A new statistic and practical guidelines for nonparametric Granger causality testing

Cees Diks & Valentyn Panchenko

Center for Nonlinear Dynamics in Economics and Finance

Department of Economics, University of Amsterdam, Roetersstraat 11

1018 WB Amsterdam, The Netherlands

October 1, 2004

Abstract

Upon illustrating how smoothing may cause over-rejection in nonparametric tests for Granger non-causality, we propose a new test statistic for which problems of this type can be avoided. We develop asymptotic theory for the new test statistic, and perform a simulation study to investigate the properties of the new test in comparison with its natural counterpart, the Hiemstra-Jones test. Our simulation results indicate that, if the bandwidth tends to zero at the appropriate rate as the sample size increases, the size of the new test remains close to nominal, while the power remains large. Transforming the time series to uniform marginals improves the behavior of both tests. In applications to Standard and Poor's index volumes and returns, the Hiemstra-Jones test suggests that volume Granger-causes returns. However, the evidence for this gets weaker if we carefully apply the recommendations suggested by our simulation study.

Keywords: Financial time series; Granger causality; Nonparametric; Hypothesis testing; Size distortion; U-statistics;

JEL classification: C12, C51, E3

1 Introduction

Granger causality has turned out to be a useful notion for characterizing dependence relations between time series in economics and econometrics. While traditional parametric tests for Granger causality have reached a mature status, and have become part of the standard toolbox of economists, the recent literature attests to an increasing interest in nonparametric counterparts of these tests. Generally speaking, the availability of ever cheaper computational power has been accompanied by a growing literature on nonparametric statistics. In the particular case of Granger causality, an increasing number of studies report evidence for causality between economic variables on the basis of the Hiemstra and Jones (1994) test (hereafter often referred to as HJ test), which has become popular among practitioners. Alternative nonparametric tests, such as the non-causality test proposed by Bell *et al.* (1996) using additive models, and the test for conditional independence recently proposed by Su and White (2003), based on weighted Hellinger distances, may also be applied as a non-causality tests in economics and finance. Here we limit ourselves to a discussion of the HJ test and our proposed modification of it.

Intuitively, for a bivariate process $\{(X_t, Y_t)\}$, $\{X_t\}$ is a Granger cause of $\{Y_t\}$ if past and current values of X contain additional information on the distribution of future values of Y that is not contained in past Y-values alone. Applications of the HJ test on economic observables has indicated the presence of various causal relations. For instance, Brooks (1998) reports bidirectional Granger causality between volume and volatility on the New York Stock Exchange, Abhyankar (1998) and Silvapulla and Moosa (1999) find causality in futures markets, and Ma and Kanas (2000) report evidence for nonlinear Granger causality from French money to the FFr/DM exchange rate. This list is not intended to be complete, and can be easily extended to include applications in other areas, such as the interaction of the energy market and stock market prices (Ciner, 2001), real estate and stock market prices (Okunev *et al.*, 2000, 2002) or London Metal Exchange cash prices and some of its possible predictors (Chen and Lin, 2004).

This paper is organized in the following way. In the next section we restate the null hypothesis in terms of probability density functions, and show that if we try to interpret the Hiemstra-Jones test as a nonparametric test of the null hypothesis, it suffers from a bias which can lead to over-

rejection. This is related to an example presented in Diks and Panchenko (2004) for which the actual rejection rate of the HJ test under the null tends to one as the sample size increases. After a discussion of the nature of the bias we suggest a class of alternative test statistics, with a reduced risk of over-rejecting. Section 3 presents a simulation study in which we analyze the size and power of the HJ test and our new test in detail, and examine the dependence of the bandwidth on the sample size. In Section 4 we first reproduce evidence for volume Granger-causing returns, but show that a careful application of the recommendations suggested by our simulation study weakens the evidence for volume causing returns considerably. Section 5 summarizes our main results and concludes.

2   Nonparametric Granger Causality Tests

For a bivariate process $\{(X_t, Y_t)\}$, we say that X is a Granger cause of Y, if the distribution of Y_t given past observations of X and Y, differs from the distribution of Y_t given past observations of Y only. If we denote the information contained in past observations X_s and Y_s, $s < t$, by $\mathcal{F}_{X,t}$ and $\mathcal{F}_{Y,t}$, respectively, this can be formalized as: $\{X_t\}$ is a nonlinear Granger cause of $\{Y_t\}$ if

$$Y_t|(\mathcal{F}_{X,t}, \mathcal{F}_{Y,t}) \not\sim Y_t|\mathcal{F}_{Y,t},$$

where "$\sim$" represents equivalence of distributions.

In testing for Granger non-causality, the aim is to detect evidence against the null hypothesis

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

Under the null hypothesis Y_t is conditionally independent on $X_{t-1}, X_{t-2}, \ldots$, given $Y_{t-1}, Y_{t-2}, \ldots$. In a nonparametric setting, conditioning on the infinite past is impossible without a model restriction, such as an assumption that the order of the process is finite. Therefore, in practice conditional independence is tested using finite lags l_X and l_Y:

$$Y_t|(X_{t-1}, \ldots, X_{t-l_X}; Y_{t-1}, \ldots, Y_{t-l_Y}) \sim Y_t|(Y_{t-1}, \ldots, Y_{t-l_Y}).$$

For a stationary bivariate time series $\{(X_t, Y_t)\}$ this is a statement about the invariant distribution of the $l_X + l_Y + 1$-dimensional vector $W_t = (X_t^{l_X}, Y_t^{l_Y}, Z_t)$, with $X_t^{l_X} = (X_{t-l_X}, \ldots, X_{t-1})$,

3

$Y_t^{l_Y} = (Y_{t-L_Y}, \ldots, Y_{t-1})$, and $Z_t = Y_t$. To keep the notation compact, and to bring about the fact that the null hypothesis is a statement about the invariant distribution of W_t, we often drop the time index and just write $W = (X, Y, Z)$, where the latter is a random vector with the invariant distribution of $(X_t^{l_X}, Y_t^{l_Y}, Y_t)$. In this paper we often consider the choice $l_X = l_Y = 1$, in which case (X, Y, Z) denotes a three-variate random variable, distributed as (X_{t-1}, Y_{t-1}, Y_t).

The null hypothesis states that the distribution of Z given $(X, Y) = (x, y)$ is the same as that of Z given $Y = y$ (and thus independent of x). Therefore, if we assume that $W = (X, Y, Z)$ is a continuous random variable, its probability density function must be of the form

$$f_{X,Y,Z}(x, y, z) = f_{X,Z}(x, z) f_{Z|X,Y}(z|x, y) = f_{X,Z}(x, z) f_{Z|Y}(z|y)$$

Equivalent but for our purposes more convenient forms are

$$\frac{f_{X,Y,Z}(x, y, z)}{f_{X,Y}(x, y)} = \frac{f_{Y,Z}(y, z)}{f_Y(y)} \tag{1a}$$

and

$$\frac{f_{X,Y,Z}(x, y, z)}{f_Y(y)} = \frac{f_{X,Y}(x, y)}{f_Y(y)} \frac{f_{Y,Z}(y, z)}{f_Y(y)}. \tag{1b}$$

The last equation is identical to

$$f_{X,Z|Y}(x, z|y) = f_{X|Y}(x|y) f_{Z|Y}(z|y),$$

which explicitly states that X and Z are independent conditionally on $Y = y$, for each fixed value of y.

2.1 The Hiemstra-Jones test

The HJ test is a modification of the Baek and Brock (1992) test for non-causality, based on asymptotic theory of the test statistic for weakly mixing processes. The test statistic of these tests is the same and based on comparing two ratios of correlation integrals. For a multivariate random vector W the associated correlation integral $C_W(\epsilon)$ is the probability of finding two independent realizations of the vector at a distance smaller than or equal to ϵ:

$$\begin{aligned} C_W(\epsilon) &= P[\|W_1 - W_2\| \le \epsilon], \qquad W_1, W_2 \text{ indep.} \sim W \\ &= \int \int I(\|s_1 - s_2\| \le \epsilon) f_W(s_1) f_W(s_2) \, \mathrm{d}s_2 \, \mathrm{d}s_1 \end{aligned}$$

4

where the integrals are taken over the sample space of W, and $\|x\| = \sup_{i=1,\ldots,d_W} |x_i|$ denotes the supremum norm, with d_W the dimension of the sample space of W. Hiemstra and Jones (1994) argue that Equation (1a) implies for any $\epsilon > 0$:

$$\frac{C_{X,Y,Z}(\epsilon)}{C_{X,Y}(\epsilon)} = \frac{C_{Y,Z}(\epsilon)}{C_Y(\epsilon)} \tag{2a}$$

or equivalently

$$\frac{C_{X,Y,Z}(\epsilon)}{C_Y(\epsilon)} = \frac{C_{X,Y}(\epsilon)}{C_Y(\epsilon)} \frac{C_{Y,Z}(\epsilon)}{C_Y(\epsilon)}. \tag{2b}$$

The HJ test consists of calculating sample versions of the correlation integrals in (2a), and then testing whether the left-hand- and right-hand-side ratios differ significantly or not. The estimators for each of the correlation integrals take the form

$$C_{W,n}(\epsilon) = \frac{2}{n(n-1)} \sum_{i<j} \sum I_{ij}^W,$$

where $I_{ij}^W = I(\|W_i - W_j\| \leq \epsilon)$. For the asymptotic theory we refer to Hiemstra and Jones (1994).

We argue that, although there may be some interest in testing relation (2a), it is not equivalent to testing the null hypothesis (1b). Since it is hard, if not impossible, to derive general closed form analytic expressions for correlation integrals, we focus on the behavior of the fractions in (2a) for small values of the bandwidth ϵ. For continuous distributions the following small ϵ approximation is useful:

$$\begin{aligned}
C_W(\epsilon) &= \int\int I(|s_1 - s_2| \leq \epsilon) f_W(s_1) f_W(s_2) \, \mathrm{d}s_1 \, \mathrm{d}s_2 \\
&= \int \epsilon^{d_W} f_W(s_1) f_W(s_1) \, \mathrm{d}s_1 + \mathcal{O}(\epsilon^{d_W+1}) \\
&= (2\epsilon)^{d_W} \int f_W^2(s) \, \mathrm{d}s + \mathcal{O}(\epsilon^{d_W+1}) \\
&= (2\epsilon)^{d_W} E\left[f_W(W)\right] + \mathcal{O}(\epsilon^{d_W+1}),
\end{aligned} \tag{3}$$

which shows that for small ϵ, the leading term in powers of ϵ is proportional to $E\left[f_W(W)\right]$.

The expectation $E\left[f_W(W)\right] = \int f_W^2(s) \, \mathrm{d}s$ can be considered as a concentration measure of W. To illustrate this, for a family of univariate pdfs with scale parameter θ, that is, $f_W(w;\theta) = \theta^{-1} g(\theta^{-1} w)$ for some pdf $g(\cdot)$, one readily finds

$$\int f_W^2(s;\theta) \, \mathrm{d}s = \frac{1}{\theta} \int g^2(x) \, \mathrm{d}x = \frac{\text{cnst.}}{\theta},$$

5

which shows that, in the univariate case, the concentration measure is inversely proportional to the scale parameter θ. For later convenience we introduce the notation H_W for the concentration measure of the random variable W,

$$H_W = \int f_W^2(w) \, \mathrm{d}w, \tag{4}$$

and $H_W(y)$ for the conditional concentration of W given $Y = y$:

$$H_W(y) = \int f_{W|Y}^2(w|y) \, \mathrm{d}w = \frac{\int f_{W,Y}^2(w, y) \, \mathrm{d}w}{f_Y^2(y)}.$$

By comparing the leading terms of the expansion in powers of ϵ in equations (2b) and (3), we find that

$$\frac{E[f_{X,Y,Z}(X,Y,Z)]}{E[f_Y(Y)]} = \frac{E[f_{X,Y}(X,Y)]}{E[f_Y(Y)]} \frac{E[f_{Y,Z}(Y,Z)]}{E[f_Y(Y)]}. \tag{5}$$

That is, for ϵ small, testing the equivalence of the ratios in (2a) amounts to testing (5) instead of the null hypothesis. Unless some additional conditions hold, this will typically not be equivalent to testing the null hypothesis.

The exact conditions under which (5) is consistent with the null hypothesis are important, since in that case it might still be possible to obtain consistency of the HJ test by letting ϵ tend to zero sufficiently slowly with increasing sample size. We therefore examine the necessary conditions in detail. For the left-hand-side of (5) one can write

$$
\begin{aligned}
\frac{E\left[f_{X,Y,Z}(X,Y,Z)\right]}{E[f_Y(X,Y)]} &= \frac{E_Y\left[E_{X,Z|Y}[f_{X,Z|Y}(X,Z|Y)f(Y)]\right]}{E[f_Y(Y)]} \\
&= \int E_{X,Z|Y=y}[f_{X,Z|Y}(X,Z|y)]w(y) \, \mathrm{d}y \\
&= \int H_{X,Z}(y)w(y) \, \mathrm{d}y,
\end{aligned}
$$

where $w(y)$ is a weight function given by $w(y) = f_Y^2(y)/\int f_Y^2(s) \, \mathrm{d}s$. This brings about the fact that the ratio on the left-hand-side of (5) for small ϵ is proportional to a weighted average of the conditional concentration $H_{X,Z}(y)$, with weight function $w(y)$. In a similar fashion, for the terms on the right-hand-side one derives

$$\frac{E\left[f_{X,Y}(X,Y)\right]}{E[f_Y(Y)]} = \int H_X(y)w(y) \, \mathrm{d}y, \qquad \text{and} \qquad \frac{E\left[f_{Y,Z}(Y,Z)\right]}{E[f_Y(Y)]} = \int H_Z(y)w(y) \, \mathrm{d}y,$$

6

Under the null hypothesis, Z is conditionally independent of X given Y, so that $H_{X,Z}(y)$ is equal to $H_X(y)H_Z(y)$, for all y. It follows that the left and right hand sides of (5) coincide under the null if and only if

$$\int H_X(y)H_Z(y)w(y)\,\mathrm{d}y - \int H_X(y)w(y)\,\mathrm{d}y \int H_Z(y)w(y)\,\mathrm{d}y = 0,$$

or

$$\mathrm{Cov}(H_X(S), H_Z(S)) = 0, \tag{6}$$

where S is a random variable with pdf $w(y)$. Only under specific conditions, such as either $H_X(y)$ or $H_Z(y)$ being independent on y, (5) holds under the null, and hence (2a) as ϵ tends to zero. Also if $H_X(y)$ and $H_Z(y)$ depend on y, (5) may hold, but this is an exception rather than the rule. Typically the covariance will not vanish, inducing a bias in the HJ test for small ϵ.

The fact that the conditional concentration measures of $X_{t-1}^{l_X}$ and Y_t given $Y_{t-1}^{l_Y}$ plays an important role is interesting. In economics and finance, stochastic processes with conditional heteroskedasticity are very common. If the models used to filter out the heteroskedasticity (usually this is done with estimated ARCH or GARCH models) are misspecified, then some covariance between the conditional concentration of $X_{t-1}^{l_X}$ and Y_t may remain, and the sensitivity of the HJ test to this covariance may lead to over-rejection, even after GARCH filtering. We examine such cases in detail in the simulation study described later.

2.2 A modified test statistic

In comparing equations (1b) and (5) it can be noticed that although (1b) holds point-wise for any triple (x, y, z) in the support of $f_{X,Y,Z}(x, y, z)$, (5) contains averages of the terms (1b), which do not respect the fact that the y-values on the rhs of (1b) should be identical. Because (1b) holds point-wise, rather than (5), the null hypothesis implies

$$E\left[\frac{f_{X,Y,Z}(U, V, W)}{f_Y(V)} - \frac{f_{X,Y}(U, V)}{f_Y(V)}\frac{f_{Y,Z}(V, W)}{f_Y(V)}\right] = 0$$

for any triple of random variables (U, V, W) taking values in the support of $f_{X,Y,Z}(x, y, z)$. Taking this equation as a starting point, we can base a test statistic on the functional

$$q = E\left[\left(\frac{f_{X,Y,Z}(X, Y, Z)}{f_Y(Y)} - \frac{f_{X,Y}(X, Y)}{f_Y(Y)}\frac{f_{Y,Z}(Y, Z)}{f_Y(Y)}\right)g(X, Y, Z)\right]$$

where $g(x, y, z)$ is a positive weight function. Under the null hypothesis the term within the round brackets vanishes, so that $q = 0$. Although q is not semi positive definite, a one-sided test, rejecting when its estimated value is too large, in practice is often found to have larger power than a two-sided test. In tests for serial dependence Skaug and Tjøstheim (1993) report good performance of a closely related unconditional test statistic (their dependence measure I_4 is an unconditional version of our term in round brackets).

We have considered several possible choices of the weight function g, being (1) $g_1(x, y, z) = f_Y(y)$, (2) $g_2(x, y, z) = f_Y^2(y)$ and (3) $g_3(x, y, z) = f_Y(y)/f_{X,Y}(x, y)$. Preliminary Monte Carlo simulations (not presented here) using the stationary bootstrap (Politis and Romano, 1994) indicated that g_1 and g_2 behave similarly and are more stable than g_3. We will focus on g_2 in the sequel of this paper. The main advantage of g_2 over g_1 is that the corresponding estimator has a representation as a U-statistic, allowing the asymptotic distribution to be derived straightforwardly for weakly dependent data. For the choice $g(x, y, z) = f_Y^2(y)$, we refer to the corresponding functional as $\tilde{q}$:

$$\tilde{q} = E\left[f_{X,Y,Z}(X, Y)f_Y(Y) - f_{X,Y}(X, Y)f_{Y,Z}(Y, Z)\right].$$

Note that $\tilde{q}$ defined as above, for small ϵ is proportional to $(2\epsilon)^{d_X + 2d_Y + d_Z}$. A natural estimator of $\tilde{q}$ is given by

$$\tilde{q}_n = (2\epsilon)^{-d_X - 2d_Y - d_Z} T_n$$

where

$$T_n = \frac{1}{n(n-1)(n-2)} \sum_i \left[\sum_{k, k \neq i} \sum_{j, j \neq i} \left(I_{ik}^{XYZ} I_{ij}^Y - I_{ik}^{XY} I_{ij}^{YZ}\right)\right],$$

where $I_{ij}^W = I(\|W_i - W_j\| < \epsilon)$. Note that the terms with $k = j$ need not be excluded explicitly as these each contribute zero to the test statistic. The test statistic can be interpreted as an average over local BDS test statistics, for the conditional distribution of X and Z, given $Y = y_i$.

If we define *local correlation integrals* of (X_i, Y_i, Z_i) as

$$C_i^W = \frac{1}{n-1} \sum_{j, j \neq i} I_{ij}^W,$$

the test statistic can be simplified to

$$T_n = \frac{(n-1)}{n(n-2)} \sum_i (C_i^{XYZ} C_i^Y - C_i^{XY} C_i^{YZ}).$$

This representation shows that the test statistic consist of a sum of local contributions, each of which has zero expectation under the null.

Using the approach followed by Hiemstra and Jones (1994) under the same stationarity and weak mixing conditions we obtain:

Theorem 1

$$\sqrt{n} \frac{(T_n - (2\epsilon)^m \tilde{q})}{S_n} \xrightarrow{d} N(0,1).$$

With S_n as given in the appendix, where also the proof is given.

Although the weight function $g_2(x,y,z) = f_Y^2(y)$ is convenient for calculating the test statistic, if $f_Y(y)$ varies considerably in the sample space, the weight function $g_2(x,y,z)$ may put too much weight on those parts of the sample space where the density of y is large, and ignore contributions from regions where the density of Y is low. To avoid a possible associated loss of power, we consider transforming the original time series data to a uniform marginal distribution. In particular when the conditioning variable Y contains only one lagged value, its marginal density becomes uniform under this transform. Pompe (1993) argued for a similar transform of the data to avoid similar problems in his serial independence test. Note that the weight functions $g_1(x,y,z) = f_Y(y)$ and $g_2(x,y,z) = f_Y^2(y)$ become equivalent in the case where Y has a uniform distribution.

3   Simulations

We use numerical simulations to investigate the behavior of the proposed T_n test and to compare it with the HJ test. Relying on the outcomes of the simulations we also develop some recommendations regarding practical implementations of the nonlinear test for Granger causality. The underlying processes for the simulations were chosen from the family of bivariate conditional heteroskedastic processes. The interest in these processes was stipulated by their relevance to

econometrics and financial time series. We start with a first order process $\{Y_t\}$ with conditional heteroskedasticity. The process $\{X_t\}$ is driven by $\{Y_t\}$ via variance. Two cases are considered:

- the first with instantaneous dependence

$$
\begin{aligned}
X_{t-1} &\sim N\left(0, a + bY_{t-1}^2 \exp(-cY_{t-1}^2)\right) \\
Y_t &\sim N\left(0, a + bY_{t-1}^2 \exp(-cY_{t-1}^2)\right).
\end{aligned}
\tag{7}
$$

- and the second with lag one dependence

$$
\begin{aligned}
X_t &\sim N\left(0, a + bY_{t-1}^2 \exp(-cY_{t-1}^2)\right) \\
Y_t &\sim N\left(0, a + bY_{t-1}^2 \exp(-cY_{t-1}^2)\right).
\end{aligned}
\tag{8}
$$

Both specifications (7) and (8) satisfy the null hypothesis: $\{X_t\}$ is not Granger causing $\{Y_t\}$. The values for the coefficients a, b and c are chosen in such a way that process $\{X_t\}$ remains stationary and ergodic (e.g. if $a > 0$, $0 < b < 1$ and $c = 0$, $\{Y_t\}$ boils down to a stable ARCH(1) process). We introduce an exotic exponential term in the conditional variance specification. This is done to reflect the fact that the true conditional variance may take forms that are not adequately captured by traditional (G)ARCH models. To emphasize certain points of interest we chose different values for the coefficients a, b and c. The lags L_x and L_y are set to one for both tests, and we consider one-sided testing, rejecting when the test statistics are too large.

Size distortions

Diks and Panchenko (2004) demonstrated that for the process with instantaneous dependence (7) and coefficients $a = 1$, $b = 0.4$, and $c = 0$, the actual size of the HJ test was severely distorted. The distortion occurs because of the positive covariance between the concentration measures H_X and H_Z as explained in subsection 2.1. Recent simulations (not shown) with the same underlying process revealed that the T_n test statistic exhibits less bias. Here we want to illustrate the same point for the process (8) with the same coefficients but without instantaneous dependence. As before, we set the bandwidth to $\epsilon = 1$, and determined the actual size for a nominal size of 0.05.
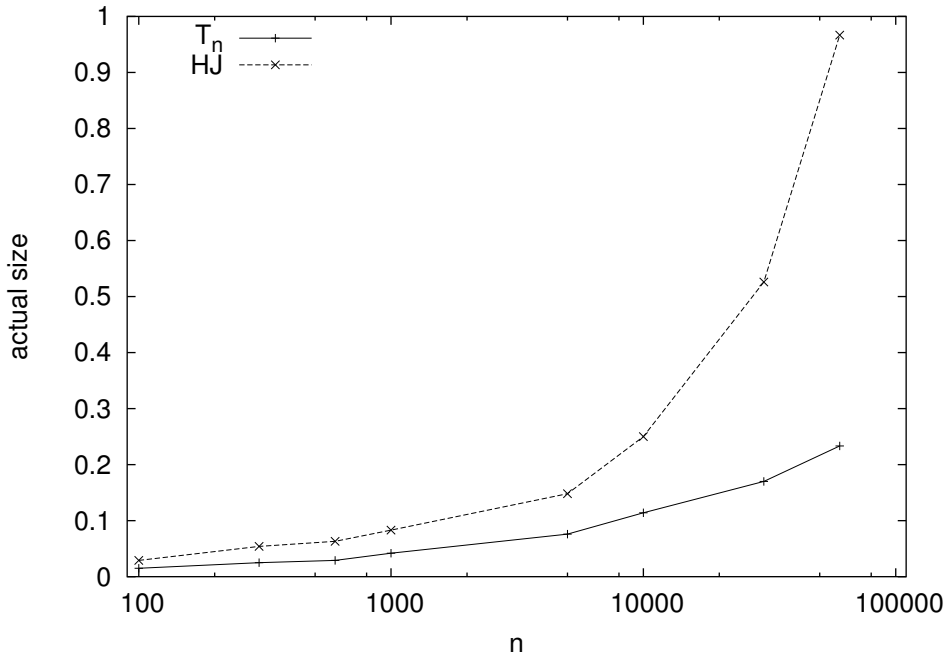
Figure 1: *Simulated size of T_n and the HJ test ($\epsilon = 1$) for the bivariate ARCH process given in (8) as a function of the time series length n (nominal size 0.05). Number of realizations: $1\,000$ for $n < 10\,000$, and $500, 100, 30$ for $n = 10\,000, 30\,000$, and $60\,000$ respectively.*

Figure 1 shows the rejection rates found as a function of the time series length n. The actual size of the T_n based test is close to nominal for series of length $n \leq 1\,000$. It becomes larger for longer series. The size of the HJ test is more heavily distorted. It is close to nominal only for time series length $n = 100$ and is close to one when $n = 60\,000$. In comparison with the process (7) considered in Diks and Panchenko (2004) the current process (8) indicates less size distortion. This is due to the weaker covariance between the concentration measures H_X and H_Z for the current process.

GARCH filtering

One might argue that it is possible to filter out the conditional heteroskedasticity using a univariate (G)ARCH specification. This would remove the bias caused by the conditional heteroskedasticity in the HJ test. However such a filtering procedure has several drawbacks. First, it may
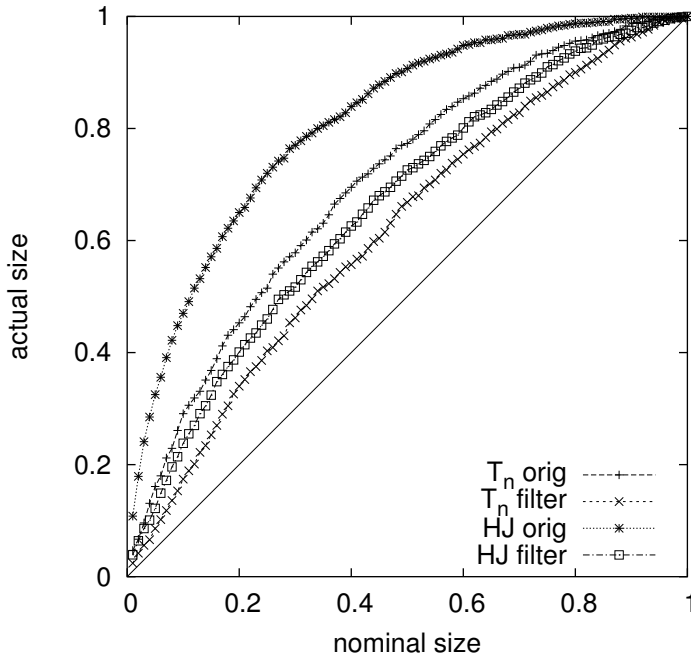
Figure 2: *Size-size plot of T_n and the HJ test ($\epsilon = 1.0$) for the original bivariate process (7) and GARCH$(1, 1)$ filtered series (nominal size 0.05). Time series length $n = 2\,000$ and number of realizations $1\,000$.*

affect the dependence structure and consequently the power of the test. Second, without knowing the precise functional form of the process, a simple (G)ARCH filter may not fully remove the conditional heteroskedasticity in the residuals. To illustrate the latter point we considered a process of the form (7) with coefficients $a = 0.65$, $b = 0.9$ and $c = 0.5$. A univariate GARCH(1,1) model is misspecified for this process. First, we took a long time series $n = 10\,000$ and estimated the parameters of a (misspecified) GARCH$(1, 1)$ model. No significant GARCH structure was found for the process $\{X_t\}$. The parameter estimates for $\{Y_t\}$ were a constant term of 0.8 and an ARCH(1) term of 0.18. The GARCH(1) term was not significantly different from zero. This estimated model was used to obtain GARCH$(1, 1)$ filtered time series. The T_n and HJ tests were applied to the original and filtered time series of length n, with ϵ set to unity.

Figure 2 shows a size-size plot of the original and filtered series. Both tests after (G)ARCH filtering exhibit an actual size closer to the nominal. Remarkably, the size of the T_n test on the original series is very close to the size of the HJ test on the filtered data. Since the GARCH(1,1)

12
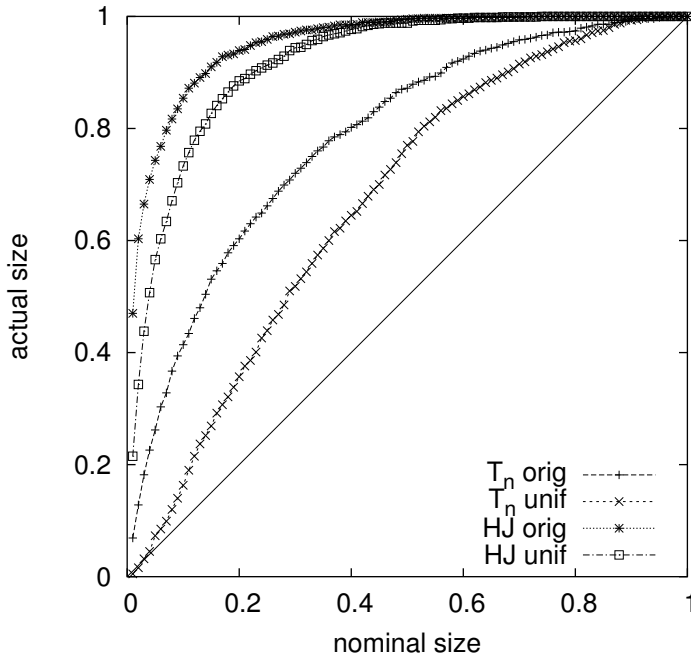
Figure 3: *Size-size plot of T_n and the HJ test ($\epsilon = 1$) for the original process 7 and series transformed to uniform marginals (nominal size 0.05). Time series length $n = 2\,000$ and number of realizations $1\,000$.*

model was not able to capture the complete structure of the conditional variance, both tests still over-reject after filtering. As before, the size distortion for the T_n statistic is found to be less than for the HJ test.

Uniform transformation

As suggested at the end of section 2, transforming the time series to a uniform marginal density, hereafter called uniform transformation, may improve the performance of the test. Here we investigate how it affects the size of the test. We take the process (7) with coefficients $a = 0.65$, $b = 0.2$ and $c = 0.5$. A series length of $n = 2\,000$ was used and the bandwidth set to $\epsilon = 1$.

Figure 3 shows the actual size as a function of the nominal size. The actual size for the transformed series is closer to the nominal size. Therefore, we can conclude that for process (7) with instantaneous dependence, the uniform transformation improves the performance of both tests. Nevertheless, the HJ test still heavily over-rejects.

test	T_n	HJ	T_n	HJ	T_n	HJ	T_n	HJ	T_n	HJ
n	1 000		5 000		10 000		30 000		60 000	
ϵ^*	1.00	0.50	0.70	0.20	0.60	0.15	0.44	0.06	0.35	0.05
size	0.074	0.080	0.057	0.052	0.050	0.065	0.040	0.030	0.033	0.560
power	0.551	0.418	0.994	0.884	1.000	0.944	1.000	0.093	1.000	1.000

Table 1: *Simulated size and power of T_n and the HJ test for bivariate ARCH process (8) as a function of the time series length n and adapted bandwidth ϵ^* (nominal size 0.05). Number of realizations: 1 000 for $n < 10\,000$, and 500, 100, 30 for $n = 10\,000, 30\,000$, and 60 000 respectively.*

We also considered specifications from the class (8), but the uniform transform showed either marginal or no size improvements for both tests in this case.

Bandwidth adaptation

Despite the better general performance of the test based on T_n, its actual size still becomes distorted for long series lengths (see Figure 1). To reduce the bias of T_n we considered an adaptation of the bandwidth ϵ to the time series length n. We conducted an extensive amount of simulations for different values of ϵ and n of 500, 1 000, 5 000 and 10 000. For each n we selected the bandwidth ϵ that produced the size closest to the commonly used nominal size of 0.05. On the basis of the selected ϵ - n correspondences we detected an approximate rate at which the optimal bandwidth ϵ^* should go to zero as n increases:

$$\epsilon^* \simeq kn^{-\alpha}. \tag{9}$$

For the process (8) with the coefficients $a = 1$, $b = 0.4$, $c = 0$ we find the constant term $k \simeq 7.0$ and the rate $\alpha \simeq 0.27$. As we will show later this rate of convergence towards zero as n increases is sufficiently slow for the power of the T_n to remain large. We tried to derive a similar relation for the HJ test. However, the rate $\alpha \simeq 0.6$ (with a constant term $k \simeq 35$) was still not high enough to prevent a bias for large n.

Table 1 reports the T_n and HJ test rejection rates (both size and power) for increasing series

length n and bandwidths equal to ϵ^* given in equation (9), for a nominal size of 0.05. The size computations were based on the process (8) and the coefficients $a = 1$, $b = 0.4$ and $c = 0$. To compute the power we took the same process and reversed the roles of $\{X_t\}$ and $\{Y_t\}$, so that the relation tested became: $\{Y_t\}$ is not Granger causing $\{X_t\}$. Coefficient b was reduced to 0.1 to make the simulation more informative (for higher b power was one in nearly all cases). The actual size of the T_n test is close to the nominal size for all n. The power of the test is not reduced by the bandwidth adaptation. The size of the HJ test remains close to the nominal size for $n \leq 30\,000$ and departs from it for higher n. This suggests that an even faster rate α should be applied. This in turn may reduce the power of the HJ test which is already smaller than the power of T_n test under the adaptive bandwidth procedure.

To demonstrate that the bandwidth adaptation is robust for other coefficients and different values of nominal size we consider the process (8) with a different set of the coefficients $a = 0.65$, $b = 0.2$ and $c = 0.5$ and generate size-size and power-size plots. As before, the power is the rejection rate for the reversed null hypothesis. For series of length $n = 2\,000$, according to equation (9) we find $\epsilon^* = 0.9$ for the T_n test and $\epsilon^* = 0.4$ for the HJ test. Figure 4 suggests that both tests have an actual size close to nominal. In terms of power T_n shows better performance.

Regarding the processes with instantaneous dependence of the form (7) which exhibited even more bias we find that it is necessary to use a higher rate α in the bandwidth adaptation equation (9).

It should be noted that there is related literature on bandwidth selection for nonparametric hypothesis tests. For example, as in Horowitz and Spokoiny (2001) one might consider a new test statistic which is the largest value of the standardized and studentized test statistics for a range of bandwidths, and use bootstrap methods to assess the critical values. Although it is an interesting question whether these techniques can also be developed in the present context, we consider this issue beyond the scope of the this paper.
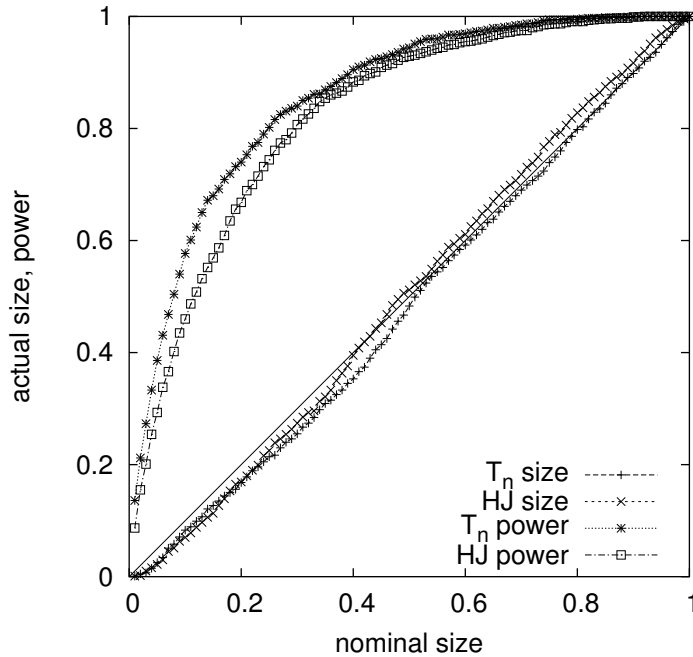
Figure 4: *Size-size and power-size plots of T_n and the HJ test for process (8) ($\epsilon = \epsilon^* = 0.9$) for T_n test, and ($\epsilon = \epsilon^* = 0.4$ for the HJ test, nominal size 0.05). Time series length $n = 2\,000$ and number of realizations $1\,000$.*

Practical recommendations

In summary, by following the procedure of this section we developed practical recommendations regarding the application of the T_n test for Granger causality. The main focus was on processes exhibiting conditional heteroskedasticity, a common stylized fact in financial time series. We observed that GARCH$(1, 1)$ filtering reduces the bias of the test. However, care should be taken since the filtering procedure with an incorrect model specification may be insufficient for preventing a size distortion. If there is an indication of instantaneous dependence between the time series, transformation to uniform marginal densities proved to be useful in making the T_n test more precise. For relatively long times series it is highly desirable to take into account a dependence of the bandwidth on the length of the series n according to equation (9). It was demonstrated that that procedure yields an accurate test without loss of power.

16

4 Applications

We consider an application to daily volume and returns data for the Standard and Poor's 500 index in the period between January 1950 and December 1990. We deliberately have chosen this period to roughly correspond to the period for which Hiemstra and Jones (1994) found strong evidence for volume Granger-causing returns (1947 – 1990) for the Dow Jones index. To keep our results comparable with those of Hiemstra and Jones, we closely followed their procedure. That is, we adjusted for day-of-the-week and month-of-the-year effects on returns and percentage volume changes, using a two-step procedure in which we first adjust for effects in the mean, and subsequently in the variance. The calendar adjusted, standardized, returns and percentage volume change data were used to estimate a linear bivariate VAR model. Finally, the two VAR residual series were EGARCH filtered, and the resulting bivariate EGARCH residual series served as the input for the causality tests presented below.

We applied the HJ test and the T_n test both with and without transforming the data to uniform marginals. Applying the tests with $\epsilon = 1.5$ (the value used by Hiemstra and Jones (1994)). leads to strong indications of a bi-directional causal relationship between volume changes and returns. The evidence for returns affecting volume (not shown) are extremely strong, (p-values smaller than 0.001 at all lags) for each of the two tests, before and after uniform transforms. However, the relation in the other direction (volume Granger causing returns) is more interesting, since evidence of Granger causality in that direction might (but need not) indicate an inefficiency in the market. Table 4 shows the p-values for the four tests at equal lags $l_X = l_Y$ ranging from 1 to 5, for $\epsilon = 1.5$. Clearly, the tests indicate a strong causal relation from relative volume changes to returns.

However, the previous simulation section indicates that, in order to avoid bias problems, one should not choose the bandwidth ϵ too large. For a time series length of $10\,000$ a bandwidth of about 0.15 and 0.6 should be used for the HJ test and T_n, respectively. The p-values calculated for these bandwidth values are also shown in Table 4. The results for the HJ test no longer indicate evidence for volume Granger-causing returns. This might be related to a loss of power of the HJ test under the adaptive bandwidth procedure, as we observed in the simulations. Using

$l_X = l_Y$	HJ	T_n	HJ	HJ (UNIF)	T_n	T_n (UNIF)
	$\epsilon = 1.5$		$\epsilon^* = 0.15$		$\epsilon^* = 0.6$	
1	0.052	0.057	0.229	0.562	0.049	0.032
2	0.001	0.001	0.143	0.597	0.002	0.028
3	0.048	0.041	0.204	0.201	0.024	0.175
4	0.017	0.014	1.000	1.000	0.063	0.198
5	0.012	0.011	—	—	0.344	0.143

Table 2: *P-values for the test of the null hypothesis that volume does not Granger cause returns. Results are shown for the HJ test statistic and for the modified test statistic T_n with bandwidth of $\epsilon = 1.5$, which is of the same order of magnitude as the values used in the literature, and for adaptive bandwidths suggested by the simulation results of the previous section. The dashes indicate a lack of inter-point distances leading to an empirical correlation integral equal to zero in the denominator.*

the adaptive bandwidth, the T_n statistic still indicates some evidence for Granger causality, albeit weaker than before. Table 4 also shows the results for uniform transforms of the data, which are very similar to those obtained without the transforms.

In summary, our findings on the basis of the Standard and Poor's data are consistent with our simulation based result that a larger value of the bandwidth increases power, but at the cost of an increased bias. If the tests are performed for bandwidth values for which the simulations suggested that the actual size remains approximately equal to nominal, the evidence for relative volume changes Granger causing returns becomes weaker for T_n and even vanishes for the HJ test.

5   Concluding Remarks

By analyzing the HJ test in detail we find that it is biased even as the bandwidth tends to zero. This shows that there are problems in interpreting the HJ test as a nonparametric test for Granger causality. Based on the analysis we proposed a new test statistic T_n that does not suffer from this problem. By symmetrizing the test statistic, it can be expressed as a U-statistic for which we

obtained the asymptotic theory for fixed bandwidth values. The simulations section presented a large number of size ans power calculations for processes with autoregressive heteroskedasticity of a nonstandard form. The simulations showed that the fixed bandwidth T_n test, like the HJ test, over-rejects for large sample sizes, albeit to a lesser extent. We show that an adapted bandwidth choice which tends to zero as the sample size increases, reduces the size distortion while retaining large power. In an application to relative volume changes and returns for historic Standard and Poor's index data we found that an initial naive application of the HJ test or the T_n test would lead to strong rejection of the null hypothesis that volume changes are not Granger causing returns. However, application of our recommendations based on the simulation study strongly weakens the evidence against the null hypothesis, which suggests that the strong rejections initially found may be spurious.

A Asymptotic distribution of T_n

The test statistic T_n can be written in terms of a U-statistic by symmetrization with respect to the three different indices. This gives

$$T_n = \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k \neq i} K(W_i, W_j, W_k)$$

with $W_i = (X_i^{l_X}, Y_i^{l_Y}, Z_i)$, $i = 1, \ldots, n$ and

$$K(W_i, W_j, W_k) = \frac{1}{6} \left(\begin{array}{l} \left(I_{ik}^{XYZ} I_{ij}^Y - I_{ik}^{XY} I_{ij}^{YZ}\right) + \left(I_{ij}^{XYZ} I_{ik}^Y - I_{ij}^{XY} I_{ik}^{YZ}\right) + \\ \left(I_{jk}^{XYZ} I_{ji}^Y - I_{jk}^{XY} I_{ji}^{YZ}\right) + \left(I_{ji}^{XYZ} I_{jk}^Y - I_{ji}^{XY} I_{jk}^{YZ}\right) + \\ \left(I_{ki}^{XYZ} I_{kj}^Y - I_{ki}^{XY} I_{kj}^{YZ}\right) + \left(I_{kj}^{XYZ} I_{ki}^Y - I_{kj}^{XY} I_{ki}^{YZ}\right) \end{array} \right)$$

Notice that, in contrast with the correlation integral, which is a second order U-statistic, T_n is a third order U-statistic. According to Denker and Keller (1983), T_n is asymptotically $N(\mu, \frac{\sigma^2}{n})$ distributed, with $\mu = (2\epsilon)^{d_X + 2d_Y + d_Z} \tilde{q}$ and

$$\sigma^2 = 9 \left[\text{Var}(h(W_1)) + 2 \sum_{k \geq 2} \text{Cov}(h(W_1), h(W_{1+k})) \right]$$

with

$$h(w) = \int \int K(w, w_1, w_2) f_W(w_1) f_W(w_2) \, dw_1 \, dw_2.$$

If we estimate $h(W_i)$ as

$$\widehat{h}_i = \frac{1}{(n-1)(n-2)} \sum_{j, j \neq i} \sum_{k, k \neq i} K(W_i, W_j, W_k),$$

an autocorrelation consistent estimator for σ^2 is given by (Newey and West, 1987):

$$S_n^2 = \sum_{k=1}^K R_k \omega_k,$$

where $R_k = \frac{1}{n-k} \sum_{i=1}^{n-k} (\widehat{h}_i - T)(\widehat{h}_{i+k} - T)$ is the sample autocovariance function of $h(W_i)$, and ω_k as in Hiemstra and Jones (1994). It follows that

$$\sqrt{n} \frac{(T_n - (2\epsilon)^m \tilde{q})}{S_n} \xrightarrow{d} N(0, 1),$$

which proves Theorem 1.

Although T_n is a third order U-statistic, both T_n and the asymptotic variance S_n^2 can be determined in $\mathcal{O}(n^2)$ computational time. For each i, the calculation of the C_i^W's and the I_{ij}^W's is $\mathcal{O}(n)$. A second $\mathcal{O}(n)$ calculation then provides the $\widehat{h}_i$ through

$$\widehat{h}_i = \frac{1}{3}\left(C_i^{XYZ}C_i^Y - C_i^{XY}C_i^{YZ}\right) + \frac{1}{3n}\sum_j\left(C_j^{XYZ}I_{ij}^Y + I_{ij}^{XYZ}C_j^Y - C_j^{XY}I_{ij}^{YZ} - I_{ij}^{XY}C_j^{YZ}\right),$$

a result which follows from straightforward calculation from the definition of $\widehat{h}_i$. C-code can be obtained from the authors upon request.

References

Abhyankar, A. (1998). Linear and non-linear granger causality: Evidence from the U.K. stock index futures market. *Journal of Futures Markets*, **18,** 519–540.

Baek, E. and Brock, W. (1992). A general test for granger causality: Bivariate model. Technical Report. Iowa State University and University of Wisconsin, Madison.

Bell, D., Kay, J. and Malley, J. (1996). A non-parametric approach to non-linear causality testing. *Economics Letters*, **51,** 7–18.

Brooks, C. (1998). Predicting stock index volatility: Can market volume help? *Journal of Forecasting*, **17,** 59–80.

Chen, A.-S. and Lin, J. W. (2004). Cointegration and detectable linear and nonlinear causality: analysis using the London Metal Exchange lead contract. *Applied Economics*, **36,** 1157–1167.

Ciner, C. (2001). Energy shocks and financial markets: Nonlinear linkages. *Studies in Nonlinear Dynamics and Econometrics*, **5,** number 3, 203–212.

Denker, M. and Keller, G. (1983). On *U*-statistics and v. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **64,** 505–522.

Diks, C. and Panchenko, V. (2004). A note on the Hiemstra-Jones test for Granger non-causality. *Studies in Nonlinear Dynamics and Econometrics*; forthcoming.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, **49,** number 5, 1639–1664.

Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive rate-optimal test of a parametric mean-regression model against a non-parametric alternative. *Econometrica*, **69,** 599–631.

Ma, Y. and Kanas, A. (2000). Testing for a nonlinear relationship among fundamentals and the exchange rates in the erm. *Journal of International Money and Finance*, **19,** 135–152.

Newey, W. and West, K. (1987). A simple, positive semi-definite, heteroskedasticity and auto-correlation consistent covariance matrix. *Econometrica*, **55,** 703–708.

Okunev, J., Wilson, P. and Zurbruegg, R. (2000). The causal relationship between real estate and stock markets. *Journal of Real Estate Finance and Economics*, **21,** number 3, 251–261.

Okunev, J., Wilson, P. and Zurbruegg, R. (2002). Relationships between australian real estate and stock market prices - a case of market inefficiency. *Journal of Forecasting*, **21,** 181–192.

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

Pompe, B. (1993). Measuring statistical dependences in time series. *Journal of Statistical Physics*, **73,** 587–610.

Silvapulla, P. and Moosa, I. A. (1999). The relationship between spot and futures prices: Evidence from the crude oil market. *Journal of Futures Markets*, **19,** 157–193.

Skaug, H. J. and Tjøstheim, D. (1993). Nonparametric tests of serial independence. In *Developments in time series analysis* (ed. T. Subba Rao), chapter 15. Chapman and Hall, London.

Su, L. and White, H. (2003). A nonparametric Hellinger metric test for conditional independence. Technical Report. Department of Economics, UCSD.