Journal of Econometrics 180 (2014) 251–264

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Nonparametric estimation and inference for conditional density based Granger causality measures☆

Abderrahim Taamouti [a,*], Taoufik Bouezmarni [b], Anouar El Ghouch [c]

[a] Economics Department, Universidad Carlos III de Madrid, Spain
[b] Département de mathématiques, Université de Sherbrooke, Canada
[c] ISBA, Université Catholique de Louvain, Belgium

ARTICLE INFO

Article history:
Received 2 April 2012
Received in revised form
3 March 2014
Accepted 3 March 2014
Available online 15 March 2014

JEL classification:
C12
C14
C15
C19
G1
G12
E3
E4

Keywords:
Causality measures
Nonparametric estimation
Time series
Bernstein copula density
Local bootstrap
Exchange rates
Volatility index
Dividend–price ratio
Liquidity stock returns

ABSTRACT

We propose a nonparametric estimation and inference for conditional density based Granger causality *measures* that *quantify* linear and nonlinear Granger causalities. We first show how to write the causality measures in terms of *copula densities*. Thereafter, we suggest consistent estimators for these measures based on a consistent nonparametric estimator of copula densities. Furthermore, we establish the asymptotic normality of these nonparametric estimators and discuss the validity of a local smoothed bootstrap that we use in finite sample settings to compute a bootstrap bias-corrected estimator and to perform statistical tests. A Monte Carlo simulation study reveals that the bootstrap bias-corrected estimator behaves well and the corresponding test has quite good finite sample size and power properties for a variety of typical data generating processes and different sample sizes. Finally, two empirical applications are considered to illustrate the practical relevance of nonparametric causality measures.

1. Introduction

Much research has been devoted to building and applying tests of non-causality. However, once we have concluded that a "causal relation" (in the sense of Granger) is present, it is usually important to assess the strength of this relationship. Only few papers have been proposed to measure the causality between random variables. Furthermore, although the concept of causality is naturally defined in terms of conditional distributions, the estimation of the existing causality measures has been done using parametric mean regression models in which the causal relations are linear. Consequently, one simply cannot use the existing measures to quantify the strength of nonlinear causalities. The present paper aims to propose a *nonparametric* estimation and inference for Granger causality measures. The proposed approach is model-free

☆ The authors thank two anonymous referees, an Associate Editor, and the Editor Peter Robinson for several useful comments. We would also like to thank the seminar participants at Humboldt University of Berlin and Institut de statistique, Biostatistique et science actuarielles (ISBA) at Université catholique de Louvain for their excellent comments. Taamouti acknowledges financial support from the Spanish Ministry of Education through grants SEJ 2007-63098 and ECO2012-19357. The research work of Bouezmarni is supported by NSERC of Canada. El Ghouch acknowledges financial support from IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

* Corresponding author. Tel.: +34 91 6249863; fax: +34 91 6249875.
E-mail addresses: ataamout@eco.uc3m.es (A. Taamouti), Taoufik.Bouezmarni@usherbrooke.ca (T. Bouezmarni), Anouar.Elghouch@uclouvain.be (A. El Ghouch).

http://dx.doi.org/10.1016/j.jeconom.2014.03.001
0304-4076/

and allows us to quantify nonlinear causalities and the causalities that show up in conditional quantiles as well as higher order conditional moments (such as volatilities, skewness, kurtosis, etc.).

The concept of causality introduced by Wiener (1956) and Granger (1969) constitutes a basic notion for studying dynamic relationships between time series. This concept is defined in terms of predictability at horizon one of a (vector) variable Y from its own past, the past of another (vector) variable X, and possibly a vector Z of auxiliary variables. The theory of Wiener–Granger causality has generated a considerable literature; for review see Dufour and Taamouti (2010). Wiener–Granger analysis distinguishes between three basic types of causality: from Y to X, from X to Y, and instantaneous causality. In practice, it is possible that all three causality relations coexist, hence the importance of finding means to quantify their degree. Unfortunately, causality tests fail to accomplish this task, because they only provide evidence on the presence or the absence of causality, and statistical significance depends on the available data and test power. A large effect may not be statistically significant (at a given level), and a statistically significant effect may not be "large" from an economic viewpoint (or more generally from the viewpoint of the subject at hand) or relevant for decision making. Hence, it is crucial to distinguish between the numerical value of a parameter and its statistical significance (see McCloskey and Ziliak (1996)).

Thus, beyond accepting or rejecting non-causality hypotheses – which state that certain variables do not help forecasting other variables – we wish to assess the magnitude of the forecast improvement, where the latter is defined in terms of some loss function (Kullback distance). Even if the hypothesis of no improvement (non-causality) cannot be rejected from looking at the available data (for example, because the sample size or the structure of the process does allow for high test power), sizeable improvements may remain consistent with the same data. Or, by contrast, a statistically significant improvement – which may easily be produced by a large data set – may not be relevant from a practical viewpoint.

The topic of measuring the causality has attracted much less attention. Geweke (1982, 1984b) introduced measures of causality based on mean-square forecast errors. Gouriéroux et al. (1987) proposed causality measures based on the Kullback information criterion and provided a *parametric* estimation for their measures. Polasek (1994, 2002) showed how causality measures can be computed using the Akaike Information Criterion (*AIC*) and a Bayesian approach. Dufour and Taamouti (2010) proposed short and long run causality measures based on vector autoregressive and moving average models. The estimation of most existing causality measures has been done based on parametric mean regression models. However, the misspecification of parametric model may affect the structure of the causality between the variables of interest. In addition, the dependence in the mean-regression is only due to the mean dependence, and thus it ignores the dependence that show up in conditional quantiles as well as higher order conditional moments. Finally, as shown in many theoretical and empirical papers, several "causal relations" are nonlinear; see for example Gabaix et al. (2003), Bouezmarni et al. (2012) and Bouezmarni and Taamouti (2011), and references therein. Hence, the existing estimation methods for causality measures cannot be used to quantify nonlinear causalities. An exception is the paper of Zheng et al. (2012) who study linear and nonlinear strength of dependence without making any parametric assumptions on the data. However, their approach only focuses on the dependence in the mean, whereas our approach deals with any type of dependence.

We propose a nonparametric estimator for Granger causality measures that quantify nonlinear causalities and causalities that show up in higher order conditional moments. The nonparametric estimator is model-free and therefore it does not require the specification of the model linking the variables of interest. We write the theoretical Granger causality measures in terms of copula densities. Copula is a tool that fully quantifies the dependence among the variables of interest, and thus it can be used to characterize the conditional probability density based Granger causality that we consider in this paper. So, it seems natural to define the measures of Granger causality in distribution using copulas. An advantage of such an approach is that it allows us to completely separate the marginal structure from the dependence structure. As noted by Chen and Fan (2006), separate modeling of the temporal dependence and the marginal behavior is particularly important when the dependence structure and the marginal properties of a time series are affected by different exogenous variables.

Thereafter, the causality measures are estimated by replacing the unknown copula densities by their nonparametric estimates. The copula densities are estimated nonparametrically using Bernstein polynomials. For i.i.d. data, Sancetta and Satchell (2004) show that, under some regularity conditions, any copula can be represented by a Bernstein copula. Bouezmarni et al. (2010) provide the asymptotic properties of the Bernstein copula density estimator for dependent data. The nonparametric Bernstein copula density estimates are guaranteed to be non-negative. Since the causality measures are defined using the Kullback distance, the non-negativity of the Bernstein estimators avoids having negative values inside the logarithmic function. Furthermore, there is no boundary bias problem when we use the Bernstein estimator, because by smoothing with beta densities the Bernstein copula density does not assign weights outside its support. Chen and Huang (2007) propose a *bivariate* kernel copula estimator based on local linear kernels that also removes the boundary bias. For the review of how to remove boundary bias in nonparametric estimation, see for example Brown and Chen (1999) and Chen (2000).

We establish the asymptotic normality of the proposed nonparametric estimator. This result is used to build tests for the statistical significance of causality measures. The asymptotic normality is achieved by subtracting some bias terms and then rescale the estimator by the proper variance. We also discuss the validity of local smoothed bootstrap that we use in finite sample settings to compute a bootstrap bias-corrected estimator and to perform statistical test for Granger causality measures. A Monte Carlo simulation study reveals that the bootstrap bias-corrected estimator behaves well and that the test has good power for a variety of typical data generating processes and different sample sizes.

Finally, the empirical importance of measuring nonlinear causalities is illustrated. In a first empirical application we quantify the causality between S&P500 Index returns and many exchange rates (US/Canada, US/UK and US/Japan exchange rates). We find that both exchange rates and stock prices could have a significant impact on each other. We also find that the impact of stock returns on exchange rates is much stronger than the impact of exchange rates on stock returns. In a second application we compare the predictive content of dividend–price ratio, volatility index (VIX) and liquidity factor for stock market returns. The results show that both dividend–price ratio and VIX help to predict stock market returns. The comparison of causality measure estimates indicates that VIX has more predictive content than dividend–price ratio. We also find that liquidity factor of Pastor and Stambaugh (2003) does not help to predict the time-series of stock returns.

The plan of the paper is as follows. Section 2 provides the motivation for considering a nonparametric causality measures. Sections 3 and 4 present the theoretical framework which underlies the definitions of causality measures using probability and copula density functions. In Section 5 we introduce a consistent nonparametric estimator of causality measures based on Bernstein polynomial. We also establish the asymptotic distribution of our estimator and discuss the asymptotic validity of a local bootstrap finite sample test. In Section 6 we extend our results to the case where the

random variables of interest are multivariate. In Section 7 we propose a bootstrap bias-corrected estimator of causality measures and provide a simulation exercise to evaluate the bias-correction and investigate the finite sample properties of local bootstrap-based test for causality measures. Section 8 is devoted to two empirical applications and the conclusion relating to the results is given in Section 9. Proofs appear in the Appendix.

2. Motivation

The causality measures that we consider here constitute a generalization of those developed by Geweke (1982, 1984b) and others. The existing measures quantify the effect of one variable Y on another variable X assuming that the regression function linking the two variables of interest is known and *linear*. Furthermore, these measures focus on quantifying the causality in mean and ignore the causalities that show up in conditional quantiles as well as higher order conditional moments (such as volatilities, skewness, kurtosis, etc.). Hence, the significance of such measures is limited in the presence of unknown regression functions and in the presence of nonlinear and high-order moment causalities.

We propose measures of Granger causality between random variables based on copula densities. Such measures detect and quantify the causalities in quantiles as well as higher order conditional moments. To see the importance of such causality measures, consider the following examples.

Example 1. Assume that the joint process $(X, Y)'$ follows a stationary VAR(1) model:

$$\begin{bmatrix} X_{t+1} \\ Y_{t+1} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.0 \\ 0.4 & 0.35 \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} + \underbrace{\begin{bmatrix} \varepsilon^X_{t+1} \\ \varepsilon^Y_{t+1} \end{bmatrix}}_{\varepsilon_{t+1}}, \tag{1}$$

where

$$\varepsilon_{t+1} \mid X_t, Y_t \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma^2_{X,t} & 0 \\ 0 & \sigma^2_y \end{bmatrix} \right)$$

with

$$\sigma^2_{X,t} = 0.01 + 0.5Y_t^2 + 0.25X_t^2.$$

Since the coefficient of Y_t in the first equation of (1) is zero, we can conclude that Y does not Granger cause X in the mean. However, if consider the causality in the variance we get

$$V(X_{t+1} \mid X_t, Y_t) = 0.01 + 0.5Y_t^2 + 0.25X_t^2,$$

where now Y Granger causes X in the variance. This example illustrates the case where the causality in the variance does exist even if there is no causality in the mean. But, how can we measure the degree of the causality in the variance? Existing measures do not answer this question.

Example 2. Suppose now X is given by the following process:

$$X_{t+1} = \mu_X + 0.5X_t + \varepsilon^X_{t+1},$$

where the error term ε^X_{t+1} follows Lévy skew stable probability distribution defined by the Fourier transform of its characteristic function $\varphi(u)$:

$$\varepsilon^X_{t+1} \mid X_t, Y_t \sim f\left(\varepsilon^X_{t+1} \mid X_t, Y_t\right) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(\beta_t, u) e^{-iu\varepsilon^X_{t+1}} du,$$

where $\varphi(\beta_t, u) = \exp\left[- \mid u \mid \left(1 + i \frac{2\beta_t \, sgn(u)}{\pi} \log(u)\right)\right]$, $sgn(u)$ is the sign of u and β_t is the time-varying skewness that depends on Y:

$$\beta_t = \lambda + \rho Y_t.$$

In this model, Y does not affect the mean and variance of X, but it does affect its skewness. Again, how can we measure the degree of the causality in skewness? Existing measures do not answer this question.

3. Granger causality measures

Let $\left\{(X_t, Y_t) \in \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2, t = 0, \ldots, T\right\}$ be a sample of stationary stochastic process in $\mathbb{R}^2$, with joint distribution function F_{XY} and density function f_{XY}. For simplicity of exposition, here we consider *univariate Markov processes of order one*. Later, see Section 6, we will extend the results to the case where the variables of interest X and Y can be *multivariate Markov processes of order p*, for $p \geq 1$.

Following Gouriéroux et al. (1987), we define the measure of Granger causality from X to Y by[1]

$$C(X \to Y) = E\left\{ \log\left(\frac{f(Y_t \mid Y_{t-1}, X_{t-1})}{f(Y_t \mid Y_{t-1})} \right) \right\}, \tag{2}$$

where the expected value is taken on the joint distribution of X_{t-1}, Y_{t-1} and Y_t. Important properties of this measure include: **(1)** is nonnegative, and **(2)** cancels only when there is no causality. The property **(2)** implies that the null hypothesis of Granger non-causality from X to Y,

$$H_0^{x \to y} : f(y_t \mid x_{t-1}, y_{t-1}) = f(y_t \mid y_{t-1}), \quad 1 \leq t \leq T, \tag{3}$$

is equivalent to $C(X \to Y) = 0$. Thus, the measure $C(X \to Y)$ defines a "distance" between the left and right hand sides of the null $H_0^{x \to y}$. It quantifies the difference between the conditional densities $f(y_t \mid x_{t-1}, y_{t-1})$ and $f(y_t \mid y_{t-1})$. Hence, high values of measure $C(X \to Y)$ will be interpreted as indicating "strong" causality from X to Y. The measure of Granger causality from Y to X, $C(Y \to X)$, is defined similarly. Moreover, the instantaneous Granger *non-causality* between X and Y can also be characterized in terms of probability density functions using the equivalent null hypotheses:

$$\begin{aligned} H_0^{x \leftrightarrow y} &: f(y_t \mid x_t, x_{t-1}, y_{t-1}) = f(y_t \mid x_{t-1}, y_{t-1}), \quad 1 \leq t \leq T, \\ H_0^{x \leftrightarrow y} &: f(x_t \mid y_t, x_{t-1}, y_{t-1}) = f(x_t \mid x_{t-1}, y_{t-1}), \quad 1 \leq t \leq T, \\ H_0^{x \leftrightarrow y} &: f(y_t, x_t \mid x_{t-1}, y_{t-1}) = f(y_t \mid x_{t-1}, y_{t-1}) f(x_t \mid x_{t-1}, y_{t-1}), \\ &\qquad 1 \leq t \leq T. \end{aligned} \tag{4}$$

From the third null hypothesis $H_0^{x \leftrightarrow y}$, the instantaneous Granger non-causality between X and Y can be viewed as the conditional independence between X_t and Y_t conditional to the past X_{t-1} and Y_{t-1}. Thus, the instantaneous causality between X and Y can be quantified using the following equivalent measures:

$$\begin{aligned} C(Y \longleftrightarrow X) &= E\left\{ \log\left(\frac{f(Y_t \mid X_t, X_{t-1}, Y_{t-1})}{f(Y_t \mid X_{t-1}, Y_{t-1})} \right) \right\} \\ &= E\left\{ \log\left(\frac{f(X_t \mid Y_t, X_{t-1}, Y_{t-1})}{f(X_t \mid X_{t-1}, Y_{t-1})} \right) \right\} \\ &= E\left\{ \log\left(\frac{f(Y_t, X_t \mid X_{t-1}, Y_{t-1})}{f(Y_t \mid X_{t-1}, Y_{t-1}) f(X_t \mid X_{t-1}, Y_{t-1})} \right) \right\}. \end{aligned} \tag{5}$$

Finally, observe that:

$$\begin{aligned} &C(X \to Y) + C(Y \to X) + C(Y \longleftrightarrow X) \\ &\quad = C(Y, X) = E\left\{ \log\left(\frac{f(Y_t, X_t \mid X_{t-1}, Y_{t-1})}{f(X_t \mid X_{t-1}) f(Y_t \mid Y_{t-1})} \right) \right\}, \end{aligned} \tag{6}$$

[1] Details of the derivation of Granger causality measure in (2) can be found in Sections 4.b and 4.c of Gouriéroux et al. (1987).

where the right-hand side of Eq. (6) defines a measure of *dependence* between X and Y, denoted by $C(Y, X)$. The measure of dependence $C(Y, X)$ decomposes the dependence between X and Y to measures of *feedbacks* (time-lagged causal effect) from X to Y ($C(X \to Y)$) and from Y to X ($C(Y \to X)$) and a measure of instantaneous causality between X and Y ($C(Y \longleftrightarrow X)$). This measure will enable one to check whether the processes X and Y must be considered together or whether they can be treated separately.

4. Copula-based Granger causality measures

Here we show that the above Granger causality measures can be rewritten in terms of copula densities. This will allow us to keep only the terms that involve the dependence among the random variables.

It is well known from Sklar (1959) that the distribution function of any joint process $(U, V, W) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ can be expressed via a copula

$$F(u, v, w) = C\left(F_U(u), F_V(v), F_W(w)\right), \tag{7}$$

where $F_Q(.)$, for $Q = U, V, W$, is the marginal distribution function of the random variable Q, and $C\left(F_U(.), F_V(.), F_W(.)\right)$ is a copula function defined on $[0, 1]^3$ which captures the dependence of (U, V, W). If we differentiate Eq. (7) with respect to (u, v, w), we obtain the joint density function of (U, V, W):

$$f(u, v, w) = f_U(u) \times f_V(v) \times f_W(w) \times c(F_U(u), F_V(v), F_W(w)), \tag{8}$$

where $f_Q(.)$, for $Q = U, V, W$, is the marginal density of random variable Q and $c\left(F_U(.), F_V(.), F_W(.)\right)$ is the copula density of (U, V, W) defined on $[0, 1]^3$.

Using Eq. (8), the measure of causality in (2) can be rewritten in terms of copula densities as follows:

$$C(X \to Y) = E\left\{\log\left(\frac{c(F_{Y_t}(Y_t), F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1}))}{c(F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1}))c(F_{Y_t}(Y_t), F_{Y_{t-1}}(Y_{t-1}))}\right)\right\}, \tag{9}$$

where $c\left(F_{Y_t}(.), F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$, $c\left(F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$ and $c\left(F_{Y_t}(.), F_{Y_{t-1}}(.)\right)$ are the copula densities of (Y_t, Y_{t-1}, X_{t-1}), (Y_{t-1}, X_{t-1}), and (Y_t, Y_{t-1}), respectively. The measure of causality from Y to X, $C(Y \to X)$, can similarly defined in terms of copula densities $c\left(F_{X_t}(.), F_{X_{t-1}}(.), F_{Y_{t-1}}(.)\right)$, $c\left(F_{X_{t-1}}(.), F_{Y_{t-1}}(.)\right)$ and $c\left(F_{X_t}(.), F_{X_{t-1}}(.)\right)$.

Finally, the first measure in (5) of the instantaneous Granger causality between X and Y can be rewritten in terms of copula densities as in Box I. We can, in a similar way, rewrite the last two measures in (5) of the instantaneous causality in terms of copula densities.

5. Estimation and inference

Since we are interested in time series data, we need to specify the dependence in the process of interest. In what follows, we consider β-mixing dependent variables. The β-mixing condition is required to show the asymptotic normality of the nonparametric estimator of our causality measures. To establish the asymptotic normality, we also need to apply the results of Bouezmarni et al. (2010, 2012). Now let us recall the definition of a β-mixing process (see e.g., Doukhan (1994), Fan and Yao (2003), among others). For $\left\{\mathcal{W}_t = (X_t, Y_t)' ; t \geq 0\right\}$ a strictly stationary stochastic process and $\mathcal{F}_t^s$ a sigma algebra generated by $(\mathcal{W}_s, \ldots, \mathcal{W}_t)$ for $s \leq t$, the process $\mathcal{W}$ is called β-mixing or absolutely regular, if

$$\beta(l) = \sup_{s \in \mathbb{N}} \mathbb{E}\left[\sup_{\mathcal{A} \in \mathcal{F}_{s+l}^{+\infty}} \left|P\left(\mathcal{A} \middle| \mathcal{F}_{-\infty}^{s}\right) - P(\mathcal{A})\right|\right] \to 0,$$

a.s. $l \to \infty$.

Although the β-mixing condition is required to show the asymptotic normality of our nonparametric estimator (see Tenreiro (1997) and Fan and Li (1999)), the consistency of this estimator can be established under α-mixing condition.

5.1. Estimation

We have shown in Section 4 that Granger causality measures can be rewritten in terms of copula densities. Thus, these measures can be estimated by replacing the unknown copula densities by their nonparametric estimates from a finite sample. Hereafter, we focus on the estimation of Granger causality measure from X to Y, $C(X \to Y)$, which is defined in (9). However, we can similarly propose estimators of measures of Granger causality from Y to X and of the instantaneous causality between X and Y defined in Box I.

To estimate $C(X \to Y)$, we first need to estimate the copula densities $c\left(F_{Y_t}(.), F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$, $c\left(F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$ and $c\left(F_{Y_t}(.), F_{X_{t-1}}(.)\right)$. This can be done using the Bernstein copula density estimators defined below and studied in Bouezmarni et al. (2010). We first set the following additional notations. We denote by

$$G_t = (G_{t1}, G_{t2}, G_{t3}) = (F_{Y_t}(Y_t), F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1}))$$

and its empirical analog

$$\hat{G}_t = (\hat{G}_{t1}, \hat{G}_{t2}, \hat{G}_{t3}) = (F_{Y_t,T}(Y_t), F_{Y_{t-1},T}(Y_{t-1}), F_{X_{t-1},T}(X_{t-1})),$$

where $F_{Y_t,T}(.)$, $F_{Y_{t-1},T}(.)$, and $F_{X_{t-1},T}(.)$ with subscript T is to indicate the empirical analog of the distribution functions $F_{Y_t}(.)$, $F_{Y_{t-1}}(.)$, and $F_{X_{t-1}}(.)$, respectively. The Bernstein copula density estimator of $c\left(F_{Y_t}(.), F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$ at a given value $g = (g_1, g_2, g_3)$ is defined by

$$\hat{c}(g_1, g_2, g_3) = \hat{c}(g) = \frac{1}{T}\sum_{t=1}^{T} K_k(g, \hat{G}_t), \tag{11}$$

where

$$K_k(g, \hat{G}_t) = k^3 \sum_{k_1=0}^{k-1}\sum_{k_2=0}^{k-1}\sum_{k_3=0}^{k-1} A_{\hat{G}_t,k} \prod_{j=1}^{3} p_{k_j}(g_j),$$

the integer k represents the bandwidth parameter, $p_{k_j}(g_j)$ is the binomial distribution

$$p_{k_j}(g_j) = \binom{k-1}{k_j} g_j^{k_j}(1-g_j)^{k-k_j-1}, \quad \text{for } k_j = 0, \ldots, k-1,$$

and $A_{\hat{G}_t,k}$ is an indicator function

$$A_{\hat{G}_t,k} = \mathbf{1}_{\left\{\hat{G}_t \in B_k\right\}},$$

$$\text{with } B_k = \left[\frac{k_1}{k}, \frac{k_1+1}{k}\right] \times \left[\frac{k_2}{k}, \frac{k_2+1}{k}\right] \times \left[\frac{k_3}{k}, \frac{k_3+1}{k}\right].$$

The Bernstein estimators $\hat{c}\left(F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$ and $\hat{c}\left(F_{Y_t}(.), F_{X_{t-1}}(.)\right)$ of $c\left(F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$ and $c\left(F_{Y_t}(.), F_{X_{t-1}}(.)\right)$, respectively, are defined similarly. Observe that the kernel $K_k(g, \hat{G}_t)$ can be rewritten as

$$K_k(g, \hat{G}_t) = \sum_{k_1=0}^{k-1}\sum_{k_2=0}^{k-1}\sum_{k_3=0}^{k-1} A_{\hat{G}_t,k} \prod_{j=1}^{3} \mathcal{B}(g_j, k_j+1, k-k_j),$$

where $\mathcal{B}(g_j, k_j+1, k-k_j)$ is a beta density with shape parameters k_j+1 and $k-k_j$ evaluated at g_j. The Bernstein copula density estimator in (11) is easy to implement, non-negative, integrates to one and is free from the boundary bias problem which often occurs with conventional nonparametric kernel estimators; see Sancetta and Satchell (2004) in the case of i.i.d. data and Bouezmarni et al. (2010) for α-mixing data. Bouezmarni et al. (2010) establish the

$$C(X \leftrightarrow Y) = E\left\{\log\left(\frac{c\left(F_{Y_t}(Y_t), F_{X_t}(X_t), F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1})\right) c\left(F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1})\right)}{c\left(F_{X_t}(X_t), F_{X_{t-1}}(X_{t-1}), F_{Y_{t-1}}(Y_{t-1})\right) c\left(F_{Y_t}(Y_t), F_{X_{t-1}}(X_{t-1}), F_{Y_{t-1}}(Y_{t-1})\right)}\right)\right\} \tag{10}$$

where $c(F_{Y_t}(Y_t), F_{X_t}(X_t), F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1}))$, $c\left(F_{X_t}(X_t), F_{X_{t-1}}(X_{t-1}), F_{Y_{t-1}}(Y_{t-1})\right)$ and $c(F_{Y_t}(Y_t), F_{X_{t-1}}(X_{t-1}), F_{Y_{t-1}}(Y_{t-1}))$ are the copula densities of $(Y_t, X_t, Y_{t-1}, X_{t-1})$, (X_t, X_{t-1}, Y_{t-1}), and (Y_t, X_{t-1}, Y_{t-1}), respectively.

Box I.

asymptotic bias, variance and the uniform almost convergence of Bernstein copula density estimator for α-mixing data. These properties are needed to prove the consistency and the asymptotic normality of the estimators of causality measures.

Based on the previous nonparametric estimators of copula densities, an estimator of Granger causality measure $C(X \to Y)$ is given by

$$\hat{C}(X \to Y) = \frac{1}{T}\sum_{t=1}^{T}\left\{\log\left(\frac{\hat{c}\left(F_{Y_t,T}(Y_t), F_{Y_{t-1},T}(Y_{t-1}), F_{X_{t-1},T}(X_{t-1})\right)}{\hat{c}\left(F_{Y_{t-1},T}(Y_{t-1}), F_{X_{t-1},T}(X_{t-1})\right)\hat{c}\left(F_{Y_t,T}(Y_t), F_{X_{t-1},T}(X_{t-1})\right)}\right)\right\}. \tag{12}$$

The most basic property that the above estimator should have is *consistency*. To prove consistency, some regularity assumptions are needed. We consider a set of standard assumptions on the stochastic process and bandwidth parameter of the Bernstein copula density estimator.

Assumptions on the stochastic process

(A1.1) $\{(Y_t, X_t) \in \mathbb{R}\times\mathbb{R} \equiv \mathbb{R}^2, t \geq 0\}$, is a strictly stationary β-mixing process with coefficient $\beta(l) = O(\rho^l)$, for some $0 < \rho < 1$.

(A1.2) The copula density $c\left(F_{Y_t}(.), F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$ is assumed to be twice continuously differentiable on $(0, 1)^3$ and bounded away from zero and bounded above.

Assumptions on the bandwidth parameter

(A1.3) We assume that for $k \to \infty$, $T k^{-7/2} \to 0$ and $T^{-1/2}k^{3/4}\ln(T) \to 0$.

Assumption (A1.1) is satisfied by many processes such as ARMA and ARCH processes as documented for example by Carrasco and Chen (2002) and Meitz and Saikkonen (2008). In Assumption (A1.2), the second differentiability of $c\left(F_{Y_t}(.), F_{Y_{t-1}}(.), F_{X_{t-1}}(.)\right)$ is required by Bouezmarni et al. (2010) in order to calculate the bias of the Bernstein copula estimator. Assumption (A1.3) is needed to cancel out some bias terms and for the almost sure convergence of the Bernstein copula estimator. Note that the bandwidth parameter k plays the inverse role compared to that of the standard nonparametric kernel, that is a large value of k reduces the bias but increases the variance. We now state the consistency of the nonparametric estimator in (12) (see the proof of Proposition 1 in Appendix A).

Proposition 1. *Under Assumptions* (A1.1)–(A1.3), *the estimator* $\hat{C}(X \to Y)$ *defined in* (12) *converges in probability to* $C(X \to Y)$.

5.2. Inference

The measures proposed in the previous sections can also be used to test for Granger non-causality between random variables. Hereafter, the null hypothesis of interest is given by

$$H_0 : C(X \to Y) = 0.$$

In this section, we provide the asymptotic normality of our nonparametric estimator in (12), and we establish the consistency of the test statistic used for testing H_0. Again, here we focus on the Granger causality measure from X to Y, but similar results can be obtained for the measures of Granger causality from Y to X and of the instantaneous causality between X and Y.

Theorem 1. *Under Assumptions* (A1.1)–(A1.3) *and* H_0, *we have*

$$TBE = T k^{-3/2}(2\hat{C}(X \to Y) - T^{-1}k^{3/2}\xi)/\sigma \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\sigma = \sqrt{2}(\pi/4)^{3/2}$ *and* $\xi = -\frac{\pi^{3/2}}{8} + \frac{\pi}{2}k^{-1/2} - k^{-1}(\pi^{1/2} - 1)$.

To prove the above Theorem (see the proof in Appendix A), we follow the proof of Theorem 1 in Bouezmarni et al. (2012). However, it is important to notice that the bias terms, B_1, B_2 and B_3 in Theorem 1 of Bouezmarni et al. (2012) are estimated, whereas in the present paper these terms are calculated exactly. For a given significance level α, we reject the null hypothesis H_0 when $TBE > z_\alpha$, where z_α is the critical value from the standard normal distribution.

The nonparametric estimator of Granger causality measure can be biased in small samples, and this may arise from bias in copula density estimates. This bias can, in turn, affect the finite sample properties of the *TBE* test statistic in Theorem 1. To correct the finite sample bias, we suggest to use the smooth bootstrap method proposed by Paparoditis and Politis (2000). The details of the bootstrap procedure that we use are provided in Section 7.1.1. The validity of smoothed bootstrap that corresponds to a test statistic which is similar to ours is established in Bouezmarni et al. (2012) (see Proposition 3 of Bouezmarni et al. (2012)). Under some regular assumptions on the bootstrap kernel and the bandwidth parameter, one can show that $TBE^* \xrightarrow{d} \mathcal{N}(0, 1)$, where TBE^* is the smoothed bootstrap version of *TBE*.

Notice that the derivation of Theorem 1 requires the boundedness of the copula density in Assumption (A1.2). It is true that many common families of copula are unbounded at the corners, Clayton, Gumbel, Gaussian and Student copulas being important examples. However, following Bouezmarni et al. (2012), we can show that the result in Theorem 1 is still valid for unbounded copula densities, if the following condition is fulfilled:

$$c(g_1, g_2, g_3) = O\left(\frac{1}{\sqrt{\prod_{j=1}^{3} g_j(1 - g_j)}}\right), \tag{13}$$

where $c(g_1, g_2, g_3)$ is the copula density function of (Y_t, Y_{t-1}, X_{t-1}). Condition (13) is satisfied by many common copula densities, see for example Omelka et al. (2009).

Finally, the following proposition establishes the consistency of the *TBE* test in Theorem 1 (see the proof of Proposition 2 in Appendix A).

Proposition 2. *Under Assumptions* (A1.1)–(A1.3), *the test defined in Theorem* 1 *is consistent if*

$$\int \log\left\{\frac{c(u, v, w)}{c(u, v)c(u, w)}\right\} dC(u, v, w) > 0,$$

where $c(u, v, w)$, $c(u, v)$ *and* $c(u, w)$ *are the copula densities of* (Y_t, Y_{t-1}, X_{t-1}), (Y_{t-1}, X_{t-1}), *and* (Y_t, X_{t-1}), *respectively, and* $C(u, v, w)$ *is the copula distribution function of* (Y_t, Y_{t-1}, X_{t-1}).

6. Measuring causality between high dimensional variables

Let $\{(X_t, Y_t) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \equiv \mathbb{R}^d, t = 0, \ldots, T\}$ be a sample of stationary stochastic process in $\mathbb{R}^d$, where $d = d_1 + d_2$, for $d_1, d_2 \geq 1$, with joint distribution function F_{XY} and density function f_{XY}. For (X_t, Y_t) a Markov process of order p, the null hypothesis of Granger non-causality from vector X to vector Y is given by

$$H_0^{X\to Y} : f(y_t \mid \underline{x}_{t-1}, \underline{y}_{t-1}) = f(y_t \mid \underline{y}_{t-1}), \quad 1 \leq t \leq T, \tag{14}$$

where $\underline{x}_{t-1} = \{x_{t-s}, 1 \leq s \leq p\}$, $\underline{y}_{t-1} = \{y_{t-s}, 1 \leq s \leq p\}$, with $y_t = (y_{1,t}, \ldots, y_{d_2,t})$ and $x_t = (x_{1,t}, \ldots, x_{d_1,t})$. Similarly, the null hypothesis of Granger non-causality from vector Y to vector X is given by

$$H_0^{Y\to X} : f(x_t \mid \underline{y}_{t-1}, \underline{x}_{t-1}) = f(x_t \mid \underline{x}_{t-1}), \quad 1 \leq t \leq T, \tag{15}$$

where $\underline{y}_{t-1}$ and $\underline{x}_{t-1}$ are defined above.

Moreover, the instantaneous *non-causality* between X and Y is characterized by the following equivalent null hypotheses:

$$\begin{aligned} H_0^{Y\leftrightarrow X} &: f(y_t \mid \underline{x}_t, \underline{y}_{t-1}) = f(y_t \mid \underline{x}_{t-1}, \underline{y}_{t-1}), \quad 1 \leq t \leq T, \\ H_0^{Y\leftrightarrow X} &: f(x_t \mid \underline{x}_{t-1}, \underline{y}_t) = f(x_t \mid \underline{x}_{t-1}, \underline{y}_{t-1}), \quad 1 \leq t \leq T, \\ H_0^{Y\leftrightarrow X} &: f(y_t, x_t \mid \underline{x}_{t-1}, \underline{y}_{t-1}) \\ &\quad = f(y_t \mid \underline{x}_{t-1}, \underline{y}_{t-1}) f(x_t \mid \underline{x}_{t-1}, \underline{y}_{t-1}), \quad 1 \leq t \leq T, \end{aligned} \tag{16}$$

where $\underline{x}_t = \{x_{t-s}, 0 \leq s \leq p\}$ and $\underline{y}_t = \{y_{t-s}, 0 \leq s \leq p\}$.

As in Section 4, we define the following *copula-based measure* of Granger causality from vector X to vector Y:

$$C(X \to Y) = E\left\{\log\left(\frac{c\left(\bar{F}_Y(Y_t), \bar{F}_{\underline{Y}}(\underline{Y}_{t-1}), \bar{F}_{\underline{X}}(\underline{X}_{t-1})\right) c\left(\bar{F}_{\underline{Y}}(\underline{Y}_{t-1})\right)}{c\left(\bar{F}_{\underline{Y}}(\underline{Y}_{t-1}), \bar{F}_{\underline{X}}(\underline{X}_{t-1})\right) c\left(\bar{F}_Y(Y_t), \bar{F}_{\underline{Y}}(\underline{Y}_{t-1})\right)}\right)\right\},$$

where, for simplicity of notation, we denote $\bar{F}_{\underline{Y}}(\underline{Y}_{t-1}) \equiv (F_{Y_1}(Y_{1,t-1}), \ldots, F_{Y_{d_2}}(Y_{d_2,t-1}), \ldots, F_{Y_1}(Y_{1,t-p}), \ldots, F_{Y_{d_2}}(Y_{d_2,t-p}))$, $\bar{F}_Y(Y_t) \equiv (F_{Y_1}(Y_{1,t}), \ldots, F_{Y_{d_2}}(Y_{d_2,t}))$ and $\bar{F}_{\underline{X}}(\underline{X}_{t-1}) \equiv (F_{X_1}(X_{1,t-1}), \ldots, F_{X_{d_1}}(X_{d_1,t-1}), \ldots, F_{X_1}(X_{1,t-p}), \ldots, F_{X_{d_1}}(X_{d_1,t-p}))$, with $F_{Q_i}(.)$ for $Q = X, Y$, is the marginal distribution function of the ith element of the random vector Q, and $c\left(\bar{F}_Y(.), \bar{F}_{\underline{Y}}(.), \bar{F}_{\underline{X}}(.)\right)$, $c\left(\bar{F}_{\underline{Y}}(.), \bar{F}_{\underline{X}}(.)\right)$, $c\left(\bar{F}_Y(.), \bar{F}_{\underline{Y}}(.)\right)$, and $c\left(\bar{F}_{\underline{Y}}(.)\right)$ are the copula densities of $(Y_t, \underline{Y}_{t-1}, \underline{X}_{t-1})$, $(\underline{Y}_{t-1}, \underline{X}_{t-1})$, and $(Y_t, \underline{Y}_{t-1})$, $\underline{Y}_{t-1}$, respectively. The *copula-based measure* of Granger causality from vector Y to vector X can be defined in a similar way.

Now, the *copula-based measure* of the instantaneous Granger causality between vectors X and Y is given by:

$$C(X \leftrightarrow Y) = E\left\{\log\left(\frac{c\left(\bar{F}_Y(Y_t), \bar{F}_X(X_t), \bar{F}_{\underline{X}}(\underline{X}_{t-1}), \bar{F}_{\underline{Y}}(\underline{Y}_{t-1})\right) c\left(\bar{F}_{\underline{X}}(\underline{X}_{t-1}), \bar{F}_{\underline{Y}}(\underline{Y}_{t-1})\right)}{c\left(\bar{F}_Y(Y_t), \bar{F}_{\underline{X}}(\underline{X}_{t-1}), \bar{F}_{\underline{Y}}(\underline{Y}_{t-1})\right) c\left(\bar{F}_X(Y_t), \bar{F}_{\underline{X}}(\underline{X}_{t-1}), \bar{F}_{\underline{Y}}(\underline{Y}_{t-1})\right)}\right)\right\},$$

where $\bar{F}_X(X_t) = (F_{X_1}(X_{1,t}), \ldots, F_{X_{d_1}}(X_{d_1,t}))$, and $\bar{F}_Y(Y_t)$, $\bar{F}_{\underline{Y}}(\underline{Y}_{t-1})$, and $\bar{F}_{\underline{X}}(\underline{X}_{t-1})$ are defined as above, $c\left(\bar{F}_Y(.), \bar{F}_X(.), \bar{F}_{\underline{X}}(.), \bar{F}_{\underline{Y}}(.)\right)$, $c\left(\bar{F}_Y(.), \bar{F}_{\underline{X}}(.), \bar{F}_{\underline{Y}}(.)\right)$, $c\left(\bar{F}_X(.), \bar{F}_{\underline{X}}(.), \bar{F}_{\underline{Y}}(.)\right)$, and $c(\bar{F}_{\underline{X}}(.), \bar{F}_{\underline{Y}}(.))$ are the copula densities of $(Y_t, X_t, \underline{X}_{t-1}, \underline{Y}_{t-1})$, $(Y_t, \underline{X}_{t-1}, \underline{Y}_{t-1})$, $(X_t, \underline{X}_{t-1}, \underline{Y}_{t-1})$ and $(\underline{X}_{t-1}, \underline{Y}_{t-1})$, respectively.

Following Section 5.1, an estimator of Granger causality measure from vector X to vector Y is given by

$$\hat{C}(X \to Y) = \frac{1}{T}\sum_{t=1}^{T}\left\{\log\left(\frac{\hat{c}\left(\bar{F}_{Y,T}(y_t), \bar{F}_{\underline{Y},T}(\underline{y}_{t-1}), \bar{F}_{\underline{X},T}(\underline{x}_{t-1})\right)\hat{c}\left(\bar{F}_{\underline{Y},T}(\underline{y}_{t-1})\right)}{\hat{c}\left(\bar{F}_{\underline{Y},T}(\underline{y}_{t-1}), \bar{F}_{\underline{X},T}(\underline{x}_{t-1})\right)\hat{c}\left(\bar{F}_{Y,T}(y_t), \bar{F}_{\underline{Y},T}(\underline{y}_{t-1})\right)}\right)\right\}, \tag{17}$$

where $\bar{F}_{Y,T}(y_t)$, $\bar{F}_{\underline{Y},T}(\underline{y}_{t-1})$, and $\bar{F}_{\underline{X},T}(\underline{x}_{t-1})$ with subscript T is to indicate the empirical analog of the terms $\bar{F}_Y(y_t)$, $\bar{F}_{\underline{Y}}(\underline{y}_{t-1})$, and $\bar{F}_{\underline{X}}(\underline{x}_{t-1})$ defined above, and $\hat{c}\left(\bar{F}_{Y,T}(y_t), \bar{F}_{\underline{Y},T}(\underline{y}_{t-1}), \bar{F}_{\underline{X},T}(\underline{x}_{t-1})\right)$, $\hat{c}\left(\bar{F}_{\underline{Y},T}(\underline{y}_{t-1}), \bar{F}_{\underline{X},T}(\underline{x}_{t-1})\right)$, $\hat{c}\left(\bar{F}_{Y,T}(y_t), \bar{F}_{\underline{Y},T}(\underline{y}_{t-1})\right)$, and $\hat{c}\left(\bar{F}_{\underline{Y},T}(\underline{y}_{t-1})\right)$ are the Bernstein copula density estimators of $c\left(\bar{F}_Y(y_t), \bar{F}_{\underline{Y}}(\underline{y}_{t-1}), \bar{F}_{\underline{X}}(\underline{x}_{t-1})\right)$, $c\left(\bar{F}_{\underline{Y}}(\underline{y}_{t-1}), \bar{F}_{\underline{X}}(\underline{x}_{t-1})\right)$, $c\left(\bar{F}_Y(y_t), \bar{F}_{\underline{Y}}(\underline{y}_{t-1})\right)$, and $c\left(\bar{F}_{\underline{Y}}(\underline{y}_{t-1})\right)$. To prove the consistency of the estimator in (17), similar assumptions to the ones in Section 5.1 are required. We consider a set of standard assumptions on the stochastic process and bandwidth parameter of the Bernstein copula density estimator.

Assumptions on the stochastic process

(A2.1) $\{(X_t, Y_t) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \equiv \mathbb{R}^d, t \geq 0\}$ is a strictly stationary β-mixing process with coefficient $\beta(l) = O(\rho^l)$, for some $0 < \rho < 1$.

(A2.2) The copula density $c\left(\bar{F}_Y(.), \bar{F}_{\underline{Y}}(.), \bar{F}_{\underline{X}}(.)\right)$ is assumed to be twice continuously differentiable on $(0, 1)^{d_2+pd}$ and bounded away from zero and bounded above.

Assumptions on the bandwidth parameter

(A2.3) We assume that for $k \to \infty$, $T k^{-((d_2+pd)/2)-2} \to 0$ and $T^{-1/2}k^{(d_2+pd)/4}\ln(T) \to 0$.

We now state the consistency of the nonparametric estimator in (17). The proof of Proposition 3 below is similar to the one of Proposition 1, thus we decided to do not include it in the paper.

Proposition 3. *Under Assumptions* (A.2.1)–(A.2.3), *the estimator* $\hat{C}(X \to Y)$ *defined in* (17) *converges almost surely to the true causality measure* $C(X \to Y)$.

We wish now to test the following null hypothesis:

$$H_0 : C(X \to Y) = 0.$$

Under the conditions (A2.1)–(A2.3), the following theorem provides the asymptotic normality of the estimator in (17) under the null H_0 (see the proof of Theorem 2 in Appendix A).

Theorem 2. *Under Assumptions* (A2.1)–(A2.3) *and* H_0, *we have*

$$TBE = T k^{-(d_2+pd)/2}(2\hat{C}(X \to Y) - T^{-1}k^{(d_2+pd)/2}\xi)/\sigma \xrightarrow{d} \mathcal{N}(0, 1),$$

where $\sigma = \sqrt{2}(\pi/4)^{(d_2+pd)/2}$ *and*

$$\xi = -2^{-(d_2+pd)}\pi^{(d_2+pd)/2} + 2^{-(pd_2+d_2)}\pi^{(pd_2+d_2)/2}k^{-(pd_1)/2} + 2^{-pd}\pi^{pd/2}k^{-d_2/2} + 2(2^{-pd_2} - 1)\pi^{pd_2/2}k^{-(pd_1+d_2)/2}.$$

The proof of Theorem 2 is similar to the one of Theorem 1, thus in Appendix A we only computed the bias terms. For a given significance level α, we reject the null hypothesis H_0 when $TBE > z_\alpha$, where z_α is the critical value from the standard normal distribution. To perform the test and make our decision, we can also use the smoothed bootstrap technique as in the bivariate case (see Section 5.2).

7. Monte Carlo simulations

Here we examine the finite sample bias in the nonparametric estimation of Granger causality measures and we suggest a bootstrap bias-corrected estimator. We also investigate the finite sample properties (size and power) of the TBE test in Theorem 1.

7.1. Bootstrap bias-corrected estimator of Granger causality measures

7.1.1. Bootstrap bias-correction

We first use bootstrap to compute the small sample bias in the nonparametric estimator of Granger causality measure. Thereafter,

Table 1
Data-generating processes (DGPs) used in the simulations.

DGPs	Variables of interest		Direction of causality in the DGP
	Y_t	X_t	
DGP1	ε_{1t} white noise	ε_{2t} white noise	$X \nrightarrow Y, Y \nrightarrow X$
DGP2	$Y_t = 0.5Y_{t-1} + \varepsilon_{1t}$	$X_t = 0.5X_{t-1} + \varepsilon_{2t}$	$X \nrightarrow Y, Y \nrightarrow X$
DGP3	$Y_t = (0.01 + 0.5Y_{t-1}^2)^{0.5}\varepsilon_{1t}$	$X_t = 0.5X_{t-1} + \varepsilon_{2t}$	$X \nrightarrow Y, Y \nrightarrow X$
DGP4	$Y_t = 0.5Y_{t-1} + 0.5X_{t-1} + \varepsilon_{1t}$	$X_t = 0.5X_{t-1} + \varepsilon_{2t}$	$X \to Y, Y \nrightarrow X$
DGP5	$Y_t = 0.5Y_{t-1} + 0.5X_{t-1}^2 + \varepsilon_{1t}$	$X_t = 0.5X_{t-1} + \varepsilon_{2t}$	$X \to Y, Y \nrightarrow X$
DGP6	$Y_t = 0.5Y_{t-1}X_{t-1} + \varepsilon_{1t}$	$X_t = 0.5X_{t-1} + \varepsilon_{2t}$	$X \to Y, Y \nrightarrow X$
DGP7	$Y_t = 0.5Y_{t-1} + 0.5X_{t-1}\varepsilon_{1t}$	$X_t = 0.5X_{t-1} + \varepsilon_{2t}$	$X \to Y, Y \nrightarrow X$
DGP8	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$, $h_{1,t} = 0.01 + 0.5Y_{t-1}^2 + 0.25X_{t-1}^2$	$X_t = 0.5X_{t-1} + \varepsilon_{2t}$	$X \to Y, Y \nrightarrow X$
DGP9	$\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.3 \end{pmatrix} + \begin{bmatrix} 0.1 & 0.8 \\ 0.7 & 0.15 \end{bmatrix} \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^Y \\ \varepsilon_t^X \end{pmatrix}$, with $\begin{pmatrix} \varepsilon_t^Y \\ \varepsilon_t^X \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right]$		$X \to Y, Y \to X, Y \longleftrightarrow X$

Note: This table summarizes the data generating processes that we consider in the simulation study to investigate the bias in the nonparametric estimation of Granger causality measures and to examine the finite sample properties (size and power) of nonparametric test for these causality measures. We simulate (Y_t, X_t), for $t = 1, \ldots, T$, under the assumption that $(\varepsilon_{1t}, \varepsilon_{2t})'$ are i.i.d. from $N(0, I_2)$. The last column of the table summarizes the directions of causality and non-causality in each DGP. The symbols "$\to$" and "$\nrightarrow$" refer to Granger causality and Granger non-causality, respectively.

we subtract the bias term to obtain a bootstrap bias-corrected estimate. Since a simple resampling from the empirical distribution will not conserve the conditional dependence structure in the data, see for example Remark 2.1 in Singh (1981), we suggest to use the local smoothed bootstrap proposed by Paparoditis and Politis (2000).

Hereafter, we discuss the implementation of local smoothed bootstrap. For simplicity of exposition, we consider the case of univariate Markov processes of order one. The method is easy to implement in the following five steps:

(1) We draw the sample of Y_{t-1}^* using the sum

$$Y_{t-1}^* = Y_{t-1}^+ + h\eta_{t-1},$$

where h is a bandwidth parameter and the random variables Y_{t-1}^+ and η_{t-1} are drawn from the empirical distribution of Y_{t-1} and a kernel density, L, respectively. In our simulations, L is given by a univariate normal density and h is computed using the standard rule of thumbs. Similar approach is used to simulate X_{t-1}^*;

(2) Conditional to Y_{t-1}^*, we draw Y_t^* and X_{t-1}^* independently. The sample of X_{t-1}^* conditional to Y_{t-1}^* is generated from the kernel estimator given by formula (2.5) of Paparoditis and Politis (2000). Similar approach is used to draw X_t^* and Y_{t-1}^* independently conditional to X_{t-1}^*;

(3) Based on the bootstrap sample, we compute the bootstrap Granger causality measure $\hat{C}^*(X \to Y)$;

(4) We repeat the steps **(1)**–**(3)** B times so that we obtain $\hat{C}_j^*(X \to Y)$, for $j = 1, \ldots, B$.

(5) We approximate the bias term $Bias = \mathsf{E}[\hat{C}(X \to Y) - C(X \to Y)]$ by the corresponding bootstrap bias $Bias^* = \mathsf{E}^*[\hat{C}^*(X \to Y) - \hat{C}(X \to Y)]$, where E^* is the expectation based on the bootstrap distribution of $\hat{C}^*(X \to Y)$, and $\hat{C}(X \to Y)$ is the estimate of $C(X \to Y)$ using the original sample. This suggests the bias estimate

$$\widehat{Bias}^* = \frac{1}{B}\sum_{j=1}^{B} \hat{C}_j^*(X \to Y) - \hat{C}(X \to Y).$$

Hence, a bootstrap bias-corrected estimator of measure of Granger causality from X to Y is given by

$$\hat{C}_{BC}^*(X \to Y) = \hat{C}(X \to Y) - \widehat{Bias}^*. \tag{18}$$

In practice and especially when the true value of causality measure is zero or close to zero, it is possible that for some bootstrap samples the quantity $\hat{C}_{BC}^*(X \to Y)$ becomes negative. In this case we follow Dufour and Taamouti (2010) and suggest to impose the following non-negativity truncation:

$$\hat{C}_{BC}^*(X \to Y) = \max\left\{\hat{C}_{BC}^*(X \to Y), 0\right\}. \tag{19}$$

We can similarly define the bootstrap bias-corrected estimators for measures of Granger causality from Y to X and of the instantaneous causality between X and Y.

To achieve the validity of the above local smoothed bootstrap, we need the following additional assumptions on the kernel L and bandwidth h (see Paparoditis and Politis (2000)):

Assumptions on bootstrap kernel and bandwidth

(A3.1) The kernel L is a product kernel of a bounded symmetric kernel density l.

(A3.2) l is r times continuously differentiable such that $\int u^j l^{(r)}(u)\,du = 0$ for $j = 0, \ldots, r-1$ and $\int u^r l^{(r)}(u)du < \infty$, where $l^{(r)}$ is the rth-derivative of l.

(A3.3) As $T \to \infty$, $h \to 0$, and $Th^{3+2r}/(\ln T)^{\gamma} \to C > 0$, for some $\gamma > 0$.

7.1.2. Simulation study

We run a Monte Carlo experiment to investigate possible bias and bias-correction in the nonparametric estimation of Granger causality measures. We consider two groups of data generating processes (DGPs) that represent linear and nonlinear regression models with different forms of heteroskedasticity. Table 1 presents the DGPs used in our simulation study and its last column summarizes the directions of causality and non-causality in these DGPs. The first three DGPs of Y, DGP1 to DGP8 of X, and DGP9 of (X, Y) are used to evaluate the bias in the nonparametric estimation. In these DGPs the true values of causality measures are known (equal to zero) or can be easily computed. For example, in the first three DGPs of X and Y, we have X and Y are by construction independent: Y does not cause X and X does not cause Y. Thus, we expect that the true measures of causality in these DGPs will be equal to zero. However, the causality from X to Y exists in DGP4 to DGP9 of Y, thus the true measures of causality from X to Y in these DGPs will not be equal to zero.

The nonparametric estimators of causality measures depend on the bandwidth parameter k, which is needed to estimate the copula densities. Here we take k equal to the integer part of $T^{1/2}$, but in Section 7.2 we consider various values of k to evaluate the sensitivity of the TBE test. To keep the computing time in our simulations reasonable, we consider two sample sizes $T = 200, 300$. We perform 250 bootstrap replications and 500 simulations to compute the bias terms and the average values of the bootstrap bias-corrected causality measures. Finally, in the simulations the data are rescaled such that the variables have zero mean and variance equal to one.

Table 2
Bootstrap bias-corrected estimation of Granger causality measures.

	Measure	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8	DGP9
Sample size: $T = 200$										
$Y \to X$		No	No	No	No	No	No	No	No	Yes
True	$C(Y \to X)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4702
Bias-corrected	$\hat{C}^*_{BC}(Y \to X)$	0.0000 (0.0147)	0.0000 (0.0146)	0.0000 (0.0147)	0.0000 (0.0143)	0.0000 (0.0140)	0.0000 (0.0144)	0.0002 (0.0146)	0.0000 (0.0140)	0.2253 (0.0139)
$X \to Y$		No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C(X \to Y)$	0.0000	0.0000	0.0000	–	–	–	–	–	0.3819
Bias-corrected	$\hat{C}^*_{BC}(X \to Y)$	0.0016 (0.0147)	0.0000 (0.0130)	0.0002 (0.0201)	0.0957 (0.0142)	0.1818 (0.0143)	0.1024 (0.0149)	0.0984 (0.0145)	0.0580 (0.0150)	0.2528 (0.0139)
$X \longleftrightarrow Y$		–	–	–	–	–	–	–	–	Yes
True	$C(X \longleftrightarrow Y)$	–	–	–	–	–	–	–	–	0.0408
Bias-Corrected	$\hat{C}^*_{BC}(X \longleftrightarrow Y)$	–	–	–	–	–	–	–	–	0.056 (0.012)
Sample size: $T = 300$										
$Y \to X$		No	No	No	No	No	No	No	No	Yes
True	$C(Y \to X)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4702
Bias-Corrected	$\hat{C}^*_{BC}(Y \to X)$	0.0012 (0.0085)	0.0000 (0.0078)	0.0000 (0.0081)	0.0000 (0.0076)	0.0000 (0.0083)	0.0000 (0.0078)	0.0000 (0.0061)	0.0000 (0.0078)	0.2447 (0.0076)
$X \to Y$		No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C(X \to Y)$	0.0000	0.0000	0.0000	–	–	–	–	–	0.3819
Bias-Corrected	$\hat{C}^*_{BC}(X \to Y)$	0.0009 (0.0087)	0.0000 (0.0040)	0.0003 (0.0097)	0.0993 (0.0077)	0.2069 (0.0078)	0.1094 (0.0089)	0.1150 (0.0082)	0.0637 (0.0087)	0.2769 (0.0077)
$X \longleftrightarrow Y$		–	–	–	–	–	–	–	–	Yes
True	$C(X \longleftrightarrow Y)$	–	–	–	–	–	–	–	–	0.0408
Bias-Corrected	$\hat{C}^*_{BC}(X \longleftrightarrow Y)$	–	–	–	–	–	–	–	–	0.054 (0.0093)

Note: This table shows the average values of bootstrap bias-corrected ($\hat{C}^*_{BC}(Y \to X)$, $\hat{C}^*_{BC}(X \to Y)$) estimates of causality measures from Y to X ($C(Y \to X)$) and from X to Y ($\hat{C}(X \to Y)$). "True" indicates the true value of causality measure, "Bias-Corrected" indicates the average value of the estimate of causality measure after bootstrap bias correction, and "—" means that the true value of causality measure is unknown. Eq. (18) is used to calculate the bootstrap bias-correction estimates of causality measures. The number of simulations used to compute the averaged values of the estimates of causality measures and the number of bootstrap replications used to calculate the bias-corrected estimates are equal to 500 and 250, respectively. "No" indicates non-causality in the true DGP (given in the first row of the table) and "Yes" means that there is causality in the true GDP. The data generating processes (DGPs) in the first row of the table are described in detail in Table 1. In parenthesis is the standard deviation of the estimated values.

The simulation results are presented in Table 2. From this, we see that the nonparametric estimators of measures are biased, possibly due to the finite sample bias in the nonparametric estimators of Bernstein copula densities. Interestingly, we find that there is a big improvement when one uses the bootstrap bias-corrected estimators.

7.2. Empirical size and power of the TBE test

We study the finite sample performance of nonparametric test proposed in Theorem 1. We examine its size and power properties using the data generating processes (DGPs) introduced in Section 7.1.2 (see Table 1). The first three DGPs of Y and DGP1 to DGP8 of X are used to investigate the size property, since in these DGPs the null hypothesis of non-causality is satisfied. However, in DGP4 to DGP9 of Y and GDP9 of X the null hypothesis is not satisfied, and therefore these GDPs serve to illustrate the power of the test.

Recall that Theorem 1 is valid only asymptotically. For finite samples and in order to improve the size and power of the proposed nonparametric test, bootstrap is used to compute the test statistics and *p-values*. As we mentioned in Section 7.1.2, a simple bootstrap, i.e. resampling from the empirical distribution, will not conserve the conditional dependence structure in the data, and hence sampling under the null hypothesis is not guaranteed. To prevent this from occurring, we use the local smoothed bootstrap of Paparoditis and Politis (2000). From the bootstrap causality measure $\hat{C}^*_j(X \to Y)$, as defined in Section 7.1.1, we compute the bootstrap test statistic TBE^*_j, for $j = 1, \ldots, B$. The bootstrap *p-value* is computed as $p^* = B^{-1} \sum_{j=1}^{B} 1_{\{TBE^*_j > TBE\}}$. Then, for given significance level α, we reject the null hypothesis if $p^* < \alpha$.

TBE test depends on the bandwidth parameter k, which is used to estimate the copula densities. In the simulation study we take k equal to the integer part of $\delta T^{1/2}$, for $\delta = 1, 1.5, 2$. To keep the computing time in the simulations reasonable, we consider two sample sizes $T = 200, 300$ and $B = 250$ bootstrap replications with resampling bandwidths chosen by the standard rule of thumb. Finally, we use 500 simulations to compute the empirical size and power of the test.

The empirical size and power for the sample sizes 200 and 300 are given in Tables 3 and 4, respectively. For 5% and 10% significance levels and for both $T = 200$ and $T = 300$, we see that the TBE^* test controls quite well its size and has good power. For DGP1 the test tends to be slightly oversized and is conservative for DGP3, DGP5 and DGP6. In most cases, the power is quite good and close to 100%.

8. Empirical applications

8.1. Stock market returns and exchange rates

The causal relationship between exchange rates and stock prices have been the focus of most economic literature for quite some time. In the literature, there is no academic consensus about this relationship and the results are somewhat mixed as to whether stock indexes lead exchange rates or vise versa. To examine those causal links, early studies were using simple correlations. Aggarwal (1981), using monthly data from 1974 to 1978, found that there is a positive and significant correlation between US stock prices and the trade-weighted US dollar that is equal to the average value of the US dollar weighted by US trade with its 46 largest trading partners.[2] Moreover, using monthly data from 1980 to 1986 on the

[2] In Aggarwal's study stock prices are given by the prices of New York Exchange Index (NYSE), the Standard and Poor's 500 Stock Index (S&P 500), and the

Table 3
Size and power properties for sample size $T = 200$.

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8	DGP9
	$T = 200, \alpha = 5\%$								
$Y \to X$	Size	Size	Size	Size	Size	Size	Size	Size	Power
$c = 1$	5.20	6.40	5.60	5.20	5.60	4.00	4.40	4.40	100
$c = 1.5$	6.00	6.00	4.00	4.80	3.20	3.60	4.00	3.20	100
$c = 2$	6.88	4.80	2.40	3.60	4.40	4.00	4.00	4.00	100
$X \to Y$	Size	Size	Size	Power	Power	Power	Power	Power	Power
$c = 1$	6.66	4.80	5.20	100	100	100	100	99.2	100
$c = 1.5$	6.40	4.80	5.60	100	100	98.8	100	99.2	100
$c = 2$	6.95	5.20	5.60	99.2	100	98.0	100	99.6	100
	$T = 200, \alpha = 10\%$								
$Y \to X$	Size	Size	Size	Size	Size	Size	Size	Size	Power
$c = 1$	8.00	10.00	10.80	8.80	8.40	5.60	10.00	10.40	100
$c = 1.5$	11.20	8.40	8.40	6.00	8.40	6.00	6.80	8.80	100
$c = 2$	13.20	8.00	7.60	6.40	7.60	6.40	9.20	10.00	100
$X \to Y$	Size	Size	Size	Power	Power	Power	Power	Power	Power
$c = 1$	10.80	10.80	9.60	100	100	100	100	99.6	100
$c = 1.5$	12.00	9.60	10.00	100	100	100	100	100	100
$c = 2$	13.40	8.80	9.60	100	100	99.2	100	99.6	100

Empirical size and power at the α level based on 500 replications. The sample size is $T = 200$ and the number of bootstrap resamples is $B = 250$. The bandwidth k is the integer part of $cT^{1/2}$.

Table 4
Size and power properties for sample size $T = 300$.

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8	DGP9
	$T = 300, \alpha = 5\%$								
$Y \to X$	Size	Size	Size	Size	Size	Size	Size	Size	Power
$c = 1$	6.80	3.60	2.40	5.20	4.40	4.40	3.60	5.600	100
$c = 1.5$	7.20	3.60	2.40	6.00	2.00	2.40	3.20	5.600	100
$c = 2$	5.80	3.80	2.60	5.60	3.40	2.60	4.40	4.50	100
$X \to Y$	Size	Size	Size	Power	Power	Power	Power	Power	Power
$c = 1$	6.40	6.20	6.00	100	100	100	100	100	100
$c = 1.5$	7.60	3.60	6.80	100	100	100	100	100	100
$c = 2$	5.10	4.50	5.80	100	100	100	100	100	100
	$T = 300, \alpha = 10\%$								
$Y \to X$	Size	Size	Size	Size	Size	Size	Size	Size	Power
$c = 1$	12.40	5.20	6.00	9.20	7.20	8.40	6.80	10.40	100
$c = 1.5$	11.60	6.80	6.40	9.60	5.20	5.60	9.20	10.40	100
$c = 2$	10.90	7.40	6.90	10.20	6.30	7.50	10.10	9.80	100
$X \to Y$	Size	Size	Size	Power	Power	Power	Power	Power	Power
$c = 1$	10.80	10.40	11.60	100	100	100	100	100	100
$c = 1.5$	11.60	8.80	11.20	100	100	100	100	100	100
$c = 2$	10.50	10.20	11.20	100	100	100	100	100	100

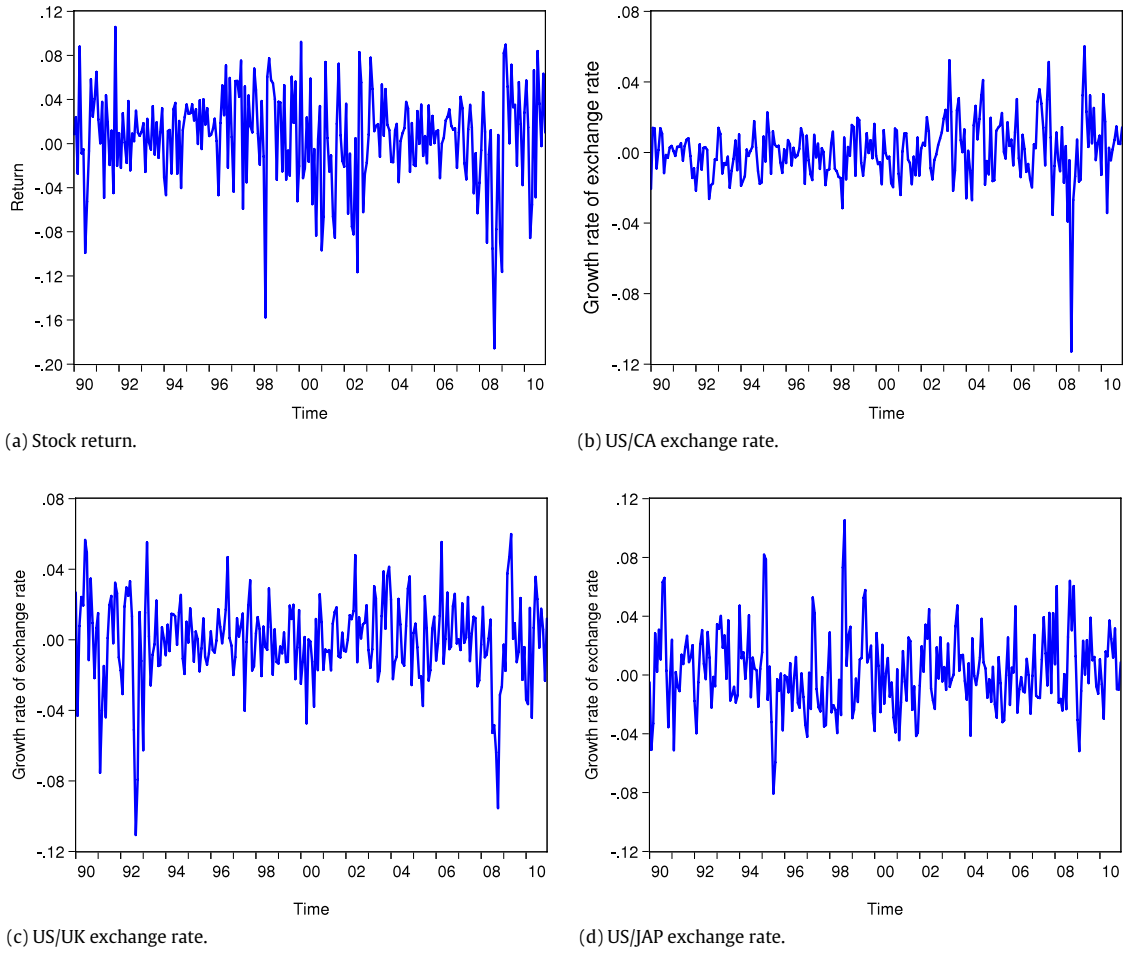
Empirical size and power at the α level based on 500 replications. The sample size is $T = 300$ and the number of bootstrap resamples is $B = 250$. The bandwidth k is the integer part of $cT^{1/2}$.

US stock prices and the effective exchange rate of the US dollar weighted against 15 other major currencies, Soenen and Henniga (1981) found that the correlation is negative and statistically significant.[3] Finally, Soenen and Aggarwal (1989) found mixed results among industrialized countries.

Many recent studies have used more sophisticated econometric techniques to study stock prices–exchange rates relationships. Bahmani-Oskooee and Sohrabian (1992) using cointegration models along with Granger causality tests, found that there is bidirectional causality between stock prices measured by S&P 500 index and the effective exchange rate of the dollar, at least in the short-run. Since Bahmani-Oskooee and Sohrabian (1992) several papers have examined different directions of causality between stock prices and exchange rates using these econometric technique and data from both industrial and developing countries. The direction of causality, similar to earlier correlation studies, appears mixed. Mok (1993), using ARIMA approach and Granger causality tests, found that the Hong Kong market efficiently incorporated much of exchange rate information in its price changes both at daily market close and open. Abdalla and Murinde (1997) found out that the results for India, Korea and Pakistan suggest that exchange rates Granger cause stock prices, which is consistent with earlier study by Aggarwal (1981). But, for the Philippines, they found that the stock prices lead the exchange rates. Granger et al. (2000), using unit root and cointegration models, found that data from South Korea are in agreement with the traditional approach. That is, exchange rates lead stock prices. On the other hand, using data of the Philippines, they found that stock prices lead exchange rates with negative correlation. Further, they found that the data from Hong Kong, Malaysia, Singapore, Thailand, and Taiwan indicate strong feedback relations, whereas that of Indonesia and Japan fail to reveal any recognizable pattern. Finally, Nieh and Lee (2001), first found that there is no long-run significant relationship between stock prices and exchange rates in the G-7 countries. This result

Department of Commerce Index of 500 Stocks (DC 500). Furthermore, the weights used to compute the variable trade-weighted US dollar represent each country's share of the total trade (measured by the sum of imports plus exports).

[3] In Soenen and Henniga's study stock prices are given by the prices of New York Exchange Index (NYSE) and the Standard and Poor's 500 Stock Index (S&P 500).



(a) Stock return.

(b) US/CA exchange rate.

(c) US/UK exchange rate.

(d) US/JAP exchange rate.

Fig. 1. S&P 500 stock returns and growth rates of US/Canada, US/UK, and US/Japan exchange rates. The sample runs from January 1990 to January 2011 for a total of 253 observations.

interfaces with Bahmani-Oskooee and Sohrabian's (1992) finding, but contrasts with the studies that suggest that there is a significant relationship between these two financial variables. Second, they found that the short-run significant relationship has only been found for one day in certain G-7 countries.

Most of the conclusions on the relationship between exchange rates and stock prices were obtained using *linear mean regression-based* tests. Although such tests have high power in uncovering linear causal relations, their power against nonlinear causal relations can be very low (see Bouezmarni et al. (2012) and Bouezmarni and Taamouti (2011), and references therein). Hence, traditional Granger causality tests might overlook a significant nonlinear relation between stock prices and exchange rates. In this section, we apply our nonparametric Granger causality measures to reexamine and quantify the causal relationship between the two financial variables in a broader framework that allows us to leave free the specification of the underlying model.

8.1.1. Data description

The data sets consist of monthly observations on S&P 500 Index and US/Canada, US/UK and US/Japan exchange rates and come from St. Louis Fed (S&P 500 Index) and Yahoo Finance (exchange rates). The sample runs from January 1990 to January 2011 for a total of 253 observations, see Fig. 1 for the series in growth rates. We perform Augmented Dickey–Fuller tests (hereafter ADF-tests) for nonstationarity of the logarithmic price and exchange rates and their first differences. Using *ADF*-tests with only an intercept and with both a trend and an intercept, the results show that all variables in logarithmic form are nonstationary. However, their first differences are stationary. The test statistics with both a trend and an intercept for the first differences of log price and log US/Canada, US/UK and US/Japan exchange rates are -14.666, -12.164, -11.390, -11.666, respectively, and the corresponding 5% critical value is -3.427. Using *ADF*-tests with only intercept leads to the same conclusions. Thus, based on the above stationarity tests we model the first difference of logarithmic price and exchange rates rather than their level. Consequently, the causality relations have to be interpreted in terms of growth rates.

8.1.2. Results and comments

We have applied the nonparametric estimator and TBE test of copula-based Granger causality measures to quantify the causality between stock market return, say r, and *US/Canada*, *US/UK* and *US/Japan* exchange rates. The empirical results are reported in Table 5 where the zero-values (0.0000) of the causality measure estimates are due to the non-negative truncation given by Eq. (19).

In Panel A of Table 5 we see that the estimates of measures of Granger causality from *US/Canada* and *US/UK* exchange rates to stock market returns are equal to zero. This indicates that the causal effects from *US/Canada* and *US/UK* exchange rates to stock returns are *economically weak*. These effects are also statistically *insignificant* at 5% significance level. Furthermore, we find that there is a causal effect from *US/Japan* exchange rate to stock market returns. This effect is statistically *significant* at 5% significance level.

Panel B of Table 5 shows that the causal effects of stock market returns on *US/UK* and *US/Japan* exchange rates are *economically weak* and statistically *insignificant* at 5% significance level. We

Table 5
Measures of causality between exchange rates and stock returns.

Direction of causality	Bandwidth $k = \delta T^{1/2}$	Bias-corrected estimate of measure	p-values (TBE test)
Panel A			
$US/Canada \rightarrow r$			
	$\delta = 1$	0.0019	0.5562
	$\delta = 1.5$	0.0000	0.6370
	$\delta = 2$	0.0000	0.7600
$US/UK \rightarrow r$			
	$\delta = 1$	0.0000	0.9601
	$\delta = 1.5$	0.0000	0.9245
	$\delta = 2$	0.0000	0.7101
$US/Japan \rightarrow r$			
	$\delta = 1$	0.0293	0.0025
	$\delta = 1.5$	0.0345	0.0051
	$\delta = 2$	0.0347	0.0226
Panel B			
$r \rightarrow US/Canada$			
	$\delta = 1$	0.0357	0.0004
	$\delta = 1.5$	0.0330	0.0090
	$\delta = 2$	0.0284	0.0501
$r \rightarrow US/UK$			
	$\delta = 1$	0.0105	0.2353
	$\delta = 1.5$	0.0095	0.3000
	$\delta = 2$	0.0097	0.3640
$r \rightarrow US/Japan$			
	$\delta = 1$	0.0083	0.2700
	$\delta = 1.5$	0.0071	0.4130
	$\delta = 2$	0.0082	0.4480

Note: This table reports the results of the bootstrap bias-corrected estimation and p-values from TBE test for measures of Granger causality from exchange rates ($US/Canada$, US/UK and $US/Japan$) to stock returns (Panel A) and from stock returns to exchange rates (Panel B).

also find that there is a causal effect from stock market returns to $US/Canada$ exchange rate, which is statistically significant at 5% significance level. Finally, it seems that the impact of stock market returns on exchange rates is more apparent than the impact of exchange rates on stock market returns.

8.2. Comparing stock return predictability using dividend–price ratio, VIX and liquidity

Many empirical studies have investigated whether stock excess returns can be predictable. The econometric method used in this context is an ordinary least squares regression of stock returns onto the past of some financial variables. Fama and French (1988) argue that using the lagged dividend–price ratio as a predictor variable has a significant effect on stock returns. Campbell and Shiller (1988) find that the lagged dividend–price ratio together with the lagged dividend growth rate have a significant predictive power on stock returns. Since the publication of Fama and French (1988) and Campbell and Shiller (1988), the question of whether stock returns are predictable or not has attracted much more attention from economists; for review see Lewellen (2004). The finding of Campbell and Shiller (1988) and Fama and French (1988) was confirmed by subsequent studies and considered to be a new stylized fact by Cochrane (1999) and Campbell (1999).

In this section, we use the nonparametric-based Granger causality measures to quantify and compare the predictive power of three financial variables (dividend–price ratio, VIX and liquidity factor) for stock market returns. The nonparametric approach does not impose any restriction on the model linking the dependent variable (here stock return) to the independent variables (dividend–price ratio, VIX or liquidity factor). In addition to dividend–price ratio, we use VIX and liquidity factor of Pastor and Stambaugh (2003) to predict stock returns. Bollerslev et al. (2009) show that the difference between VIX and realized variation, called *variance risk premium*, is able to explain a non-trivial fraction of the time series variation in post 1990 aggregate stock market returns, with high (low) premia predicting high (low) future returns. Further, Pastor and Stambaugh (2003) find that expected stock returns are related cross-sectionally to the sensitivities of returns to fluctuations in aggregate liquidity. They find that over a 34-year period, the average return on stocks with high sensitivities to liquidity exceeds that for stocks with low sensitivities by 7.5% annually, adjusted for exposures to the market return as well as size, value, and momentum factors.

8.2.1. Data description

We consider monthly aggregate S&P 500 composite index over the period January 1996 to September 2008 (153 trading months). Our empirical analysis is based on the logarithmic return on the S&P 500 in excess of the 3-month T-bill rate. The excess returns are annualized. We also consider the following monthly financial variables: dividend–price ratio, VIX and liquidity factors of Pastor and Stambaugh (2003). The monthly dividend–price ratio is computed from the Center for Research in Security Prices (CRSP) indices for the S&P 500 universe which contains monthly index files with value-weighted returns, with and without dividends. We also consider monthly data for VIX index. The VIX volatility index is an indication of the expected volatility of the S&P 500 stock index for the next thirty days. The VIX is provided by the Chicago Board Options Exchange (CBOE) in the US, and is calculated using the near term S&P 500 options markets. It is based on the highly liquid S&P500 index options along with the "model-free" approach. Finally, we consider the liquidity factor of Pastor and Stambaugh (2003). This factor can be downloaded from Stambaugh's website.

8.2.2. Results and comments

Table 6 reports the estimates and the corresponding p-values from TBE test of measures of Granger causality from dividend–price ratio, VIX and liquidity factor to stock market returns. The results show that both dividend–price ratio and VIX help to predict stock market returns. The estimates of Granger causality from dividend–price ratio to stock returns and from VIX to stock returns are statistically significant at the conventional levels of significance. The comparison of Granger causality measures estimates indicates that VIX has more predictive content than the dividend–price ratio. Similarly, we can say that the impact of VIX on

Table 6
Measures of causality between stock return and dividend–price ratio, VIX and liquidity.

Direction of causality	Bandwidth $k = \delta T^{1/2}$	Bias-corrected estimate of measure	p-values (TBE test)
Dividend–price ratio $\to r$			
	$\delta = 1$	0.0236	0.0735
	$\delta = 1.5$	0.0355	0.0560
	$\delta = 2$	0.0401	0.0320
Volatility Index $\to r$			
	$\delta = 1$	0.0504	0.0001
	$\delta = 1.5$	0.0516	0.0012
	$\delta = 2$	0.0501	0.0095
Liquidity $\to r$			
	$\delta = 1$	0.0000	0.7230
	$\delta = 1.5$	0.0108	0.4320
	$c = 2$	0.0184	0.2317

Note: This table reports the results of the bootstrap bias-corrected estimation and p-values from TBE test for measures of Granger causality from dividend–price ratio, VIX and liquidity factor to stock market returns.

stock returns is "much" stronger than the impact of dividend–price ratio on stock returns. Finally, it seems that the liquidity factor of Pastor and Stambaugh (2003) cannot help to predict stock returns. Its predictive content is weak and is statistically insignificance. Hence, using this liquidity factor will not help to predict the *time series* of stock market returns. Using cross-section data, Pastor and Stambaugh (2003) argue that this liquidity factor helps to explain the *cross-section* of individual stock returns: they find that over a 34-year period, the average return on stocks with high sensitivities to liquidity exceeds that for stocks with low sensitivities by 7.5% annually. To conclude, liquidity factors explains the variation in the cross-section of individual stock returns, but not the variation in the market stock returns (S&P 500).

9. Conclusion

We proposed a nonparametric estimator and a nonparametric test for conditional density based Granger causality measures that quantify linear and nonlinear causality between random variables. We first showed that the Granger causality measures can be rewritten in terms of copula densities. Thereafter, we proposed consistent nonparametric estimators for these Granger causality measures based on consistent nonparametric estimators of copula densities. We proved that the nonparametric estimators of the measures are asymptotically normally distributed and we discussed the validity of a local smoothed bootstrap that can be used in finite sample settings to compute bootstrap bias-corrected estimators and build tests for Granger causality measures. A simulation study revealed that the bootstrap bias-corrected estimator of causality measures behaves well and that the test has quite good finite sample properties for a variety of typical data generating processes and different sample sizes. Finally, we illustrated the practical relevance of nonparametric causality measures by quantifying the Granger causality between S&P500 Index returns and many exchange rates: US/Canada, US/UK and US/Japan exchange rates. We also compared the predictive content of dividend–price ratio, VIX and liquidity factor for stock market returns.

Appendix. Proofs

This Appendix provides the proofs of the theoretical results developed in Sections 5 and 6. Except for the proof of Proposition 1, most of the rest of the proofs here are inspired from the paper Bouezmarni et al. (2012).

Proof of Proposition 1. Put $\xi_t = \left(F_{Y_t}(y_t), F_{Y_{t-1}}(y_{t-1}), F_{X_{t-1}}(x_{t-1})\right)$ and $\xi_{t,T} = \left(F_{Y_t,T}(y_t), F_{Y_{t-1},T}(y_{t-1}), F_{X_{t-1},T}(x_{t-1})\right)$. Using Taylor expansion and the fact that $|\xi_{t,T} - \xi_t| = O_P(T^{-1/2})$ uniformly, we obtain

$$\log(\hat{c}(\xi_{t,T})) = \log(\hat{c}(\xi_t)) + O_P(T^{-1/2}). \tag{20}$$

Second, using Taylor again and the fact that $|\hat{c}(\xi_t) - c(\xi_t)| = O_P(k^{-1}+T^{-1/2}k^{3/4}\ln(T))$ uniformly (see Bouezmarni et al. (2010)), we have

$$\log(\hat{c}(\xi_t)) = \log(\hat{c}(\xi_t)) + O_P(k^{-1} + T^{-1/2}k^{3/4}\ln(T)). \tag{21}$$

From (20) and (21), we obtain

$$\begin{aligned}&\hat{C}^c(X \to Y)\\ &= \frac{1}{T}\sum_{t=1}^{T}\left\{\log\left(\frac{c\left(F_{Y_t}(Y_t), F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1})\right)}{c\left(F_{Y_{t-1}}(Y_{t-1}), F_{X_{t-1}}(X_{t-1})\right)c\left(F_{Y_t}(Y_t), F_{X_{t-1}}(X_{t-1})\right)}\right)\right\}\\ &\quad + O_P(\eta(k,T)),\end{aligned}$$

where $\eta(k,T) = T^{-1/2} + k^{-1} + T^{-1/2}k^{3/4}\ln(T)$. Hence, the results of Proposition 1 can be deduced from the law of large numbers. ■

Proof of Theorem 1. In what follows, $\bar{F}_\zeta(.)$ ($\bar{F}_{\zeta,T}(.)$) denote the distribution function of ζ (resp. the empirical distribution function of ζ), with ζ is either Y, $\underline{Y}$ or $\underline{X}$. Also, put $G_t = (\bar{F}_Y(Y_t), \bar{F}_{\underline{Y}_t}(\underline{Y}_{t-1}), \bar{F}_{\underline{X}}(\underline{X}_{t-1}))$, $U_t = (\bar{F}_Y(Y_t), \bar{F}_{\underline{Y}}(\underline{Y}_{t-1}))$, and $V_t = (\bar{F}_{\underline{Y}}(\underline{Y}_{t-1}), \bar{F}_{\underline{X}}(\underline{X}_{t-1}))$. Since

$$\hat{c}(\bar{F}_{Y,T}(y), \bar{F}_{\underline{Y},T}(\underline{y}), \bar{F}_{\underline{X},T}(\underline{x})) = \hat{c}(\bar{F}_Y(y), \bar{F}_{\underline{Y}}(\underline{y}), \bar{F}_{\underline{X}}(\underline{x})) + O_P(T^{-1}),$$

uniform in $(0,1)^d$, studding the asymptotic distribution of $\hat{C}^c(X \to Y)$ reduces to the study of

$$H(\hat{c}, \hat{C}) := \frac{1}{T}\sum_{t=1}^{T}\log\left\{\frac{\hat{c}(G_t)}{\hat{c}(U_t)\hat{c}(V_t)}\right\}.$$

Let first consider $H(\hat{c}, C) := \int \log\left\{\frac{\hat{c}(u,v,w)}{\hat{c}(u,v)\hat{c}(u,w)}\right\} dC(u,v,w)$. Using Taylor expansion, we obtain

$$\begin{aligned}H(\hat{c}, C) &\approx \int\left(\frac{\hat{c}(u,v,w)}{\hat{c}(u,v)\hat{c}(u,w)} - 1\right) dC(u,v,w)\\ &\quad - \frac{1}{2}\int\left(\frac{\hat{c}(u,v,w)}{\hat{c}(u,v)\hat{c}(u,w)} - 1\right)^2 dC(u,v,w)\\ &\quad + \frac{1}{6}\int\left(\frac{\hat{c}(u,v,w)}{\hat{c}(u,v)\hat{c}(u,w)} - 1\right)^3 dC(u,v,w)\\ &= I_1 + I_2 + I_3 ::: (say).\end{aligned}$$

Define $\phi(\alpha) = \frac{\phi_3(\alpha)}{\phi_1(\alpha)\phi_2(\alpha)} - 1$, where

$$\phi_1(\alpha) = c(u,v) + \alpha c^*(u,v),\ \phi_2(\alpha) = c(u,w) + \alpha c^*(u,w), \quad \text{and}$$
$$\phi_3(\alpha) = c(u,v,w) + \alpha c^*(u,v,w),$$

with $c^*(u,v,w)$, $c^*(u,v)$ and $c^*(u,w)$ being functions in Γ_i, for $i = 1, 2$ and 3, respectively, and Γ_i is a set defined as

$$\Gamma_i = \left\{\gamma : [0,1]^{q_i} \to \mathbb{R},\ \gamma \text{ is bounded}, \int \gamma = 0\right\}, : \text{with} :$$
$$q_1 = 3, : \text{and} : q_2 = q_3 = 2.$$

Using Taylor's expansion, we have that, for any $\alpha \geq 0$,

$$\phi(\alpha) = \phi(0) + \alpha\phi'(0) + \frac{1}{2}\alpha^2\phi''(\alpha^*), \quad \text{for } \alpha^* \in [0, \alpha].$$

One can check that,

$$\phi'(\alpha) = \frac{c^*(u,v,w)\phi_1(\alpha)\phi_2(\alpha) - c^*(u,v)\phi_2(\alpha)\phi_3(\alpha) - c^*(u,w)\phi_1(\alpha)\phi_3(\alpha)}{\phi_1^2(\alpha)\phi_2^2(\alpha)},$$

and

$$\phi''(\alpha) = O(\|c^*(u,v)c^*(u,v,w)\|_\infty + \|c^*(u,w)c^*(u,v,w)\|_\infty).$$

Next, we consider $\alpha = 1$, $c^*(u,v,w) = \hat{c}(u,v,w) - c(u,v,w)$, $c^*(u,v) = \hat{c}(u,v) - c(u,v)$, and $c^*(u,w) = \hat{c}(u,w) - c(u,w)$. Using the results of Bouezmarni et al. (2010), we get

$$\phi''(\alpha) = O_p\left(\zeta(k,T)\right),$$

where $\zeta(k,T) = T^{-1}k^{5/4}\ln^2(T) + k^{-2}$. Under H_0, we have $\phi(0) = 0$ and

$$\phi'(0) = \frac{\hat{c}^*(u,v,w)}{c(u,v,w)} - \frac{\hat{c}(u,v)}{c(u,v)} - \frac{\hat{c}(u,w)}{c(u,w)} + 1.$$

Hence, we have

$$\begin{aligned} I_1 &= \int \left(\frac{\hat{c}(u,v,w)}{c(u,v,w)} - \frac{\hat{c}(u,v)}{c(u,v)} - \frac{\hat{c}(u,w)}{c(u,w)} + 1\right) dC(u,v,w) \\ &\quad + O_p\left(\zeta(k,T)\right) \\ &= 2 - \int \frac{\hat{c}(u,v)}{c(u,v)} c(u,v,w)dudvdw \\ &\quad - \int \frac{\hat{c}(u,w)}{c(u,w)} c(u,v,w)dudvdw + O_p\left(\zeta(k,T)\right) \\ &= O_p(\zeta(k,T)). \end{aligned}$$

Similarly, one can show that $I_3 = O_p\left(\zeta(k,T)\right)$. Hence, the asymptotic distribution of $H(\hat{c}, C)$ follows from the fact that, see Bouezmarni et al. (2012),

$$\begin{aligned} &\frac{T\,k^{-3/2}}{\sigma}\left(2I_2 - C_1T^{-1}k^{3/2} - B_1T^{-1}k - B_2T^{-1}k - B_3T^{-1}k^{1/2}\right) \\ &\quad \to \mathcal{N}(0,1), \end{aligned}$$

where $C_1 = -2^{-3}\pi^{3/2}$, $B_1 = B_2 = \frac{\pi}{4}$ and $B_3 = 1 - \pi^{1/2}k^{1/2}$, and $\sigma = \sqrt{2}\,(\pi/4)^{3/2}$ and the fact that

$$Tk^{-3/2}\left(H(\hat{c},\hat{C}) - H(\hat{c},C)\right) = o_p(1). :::\ \blacksquare$$

Proof of Proposition 2. The proof of Proposition 2 can be deduced from the proof of Theorem 1 by observing that

$$\phi(0) = \frac{c(u,v,w)}{c(u,v)c(u,w)} - 1 \approx \log\left(\frac{c(u,v,w)}{c(u,v)c(u,w)}\right) > 0.$$

So, in such a case, our test statistic *TBE* converges to infinity. ■

Proof of Theorem 2. The proof of the asymptotic normality for the high dimensional case is similar to the proof of Theorem 1 given above. The key ingredient is the fact that

$$\begin{aligned} &\frac{T\,k^{-l/2}}{\sigma}(2I_2 - C_1T^{-1}k^{l/2} - B_1T^{-1}k^{(l_1+l_2)/2} \\ &\quad - B_2T^{-1}k^{(l_1+l_3)/2} - B_3T^{-1}k^{l_1/2}) \xrightarrow{d} \mathcal{N}(0,1), \end{aligned}$$

where $l_1 = pd_2$, $l_2 = d_2$, $l_3 = pd_1$, $l = l_1 + l_2 + l_3$, $C_1 = -2^{-l}\pi^{l/2}$, $\sigma = \sqrt{2}\,(\pi/4)^{l/2}$, $B_1 = 2^{-(l_1+l_2)}\pi^{(l_1+l_2)/2}$, $B_2 = 2^{-(l_1+l_3)}\pi^{(l_1+l_3)/2}$, and $B_3 = -(2 - 2^{-l_1+1})\pi^{l_1/2}k^{l_1/2}$; see the technical Appendix of Bouezmarni et al. (2012) for more details. ■

References

Abdalla, I., Murinde, V., 1997. Exchange rate and stock price interactions in emerging financial markets: evidence on India, Korea, Pakistan, and Philippines. Appl. Financ. Econ. 7, 25–35.

Aggarwal, R., 1981. Exchange rates and stock prices: a study of the US capital markets under floating exchange rates. Akron Business Econ. Rev. 12, 7–12.

Bahmani-Oskooee, M., Sohrabian, A., 1992. Stock prices and the effective exchange rate of the dollar. Appl. Econ. 24, 459–464.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financial Studies 22, 4463–4492.

Bouezmarni, T., Rombouts, J., Taamouti, A., 2010. Asymptotic properties of the Bernstein density copula estimator for alpha-mixing data. J. Multivariate Anal. 101, 1–10.

Bouezmarni, T., Rombouts, J., Taamouti, A., 2012. A nonparametric copula based test for conditional independence with applications to granger causality. J. Business Econom. Statist. 30, 275–287.

Bouezmarni, T., Taamouti, A., 2011. Nonparametric Tests for Conditional Independence Using Conditional Distribution, Universidad Carlos III de Madrid Working Paper.

Brown, B.M., Chen, S.X., 1999. Beta-Bernstein smoothing for regression curves with compact support. Scand. J. Stat. 26, 47–59.

Campbell, J.Y., Shiller, R., 1988. The dividend-price ratio and expectation of future dividends and discount factors. Rev. Financ. Stud. 1, 195–228.

Campbell, J.Y., 1999. Asset prices, consumption, and the business cycle. Handb. of Macroeconomics 1, 1231–1303.

Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric. Theory 18, 17–39.

Chen, S.X., 2000. Beta Kernel smoothers for regression curves. Statist. Sinica 10, 73–92.

Chen, X., Fan, Y., 2006. Estimation of copula-based semiparametric time series models. J. Econometrics 130, 307–335.

Chen, S.X., Huang, T.-M., 2007. Nonparametric estimation of copula functions for dependence modelling. Canad. J. Statist. 35, 265–282.

Cochrane, J., 1999. New facts in finance. Economic Perspectives XXIII (3), 36–58.

Dufour, J.-M., Taamouti, A., 2010. Short and long run causality measures: theory and inference. J. Econometrics 154, 42–58.

Doukhan, P., 1994. Mixing: Properties and Examples. Springer-Verlag, New York.

Fama, E.F., French, K., 1988. Dividend yields and expected stock returns. J. Financ. Econ. 22, 3–25.

Fan, Y., Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification tests. J. Nonparametric Stat. 10, 245–271.

Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York.

Gabaix, X., Gopikrishnan, P., Plerou, H.V., Stanley, E., 2003. A theory of power law distributions in financial market fluctuations. Nature 423, 267–270.

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. J. Amer. Statist. Assoc. 77, 304–313.

Geweke, J., 1984. Measures of conditional linear dependence and feedback between time series. J. Amer. Statist. Assoc. 79, 907–915.

Gouriéroux, C., Monfort, A., Renault, E., 1987. Kullback causality measures. Ann. Econom. Statist. 6–7, 369–410.

Granger, C., Huang, B., Yang, C., 2000. A bivariate causality between stock prices and exchange rates: evidence from recent Asian flu. Quart. Rev. Econ. Fin. 40, 337–354.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–459.

Lewellen, J., 2004. Predicting returns with financial ratios. J. Financ. Econom. 74, 209–235.

McCloskey, D.N., Ziliak, S., 1996. The standard error of regressions. J. Economic Literature XXXIV, 97–114.

Meitz, M., Saikkonen, P., 2008. Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. Econometric Theory 24, 1291–1320.

Mok, H., 1993. Causality of interest rate, exchange rate and stock prices at stock market open and close in Hong Kong. Asia Pacific J. Management 10, 123–143.

Nieh, C., Lee, C., 2001. Dynamic relationship between stock prices and exchange rates for G7 countries. Quart. Rev. Econ. Fin. 41, 477490.

Omelka, M., Gijbels, I., Veraverbek, N., 2009. Improved Kernel estimation of copulas: Weak convergence and goodnees-of-fit testing. Ann. Statist. 37, 3023–3058.

Paparoditis, E., Politis, D., 2000. The local bootstrap for Kernel estimators under general dependence conditions. Ann. Inst. Statist. Math. 52, 139–159.

Pastor, L., Stambaugh, R., 2003. Liquidity risk and expected stock returns. J. Political Economy 111, 642–685.

Polasek, W., 1994. Temporal causality measures based on AIC. In: Bozdogan, H. (Ed.), Proceedings of the Frontier of Statistical Modeling: An Informal Approach. Kluwer, Dordrecht, The Netherlands, pp. 159–168.

Polasek, W., 2002. Bayesian causality measures for multiple ARCH models using marginal likelihoods. In: Technical Report. Institute of Statistics and Econometrics, Univeristy of Basel, Switzerland.

Sancetta, A., Satchell, S., 2004. The Bernstein copula and its applications to modeling and approximating of multivariate distributions. Econometric Theory 20, 535–562.

Singh, K., 1981. On the asymptotic accuracy of Efron's bootstrap. Annals of Statistics 9, 1187–1195.

Sklar, A., 1959. Fonctions de Répartition à n Dimensions et leurs Marges. Publications de l'Institut de Statistique de l'Université de Paris 8, 229–231.
Soenen, L., Aggarwal, R., 1989. financial prices as determinants of changes in currency values, Paper presented at the 25th Annual Meetings of Eastern Finance Association, Philadelphia.
Soenen, L., Henniga, R.E., 1981. An analysis of exchange rates and stock prices: The US experience between 1980 and 1986. Akron Business Economic Rev. 19, 71–76.
Tenreiro, C., 1997. Loi Asymptotique des Erreurs Quadratiques Intégrées des Estimateurs à Noyau de la Densité et de la Régression sous des Conditions de dépendance. Port. Math. 54, 197–213.
Wiener, N., 1956. In: Beckenback, E.F. (Ed.), The Theory of Prediction. McGraw-Hill, New York, (Chapter 8).
Zheng, S., Shi, N.-Z., Zhang, Z., 2012. Generalized measures of correlation for asymmetry, nonlinearity, and beyond. J. American Statistical Association 107, 1239–1252.