



## UvA-DARE (Digital Academic Repository)

### Transfer Entropy for Nonparametric Granger Causality Detection

*An Evaluation of Different Resampling Methods*

Diks, C.; Fang, H.

**DOI**

[10.3390/e19070372](https://doi.org/10.3390/e19070372)

**Publication date**

2017

**Document Version**

Final published version

**Published in**

Entropy

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Diks, C., & Fang, H. (2017). Transfer Entropy for Nonparametric Granger Causality Detection: An Evaluation of Different Resampling Methods. *Entropy*, 19(7), [372].  
<https://doi.org/10.3390/e19070372>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Article

# Transfer Entropy for Nonparametric Granger Causality Detection: An Evaluation of Different Resampling Methods

Cees Diks <sup>1,2</sup> and Hao Fang <sup>1,2,\*</sup>,<sup>†</sup>

<sup>1</sup> CeNDEF, Amsterdam School of Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands; c.g.h.diks@uva.nl

<sup>2</sup> Tinbergen Institute, 1082 MS Amsterdam, The Netherlands

\* Correspondence: h.fang@uva.nl; Tel.: +31-06-8412-4179

<sup>†</sup> Current address: Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands.

Academic Editor: Michael (Mike) Stutzer

Received: 16 May 2017; Accepted: 17 July 2017; Published: 21 July 2017

**Abstract:** The information-theoretical concept transfer entropy is an ideal measure for detecting conditional independence, or Granger causality in a time series setting. The recent literature indeed witnesses an increased interest in applications of entropy-based tests in this direction. However, those tests are typically based on nonparametric entropy estimates for which the development of formal asymptotic theory turns out to be challenging. In this paper, we provide numerical comparisons for simulation-based tests to gain some insights into the statistical behavior of nonparametric transfer entropy-based tests. In particular, surrogate algorithms and smoothed bootstrap procedures are described and compared. We conclude this paper with a financial application to the detection of spillover effects in the global equity market.

**Keywords:** transfer entropy; Granger causality; nonparametric test; bootstrap; surrogate data

**MSC:** 62G09; 62G10; 91B84

## 1. Introduction

Entropy, introduced by Shannon [1,2], is an information theoretical concept with several appealing properties, and therefore wide applications in information theory, thermodynamics and time series analysis. Based on this classical measure, *transfer entropy* (TE) has become a popular information theoretical measure for quantifying the flow of information. This concept, which was coined by Schreiber [3], was applied to distinguish a possible asymmetric information exchange between the variables of a bivariate system. When based on appropriate non-parametric density estimates, the TE is a flexible non-parametric measure for conditional dependence, coupling structure, or Granger causality in a general sense.

The notion of Granger causality was developed by the pioneering work of Granger [4] to capture causal interactions in a linear system. In a more general model-free world, the Granger causal effect can be interpreted as the impact of incorporating the history of another variable on the conditional distribution of a future variable in addition to its own history. Recently, various nonparametric measures have been developed to capture such difference between conditional distributions in a more complex, and typically nonlinear system. There is a growing list of such methods, based on, among others, correlation integrals [5], kernel density estimation [6], Hellinger distances [7], copula functions [8] and empirical likelihood [9].

In contrast with the above-mentioned methods, TE-based causality tests do not attempt to capture the difference between two conditional distributions explicitly. Instead, with the information theoretical interpretation, the TE offers a natural way to measure directional information transfer and Granger causality. We refer to [10,11] for detailed reviews of the relation between Granger causality and directed information theory. However, the direct application of entropy and its variants, though attractive, turns out to be difficult, if not impossible altogether, due to the lack of asymptotic distribution theory for the test statistics. For example, Granger and Lin normalize the entropy to detect serial dependence with critical values obtained from simulations [12]. Hong and White provide the asymptotic distribution for the Granger–Lin statistic with a specific kernel function [13]. Barnett and Bossomaier derive a  $\chi^2$  distribution for the TE at the cost of the model-free property [14].

On the other hand, to obviate the asymptotic problem, several resampling methods on TE have been developed for providing empirical distributions of the test statistics. Two popular techniques are bootstrapping and surrogate data. Bootstrapping is a random resampling technique proposed by Efron [15] to estimate the properties of an estimator by measuring those properties from approximating distributions. The “surrogate” approach developed by Theiler et al. [16] is another randomization method initially employing Fourier transforms to provide a benchmark in detecting nonlinearity in a time series setting. It is worth mentioning that the two methods are different with respect to the statistical properties of the resampled data. For the surrogate method the null hypothesis is maintained, while the bootstrap method does not seek to impose the null hypothesis on the bootstrapped samples. We refer to [17,18] for detailed applications of the two methods.

However, not all resampling methods are suitable for entropy-based dependence measures. As Hong and White [13] put it, a standard bootstrap fails to deliver a consistent entropy-based statistic because it does not preserve the statistical properties of a degenerate  $U$ -statistic [13]. Similarly, with respect to traditional surrogates based on phase randomization of the Fourier transform, Hinich et al. [19] criticize the particularly restrictive assumption of linear Gaussian process, and Faes et al. [20] point out that it cannot preserve the whole statistical structure of the original time series.

As far as we are aware, there are several applications of both methods in entropy-based tests, for example, Su and White [7] propose a smoothed local bootstrap for entropy-based test for serial dependence, Papanas et al. [21] apply stationary bootstrap in partial TE estimation, Quiroga et al. [22] use time-shifted surrogates to test the significance of the asymmetry of directional measures of coupling, and Marschinski and Kantz [23] introduce the effective TE, which relies on random shuffling surrogate in estimation. Kugiumtzis [24] and Papanas et al. [21] provide some comparisons between bootstrap and surrogate methods for entropy-based tests.

In this paper, we adopt TE as a test statistic for measuring conditional independence (Granger non-causality). Being aware of the fact that the analytical null distribution may not always be accurate or available in an analytically closed form, we resort to resampling techniques for constructing the empirical null distribution. The techniques under consideration include smoothed local bootstrap, stationary bootstrap and time-shifted surrogates, all of which are shown in literature to be applicable to entropy-based test statistics. Using different dependence structures, the size and power performance of all methods are examined in simulations.

The remainder of this paper is organized as follows. Section 2 first provides the TE-based testing framework and a short introduction to kernel density estimation; then bandwidth selection rules are discussed. After presenting the resampling methods, including the smoothed local bootstrap and time-shifted surrogates for different dependence structure settings. Section 3 examines the empirical performance of different resampling methods, presenting the size and power of the tests. Section 4 considers a financial application with the TE-based nonparametric test and Section 5 summarizes.

## 2. Methodology

### 2.1. Transfer Entropy and Its Estimator

Information theory is a branch of applied mathematical theory of probability and statistics. The central problem of classical information theory is to measure transmission of information over a noisy channel. Entropy, also referred to as Shannon entropy, is one key measure in the field of information theory brought by [1,2]. Entropy measures the uncertainty and randomness associated with a random variable. Supposing that  $S$  is a random vector with density  $f_S(s)$ , its Shannon entropy is defined as

$$H(s) = - \int f_S(s) \log(f_S(s)) ds. \quad (1)$$

There is a long history of applying information theoretical measures in time series analysis. For example, Robinson [25] applies the Kullback–Leibler information criterion [26] to construct a one-sided test for serial independence. Since then, nonparametric tests using entropy measures for dependence between two time series are becoming prevalent. Granger and Lin [12] normalize the entropy measure to identify the lags in a nonlinear bivariate time series model. Granger et al. [27] study dependence with a transformed metric entropy, which turns out to be a proper measure of distance. Hong and White [13] provide a new entropy-based test for serial dependence, and the test statistic follows a standard normal distribution asymptotically.

Although those heuristic approaches work for entropy-based measures of dependence, these methodologies do not carry over directly to measures of *conditional* dependence, i.e., Granger causality. The term TE was coined by Schreiber [3], although it appeared in the literature earlier under different names, is a suitable measure to serve this purpose. The TE quantifies the amount of information contained in one series at  $k$  steps ahead from the state of another series, given the current and past state of itself. Suppose we have two series  $\{X_t\}$  and  $\{Y_t\}$ , for brevity put  $X = \{X_t\}$ ,  $Y = \{Y_t\}$  and  $Z = \{Y_{t+k}\}$ , further we define a three-variate vector  $W_t$  as  $W_t = (X_t, Y_t, Z_t)$ , where  $Z_t = Y_{t+k}$ ; and  $W = (X, Y, Z)$  is used when there is no danger of confusion. Within this bivariate setting,  $W$  is a three dimensional continuous vector. In this paper, we limit ourselves to  $k = 1$  for simplicity, but the method can be generalized into multiple steps easily. The quantity  $TE_{X \rightarrow Y}$  is a nonlinear measure for the amount of information explained in  $Z$  (future  $Y$ ) by  $X$ , accounting for the information on  $Z$  already contained in  $Y$ . Although TE defined in [3] applies to discrete variables, it is easily generalized to continuous variables. Conditional on  $Y$ ,  $TE_{X \rightarrow Y}$  is defined as

$$\begin{aligned} TE_{X \rightarrow Y} &= E_W \left( \log \frac{f_{Z,X|Y}(Z, X|Y)}{f_{X|Y}(X|Y)f_{Z|Y}(Z|Y)} \right) \\ &= \int \int \int f_{X,Y,Z}(X, Y, Z) \log \frac{f_{Z,X|Y}(Z, X|Y)}{f_{X|Y}(X|Y)f_{Z|Y}(Z|Y)} dx dy dz \\ &= E_W \left( \log \frac{f_{X,Y,Z}(X, Y, Z)}{f_Y(Y)} - \log \frac{f_{X,Y}(X, Y)}{f_Y(Y)} - \log \frac{f_{Y,Z}(Y, Z)}{f_Y(Y)} \right) \\ &= E_W (\log f_{X,Y,Z}(X, Y, Z) + \log f_Y(Y) - \log f_{X,Y}(X, Y) - \log f_{Y,Z}(Y, Z)). \end{aligned} \quad (2)$$

Using conditional mutual information  $I(Z, X|Y = y)$ , the TE can be equivalently formulated in terms of four Shannon entropy terms as

$$\begin{aligned} TE_{X \rightarrow Y} &= I(Z, X|Y) \\ &= H(Z|Y) - H(Z|X, Y) \\ &= H(Z, Y) - H(Y) - H(Z, X, Y) + H(X, Y). \end{aligned} \quad (3)$$

In order to construct a test for Granger causality based on the TE, one first needs to show quantitatively that the TE is a proper basis for detecting whether the null hypothesis is satisfied.

The following theorem, as a direct application of the Kullback–Leibler criterion, lays the quantitative foundation for testing based on the TE.

**Theorem 1.**  $TE_{X \rightarrow Y} \geq 0$  with equality if and only if  $f_{Z, X|Y}(Z, X|Y) = f_{X|Y}(X|Y)f_{Z|Y}(Z|Y)$ .

**Proof.** The proof of Theorem 1 is given in [28].  $\square$

It is not difficult to verify that the condition for  $TE_{X \rightarrow Y} = 0$  coincides with the null hypothesis of Granger non-causality defined in Equation (4), also referred to as conditional independence or no coupling. Mathematically speaking, the null hypothesis of Granger non-causality,  $H_0 : \{X_t\}$  is not a Granger cause of  $\{Y_t\}$ , can be phrased as

$$H_0 : \frac{f_{X,Y,Z}(x,y,z)}{f_Y(y)} = \frac{f_{Y,Z}(y,z)}{f_Y(y)} \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad (4)$$

for  $(x, y, z)$  in the support of  $W$ . A nonparametric test for Granger non-causality seeks to find statistical evidence of violation of Equation (4). There are many nonparametric measures available for this purpose, some of which are mentioned above. Equation (4) provides the basis for a model-free test without imposing any parametric assumptions about the data generating process or underlying distributions for  $\{X_t\}$  and  $\{Y_t\}$ . We only assume two things here. First,  $\{X_t, Y_t\}$  is a strictly stationary bivariate process. Second, the process has finite memory, i.e., variable lags  $l_X, l_Y \ll \infty$ . The second (finite Markov order) assumption is needed in this nonparametric setting to make conditioning on past information feasible by conditioning on a finite number of past observations. Moreover, strict stationarity and the mixing properties implied by the finite Markov order assumption ensure that the transfer entropy can be estimated consistently through kernel density estimation of the underlying densities.

As far as we are aware, the direct use of TE to test Granger non-causality in nonparametric setting is difficult, if not impossible at all, due to the lack of asymptotic theory for the test statistic. As Granger and Lin [12] put it, very few asymptotic distribution results for entropy-based estimators are available. Although over the years several break-throughs have been made with application of entropy to testing serial independence, the limiting distribution of TE statistic is still unknown. One may wish to use simulation techniques to overcome the lack of asymptotic distributions. However, as noted by Su and White [7], there are estimation biases of the TE statistics for non-parametric dependence measures under the smoothed bootstrap procedure. Even for the parametric test statistic used by Barnett and Bossomaier [14], the authors noticed that the TE-based estimator is generally biased.

## 2.2. Density Estimation and Bandwidth Selection

The non-negativity property in Theorem 1 makes  $TE_{X \rightarrow Y}$  a desirable measure for constructing a one-sided test of conditional independence; any positive divergence from zero is a sign of conditional dependence of  $Y$  on  $X$ . To estimate TE there are several different approaches, such as histogram-based estimators [29], correlation sums [30] and nearest neighbor estimators [31]. However, the optimal rule for the number of neighbor points is unclear, and as Kraskov et al. [31] comment, a small value of neighbor points may lead to large statistical errors. A more natural method, kernel density estimators, the properties of which have been well studied, is applied in this paper. With the plug-in kernel estimates of densities, we may replace the expectation in Equation (2) by a sample average to get an estimate for  $TE_{X \rightarrow Y}$ .

A local density estimator of a  $d_W$ -variate random vector  $W$  at  $W_i$  is given by

$$\hat{f}_W(W_i) = ((n-1)h)^{-d_W} \sum_{j,j \neq i}^n \mathbb{K} \left( \frac{W_i - W_j}{h} \right), \quad (5)$$

where  $\mathbb{K}$  is a kernel function and  $h$  is the bandwidth. We take  $\mathbb{K}(\cdot)$  to be a product kernel function defined as  $\mathbb{K}(W) = \prod_{s=1}^{d_W} \kappa(w^s)$ , where  $w^s$  is  $s^{th}$  element in  $W$ . Using a standard univariate Gaussian kernel,  $\kappa(w^s) = (2\pi)^{-1/2} e^{-\frac{1}{2}(w^s)^2}$ ,  $\mathbb{K}(\cdot)$  is the standard multivariate Gaussian kernel as described by Wand and Jones [32] and Silverman [33]. Using Equation (5) as the plug-in density estimator, and replacing the expectation by the sample mean, we obtain the estimator for the TE given by

$$\hat{I}(Z, X|Y) = \frac{1}{n} \sum_{i=1}^n \left( \log \hat{f}(x_i, y_i, z_i) + \log \hat{f}(y_i) - \log \hat{f}(x_i, y_i) - \log \hat{f}(y_i, z_i) \right). \quad (6a)$$

If we estimate the Shannon entropy in Equation (1) based on a sample of size  $n$  from the  $d_W$ -dimensional random vector  $W$ , by the sample average of the plug-in density estimates, we obtain

$$\hat{H}(W) = \frac{1}{n} \sum_{i=1}^n \left( \log(\hat{f}_W(W_i)) \right), \quad (6b)$$

then Equation (6a) can be equivalently expressed in terms of four entropy estimators, that is,

$$\hat{I}(Z, X|Y) = -\hat{H}(x_i, y_i, z_i) - \hat{H}(y_i) + \hat{H}(x_i, y_i) + \hat{H}(y_i, z_i). \quad (6c)$$

To construct a statistical test, we develop the asymptotic properties of  $\hat{I}(Z, X|Y)$  defined in Equations (6a) and (6c) through two steps. In the first step, given the density estimates the consistency of entropy estimates is achieved and then the linear combination of four entropy estimates would converge in probability to the true value. The following two theorems ensure the consistency of  $\hat{I}(Z, X|Y)$ .

**Theorem 2.** Given the kernel density estimate  $\hat{f}_W(W_i)$  for  $f_W(W_i)$ , where  $W$  is a  $d_W$ -dimensional random vector with length  $n$ , let  $\hat{H}(W)$  be the plug-in estimate for the Shannon entropy as defined in Equation (6b). Then  $\hat{H}(W) \xrightarrow{P} H(W)$ .

**Proof.** The proof of Theorem 2 is given in [12] using results from [34].  $\square$

The basic idea of the proof is to take the Taylor series expansion of  $\log(\hat{f}_W(W_i))$  around the true value  $\log(f_W(W_i))$  and use the fact that  $\hat{f}_W(W_i)$ , given an appropriate bandwidth sequence, converges to  $f_W(W_i)$  pointwise to obtain consistency. In the next step, the consistency of  $\hat{I}(Z, X|Y)$  is provided by the continuous mapping theorem.

**Theorem 3.** Given  $\hat{H}(W) \xrightarrow{P} H(W)$ , with  $\hat{I}(Z, X|Y)$  defined as in Equation (6c),  $\hat{I}(Z, X|Y) \xrightarrow{P} I(Z, X|Y)$ .

**Proof.** The proof is straightforward if one applies the Continuous Mapping Theorem. See Theorem 2.3 in [35].  $\square$

Before we move to the next section, it is worth having a careful look at the issue of bandwidth selection. The bandwidth  $h$  for kernel estimation determines how smooth the density estimation is; a smaller bandwidth reveals more structure of the data, whereas a larger bandwidth delivers a smoother density estimate. Bandwidth selection is essentially a trade-off between the bias and variance in density estimation. A very small value of  $h$  could eliminate estimation bias, with a large variance. On the other hand, a large bandwidth reduces estimation variance at the expense of incorporating more bias. See Chapter 3 in [32] and [36] for details.

However, the bandwidth selection for the TE statistic is more involved. To the best of our knowledge, there is a blank in the field of optimal bandwidth selection in kernel-based TE estimator. As He et al. [37] show, when estimating the entropy estimator, two types of errors would be generated, one is from entropy estimation and the other from density estimation, and the optimal bandwidth for

density estimation may not coincide with the optimal one for entropy estimation. Thus, rather than the rule-of-thumb bandwidth in [33], which aims at optimal density estimation, the bandwidth in our study should provide an accurate estimator for  $I(Z, X|Y)$  in the minimal mean squared error (MSE) sense, say. In [28] we develop such a bandwidth rule for a TE-based estimator. In that paper, rather than directly developing asymptotic properties for TE, we study the properties of the first order Taylor expansion of TE under the null hypothesis. The suggested bandwidth is shown to be MSE optimal and allows us to obtain asymptotic normality for the test statistic. In principal, the convergence rate of the TE estimator should be the same as the leading term of its Taylor approximation. We therefore propose to use the same rate also here, giving

$$h = Cn^{-2/7}, \quad (7)$$

where  $C$  is an unknown parameter. This bandwidth would deliver a consistent test since the variance of local estimate of  $\hat{I}(Z_i, X_i|Y_i)$  will dominate the MSE. In [28] we suggest to use  $C = 4.8$  based on simulations, while Diks and Panchenko suggest to set  $C \approx 8$  for autoregressive conditional heteroskedasticity (ARCH) processes [6]. Our simulations here also may possibly prefer a larger value of  $C$  because the squared bias is of higher order and hence less concern for the TE-based statistic. A larger bandwidth could better control the estimation variance and deliver a more powerful test. As a robustness check, we adopt  $C = 8$  as well as  $C = 4.8$  suggested by our other simulation study [28]. To match the Gaussian kernel, we standardize the data before estimate Equation (6a)–(6c) such that the transformed time series have mean zero and unit variance; very similar results are obtained by matching the mean absolute deviation instead of the variance of the standard Gaussian kernel for TE estimation.

### 2.3. Resampling Methods

To develop simulation-based tests for the null hypothesis, given in Equation (4), of no Granger causality from  $X$  to  $Y$ , or equivalently, for conditional independence, we consider three resampling techniques, i.e., (1) time shifted surrogates developed by Quiroga et al. [22], (2) the smoothed bootstrap of Su and White [7] and (3) the stationary bootstrap introduced by Politis and Romano [38]. The first technique is widely applied in coupling measures, as for example by Kugiumtzis [39] and Papanas et al. [40], while the latter two have already been used for detecting conditional independence for decades. It worth mentioning that the surrogates and bootstrap methods treat the null quite differently. Surrogate data are supposed to preserve the dependence structure imposed by  $H_0$  while bootstrap data are not restricted to  $H_0$ . It is possible to bootstrap the dataset without imposing the conditional independence structure of  $\{X, Y, Z\}$  implied by the null hypothesis; see, for instance, [41] for more details. To avoid resampling errors and to make different methods more comparable, we limit ourselves to methods that impose the null hypothesis on the resampled data. The following three different resampling methods are implemented with different sampling details.

#### 1. Time-Shifted Surrogates

- (TS.a) The first resampling method only deals with the driving variable  $X$ . Suppose we have observations  $\{x_1, \dots, x_n\}$ , the time-shifted surrogates are generated by cyclically time-shifting the components of the time series. Specifically, an integer  $d$  is randomly generated within the interval  $([0.05n], [0.95n])$ , and then the first  $d$  values of  $\{x_1, \dots, x_n\}$  would be moved to the end of the series, to deliver the surrogate sample  $X^* = \{x_{d+1}, \dots, x_n, x_1, \dots, x_d\}$ . Compared with the traditional surrogates based on phase randomization of the Fourier transform, the time-shifted surrogates can preserve the whole statistical structure in  $X$ . The couplings between  $X$  and  $Y$  are destroyed, although the null hypothesis of  $X$  not causing  $Y$  is imposed.
- (TS.b) The second scheme resamples both the driving variable  $X$  and the response variable  $Y$  separately. Similar to (TS.a),  $Y^* = \{y_{c+1}, \dots, y_n, y_1, \dots, y_c\}$  is created given another random

integer  $c$  from the range  $([0.05n], [0.95n])$ . In contrast with the standard time-shifted surrogates described in (TS.a), in this setting we add more noise to the coupling between  $X$  and  $Y$ .

## 2. Smoothed Local Bootstrap

The smoothed bootstrap selects samples from a smoothed distribution instead of drawing observations from the empirical distribution directly. See [42] for a discussion of the smoothed bootstrap procedure. Based on rather mild assumptions, Neumann and Paparoditis [43] show that there is no need to reproduce the whole dependence structure of the stochastic process to get an asymptotically correct nonparametric dependence estimator. Hence a smoothed bootstrap from the estimated conditional density is able to deliver a consistent statistic. Specifically, we consider two versions of the smoothed bootstrap that are different in dependence structure to some extent.

- (SMB.a) In the first setting,  $Y^*$  is resampled without replacement from the smoothed local bootstrap. Given the sample  $Y = \{y_1, \dots, y_n\}$ , the bootstrap sample is generated by adding a smoothing noise term  $\varepsilon_i^Y$  such that  $\tilde{y}_i^* = y_i^* + h_b \varepsilon_i^Y$ , where  $h_b > 0$  is the bandwidth used in bootstrap procedure,  $\varepsilon_i^Y$  represents a sequence of i.i.d.  $N(0, 1)$  random variables. Without random replacement from the original time series, this procedure does not disturb the original dynamics of  $Y = \{y_1, \dots, y_n\}$  at all. After  $Y^*$  is resampled, both  $X^*$  and  $Z^*$  are drawn from the smoothed conditional densities  $f(x|Y^*)$  and  $f(z|Y^*)$  as described in [44].
- (SMB.b) Secondly, we implement the smoothed local bootstrap as in [7]. The only difference between this setting and (SMB.a) is that the bootstrap sample  $Y^*$  is drawn with replacement from the smoothed kernel density.

## 3. Stationary Bootstrap

Politis and Romano [38] propose the stationary bootstrap to maintain serial dependence within the bootstrap time series. This method replicates the time dependence of original data by resampling blocks of the data with randomly varying block length. The lengths of the bootstrap blocks follows a geometric distribution. Given a fixed probability  $p$ , the length  $L_i$  of block  $i$  is decided as  $P(L_i = k) = (1 - p)^{k-1} p$  for  $k = 1, 2, \dots$ , and the starting points of block  $i$  are randomly and uniformly drawn from the original  $n$  observations. To restore the dependence structure exactly under the null, we combine the stationary bootstrap with the smoothed local bootstrap for our simulations.

- (STB) In short, firstly  $y_1^*$  is picked randomly from the original  $n$  observations of  $Y = \{y_1, \dots, y_n\}$ , denoted as  $y_1^* = y_s$  where  $s \in [1, n]$ . With probability  $p$ ,  $y_2^*$  is picked at random from the data set; and with probability  $1 - p$ ,  $y_2^* = y_{s+1}$ , so that  $y_2^*$  would be the next observation to  $y_s$  in original series  $Y = \{y_1, \dots, y_n\}$ . Proceeding in this way,  $\{y_1^*, \dots, y_n^*\}$  can be generated. If  $y_i^* = y_s$  and  $s = n$ , the “circular boundary condition” would kick in, so that  $y_{i+1}^* = y_1$ . After  $Y^* = \{y_1^*, \dots, y_n^*\}$  is generated, both  $X^*$  and  $Z^*$  are randomly drawn from the smoothed conditional densities  $f(x|Y^*)$  and  $f(z|Y^*)$  as in (SMB.b).

The resampling procedure works as the follows: once the TE statistic  $\hat{I}$  for the original data  $\mathbf{W} = \{(X_i, Y_i, Z_i), i = 1, \dots, n\}$  is estimated according to Equation (6a)–(6c), we start to generate the resampled data set, which is denoted by  $\mathbf{W}_j^*$  with  $j = 1, \dots, B$ , where  $B$  is the number of simulations. Using the simulated sample, for each  $j$  we compute the TE statistic  $\hat{I}_j^*$ , in exactly the same way as  $\hat{I}$  was computed. The  $p$ -value for the one-sided test is calculated as

$$\hat{p} = \frac{1}{B+1} \left( 1 + \sum_{j=1}^B \mathbf{1}(\hat{I}_j^* \geq \hat{I}) \right), \quad (8)$$



where the constant 1 is added to avoid  $p$ -values equal to zero.

A final remark concerns the difference between this paper and [21]; there both time-shifted surrogate and the stationary bootstrap are implemented for an entropy-based causality test. However, our paper provides additional insights into several aspects. Firstly, the smoothed bootstrap, being shown in the literature to work for nonparametric kernel estimators under general dependence structure, is applied in our paper. Secondly, they treat the bootstrap and surrogate sample in a similar way, but as we noted above, the bootstrap method is not designed to impose the null hypothesis but designed to keep the dependence structure present in the original data. The stationary bootstrap procedure in [21] might be incompatible with the null hypothesis of conditional independence since it destroys the dependence completely. Because they restore independence between  $X$  and  $Y$  rather than conditional independence between  $X|Y$  and  $Z|Y$  during resampling, the distribution of the estimated statistics from the resampled data may not necessary correspond to that of the statistic under the null of only conditional independence. Thirdly, we provide rigorous size and power result in our simulations, which is missing in their paper.

### 3. Simulation Study

In this section, we investigate the performance of the five resampling methods in detecting conditional dependence for several data generating processes. In Equations (9)–(16), we use a single parameter  $a$  to control the strength of the conditional dependence. The size assessment is obtained based on testing Granger non-causality from  $\{X_t\}$  to  $\{Y_t\}$ , and for the power we use the same process but we test for Granger non-causality from  $\{Y_t\}$  to  $\{X_t\}$ . We set  $a = 0.4$  to represent moderate dependence in the size performance investigation and  $a = 0.1$  to evaluate the power of the tests. Further, Equation (17) represents a stationary autoregressive process with regime switching, and Equation (18) is included to investigate the power performance in the presence of two-way causal linkages, where the two control parameters are  $b = -0.2$  and  $c = 0.1$ .

In each experiment, we run 500 simulations for sample sizes  $n = \{200, 500, 1000, 2000\}$ . The surrogate and the bootstrap sample size is set to  $B = 999$ . For fair comparisons between (TS.a) and (TS.b), as well as between (SMB.a) and (SMB.b), we fix the seeds of the random number generator in the resampling functions to eliminate the potential effect of randomness. Besides, we use the empirical standard deviation of  $\{Y_t\}$  as the bootstrapping bandwidth and  $C = \{4.8, 8\}$  in the bandwidth equation Equation (7) for the kernel density estimation.

The processes under consideration include a linear vector autoregressive process (VAR) in Equation (9), a nonlinear VAR process in Equation (10), a bivariate ARCH process in Equation (11), a bilinear process in Equation (12), a bivariate AR(2)-GARCH process in Equation (13) where “GARCH” stands for “generalized ARCH”, a bivariate autoregressive-moving average (ARMA)-GARCH process in Equation (14), a bivariate AR(1)-EGARCH process in Equation (15) where “EGARCH” represents “exponential GARCH”, a vector error correction (VECM) process in Equation (16), a threshold AR(1) process in Equation (17), and a two-way VAR process in Equation (18).

It is worth mentioning that the data generating processes in Equations (9)–(12), (17) and (18) are stationary and of finite memory as we assumed earlier. However, it is also important to be aware of the behavior of the proposed non-parametric test for robustness consideration when the two assumptions are not satisfied. The finite memory assumption is violated in Equations (13)–(15) since the GARCH process, being equivalent to an infinite ARCH process, strictly speaking is of infinite Markov order; and the stationarity assumption does not hold in Equation (16) where  $X_t$  and  $Y_t$  are cointegrated of order one. Since for the VECM process the two time series  $\{X_t\}$  and  $\{Y_t\}$  are not stationary, we can not directly apply our nonparametric test. In this case, we perform the Engle–Granger approach [45] first to eliminate the influence of the co-integration, and then perform the nonparametric test on the collected stationary residuals from the linear regression of  $\Delta X_t$  and  $\Delta Y_t$  on a constant and the co-integration term. The procedure is similar to that in [46].

1. Linear vector autoregressive process (VAR).

$$\begin{aligned} X_t &= aY_{t-1} + \varepsilon_{x,t}, & \varepsilon_{x,t} &\sim N(0,1) \\ Y_t &= aY_{t-1} + \varepsilon_{y,t}, & \varepsilon_{y,t} &\sim N(0,1). \end{aligned} \quad (9)$$

2. Nonlinear VAR. This process is considered in [47] to show the failure of linear Granger causality test.

$$\begin{aligned} X_t &= aX_{t-1}Y_{t-1} + \varepsilon_{x,t}, & \varepsilon_{x,t} &\sim N(0,1) \\ Y_t &= 0.6Y_{t-1} + \varepsilon_{y,t}, & \varepsilon_{y,t} &\sim N(0,1). \end{aligned} \quad (10)$$

3. Bivariate ARCH process.

$$\begin{aligned} X_t &\sim N(0, 1 + aY_{t-1}^2) \\ Y_t &\sim N(0, 1 + aY_{t-1}^2). \end{aligned} \quad (11)$$

4. Bilinear process considered in [48].

$$\begin{aligned} X_t &= 0.3X_{t-1} + aY_{t-1}\varepsilon_{y,t-1} + \varepsilon_{x,t}, & \varepsilon_{x,t} &\sim N(0,1) \\ Y_t &= 0.4Y_{t-1} + \varepsilon_{y,t}, & \varepsilon_{y,t} &\sim N(0,1). \end{aligned} \quad (12)$$

5. Bivariate AR(2)-GARCH process.

$$\begin{aligned} X_t &= 0.5X_{t-1} - 0.2X_{t-2} + \varepsilon_{x,t}, \\ \varepsilon_{x,t} &= \sqrt{h_{x,t}}v_{x,t}, & v_{x,t} &\sim t(5), \\ h_{x,t} &= 0.9 + a\varepsilon_{y,t-1}^2 + 0.2h_{x,t-1}; \\ Y_t &= 0.3Y_{t-1} + 0.2Y_{t-2} + \varepsilon_{y,t}, \\ \varepsilon_{y,t} &= \sqrt{h_{y,t}}v_{y,t}, & v_{y,t} &\sim t(5), \\ h_{y,t} &= 0.3 + 0.1\varepsilon_{y,t-1}^2 + 0.2h_{y,t-1}. \end{aligned} \quad (13)$$

6. Bivariate ARMA-GARCH process.

$$\begin{aligned} X_t &= 0.3X_{t-1} + 0.3\varepsilon_{x,t-1} + \varepsilon_{x,t}, \\ \varepsilon_{x,t} &= \sqrt{h_{x,t}}v_{x,t}, & v_{x,t} &\sim t(5), \\ h_{x,t} &= 0.5 + a\varepsilon_{y,t-1}^2 + 0.3h_{x,t-1}; \\ Y_t &= 0.4Y_{t-1} - 0.2\varepsilon_{y,t-1} + \varepsilon_{y,t}, \\ \varepsilon_{y,t} &= \sqrt{h_{y,t}}v_{y,t}, & v_{y,t} &\sim t(5), \\ h_{y,t} &= 0.8 + 0.05\varepsilon_{y,t-1}^2 + 0.4h_{y,t-1}. \end{aligned} \quad (14)$$

7. Bivariate AR(1)-EGARCH process.

$$\begin{aligned} X_t &= 0.5X_{t-1} + \varepsilon_{x,t}, \\ \varepsilon_{x,t} &= \sqrt{h_{x,t}}v_{x,t}, & v_{x,t} &\sim t(5), \\ \log(h_{x,t}) &= -0.5 + a\left|\frac{\varepsilon_{y,t-1}^2}{\sqrt{h_{x,t-1}}}\right| + 0.2\frac{\varepsilon_{y,t-1}^2}{\sqrt{h_{x,t-1}}} + 0.9\log(h_{x,t-1}); \\ Y_t &= 0.6Y_{t-1} + \varepsilon_{y,t}, \\ \varepsilon_{y,t} &= \sqrt{h_{y,t}}v_{y,t}, & v_{y,t} &\sim t(5), \\ \log(h_{y,t}) &= -0.6 + 0.05\left|\frac{\varepsilon_{y,t-1}^2}{\sqrt{h_{x,t-1}}}\right| + 0.02\frac{\varepsilon_{y,t-1}^2}{\sqrt{h_{x,t-1}}} + 0.8\log(h_{x,t-1}). \end{aligned} \quad (15)$$

8. VECM process. Note that in this situation both  $\{X_t\}$  and  $\{Y_t\}$  are not stationary.

$$\begin{aligned} X_t &= 1.2 + 0.6Y_{t-1} + \varepsilon_{x,t} \\ \varepsilon_{x,t} &= \sqrt{1 - a^2}v_{x,t} + a\varepsilon_{y,t-1}, & v_{x,t} &\sim N(0,1) \\ Y_t &= Y_{t-1} + \varepsilon_{y,t}, & \varepsilon_{y,t} &\sim N(0,1). \end{aligned} \quad (16)$$

## 9. Threshold AR(1) process.

$$\begin{aligned} X_t &= \begin{cases} 0.2X_{t-1} + \varepsilon_{x,t} & \varepsilon_{x,t} \sim N(0, 1), & \text{if } Y_{t-1} < 0 \\ 0.9X_{t-1} + \varepsilon_{x,t} & \varepsilon_{x,t} \sim N(0, 1), & \text{if } Y_{t-1} \geq 0 \end{cases} \\ Y_t &= \varepsilon_{y,t}, \quad \varepsilon_{y,t} \sim N(0, 1). \end{aligned} \quad (17)$$

## 10. Two-way VAR process.

$$\begin{aligned} X_t &= 0.7X_{t-1} + bY_{t-1} + \varepsilon_{x,t}, & \varepsilon_{x,t} &\sim N(0, 1) \\ Y_t &= cX_{t-1} + 0.5Y_{t-1} + \varepsilon_{y,t}, & \varepsilon_{y,t} &\sim N(0, 1). \end{aligned} \quad (18)$$

The empirical rejection rates are summarized in Tables 1–10. The top panels in each table summarize the empirical rejection rates obtained for the 5% and 10% nominal significance levels for processes (9)–(18) under the null hypothesis, and the bottom panels report the corresponding empirical power under the alternatives. Generally speaking, the size and power are quite satisfactory for almost all combinations of the constant  $C$ , sample size  $n$  and nominal significance level. The performance differences for the various resampling schemes are not substantial.

With respect to the size performance, most of the time we see that the realized rejection rates stay in line with the nominal size. Besides, the bootstrap methods outperform the time-shifted surrogate methods in that their empirical size is slightly closer to the nominal size. Lastly, the size of the tests is not very sensitive to the choice of the constant  $C$  apart from the cases for the models given in Equations (13)–(15), where the data generating process has infinite memory.

From the point of view of power, (TS.a) and (SMB.a) seem to outperform their counterparts, yet, the differences are subtle. Along the dimension of the sample size, clearly we see that the empirical power increases in the sample size in most cases. Furthermore, the results are very robust with respect to choices for the constant  $C$  in the kernel density estimation bandwidth. For the VAR and nonlinear processes given by Equations (9), (10) and (12) a smaller  $C$  seems to give more powerful tests while a larger  $C$  is more beneficial for detecting conditional dependence structure in the (G)ARCH processes of Equations 11 and (13)–(15).

Finally, Table 10 presents the empirical power for the two-way VAR process, where the two variables  $\{X\}$  and  $\{Y\}$  are inter-tangled with each other. Due to the setting  $b = -0.2$  and  $c = 0.1$  in Equation (18), it is obvious that  $\{Y\}$  is a stronger Granger cause for  $\{X\}$  than the other way around. As a consequence, the reported rejection rates in Table 10 are overall higher when testing  $Y \rightarrow X$  than  $X \rightarrow Y$ .

To visualize the simulation results, Figures 1–10 report the empirical size and power against the nominal size. Since the performance of the five difference resampling methods is quite similar, we only show the results for (SMB.a) for simplicity. In each figure, the left (right) panels show the realized size (power), and we choose  $C = 4.8$  ( $C = 8$ ) for the top (bottom) two panels. We can see from the figures that the empirical performance of the TE test are overall satisfactory, apart for those (G)ARCH processes where a small  $C$  may lead to conservative testing size for large sample sizes (see Figures 3a, 5a, 6a and 7a). The under-rejection problem is caused by the inappropriate choice  $C = 4.8$ , which makes the bandwidth for kernel estimation too small. The influence of an inappropriately small bandwidth can also be seen in Figures 5b and 6b, where the test has limited power for the alternative.

**Table 1.** Observed size and power of the TE-based test for the linear VAR process in Equation (9).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0560	0.0500	0.0460	0.0420	0.0440	0.0820	0.0740	0.0740	0.0780	0.0780
	500	0.0740	0.0680	0.0660	0.0700	0.0660	0.1160	0.1120	0.1200	0.1160	0.1220
	1000	0.0620	0.0560	0.0600	0.0560	0.0560	0.0940	0.0920	0.0980	0.0960	0.0980
	2000	0.0380	0.0340	0.0380	0.0460	0.0460	0.0940	0.0920	0.0960	0.0980	0.0980
C = 8	200	0.0500	0.0440	0.0460	0.0460	0.0440	0.1120	0.1000	0.0960	0.0960	0.0920
	500	0.0840	0.0760	0.0760	0.0720	0.0660	0.1360	0.1300	0.1060	0.1160	0.1140
	1000	0.0720	0.0680	0.0620	0.0560	0.0580	0.1280	0.1280	0.1160	0.1260	0.1200
	2000	0.0880	0.0780	0.0820	0.0760	0.0820	0.1420	0.1380	0.1320	0.1340	0.1440
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.1880	0.1980	0.1900	0.1920	0.1900	0.2780	0.2780	0.2880	0.2860	0.2920
	500	0.3460	0.3460	0.3400	0.3480	0.3420	0.4520	0.4460	0.4580	0.4500	0.4500
	1000	0.5440	0.5340	0.5320	0.5340	0.5280	0.6400	0.6520	0.6460	0.6480	0.6460
	2000	0.7500	0.7420	0.7500	0.7460	0.7460	0.8160	0.8080	0.8100	0.8120	0.8120
C = 8	200	0.1660	0.1680	0.1640	0.1680	0.1700	0.2660	0.2640	0.2680	0.2740	0.2740
	500	0.2900	0.2900	0.3020	0.3040	0.3020	0.4020	0.3960	0.3940	0.4020	0.3980
	1000	0.4980	0.4980	0.5000	0.4900	0.4980	0.6040	0.6120	0.6120	0.6140	0.6120
	2000	0.8420	0.8380	0.8400	0.8340	0.8460	0.8960	0.8880	0.8900	0.8900	0.8900

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (9) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study we consider  $C = 4.8$  and  $C = 8$ .

**Table 2.** Observed size and power of the TE-based test for the nonlinear VAR process in Equation (10).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0500	0.0360	0.0340	0.0380	0.0360	0.0760	0.0720	0.0780	0.0760	0.0720
	500	0.0580	0.0580	0.0580	0.0580	0.0580	0.0960	0.0980	0.1040	0.1040	0.0980
	1000	0.0340	0.0360	0.0380	0.0360	0.0380	0.0620	0.0580	0.0740	0.0700	0.0720
	2000	0.0380	0.0320	0.0440	0.0460	0.0420	0.0780	0.0700	0.0960	0.0920	0.0880
C = 8	200	0.0560	0.0480	0.0460	0.0520	0.0480	0.1000	0.0800	0.0940	0.0940	0.0940
	500	0.0640	0.0620	0.0500	0.0480	0.0540	0.1120	0.1100	0.1080	0.1040	0.1040
	1000	0.0440	0.0360	0.0320	0.0300	0.0280	0.0900	0.0860	0.0820	0.0780	0.0800
	2000	0.0260	0.0300	0.0280	0.0280	0.0260	0.0700	0.0640	0.0740	0.0660	0.0640
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.1400	0.1360	0.1380	0.1400	0.1440	0.2480	0.2420	0.2300	0.2360	0.2300
	500	0.3380	0.3360	0.3340	0.3360	0.3320	0.4400	0.4400	0.4440	0.4300	0.4360
	1000	0.6060	0.6040	0.6240	0.6220	0.6220	0.7180	0.7260	0.7140	0.7180	0.7160
	2000	0.8760	0.8760	0.8780	0.8740	0.8780	0.9440	0.9300	0.9320	0.9300	0.9280
C = 8	200	0.0940	0.0900	0.0900	0.0900	0.0880	0.1800	0.1740	0.1700	0.1680	0.1760
	500	0.1880	0.1800	0.1800	0.1760	0.1780	0.2960	0.2960	0.3000	0.2980	0.3020
	1000	0.3800	0.3940	0.3900	0.3840	0.3820	0.5520	0.5440	0.5480	0.5520	0.5480
	2000	0.8340	0.8300	0.8280	0.8220	0.8300	0.9040	0.9040	0.9060	0.8980	0.9040

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (10) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .

**Table 3.** Observed size and power of the TE-based test for the bivariate ARCH process in Equation (11).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0660	0.0580	0.0620	0.0640	0.0620	0.1420	0.1280	0.1260	0.1340	0.1280
	500	0.0640	0.0560	0.0560	0.0560	0.0580	0.1120	0.1060	0.1080	0.1020	0.1000
	1000	0.0480	0.0460	0.0400	0.0420	0.0340	0.0740	0.0700	0.0700	0.0660	0.0620
	2000	0.0220	0.0200	0.0080	0.0080	0.0080	0.0460	0.0500	0.0220	0.0260	0.0180
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.0920	0.0760	0.0760	0.0720	0.0840	0.1540	0.1360	0.1280	0.1260	0.1320
	500	0.1100	0.0900	0.0720	0.0760	0.0800	0.1980	0.1860	0.1620	0.1480	0.1600
	1000	0.0920	0.0960	0.0740	0.0720	0.0800	0.1500	0.1500	0.1220	0.1180	0.1240
	2000	0.0780	0.0740	0.0660	0.0620	0.0600	0.1260	0.1180	0.1140	0.1160	0.1120
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.2320	0.2240	0.2180	0.2240	0.2160	0.3320	0.3340	0.3320	0.3520	0.3460
	500	0.3520	0.3420	0.3520	0.3540	0.3540	0.4800	0.4800	0.4740	0.4780	0.4780
	1000	0.5020	0.5060	0.5120	0.5120	0.5000	0.6340	0.6320	0.6340	0.6280	0.6300
	2000	0.6020	0.6060	0.5940	0.5920	0.5960	0.7340	0.7280	0.7360	0.7300	0.7280
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.2940	0.2880	0.3220	0.3180	0.3080	0.4360	0.4320	0.4380	0.4400	0.4340
	500	0.5520	0.5500	0.5620	0.5720	0.5680	0.6880	0.6940	0.6880	0.6900	0.6920
	1000	0.7720	0.7720	0.7820	0.7800	0.7740	0.8600	0.8560	0.8640	0.8600	0.8640
	2000	0.9780	0.9720	0.9780	0.9740	0.9760	0.9900	0.9900	0.9920	0.9920	0.9920

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (11) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .

**Table 4.** Observed size and power of the TE-based test for the bilinear process in Equation (12).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	$n$	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0460	0.0420	0.0420	0.0420	0.0460	0.0820	0.0780	0.0900	0.0860	0.0900
	500	0.0540	0.0480	0.0480	0.0500	0.0480	0.0980	0.0940	0.1040	0.1040	0.1040
	1000	0.0440	0.0440	0.0440	0.0420	0.0500	0.1020	0.1020	0.1060	0.1000	0.1040
	2000	0.0460	0.0480	0.0480	0.0480	0.0520	0.0940	0.0940	0.1000	0.1000	0.1020
	$n$	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.0440	0.0400	0.0460	0.0460	0.0460	0.0940	0.0860	0.0940	0.0900	0.0940
	500	0.0500	0.0480	0.0440	0.0420	0.0420	0.0840	0.0880	0.0820	0.0820	0.0820
	1000	0.0660	0.0620	0.0540	0.0520	0.0560	0.1220	0.1200	0.1180	0.1140	0.1100
	2000	0.0360	0.0380	0.0400	0.0340	0.0360	0.0880	0.0920	0.1000	0.0980	0.0960
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	$n$	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.1800	0.1800	0.2040	0.2000	0.1960	0.3000	0.2900	0.3020	0.3000	0.3060
	500	0.4620	0.4580	0.4680	0.4460	0.4640	0.5920	0.5940	0.6020	0.6060	0.6100
	1000	0.7700	0.7800	0.7780	0.7820	0.7820	0.8620	0.8560	0.8640	0.8580	0.8600
	2000	0.9780	0.9820	0.9800	0.9800	0.9780	0.9860	0.9880	0.9880	0.9880	0.9880
	$n$	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.1440	0.1380	0.1540	0.1440	0.1460	0.2340	0.2320	0.2540	0.2420	0.2440
	500	0.3240	0.3200	0.3320	0.3320	0.3320	0.4620	0.4680	0.4680	0.4700	0.4700
	1000	0.6400	0.6360	0.6400	0.6240	0.6260	0.7520	0.7480	0.7460	0.7500	0.7460
	2000	0.9620	0.9580	0.9620	0.9600	0.9640	0.9880	0.9900	0.9860	0.9900	0.9840

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (12) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .

**Table 5.** Observed size and power of the TE-based test for the AR(2)-GARCH process in Equation (13).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0600	0.0540	0.0840	0.0780	0.0740	0.1280	0.1280	0.1520	0.1500	0.1520
	500	0.0580	0.0560	0.0740	0.0720	0.0700	0.1100	0.1060	0.1380	0.1320	0.1300
	1000	0.0420	0.0480	0.0440	0.0540	0.0540	0.0900	0.0840	0.0920	0.0920	0.0900
	2000	0.0480	0.0440	0.0320	0.0340	0.0320	0.0800	0.0800	0.0660	0.0680	0.0660
C = 8	200	0.0680	0.0620	0.0920	0.0880	0.0860	0.1280	0.1200	0.1440	0.1400	0.1340
	500	0.0880	0.0880	0.1120	0.1060	0.1040	0.1400	0.1400	0.1480	0.1520	0.1520
	1000	0.0600	0.0620	0.0660	0.0680	0.0680	0.1120	0.1140	0.1080	0.1160	0.1140
	2000	0.0820	0.0800	0.0800	0.0840	0.0760	0.1380	0.1380	0.1360	0.1300	0.1300
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0860	0.0800	0.0860	0.0840	0.0840	0.1420	0.1460	0.1500	0.1440	0.1400
	500	0.0980	0.0920	0.0980	0.0840	0.0920	0.1700	0.1700	0.1500	0.1480	0.1580
	1000	0.0760	0.0800	0.0520	0.0540	0.0620	0.1500	0.1480	0.1220	0.1160	0.1080
	2000	0.0540	0.0480	0.0280	0.0260	0.0260	0.1080	0.1200	0.0460	0.0460	0.0500
C = 8	200	0.1120	0.1020	0.1320	0.1300	0.1360	0.1900	0.1860	0.2200	0.2120	0.2140
	500	0.1540	0.1560	0.1740	0.1660	0.1680	0.2560	0.2480	0.2680	0.2640	0.2560
	1000	0.2180	0.2100	0.2080	0.2000	0.2020	0.3080	0.3140	0.3020	0.2960	0.3020
	2000	0.2740	0.2820	0.2540	0.2500	0.2560	0.3900	0.3900	0.3420	0.3460	0.3540

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (13) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .



**Table 6.** Observed size and power of the TE-based test for the ARMA-GARCH process in Equation (14).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0580	0.0480	0.0600	0.0560	0.0600	0.1080	0.0980	0.1160	0.1160	0.1080
	500	0.0640	0.0640	0.0660	0.0640	0.0700	0.1060	0.0960	0.1120	0.1100	0.1080
	1000	0.0540	0.0500	0.0460	0.0420	0.0440	0.0940	0.0980	0.0760	0.0800	0.0820
	2000	0.0240	0.0260	0.0160	0.0180	0.0200	0.0660	0.0720	0.0280	0.0300	0.0280
C = 8	200	0.0540	0.0380	0.0780	0.0760	0.0700	0.1160	0.1000	0.1520	0.1480	0.1520
	500	0.0900	0.0820	0.1020	0.1060	0.0940	0.1480	0.1240	0.1600	0.1520	0.1520
	1000	0.0540	0.0580	0.0600	0.0620	0.0660	0.1220	0.1140	0.1220	0.1220	0.1260
	2000	0.0620	0.0640	0.0640	0.0660	0.0680	0.1220	0.1220	0.1180	0.1140	0.1080
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.2680	0.2620	0.2560	0.2600	0.2640	0.3840	0.3820	0.3800	0.3720	0.3760
	500	0.3540	0.3460	0.3540	0.3600	0.3580	0.4580	0.4500	0.4600	0.4640	0.4520
	1000	0.3380	0.3420	0.3280	0.3240	0.3260	0.4200	0.4260	0.3880	0.3860	0.3900
	2000	0.2740	0.2800	0.2300	0.2320	0.2320	0.3360	0.3380	0.2820	0.2820	0.2880
C = 8	200	0.3860	0.3800	0.4140	0.4060	0.4100	0.5180	0.5000	0.5320	0.5260	0.5180
	500	0.7260	0.7280	0.7220	0.7200	0.7180	0.8060	0.8020	0.8040	0.8000	0.7980
	1000	0.8500	0.8440	0.8400	0.8380	0.8320	0.8780	0.8740	0.8700	0.8700	0.8700
	2000	0.8900	0.8960	0.8860	0.8880	0.8900	0.9240	0.9200	0.9160	0.9160	0.9140

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (14) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .

**Table 7.** Observed size and power of the TE-based test for the AR(1)-EGARCH process in Equation (15).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0420	0.0420	0.0740	0.0740	0.0740	0.1000	0.0940	0.1560	0.1500	0.1480
	500	0.0640	0.0520	0.0820	0.0860	0.0860	0.1040	0.0940	0.1460	0.1460	0.1460
	1000	0.0400	0.0360	0.0460	0.0440	0.0480	0.0760	0.0740	0.0880	0.0880	0.0900
	2000	0.0360	0.0400	0.0320	0.0380	0.0360	0.0680	0.0680	0.0520	0.0520	0.0580
C = 8	200	0.0540	0.0480	0.0860	0.0880	0.0860	0.1140	0.1060	0.1460	0.1440	0.1460
	500	0.0900	0.0740	0.1020	0.1020	0.1020	0.1440	0.1340	0.1760	0.1660	0.1580
	1000	0.0720	0.0640	0.0800	0.0800	0.0800	0.1240	0.1200	0.1380	0.1320	0.1340
	2000	0.0660	0.0640	0.0680	0.0700	0.0720	0.1020	0.1080	0.1100	0.1100	0.1160
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.1660	0.1460	0.2200	0.2220	0.2280	0.2580	0.2520	0.3280	0.3280	0.3260
	500	0.2500	0.2420	0.2940	0.2960	0.2960	0.3740	0.3720	0.4140	0.4080	0.4140
	1000	0.2860	0.2860	0.3040	0.3100	0.3000	0.3840	0.3820	0.4000	0.4020	0.3960
	2000	0.3180	0.3020	0.2900	0.2900	0.2900	0.3780	0.3840	0.3520	0.3440	0.3600
C = 8	200	0.2160	0.2000	0.2740	0.2660	0.2640	0.3000	0.2880	0.3560	0.3580	0.3600
	500	0.3660	0.3520	0.4040	0.4080	0.4020	0.4900	0.4720	0.5720	0.5600	0.5580
	1000	0.6200	0.6020	0.6600	0.6520	0.6520	0.7280	0.7140	0.7420	0.7440	0.7460
	2000	0.8300	0.8360	0.8460	0.8420	0.8380	0.8840	0.8820	0.8880	0.8860	0.8920

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (15) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .

**Table 8.** Observed size and power of the TE-based test for the VECM process in Equation (16).

		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.0060	0.0060	0.0040	0.0040	0.0040	0.0220	0.0200	0.0220	0.0220	0.0200
	500	0.0220	0.0220	0.0240	0.0240	0.0220	0.0460	0.0480	0.0500	0.0480	0.0560
	1000	0.0400	0.0400	0.0400	0.0400	0.0420	0.0840	0.0860	0.0900	0.0840	0.0860
	2000	0.0440	0.0420	0.0380	0.0360	0.0360	0.0800	0.0780	0.0820	0.0740	0.0700
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.0080	0.0100	0.0100	0.0080	0.0100	0.0340	0.0300	0.0300	0.0280	0.0340
	500	0.0240	0.0220	0.0220	0.0200	0.0240	0.0680	0.0660	0.0680	0.0680	0.0660
	1000	0.0480	0.0420	0.0420	0.0440	0.0440	0.0860	0.0840	0.0840	0.0820	0.0880
	2000	0.0200	0.0220	0.0200	0.0200	0.0220	0.0560	0.0540	0.0520	0.0560	0.0540
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.2540	0.2520	0.2600	0.2600	0.2640	0.3260	0.3300	0.3400	0.3380	0.3440
	500	0.5140	0.5120	0.5220	0.5280	0.5120	0.6120	0.5920	0.5980	0.6060	0.6080
	1000	0.7840	0.7840	0.7800	0.7840	0.7800	0.8340	0.8320	0.8420	0.8440	0.8420
	2000	0.9240	0.9260	0.9280	0.9260	0.9280	0.9560	0.9560	0.9520	0.9540	0.9520
		$\alpha = 0.05$					$\alpha = 0.10$				
	<i>n</i>	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.2500	0.2440	0.2500	0.2560	0.2480	0.3100	0.3000	0.3140	0.3160	0.3140
	500	0.4600	0.4640	0.4680	0.4700	0.4740	0.5600	0.5560	0.5540	0.5600	0.5580
	1000	0.7620	0.7660	0.7780	0.7680	0.7760	0.8220	0.8180	0.8240	0.8240	0.8280
	2000	0.9520	0.9560	0.9560	0.9540	0.9520	0.9780	0.9700	0.9780	0.9780	0.9780

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (16) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .

**Table 9.** Observed size and power of the TE-based test for the threshold AR(1) process in Equation (17).

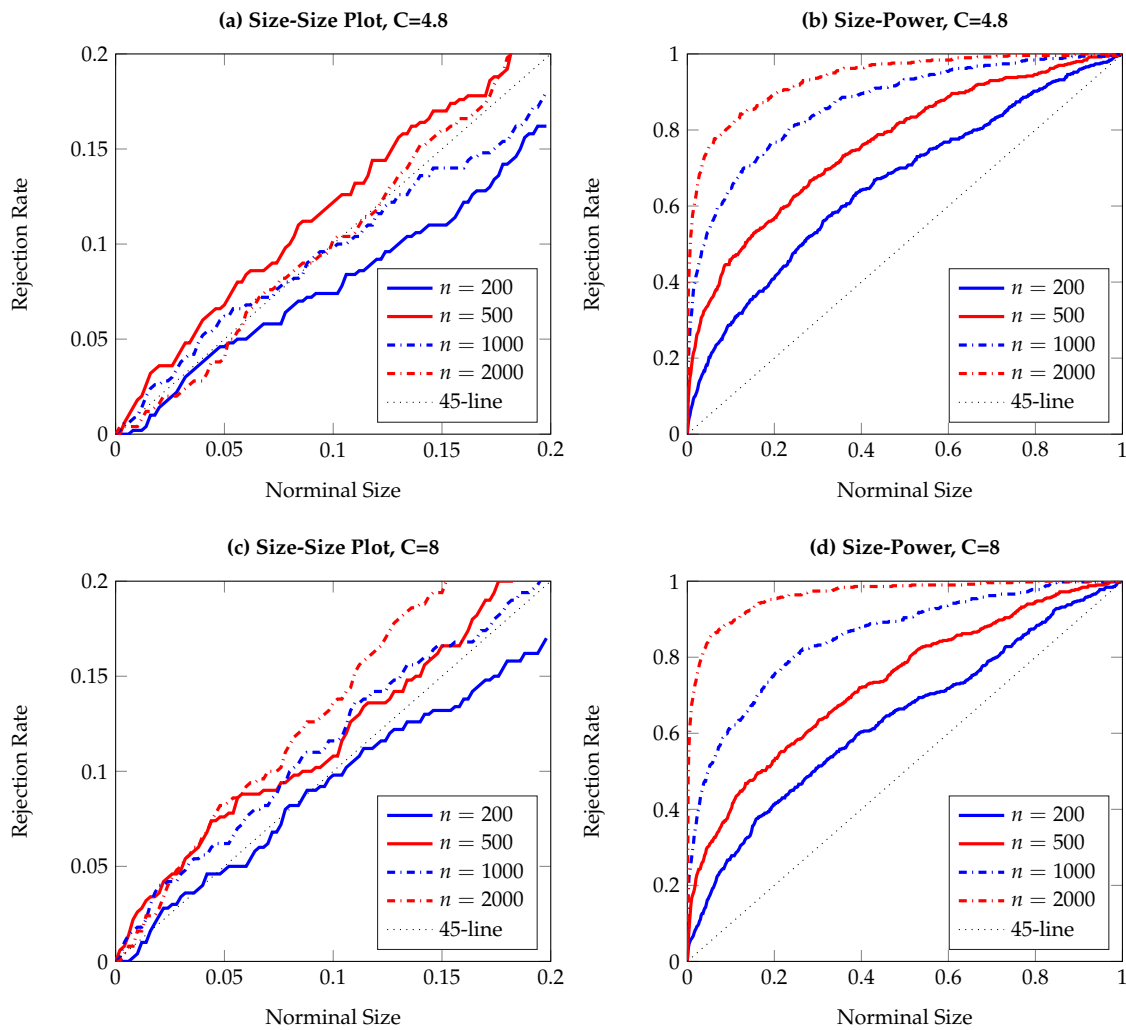
		Size									
		$\alpha = 0.05$					$\alpha = 0.10$				
	$n$	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
$C = 4.8$	200	0.0420	0.0360	0.0420	0.0360	0.0360	0.0780	0.0680	0.0720	0.0700	0.0720
	500	0.0460	0.0420	0.0440	0.0460	0.0480	0.0880	0.0880	0.0860	0.0900	0.0940
	1000	0.0380	0.0420	0.0420	0.0400	0.0360	0.0980	0.0900	0.0760	0.0840	0.0800
	2000	0.0440	0.0480	0.0420	0.0400	0.0400	0.0860	0.0900	0.0780	0.0800	0.0780
$C = 8$	200	0.0400	0.0400	0.0380	0.0400	0.0400	0.0800	0.0800	0.0920	0.0800	0.0860
	500	0.0420	0.0380	0.0440	0.0560	0.0560	0.1080	0.1020	0.1020	0.1040	0.1060
	1000	0.0480	0.0440	0.0440	0.0440	0.0500	0.1020	0.0980	0.0960	0.1080	0.1000
	2000	0.0440	0.0460	0.0480	0.0440	0.0460	0.0940	0.0860	0.0840	0.0880	0.0920
		Power									
		$\alpha = 0.05$					$\alpha = 0.10$				
	$n$	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
$C = 4.8$	200	0.8520	0.8460	0.8200	0.8280	0.8120	0.9080	0.9120	0.8940	0.8960	0.8920
	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$C = 8$	200	0.3900	0.3740	0.3120	0.3160	0.3180	0.4980	0.5000	0.4420	0.4460	0.4440
	500	0.9460	0.9440	0.9220	0.9240	0.9200	0.9760	0.9760	0.9600	0.9640	0.9660
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Note: Empirical size and power of the TE-based test at 5% and 10% significance levels for process Equation (17) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $a = 0.4$  for size evaluation and  $a = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .

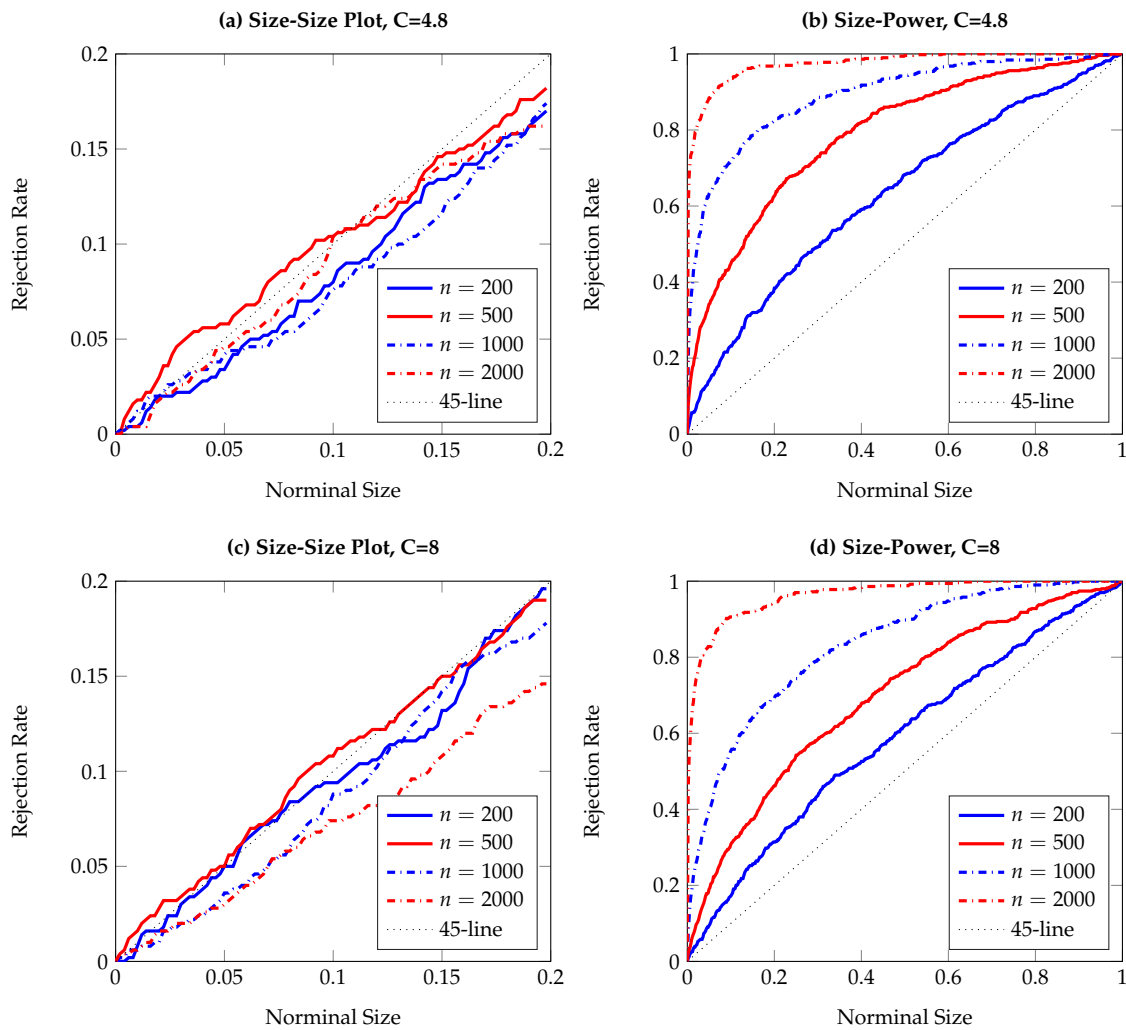
**Table 10.** Observed Power of the TE-based test for the two-Way VAR process in Equation (18).

		X → Y									
		α = 0.05					α = 0.10				
	n	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.2780	0.2600	0.3540	0.3520	0.3560	0.3980	0.3760	0.4560	0.4560	0.4660
	500	0.5640	0.5360	0.6260	0.6240	0.6260	0.6780	0.6680	0.7420	0.7400	0.7460
	1000	0.8380	0.8320	0.8800	0.8740	0.8780	0.9040	0.8980	0.9200	0.9240	0.9200
	2000	0.9900	0.9900	0.9980	0.9960	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		α = 0.05					α = 0.10				
	n	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.2480	0.2240	0.3080	0.3180	0.3100	0.3540	0.3360	0.3900	0.3940	0.3920
	500	0.4180	0.4020	0.4740	0.4700	0.4700	0.5400	0.5320	0.5880	0.5920	0.5800
	1000	0.7960	0.7840	0.8140	0.8140	0.8160	0.8560	0.8540	0.8740	0.8720	0.8720
	2000	0.9800	0.9780	0.9820	0.9800	0.9800	0.9900	0.9880	0.9900	0.9900	0.9920
		Y → X									
		α = 0.05					α = 0.10				
	n	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 4.8	200	0.5140	0.5060	0.5480	0.5560	0.5480	0.6120	0.5960	0.6920	0.6920	0.6940
	500	0.9400	0.9340	0.9480	0.9440	0.9460	0.9620	0.9580	0.9720	0.9700	0.9740
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		α = 0.05					α = 0.10				
	n	TS.a	TS.b	SMB.a	SMB.b	STB	TS.a	TS.b	SMB.a	SMB.b	STB
C = 8	200	0.4040	0.3700	0.4420	0.4400	0.4280	0.5060	0.4900	0.5400	0.5400	0.5320
	500	0.8260	0.8220	0.8220	0.8200	0.8160	0.8740	0.8740	0.8700	0.8720	0.8680
	1000	0.9820	0.9840	0.9820	0.9800	0.9800	0.9940	0.9940	0.9940	0.9920	0.9940
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

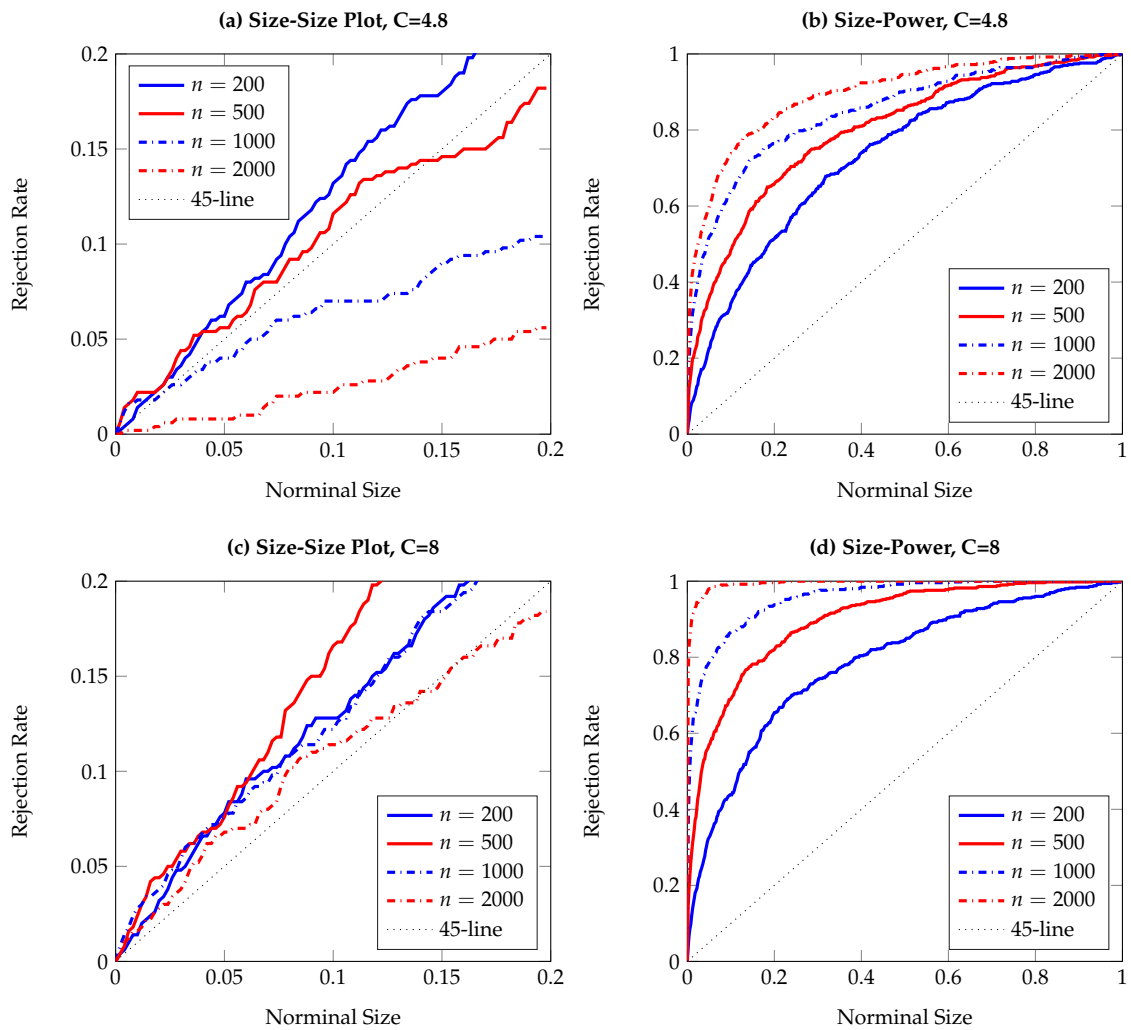
Note: Empirical power of the TE-based test at 5% and 10% significance levels for process Equation (18) for different resampling methods. The values represent observed rejection rates over 500 realizations for nominal size 0.05. Sample sizes go from 200 to 2000. The control parameter  $b = -0.2$  and  $c = 0.1$  for establishing powers. For this simulation study, we consider  $C = 4.8$  and  $C = 8$ .



**Figure 1.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The data generating process (DGP) is the bivariate VAR process in Equation (9), with  $Y$  affecting  $X$ . The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .

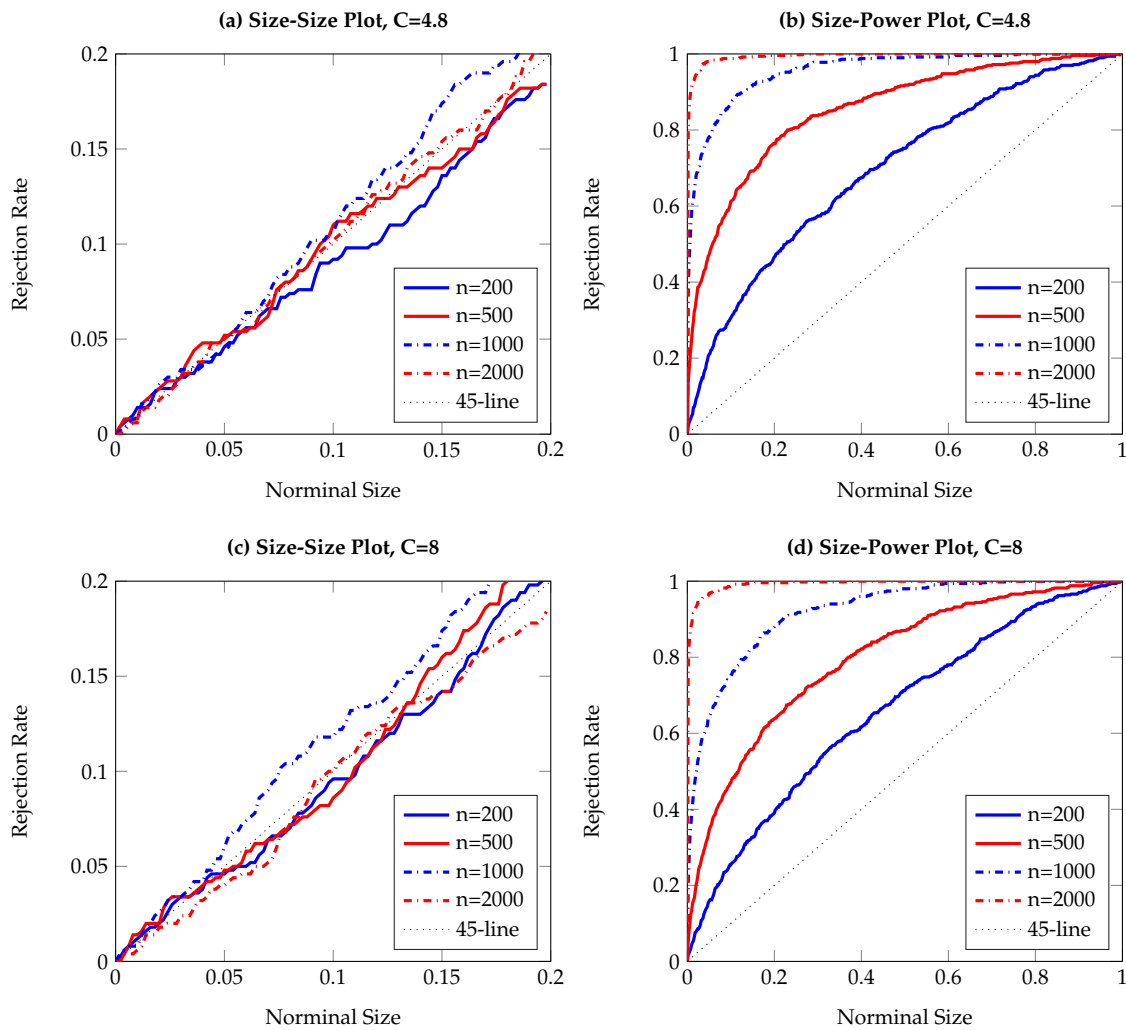


**Figure 2.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the bivariate non-linear VAR process in Equation (10), with  $Y$  affecting  $X$ . The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .

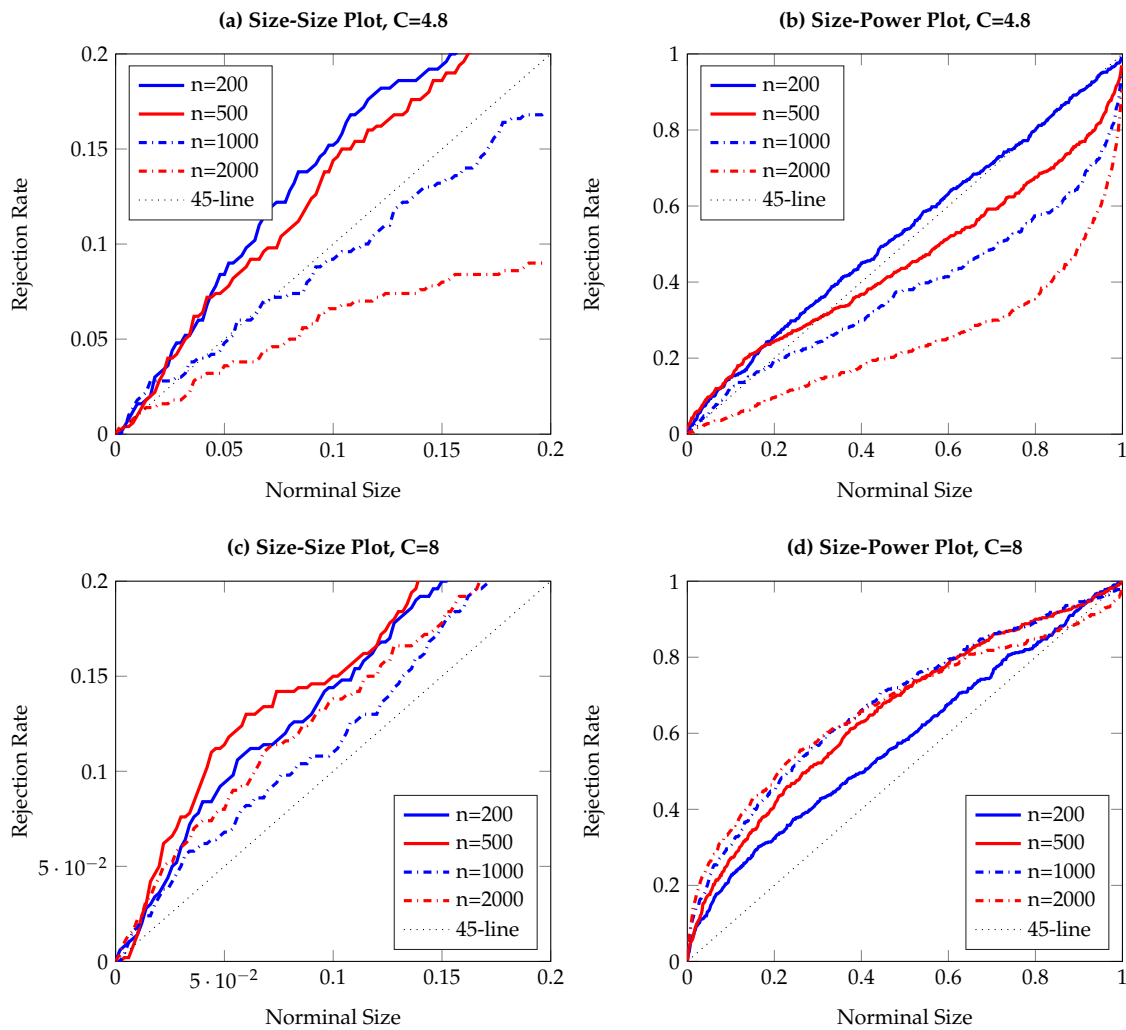


**Figure 3.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the bivariate ARCH process in Equation (11), with Y affecting X. The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .

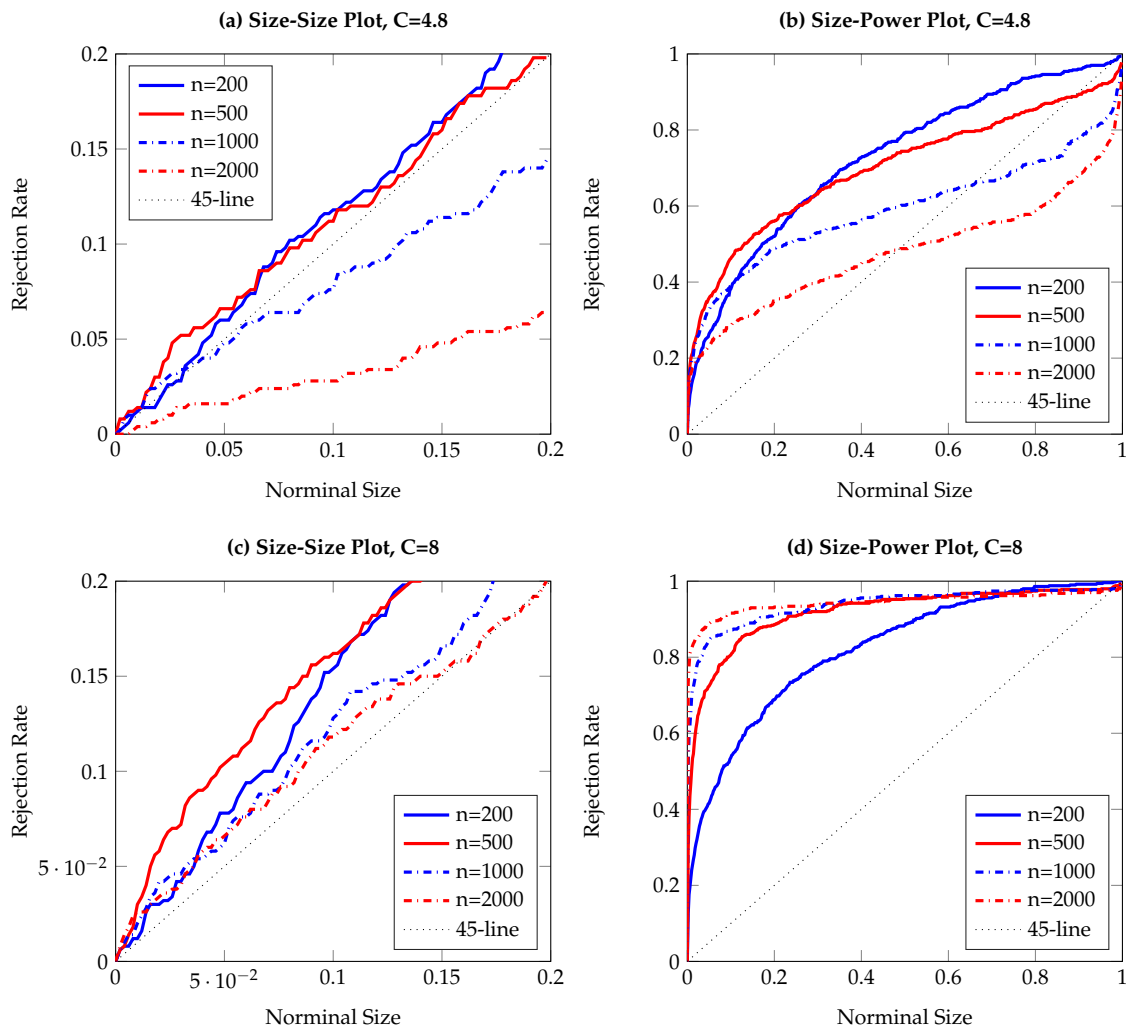




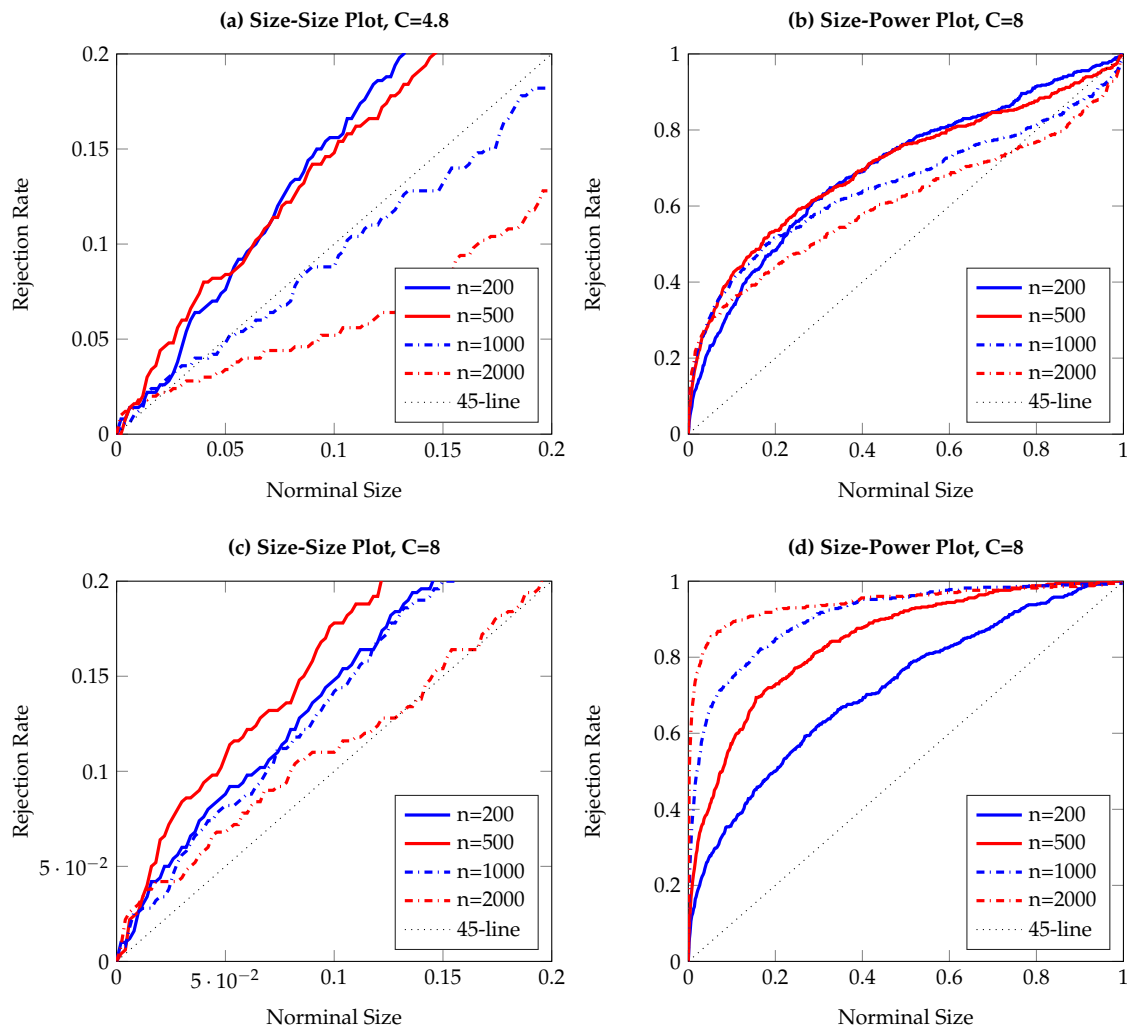
**Figure 4.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the bilinear process in Equation (12), with  $Y$  affecting  $X$ . The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .



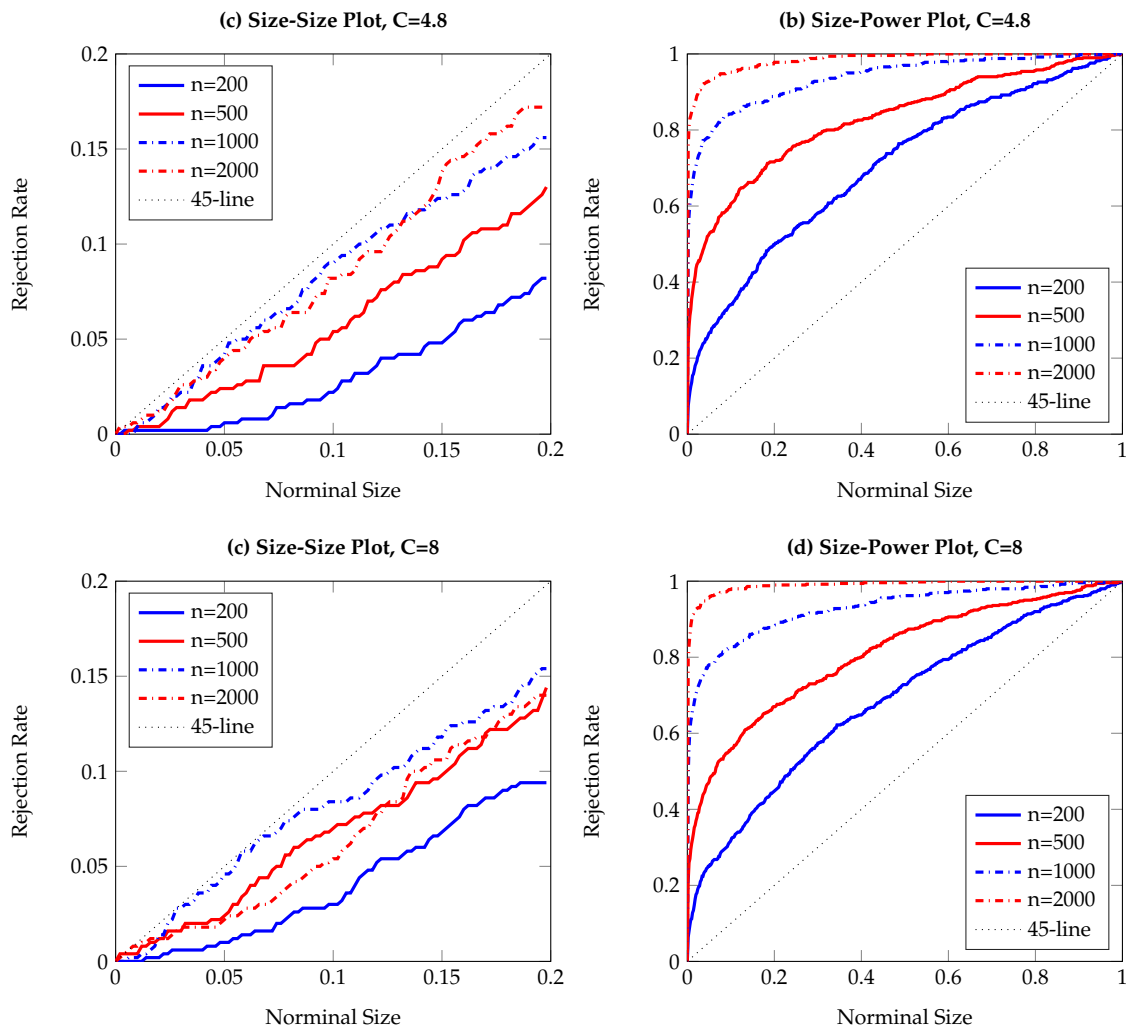
**Figure 5.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the bivariate AR2-GARCH process in Equation (13), with Y affecting X. The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .



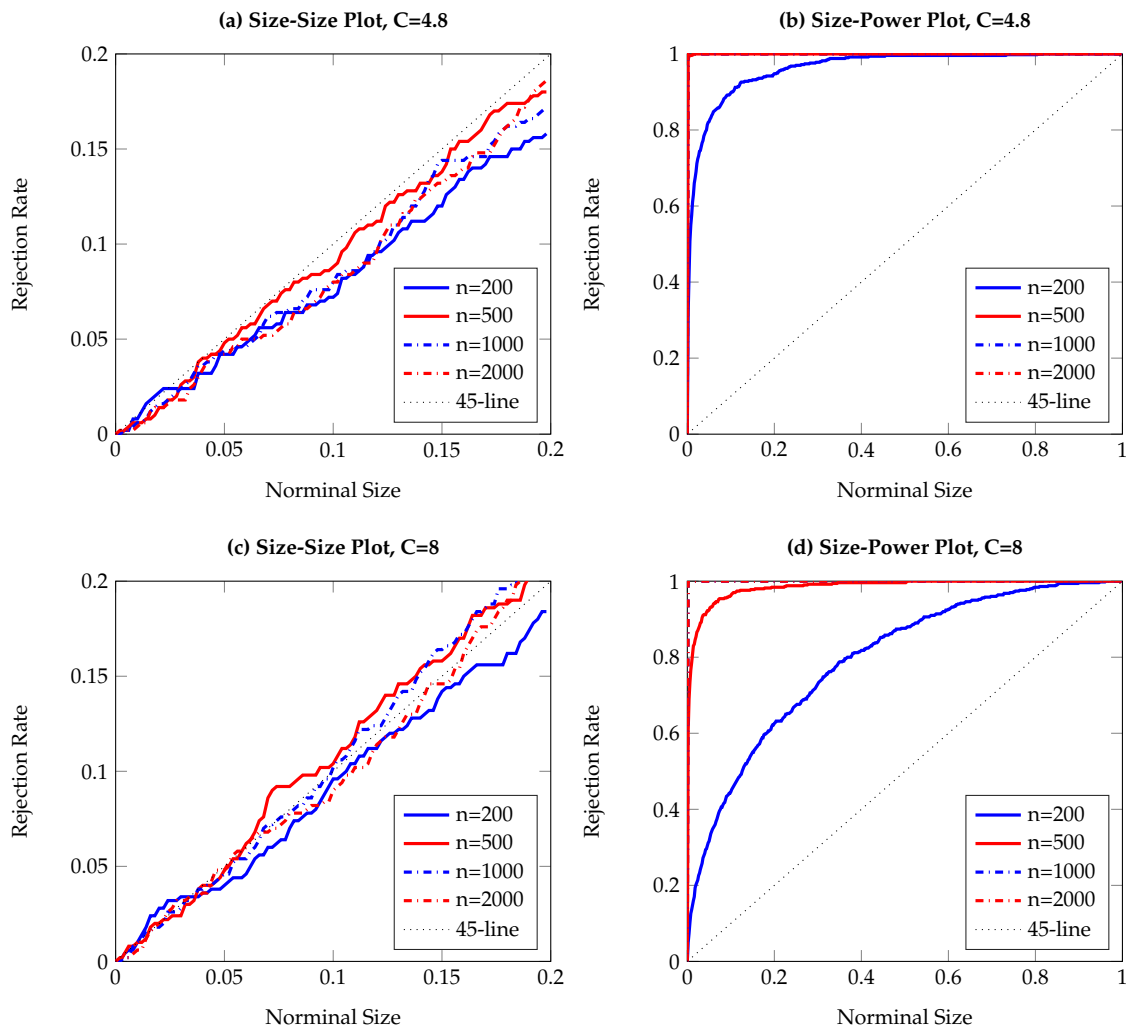
**Figure 6.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the bivariate ARMA-GARCH process in Equation (14), with  $Y$  affecting  $X$ . The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .



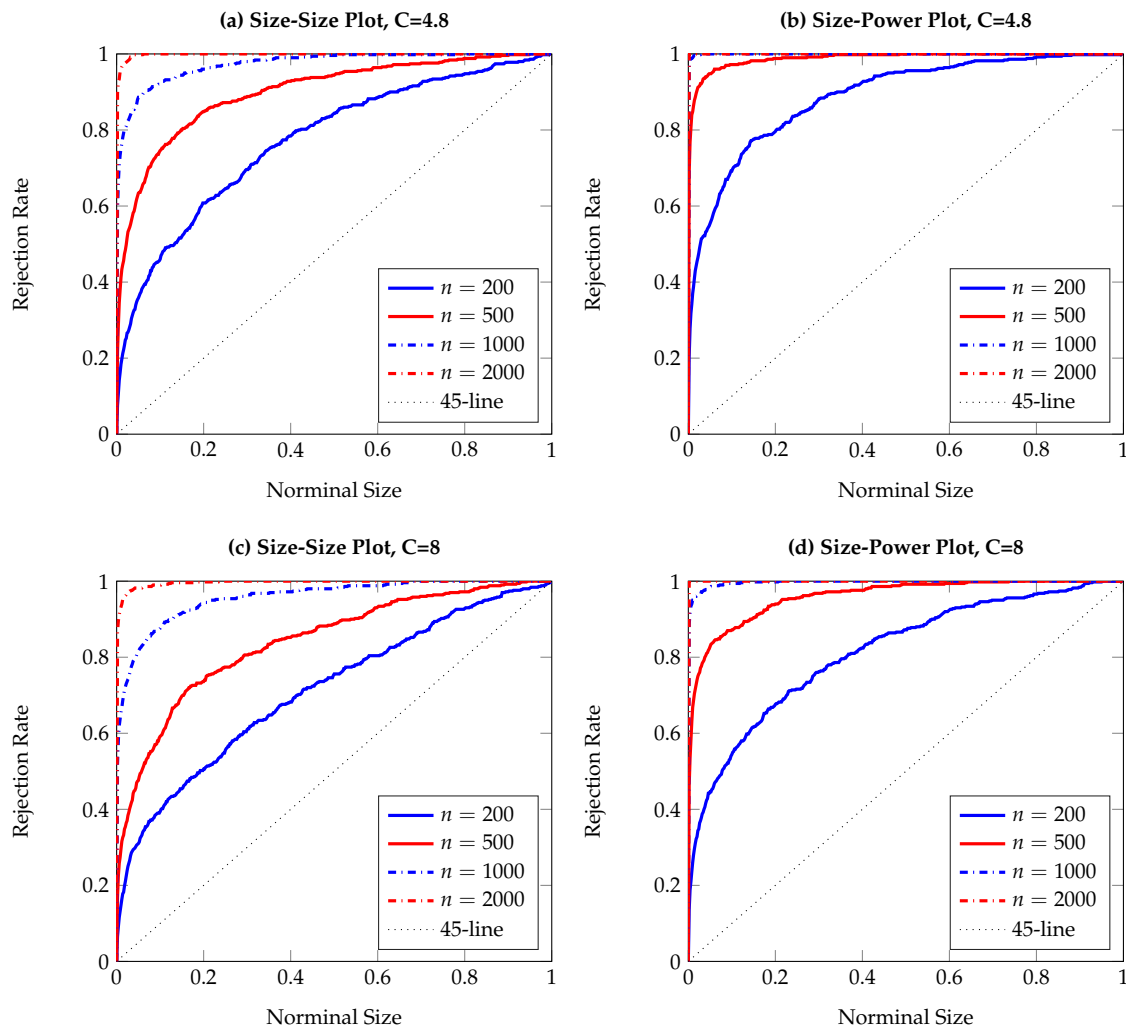
**Figure 7.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the bivariate AR1-EGARCH process in Equation (15), with Y affecting X. The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .



**Figure 8.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the VECM process in Equation (16), with  $Y$  affecting  $X$ . The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .



**Figure 9.** Size-size and size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the bivariate threshold AR(1) process in Equation (17), with Y affecting X. The left (right) column shows observed rejection rates under the null (alternative) hypothesis. The sample size varies from  $n = 200$  to  $n = 2000$ .



**Figure 10.** Size-power plots of Granger non-causality tests, based on 500 replications and smoothed local bootstrap (a). The DGP is the two-way VAR process in Equation (18), with  $X$  affecting  $Y$  and  $Y$  affecting  $X$ . The left (right) column shows observed rejection rates for testing  $X$  ( $Y$ ) causing  $Y$  ( $X$ ). The sample size varies from  $n = 200$  to  $n = 2000$ .

#### 4. Application

In this section, we apply the TE-based nonparametric test on detecting financial market interdependence, in terms of both return and volatility. Diebold and Yilmaz [49] performed a variance decomposition of the covariance matrix of the error terms from a reduced-form VAR model to investigate the spillover effect in the global equity market. More recently, Gamba-Santamaria et al. [50] extended the framework and considered the time-varying feature in global volatility spillovers. Their research, although providing simple and intuitive methods for measuring directional linkages between global stock markets, may suffer from the limitation of the linear parametric modeling, as discussed above. We revisit the topic of spillovers in the global equity market by the nonparametric method.

For our analysis, we use daily nominal stock market indexes from January 1992 to March 2017, obtained from Datastream, for six developed countries including the US (DJIA), Japan (Nikkei 225), Hong Kong (Hangseng), the UK (FTSE 100), Germany (DAX 30) and France (CAC 40). The target series are weekly return and volatility for each index. The weekly returns are calculated in terms of differenced log prices multiplied by 100, from Friday to Friday. Where the price for Friday is not available due to a public holiday, we use the Thursday price instead.

The weekly volatility series are generated following [49] by making use of the weekly high, low, opening and closing prices, obtained from the underlying daily high, low, opening and closing data. The volatility  $\sigma_t^2$  for week  $t$  is estimated as

$$\hat{\sigma}_t^2 = 0.511(H_t - L_t)^2 - 0.019[(C_t - O_t)(H_t + L_t - 2O_t) - 2(H_t - O_t)(L_t - O_t)] - 0.383(C_t - O_t)^2, \quad (19)$$

where  $H_t$  is the Monday-Friday high,  $L_t$  is the Monday-Friday low,  $O_t$  is the Monday-Friday open and  $C_t$  is the Monday-Friday close (in natural logarithms multiplied by 100). Further, after deleting the volatility estimates for the New Year week in 2002, 2008 and 2013 due to the lack of observations for Nikkei 225 index, we have 1313 observations in total for weekly returns volatilities. The descriptive statistics, Ljung Box (LB) test statistics and Augmented Dickey Fuller (ADF) test statistics for both series are summarized in Table 11. From the ADF test results, it is clear that all time series are stationary for further analysis (we also performed the Johansen cointegration test pair-wisely on the price levels and no cointegration was found for the six market indexes.).

**Table 11.** Descriptive statistics for global stock market return and volatility.

Return						
	DJIA	Nikkei	Hangseng	FTSE	DAX	CAC
<b>Mean</b>	0.1433	−0.0126	0.1294	0.0823	0.1535	0.0790
<b>Median</b>	0.2911	0.1472	0.2627	0.2121	0.4029	0.1984
<b>Maximum</b>	10.6977	11.4496	13.9169	12.5845	14.9421	12.4321
<b>Minimum</b>	−20.0298	−27.8844	−19.9212	−23.6317	−24.3470	−25.0504
<b>Std. Dev.</b>	2.2321	3.0521	3.3819	2.3367	3.0972	2.9376
<b>Skewness</b>	−0.8851	−0.6978	−0.3773	−0.8643	−0.6398	−0.6803
<b>Kurtosis</b>	10.8430	8.9250	5.9522	13.2777	7.9186	8.0780
<b>LB Test</b>	49.9368 **	15.4577	28.7922	61.0916 **	28.0474	43.5004 **
<b>ADF Test</b>	−38.9512 **	−37.1989 **	−35.4160 **	−38.8850 **	−37.2015 **	−38.7114 **
Volatility						
	DJIA	Nikkei	Hangseng	FTSE	DAX	CAC
<b>Mean</b>	4.5983	7.7167	9.6164	5.5306	8.5698	8.2629
<b>Median</b>	2.2155	4.6208	4.5827	2.7161	4.0596	4.7122
<b>Maximum</b>	208.2227	265.9300	379.4385	149.1572	175.0968	179.8414
<b>Minimum</b>	0.0636	0.1882	0.1554	0.1154	0.1263	0.2904
<b>Std. Dev.</b>	9.9961	13.5154	21.3838	10.1167	15.2845	12.7872
<b>Skewness</b>	10.9980	9.6361	10.2868	6.7179	5.3602	6.0357
<b>Kurtosis</b>	180.0844	140.7606	152.4263	67.0128	42.0810	58.3263
<b>LB Test</b>	1924.0870 **	933.3972 **	1198.6872 **	1970.7366 **	2770.5973 **	1982.4141 **
<b>ADF Test</b>	−16.0378 **	−17.9044 **	−17.4896 **	−14.0329 **	−14.1928 **	−13.6136 **

Note: Descriptive statistics for six globally leading indexes. The sample size is 1313 for both Returns and Volatilities. The nominal returns are measured by weekly Friday-to-Friday log price difference multiplied by 100 and the Monday-to-Friday volatilities are calculated following [49]. For the LB test and the ADF test statistics, the asterisks indicate the significance of the corresponding  $p$ -value at 1% (\*\*) levels.

We firstly provide a full-sample analysis of global stock market return and volatility spillovers over the period from January 1992 to March 2017, summarized in Tables 12 and 13. The two tables report the pairwise test statistics for conditional independence between index  $X$  and index  $Y$ , given the constant  $C$  in the bandwidth for kernel estimation is 4.8 or 8 and 999 resampling time series. In other words, we test for the absence of the one-week-ahead directional linkage from index  $X$  to  $Y$  by using the five resampling methods described in Section 2. For example, the first line in the top panel in Table 12 reports the one-week-ahead influence of DJIA returns upon other indexes by using the first time-shifted surrogates method (TS.a). Given  $C = 8$ , DJIA return is shown to be a strong Granger cause for Nikkei, FTSE and CAC at the 1% level, and for DAX at the 5% level.



**Table 12.** Detection of Conditional Dependence in Global Stock Returns.

		To											
		DJIA		Nikkei		Hangseng		FTSE		DAX		CAC	
From		C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8
		TS.a	DJIA	-		0.367	0.002**	0.972	0.273	0.609	0.009**	0.533	0.012*
Nikkei	0.231		0.081	-		0.997	0.951	0.883	0.059	0.868	0.242	0.004**	0.001**
Hangseng	0.898		0.197	0.963	0.407	-		0.701	0.035*	0.969	0.174	0.640	0.005**
FTSE	0.483		0.004**	0.004**	0.001**	0.419	0.001**	-		0.839	0.185	0.917	0.615
DAX	0.977		0.149	0.027*	0.001**	0.009**	0.001**	0.004**	0.001**	-		0.695	0.025*
CAC	0.995		0.713	0.001**	0.001**	0.294	0.001**	0.918	0.741	0.639	0.203	-	
TS.b	DJIA	-		0.402	0.004**	0.967	0.341	0.595	0.020*	0.569	0.049*	0.288	0.012*
	Nikkei	0.219	0.110	-		0.999	0.956	0.853	0.103	0.855	0.255	0.009**	0.006**
	Hangseng	0.899	0.269	0.975	0.485	-		0.696	0.088	0.971	0.231	0.664	0.023*
	FTSE	0.477	0.016*	0.006**	0.003**	0.404	0.018*	-		0.847	0.245	0.896	0.642
	DAX	0.976	0.221	0.027*	0.002**	0.009**	0.005**	0.011*	0.010**	-		0.692	0.065
	CAC	0.993	0.729	0.002**	0.002**	0.346	0.016*	0.907	0.763	0.650	0.244	-	
SMB.a	DJIA	-		0.564	0.001**	0.999	0.349	0.796	0.003**	0.817	0.014*	0.425	0.001**
	Nikkei	0.321	0.147	-		0.988	0.946	0.957	0.085	0.944	0.321	0.006**	0.002**
	Hangseng	0.946	0.273	0.967	0.483	-		0.860	0.016*	0.994	0.188	0.793	0.004**
	FTSE	0.579	0.005**	0.021*	0.001**	0.701	0.003**	-		0.947	0.297	0.943	0.739
	DAX	0.988	0.240	0.044*	0.001**	0.012*	0.001**	0.006**	0.001**	-		0.788	0.020*
	CAC	0.993	0.762	0.006**	0.001**	0.594	0.001**	0.946	0.861	0.842	0.270	-	
SMB.b	DJIA	-		0.583	0.002**	0.994	0.334	0.797	0.001**	0.855	0.014*	0.413	0.001**
	Nikkei	0.351	0.155	-		0.992	0.940	0.952	0.088	0.965	0.322	0.002**	0.003**
	Hangseng	0.945	0.276	0.970	0.506	-		0.866	0.015*	0.997	0.215	0.829	0.004**
	FTSE	0.637	0.006**	0.008**	0.001**	0.714	0.003**	-		0.954	0.345	0.953	0.789
	DAX	0.980	0.256	0.034*	0.001**	0.011*	0.001**	0.009**	0.001**	-		0.831	0.042*
	CAC	0.993	0.825	0.003**	0.001**	0.591	0.002**	0.980	0.898	0.862	0.304	-	
STB	DJIA	-		0.645	0.001**	0.996	0.334	0.787	0.002**	0.823	0.015*	0.430	0.001**
	Nikkei	0.363	0.114	-		0.989	0.944	0.946	0.079	0.955	0.296	0.003**	0.002**
	Hangseng	0.964	0.272	0.984	0.491	-		0.859	0.013*	0.987	0.197	0.786	0.004**
	FTSE	0.652	0.005**	0.016*	0.001**	0.688	0.003**	-		0.940	0.262	0.965	0.761
	DAX	0.982	0.234	0.048*	0.001**	0.017*	0.001**	0.006**	0.001**	-		0.828	0.029*
	CAC	0.996	0.814	0.007**	0.001**	0.578	0.001**	0.967	0.835	0.799	0.265	-	

Note: Statistics for pairwise TE-based test on returns of global stock-indexes for one-week ahead conditional non-independence. The results are shown both for the five different resampling methods in Section 2.3. The constant C takes value 4.8 and 8 for robustness-check. The asterisks indicate the significance of the corresponding p-value, at the 5% (\*) and 1% (\*\*) levels.

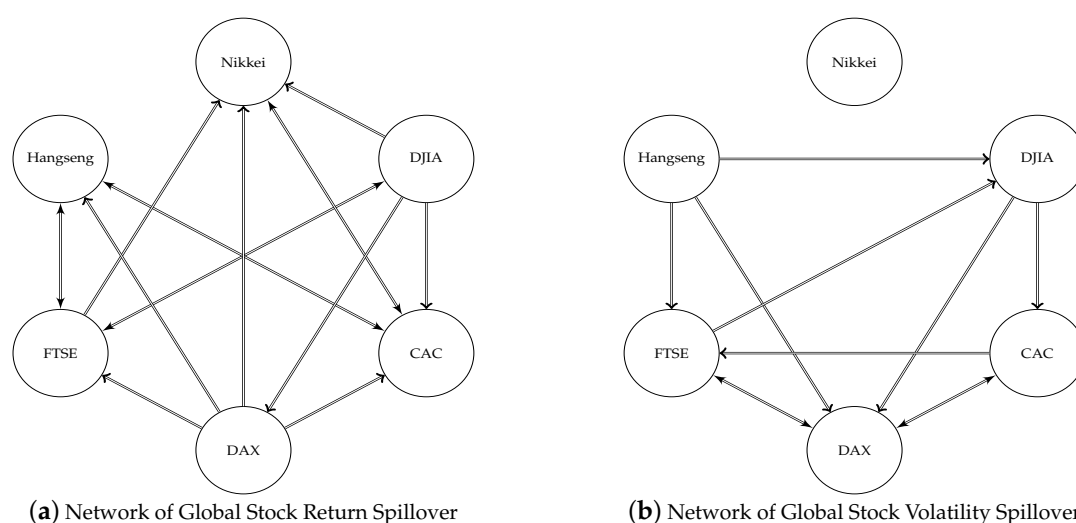
**Table 13.** Detection of Conditional Dependence in Global Stock Volatilities.

		To		DJIA		Nikkei		Hangseng		FTSE		DAX		CAC	
		From		C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8	C = 4.8	C = 8
TS.a	DJIA	-		0.997	0.998	0.998	0.994	0.974	0.853	0.828	0.001**	0.005**	0.001**		
	Nikkei	0.998	0.995	-		0.996	1.000	0.997	0.994	0.971	0.950	1.000	0.999		
	Hangseng	0.943	0.003**	0.989	0.992	-		0.822	0.001**	0.001**	0.001**	0.973	0.953		
	FTSE	0.010**	0.001**	0.955	0.944	0.997	1.000	-		0.806	0.001**	0.985	0.946		
	DAX	0.975	0.931	0.934	0.898	0.999	0.995	0.001**	0.001**	-		0.072	0.001**		
	CAC	0.996	0.944	0.988	0.987	0.999	0.997	0.003**	0.001**	0.054	0.001**	-			
TS.b	DJIA	-		0.993	0.994	0.996	0.992	0.958	0.857	0.806	0.010**	0.018*	0.005**		
	Nikkei	0.986	0.994	-		0.996	0.997	0.988	0.988	0.942	0.926	0.998	0.996		
	Hangseng	0.919	0.011*	0.976	0.966	-		0.819	0.002**	0.001**	0.001**	0.965	0.941		
	FTSE	0.019*	0.003**	0.950	0.927	0.994	0.997	-		0.785	0.002**	0.982	0.932		
	DAX	0.976	0.904	0.943	0.890	0.997	0.999	0.001**	0.001**	-		0.113	0.009**		
	CAC	0.998	0.951	0.979	0.983	0.999	0.998	0.009**	0.002**	0.077	0.003**	-			
SMB.a	DJIA	-		0.823	0.786	0.984	0.968	0.468	0.189	0.210	0.004**	0.027*	0.002**		
	Nikkei	0.957	0.941	-		1.000	1.000	0.843	0.800	0.743	0.614	0.998	0.993		
	Hangseng	0.458	0.030*	0.888	0.808	-		0.165	0.007**	0.001**	0.001**	0.839	0.736		
	FTSE	0.034*	0.001**	0.557	0.417	1.000	1.000	-		0.358	0.001**	0.967	0.793		
	DAX	0.702	0.255	0.448	0.327	1.000	1.000	0.001**	0.001**	-		0.030*	0.001**		
	CAC	0.887	0.241	0.673	0.643	1.000	1.000	0.005**	0.001**	0.012*	0.001**	-			
SMB.b	DJIA	-		0.851	0.806	0.973	0.969	0.563	0.269	0.226	0.004**	0.032*	0.001**		
	Nikkei	0.950	0.940	-		1.000	1.000	0.888	0.849	0.792	0.675	0.999	0.993		
	Hangseng	0.451	0.022*	0.919	0.869	-		0.209	0.005**	0.002**	0.001**	0.868	0.793		
	FTSE	0.023*	0.001**	0.561	0.457	1.000	1.000	-		0.434	0.001**	0.974	0.840		
	DAX	0.819	0.442	0.491	0.356	1.000	1.000	0.001**	0.001**	-		0.030*	0.001**		
	CAC	0.958	0.554	0.781	0.696	1.000	1.000	0.009**	0.001**	0.014*	0.001**	-			
STB	DJIA	-		0.847	0.813	0.972	0.964	0.602	0.278	0.221	0.009**	0.045*	0.004**		
	Nikkei	0.967	0.956	-		1.000	1.000	0.835	0.789	0.777	0.624	0.998	0.997		
	Hangseng	0.527	0.033*	0.933	0.887	-		0.222	0.010**	0.003**	0.001**	0.860	0.762		
	FTSE	0.024*	0.002**	0.635	0.497	1.000	1.000	-		0.405	0.001**	0.977	0.826		
	DAX	0.795	0.392	0.561	0.410	1.000	1.000	0.004**	0.001**	-		0.027*	0.001**		
	CAC	0.903	0.561	0.809	0.748	1.000	1.000	0.020*	0.001**	0.010**	0.001**	-			

Note: Statistics for pairwise TE-based test on volatilities of global stock-indexes for one-week ahead conditional non-independence. The results are shown both for the five different resampling methods in Section 2.3. The constant C takes value 4.8 and 8 for robustness-check. The asterisks indicate the significance of the corresponding p-values at the 5% (\*) and 1% (\*\*) levels.

Based on Tables 12 and 13, we may draw several conclusions. Firstly, the US index and German index are the most important return transmitters and Hong Kong is the largest source for volatility spillover, judged by the numbers of significant linkages. Note that this finding is similar as the result in [49], where the total return (volatility) spillovers from US (Hong Kong) to others are found to be much higher than from any other country. Figure 11 provides a graphical illustration of the global spillover network based on the result of (SMB.a) from Tables 12 and 13. Apart from the main transmitters, we can clearly see that Nikkei and CAC are the main receivers in the global return spillover network, while DAX is the main receiver of global volatility transmission.

Secondly, the result obtained is very robust, no matter which re-sampling method is applied. Although the differences between the five resampling methods are small, (TS.a) is seen to be slightly more powerful than (TS.b) in Table 12. The three different bootstrap methods are very consistent almost all the time, similarly to what we observed in Section 3.

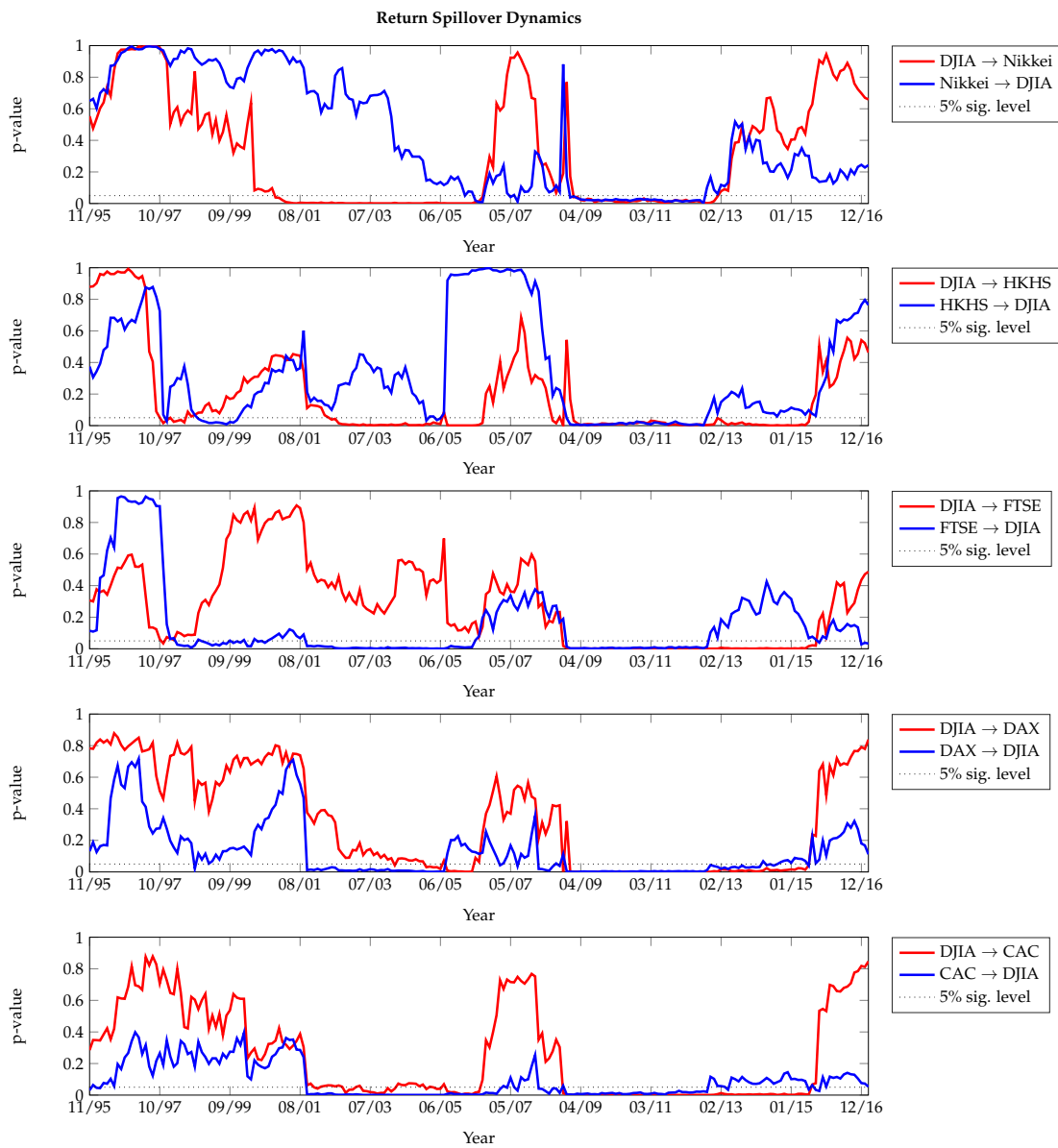


**Figure 11.** Graphical representation of pairwise causalities on global stock returns and volatilities. All " $\rightarrow$ " in the graph indicate a significant directional causality at the 5% level.

However, the summary results in Tables 12 and 13 are static in the sense that they do not take into account possible time-variation. The statistics are measurements for averaged-out directional linkages over the whole period from 1992 to 2017. The conditional dependence structure of the time series, at any point in time, can be very different. Hence, the full-sample analysis is very likely to oversee the cyclical dynamics between each pair of stock indices. To investigate the dynamics in the global stock market, we now move from the full-sample analysis to a rolling-window study. Considering a 200-week rolling window starting from the beginning of the sample and admitting a 5-week forward step for the iterative evaluation of the conditional dependence, we can assess the variation of the spillover in the global equity market over time.

Taking the return series of the DJIA as an illustration, we iteratively exploit the local smoothed bootstrap method for detecting Granger causality from and to the DJIA return in Figure 12 (all volatility series are extremely skewed, see Table 11. In a small sample analysis, the test statistics turn out to be sensitive to the clustering outliers, which typically occur during the market turmoil. As a result, the volatility dynamics are more radical and less informative than that of returns). The red line represents the  $p$ -values for the TE-based test on DJIA weekly return as the information transmitter while the blue line shows the  $p$ -values associated with testing on DJIA as a receiver of information spillover on a weekly basis from others. The plots displays an event-dependent pattern, particularly for the recent financial crisis; from early 2009 until the end of 2012, all pairwise tests show the presence

of a strong bi-directional linkage. Besides, the DJIA is strongly leading the Nikkei, Hangseng and CAC during the first decade of this century.



**Figure 12.** Time-varying  $p$ -values for the TE-based Granger causality test in Return series. The causal linkages from DJIA to other markets, as well as the linkages from other markets to DJIA are tested.

Further, we see that the influence from other indices to the DJIA are different, typically responding to economics events. For example, the blue line in the second panel of Figure 12 plunges below the 5% level twice before the recent financial crisis, meaning that the Hong Kong Hangseng index causes fluctuations in the DJIA during those two periods; first in the late 90’s and again by the end of 2004. The timing of the first fall matches the 1997 Asian currency crisis and the latter one was in fact caused by China’s austerity policy in October 2004.

Finally, the dynamic plots provide additional insights into the sample period that the full sample analysis may omit. The DAX and CAC are found to be less relevant for future fluctuations in the weekly return of the DJIA, according to Table 12 and Figure 11. However, one can clearly see that since

2001, the  $p$ -values for DAX→DJIA and CAC→DJIA are consistently below 5% for most of the time, suggesting an increase of the integration of global financial markets.

## 5. Conclusions

This paper provides guidelines for the practical application of TE in detecting conditional dependence, i.e., Granger causality in a more general sense, between two time series. Although there already is a tremendous literature that tried to apply the TE in this context, the asymptotics of the statistic and the performance of the resampling-based measures are still not understood well. We have considered tests based on five different resampling methods, all of which were shown in the literature to be suitable for entropy-related tests, and investigated the size and power of the associated tests numerically. Two time-shifted surrogates and three smoothed bootstrap methods are tested on simulated data from several processes. The simulation results in this controlled environment suggest that all five measures achieve reasonable rejection rates under the null as well as the alternative hypotheses. Our results are very robust with respect to the density estimation method, including the procedure used for standardizing the location and scale of the data and the choice of the bandwidth parameter, as long as the convergence rate of the kernel estimator of TE is consistent with its first order Taylor expansion.

In the empirical application, we have shown how the proposed resampling techniques can be used on real world data for detecting conditional dependence in the data set. We use global equity data to carry out the detection in pairwise causalities in the return and volatility series among the world leading stock indexes. Our work can be viewed as a nonparametric extension of the spillover measures considered by Diebold and Yilmaz [49]. In accordance with them, we found evidence that the DJIA and the DAX are the most important return transmitters and Hong Kong is the largest source for volatility spillover. Furthermore, the rolling window-based test for Granger causality in pairwise return series demonstrated that the causal linkages in the global equity market are time-varying rather than static. The overall dependence is more tight during the most recent financial crisis, and the fluctuations of the  $p$ -values are shown to be event dependent.

As for future work, there are several directions for potential extensions. On the theoretical side, it would be practically meaningful to consider causal linkage detection beyond the single period lag and to deal with the infinite order issue in a nonparametric setting. Further nonparametric techniques need to be developed to play a similar role as the information criterion does for order selection of an estimation model in the parametric world. On the empirical side, it will be interesting to further exploit entropy-based statistics in testing conditional dependence when there exists a so-called common factor, i.e., looking at multivariate systems with more than two variables. One potential candidate for this type of test in the partial TE has been coined by Vakorin et al. [51], but its statistical properties still need to be thoroughly studied yet.

**Acknowledgments:** We thank the editor and two anonymous referees for their help and constructive comments. We would also like to thank seminar participants at the University of Amsterdam and the Tinbergen Institute. The authors also wish to extend particular gratitude to Simon Broda and Chen Zhou for their comments on an earlier version of this paper. We also acknowledge SURFsara for providing the computational environment of the LISA cluster.

**Author Contributions:** Both authors contributed equally to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423; 623–656.
2. Shannon, C.E. Prediction and entropy of printed English. *Bell Syst. Tech. J.* **1951**, *30*, 50–64.
3. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.
4. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **1969**, *37*, 424–438.

5. Hiemstra, C.; Jones, J.D. Testing for linear and nonlinear Granger causality in the stock price-volume relation. *J. Financ.* **1994**, *49*, 1639–1664.
6. Diks, C.; Panchenko, V. A new statistic and practical guidelines for nonparametric Granger causality testing. *J. Econ. Dyn. Control* **2006**, *30*, 1647–1669.
7. Su, L.; White, H. A nonparametric Hellinger metric test for conditional independence. *Econom. Theory* **2008**, *24*, 829–864.
8. Bouezmarni, T.; Rombouts, J.V.; Taamouti, A. Nonparametric copula-based test for conditional independence with applications to Granger causality. *J. Bus. Econ. Stat.* **2012**, *30*, 275–287.
9. Su, L.; White, H. Testing conditional independence via empirical likelihood. *J. Econom.* **2014**, *182*, 27–44.
10. Hlaváčková-Schindler, K.; Paluš, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **2007**, *441*, 1–46.
11. Amblard, P.O.; Michel, O.J. The relation between Granger causality and directed information theory: A review. *Entropy* **2012**, *15*, 113–143.
12. Granger, C.; Lin, J.L. Using the mutual information coefficient to identify lags in nonlinear models. *J. Time Ser. Anal.* **1994**, *15*, 371–384.
13. Hong, Y.; White, H. Asymptotic distribution theory for nonparametric entropy measures of serial dependence. *Econometrica* **2005**, *73*, 837–901.
14. Barnett, L.; Bossomaier, T. Transfer entropy as a log-likelihood ratio. *Phys. Rev. Lett.* **2012**, *109*, 138105.
15. Efron, B. Computers and the theory of statistics: Thinking the unthinkable. *SIAM Rev.* **1979**, *21*, 460–480.
16. Theiler, J.; Eubank, S.; Longtin, A.; Galdrikian, B.; Farmer, J.D. Testing for nonlinearity in time series: The method of surrogate data. *Phys. D Nonlinear Phenom.* **1992**, *58*, 77–94.
17. Schreiber, T.; Schmitz, A. Surrogate time series. *Phys. D Nonlinear Phenom.* **2000**, *142*, 346–382.
18. Horowitz, J.L. The Bootstrap. *Handb. Econ.* **2001**, *5*, 3159–3228.
19. Hinich, M.J.; Mendes, E.M.; Stone, L. Detecting nonlinearity in time series: Surrogate and bootstrap approaches. *Stud. Nonlinear Dyn. Econom.* **2005**, *9*, doi:10.2202/1558-3708.1268.
20. Faes, L.; Porta, A.; Nollo, G. Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in bivariate time series: Comparison among different strategies based on k nearest neighbors. *Phys. Rev. E* **2008**, *78*, 026201.
21. Papan, A.; Kyrtsov, K.; Kugiumtzis, D.; Diks, C. Assessment of resampling methods for causality testing: A note on the US inflation behavior. *PLoS ONE* **2017**, *12*, e0180852, doi:10.1371/journal.pone.0180852.
22. Quiroga, R.Q.; Kraskov, A.; Kreuz, T.; Grassberger, P. Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E* **2002**, *65*, 041903.
23. Marschinski, R.; Kantz, H. Analysing the information flow between financial time series. *Eur. Phys. J. B Condens. Matter Complex Syst.* **2002**, *30*, 275–281.
24. Kugiumtzis, D. Evaluation of surrogate and bootstrap tests for nonlinearity in time series. *Stud. Nonlinear Dyn. Econom.* **2008**, doi:10.2202/1558-3708.1474.
25. Robinson, P.M. Consistent nonparametric entropy-based testing. *Rev. Econ. Stud.* **1991**, *58*, 437–453.
26. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
27. Granger, C.; Maasoumi, E.; Racine, J. A dependence metric for possibly nonlinear processes. *J. Time Ser. Anal.* **2004**, *25*, 649–669.
28. Diks, C.; Fang, H. *Detecting Granger Causality with a Nonparametric Information-Based Statistic*; CeNDEF Working Paper No. 17-03; University of Amsterdam: Amsterdam, The Netherlands, 2017.
29. Moddemeijer, R. On estimation of entropy and mutual information of continuous distributions. *Signal Process.* **1989**, *16*, 233–248.
30. Diks, C.; Manzan, S. Tests for serial independence and linearity based on correlation integrals. *Stud. Nonlinear Dyn. Econom.* **2002**, *6*, 1–20, doi:10.2202/1558-3708.1005.
31. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138.
32. Wand, M.P.; Jones, M.C. *Kernel Smoothing*; CRC Press: Boca Raton, FL, USA, 1994; Volume 60.
33. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, USA, 1986; Volume 26.
34. Joe, H. Estimation of entropy and other functionals of a multivariate density. *Ann. Inst. Stat. Math.* **1989**, *41*, 683–697.
35. Van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 2000.

36. Jones, M.C.; Marron, J.S.; Sheather, S.J. A brief survey of bandwidth selection for density estimation. *J. Am. Stat. Assoc.* **1996**, *91*, 401–407.
37. He, Y.L.; Liu, J.N.; Wang, X.Z.; Hu, Y.X. Optimal bandwidth selection for re-substitution entropy estimation. *Appl. Math. Comput.* **2012**, *219*, 3425–3460.
38. Politis, D.N.; Romano, J.P. The stationary bootstrap. *J. Am. Stat. Assoc.* **1994**, *89*, 1303–1313.
39. Kugiumtzis, D. Direct-coupling information measure from nonuniform embedding. *Phys. Rev. E* **2013**, *87*, 062918.
40. Papan, A.; Kyrtsou, C.; Kugiumtzis, D.; Diks, C. Simulation study of direct causality measures in multivariate time series. *Entropy* **2013**, *15*, 2635–2661.
41. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; CRC Press: Boca Raton, FL, USA, 1994.
42. Shao, J.; Tu, D. *The Jackknife and Bootstrap*; Springer: Berlin/Heidelberg, Germany, 2012.
43. Neumann, M.H.; Paparoditis, E. On bootstrapping L2-type statistics in density testing. *Stat. Probab. Lett.* **2000**, *50*, 137–147.
44. Paparoditis, E.; Politis, D.N. The local bootstrap for kernel estimators under general dependence conditions. *Ann. Inst. Stat. Math.* **2000**, *52*, 139–159.
45. Engle, R.; Granger, C. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **1987**, *2*, 251–276.
46. Bekiros, S.; Diks, C. The nonlinear dynamic relationship of exchange rates: Parametric and nonparametric causality testing. *J. Macroecon.* **2008**, *30*, 1641–1650.
47. Baek, E.G.; Brock, W.A. *A General Test for Nonlinear Granger Causality: Bivariate Model*; Working Paper; Iowa State University and University of Wisconsin: Madison, WI, USA, 1992.
48. Davidson, J. Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *J. Econom.* **2002**, *2*, 243–269.
49. Diebold, F.X.; Yilmaz, K. Measuring financial asset return and volatility spillovers, with application to global equity markets. *Econ. J.* **2009**, *119*, 158–171.
50. Gamba-Santamaria, S.; Gomez-Gonzalez, J.E.; Hurtado-Guarin, J.L.; Melo-Velandia, L.F. *Volatility Spillovers Among Global Stock Markets: Measuring Total and Directional Effects*; Technical Report No. 983; Banco de la Republica de Colombia: Bogotá, Colombia, 2017.
51. Vakorin, V.A.; Krakovska, O.A.; McIntosh, A.R. Confounding effects of indirect connections on causality estimation. *J. Neurosci. Methods* **2009**, *184*, 152–160.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).