in many applications.

As a convenient way of overcoming these difficulties, we propose to use suitable bootstrap procedures to approximate the sampling distributions of the test statistics of interest under the null hypothesis, and thus obtain P-values and/or critical values for the associated symmetry tests. More specifically, letting $S = S(\mathcal{X}_n)$ be a statistic for testing the symmetry hypothesis (1), a bootstrap approximation to its null distribution is provided by the conditional distribution, given $\mathcal{X}_n$, of the bootstrap analogue $S^* = S(\mathcal{X}_n^*)$ of S; here, $\mathcal{X}_n^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ are bootstrap pseudo-observations generated from an estimate of the distribution of $\mathcal{X}_n$ such that the conditional distribution of each X_t^*, given $\mathcal{X}_n$, is symmetric with centre $\bar{X}$. In practice, the bootstrap distribution of S is further approximated by numerical simulation, which amounts to constructing B replicates $\{S_1^*, S_2^*, \ldots, S_B^*\}$ of S^* from B independent sets of bootstrap pseudo-observations $\mathcal{X}_n^*$; the empirical distribution of $\{S_1^*, S_2^*, \ldots, S_B^*\}$ then serves as the (simulated) bootstrap approximation to the null sampling distribution of S. Consequently, the (simulated) bootstrap P-value of a test that rejects the null hypothesis (1) for large values of $|S|$ is computed as $P^*(\check{S}) = B^{-1} \sum_{i=1}^{B} \mathbb{I}(|S_i^*| > |\check{S}|)$, where $\check{S}$ is the observed value of S. The bootstrap test of nominal level $\alpha \in (0, 1)$ rejects symmetry if $P^*(\check{S}) \leq \alpha$ or, equivalently, if $|\check{S}|$ exceeds the $(\lceil B(1-\alpha) \rceil)$-th order statistic of $\{|S_1^*|, |S_2^*|, \ldots, |S_B^*|\}$. It is worth stressing that the requirement that the bootstrap pseudo-data $\mathcal{X}_n^*$ reflect the symmetry hypothesis under test, even though $\mathcal{X}$ may not satisfy (1), is essential for ensuring that the bootstrap test has reasonable power against departures from the null hypothesis (see, e.g., Lehmann and Romano (2005, Sec. 15.6)). In the sequel, we consider two different resampling schemes to generate bootstrap pseudo-observations $\mathcal{X}_n^*$, namely symmetrized versions of the schemes associated with the ARSB and STB procedures.

The typical assumption underlying the ARSB is that $\mathcal{X}$ admits the representation

$$X_t - \mu = \sum_{j=1}^{\infty} \phi_j (X_{t-j} - \mu) + \varepsilon_t, \tag{21}$$

where $\{\phi_j\}_{j=1}^{\infty}$ is an absolutely summable sequence of real numbers and $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ are i.i.d., real-valued, zero-mean random variables with finite, positive variance. The idea is to approximate (21) by a finite-order autoregressive model, the order of which increases simultaneously with the sample size at an appropriate rate, and use this model as the basis of a semi-parametric, residual-based bootstrap scheme (see, e.g., Kreiss (1992), Bühlmann (1997), Choi and Hall (2000), Kreiss et al. (2011)). More specifically, for some positive integer h (chosen as a function of n so that h increases with n but at a slower rate), let $(\hat{\phi}_{h1}, \ldots, \hat{\phi}_{hh})$ be the h-th order least-squares estimates of the autoregressive coefficients for $\mathcal{X}$, obtained by minimizing

$$(n-h)^{-1} \sum_{t=h+1}^{n} \left\{ (X_t - \bar{X}) - \sum_{j=1}^{h} \phi_{hj} (X_{t-j} - \bar{X}) \right\}^2, \tag{22}$$

and put

$$\hat{\varepsilon}_t = X_t - \bar{X} - \sum_{j=1}^{h} \hat{\phi}_{hj} (X_{t-j} - \bar{X}), \qquad t = h+1, \ldots, n.$$

Then, given some initial values $(X_{-h+1}^*, \ldots, X_0^*)$, bootstrap pseudo-observations $\mathcal{X}_n^*$ are obtained via the recursion

$$X_t^* - \bar{X} = \sum_{j=1}^{h} \hat{\phi}_{hj} (X_{t-j}^* - \bar{X}) + \varepsilon_t^*, \qquad t = 1, 2, \ldots, \tag{23}$$

where $\{\varepsilon_t^*\}$ are conditionally i.i.d. random variables, given $\mathcal{X}_n$, the distribution of which is the symmetrized empirical distribution of the centred residuals $\tilde{\varepsilon}_t = \hat{\varepsilon}_t - (n-h)^{-1} \sum_{t=h+1}^{n} \hat{\varepsilon}_t$, that is, the discrete distribution assigning mass $[2(n-h)]^{-1}$ to each $\pm\tilde{\varepsilon}_t$ $(t = h+1, \ldots, n)$.

We note that, although least-squares estimates $(\hat{\phi}_{h1}, \ldots, \hat{\phi}_{hh})$ of the parameters of the approximating autoregression are used in (23) to construct X_t^*, asymptotically equivalent estimates, such as those obtained from the empirical Yule–Walker equations, may alternatively be used. The Yule–Walker estimator is theoretically attractive because its use guarantees that the bootstrap pseudo-observations $\mathcal{X}_n^*$ are generated from a causal autoregressive process, but it is known to be significantly biased in small samples compared to the least-squares estimator (see, e.g., Tjøstheim and Paulsen (1983), Paulsen and Tjøstheim (1985)). Also note that, following Bühlmann (1997), $\mathcal{X}_n^*$ are obtained here by setting $X_t^* = \bar{X}$ for $t \leq 0$, generating $n + 100$ bootstrap replicates according to (23), and then discarding the first 100 replicates to minimize the effect of initial values. Finally, the order h of the autoregressive sieve is selected by means of Akaike's information criterion so as to minimize $\log \hat{\omega}_h^2 + 2(n-h)^{-1}h$ over the range $1 \leq h \leq \lfloor 10 \log_{10} n \rfloor$, $\hat{\omega}_h^2$ being the minimum value of (22). Under mild regularity conditions, a data-dependent choice of h based on Akaike's criterion is asymptotically efficient (e.g., Lee and Karagrigoriou (2001), Poskitt (2007)).

To describe the symmetrized STB procedure, let $Y_t = X_t$ for $1 \leq t \leq n$, $Y_t = 2\bar{X} - X_{t-n}$ for $n + 1 \leq t \leq 2n$, and $Y_t = Y_{t-2n}$ for $t > 2n$. The STB scheme involves sampling randomly, with replacement, from a collection of blocks of random length consisting of consecutive observations from $\{Y_t\}_{t\geq 1}$ (cf. Politis and Romano (1994)). More specifically, for some $p \in (0, 1)$ (chosen as a function of n so that p decreases with n but at a slower rate), let $\{L_i\}_{i\geq 1}$ be conditionally i.i.d. (positive) random variables, given $\mathcal{X}_n$, having the geometric distribution with mean $1/p$. Further, with $\bar{\tau} = \min\{\tau \geq 1 :_{i=1}^{\tau} L_i \geq n\}$, let $U_1, \ldots, U_{\bar{\tau}}$ be i.i.d. random variables, independent of $\{L_i\}$ and $\mathcal{X}_n$, having the discrete uniform distribution on $\{1, 2, \ldots, 2n\}$. Then, bootstrap pseudo-observations $\mathcal{X}_n^*$ are obtained by arranging in a sequence the first n elements of $\mathcal{B}(U_1, L_1), \ldots, \mathcal{B}(U_{\bar{\tau}}, L_{\bar{\tau}})$, where $\mathcal{B}(t, l) = (Y_t, \ldots, Y_{t+l-1})$ denotes a data block with starting point $t \geq 1$ and length $l \geq 1$.

In our implementation of the STB, we set $p = \min\{|2\hat{\rho}/(1 - \hat{\rho}^2)|^{-2/3} n^{-1/3}, 0.9999\}$, where $\hat{\rho}$ is the lag-1 sample autocorrelation of $\mathcal{X}_n$ (cf. Carlstein (1986)). This provides a computationally convenient choice for the expected block length $1/p$, and is motivated by the observation that STB variance estimators have the same asymptotic accuracy as block-bootstrap estimators based on non-overlapping data blocks of fixed length (Nordman (2009)). More intricate, and potentially more accurate, ways of choosing the optimal (expected) block length empirically are discussed in Lahiri (2003, Ch. 7); we do not use them here because of their high computational cost in the context of simulation experiments.

We conclude this section by noting that the linear structure assumed in (21) may arguably be considered as somewhat restrictive. However, the results of Bickel and Bühlmann

(1997) indicate that linearity may not be too onerous a requirement, in the sense that the closure (with respect to certain metrics) of the class of causal linear processes is quite large; roughly speaking, for any strictly stationary nonlinear process, there exists another process in the closure of causal linear processes having identical sample paths with probability exceeding 0.36. This also suggests that the ARSB is likely to yield reasonably good approximations within a class of processes larger than that associated with (21). In fact, Kreiss et al. (2011) have demonstrated that the ARSB is asymptotically valid for a general class of statistics associated with strictly stationary, weakly dependent, regular processes having spectral densities that are bounded away from zero and infinity. Such processes can always be represented in the form (21), with $\{\varepsilon_t\}$ being a strictly stationary sequence of uncorrelated (although not necessarily independent) random variables. Then, the autoregressive coefficients $(\phi_{h1}, \ldots, \phi_{hh})$ in (22) may also be thought of as the coefficients of the optimal (in a mean-square sense) linear predictor of $X_t - \mu$ based on the finite past $\{X_{t-1} - \mu, \ldots, X_{t-h} - \mu\}$. The finite-predictor coefficients are uniquely determined for each fixed integer $h \geq 1$ as long as $\sigma^2 > 0$ and $\gamma_m \to 0$ as $|m| \to \infty$, and converge to the corresponding infinite-predictor coefficients as h tends to infinity (see, e.g, Pourahmadi (2001, Sec. 7.6), Kreiss et al. (2011)).

The STB is known to provide asymptotically valid approximations under general conditions that allow for weak dependence, such as strong mixing (Politis and Romano (1994), Lahiri (2003)), near-epoch dependence (Gonçalves and White (2002), Gonçalves and de Jong (2003)), and ψ-weak dependence (Hwang and Shin (2012)).

4 Simulation Study

In this section, we report and discuss the results of a simulation study on the finite-sample properties of bootstrap-assisted tests for symmetry under various data-generating mechanisms.

4.1 Experimental Design

In the first set of experiments, we examine the performance of symmetry tests under linear dependence by considering artificial data generated according to the models:

M1: $X_t = 0.8X_{t-1} + \varepsilon_t$,

M2: $X_t = 0.6X_{t-1} - 0.5X_{t-2} + \varepsilon_t$,

M3: $X_t = 0.6X_{t-1} + 0.3\varepsilon_{t-1} + \varepsilon_t$.

Throughout this section, $\{\varepsilon_t\}$ are i.i.d. random variables the common distribution of which is either standard normal (labelled N in the various tables) or generalized lambda with inverse distribution function $F^{-1}(u) = \lambda_1 + (1/\lambda_2)\{u^{\lambda_3} - (1-u)^{\lambda_4}\}$, $0 < u < 1$ (standardized to have zero mean and unit variance). The values of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ used in the experiments are taken from Bai and Ng (2005) and can be found in Table 1; distributions S1–S3 are symmetric (but leptokurtic), whereas A1–A4 are asymmetric.

In a second set of experiments, we assess the performance of tests under nonlinear dependence by using artificial data from the models:

M4: $X_t = 0.9X_{t-1}\mathbb{I}(|X_{t-1}| \leq 1) - 0.3X_{t-1}\mathbb{I}(|X_{t-1}| > 1) + \varepsilon_t$,

M5: $X_t = \eta_t\varepsilon_t, \quad \eta_t^2 = 0.05 + (0.1\varepsilon_{t-1}^2 + 0.85)\eta_{t-1}^2$,

M6: $X_t = 0.7X_{t-2}\varepsilon_{t-1} + \varepsilon_t$.

Model M4 is a threshold autoregressive model, M5 is a generalized autoregressive conditionally heteroskedastic model, and M6 is a bilinear model. In all three cases, the third cumulant of X_t is zero if ε_t is symmetric (cf. Tong (1990, pp. 166–167), Martins (1999)), and $\{X_t\}$ does not admit the representation (21) with respect to an i.i.d. noise sequence.

For each design point, 1,000 independent realizations of $\{X_t\}$ of length $100 + n$, with $n \in \{150, 300\}$, are generated. (Results for $n = 500$ are not reported here in order to save space, but are available from the authors upon request.) The first 100 data points of each realization are then discarded in order to eliminate start-up effects and the remaining n data points are used to compute the values of the test statistics (2)–(20). The number of bootstrap replications is set to $B = 199$. We note that using a larger number of bootstrap replications does not change the results substantially. Hall (1986) and Jöckel (1986) provide theoretical explanations of the ability of simulation-based inference procedures to yield good results for relatively small values of the simulation size.

Table 1: Noise Distributions

	λ_1	λ_2	λ_3	λ_4	Skewness	Kurtosis
N	–	–	–	–	0.0	3.0
S1	0.000000	-1.000000	-0.080000	-0.080000	0.0	6.0
S2	0.000000	-0.397912	-0.160000	-0.160000	0.0	11.6
S3	0.000000	-1.000000	-0.240000	-0.240000	0.0	126.0
A1	0.000000	-1.000000	-0.007500	-0.030000	1.5	7.5
A2	0.000000	-1.000000	-0.100900	-0.180200	2.0	21.1
A3	0.000000	-1.000000	-0.001000	-0.130000	3.2	23.8
A4	0.000000	-1.000000	-0.000100	-0.170000	3.8	40.7

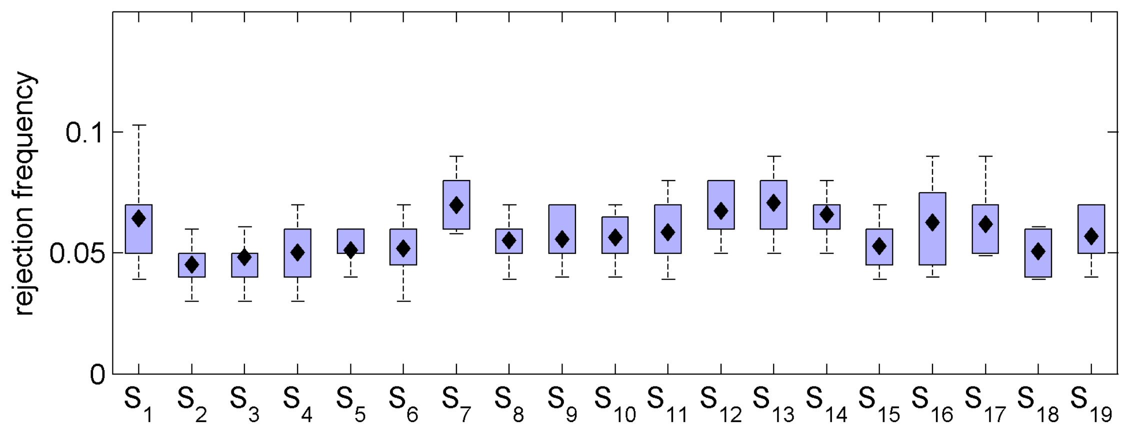
4.2 Simulation Results

Results over all 24 design points which satisfy the null hypothesis of symmetry are summarized graphically in Figures 1 and 2. These show boxplots of the Monte Carlo rejection frequencies of ARSB-assisted and STB-assisted tests of nominal level $\alpha = 0.05$. The top and bottom of each coloured box represent the 25th and 75th percentiles, respectively, of the empirical rejection frequencies, the black diamond inside the box indicates the mean value, and the whiskers indicate the 10th and 90th percentiles. Detailed results for individual design points can be found in Tables 2–20 in the Appendix.

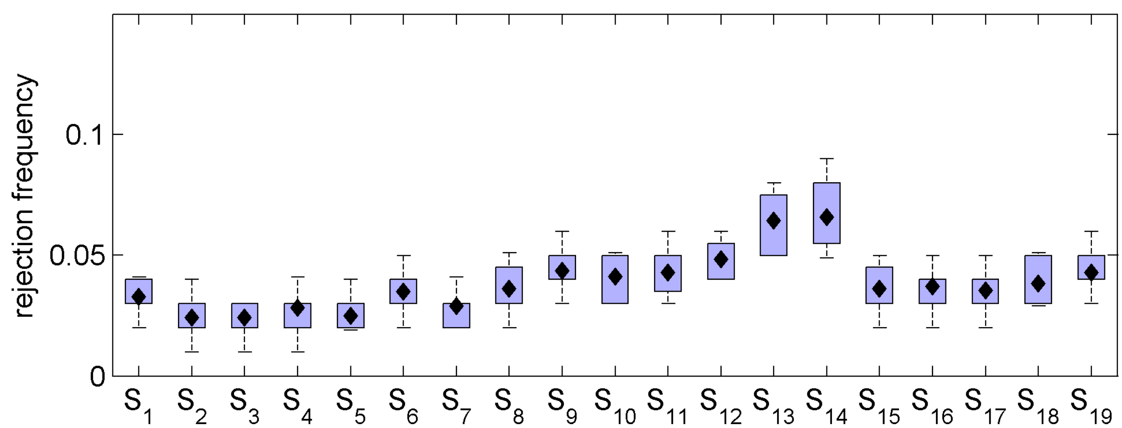
Although both bootstrap procedures are generally successful at controlling the discrepancy between the exact and nominal probabilities of Type I error, the ARSB seems to have an advantage, especially for the smaller of the two sample sizes considered. Even in those cases, however, where some level distortion is observed, it is not of a magnitude that makes the tests unattractive for application. It is possible that the level properties of STB-assisted tests can be improved further by using a more sophisticated data-driven method to select the (expected) block length.

Boxplots of the empirical rejection frequencies of tests of nominal level $\alpha = 0.05$ over the 24 design points which do not satisfy the null hypothesis of symmetry are shown in Figures 3 and 4. Results for individual design points are reported in Tables 2–20 in the

Figure 1: Empirical Rejection Frequencies of Symmetry Tests: $n = 150$

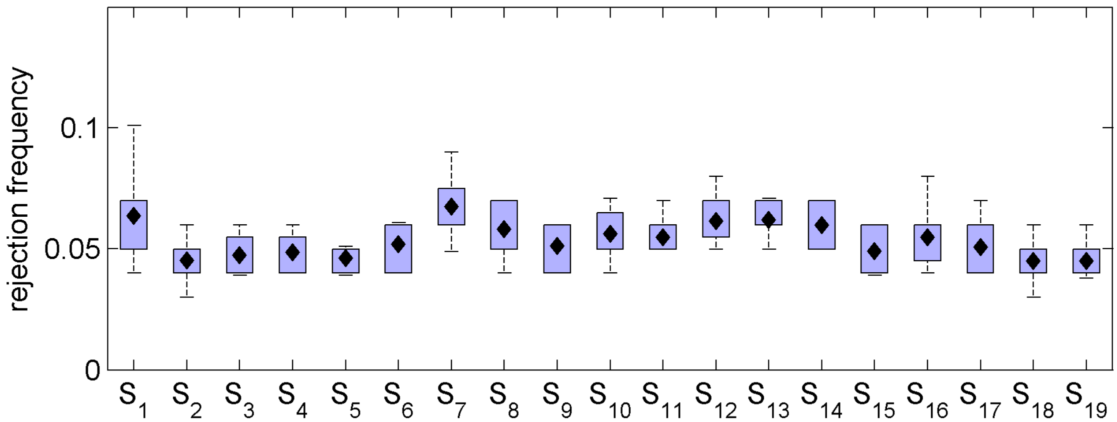


(a) Autoregressive sieve bootstrap

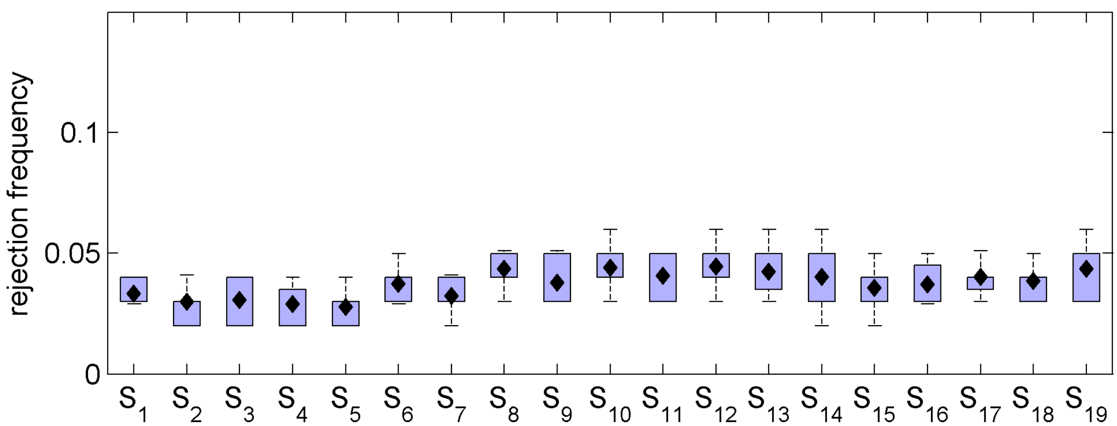


(b) Stationary bootstrap

Figure 2: Empirical Rejection Frequencies of Symmetry Tests: $n = 300$

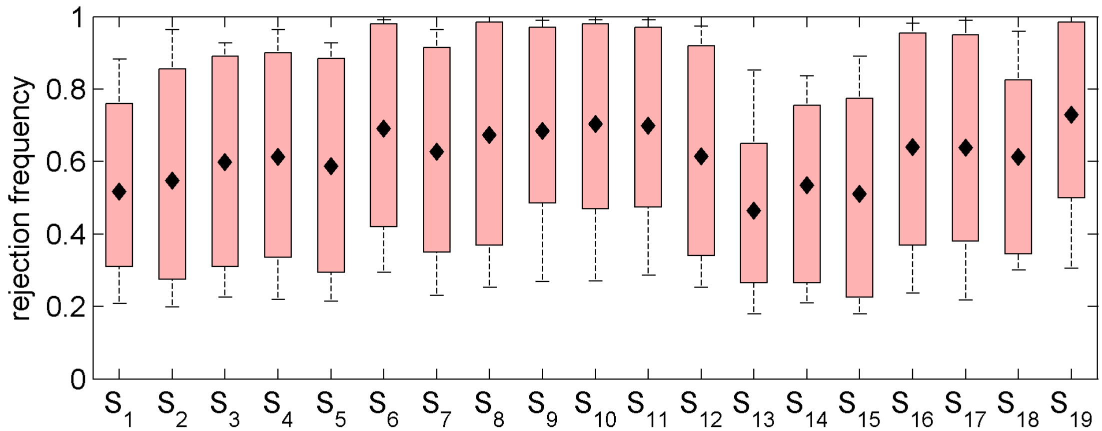


(a) Autoregressive sieve bootstrap

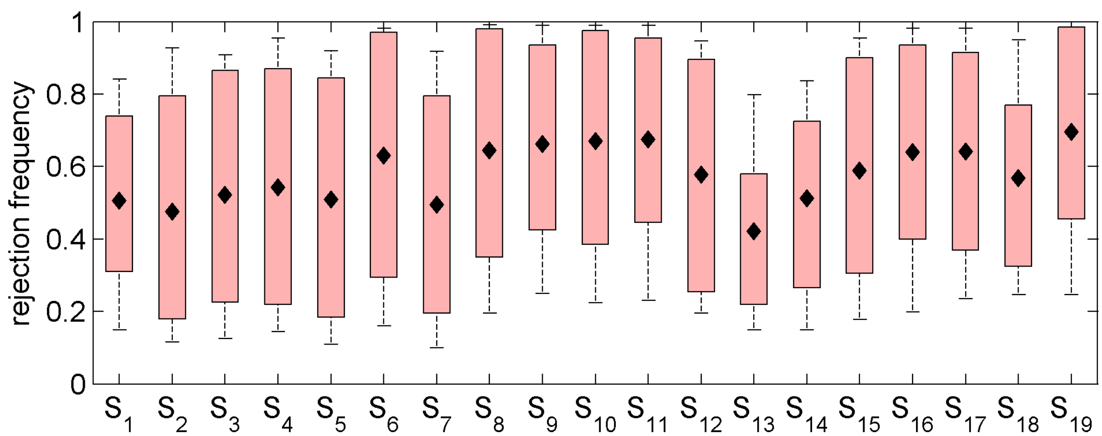


(b) Stationary bootstrap

Figure 3: Empirical Rejection Frequencies of Symmetry Tests: $n = 150$

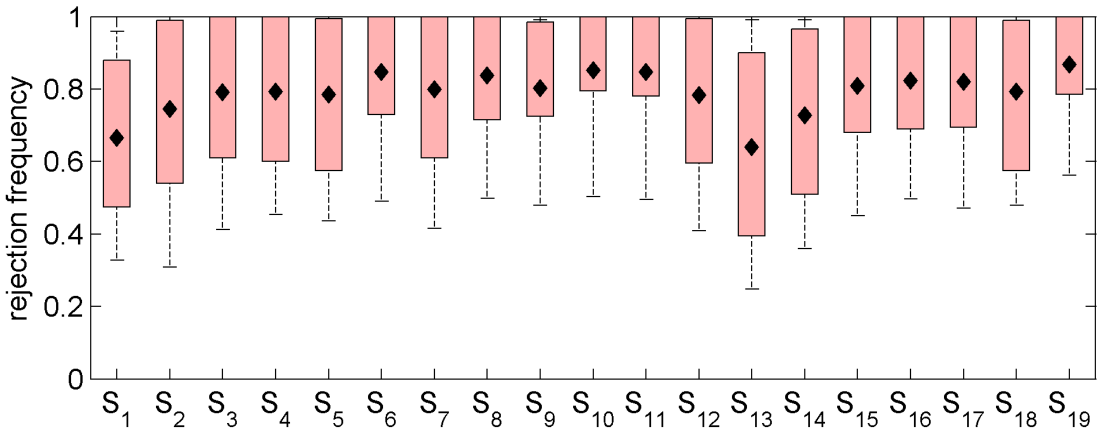


(a) Autoregressive sieve bootstrap

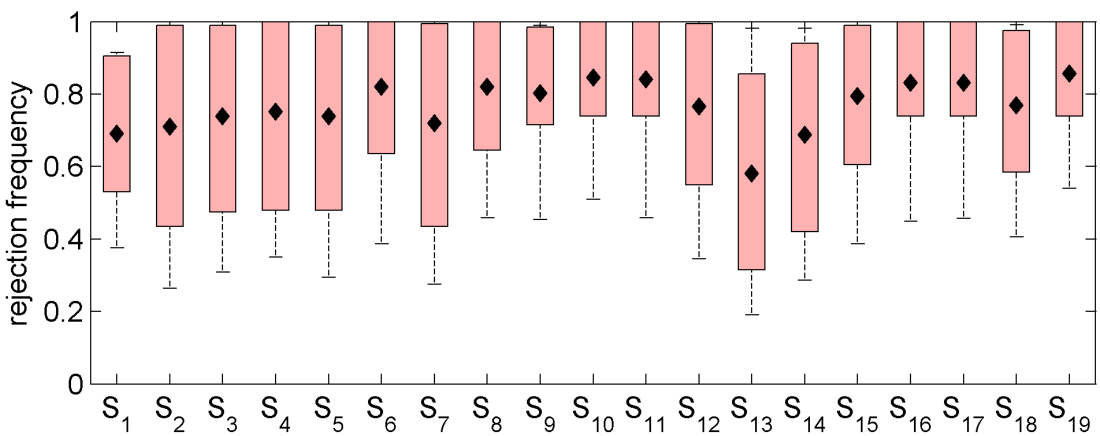


(b) Stationary bootstrap

Figure 4: Empirical Rejection Frequencies of Symmetry Tests: $n = 300$



(a) Autoregressive sieve bootstrap



(b) Stationary bootstrap

Appendix. Regardless of which of the two bootstrap procedures is used, the triples test based on $\mathcal{S}_{19}$ tends to be the top performer in terms of average empirical power (indicated by black diamonds), albeit only marginally in some cases. Among the other tests, $\mathcal{S}_{10}$ and $\mathcal{S}_{11}$ are also competitive, and outperform the classical skewness test $\mathcal{S}_1$, which is also based on a linear statistic involving an odd function of standardized data (this is consistent with findings reported in Psaradakis (2016)). Tests based on $\mathcal{S}_6$, $\mathcal{S}_8$ and $\mathcal{S}_9$ are almost as powerful overall as those based on $\mathcal{S}_{10}$ and $\mathcal{S}_{11}$. Among the nonparametric tests, the Wilcoxon-type tests based on $\mathcal{S}_{16}$ and $\mathcal{S}_{17}$ tend to have the highest rejection frequencies. Rather unsurprisingly, the rejection frequencies of all tests improve with increasing skewness (and leptokurtosis) in the noise distribution, as well as with an increasing sample size.

Finally, inspection of the results obtained under models M4–M6 reveals that deviations from the linearity assumption which underlines the ARSB procedure do not have an adverse effect on the properties of bootstrap tests. ARSB-assisted tests generally work well even for data that are generated by processes which are not representable as in (21). It can also be seen in Tables 2–20 that test rejection frequencies are higher for data with asymmetric marginal distributions generated according to the nonlinear models.

5 Summary

This paper has considered the problem of testing for symmetry (around an unspecified centre) of the one-dimensional marginal distribution of a strictly stationary and weakly dependent stochastic process. We have examined the properties of nineteen different tests, most of which have been proposed for i.i.d. observations. Since conventional large-sample approximations to the null distributions of many of the test statistics are either unavailable under dependence or involve unknown quantities (including the marginal distribution of the data), we have explored how the ARSB and the STB procedures may be used to obtain P-values and/or critical values for the tests. Such bootstrap-assisted tests are straightforward to implement and require no prior estimation of asymptotic variances. Our simulation study has revealed that the ARSB-assisted version of the well-known triples test provides the best overall performance in terms of level accuracy and power.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

Antille, A. and G. Kersting (1977). Tests for symmetry. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 39*, 235–255.

Antille, A., G. Kersting, and W. Zucchini (1982). Testing symmetry. *Journal of the American Statistical Association 77*, 639–646.

Arcones, M. A. and E. Giné (1991). Some bootstrap tests of symmetry for univariate continuous distributions. *Annals of Statistics 19*, 1496–1511.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business and Economic Statistics 23*, 49–60.

Bhattacharya, P. K., J. L. Gastwirth, and A. L. Wright (1982). Two modified Wilcoxon tests for symmetry about an unknown location parameter. *Biometrika 69*, 377–382.

Bickel, P. J. and P. Bühlmann (1997). Closure of linear processes. *Journal of Theoretical Probability 10*, 445–479.

Boos, D. D. (1982). A test for asymmetry associated with the Hodges–Lehmann estimator. *Journal of the American Statistical Association 77*, 647–651.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli 3*, 123–148.

Cabilio, P. and J. Masaro (1996). A simple test of symmetry about an unknown median. *Canadian Journal of Statistics 24*, 349–361.

Carlstein, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics 14*, 1171–1179.

Chen, Y.-T. (2001). Testing conditional symmetry with an application to stock returns. Institute for Social Sciences and Philosophy, Academia Sinica, Taiwan.

Chen, Y.-T., R. Y. Chou, and C.-M. Kuan (2000). Testing time reversibility without moment restrictions. *Journal of Econometrics 95*, 199–218.

Choi, E. and P. Hall (2000). Bootstrap confidence regions constructed from autoregressions of arbitrary order. *Journal of the Royal Statstical Society, Ser. B 62*, 461–477.

Davis, C. E. and D. Quade (1978). U-statistics for skewness or symmetry. *Communications in Statistics – Theory and Methods 7*, 413–418.

Ekström, M. and S. R. Jammalamadaka (2007). An asymptotically distribution-free test of symmetry. *Journal of Statistical Planning and Inference 137*, 799–810.

Ekström, M. and S. R. Jammalamadaka (2012). A general measure of skewness. *Statistics and Probability Letters 82*, 1559–1568.

Finch, S. J. (1977). Robust univariate test of symmetry. *Journal of the American Statistical Association 72*, 387–392.

Gastwirth, J. L. (1971). On the sign test for symmetry. *Journal of the American Statistical Association 66*, 821–823.

Gonçalves, S. and R. de Jong (2003). Consistency of the stationary bootstrap under weak moment conditions. *Economics Letters 81*, 273–278.

Gonçalves, S. and H. White (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econometric Theory 18*, 1367–1384.

Gupta, M. K. (1967). An asymptotically nonparametric test of symmetry. *Annals of Mathematical Statistics 38*, 849–866.

Hall, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics 14*, 1453–1462.

Henze, N., B. Klar, and S. G. Meintanis (2003). Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. *Journal of Multivariate Analysis 87*, 275–297.

Holgersson, H. E. T. (2007). Robust testing for skewness. *Communications in Statistics – Theory and Methods 36*, 485–498.

Hwang, E. and D. W. Shin (2012). Strong consistency of the stationary bootstrap under ψ-weak dependence. *Statistics and Probability Letters 82*, 488–495.

Jöckel, K.-H. (1986). Finite-sample properties and asymptotic efficiency of Monte Carlo tests. *Annals of Statistics 14*, 336–347.

Kreiss, J.-P. (1992). Bootstrap procedures for AR(∞) processes. In K.-H. Jöckel, G. Rothe, and W. Sendler (Eds.), *Bootstrapping and Related Techniques*, pp. 107–113. Heidelberg: Springer-Verlag.

Kreiss, J.-P., E. Paparoditis, and D. N. Politis (2011). On the range of validity of the autoregressive sieve bootstrap. *Annals of Statistics 39*, 2103–2130.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. New York: Springer-Verlag.

Lee, S. and A. Karagrigoriou (2001). An asymptotically optimal selection of the order of a linear process. *Sankhyā, Ser. A 63*, 93–106.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses*. New York: 3rd Edition, Springer.

Lyubchich, V., X. Wang, A. Heyes, and Y. R. Gel (2016). A distribution-free m-out-of-n bootstrap approach to testing symmetry about an unknown median. *Computational Statistics and Data Analysis 104*, 1–9.

Maasoumi, E. and J. S. Racine (2009). A robust entropy-based test of asymmetry for discrete and continuous processes. *Econometric Reviews 28*, 246–261.

Martins, C. M. (1999). A note on the third-order moment structure of a bilinear model with non-independent shocks. *Portugaliae Mathematica 56*, 115–125.

Miao, W., Y. R. Gel, and J. L. Gastwirth (2006). A new test of symmetry about an unknown median. In A. Hsiung, C.-H. Zhang, and Z. Ying (Eds.), *Random Walk, Sequential Analysis and Related Topics – A Festschrift in Honor of Yuan-Shih Chow*, pp. 199–214. World Scientific Publishing, Singapore.

Mira, A. (1999). Distribution-free test for symmetry based on Bonferroni's measure. *Journal of Applied Statistics 26*, 959–972.

Nordman, D. J. (2009). A note on the stationary bootstrap's variance. *Annals of Statistics 37*, 359–370.

Paulsen, J. and D. Tjøstheim (1985). On the estimation of residual variance and order in autoregressive time series. *Journal of the Royal Statistical Society, Ser. B 47*, 216–228.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*, 1303–1313.

Poskitt, D. S. (2007). Autoregressive approximation in nonstandard situations: the fractionally integrated and non-invertible cases. *Annals of the Institute of Statistical Mathematics 59*, 697–725.

Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory.* New York: Wiley.

Premaratne, G. and A. Bera (2005). A test for symmetry with leptokurtic financial data. *Journal of Financial Econometrics 3*, 169–187.

Psaradakis, Z. (2003). A bootstrap test for symmetry of dependent data based on a Kolmogorov–Smirnov type statistic. *Communications in Statistics – Simulation and Computation 32*, 113–126.

Psaradakis, Z. (2016). Using the bootstrap to test for symmetry under unknown dependence. *Journal of Business and Economic Statistics 34*, 406–415.

Psaradakis, Z. and M. Vávra (2015). A quantile-based test for symmetry of weakly dependent processes. *Journal of Time Series Analysis 36*, 587–598.

Racine, J. S. and E. Maasoumi (2007). A versatile and robust metric entropy test of time-reversibility, and other hypotheses. *Journal of Econometrics 138*, 547–567.

Randles, R. H., M. A. Fligner, G. E. Policello, and D. A. Wolfe (1980). An asymptotically distribution-free test for symmetry versus asymmetry. *Journal of the American Statistical Association 75*, 168–172.

Schuster, E. F. and R. C. Barker (1987). Using the bootstrap in testing symmetry versus asymmetry. *Communications in Statistics – Simulation and Computation 16*, 69–84.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Tjøstheim, D. and J. Paulsen (1983). Bias of some commonly-used time series estimates. *Biometrika 70*, 389–399; Corrigendum (1984), 71, 656.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

Zheng, T. and J. L. Gastwirth (2010). On bootstrap tests of symmetry about an unknown median. *Journal of Data Science 8*, 413–427.

A Appendix: Tables

Table 2: Empirical Rejection Frequencies of $\mathcal{S}_1$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.06	0.05	0.05	0.05	0.21	0.11	0.26	0.30	0.03	0.02	0.03	0.02	0.17	0.10	0.41	0.41
	M2	0.05	0.06	0.07	0.06	0.58	0.32	0.69	0.71	0.03	0.03	0.04	0.02	0.62	0.28	0.75	0.70
	M3	0.04	0.07	0.07	0.06	0.34	0.19	0.49	0.47	0.05	0.04	0.03	0.03	0.34	0.15	0.63	0.69
	M4	0.03	0.05	0.04	0.03	0.84	0.34	0.81	0.81	0.04	0.03	0.03	0.04	0.77	0.33	0.79	0.73
	M5	0.05	0.05	0.06	0.09	0.90	0.46	0.92	0.88	0.04	0.06	0.04	0.03	0.86	0.37	0.85	0.84
	M6	0.10	0.13	0.13	0.10	0.38	0.23	0.58	0.62	0.03	0.02	0.03	0.03	0.29	0.14	0.49	0.46
$n = 300$	M1	0.08	0.04	0.04	0.07	0.34	0.21	0.48	0.48	0.04	0.04	0.04	0.03	0.42	0.25	0.77	0.77
	M2	0.06	0.06	0.07	0.05	0.86	0.45	0.92	0.82	0.03	0.04	0.03	0.04	0.91	0.52	0.91	0.84
	M3	0.06	0.07	0.06	0.06	0.64	0.31	0.78	0.73	0.04	0.04	0.03	0.03	0.73	0.39	0.90	0.85
	M4	0.05	0.06	0.04	0.03	0.96	0.58	0.89	0.87	0.04	0.04	0.03	0.04	0.95	0.55	0.91	0.87
	M5	0.04	0.06	0.06	0.06	0.96	0.64	0.96	0.92	0.02	0.03	0.03	0.02	0.95	0.61	0.91	0.86
	M6	0.08	0.10	0.12	0.11	0.47	0.33	0.67	0.70	0.03	0.03	0.03	0.03	0.41	0.23	0.54	0.56

Table 3: Empirical Rejection Frequencies of $\mathcal{S}_2$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.06	0.04	0.05	0.06	0.18	0.11	0.31	0.35	0.02	0.02	0.03	0.04	0.08	0.07	0.24	0.26
	M2	0.05	0.04	0.04	0.03	0.38	0.22	0.85	0.90	0.03	0.01	0.02	0.02	0.27	0.12	0.76	0.83
	M3	0.05	0.04	0.05	0.05	0.32	0.20	0.62	0.67	0.02	0.02	0.02	0.02	0.16	0.12	0.46	0.52
	M4	0.05	0.04	0.05	0.04	0.43	0.22	0.74	0.71	0.01	0.03	0.02	0.03	0.36	0.14	0.72	0.70
	M5	0.03	0.04	0.03	0.06	0.86	0.48	1.00	1.00	0.02	0.03	0.03	0.01	0.84	0.38	1.00	1.00
	M6	0.04	0.04	0.05	0.06	0.45	0.24	0.96	0.96	0.03	0.02	0.04	0.04	0.39	0.20	0.90	0.92
$n = 300$	M1	0.04	0.05	0.05	0.06	0.30	0.20	0.60	0.66	0.02	0.05	0.02	0.04	0.21	0.18	0.59	0.63
	M2	0.03	0.06	0.06	0.05	0.73	0.43	1.00	0.99	0.02	0.02	0.03	0.02	0.68	0.31	0.99	1.00
	M3	0.05	0.03	0.05	0.03	0.61	0.38	0.94	0.97	0.02	0.02	0.03	0.04	0.43	0.27	0.91	0.94
	M4	0.04	0.03	0.05	0.07	0.75	0.31	0.98	0.98	0.03	0.03	0.03	0.04	0.76	0.27	0.98	0.97
	M5	0.02	0.05	0.04	0.04	0.99	0.78	1.00	1.00	0.03	0.03	0.03	0.03	0.99	0.73	1.00	1.00
	M6	0.04	0.04	0.05	0.06	0.81	0.48	1.00	1.00	0.03	0.05	0.03	0.03	0.78	0.44	1.00	1.00

Table 4: Empirical Rejection Frequencies of $\mathcal{S}_3$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.05	0.05	0.06	0.07	0.18	0.15	0.46	0.47	0.02	0.03	0.03	0.03	0.08	0.07	0.26	0.28
	M2	0.07	0.05	0.05	0.04	0.48	0.26	0.90	0.92	0.02	0.01	0.03	0.02	0.35	0.15	0.81	0.87
	M3	0.05	0.06	0.05	0.05	0.36	0.25	0.80	0.80	0.03	0.01	0.02	0.03	0.23	0.13	0.61	0.57
	M4	0.05	0.04	0.04	0.05	0.53	0.23	0.83	0.84	0.02	0.02	0.03	0.04	0.46	0.22	0.86	0.84
	M5	0.03	0.03	0.03	0.05	0.88	0.50	1.00	1.00	0.02	0.03	0.03	0.03	0.87	0.46	1.00	0.99
	M6	0.04	0.04	0.05	0.06	0.47	0.24	0.90	0.92	0.02	0.01	0.03	0.02	0.43	0.22	0.87	0.90
$n = 300$	M1	0.06	0.07	0.05	0.05	0.35	0.28	0.81	0.84	0.02	0.02	0.03	0.04	0.21	0.16	0.63	0.70
	M2	0.05	0.03	0.06	0.06	0.80	0.51	1.00	1.00	0.02	0.02	0.03	0.04	0.72	0.38	1.00	0.99
	M3	0.05	0.06	0.04	0.05	0.71	0.42	1.00	0.99	0.03	0.03	0.02	0.03	0.53	0.32	0.95	0.98
	M4	0.04	0.04	0.04	0.04	0.86	0.44	1.00	1.00	0.03	0.04	0.02	0.03	0.83	0.41	1.00	0.99
	M5	0.03	0.04	0.04	0.05	0.99	0.84	1.00	1.00	0.03	0.02	0.03	0.04	1.00	0.83	1.00	1.00
	M6	0.06	0.04	0.05	0.04	0.72	0.45	0.99	1.00	0.04	0.04	0.04	0.05	0.72	0.42	0.99	0.98

Table 5: Empirical Rejection Frequencies of $\mathcal{S}_4$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.07	0.05	0.05	0.07	0.23	0.15	0.41	0.47	0.02	0.03	0.05	0.05	0.09	0.10	0.29	0.35
	M2	0.05	0.03	0.04	0.03	0.52	0.29	0.90	0.94	0.02	0.01	0.01	0.02	0.37	0.18	0.85	0.87
	M3	0.05	0.06	0.06	0.08	0.38	0.21	0.80	0.87	0.03	0.01	0.03	0.03	0.20	0.15	0.62	0.69
	M4	0.05	0.05	0.04	0.04	0.54	0.22	0.88	0.83	0.04	0.02	0.03	0.03	0.47	0.19	0.87	0.86
	M5	0.03	0.04	0.06	0.05	0.90	0.55	1.00	1.00	0.03	0.03	0.04	0.03	0.87	0.48	1.00	1.00
	M6	0.04	0.06	0.06	0.05	0.48	0.26	0.93	0.96	0.02	0.03	0.04	0.03	0.44	0.24	0.89	0.95
$n = 300$	M1	0.05	0.06	0.05	0.05	0.39	0.27	0.77	0.77	0.03	0.02	0.03	0.03	0.26	0.17	0.69	0.70
	M2	0.04	0.04	0.05	0.06	0.80	0.51	1.00	1.00	0.03	0.02	0.02	0.02	0.74	0.44	0.99	1.00
	M3	0.06	0.05	0.04	0.04	0.69	0.47	0.98	0.98	0.02	0.02	0.04	0.03	0.50	0.36	0.95	0.99
	M4	0.03	0.05	0.06	0.06	0.87	0.46	1.00	0.99	0.03	0.02	0.04	0.02	0.83	0.42	1.00	1.00
	M5	0.05	0.04	0.04	0.06	1.00	0.85	1.00	1.00	0.03	0.04	0.02	0.04	1.00	0.81	1.00	1.00
	M6	0.04	0.05	0.05	0.05	0.77	0.47	0.99	1.00	0.03	0.03	0.04	0.05	0.75	0.46	1.00	1.00

Table 6: Empirical Rejection Frequencies of $\mathcal{S}_5$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.06	0.05	0.05	0.06	0.17	0.13	0.41	0.42	0.04	0.03	0.02	0.04	0.10	0.06	0.23	0.28
	M2	0.05	0.06	0.05	0.04	0.47	0.24	0.90	0.92	0.02	0.02	0.02	0.02	0.36	0.15	0.83	0.80
	M3	0.05	0.06	0.07	0.05	0.35	0.24	0.79	0.79	0.03	0.02	0.01	0.02	0.17	0.11	0.48	0.58
	M4	0.05	0.05	0.06	0.06	0.53	0.22	0.83	0.82	0.01	0.02	0.03	0.02	0.48	0.19	0.86	0.82
	M5	0.04	0.04	0.05	0.04	0.88	0.48	1.00	1.00	0.02	0.03	0.03	0.03	0.90	0.47	1.00	1.00
	M6	0.05	0.04	0.05	0.05	0.46	0.24	0.89	0.92	0.04	0.03	0.02	0.03	0.39	0.18	0.89	0.91
$n = 300$	M1	0.06	0.05	0.05	0.05	0.40	0.24	0.77	0.81	0.03	0.02	0.03	0.01	0.24	0.13	0.62	0.69
	M2	0.05	0.05	0.04	0.04	0.80	0.51	1.00	1.00	0.02	0.02	0.02	0.03	0.75	0.40	1.00	1.00
	M3	0.03	0.05	0.05	0.05	0.64	0.44	0.98	0.99	0.03	0.02	0.02	0.04	0.50	0.30	0.94	0.96
	M4	0.06	0.05	0.05	0.05	0.85	0.44	1.00	0.99	0.03	0.04	0.02	0.03	0.83	0.46	0.99	0.99
	M5	0.05	0.02	0.04	0.05	1.00	0.81	1.00	1.00	0.03	0.03	0.03	0.04	1.00	0.80	1.00	1.00
	M6	0.04	0.05	0.04	0.04	0.75	0.45	0.99	0.99	0.04	0.02	0.03	0.04	0.76	0.42	0.99	0.99

Table 7: Empirical Rejection Frequencies of $\mathcal{S}_6$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.07	0.06	0.07	0.06	0.24	0.13	0.53	0.58	0.02	0.03	0.03	0.04	0.16	0.10	0.39	0.47
	M2	0.05	0.06	0.06	0.04	0.69	0.31	0.98	0.99	0.03	0.03	0.03	0.02	0.55	0.25	0.97	0.96
	M3	0.05	0.06	0.06	0.05	0.49	0.30	0.93	0.94	0.02	0.03	0.04	0.04	0.31	0.16	0.75	0.82
	M4	0.05	0.06	0.07	0.06	0.76	0.35	0.98	0.98	0.05	0.04	0.04	0.03	0.71	0.28	0.98	0.97
	M5	0.04	0.02	0.05	0.05	0.97	0.57	1.00	1.00	0.05	0.05	0.03	0.04	0.95	0.55	1.00	1.00
	M6	0.03	0.04	0.03	0.06	0.59	0.32	0.98	0.98	0.04	0.04	0.03	0.04	0.58	0.27	0.97	0.97
$n = 300$	M1	0.07	0.06	0.06	0.04	0.41	0.31	0.87	0.91	0.04	0.04	0.03	0.04	0.35	0.25	0.80	0.86
	M2	0.06	0.05	0.04	0.05	0.94	0.66	1.00	1.00	0.04	0.03	0.02	0.04	0.90	0.55	1.00	1.00
	M3	0.06	0.04	0.06	0.07	0.79	0.50	1.00	1.00	0.04	0.04	0.02	0.04	0.70	0.39	1.00	1.00
	M4	0.04	0.05	0.06	0.06	0.99	0.67	1.00	1.00	0.03	0.04	0.03	0.05	0.99	0.57	1.00	1.00
	M5	0.04	0.05	0.04	0.06	1.00	0.85	1.00	1.00	0.05	0.04	0.04	0.04	1.00	0.87	1.00	1.00
	M6	0.05	0.04	0.05	0.05	0.90	0.55	1.00	1.00	0.03	0.05	0.04	0.04	0.90	0.55	1.00	1.00

Table 8: Empirical Rejection Frequencies of $\mathcal{S}_7$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.07	0.08	0.08	0.07	0.23	0.17	0.49	0.55	0.03	0.03	0.02	0.03	0.09	0.07	0.24	0.29
	M2	0.06	0.07	0.07	0.09	0.55	0.26	0.90	0.95	0.02	0.02	0.02	0.03	0.30	0.13	0.72	0.76
	M3	0.08	0.08	0.07	0.09	0.46	0.26	0.90	0.93	0.02	0.03	0.02	0.03	0.22	0.10	0.62	0.68
	M4	0.07	0.07	0.08	0.10	0.50	0.23	0.88	0.85	0.02	0.05	0.04	0.06	0.41	0.13	0.83	0.80
	M5	0.03	0.06	0.06	0.04	0.87	0.44	1.00	1.00	0.02	0.03	0.02	0.03	0.79	0.35	0.99	0.99
	M6	0.07	0.07	0.06	0.06	0.52	0.23	0.96	0.95	0.04	0.03	0.03	0.03	0.41	0.17	0.89	0.91
$n = 300$	M1	0.04	0.07	0.07	0.06	0.42	0.27	0.85	0.84	0.01	0.03	0.04	0.03	0.23	0.12	0.65	0.71
	M2	0.06	0.06	0.07	0.09	0.83	0.46	0.99	1.00	0.03	0.02	0.03	0.05	0.69	0.28	0.99	1.00
	M3	0.09	0.07	0.08	0.07	0.75	0.49	1.00	1.00	0.03	0.02	0.04	0.02	0.51	0.29	0.97	0.98
	M4	0.07	0.09	0.08	0.07	0.87	0.38	1.00	1.00	0.03	0.04	0.04	0.02	0.80	0.28	0.99	1.00
	M5	0.04	0.07	0.06	0.05	1.00	0.73	1.00	1.00	0.03	0.03	0.04	0.03	0.99	0.66	1.00	1.00
	M6	0.05	0.08	0.07	0.06	0.85	0.46	1.00	1.00	0.04	0.05	0.04	0.04	0.78	0.36	1.00	1.00

Table 9: Empirical Rejection Frequencies of $\mathcal{S}_8$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.05	0.05	0.05	0.05	0.19	0.15	0.35	0.42	0.02	0.03	0.03	0.04	0.15	0.11	0.37	0.42
	M2	0.04	0.05	0.05	0.06	0.63	0.36	0.98	0.99	0.03	0.03	0.03	0.03	0.58	0.27	0.96	0.98
	M3	0.04	0.06	0.06	0.07	0.45	0.26	0.84	0.85	0.02	0.03	0.03	0.04	0.37	0.20	0.80	0.83
	M4	0.03	0.06	0.05	0.05	0.73	0.37	1.00	0.98	0.03	0.05	0.04	0.05	0.72	0.33	0.99	0.98
	M5	0.03	0.06	0.07	0.06	0.97	0.63	1.00	1.00	0.02	0.05	0.06	0.03	0.97	0.59	1.00	1.00
	M6	0.06	0.05	0.06	0.12	0.69	0.37	0.99	0.99	0.05	0.03	0.04	0.06	0.63	0.28	0.98	0.98
$n = 300$	M1	0.05	0.07	0.05	0.05	0.39	0.27	0.79	0.79	0.03	0.04	0.05	0.04	0.35	0.22	0.78	0.85
	M2	0.03	0.05	0.06	0.06	0.92	0.65	1.00	1.00	0.04	0.03	0.05	0.04	0.91	0.55	1.00	1.00
	M3	0.07	0.06	0.07	0.07	0.78	0.51	1.00	0.99	0.04	0.04	0.05	0.06	0.71	0.47	1.00	1.00
	M4	0.05	0.07	0.05	0.06	0.97	0.64	1.00	1.00	0.05	0.06	0.05	0.04	0.96	0.58	1.00	1.00
	M5	0.04	0.05	0.07	0.07	1.00	0.89	1.00	1.00	0.03	0.05	0.05	0.04	1.00	0.88	1.00	1.00
	M6	0.04	0.07	0.06	0.08	0.91	0.60	1.00	1.00	0.04	0.04	0.04	0.05	0.88	0.57	1.00	1.00

Table 10: Empirical Rejection Frequencies of $\mathcal{S}_9$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.05	0.07	0.05	0.06	0.26	0.17	0.47	0.51	0.05	0.04	0.06	0.06	0.25	0.15	0.49	0.54
	M2	0.05	0.06	0.04	0.06	0.76	0.42	0.98	0.96	0.03	0.03	0.04	0.04	0.72	0.38	0.94	0.93
	M3	0.07	0.06	0.07	0.07	0.50	0.30	0.89	0.85	0.03	0.04	0.03	0.05	0.43	0.24	0.81	0.84
	M4	0.04	0.04	0.03	0.05	0.90	0.52	0.99	0.98	0.04	0.04	0.06	0.04	0.89	0.42	0.99	0.97
	M5	0.07	0.05	0.05	0.05	0.98	0.63	1.00	0.99	0.04	0.05	0.05	0.04	0.98	0.59	0.99	0.99
	M6	0.06	0.05	0.07	0.07	0.51	0.27	0.83	0.78	0.06	0.06	0.04	0.03	0.53	0.25	0.77	0.80
$n = 300$	M1	0.05	0.05	0.05	0.04	0.49	0.27	0.76	0.77	0.03	0.03	0.01	0.05	0.48	0.31	0.86	0.88
	M2	0.06	0.05	0.06	0.04	0.98	0.69	0.99	0.98	0.03	0.03	0.05	0.03	0.97	0.65	0.98	0.98
	M3	0.05	0.06	0.03	0.07	0.78	0.49	0.97	0.96	0.03	0.04	0.07	0.03	0.80	0.47	0.94	0.93
	M4	0.06	0.06	0.04	0.04	1.00	0.76	0.99	0.98	0.03	0.03	0.06	0.03	0.99	0.78	1.00	0.99
	M5	0.05	0.04	0.05	0.06	0.99	0.79	1.00	0.99	0.04	0.03	0.05	0.05	0.99	0.80	0.99	0.99
	M6	0.04	0.06	0.06	0.06	0.60	0.38	0.82	0.83	0.05	0.03	0.04	0.04	0.58	0.31	0.81	0.80

Table 11: Empirical Rejection Frequencies of $\mathcal{S}_{10}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.03	0.06	0.04	0.06	0.27	0.15	0.49	0.52	0.04	0.03	0.05	0.04	0.17	0.13	0.48	0.56
	M2	0.04	0.05	0.05	0.05	0.73	0.44	0.98	0.98	0.03	0.04	0.06	0.04	0.65	0.34	0.98	0.97
	M3	0.06	0.05	0.07	0.05	0.50	0.27	0.87	0.91	0.05	0.04	0.04	0.03	0.40	0.23	0.84	0.87
	M4	0.06	0.05	0.07	0.06	0.84	0.45	0.98	0.99	0.04	0.05	0.03	0.06	0.83	0.37	0.99	0.98
	M5	0.05	0.06	0.05	0.07	0.98	0.69	1.00	1.00	0.03	0.03	0.05	0.04	0.98	0.63	1.00	0.99
	M6	0.06	0.07	0.07	0.08	0.65	0.37	0.91	0.94	0.05	0.04	0.05	0.03	0.60	0.31	0.90	0.90
$n = 300$	M1	0.05	0.04	0.07	0.05	0.44	0.30	0.84	0.86	0.04	0.03	0.04	0.04	0.42	0.29	0.86	0.90
	M2	0.06	0.06	0.05	0.06	0.97	0.74	1.00	1.00	0.03	0.05	0.04	0.05	0.97	0.67	1.00	1.00
	M3	0.04	0.05	0.05	0.04	0.80	0.52	1.00	1.00	0.04	0.04	0.03	0.04	0.77	0.52	1.00	0.99
	M4	0.06	0.06	0.05	0.07	0.99	0.79	1.00	1.00	0.05	0.06	0.03	0.04	1.00	0.71	1.00	1.00
	M5	0.03	0.05	0.07	0.05	1.00	0.90	1.00	1.00	0.06	0.08	0.05	0.04	1.00	0.90	1.00	1.00
	M6	0.08	0.06	0.08	0.07	0.85	0.51	0.98	0.98	0.03	0.05	0.05	0.05	0.84	0.53	0.97	0.98

Table 12: Empirical Rejection Frequencies of $\mathcal{S}_{11}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.05	0.06	0.05	0.05	0.25	0.20	0.47	0.49	0.03	0.05	0.05	0.04	0.23	0.18	0.50	0.54
	M2	0.04	0.06	0.07	0.07	0.78	0.45	0.97	0.97	0.05	0.04	0.03	0.04	0.76	0.34	0.96	0.95
	M3	0.03	0.05	0.07	0.05	0.48	0.29	0.88	0.88	0.03	0.04	0.04	0.05	0.45	0.23	0.85	0.88
	M4	0.05	0.03	0.07	0.05	0.89	0.50	0.98	0.99	0.06	0.03	0.05	0.03	0.89	0.44	0.99	0.98
	M5	0.05	0.06	0.08	0.06	0.99	0.69	1.00	1.00	0.04	0.05	0.06	0.05	0.97	0.62	0.99	1.00
	M6	0.07	0.08	0.07	0.09	0.54	0.34	0.87	0.87	0.03	0.06	0.04	0.04	0.54	0.28	0.81	0.82
$n = 300$	M1	0.05	0.05	0.06	0.07	0.46	0.29	0.87	0.87	0.04	0.05	0.05	0.06	0.46	0.34	0.87	0.91
	M2	0.06	0.06	0.05	0.05	0.97	0.68	1.00	1.00	0.03	0.04	0.03	0.04	0.97	0.70	1.00	1.00
	M3	0.05	0.05	0.06	0.05	0.84	0.54	1.00	0.99	0.03	0.03	0.05	0.04	0.80	0.50	1.00	0.99
	M4	0.07	0.05	0.05	0.05	1.00	0.79	1.00	1.00	0.04	0.04	0.05	0.03	1.00	0.76	1.00	1.00
	M5	0.04	0.06	0.05	0.05	1.00	0.87	1.00	1.00	0.04	0.05	0.03	0.05	0.99	0.89	1.00	1.00
	M6	0.05	0.06	0.05	0.08	0.77	0.50	0.94	0.95	0.04	0.04	0.03	0.05	0.72	0.44	0.93	0.94

Table 13: Empirical Rejection Frequencies of $\mathcal{S}_{12}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.08	0.06	0.06	0.06	0.19	0.15	0.33	0.36	0.04	0.05	0.04	0.06	0.16	0.12	0.37	0.40
	M2	0.07	0.07	0.07	0.08	0.49	0.35	0.91	0.93	0.06	0.06	0.05	0.04	0.40	0.26	0.88	0.90
	M3	0.07	0.05	0.08	0.06	0.36	0.26	0.70	0.70	0.05	0.05	0.04	0.05	0.25	0.20	0.63	0.68
	M4	0.06	0.05	0.05	0.09	0.59	0.26	0.92	0.89	0.04	0.04	0.04	0.06	0.52	0.25	0.89	0.90
	M5	0.04	0.07	0.08	0.08	0.92	0.61	1.00	1.00	0.04	0.07	0.04	0.06	0.88	0.56	1.00	1.00
	M6	0.07	0.07	0.07	0.08	0.59	0.31	0.95	0.97	0.04	0.05	0.05	0.04	0.51	0.25	0.94	0.94
$n = 300$	M1	0.08	0.05	0.04	0.06	0.31	0.25	0.70	0.72	0.03	0.05	0.07	0.04	0.30	0.23	0.72	0.74
	M2	0.08	0.07	0.06	0.05	0.79	0.55	1.00	1.00	0.03	0.05	0.04	0.04	0.72	0.44	1.00	1.00
	M3	0.06	0.06	0.05	0.05	0.64	0.42	0.97	0.96	0.04	0.03	0.05	0.06	0.60	0.35	0.94	0.96
	M4	0.06	0.06	0.07	0.07	0.86	0.51	0.99	0.99	0.05	0.04	0.05	0.04	0.83	0.47	0.99	1.00
	M5	0.06	0.06	0.07	0.08	1.00	0.84	1.00	1.00	0.06	0.05	0.04	0.03	0.99	0.84	1.00	1.00
	M6	0.05	0.07	0.06	0.06	0.80	0.54	0.99	1.00	0.05	0.05	0.04	0.04	0.77	0.50	1.00	0.99

Table 14: Empirical Rejection Frequencies of $\mathcal{S}_{13}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.06	0.09	0.05	0.07	0.17	0.15	0.35	0.35	0.05	0.07	0.08	0.06	0.15	0.10	0.28	0.31
	M2	0.08	0.07	0.09	0.08	0.33	0.21	0.63	0.70	0.06	0.08	0.07	0.07	0.27	0.14	0.55	0.61
	M3	0.06	0.07	0.07	0.06	0.32	0.18	0.60	0.67	0.05	0.05	0.05	0.05	0.24	0.18	0.46	0.54
	M4	0.08	0.08	0.11	0.07	0.32	0.18	0.48	0.46	0.05	0.07	0.06	0.08	0.27	0.16	0.45	0.47
	M5	0.06	0.05	0.05	0.08	0.62	0.32	0.97	0.98	0.06	0.04	0.06	0.08	0.61	0.31	0.96	0.96
	M6	0.05	0.07	0.08	0.07	0.36	0.20	0.75	0.84	0.07	0.07	0.09	0.08	0.36	0.20	0.78	0.78
$n = 300$	M1	0.06	0.06	0.08	0.06	0.27	0.23	0.60	0.64	0.02	0.05	0.04	0.05	0.19	0.14	0.50	0.54
	M2	0.06	0.05	0.05	0.06	0.53	0.25	0.91	0.91	0.03	0.05	0.04	0.03	0.39	0.22	0.85	0.86
	M3	0.06	0.06	0.07	0.06	0.45	0.32	0.87	0.89	0.03	0.03	0.04	0.04	0.34	0.23	0.76	0.80
	M4	0.05	0.06	0.06	0.07	0.52	0.23	0.77	0.76	0.03	0.04	0.04	0.05	0.46	0.19	0.69	0.68
	M5	0.04	0.08	0.06	0.07	0.88	0.51	1.00	1.00	0.04	0.04	0.05	0.06	0.86	0.48	1.00	1.00
	M6	0.07	0.07	0.07	0.06	0.56	0.34	0.95	0.99	0.05	0.06	0.05	0.06	0.56	0.29	0.94	0.98

Table 15: Empirical Rejection Frequencies of $\mathcal{S}_{14}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.06	0.07	0.07	0.05	0.24	0.15	0.44	0.49	0.04	0.06	0.05	0.09	0.15	0.13	0.45	0.44
	M2	0.07	0.07	0.08	0.07	0.43	0.21	0.74	0.79	0.06	0.08	0.08	0.06	0.36	0.21	0.70	0.73
	M3	0.05	0.06	0.04	0.07	0.29	0.21	0.70	0.77	0.05	0.07	0.09	0.06	0.29	0.14	0.65	0.69
	M4	0.07	0.06	0.07	0.07	0.51	0.23	0.65	0.68	0.05	0.06	0.08	0.05	0.39	0.24	0.72	0.68
	M5	0.07	0.08	0.07	0.05	0.73	0.38	0.99	0.99	0.04	0.07	0.06	0.07	0.75	0.39	0.99	0.99
	M6	0.07	0.06	0.08	0.08	0.39	0.23	0.80	0.82	0.07	0.08	0.06	0.10	0.38	0.23	0.77	0.82
$n = 300$	M1	0.06	0.06	0.06	0.07	0.36	0.25	0.70	0.81	0.02	0.04	0.04	0.05	0.26	0.22	0.71	0.74
	M2	0.05	0.05	0.05	0.05	0.65	0.42	0.96	0.97	0.05	0.02	0.03	0.05	0.57	0.32	0.91	0.96
	M3	0.06	0.05	0.06	0.07	0.60	0.37	0.96	0.97	0.04	0.04	0.06	0.07	0.50	0.34	0.92	0.95
	M4	0.04	0.06	0.06	0.07	0.70	0.36	0.94	0.93	0.02	0.03	0.04	0.04	0.66	0.29	0.92	0.90
	M5	0.05	0.07	0.06	0.08	0.95	0.64	1.00	1.00	0.04	0.03	0.03	0.04	0.93	0.59	1.00	1.00
	M6	0.06	0.07	0.06	0.07	0.60	0.39	0.97	0.99	0.05	0.04	0.04	0.06	0.56	0.32	0.97	0.98

Table 16: Empirical Rejection Frequencies of $\mathcal{S}_{15}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.06	0.07	0.06	0.06	0.17	0.11	0.39	0.40	0.02	0.03	0.03	0.03	0.16	0.09	0.37	0.43
	M2	0.06	0.04	0.07	0.05	0.36	0.20	0.82	0.81	0.01	0.03	0.02	0.03	0.45	0.26	0.90	0.90
	M3	0.04	0.06	0.06	0.05	0.25	0.19	0.67	0.74	0.04	0.03	0.03	0.05	0.35	0.18	0.64	0.72
	M4	0.03	0.05	0.05	0.07	0.34	0.18	0.64	0.58	0.04	0.05	0.06	0.04	0.69	0.26	0.95	0.92
	M5	0.03	0.04	0.06	0.05	0.74	0.37	0.99	0.99	0.04	0.02	0.04	0.04	0.94	0.50	0.99	0.99
	M6	0.05	0.04	0.07	0.05	0.38	0.18	0.88	0.88	0.04	0.05	0.05	0.05	0.48	0.21	0.86	0.89
$n = 300$	M1	0.05	0.03	0.06	0.04	0.36	0.22	0.75	0.78	0.05	0.03	0.04	0.05	0.36	0.25	0.77	0.77
	M2	0.05	0.06	0.06	0.07	0.87	0.55	1.00	1.00	0.02	0.03	0.03	0.01	0.87	0.50	1.00	1.00
	M3	0.04	0.05	0.05	0.06	0.73	0.46	0.99	0.99	0.04	0.02	0.03	0.03	0.67	0.39	0.98	0.98
	M4	0.06	0.06	0.06	0.06	0.97	0.63	1.00	1.00	0.04	0.05	0.05	0.04	0.96	0.54	1.00	1.00
	M5	0.04	0.05	0.04	0.04	1.00	0.80	1.00	1.00	0.03	0.03	0.06	0.04	0.99	0.76	1.00	0.99
	M6	0.04	0.04	0.03	0.04	0.84	0.49	0.99	0.99	0.03	0.03	0.04	0.04	0.84	0.49	0.99	0.99

Table 17: Empirical Rejection Frequencies of $\mathcal{S}_{16}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.04	0.08	0.06	0.06	0.21	0.13	0.27	0.28	0.04	0.04	0.04	0.05	0.19	0.14	0.43	0.49
	M2	0.05	0.06	0.06	0.06	0.70	0.40	0.95	0.96	0.05	0.02	0.03	0.04	0.68	0.30	0.95	0.92
	M3	0.06	0.07	0.06	0.07	0.36	0.24	0.60	0.62	0.03	0.03	0.03	0.03	0.38	0.20	0.76	0.72
	M4	0.04	0.04	0.03	0.06	0.87	0.44	0.97	0.97	0.05	0.04	0.03	0.03	0.87	0.42	0.98	0.97
	M5	0.04	0.04	0.09	0.08	0.98	0.66	0.99	1.00	0.05	0.04	0.06	0.02	0.97	0.62	0.99	0.99
	M6	0.06	0.09	0.09	0.12	0.56	0.38	0.88	0.93	0.02	0.04	0.04	0.04	0.49	0.25	0.82	0.85
$n = 300$	M1	0.05	0.06	0.05	0.05	0.38	0.26	0.62	0.66	0.03	0.03	0.02	0.04	0.44	0.26	0.85	0.85
	M2	0.04	0.05	0.06	0.06	0.98	0.72	1.00	1.00	0.03	0.03	0.05	0.03	0.98	0.66	1.00	1.00
	M3	0.06	0.05	0.04	0.05	0.78	0.51	0.97	0.96	0.03	0.03	0.05	0.03	0.81	0.46	0.97	0.97
	M4	0.05	0.04	0.05	0.04	1.00	0.78	1.00	1.00	0.05	0.04	0.04	0.03	1.00	0.75	1.00	1.00
	M5	0.04	0.07	0.04	0.07	1.00	0.91	1.00	1.00	0.04	0.05	0.04	0.04	0.99	0.88	1.00	1.00
	M6	0.05	0.08	0.09	0.08	0.77	0.52	0.97	0.97	0.05	0.06	0.02	0.03	0.73	0.45	0.96	0.96

Table 18: Empirical Rejection Frequencies of $\mathcal{S}_{17}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.05	0.06	0.06	0.07	0.20	0.13	0.28	0.27	0.03	0.04	0.02	0.04	0.20	0.14	0.45	0.46
	M2	0.04	0.07	0.05	0.07	0.72	0.41	0.94	0.96	0.04	0.02	0.03	0.04	0.69	0.31	0.92	0.91
	M3	0.05	0.06	0.06	0.05	0.38	0.22	0.63	0.63	0.02	0.05	0.03	0.04	0.34	0.25	0.74	0.80
	M4	0.04	0.05	0.06	0.06	0.88	0.45	0.99	0.98	0.04	0.03	0.05	0.04	0.84	0.40	0.98	0.96
	M5	0.05	0.09	0.06	0.07	0.98	0.61	0.99	0.99	0.03	0.05	0.03	0.05	0.95	0.61	0.99	0.99
	M6	0.05	0.09	0.08	0.10	0.57	0.38	0.85	0.90	0.03	0.03	0.03	0.04	0.51	0.24	0.87	0.86
$n = 300$	M1	0.07	0.04	0.05	0.07	0.39	0.26	0.65	0.62	0.03	0.04	0.05	0.03	0.43	0.28	0.82	0.86
	M2	0.05	0.05	0.04	0.05	0.97	0.74	1.00	1.00	0.06	0.03	0.03	0.04	0.98	0.63	1.00	1.00
	M3	0.05	0.06	0.04	0.04	0.77	0.48	0.95	0.94	0.04	0.04	0.03	0.05	0.77	0.52	0.97	0.97
	M4	0.05	0.04	0.03	0.04	1.00	0.79	1.00	1.00	0.06	0.04	0.04	0.04	0.99	0.75	1.00	1.00
	M5	0.04	0.06	0.05	0.05	1.00	0.90	1.00	1.00	0.04	0.04	0.04	0.05	1.00	0.87	1.00	1.00
	M6	0.06	0.05	0.07	0.07	0.77	0.52	0.98	0.98	0.04	0.03	0.04	0.04	0.73	0.46	0.96	0.96

Table 19: Empirical Rejection Frequencies of $\mathcal{S}_{18}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.05	0.06	0.05	0.06	0.31	0.22	0.63	0.67	0.05	0.06	0.05	0.05	0.26	0.21	0.49	0.58
	M2	0.06	0.04	0.05	0.05	0.50	0.31	0.81	0.81	0.02	0.03	0.03	0.04	0.47	0.26	0.76	0.74
	M3	0.03	0.05	0.04	0.06	0.49	0.31	0.84	0.90	0.04	0.03	0.04	0.07	0.35	0.25	0.77	0.80
	M4	0.06	0.07	0.06	0.06	0.71	0.36	0.96	0.94	0.03	0.05	0.04	0.03	0.64	0.35	0.95	0.95
	M5	0.05	0.05	0.04	0.03	0.79	0.41	0.97	0.96	0.04	0.03	0.03	0.05	0.77	0.40	0.96	0.95
	M6	0.07	0.04	0.05	0.04	0.33	0.22	0.62	0.66	0.02	0.03	0.03	0.03	0.30	0.21	0.59	0.65
$n = 300$	M1	0.05	0.04	0.04	0.06	0.50	0.39	0.92	0.92	0.04	0.04	0.04	0.04	0.41	0.29	0.87	0.85
	M2	0.05	0.03	0.04	0.06	0.82	0.51	0.98	0.99	0.04	0.03	0.05	0.03	0.77	0.46	0.97	0.97
	M3	0.05	0.06	0.05	0.05	0.75	0.49	0.99	1.00	0.02	0.05	0.06	0.03	0.70	0.43	0.98	0.99
	M4	0.05	0.03	0.05	0.04	0.90	0.60	1.00	1.00	0.04	0.03	0.04	0.05	0.91	0.63	1.00	1.00
	M5	0.05	0.03	0.04	0.04	0.95	0.67	0.99	1.00	0.04	0.03	0.04	0.05	0.95	0.65	0.99	0.99
	M6	0.04	0.04	0.05	0.04	0.55	0.38	0.85	0.89	0.04	0.03	0.03	0.04	0.54	0.37	0.86	0.90

Table 20: Empirical Rejection Frequencies of $\mathcal{S}_{19}$

		ARSB								STB							
		N	S1	S2	S3	A1	A2	A3	A4	N	S1	S2	S3	A1	A2	A3	A4
$n = 150$	M1	0.05	0.07	0.07	0.06	0.26	0.22	0.60	0.64	0.03	0.04	0.05	0.05	0.21	0.14	0.48	0.57
	M2	0.07	0.05	0.06	0.05	0.79	0.43	0.99	1.00	0.04	0.03	0.04	0.02	0.72	0.36	0.98	0.99
	M3	0.05	0.06	0.05	0.07	0.55	0.31	0.94	0.97	0.04	0.04	0.04	0.04	0.49	0.25	0.88	0.89
	M4	0.04	0.07	0.04	0.06	0.82	0.45	0.99	0.98	0.05	0.06	0.05	0.05	0.83	0.43	1.00	0.99
	M5	0.05	0.05	0.07	0.04	0.98	0.61	1.00	1.00	0.03	0.04	0.05	0.03	0.97	0.60	1.00	1.00
	M6	0.06	0.05	0.06	0.07	0.66	0.35	0.98	0.99	0.06	0.04	0.05	0.06	0.64	0.30	0.98	0.99
$n = 300$	M1	0.04	0.04	0.04	0.06	0.49	0.35	0.89	0.94	0.05	0.03	0.06	0.04	0.45	0.32	0.88	0.90
	M2	0.02	0.04	0.04	0.06	0.97	0.71	1.00	1.00	0.04	0.04	0.03	0.03	0.95	0.64	1.00	1.00
	M3	0.05	0.04	0.04	0.05	0.85	0.57	1.00	1.00	0.03	0.04	0.05	0.04	0.79	0.55	1.00	1.00
	M4	0.04	0.05	0.05	0.05	0.99	0.72	1.00	1.00	0.05	0.05	0.05	0.05	1.00	0.69	1.00	1.00
	M5	0.05	0.04	0.06	0.05	1.00	0.86	1.00	1.00	0.05	0.03	0.03	0.03	1.00	0.89	1.00	1.00
	M6	0.05	0.06	0.04	0.02	0.90	0.62	1.00	1.00	0.06	0.05	0.06	0.06	0.91	0.59	1.00	1.00