Chapter 16

A general nonparametric bootstrap test for Granger causality

Cees Diks
University of Amsterdam

Jacob DeGoede
University of Leiden

We introduce an information theoretic test statistic for Granger causality, which can be estimated by means of correlation integrals. The significance of the test statistic is determined using bootstrap methods rather than asymptotic distribution theory. Several bootstrap strategies are suggested and compared by Monte Carlo simulations. All these bootstrap methods appear to work well, but only bootstrapping the hypothesized noncausing time series has the additional advantage that a simplified test statistic can be used.

16.1 Granger causality

For the case of two scalar-valued time series, $\{X_t\}$ and $\{Y_t\}$, intuitively, $\{Y_t\}$ is a Granger cause of $\{X_t\}$ if past and present values of Y contain information about the distribution of future values of X, not contained in past and present observations of X. This causality concept, which will be defined more formally later, is useful in empirical research on causal relationships among observed time series.

Tests for Granger causality originate from econometrics and have recently been applied in fields ranging from neurology [4] to epidemiology [10]. In econometrics the challenge consists of detecting and characterising dependence within a sea of noise. Koch and Koch [7] examined contemporaneous and lead–lag relationships across national equity markets, and found that the interdependence across national markets has increased over time. They also found that most dependence occurs within 24 hours, and that Japan's market influence has increased to a size comparable to that of the USA. Tan and Cheng [16] reported the presence of causal links between money and output. Since Granger causality goes beyond dependence and focuses on causal relationships, such results are of vital importance in establishing sensible control strategies in the form of government policies, supported by empirical evidence that spurious correlations can be ruled out.

The traditional approach to testing for Granger causality consists of comparing prediction errors of an autoregressive model of X with the prediction errors obtained by a model which regresses X on past and current values of both X and Y. This approach is appealing, since the test reduces to determining the significance of the coefficients of the terms in the regression that depend on past and current values of Y. There are, however, two disadvantages. First, parametric tests require modelling assumptions such as linearity of the regression structure and, second, tests based on prediction errors will be sensitive only to causality in the mean. Higher-order structure, such as heteroskedasticity, will be ignored.

Asimakopoulos *et al* [1] examined nonlinear Granger causality in currency futures returns, and found several uni-directional causal relationships. A study by Longin and Solnik [8] confirmed the often reported empirical fact that markets are linked more strongly in periods of higher volatility. This is an indication of nonlinearity and stresses the importance of general tests for Granger causality which are sensitive also to nonlinear causal relationships. The nonlinear aspect was modelled explicitly in a parametric test for causality-in-variance proposed by Cheung and Ng [3].

In this contribution we will focus on nonparametric tests, which eliminates possible problems resulting from model misspecification. Before describing general nonparametric tests for Granger causality, it will be useful to give a more formal definition of the concept of causality. For two strictly stationary time series $\{X_t\}$ and $\{Y_t\}$, satisfying some weak mixing conditions, we define Granger causality as follows.

Definition 16.1. *$\{Y_t\}$ is a nonlinear Granger cause of $\{X_t\}$ (denoted $Y \to X$) if*

$$F_{X_{t+1}}(x|\mathcal{F}_X(t), \mathcal{F}_Y(t)) \neq F_{X_{t+1}}(x|\mathcal{F}_X(t)),$$

where $F_{X_{t+1}}(x|\mathcal{F})$ denotes the cumulative distribution function of X_{t+1} given $\mathcal{F}$, and $\mathcal{F}_X(t)$ and $\mathcal{F}_Y(t)$ denote the information sets consisting of past observations of X and Y up to and including time t.

Note that this definition relies on comparing the one-step-ahead distribution of X with and without past and current observed values of Y. Generalizations of this definition of Granger causality which consider future joint distributions of X can be formulated in a similar way. Also note that our definition should be considered an operational definition of causality, since the existence of an unobserved variable, Z say, which is causing both X and Y, leading to Y being a Granger cause of X, can never be excluded empirically. However, the great advantage of the concept of Granger causality is that it is empirically testable.

We propose an information theoretic test for Granger causality for stationary weakly dependent time series, based on conditional entropies. These entropies can be expressed in terms of correlation integrals, the nonparametric estimation of which is straightforward. Correlation integrals originate from the study of chaotic systems, where they are important means of characterizing the dynamics of deterministic processes. The contributions of Floris Takens to this field are well known [15]. The test proposed by Hiemstra and Jones [5], who reported bi-directional Granger causality (interaction) between changes in trading volume on the New York Stock Exchange and returns of the Dow Jones Industrial Average Index, is also based on correlation integrals, and closely resembles the test proposed here. However, our test statistic is motivated by information theoretic arguments and we use bootstrap methods rather than asymptotic theory for determining the significance of the test statistic.

An important new insight is recognition of the connection between correlation integrals and information theory, see e.g. Prichard and Theiler [13]. Correlation integral based information theoretic quantities require only weak assumptions on the underlying processes, and yet turn out to be powerful tools for characterizing causal relationships and quantifying information flows. A great advantage is that applications of these methods are no longer restricted to deterministic time series but are suitable for arbitrary stationary, weakly mixing processes. A famous and widely used example of a correlation integral based test for serial independence was put forward by Brock *et al* [2].

The theory of bootstraps for dependent processes also has been developed strongly recently, and with the current cheaply available computational power these bootstraps have become practically feasible even for large data sets. The combined use of correlation integral based information theoretical statistics together with recently developed time series bootstrap methods promises to provide powerful and statistically sound means for studying dynamical relationships among time series.

16.2 Information theoretic test statistic

Given two time series $\{X_t\}$ and $\{Y_t\}$ we wish to test the null hypothesis

$$H_0: \qquad \{Y_t\} \text{ is not Granger causing } \{X_t\}$$

According to the established tradition in statistics we should be speaking of testing the null hypothesis, rather than its negation. This implies that we are considering tests for Granger noncausality rather than tests for Granger causality. However, for simplicity we choose to continue this slight abuse of language.

Using information theoretic quantities, we take an approach that closely follows the definition of Granger causality in the previous section. First, from the time series $\{X_t\}$ and $\{Y_t\}$ delay vectors

$$\boldsymbol{X}_t = (X_{t-m+1}, \ldots, X_t), \qquad \boldsymbol{Y}_t = (Y_{t-l+1}, \ldots, Y_t) \tag{16.1}$$

are constructed of embedding dimension m and l, respectively. The idea is to quantify the average amount of extra information on X_{t+1} contained in the delay vector $\boldsymbol{Y}_t$, given that we already know $\boldsymbol{X}_t$.

Generally speaking, the average amount of information a random variable $\boldsymbol{X}$ contains about a random variable $\boldsymbol{Y}$ can be expressed as the generalized [13] mutual information of $\boldsymbol{X}$ and $\boldsymbol{Y}$, which, in terms of correlation integrals, reads

$$\boldsymbol{I}_q(\boldsymbol{X};\boldsymbol{Y}) = \ln C_q(\boldsymbol{X},\boldsymbol{Y}) - \ln C_q(\boldsymbol{X}) - \ln C_q(\boldsymbol{Y}) \tag{16.2}$$

where

$$C_q(\boldsymbol{X},\epsilon) = \left[\int \left(\int I_{(\|\boldsymbol{x}-\boldsymbol{y}\|\leq\epsilon)}\, \mathrm{d}\mu_{\boldsymbol{X}}(\boldsymbol{x})\right)^{q-1} \mathrm{d}\mu_{\boldsymbol{X}}(\boldsymbol{y})\right]^{\frac{1}{q-1}}. \tag{16.3}$$

Here $I_{(\cdot)}$ denotes the indicator function which is equal to one if its argument is true, and is zero otherwise, and $\|\cdot\|$ denotes the supremum norm

$$\|\boldsymbol{x}\| = \sup_{i=1,\ldots,\dim \boldsymbol{x}} |x_i|. \tag{16.4}$$

For $q = 2$, $C_2(\boldsymbol{X},\epsilon)$ is nothing but the fraction of distances between two independently chosen points, according to $\mu_{\boldsymbol{X}}$, that is smaller than or equal to ϵ. Thus the second-order ($q = 2$) correlation integral is equal to the probability that a distance between two independent realizations of $\boldsymbol{X}$ is smaller than or equal to ϵ. For computational convenience, we use $q = 2$ in our calculations, and for simplicity the index q as well as the scale parameter ϵ are omitted in the notation of the correlation integrals in the sequel. When $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent, the joint correlation integral factorizes, $C(\boldsymbol{X},\boldsymbol{Y}) = C(\boldsymbol{X})C(\boldsymbol{Y})$, and $\boldsymbol{I}(\boldsymbol{X},\boldsymbol{Y}) = 0$. In the extreme case where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are identical, one has $\ln C(\boldsymbol{X},\boldsymbol{Y}) = \ln C(\boldsymbol{X}) = \ln C(\boldsymbol{Y})$, so that $\boldsymbol{I}(\boldsymbol{X},\boldsymbol{Y}) = \ln C(\boldsymbol{X})$.

In a time series setting the average information about X_{t+1} contained in $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ jointly, is given by

$$\boldsymbol{I}(\boldsymbol{X}_t,\boldsymbol{Y}_t;X_{t+1}) = \ln C(\boldsymbol{X}_t,\boldsymbol{Y}_t,X_{t+1}) - \ln C(\boldsymbol{X}_t,\boldsymbol{Y}_t) - \ln C(X_{t+1}) \tag{16.5}$$

while the average information about X_{t+1} in $\boldsymbol{X}_t$ only is given by

$$\boldsymbol{I}(\boldsymbol{X}_t;X_{t+1}) = \ln C(\boldsymbol{X}_t,X_{t+1}) - \ln C(\boldsymbol{X}_t) - \ln C(X_{t+1}). \tag{16.6}$$

By subtracting these two information measures, we can quantify the average amount of extra information that $\boldsymbol{Y}_t$ contains about X_{t+1} in addition to the information already in $\boldsymbol{X}_t$.

If past observations of Y contain no extra information about future values of X, one has $\boldsymbol{I}(\boldsymbol{X}_t, \boldsymbol{Y}_t; X_{t+1}) = \boldsymbol{I}(\boldsymbol{X}_t; X_{t+1})$. If, on the other hand, past observations of Y do contain information on current and future values of X, we expect $\boldsymbol{I}(\boldsymbol{X}_t, \boldsymbol{Y}_t; X_{t+1}) > \boldsymbol{I}(\boldsymbol{X}_t; X_{t+1})$. As our test statistic we use a correlation integral based estimator of

$$\begin{aligned} Q &= \boldsymbol{I}(\boldsymbol{X}_t, \boldsymbol{Y}_t; X_{t+1}) - \boldsymbol{I}(\boldsymbol{X}_t; X_{t+1}) \\ &= \ln C(\boldsymbol{X}_t, \boldsymbol{Y}_t, X_{t+1}) - \ln C(\boldsymbol{X}_t, \boldsymbol{Y}_t) - \ln C(\boldsymbol{X}_t, X_{t+1}) + \ln C(\boldsymbol{X}_t) \end{aligned} \tag{16.7}$$

which gives

$$\widehat{Q} = \ln \widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t, X_{t+1}) - \ln \widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t) - \ln \widehat{C}(\boldsymbol{X}_t, X_{t+1}) + \ln \widehat{C}(\boldsymbol{X}_t) \tag{16.8}$$

where $\widehat{C}(\boldsymbol{X})$ represents the estimated correlation integral of $\boldsymbol{X}$. Rather than using asymptotic theory, bootstrap methods will be used to determine whether $\widehat{Q}$ is significantly larger than zero.

Since we expect the test statistic to be equal to zero under the null hypothesis and positive under alternatives, a one-sided test is called for. The null hypothesis is rejected only if $\widehat{Q}$ is significantly larger than zero. However, there are some subtleties involved here, and exceptions can be constructed for which the test statistic decreases in the presence of Granger causality. This is related to the fact that the correlation integral is not always larger for time series with more structure, which was pointed out to us by Floris Takens (also see [15]). Anomalies like these can be traced back to the use of the order-two correlation integral C_2 rather than the correlation integral C_1 of order one. Information theoretic quantities defined in terms of C_1 rather than C_2, have a number of nice properties which those based on C_2 are lacking. Some authors resolve this point by implementing tests based on C_1, which, however, is much more difficult to estimate. Others solve the problem by analysing ranks rather than the raw data [12]. In practice the differences between C_1 and C_2 are often small. We choose to use C_2 for efficiency reasons, at the expense of a possible loss of statistical power.

Notice that Hiemstra and Jones [5] test the relationship

$$\frac{C(\boldsymbol{X}_t, \boldsymbol{Y}_t, X_{t+1})}{C(\boldsymbol{X}_t, \boldsymbol{Y}_t)} = \frac{C(\boldsymbol{X}_t, X_{t+1})}{C(\boldsymbol{X}_t)} \tag{16.9}$$

by calculating

$$\widehat{T} = \frac{\widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t, X_{t+1})}{\widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t)} - \frac{\widehat{C}(\boldsymbol{X}_t, X_{t+1})}{\widehat{C}(\boldsymbol{X}_t)} \tag{16.10}$$

from the data rejecting the null hypothesis whenever $\widehat{T}$ is too large. Upon taking logarithms on both sides of (16.9), it can be shown that they test exactly the same equality as we do. However, their test statistic is different, and cannot be mapped to ours in a one-to-one fashion. Furthermore, they use asymptotic distribution theory rather than a bootstrap test, so that the size and power of their test need not be the same as ours.

16.3 Bootstrap procedures

The stationary bootstrap proposed by Politis and Romano [11] is simple to apply to univariate time series. We are dealing with the bivariate case, and there are several ways of bootstrapping the two time series. Before describing the various possibilities, a brief description of the stationary bootstrap will be given.

The stationary bootstrap replicates the time series by concatenating blocks of observations from the original time series. The blocks are selected randomly from the original time series and have a random length with a geometric distribution. To ensure stationarity of the bootstrap time series, whenever a block exceeds the end of the time series, one continues by adding observations starting from the beginning of the time series. The following implementation is used for constructing a bootstrap replication $\{X_t^*\}_{t=1}^N$ of $\{X_t\}_{t=1}^N$. An index, i_1, is selected randomly according to the uniform distribution on $1, \ldots, N$. The first observation of the bootstrap time series X_1^* is taken to be X_{i_1}. Then, with probability $1 - P$, with P small, one chooses $i_2 = i_1 + 1$, and with probability P, i_2 is selected randomly again from the uniform distribution on $1, \ldots, N$. The next value in the bootstrap time series is then taken to be X_{i_2}. In this way, one continues until a time series of length N is obtained. Whenever i_k becomes equal to $N + 1$, i_k is set to 1, that is, to the index pointing to the first observation of the original time series.

Let us now return to the various possibilities of bootstrapping a pair of time series in the context of testing for Granger causality. We would like to test the null hypothesis that Y is not Granger causing X. As a first attempt we could bootstrap $\{X_t\}$ and $\{Y_t\}$ independently, to obtain a distribution under the null hypothesis (the absence of Granger causality). This bootstrap method will be referred to as the XY bootstrap. Note that this destroys any dependence between X and Y rather than only Granger causality, if present. This may influence the distribution of the test statistic and, hence, under the null of no Granger causality may introduce deviations of the rejection probability from the nominal size. In the simulation studies presented later we will consider this point.

A slight modification is to bootstrap only the 'causing' time series $\{Y_t\}$. This bootstrap, referred to as the Y bootstrap, also ensures the absence of information of Y on future values of X and, hence, should also work. Since the variability of the test statistic is expected to decrease using the Y bootstrap, the power of the test may increase for Y. Also it should be examined whether the Y bootstrap method gives rise to a size which differs from the nominal size in the presence of dependence, since this dependence is also lost under the bootstrap procedure.

To preserve dependence between X and Y one can bootstrap X and Y contemporaneously. That is, whenever an index i_k is selected the bootstrap values for X_k^* and Y_k^* are selected with the same index, $X_k^* = X_{i_k}$, and $Y_k^* = Y_{i_k}$. This we call the (X, Y) bootstrap, where the brackets now indicate that the two time series are considered as one bi-variate time series. Using this approach we no longer perform bootstraps under the null hypothesis, because Granger causality in the original time series will be preserved in the bootstrap one. This actually amounts to performing a standard bootstrap procedure, in which the bootstrap distribution of the test statistics can be expected to be centred (on average) around the value of the test statistic of the original time series. Before determining the p-values from the bootstrap distribution one should first centre the bootstrap distribution of the test statistic around zero, which is the expected value of the test statistic under the null hypothesis.

In the last method considered, referred to as the $(\boldsymbol{X}, \boldsymbol{Y})$ bootstrap, delay vectors of embedding dimension $m + 1$ are bootstrapped contemporaneously instead of individual observations.

Summarizing, we examine the following bootstrap procedures:

- bootstrapping both time series independently, XY
- bootstrapping only the 'causing' time series, Y
- bootstrapping contemporaneously, (X, Y)
- bootstrapping delay vectors, $(\boldsymbol{X}, \boldsymbol{Y})$.

Strictly speaking, only the last two methods are bootstraps, and the first two methods are randomization procedures aimed at 'bootstrapping under the null'. There is some analogy with randomization tests for serial independence. Under the assumption that a time series consists of independent, identically distributed observations, one can randomize the data by permuting them randomly. Since under the assumption of independence permutations are equally likely, this randomization procedure is exact. The randomization procedures proposed here are not exact, and before they are used in practice, at least some numerical evidence is required to warrant their application. In the next section some preliminary results are presented.

16.4 Monte Carlo simulations

In this section the size and power of the bootstrap test are determined numerically by Monte Carlo simulation for various bivariate time series models. This is important since it became clear in the previous section that none of the bootstrap procedures can be expected *a priori* to have the desired size properties.

The test statistic determined from the original pair of time series $\{X_t\}$ and $\{Y_t\}$ is denoted by $\widehat{Q}_1$. The values of the $B - 1$ bootstrap replications $\{X_t^*\}$ and $\{Y_t^*\}$ of the pair of time series are referred to as $\widehat{Q}_2, \ldots, \widehat{Q}_B$. The p-value is determined as

$$\widehat{p} = \frac{\sum_{i=1}^{B} I(\widehat{Q}_i \geq \widehat{Q}_1)}{B}. \tag{16.11}$$

The numbers presented in the tables are the fractions of rejections at a nominal size $\alpha = 0.05$ for 1000 independent realizations, where we used $B = 20$. That is, the bootstrap test is applied 1000 times to independently generated realizations of the pair of time series, and the rejection rate (the relative number of times

Table 16.1. Rejection rates (size) in absence of Granger causality, $X_t \sim N(0, 1)$ and $Y_t \sim N(0, 1)$, independently; $l > 0$ corresponds to tests for $Y \to X$, and $l < 0$ to $X \to Y$.

l	XY	Y	(X, Y)	$(\boldsymbol{X}, \boldsymbol{Y})$	HJ
−2	0.043	0.049	0.034	0.021	0.025
−1	0.067	0.048	0.024	0.028	0.018
1	0.050	0.054	0.026	0.026	0.011
2	0.052	0.049	0.030	0.016	0.019

that $\widehat{p} \leq 0.05$) is quoted. The size is the relative number of rejections for processes that satify the null hypothesis. Ideally, this 'actual' size should be close to the nominal size. If the actual size is smaller than the nominal size the test is called conservative. If the actual size is larger than the nominal size, the rejection probability is larger than the nominal size for processes that satisfy the null hypothesis, that is, the type I error is larger than the nominal size. This certainly is undesirable for a statistical test. The rejection rate under alternatives, i.e. processes not satisfying the null hypothesis, is called the power of the test. Provided that the size does not exceed the nominal size, the larger the power of the test, the better.

The aim of this Monte Carlo study is twofold. We examine the size not only in cases where $\{X_t\}$ and $\{Y_t\}$ are independent, but also dependent with and without Granger causality. We also want to examine the power of the test in the presence of Granger causality. Throughout we will compare the size and the power of the test with the size and power obtained with the test of Hiemstra and Jones [5]. Our ultimate goal is to estimate the effect of dependence and to examine whether we can use the 'randomization' approaches, in which dependence is ignored.

We use the following parameter values. In testing $Y \to X$ the embedding dimension for the X time series is set to $m = 2$, while the embedding dimension for Y can take the values $l = 1, 2$. In testing $X \to Y$ the same parameter values are used but the roles of X and Y are reversed. The scale parameter is taken to be $\epsilon = 1$ (after rescaling each time series to unit variance). The time series length for the Monte Carlo simulations is $N = 100$. The switching probability used in the stationary bootstrap is set to $P = 0.05$. We used $B = 20$, which amounts to $B - 1 = 19$ bootstrap replications.

16.4.1 Size

In this subsection we determine the size for two bivariate processes without Granger causality. Cases in which the time series are independent, as well as dependent, are examined.

First we study an example in which $\{X_t\}$ and $\{Y_t\}$ are independent. We

Table 16.2. Rejection rates (size) in absence of Granger causality, but in the presence of dependence, $(X_t, Y_t) \sim \text{BVN}(0, 0, 1, 1, \frac{1}{2})$.

l	XY	Y	(X, Y)	$(\boldsymbol{X}, \boldsymbol{Y})$	HJ
−2	0.029	0.028	0.022	0.027	0.023
−1	0.030	0.038	0.023	0.024	0.017
1	0.028	0.041	0.020	0.019	0.015
2	0.047	0.034	0.031	0.024	0.027

take both processes to consist of independent, normally distributed values, that is, $X_t \sim N(0, 1)$, and $Y_t \sim N(0, 1)$. In this case there is no Granger causality and no dependence, and all bootstrap methods should work, at least asymptotically. Table 16.1 gives the sizes obtained for this example. In all tables, positive values of the lag l correspond to tests for $Y \to X$, and negative values l correspond to tests for $X \to Y$. The different rows correspond to the various methods examined. The last row HJ denotes the Hiemstra and Jones test. The size is close to the nominal size $\alpha = 0.05$ for the first two bootstrap methods, which both ignore dependence. Note, however, that the test appears to be somewhat conservative for the two bootstraps which are constructed to preserve dependence. This holds true also for the Hiemstra and Jones test.

Next we consider an example in which there is instantaneous dependence between the two time series, but no Granger causality. We take (X_t, Y_t) to be bivariate normally distributed, with a correlation coefficient of $\frac{1}{2}$, denoted by $\text{BVN}(0, 0, 1, 1, \frac{1}{2})$, a bivariate normal distribution for which the two components both have mean zero and unit variance while the correlation between the components is $\frac{1}{2}$. The resulting rejection rates, given in table 16.2 suggest that all tests are slightly conservative.

16.4.2 Size and power

As a first example with Granger causality, a case is considered with uni-directional Granger causality. We examine the previous process again, but now with the Y time series shifted in time by one time unit, so that it is running ahead. In this way, a situation is obtained in which Y Granger-causes X. The process satisfies $(X_t, Y_{t-1}) \sim \text{BVN}(0, 0, 1, 1, \frac{1}{2})$. The rejection rates for $Y \to X$ now amount to the power of the test, whereas the rejection rates for $X \to Y$ $(l < 0)$ must be interpreted as sizes. The results shown in table 16.3 suggest that again the actual size is smaller than the nominal size, and that the bootstraps preserving dependence, as well as the Hiemstra and Jones test, are slightly more conservative than the tests which ignore dependence (XY and Y). In terms of power $(l > 0)$ the tests which ignore dependence appear to perform slightly better than the methods

Table 16.3. Rejection rates (size and power) in the presence of uni-directional Granger causality, $Y \to X$, with $(X_t, Y_{t-1}) \sim \text{BVN}(0, 0, 1, 1, \frac{1}{2})$.

l	XY	Y	(X, Y)	$(\boldsymbol{X}, \boldsymbol{Y})$	HJ
−2	0.033	0.037	0.019	0.026	0.026
−1	0.037	0.039	0.023	0.024	0.021
1	0.717	0.733	0.550	0.614	0.591
2	0.452	0.479	0.384	0.617	0.355

Table 16.4. Granger causality, $Y \to X$, linear dependence.

l	XY	Y	(X, Y)	$(\boldsymbol{X}, \boldsymbol{Y})$	HJ
−2	0.027	0.035	0.035	0.027	0.034
−1	0.029	0.032	0.027	0.034	0.031
1	0.812	0.830	0.700	0.707	0.781
2	0.615	0.570	0.567	0.720	0.634

Table 16.5. Power, linear interaction case.

l	XY	Y	(X, Y)	$(\boldsymbol{X}, \boldsymbol{Y})$	HJ
−2	0.319	0.314	0.344	0.721	0.506
−1	0.524	0.524	0.448	0.728	0.652
1	0.492	0.533	0.489	0.721	0.585
2	0.290	0.310	0.337	0.719	0.444

which preserve dependence and also better than the Hiemstra and Jones test.

The time series generated by the model

$$\begin{aligned} X_t &= 0.6X_{t-1} + 0.5Y_{t-1} + \epsilon_t \\ Y_t &= 0.6Y_{t-1} + \epsilon_t' \end{aligned} \tag{16.12}$$

where ϵ_t and ϵ_t' are independent and standard normally distributed, also exhibit uni-directional Granger causality, $Y \to X$. The rejection rates for this model are given in Table 16.4. For the first lag $l = 1$ the first two bootstrap methods again appear to have slightly more power than the other bootstrap methods and the Hiemstra and Jones test.

Table 16.6. Size and power for a model with nonlinear Granger causality, $Y \to X$.

l	XY	Y	(X, Y)	$(\boldsymbol{X}, \boldsymbol{Y})$	HJ
−2	0.037	0.033	0.036	0.030	0.022
−1	0.024	0.030	0.028	0.032	0.025
1	0.965	0.956	0.889	0.910	0.926
2	0.869	0.861	0.805	0.895	0.823

Next an interaction case is considered, given by

$$\begin{aligned} X_t &= 0.5X_{t-1} + 0.4Y_{t-1} + \epsilon_t \\ Y_t &= 0.5Y_{t-1} + 0.4X_{t-1} + \epsilon'_t. \end{aligned} \tag{16.13}$$

Table 16.5 shows the obtained powers for this process. The tests which ignore dependence still have some power, but this can be seen to be considerably smaller than that for the tests which preserve dependence, and the Hiemstra and Jones test. A possible explanation for the small power of the bootstrap test could be the fact that a very small value of B was used ($B = 20$). As shown by Hope [6] and Marriott [9] the power slightly increases when a larger number of bootstrap replications are used, but there is no need to choose an excessively large value for B. According to Marriot, typically 5 is a suitable value for αB, suggesting the choice $B = 100$ rather than $B = 20$ for $\alpha = 0.05$. Indeed the power for the bootstraps was observed to improve slightly in trial runs with $B = 100$, but this increase was certainly not sufficient to change our conclusion that the Hiemstra and Jones test outperforms all but the $(\boldsymbol{X}, \boldsymbol{Y})$ bootstraps in this case.

The last process we consider exhibits nonlinear Granger causality $Y \to X$. The model is

$$\begin{aligned} X_t &\sim N(0, \sigma_t^2) \\ \sigma_t &= Y_{t-1} \sim N(0, 1). \end{aligned} \tag{16.14}$$

This is a simple model of bivariate conditional heteroskedasticity. Processes with conditional heteroskedasticity play an important role in econometrics, where they are frequently used to model a phenomenon referred to as volatility clustering. This is the tendency of stock prices to show larger price movements after periods of large price movements, and smaller price movements after periods with small price movements. Table 16.6 shows the estimated size and power for this example. The size and power appear to be comparable for all tests.

16.5 Summary and discussion

We argued that information theoretic quantities provide a natural means of testing for Granger causality, and we proposed an information theoretic test statistic for

Granger causality. It was shown that several alternative time series bootstrap strategies perform well in terms of size and power. All bootstrap tests, and also Hiemstra and Jones's test, were found to be conservative, at least for the examples studied here. This implies that practitioners need not fear that the rate of type I errors exceeds the nominal size. In some examples the power of the tests outperformed the Hiemstra and Jones test, whereas the opposite also occurred.

The bootstrap tests turn out to be very robust against dependence between X and Y, even when the bootstraps destroy this. A possible explanation for this phenomenon is that the test statistic picks out just the right type of dependence between two time series. The test statistic by construction is only sensitive to dependence that can be associated with Granger causality. All other dependence such as instantaneous correlation is ignored by the test statistic.

If only the noncausing time series $\{Y_t\}$ is bootstrapped, the terms $\ln \widehat{C}(\boldsymbol{X}_t^*, X_{t+1}^*)$ and $\ln \widehat{C}(\boldsymbol{X}_t^*)$ for all bootstraps are equal, and equal to the value obtained for the original time series. Therefore, p-values determined with this method will remain unchanged upon leaving out these terms from the test statistic. This bootstrap method thus has the advantage that a simplified test statistic can be used, namely $\widehat{Q}' = \ln \widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t, X_{t+1}) - \ln \widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t)$. The quantity $-Q'$ can be interpreted as the correlation entropy of X conditional on Y. Indeed, when Y Granger causes X one would expect the conditional correlation entropy of X given Y to be smaller than when Y contains no additional information on future values of X. This suggests that a Hiemstra and Jones type of test could also be designed based on $\widehat{T}' = \widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t, X_{t+1})/\widehat{C}(\boldsymbol{X}_t, \boldsymbol{Y}_t)$. If the asymptotic variance of this simplified statistic is known, conditionally on X, the significance can be determined by calculating how far, in terms of standard deviations, it is located from its value under the null hypothesis, $\widehat{T_0} = \widehat{C}(\boldsymbol{X}_t, X_{t+1})/\widehat{C}(\boldsymbol{X}_t)$.

Although the nonparametric bootstrap tests discussed here all have nice size and power properties, they remain quite uninformative about the nature of the Granger causality involved. Information on the exact lags involved is difficult to obtain from the test results, even when the results for different lags are compared. For example, when Y Granger causes X through the first lag only, the test will also detect Granger causality for $l = 2$, simply because for $l = 2$ the delay vector $\boldsymbol{Y}_t$ contains, apart from Y_{t-1}, also the lagged value Y_t which contains information on X_{t+1}. A possible solution, in the spirit of Savit and Green [14] who use a similar approach for lagged dependence in a univariate time series setting, is to compare the information about X_{t+1} contained in $(\boldsymbol{X}_t, Y_{t-1}, Y_t)$ with that contained only in $(\boldsymbol{X}_t, Y_t)$. In this way, the extra information in each of the added lagged values of Y can be examined separately.

In the examples studied the bootstrap tests and the Hiemstra and Jones test performed about equally well. The information theoretic approach, however, has the advantage that clear-cut statistical quantities can be used for time series problems that involve information flowing from one variable to another. Therefore, a future direction is the development of an asymptotic theory for the information theoretic test statistics proposed in this contribution and related statistics. Since correlation integrals are known to be asymptotically normal for stationary mixing processes, the asymptotic distributions of the estimators of the information theoretic quantities based on them, are expected to be analytically tractable.

References

[1] Asimakopoulos I, Ayling D and Mahmood W M 2000 Nonlinear Granger causality in the currency futures returns *Econom. Lett.* **68** 25–30

[2] Brock W A, Dechert W D, Scheinkman J A and LeBaron B 1996 A test for independence based on the correlation dimension *Econom. Rev.* **15** 197–235

[3] Cheung Y-W and Ng LK 1996 A causality-in-variance test and its application to financial market prices *J. Econom.* **72** 33–48

[4] Freiwald W A, Valdes P, Bosch J, Biscay R, Jimenez J C, Rodriguez L M, Rodriguez V, Kreiter A K and Singer W 1999 Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex *J. Neurosci. Meth.* **94** 105–19

[5] Hiemstra C and Jones J D 1994 Testing for linear and nonlinear Granger causality in the stock price–volume relation *J. Finance* **49** 1639–64

[6] Hope A C A 1968 A simplified Monte Carlo significance test procedure *J. R. Stat. Soc.* B **30** 582–98

[7] Koch P D and Koch T W 199) Evolution in dynamic linkages across daily national stock indices *J. Int. Money Finance* **10** 231–51

[8] Longin F and Solnik B 1995 Is the correlation in international equity returns constant: 1960–1990? *J. Int. Money Finance* **14** 3–26

[9] Marriott F H C 1979 Barnard's Monte Carlo tests: how many simulations? *Appl. Stat.* **28** 75–7

[10] Pitard A and Viel J F 1999 A model selection tool in multi-pollutant time series: the Granger-causality diagnosis *Environmetrics* **10** 53–65

[11] Politis D N and Romano J P 1994 The stationary bootstrap *J. Am. Stat. Assoc.* **89** 1303–13

[12] Pompe B, Blidh P, Hoyer D and Eiselt M 1998 Using mutual information to measure coupling in the cardiorespiratory system. New insights into nonlinear coordinations *IEEE Eng. Med. Biol. Mag.* **17** 32–9

[13] Prichard D and Theiler J 1995 Generalized redundancies for time series analysis *Physica* D **84** 476–93

[14] Savit R and Green M 1991 Time series and dependent variables *Physica* D **50** 95–116

[15] Takens F 1993 Detecting nonlinearities in stationary time series *Int. J. Bif. Chaos* **3** 241–56

[16] Tan K-G and Cheng C-S 1995 The causal nexus of money, output and prices in Malaysia *Appl. Economics* **27** 1245–51