

Nonparametric Tests for Conditional Independence Using Conditional Distributions*

Taoufik Bouezmarni[†]
Université de Sherbrooke

Abderrahim Taamouti[‡]
Universidad Carlos III de Madrid

May 25, 2014

ABSTRACT

The concept of causality is naturally defined in terms of conditional distribution, however almost all the empirical works focus on causality in mean. This paper aims to propose a nonparametric statistic to test the conditional independence and Granger non-causality between two variables conditionally on another one. The test statistic is based on the comparison of conditional distribution functions using an L_2 metric. We use Nadaraya-Watson method to estimate the conditional distribution functions. We establish the asymptotic size and power properties of the test statistic and we motivate the validity of the local bootstrap. We ran a simulation experiment to investigate the finite sample properties of the test and we illustrate its practical relevance by examining the Granger non-causality between S&P 500 Index returns and VIX volatility index. Contrary to the conventional t-test which is based on a linear mean-regression, we find that VIX index predicts excess returns both at short and long horizons.

Key words: Nonparametric tests; time series; conditional independence; Granger non-causality; Nadaraya-Watson estimator; conditional distribution function; VIX volatility index; S&P500 Index.

JEL Classification: C12; C14; C15; C19; G1; G12; E3; E4.

1 Introduction

This paper proposes a nonparametric test for conditional independence between two random variables of interest Y and Z conditionally on another variable X , based on comparison of conditional

*The authors thank two anonymous referees and the Editor-in-Chief Irène Gijbels for several useful comments. The main results of this work were obtained while the first author was a postdoctoral fellow at the Department of Mathematics and Statistics at Université de Montréal under the supervision of Professor Roch Roy. Special thanks to Roch Roy who helped us a lot to write this paper. We also thank Mohamed Ouzineb for his programming assistance. Financial support from the Natural Sciences and Engineering Research Council of Canada and from the Spanish Ministry of Education through grants SEJ 2007-63098 are also acknowledged.

[†]Département de mathématiques, Université de Sherbrooke, Sherbrooke, Quebec, Canada J1K 2R1. E-mail: taoufik.bouezmarni@usherbrooke.ca. TEL: +1-819 821 8000 #62035; FAX: +1- 819 821-7189.

[‡]*Corresponding author:* Economics Department, Universidad Carlos III de Madrid. Address: Departamento de Economía Universidad Carlos III de Madrid Calle Madrid, 126 28903 Getafe (Madrid) España. TEL: +34-91 6249863; FAX: +34-91 6249329; e-mail: ataamout@eco.uc3m.es.

cumulative distribution functions. Since the concept of causality can be viewed as a form of conditional independence, see Florens and Mouchart (1982) and Florens and Fougère (1996), tests for Granger non-causality between Y and Z conditionally on X can also be deduced from the proposed conditional independence test.

The concept of causality introduced by Granger (1969) and Wiener (1956) is now a basic notion when studying dynamic relationships between time series. This concept is defined in terms of predictability at horizon one of a variable Y from its own past, the past of another variable Z , and possibly a vector X of auxiliary variables. Following Granger (1969), the causality from Z to Y one period ahead is defined as follows: Z causes Y if observations on Z up to time $t - 1$ can help to predict Y_t given the past of Y and X up to time $t - 1$. The theory of causality has generated a considerable literature and for reviews see Pierce and Haugh (1977), Newbold (1982), Geweke (1984), Lütkepohl (1991), Boudjellaba, Dufour, and Roy (1992), Boudjellaba, Dufour, and Roy (1994), Gouriéroux and Monfort (1997, Chapter 10), Saidi and Roy (2008), Dufour and Renault (1998), Dufour and Taamouti (2010) among others.

To test non-causality, early studies often focus on the conditional mean, however the concept of causality is naturally defined in terms of conditional distribution; see Granger (1980) and Granger and Newbold (1986). Causality in distribution has been less studied in practice, but empirical evidence show that for many economic and financial variables, *e.g.* returns and output, the conditional quantiles are predictable, but not the conditional mean. Lee and Yang (2012), using U.S. monthly series on real personal income, output, and money, find that quantile forecasting for output growth, particularly in tails, is significantly improved by accounting for money. However, money-income causality in the conditional mean is quite weak and unstable. Cenesizoglu and Timmermann (2008), use quantile regression models to study whether a range of economic state variables are helpful in predicting different quantiles of stock returns. They find that many variables have an asymmetric effect on the return distribution, affecting lower, central and upper quantiles very differently. The upper quantiles of the return distribution can be predicted by means of economic state variables although the center of the return distribution is more difficult to predict. Moreover, generally speaking, it is possible to have situations where the causality in low moments (like mean) does not exist, but it does exist in high moments. Consequently, non-causality tests should be defined based on distribution functions. Further, since Granger non-causality is a form of conditional independence—see Florens and Mouchart (1982) and Florens and Fougère (1996)—these tests can be deduced from the conditional independence tests.

The literature on nonparametric conditional independence tests for continuous variables is quite recent. These tests are generally constructed in the context of i.i.d. data, α -mixing data, or β -mixing data. For i.i.d. data, Huang (2010) proposes tests for conditional independence using maximal nonlinear conditional correlation. Su and Spindler (2013) build a nonparametric test for asymmetric information based on the notion of conditional independence, which avoids the problem of either functional or distributional misspecification. Huang, Sun, and White (2013) introduce a nonparametric test for conditional independence based on an estimator of the topological "distance" between restricted and unrestricted probability measures corresponding to conditional independence or its absence, respectively. Linton and Gozalo (2014) develop a non-pivotal non-parametric empirical distribution function based test of conditional independence, the asymptotic

null distribution of which is a functional of a Gaussian process. Song (2009) proposes a Rosenblatt-transform based test of conditional independence between two random variables given a real function of a random vector. The function is supposed known up to an unknown finite dimensional parameter. Song (2009) suggests to use a wild bootstrap method in a spirit similar to Delgado and González Manteiga (2001) to approximate the distribution function of his test statistics. The latter three tests detect local alternatives to conditional independence that decay to zero at the parametric rate. Bergsma (2011) uses the partial copula to test for the conditional independence between random variables.

For α -mixing data, Su and White (2012) propose a nonparametric test for conditional independence using local polynomial quantile regression. Their test achieves the $T^{-1/2}$ convergence rate, where T is the sample size. Su and White (2014) construct a class of smoothed empirical likelihood-based tests for conditional independence, which are asymptotically normal under the null hypothesis. They derive the asymptotic distributions of the tests under a sequence of local alternatives, and they show that these tests possess a weak optimality property in large samples.

Finally, for β -mixing data, de Matos and Fernandes (2007) build a nonparametric test for testing the Markov property using the concept of conditional independence. Wang and Hong (2013) provide a characteristic function based test for conditional independence using a nonparametric regression approach. Su and White (2007) propose a nonparametric test based on conditional characteristic function. Their test statistic uses the squared Euclidean distance and requires to specify two weighting functions. Su and White (2008) propose a nonparametric test based on density functions and the weighted Hellinger distance. Their test statistic is consistent, asymptotically normal, and has power against alternatives at distance $T^{-1/2}h^{-d/4}$, where h is the bandwidth parameter and d is the dimension of the vector of all variables in the study. Bouezmarni, Rombouts, and Taamouti (2012) introduce a nonparametric test for conditional independence based on comparison of Bernstein copula densities using the Hellinger distance. Their test statistic is asymptotically pivotal under the null hypothesis.

In this paper, we propose a nonparametric statistic to test for conditional independence and Granger non-causality between two random variables. The test statistic compares the conditional cumulative distribution functions based on an L_2 metric. We use the Nadaraya-Watson (NW) estimator to estimate the conditional distribution functions. We establish the asymptotic size and power properties of the conditional independence test statistic and we motivate the validity of the local bootstrap. Theoretically, we show that our conditional distribution-based test is more powerful than Su and White's (2008) test and it has the same asymptotic power compared to the characteristic function-based test of Su and White (2007). Furthermore, our test is very simple to implement compared to the test of Su and White (2007). We also ran a simulation study to investigate the finite sample properties of the test. The simulation results show that the test behaves quite well in terms of size and power properties.

We illustrate the practical relevance of our nonparametric test by considering an empirical application where we examine the Granger non-causality between S&P 500 Index returns and VIX volatility index. Contrary to the conventional t-test based on a linear mean-regression, we find that VIX predicts excess returns both at short and long-run horizons. This presents evidence in favor of the existence of nonlinear *volatility feedback* effect that explains the well known asymmetric

relationship between returns and volatility.

The paper is organized as follows. In Section 2, we discuss the null hypotheses of conditional independence, the alternative hypotheses and we define our test statistic. In Section 3, we establish the asymptotic distribution and power properties of the proposed test statistic and we motivate the validity of the local bootstrap. In Section 4, we use Monte Carlo simulations to investigate the finite sample size and power properties. Section 5 contains an application using financial data. Section 6 concludes. The proofs of the asymptotic results are presented in Section 7.

2 Null hypothesis

Let $\mathcal{V}_T = \{V_t \equiv (X_t, Y_t, Z_t)\}_{t=1}^T$ be a sample of weakly dependent random variables in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}$, with joint distribution function F and density function f . For the remainder of the paper, we assume that $d_2 = 1$ which corresponds to the case of most practical interest. Suppose we are interested in testing the conditional independence between the random variables of interest Y and Z conditionally on X . The linear mean-regression model is widely used to capture and test the dependence between random variables and the least squares estimator is optimal when the errors in the regression model are normally distributed. However, in the mean regression the dependence is only due to the mean dependence, thus we ignore the dependence described by high-order moments. The use of conditional distribution functions will allow us to capture the dependence due to both low and high-order moments.

Testing the conditional independence between Y and Z conditionally on X , corresponds to test the null hypothesis

$$H_0 : \Pr \{F(y | X, Z) = F(y | X)\} = 1, \forall y \in \mathbb{R}^{d_2},$$

against the alternative hypothesis

$$H_1 : \Pr \{F(y | X, Z) = F(y | X)\} < 1, \text{ for some } y \in \mathbb{R}^{d_2}. \quad (1)$$

As pointed out by Su and White (2014), the null H_0 can be tested using the following null hypothesis defined for each value y

$$H_0(y) : \Pr \{F(y | X, Z) = F(y | X)\} = 1$$

and integrating over y . The latter integration can be computed numerically, but the computation can be time-consuming. Thus, instead of testing the above null hypothesis H_0 and $H_0(y)$, we test the following null hypothesis H'_0 , which is weaker than the null hypothesis H_0 , but practically is more convenient,

$$H'_0 : \Pr \{F(Y | X, Z) = F(Y | X)\} = 1, \quad (2)$$

against the alternative hypothesis

$$H'_1 : \Pr \{F(Y | X, Z) = F(Y | X)\} < 1. \quad (3)$$

Independently of our work, the hypothesis testing problem defined by H'_0 and H'_1 is also considered in the paper by Su and Spindler (2013) for testing the asymmetric information in the context of i.i.d data. Since the conditional distribution functions $F(y | X, Z)$ and $F(y | X)$ are unknown, we

use a nonparametric approach to estimate them. The kernel method is simple to implement and it is widely used to estimate nonparametric functional forms and distribution functions; for a review see Troung and Stone (1992) and Boente and Fraiman (1995).

To estimate the conditional distribution function, we use the Nadaraya-Watson approach proposed by Nadaraya (1964) and Watson (1964); for a review see Simonoff (1996), Li and Racine (2007), Hall, Wolff, and Yao (1999), and Cai (2002). If we denote $v = (x, y, z) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}$, $\bar{V} = (X, Z)$ and $\bar{v} = (x, z)$, then the Nadaraya-Watson estimator of the conditional distribution function of Y given X and Z is defined by

$$\hat{F}_{h_1}(y|\bar{v}) = \frac{\sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t) \mathbf{I}_{A_{Y_t}}(y)}{\sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t)}, \quad (4)$$

where $K_{h_1}(\cdot) = h_1^{-(d_1+d_3)} K(\cdot/h)$, for $K(\cdot)$ a kernel function, $h_1 = h_{1,n}$ is a bandwidth parameter, and $\mathbf{I}_{A_{Y_t}}(\cdot)$ is an indicator function defined on the set $A_{Y_t} = [Y_t, +\infty)$. Similarly, the Nadaraya-Watson estimator of the conditional distribution function of Y given only X is defined by:

$$\hat{F}_{h_2}(y|x) = \frac{\sum_{t=1}^T K_{h_2}^*(x - X_t) \mathbf{I}_{A_{Y_t}}(y)}{\sum_{t=1}^T K_{h_2}^*(x - X_t)}, \quad (5)$$

where $K_{h_2}^*(\cdot) = h_2^{-d_1} K^*(\cdot/h)$, for $K^*(\cdot)$ a different kernel function, and $h_2 = h_{2,n}$ is a different bandwidth parameter. Notice that the Nadaraya-Watson estimator for the conditional distribution is positive and monotone.

To test the null hypothesis (2) against the alternative hypothesis (3), we propose the following test statistic which is based on the conditional distribution function estimators

$$\hat{\Gamma} = \frac{1}{T} \sum_{t=1}^T \left\{ \hat{F}_{h_1}(Y_t|\bar{V}_t) - \hat{F}_{h_2}(Y_t|X_t) \right\}^2 w(\bar{V}_t), \quad (6)$$

where $w(\cdot)$ is a nonnegative weighting function of the data \bar{V}_t , for $1 \leq t \leq T$. In the simulation and application sections, and because we standardized the data, we consider a bounded support for the weight $w(\cdot)$. In the latter case we suggest to use a large bandwidth parameter for the estimation of the conditional distribution function in the tails. The weighting function $w(\cdot)$ could be useful for testing the causality in a specific range of data, for example to test Granger causality from some economic variables (e.g. inflation; DGP) to positive income. Moreover, to overcome a possible boundary bias in the estimation of the distribution function, we suggest to use the weighted Nadaraya-Watson (WNW) estimator of the distribution function proposed by Hall, Wolff, and Yao (1999) for β -mixing data and by Cai (2002) for α -mixing data. However, in these cases the test will be valid only when $d_1 + d_3 < 8$. Finally, observe that the test statistic $\hat{\Gamma}$ in (6) depends on the sample size T and it is close to zero if conditionally on X , the variables Y and Z are independent, and it diverges in the opposite case. Further, in the present paper we focus on the L_2 distance, however other distances like Hellinger distance, Kullback measure, and L_p distance, can also be considered.

3 Asymptotic distribution and power of the test statistic

In this section, we provide the asymptotic distribution of our test statistic $\hat{\Gamma}$ under the null hypothesis, and we derive its power function under local alternatives. We also establish the validity of the bootstrapped version of the test statistic.

Since we are interested in time series data, an assumption about the nature of the dependence in the individual time series is needed to derive the asymptotic distribution. We follow the literature on U-statistics and assume β -mixing dependent variables; see Tenreiro (1997) and Fan and Li (1999) among others. To recall the definition of a β -mixing process, let's consider $\{V_t; t \in \mathbb{Z}\}$ a strictly stationary stochastic process and denote \mathcal{F}_s^t the σ -algebra generated by the observations (V_s, \dots, V_t) , for $s \leq t$. The process $\{V_t\}$ is called β -mixing or absolutely regular if

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

For more details about mixing processes, the reader can consult Doukhan (1995). Other additional assumptions are needed to show the asymptotic normality of our test statistic. We assume a set of standard assumptions on the stochastic process and on the bandwidth parameter in the Nadaraya-Watson estimators of the conditional distribution functions.

Assumption A.1 (Stochastic Process)

A1.1 The process $\{V_t = (X_t, Y_t, Z_t) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}, t \in \mathbb{Z}\}$ is strictly stationary and absolutely regular with mixing coefficients $\beta(l)$, such that $\beta(l) = O(\nu^l)$, for some $0 < \nu < 1$.

A1.2 The conditional distribution functions $F(y|X)$ and $F(y|X, Z)$ are $(r + 1)$ times continuously differentiable with respect to X and (X, Z) , respectively, for some integer $r \geq 2$, and bounded on \mathbb{R}^d . The marginal densities of X_t and $\bar{V}_t = (X_t, Z_t)$, denoted by g^* and g respectively, are twice differentiable and bounded away from zero on the compact support of $w(\cdot)$.

Assumption A.2 (Kernel, Bandwidth, and Weight Function)

A2.1 The kernels K and K^* are the product of a univariate symmetric and bounded kernel $k : \mathbb{R} \rightarrow \mathbb{R}$, i.e. $K(\eta_1, \dots, \eta_{d_1+d_3}) = \prod_{j=1}^{d_1+d_3} k(\eta_j)$ and $K^*(\eta_1, \dots, \eta_{d_1}) = \prod_{j=1}^{d_1} k(\eta_j)$, such that $\int_{\mathbb{R}} k(\zeta) d\zeta = 1$ and $\int_{\mathbb{R}} \zeta^i k(\zeta) d\zeta = 0$ for $1 \leq i \leq r - 1$ and $\int_{\mathbb{R}} \zeta^r k(\zeta) d\zeta < \infty$.

A2.2 As $T \rightarrow \infty$, the bandwidth parameters h_1 and h_2 are such that $h_1, h_2 \rightarrow 0$, $h_2 = o(h_1)$ and $h_1^{d_1+d_3} = o(h_2^{d_1})$. Further, as $T \rightarrow \infty$, $Th_1^{2(d_1+d_3)} / \ln(T) \rightarrow \infty$ and $Th_1^{(d_1+d_3)/2+2r} \rightarrow 0$.

A2.3 The weight $w(\cdot)$ is a nonnegative function with compact support $A \subset \mathbb{R}^{d_1+d_3}$.

Assumption **A1.1** is often considered in the literature and it is satisfied by many processes such as ARMA, GARCH, ACD and stochastic volatility models; see Carrasco and Chen (2002) and Meitz and Saikkonen (2008) among others. Assuming β -mixing data, our main results in sections 3.1-3.3 can be shown using the Central Limit Theorem (CLT) for the second order U-statistic in Tenreiro

(1997). It can also be shown that these results continue to hold for α -mixing data. Dehling and Wendler (2010) provide the Central Limit Theorem for U-statistics under strong mixing conditions. Assumption **A1.2** is needed to derive the bias and variance of the Nadaraya-Watson estimators of the conditional distribution functions. The integer r in assumptions **A1.2** and **A2.1** depends on the dimension of the data, i.e., for example with $d_1 = d_2 = d_3 = 1$, we can consider the Gaussian kernel function ($r = 2$). But for a higher dimension, a higher order kernel function is required. Assumption **A2.2** implies that if the bandwidth parameters $h_1 = cst_1 T^{-1/\psi_1}$ and $h_2 = cst_2 T^{-1/\psi_2}$ are considered, then ψ_1 and ψ_2 must satisfy the conditions $d_1 + d_3 < \psi_1 < (d_1 + d_3)/2 + 2r$ and $\psi_2 < \psi_1 < \psi_2(1 + d_3/d_1)$. Assumption **A2.2** is similar to the one considered in Ait-Sahalia, Bickel, and Stoke (2001) and is slightly different from Assumption **A.2 (ii)** in Su and White (2007). The latter assume that $Th_1^{(d_1+d_3)/2}h_2^{2r} = o(1)$ instead of $Th_1^{(d_1+d_3)/2+2r} = o(1)$. Hence, our Assumption **A2.2** implies Assumption **A.2 (ii)** in Su and White (2007). However, in their Monte Carlo study, Su and White (2007) use the bandwidths $h_1 = O(T^{-1/(4+d_1+d_3)})$ and $h_2 = O(T^{-1/(4+d_1)})$, for $r = 4$, $d_2 = d_3 = 1$ and $d_1 = d \leq 2$, which satisfy our Assumption **A2.2**.

3.1 Asymptotic distribution of the test statistic

Before presenting the main results, we first define the following terms:

$$\begin{aligned}
D_1 &= C_1 h_1^{-(d_1+d_3)} \int_{v_t} \frac{w(\bar{v}_t)}{g(\bar{v}_t)} (1 - F(y_t|\bar{v}_t))^3 f(v_t) dv_t, \\
D_2 &= h_2^{-d_1} C_2 \int_{x_t, y_t} \frac{w^*(x_t)}{g^*(x_t)} (1 - F(y_t|x_t))^3 f(x_t, y_t) dx_t dy_t, \\
D_3 &= -2C_3 h_1^{-d_1} \int \frac{w(\bar{v}_t)}{g^*(x_t)} (1 - F(y_t|\bar{v}_t))^3 f(v_t) dv_t, \\
D &= (D_1 + D_2 + D_3)/T,
\end{aligned} \tag{7}$$

where

$$w^*(x_t) = \int_z w(x_t, z) g(z|x_t) dz,$$

and

$$C_1 = \frac{1}{3} \int K^2(x, z) dx dz, \quad C_2 = \frac{1}{3} \int K^{*2}(x) dx \quad \text{and} \quad C_3 = \frac{1}{3} K^*(0).$$

Further, we denote

$$\sigma^2 = 2C \int_{\bar{v}_0} w^2(\bar{v}_0) \left[\int_{\tilde{y}_0} \int_{y_0} h^2(\tilde{y}_0, y_0, \bar{v}_0) f(\tilde{y}_0|\bar{v}_0) f(y_0|\bar{v}_0) d\tilde{y}_0 dy_0 \right] g^2(\bar{v}_0) d\bar{v}_0 \tag{8}$$

for $C = \int_{a_1, a_3} \left(\int_{b_1, b_3} K(\bar{b} + \bar{a}) K(\bar{b}) db_1 db_3 \right)^2 da_1 da_3$ and

$$h(\tilde{y}_0, y_0, \bar{v}_0) = \frac{1}{3} + \frac{1}{2} [F^2(y_0|\bar{v}_0) + F^2(\tilde{y}_0|\bar{v}_0)] - F(\max(y_0, \tilde{y}_0)|\bar{v}_0).$$

The following theorem establishes the asymptotic normality of the test statistic $\hat{\Gamma}$ in (6) under the null hypothesis. In the sequel, “ \xrightarrow{d} ” stands for convergence in distribution.

Theorem 1 (Asymptotic distribution) *If Assumptions A.1 and A.2 hold, then under H_0 we have*

$$Th_1^{\frac{1}{2}(d_1+d_3)}(\hat{\Gamma} - D) \xrightarrow{d} N(0, \sigma^2), \text{ as } T \rightarrow \infty,$$

where $\hat{\Gamma}$ is defined in (6) and D and σ^2 are defined in Equations (7) and (8), respectively. ■

Theorem 1 is valid only when $d_1+d_3 < 4r$. Hence, for small dimensions, for example $d_1 = d_3 = 1$, we can use as a kernel the normal density function. However, if the test is for higher dimensions, a higher order kernel is required. Further, notice that the term D_3 in (7) is negligible for $d_3 > d_1$. Finally, for $h_1 = cst_1 T^{-1/\psi_1}$ and $h_2 = cst_2 T^{-1/\psi_2}$ with $\psi_1 < \frac{\psi_2}{2}(1 + d_3/d_2)$, the second bias term D_2 in (7) is also negligible.

The implementation of our test statistic requires the estimation of the variance term σ^2 and the bias terms D_1, D_2 , and D_3 . We propose the following estimators

$$\begin{aligned} \hat{D}_1 &= C_1 h_1^{-(d_1+d_3)} \frac{1}{T} \sum_{t=1}^T \frac{w(\bar{V}_t)}{\hat{g}(\bar{V}_t)} (1 - \hat{F}(Y_t|\bar{V}_t))^3, \\ \hat{D}_2 &= h_2^{-d_1} C_2 \frac{1}{T} \sum_{t=1}^T \frac{\hat{w}^*(X_t)}{\hat{g}^*(X_t)} (1 - \hat{F}(Y_t|X_t))^3, \\ \hat{D}_3 &= -2C_3 h_1^{-d_1} \frac{1}{T} \sum_{t=1}^T \frac{w(\bar{V}_t)}{\hat{g}^*(X_t)} (1 - \hat{F}(Y_t|\bar{V}_t))^3, \\ \hat{D} &= (\hat{D}_1 + \hat{D}_2 + \hat{D}_3)/T, \end{aligned} \tag{9}$$

where

$$\hat{w}^*(x) = \frac{1}{T} \sum_{t=1}^T w(x, Z_t),$$

and $\hat{F}_{h_1}(Y_t|\bar{V}_t)$ (resp. $\hat{F}_{h_2}(Y_t|X_t)$) is the estimator of the conditional distribution function $F_{h_1}(Y_t|\bar{V}_t)$ (resp. $F_{h_2}(Y_t|X_t)$) defined in (4) (resp. (5)), and $\hat{g}(\cdot)$ and $\hat{g}^*(\cdot)$ are the nonparametric kernel estimators of $g(\cdot)$ and $g^*(\cdot)$. Similarly, an estimator of the variance σ^2 is given by:

$$\hat{\sigma}^2 = \frac{2C}{T} \sum_{t=1}^T w^2(\bar{V}_t) \hat{a}(\bar{V}_t) g(\bar{V}_t), \tag{10}$$

where

$$\hat{a}(\bar{v}) = \frac{1}{T^2} \sum_{t,t'=1}^T Th^2(\tilde{Y}_t, Y_{t'}, \bar{v})$$

and the process $\{\tilde{Y}_t, t \geq 1\}$ is an independent copy of Y_t , conditionally on \bar{V}_t . Note that we can generate independently the data $\{\tilde{Y}_t\}$ using a kernel estimate of the conditional distribution of Y_t given \bar{V}_t .

Finally, we reject the null hypothesis when $Th_1^{\frac{1}{2}(d_1+d_3)}(\hat{\Gamma} - \hat{D})/\hat{\sigma} > z_\alpha$, where z_α is the $(1 - \alpha)$ -quantile of the $N(0, 1)$ distribution.

3.2 Power of the test statistic

Here, we study the consistency and the power of our nonparametric test statistic against fixed or local alternatives. The following proposition states the consistency of the test for a fixed alternative.

Proposition 1 (Consistency) *If Assumptions A.1 and A.2 hold, then the test based on $\hat{\Gamma}$ in (6) is consistent for any distributions $F(y | x, z)$ and $F(y | x)$ such that*

$$\int (F(y|x, z) - F(y|x))^2 w(x, z) dx dy dz > 0.$$

■

Now, we examine the power of the above proposed test against local alternatives. We consider the following sequence of local alternatives

$$H_1(\xi_T) : F^{[T]}(y | x, z) = F^{[T]}(y | x) [1 + \xi_T \Delta(x, y, z) + o(\xi_T) \Delta_T(x, y, z)], \quad (11)$$

where $F^{[T]}(y|x, z)$ (*resp.* $F^{[T]}(y|x)$) is the conditional distribution of $Y_{T,t}$ given $X_{T,t}$ and $Z_{T,t}$ (*resp.* of $Y_{T,t}$ given $X_{T,t}$) and $\xi_T \rightarrow 0$ as $T \rightarrow \infty$. The notation “[T]” in $F^{[T]}(y|x, z)$ and $F^{[T]}(y|x)$ is to say that the difference between the latter distribution functions depends on the sample size T . We suppose that $\|f^{[T]} - f\|_\infty = o(T^{-1}h_1^{-(d_1+d_3)/2})$. The $\Delta(x, y, z)$ and $\Delta_T(x, y, z)$ are such that $1 + \xi_T \Delta(x, y, z) + o_p(\xi_T) \Delta_T(x, y, z) \geq 0$, for all (x, y, z) and T , $\int \Delta(x, y, z) f(x, z) dz = 0$, $\int \Delta_T(x, y, z) f(x, z) dz = 0$,

$$\int \Delta^2(x, y, z) w(x, z) f(x, y, z) dx dy dz = \gamma < \infty, \quad (12)$$

and that

$$\int \Delta_T^2(x, y, z) w(x, z) f(x, y, z) dx dy dz < \infty,$$

We assume the following assumptions on the stochastic process $\{(X_{T,t}, Y_{T,t}, Z_{T,t})\}$ that are similar to the above assumptions **A1.1** and **A1.2**:

A1.1* Let $\{(X_{T,t}, Y_{T,t}, Z_{T,t}), t = 1, \dots, T\}$ be a strictly stationary β -mixing process with coefficients $\beta^{[T]}(l)$ such that

$$\sup_T \beta^{[T]}(l) = O(\bar{\nu}^l), \quad \text{for some } 0 < \bar{\nu} < 1.$$

A1.2* The conditional distribution functions $F^{[T]}(y|X)$ and $F^{[T]}(y|X, Z)$ are $(r + 1)$ times continuously differentiable, for some integer $r \geq 2$, and bounded on \mathbb{R}^d . The marginal densities of $X_{T,t}$ and $\bar{V}_{T,t} = (X_{T,t}, Z_{T,t})$, denoted by $g^{*[T]}$ and $g^{[T]}$ respectively, are twice differentiable and bounded away from zero on the compact support of $w(\cdot)$.

The following proposition establishes the asymptotic local power property of the test statistic $\hat{\Gamma}$ under the local alternatives in (11).

Proposition 2 (Asymptotic local power properties) *Under Assumptions **A1.1***, **A1.2*** and **A.2** and under the local alternative $H_1(\xi_T)$ with $\xi_T = T^{-1/2}h_1^{-(d_1+d_3)/4}$, we have*

$$Th_1^{\frac{1}{2}(d_1+d_3)}(\hat{\Gamma} - D) \xrightarrow{d} N(\gamma, \sigma^2), \text{ as } T \rightarrow \infty,$$

where D , σ^2 , and γ are defined in (7), (8), and (12), respectively. ■

Notice that our test has power against alternatives at distance $T^{-1/2}h_1^{-(d_1+d_3)/4}$ compared to the power of the tests of Su and White (2008) and Bouezmarni, Rombouts, and Taamouti (2012), which have power only against alternatives at distance $T^{-1/2}h_1^{-(d_1+d_2+d_3)/4}$. Further, our test has an asymptotic power at the same distance as the characteristic function-based test of Su and White (2007) and Su and White (2014).

3.3 Local bootstrap

In finite samples, the asymptotic normal distribution does not generally provide a satisfactory approximation for the exact distribution of nonparametric test statistic. To improve the finite sample properties of our test, we propose the use of a bootstrap method. In our context, in order to generate data under the null hypothesis, that is under the conditional independence, the local smoothed bootstrap suggested by Paparoditis and Politis (2000) seems appropriate.

In the sequel, $X \sim f_X$ means that the random variable X is generated from the density function f_X . Consider L_1, L_2 and L_3 three product kernels that satisfy Assumption **A2.1** and a bandwidth kernel h satisfying Assumption **A.3** below. The local smoothed bootstrap method is easy to implement in the following five steps:

(1) We draw a bootstrap sample $\{(X_t^*, Y_t^*, Z_t^*), t = 1, \dots, T\}$ as follows

$$X_t^* \sim T^{-1}h^{-d_1} \sum_{s=1}^T L_1(X_s - x)/h;$$

and conditionally on X_t^* ,

$$Y_t^* \sim h^{-d_2} \sum_{s=1}^T L_1((X_s - X_t^*)/h) L_2((Y_s - y)/h) / \sum_{s=1}^n L_1((X_s - X_t^*)/h)$$

and

$$Z_t^* \sim h^{-d_3} \sum_{s=1}^T L_1((X_s - X_t^*)/h) L_3((Z_s - z)/h) / \sum_{s=1}^T L_1((X_s - X_t^*)/h);$$

(2) based on the bootstrap sample, we compute the bootstrap test statistic $\Gamma^* = Th_1^{\frac{1}{2}(d_1+d_3)}(\hat{\Gamma}^* - \hat{D}^*)/\hat{\sigma}^*$, where $\hat{\Gamma}^*$, \hat{D}^* , and $\hat{\sigma}^*$ are analogously defined as $\hat{\Gamma}$, \hat{D} , and $\hat{\sigma}$, and computed using the bootstrap data (X_t^*, Y_t^*, Z_t^*) ;

(3) we repeat the steps (1)-(2) B times so that we obtain Γ_j^* , for $j = 1, \dots, B$;

(4) we compute the bootstrap p -value and for a given significance level α , we reject the null hypothesis if $p^* < \alpha$.

Here we take the same bandwidth parameter h , however different bandwidths could also be considered. An additional assumption concerning the bandwidth parameter h is required to validate the local bootstrap.

Assumption A.3 (Bootstrap Bandwidth)

A3.1 As $T \rightarrow \infty$, $h \rightarrow 0$ and $Th^{d+2r}/(\ln T)^\gamma \rightarrow C > 0$, for some $\gamma > 0$ and $d = d_1 + d_2 + d_3$

The following proposition establishes the consistency of the local bootstrap for the conditional independence test.

Proposition 3 (Smoothed local bootstrap) *Suppose that Assumptions A.1, A.2 and A.3 are satisfied. Then, conditionally on the observations $\mathcal{V}_T = \{V_t \equiv (X_t, Y_t, Z_t)\}_{t=1}^T$, we have*

$$\Gamma^* \xrightarrow{d} N(0, 1), \text{ as } T \rightarrow \infty.$$

■

The proofs are presented in the Appendix. The finite-sample properties of our nonparametric test are investigated in the next section.

Table 1: Data generating processes used in the simulation study.

DGP	X_t	Y_t	Z_t
DGP1	ε_{1t}	ε_{2t}	ε_{3t}
DGP2	Y_{t-1}	$Y_t = 0.5Y_{t-1} + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
DGP3	Y_{t-1}	$Y_t = (0.01 + 0.5Y_{t-1}^2)^{0.5}\varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
DGP4	Y_{t-1}	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$ $h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05Y_{t-1}^2$	$Z_t = \sqrt{h_{2,t}}\varepsilon_{2t}$ $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$
DGP5	Y_{t-1}	$Y_t = 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
DGP6	Y_{t-1}	$Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}^2 + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
DGP7	Y_{t-1}	$Y_t = 0.5Y_{t-1}Z_{t-1} + \varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
DGP8	Y_{t-1}	$Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}\varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
DGP9	Y_{t-1}	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$ $h_{1,t} = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$

4 Monte Carlo simulations: size and power

Here, we present the results of a Monte Carlo experiment to illustrate the size and power of the proposed test using reasonable sample sizes. We have limited our study to two groups of data generating processes (DGPs) that represent different linear and nonlinear regression models with

different forms of heteroscedasticity. These DGPs are described in Table 1. The first four DGPs were used to evaluate the empirical size. In these DGPs, Y and Z are, by construction, independent conditional on X . In the last five DGPs, Y and Z are, by construction, dependent conditional on X and have served to evaluate the power. We have considered three different sample sizes, $T = 200$, $T = 300$, and $T = 800$. For each DGP and for each sample size, we have generated 500 independent realizations and for each realization, 500 bootstrapped samples were obtained. For estimating the conditional distribution functions, we have used the normal density function, which is a second-order kernel, hence $C_1 = 1/2\pi$, $C_2 = 1/\sqrt{2\pi}$, $C_3 = 1/\sqrt{\pi}$, and $C = 1/4\pi$. Because the data are standardized, the weighting function in the test statistic $\hat{\Gamma}$ in (6) is given by the indicator function defined on the set $A = \{(x, z), -2 \leq x, z \leq 2\}$. Finally, for generating the bootstrap replications, we have used the normal kernel with a different bandwidth which is provided by the rule of thumb proposed in Silverman (1986).

In addition to the Silverman's rule of thumb used in this paper, at least three other ways can be used to choose the bandwidth in practice. The first one is the cross-validation bandwidth proposed by Li, Lin, and Racine (2013). The rate of the cross-validation bandwidth satisfies our Assumption **A2.2** since it is of order $T^{-1/(d_1+d_3)}$. However, strictly speaking, since the cross-validated bandwidth is random, the asymptotic theory can be justified with this random bandwidth only through certain stochastic equicontinuity argument.¹ The cross-validation technique is used in Li, Maasoumi, and Racine (2009) for testing the equality of two unconditional and conditional functions in the context of mixed categorical and continuous data. However, this approach, which is optimal for the estimation, loses the optimality for nonparametric kernel testing. The second way is given by an *adaptive-rate-optimal rule* proposed by Horowitz and Spokoiny (2001) for testing a parametric model for conditional mean function against a nonparametric alternative. The third way for selecting a practical bandwidth is introduced by Gao and Gijbels (2008). Gao and Gijbels (2008) propose, using the Edgeworth expansion of the asymptotic distribution of the test, to choose the bandwidth such that the power function of the test is maximized while the size function is controlled. The above three approaches will be investigated in future research. In this paper, we take $h_1 = c_1 T^{-1/4.75}$ and $h_2 = c_2 T^{-1/4.25}$ for various values of c_1 and c_2 , which correspond to the most practical case. These values are selected in order to satisfy our Assumption **A2.2**.

For a given DGP, the 500 independent realizations of length T were obtained as follows:

- (1) We generate $T + 200$ independent and identically distributed noise values $(\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})' \sim N(0, I_3)$;
- (2) Each noise sequence was plugged into the DGP equation to generate $(X_t, Y_t, Z_{t-1})'$, $t = 1, \dots, T + 200$. The initial values were set to zero (resp. to one) for X_t , Y_t and Z_t (resp. for $h_{1,t}$ and $h_{2,t}$). To attenuate the impact of the initial values, the first 200 observations were discarded.

Our test is valid for testing both linear and nonlinear Granger causalities and we have compared it with the commonly used t -test for linear causality. In the linear causality analysis, we have examined if the variable Z_{t-1} explains Y_t in the presence of Y_{t-1} , using the following linear mean

¹We thank an Anonymous Referee for his/her remark with respect to the randomness of the cross-validation bandwidth and the importance of using certain stochastic equicontinuity argument to justify the asymptotic theory.

regression:

$$Y_t = \mu + \beta Y_{t-1} + \alpha Z_{t-1} + \varepsilon_t.$$

The null hypothesis of Granger non-causality is given by $H_0 : \alpha = 0$ against the alternative hypothesis $H_1 : \alpha \neq 0$. To test H_0 , the t -statistic is given by $t_{\hat{\alpha}} = \frac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}}$, where $\hat{\alpha}$ is the least squares estimator of α and $\hat{\sigma}_{\hat{\alpha}}$ is the estimator of its standard error $\sigma_{\hat{\alpha}}$. In the presence of possibly dependent errors, $\hat{\sigma}_{\hat{\alpha}}$ was computed using the commonly used heteroscedasticity autocorrelation consistent (HAC) estimator suggested by Newey and West (1987).

Table 2: Empirical size of the bootstrapped nonparametric test of conditional independence.

	DGP1	DGP2	DGP3	DGP4	DGP1	DGP2	DGP3	DGP4
	$T = 200, \alpha = 5\%$				$T = 200, \alpha = 10\%$			
LIN	0.047	0.051	0.041	0.053	0.091	0.092	0.098	0.092
BT, $c_1=1, c_2=1$	0.050	0.056	0.044	0.038	0.096	0.104	0.098	0.098
BT, $c_1=0.85, c_2=0.7$	0.048	0.044	0.064	0.056	0.104	0.128	0.132	0.100
BT, $c_1=0.75, c_2=0.6$	0.036	0.048	0.052	0.052	0.096	0.088	0.120	0.088
	$T = 300, \alpha = 5\%$				$T = 300, \alpha = 10\%$			
LIN	0.051	0.060	0.051	0.048	0.095	0.104	0.108	0.110
BT, $c_1=1, c_2=1$	0.053	0.043	0.068	0.040	0.120	0.097	0.110	0.100
BT, $c_1=0.85, c_2=0.7$	0.060	0.036	0.068	0.060	0.120	0.084	0.108	0.130
BT, $c_1=0.75, c_2=0.6$	0.044	0.032	0.060	0.056	0.108	0.076	0.096	0.112
	$T = 800, \alpha = 5\%$				$T = 800, \alpha = 10\%$			
LIN	0.049	0.052	0.050	0.050	0.100	0.102	0.099	0.103
BT, $c_1=1, c_2=1$	0.049	0.051	0.056	0.047	0.105	0.099	0.100	0.103
BT, $c_1=0.85, c_2=0.7$	0.056	0.046	0.056	0.054	0.108	0.092	0.103	0.114
BT, $c_1=0.75, c_2=0.6$	0.045	0.042	0.056	0.052	0.100	0.089	0.092	0.103

Empirical sizes are based on 500 replications. LIN refers to the linear test and BT to our test. c_1 and c_2 refer to the constants in the bandwidth parameters.

The empirical sizes of the linear causality test (LIN test) and of the distribution-based test (BT test) for different values of the constants c_1 and c_2 in the bandwidth parameters are given in Table 2. Based on 500 replications, the standard error of the rejection frequencies is 0.0097 at the nominal level $\alpha = 5\%$ and 0.0134 at $\alpha = 10\%$. Globally, the sizes of both tests are fairly well controlled even with series of length $T = 200$. For $T = 800$, the empirical sizes are very close to the nominal levels $\alpha = 5\%$ and $\alpha = 10\%$, respectively. Thus, in large samples the empirical size is well controlled. For LIN test, all rejection frequencies are within 2 standard errors from the nominal levels 5% and 10%. For BT test, at 5%, all rejection frequencies are also within 2 standard errors. However, at 10%, three rejection frequencies are between 2 and 3 standard errors (two at $T = 200$ and one at $T = 300$). There is no strong evidence of overrejection or underrejection. Finally, for BT test the empirical sizes seem slightly closer to the corresponding nominal sizes when $c_1 = c_2 = 1$.

Table 3: Empirical power of the bootstrapped nonparametric test of conditional independence.

		DGP5	DGP6	DGP7	DGP8	DGP9
$\alpha = 5\%$	$T = 200$					
	LIN	0.994	0.401	0.184	0.137	0.151
	BT, $c_1=1, c_2=1$	0.996	0.812	0.852	1.000	0.936
	BT, $c_1=0.85, c_2=0.7$	0.988	0.728	0.792	1.000	0.908
	BT, $c_1=0.75, c_2=0.6$	0.976	0.719	0.808	1.000	0.896
	$T = 300$					
	LIN	1.000	0.412	0.204	0.142	0.171
	BT, $c_1=1, c_2=1$	1.000	0.976	0.966	1.000	1.000
	BT, $c_1=0.85, c_2=0.7$	1.000	0.884	0.908	1.000	0.984
	BT, $c_1=0.75, c_2=0.6$	1.000	0.784	0.868	1.000	0.960
	$T = 800$					
	LIN	1.000	0.422	0.216	0.151	0.183
	BT, $c_1=1, c_2=1$	1.000	1.000	1.000	1.000	1.000
	BT, $c_1=0.85, c_2=0.7$	1.000	0.997	1.000	1.000	1.000
	BT, $c_1=0.75, c_2=0.6$	1.000	0.978	0.995	1.000	1.000
	$\alpha = 10\%$	$T = 200$				
LIN		1.000	0.410	0.211	0.134	0.161
BT, $c_1=1, c_2=1$		0.992	0.916	0.916	0.984	0.980
BT, $c_1=0.85, c_2=0.7$		0.996	0.844	0.868	1.000	0.960
BT, $c_1=0.75, c_2=0.6$		0.984	0.831	0.854	1.000	0.964
$T = 300$						
LIN		1.000	0.432	0.224	0.159	0.187
BT, $c_1=1, c_2=1$		1.000	1.000	0.951	1.000	1.000
BT, $c_1=0.85, c_2=0.7$		1.000	0.948	0.964	1.000	1.000
BT, $c_1=0.75, c_2=0.6$		1.000	0.912	0.924	1.000	0.984
$T = 800$						
LIN		1.000	0.456	0.253	0.178	0.212
BT, $c_1=1, c_2=1$		1.000	1.000	1.000	1.000	1.000
BT, $c_1=0.85, c_2=0.7$		1.000	0.984	0.980	1.000	1.000
BT, $c_1=0.75, c_2=0.6$		1.000	0.954	0.993	1.000	1.000

Empirical powers are based on 500 replications. LIN refers to the linear test and BT to our test. c_1 and c_2 refer to the constants in the bandwidth parameters.

The empirical powers of LIN and BT tests are reported in Table 3. As expected, with the linear DGP5, LIN test performs extremely well but the nonparametric BT test performs almost as well, especially when $T = 800$. With the four nonlinear models under consideration, BT test clearly outperforms LIN test. In most cases, BT test produces the greatest power when $c_1 = c_2 = 1$. Finally, at both levels 5% and 10%, the powers increase considerably with DGP6, DGP7 and DGP9, when T goes from 200 to 800.

5 Empirical application

We use real data to illustrate the practical importance of the proposed nonparametric test. We show that using tests based on linear mean regressions may lead to wrong conclusions about the existence of a relationship between financial variables. In particular, we examine the linear and nonlinear causalities between stock market excess return and volatility index (VIX). We test whether stock market excess returns can be predictable at short and long-run horizons using the VIX index. We compare the results using the conventional t -test and the new nonparametric test.

5.1 Stock return predictability using volatility index

Many empirical studies have investigated whether stock excess returns can be predictable; see Fama and French (1988), Campbell and Shiller (1988), Kothari and Shanken (1997), Lewellen (2004), Bollerslev, Tauchen, and Zhou (2009) among many others. In most of these studies, the econometric method used is the conventional t -test based on the ordinary least squares regression of stock returns onto the past of some financial variables.² Here we examine the short and long-run stock return predictability using VIX in a broader framework that allows us to leave free the specification of the underlying model. Nonparametric tests are well suited for that since they do not impose any restriction on the model linking the dependent variable to the independent variables.

Recent works use VIX to predict stock excess returns. Bollerslev, Tauchen, and Zhou (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. In what follows, we use VIX together with nonparametric test to check whether the excess returns on S&P 500 Index are predictable. We compare our results to those obtained using the standard t -test.

5.2 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 monthly data for VIX. The latter 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

²Previous studies have also considered testing return predictability from past returns, for a review see Lo and MacKinlay (1988), French and Roll (1986), Shiller (1984), Summers (1986) among others.

the highly liquid S&P 500 index options along with the “model-free” approach. The VIX index time series also covers the period from January 1996 to September 2008 for a total of 153 observations. Finally, we performed an Augmented Dickey-Fuller test for nonstationarity of the stock return and VIX and the stationarity hypothesis was not rejected.

5.3 Causality tests

To test the linear causality between S&P 500 excess return and VIX, we consider the following linear mean regression

$$exr_{t+\tau} = \mu_\tau + \beta_\tau exr_t + \alpha_\tau VIX_t + \varepsilon_{t+\tau},$$

where $exr_{t+\tau}$ is the excess return τ months ahead and VIX_t represents VIX at time t . In the empirical application, we take $\tau = 1, 2, 3, 6,$ and 9 months. VIX does not linearly Granger cause the excess return τ periods ahead if $H_0 : \alpha_\tau = 0$. We use the standard t -statistic to test the null hypothesis H_0 . To avoid the impact of the dependence in the error terms on our inference, the t -statistic is based on the commonly used HAC robust variance estimator. The results of linear causality (predictability) tests between stock excess returns and VIX are presented in Table 4; see the second row LIN in Table 4. At 5% significance level, we find convincing evidence that excess return can not be predicted at both short and long-run horizons using VIX.

Now, to test for the presence of nonlinear causality (predictability) we consider the following null hypothesis:

$$H_0 : \Pr \{F(exr_{t+\tau} | exr_t, VIX_t) = F(exr_{t+\tau} | exr_t)\} = 1$$

that we test against the alternative hypothesis

$$H_1 : \Pr \{F(exr_{t+\tau} | exr_t, VIX_t) = F(exr_{t+\tau} | exr_t)\} < 1.$$

Table 4: P-values for linear and nonlinear causality tests between Return at different horizons and Volatility Index (VIX).

Test statistic / Horizon Return	1 Month	2 Months	3 Months	6 Months	9 Months
LIN	0.433	0.133	0.888	0.954	0.995
BT $c_1 = c_2 = 1.5$	0.000	0.000	0.010	0.000	0.000
BT $c_1 = c_2 = 1.2$	0.000	0.000	0.015	0.000	0.000
BT $c_1 = c_2 = 1$	0.000	0.005	0.025	0.010	0.000
BT $c_1 = 0.85, c_2 = 0.7$	0.000	0.010	0.035	0.036	0.000
BT $c_1 = 0.75, c_2 = 0.6$	0.000	0.045	0.085	0.061	0.005

LIN and BT correspond to linear test and our nonparametric test, respectively. c_1 and c_2 refer to the constants in the bandwidth parameters.

The results of nonlinear causality (predictability) tests between stock excess return and VIX are also presented in Table 4; see the rows BT of Table 4. Before we start discussing our results, we

have to mention that the data are standardized and the weighting function $w(\cdot)$ is the same like the one used in the simulation study; see Section 4. Further, five different combinations for the values of c_1 and c_2 are considered. We have seen in the simulation study that our nonparametric test has generally good properties (size and power) when $c_1 = c_2 = 1$. Therefore, our decision rule will be typically based on the results corresponding to $c_1 = c_2 = 1$. At 5% and even 1% significance levels, our nonparametric test show that VIX predicts stock excess returns both at short and long-run horizons.³

6 Conclusion

We propose a new statistic to test the conditional independence and Granger non-causality between two random variables. Our approach is based on the comparison of conditional distribution functions and the test statistic is defined using an L_2 metric. We use the Nadaraya-Watson approach to estimate the conditional distribution functions. We establish the asymptotic size and power properties of the new test and we motivate the validity of the local bootstrap. Our test is easy to implement, has power against alternatives at distance $T^{-1/2}h^{-(d_1+d_3)/4}$ compared to the test of Su and White (2008), which has power only for alternatives at distance $T^{-1/2}h^{-d/4}$, where $d = d_1 + d_2 + d_3$. It also has the same performance as the one of the tests of Su and White (2007) and Bouezmarni, Rombouts, and Taamouti (2012). We ran a simulation study to investigate the finite sample properties and the results show that the test behaves quite well in terms of size and power.

We illustrate the practical relevance of our nonparametric test by considering an empirical application where we examine Granger non-causality between S&P500 Index returns and volatility index (VIX). Contrary to the linear causality analysis which is based on the conventional t -test, we find that VIX predicts stock excess returns both at short and long-run horizons.

Finally, our test can be extended to data with mixed variables, i.e., continuous and discrete variables, by using the estimator proposed by Li and Racine (2008). Also, a practical bandwidth choice for the conditional test and an extensive comparison with the existing tests need further study.

7 Appendix

We provide the proofs of the theoretical results described in Section 3. The main tool in the proof of Theorem 1 and Propositions 1 and 2 is the asymptotic normality of U-statistics. To prove Theorem 1 and Proposition 2, we use Theorem 1 of Tenreiro (1997). To show the validity of the local smoothed bootstrap in Proposition 3, we use Theorem 1 of Hall (1984). The proofs are in general inspired from the ones in Ait-Sahalia, Bickel, and Stoke (2001) and Tenreiro (1997), of course with adapted calculations for our test.

³Other results about testing stock return predictability using variance risk premium are available from the authors upon request. The variance risk premium is measured by the difference between risk-neutral and physical (historical) variances. The results using our nonparametric test show that the variance risk premium helps to predict excess returns at long horizons, but not a short horizons.

Now, we establish the asymptotic normality of the test statistic $\hat{\Gamma}$ defined in (6). This test statistic can be rewritten as follows

$$\hat{\Gamma} = \int \left\{ \hat{F}_{h_1}(y|x, z) - \hat{F}_{h_2}(y|x) \right\}^2 w(x, z) dF_T(v),$$

where F_T is the empirical distribution function of the random vector V_t . Let us define the following pseudo-statistic

$$\tilde{\Gamma} = \int \left\{ \hat{F}_{h_1}(y|x, z) - \hat{F}_{h_2}(y|x) \right\}^2 w(x, z) dF(v), \quad (13)$$

where the empirical distribution function $F_T(v)$ in $\hat{\Gamma}$ is replaced by the true distribution function $F(v)$. We show, see Lemma 5, that replacing $F_T(v)$ by $F(v)$ will not affect the asymptotic normality of the test statistics $\hat{\Gamma}$. We begin by studying the asymptotic distribution of $\tilde{\Gamma}$. To do so, we consider the following term:

$$\Psi(t) = \int \phi^2(t, v) w(x, z) dF(v),$$

where

$$\phi(t, v) = \frac{F(y|x, z)g(x, z) + t\mu_1(x, y, z)}{g(x, z) + tg_1(x, z)} - \frac{F(y|x)g^*(x) + t\mu_1^*(x, y)}{g^*(x) + tg_1^*(x)}.$$

Using Taylor's expansion, we have

$$\Psi(t) = \Psi(0) + t\Psi'(0) + \frac{t^2}{2}\Psi''(0) + \frac{t^2}{6}\Psi'''(t^*),$$

where $0 \leq t^* \leq t$. Under H_0 , we have $\phi(0, v) = 0$. Then, $\Psi(0) = 0$ and $\Psi'(0) = 0$. As in Ait-Sahalia, Bickel, and Stoke (2001), we can show that

$$\Psi'''(t^*) = O(\|g_1(x, z)\|^3 + \|g_1^*(x)\|^3).$$

Therefore, at $t = 1$, we obtain

$$\Psi(1) = \int \left(\frac{d\phi(0, v)}{dt} \right)^2 w(x, z) dF(v) + O(\|g(x, z)\|^3 + \|g_1^*(x)\|^3). \quad (14)$$

Now, consider the terms:

$$\mu_1(x, y, z) = \sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t) \mathbb{I}_{A_{Y_t}}(y) - F(y|x, z)g(x, z), \quad \mu_1^*(x, y) = \sum_{t=1}^T K_{h_2}^*(x - X_t) \mathbb{I}_{A_{Y_t}}(y) - F(y|x)g^*(x)$$

and

$$g_1(x, z) = \hat{g}(x, y) - g(x, z), \quad g_1^*(x) = \hat{g}^*(x) - g^*(x).$$

Hence, using these notations and (14) we obtain

$$\tilde{\Gamma} = \int \left(\frac{1}{n} \sum_{i=1}^n J(v_t, v) \right)^2 w(x, z) dF(v) + O(\|\hat{g}(x, y) - g(x, z)\|^3 + \|\hat{g}^*(x) - g^*(x)\|^3),$$

where

$$J(V_t, v) = K_{h_1}(\bar{v} - \bar{V}_t) \left[\frac{1}{g(x, y)} \left(\mathbb{I}_{A_{Y_t}}(y) - F(y|x, z) \right) \right] - \frac{1}{g^*(x)} K_{h_2}^*(\bar{x} - \bar{X}_t) \left[\left(\mathbb{I}_{A_{Y_t}}(y) - F(y|x) \right) \right].$$

Under Assumption A.2, the term $O(\|\hat{g}(x, y) - g(x, z)\|^3 + \|\hat{g}^*(x) - g^*(x)\|^3)$ is negligible. Let us denote,

$$J^*(V_t, v) = J(V_t, v) - \mathbb{E}(J(V_t, v)).$$

If we ignore the negligible term, the pseudo-statistic Γ can be written as follows

$$\begin{aligned} \tilde{\Gamma} &\approx \frac{1}{T^2} \int \left(\sum_{t=1}^T J(V_t, v) \right)^2 w(x, z) dF(v) \\ &= \frac{2}{T^2} \sum_{t < s}^T \int J(V_t, v) J(V_s, v) w(x, z) dF(v) + \frac{1}{T^2} \left\{ \sum_{t=1}^T \int J^2(V_t, v) w(x, z) dF(v) \right\} \\ &= \frac{2}{T^2} \sum_{t < s}^T \int J^*(V_t, v) J^*(V_s, v) w(x, z) dF(v) + \frac{2}{T^2} \left\{ (T-1) \sum_{t=1}^T \int J^*(V_t, v) \mathbb{E}(J(V_1, v)) w(x, z) dF(v) \right\} \\ &\quad + \frac{1}{T^2} \left\{ T(T-1) \int \mathbb{E}^2(J(V_1, v)) w(x, z) dF(v) \right\} + \frac{1}{T^2} \left\{ \sum_{t=1}^T \int J^2(V_t, v) w(x, z) dF(v) \right\} \\ &= 2T^{-1} h_1^{-\frac{(d_1+d_3)}{2}} \left\{ T^{-1} \sum_{t < s}^T H_T(V_t, V_s) \right\} + 2T^{-1/2} (1 - T^{-1}) h_1^r \left\{ T^{-1/2} \sum_{t=1}^T G_T(V_t) \right\} + T^{-1} B_T + N_T \\ &\equiv 2T^{-1} h_1^{-\frac{(d_1+d_3)}{2}} T_{11} + 2T^{-1/2} (1 - T^{-1}) h_1^r T_{12} + T^{-1} B_T + N_T \end{aligned} \quad (15)$$

where

$$B_T = \frac{1}{T} \left\{ \sum_{t=1}^T \int J^2(V_t, v) w(x, z) dF(v) \right\}, \quad N_T = \frac{1}{T^2} \left\{ T(T-1) \int \mathbb{E}^2(J(V_1, v)) w(x, z) dF(v) \right\}$$

$$T_{11} = T^{-1} \sum_{1 \leq t < s \leq T} H_T(V_t, V_s), \quad T_{12} = T^{-1/2} \sum_{t=1}^T G_T(V_t), \quad (16)$$

with

$$H_T(a, b) = h_1^{-\frac{(d_1+d_3)}{2}} \int J^*(a, v) J^*(b, v) w(x, z) dF(v) \text{ and } G_T(a) = h_1^{-r} \int J^*(a, v) \mathbb{E}(J(V_1, v)) w(x, z) dF(v).$$

Notice that the term T_{11} is a degenerate U-statistic such that $H(a, b) = H(b, a)$ and $\mathbb{E}(H(V_t, b)) = 0$. The central limit theorem for U-statistics is developed in Yoshihara (1976), Denker and Keller (1983), Tenreiro (1997), and Fan and Li (1999) among others. We apply Theorem 1 of Tenreiro (1997) to show that the term T_{11} is asymptotically normal. The variance of T_{11} is $\tilde{\sigma}^2 = \frac{1}{2} \mathbb{E} \left[H_T(V_0, \tilde{V}_0) \right]^2$, for $\{\tilde{V}_t, t \geq 0\}$ an *i.i.d.* sequence where \tilde{V}_t is an independent copy of V_t . We also show that under Assumption **A.2.2**, T_{12} is negligible. Further, the term B_T represents the bias term in the test statistic and it is very important in finite samples, when bootstrap is used to calculate the *p-values*. The term N_T is deterministic and negligible. To sum up, the test statistic is normal with mean and variance given by B_T and σ^2 respectively.

Now, let us show the asymptotic normality of T_{11} . To do so, we need to check the conditions of Theorem 1 in Tenreiro (1997).

Lemma 1 Under Assumptions **A.1-A.2** and H_0 , we have

$$T_{11} \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}^2),$$

where

$$\tilde{\sigma}^2 = \frac{C}{2} \int_{\tilde{v}_0} w^2(\tilde{v}_0) \left[\int_{\tilde{y}_0} \int_{y_0} h^2(\tilde{y}_0, y_0, \tilde{v}_0) f(\tilde{y}_0|\tilde{v}_0) f(y_0|\tilde{v}_0) d\tilde{y}_0 dy_0 \right] g^2(\tilde{v}_0) d\tilde{v}_0$$

with $C = \int_{a_1, a_3} \left(\int_{b_1, b_3} K(\bar{b} + \bar{a}) K(\bar{b}) db_1 db_3 \right)^2 da_1 da_3$ and

$$h(\tilde{y}_0, y_0, \tilde{v}_0) = \int \left(I_{\{y_0 \leq y\}} - F(y|\tilde{v}_0) \right) \left(I_{\{\tilde{y}_0 \leq y_0\}} - F(y_0|\tilde{v}_0) \right) f(y|\tilde{v}_0) dy.$$

■

Proof. Now, let us check the conditions **(iii)**-**(vi)** in Theorem 1 of Tenreiro (1997), which are **(iii)** $u_T(4 + \delta_0) = O(T^{\gamma_0})$; **(iv)** $v_T(2) = o(1)$; **(v)** $w_T(2 + \delta_0/2) = o(T^{1/2})$; **(vi)** $z_T(2)T^{\gamma_1} = O(1)$, for some $\delta_0, \gamma_1 > 0$ and $\gamma_0 < 1/2$ and where

$$u_T(p) \equiv \max\{\max_{1 \leq t \leq T} \|h_T(V_t, V_0)\|_p, \|h_T(V_t, \tilde{V}_0)\|_p\},$$

$$v_T(p) \equiv \max\{\max_{1 \leq t \leq T} \|G_{T0}(V_t, V_0)\|_p, \|G_{T0}(V_0, \tilde{V}_0)\|_p\},$$

$$w_T(p) \equiv \|G_{T0}(V_0, V_0)\|_p,$$

$$z_T(p) \equiv \max_{\substack{1 \leq t_1 \leq T \\ 1 \leq t_2 \leq T}} \max\{\|G_{Tt_2}(V_{t_1}, V_0)\|_p, \|G_{Tt_2}(V_0, V_{t_1})\|_p, \|G_{Tt_2}(V_0, \tilde{V}_0)\|_p\},$$

with $G_{Tt}(u_1, u_2) \equiv \mathbb{E}[h_T(V_t, u_1)h_T(V_0, u_2)]$ and $\|\cdot\|_p \equiv \{\mathbb{E}|\cdot|_p\}^{1/p}$. In addition, in Tenreiro (1997) $\mathbb{E}[h_T(V_0, \tilde{V}_0)]^2 = 2\tilde{\sigma}^2 + o(1)$.

To check these conditions, we first need to calculate $\|H_T(V_t, V_0)\|_p = \mathbb{E}^{1/p}|H_T(V_t, V_0)|^p$ and $\|G_T(V_t, V_0)\|_p$, where $G_T(u, v) = \mathbb{E}(H_T(V_0, u)H_T(V_0, v))$. We have,

$$\begin{aligned} \mathbb{E}(|H_T(V_t, V_0)|^p) &\approx h_1^{\frac{p(d_1+d_3)}{2}} \int \int \left| \int \frac{1}{g^2(\bar{v})} K_{h_1}(\bar{v} - \bar{v}_t) K_{h_1}(\bar{v} - \bar{v}_0) (I_{\{y_t \leq y\}} - F(y|\bar{v})) \right. \\ &\quad \left. (I_{\{y_0 \leq y\}} - F(y|\bar{v})) w(x, z) dF(v) \right|^p f(v_t, v_0) dv_t dv_0 \\ &= h_1^{\frac{-p(d_1+d_3)}{2}} \int \int \left| \int \frac{1}{g^2(\bar{v})} K((\bar{v} - \bar{v}_t)/h_1) K((\bar{v} - \bar{v}_0)/h_1) (I_{\{y_t \leq y\}} - F(y|\bar{v})) \right. \\ &\quad \left. (I_{\{y_0 \leq y\}} - F(y|\bar{v})) w(x, z) dF(v) \right|^p f(v_t, v_0) dv_t dv_0. \end{aligned}$$

By change of variables, as for $\mathbb{E} \left[H_T(V_0, \tilde{V}_0) \right]^2$ given below, we can show that $|H_T(V_t, V_0)|^p = O\left(h_1^{(d_1+d_3)(1-p/2)}\right)$. Thus, $\|H_T(V_t, V_0)\|_p = O\left(h_1^{(d_1+d_3)(1/p-1/2)}\right)$. Using the same argument, we can show that $\|H_T(V_0, \tilde{V}_0)\|_p = O\left(h_1^{(d_1+d_3)(1/p-1/2)}\right)$. Hence, condition **(iii)** is fulfilled.

Let us now calculate the term $G_T(u, v)$,

$$\begin{aligned}
G_T(u, v) &= \mathbb{E}(H_T(V_0, u)H_T(V_0, v)) \\
&\approx h_1^{(d_1+d_3)} \mathbb{E} \left(\int \int K_{h_1}(\bar{\xi} - \bar{V}_0) (I_{\{Y_0 \leq \xi_2\}} - F(\xi_2|\bar{\xi})) K_{h_1}(\bar{\xi} - \bar{u}) (I_{\{u_2 \leq \xi_2\}} - F(\xi_2|\bar{\xi})) \right. \\
&\quad \left. K_{h_1}(\bar{\xi} - \bar{V}_0) (I_{\{Y_0 \leq \bar{\xi}_2\}} - F(\bar{\xi}_2|\bar{\xi})) K_{h_1}(\bar{\xi} - \bar{v}) (I_{\{v_2 \leq \bar{\xi}_2\}} - F(\bar{\xi}_2|\bar{\xi})) \alpha_{\bar{u}}(\xi) \alpha_{\bar{v}}(\xi) d\xi d\bar{\xi} \right) \\
&\leq Ch_1^{-3(d_1+d_3)} \int \int \int K((\bar{\xi} - \bar{\xi}_0)/h_1) K((\bar{\xi} - \bar{u})/h_1) K((\bar{\xi} - \bar{\xi}_0)/h_1) \\
&\quad K((\bar{\xi}^+ - v^+)/h_1) d\xi d\bar{\xi} d\xi_0,
\end{aligned}$$

where $\alpha_{\bar{u}}(\xi) = \frac{w(\xi)f(\xi)}{g^2(\xi)}$ and $\alpha_{\bar{v}}(\xi) = \frac{w(\bar{\xi})f(\bar{\xi})}{g^2(\bar{\xi})}$. By the change of variables, $\xi = \xi_0 + h_1\tau$, $\bar{\xi} = \xi_0 + h_1(\tau + \bar{\tau})$ and $\xi_0 = u + h_1(\tau_0 - \tau)$, we obtain

$$G_T(u, v) \leq C \int \int \int K(\tau^+) K(\tau^+ + \bar{\tau}^+) K\left(\tau_0^+ K(\tau_0^+ + \bar{\tau}^+ + \frac{u-v}{h_1})\right) d\tau d\bar{\tau} d\tau_0 + o(h_1^{d_1+d_3}).$$

Thus,

$$\|G_T(V_t, V_0)\|_p = O\left(h^{(d_1+d_3)/p}\right) \quad \text{and} \quad \|G_T(\tilde{V}_0, V_0)\|_p = O\left(h^{(d_1+d_3)/p}\right).$$

Hence, $v_T(p) = O(h^{d/p})$. Following the same steps, we can show that $w_T(p)$ is bounded and $z_T(p) \leq Ch_1^{d_1+d_3}$. Therefore, conditions **(iv)**, **(v)** and **(vi)** are fulfilled.

Now, to calculate the variance, observe that the product $J(V_t, v) \times J(V_s, v)$ is composed of four terms and that the dominant one is

$$\frac{1}{g^2(\bar{v})} K_{h_1}(\bar{v} - \bar{V}_t) \times K_{h_1}(\bar{v} - \bar{V}_s) \left(I_{A_{Y_t}}(y) - F(y|\bar{v}) \right) \left(I_{A_{Y_s}}(y) - F(y|\bar{v}) \right).$$

Thus, we have

$$\begin{aligned}
\mathbb{E} \left[H_T(V_0, \tilde{V}_0) \right]^2 &= h_1^{-3(d_1+d_3)} \int_{v_0, \tilde{v}_0} \left\{ \int_v K\left(\frac{\bar{v} - \bar{v}_0}{h_1}\right) K\left(\frac{\bar{v} - \tilde{v}_0}{h_1}\right) \right. \\
&\quad \left. (I_{\{y_0 \leq y\}} - F(y|\bar{v})) (I_{\{\tilde{y}_0 \leq y\}} - F(y|\bar{v})) \varphi(\bar{v}) f(v) dv \right\}^2 f(v_0) f(\tilde{v}_0) dv_0 d\tilde{v}_0 + o(1),
\end{aligned}$$

where $\varphi(\bar{v}) = w(\bar{v})/g^2(\bar{v})$. Two changes of variables are needed. The first one is $\tilde{v}_0 = (\tilde{x}_0, \tilde{z}_0) = \bar{v}_0 + h_1 \bar{a}$, ($d\tilde{v}_0 = h_1^{d_1+d_3} da$) with $a = (a_1, a_2, a_3)$ and $a_2 = \tilde{y}_0$. The second one is $\bar{v} = \bar{v}_0 + h_1(\bar{b} + \bar{a})$, ($dv = h_1^{d_1+d_3} db$) with $b = (b_1, b_2, b_3)$ and $b_2 = y$. Using Taylor expansion, we obtain

$$2\sigma^2 \equiv \mathbb{E} \left[H_T(V_0, \tilde{V}_0) \right]^2 = C \int_{v_0} w^2(\bar{v}_0) \int_{\tilde{y}_0} h^2(\tilde{y}_0, y_0, \bar{v}_0) f(x_0, \tilde{y}_0, z_0) f(v_0) d\tilde{y}_0 dy_0 d\bar{v}_0 + o(1),$$

where

$$\begin{aligned}
h(\tilde{y}_0, y_0, \bar{v}_0) &= \int (I_{\{y_0 \leq y\}} - F(y|\bar{v}_0)) (I_{\{\tilde{y}_0 \leq y_0\}} - F(y_0|\bar{v}_0)) f(y|\bar{v}_0) dy \\
&= \int_{\max(y_0, \tilde{y}_0)}^{\infty} f(y|\bar{v}_0) dy - \int_{y_0}^{\infty} F(y|\bar{v}_0) f(y|\bar{v}_0) dy \int_{\tilde{y}_0}^{\infty} F(y|\bar{v}_0) f(y|\bar{v}_0) dy + \int F^2(y|\bar{v}_0) f(y|\bar{v}_0) dy \\
&= \frac{1}{3} + \frac{1}{2} [F^2(y_0|\bar{v}_0) + F^2(\tilde{y}_0|\bar{v}_0)] - F(\max(y_0, \tilde{y}_0)|\bar{v}_0).
\end{aligned}$$

■

The following lemma provides the asymptotic bias of the pseudo-statistic $\tilde{\Gamma}$.

Lemma 2 *Under assumptions A.1-A.2 and H_0 , we have*

$$Th_1^{\frac{d_1+d_3}{2}} (T^{-1}B_T - D) = o_p(1),$$

where the terms D and B_T are defined in (7) and (16), respectively. ■

Proof. We start with the calculation of the expectation of B_T . We have

$$\begin{aligned} \mathbb{E}(B_T) &= \int \mathbb{E} \left(\frac{K_{h_1}(\bar{v} - \bar{V}_t) (I_{A_t}(y) - F(y|\bar{v}))}{g(\bar{v})} - \frac{K_{h_2}^*(x - X_t) (I_{A_t}(y) - F(y|x))}{g^*(x)} \right)^2 w(\bar{v})f(v)dv \\ &= \int \mathbb{E} \left(\frac{K_{h_1}(\bar{v} - \bar{V}_t) (I_{A_t}(y) - F(y|\bar{v}))}{g(\bar{v})} \right)^2 w(\bar{v})f(v)dv \\ &\quad + \int \mathbb{E} \left(\frac{K_{h_2}^*(x - X_t) (I_{A_t}(y) - F(y|x))}{g^*(x)} \right)^2 w(\bar{v})f(v)dv \\ &\quad - 2 \int \mathbb{E} \left(\frac{K_{h_1}(\bar{v} - \bar{V}_t) (I_{A_t}(y) - F(y|\bar{v}))}{g(\bar{v})} \right) \left(\frac{K_{h_2}^*(x - X_t) (I_{A_t}(y) - F(y|x))}{g^*(x)} \right) w(\bar{v})f(v)dv \\ &= D_1 + D_2 + D_3. \end{aligned}$$

First, from the change of variables, $\bar{v}' = (\bar{v} - \bar{v}_t)/h_1$ and $v' = (v'_1, v'_2, v'_3)$, with $v'_2 = y$, and using Taylor expansion, we obtain

$$\begin{aligned} D_1 &= h_1^{-(d_1+d_3)} \int \int \frac{K^2(\bar{v}') \{I_{A_t}(v'_2) - F(v'_2|\bar{v}_t)\}^2}{g(\bar{v}_t)^2} w(\bar{v}_t)f(x_t, v'_2, z_t)f(v_t) dv_t dv' + o(1) \\ &= h_1^{-(d_1+d_3)} \int K^2(\bar{v}')d\bar{v}' \int_{v_t} \frac{w(\bar{v}_t)f(v_t)}{g(\bar{v}_t)^2} \int_{v'_2} \{I_{A_t}(v'_2) - F(v'_2|\bar{v}_t)\}^2 f(x_t, v'_2, z_t)dv'_2 dv_t. \end{aligned}$$

Since

$$\begin{aligned} \int_{v'_2} \{I_{A_t}(v'_2) - F(v'_2|\bar{v}_t)\}^2 f(x_t, v'_2, z_t)dv'_2 &= g(\bar{v}_t) \int_{v'_2 \geq y_t} \{1 - F(v'_2|\bar{v}_t)\}^2 f(v'_2|\bar{v}_t)dv'_2 \\ &= \frac{1}{3}g(\bar{v}_t)(1 - F(y_t|\bar{v}_t))^3, \end{aligned}$$

we get

$$D_1 = C_1 h_1^{-(d_1+d_3)} \int_{v_t} \frac{w(\bar{v}_t)}{g(\bar{v}_t)} (1 - F(y_t|\bar{v}_t))^3 f(v_t)dv_t,$$

where $C_1 = \frac{1}{3} \int K^2(\bar{v}')d\bar{v}'$.

Second, by the change of variable $(x - x_t)/h_2 = x'$ and using Taylor expansion, we have

$$D_2 = h_2^{-d_1} \int_{x', y, z} \int_{x_t, y_t} \frac{1}{(g^*(x_t))^2} K^{*2}(x') \{I_{A_t}(y) - F(y|x)\}^2 w(x, z)f(x_t, y, z)f(x_t, y_t)dx'dydzdx_tdy_t.$$

Under H_0 , we get

$$\int_y f(x_t, y, z)(I_{A_t}(y) - F(y|x_t)) dy = \frac{1}{3}(1 - F(y_t|x_t))^3 g(x_t, z),$$

and hence

$$D_2 = h_2^{-d_1} C_2 \int_{x_t, y_t} \frac{w^*(x_t)}{g^*(x_t)} (1 - F(y_t|x_t))^3 f(x_t, y_t) dx_t dy_t,$$

for $C_2 = \frac{1}{3} \int K^{*2}(x) dx$ and $w^*(x_t) = \int w(x_t, z) g(z|x_t) dz$.

Finally, using the following change of variables $x = x_t + h_2 x'$ and $z = z_t + h_1 z'$, under H_0 , we obtain

$$\begin{aligned} -\frac{1}{2} D_3 &= \int \int \left\{ \frac{K_{h_1}(\bar{v} - \bar{v}_t) \{I_{A_t}(y) - F(y|\bar{v})\}}{g(\bar{v})} \times \frac{K_{h_2}^*(x - x_t) \{I_{A_t}(y) - F(y|x)\}}{g^*(x)} \right\} w(\bar{v}) f(v) f(v_t) dv dv_t \\ &= h_1^{-d_1} \int \int \frac{K(\frac{h_2}{h_1} x', z') K^*(x') \{I_{A_t}(y) - F(y|\bar{v}_t)\}^2}{g(\bar{v}_t) g^*(x_t)} w(\bar{v}_t) f(x_t, y, z_t) f(v_t) dv_t dx' dz' dy. \end{aligned}$$

Since $h_2 = o(h_1)$ and $\int_y \{I_{A_t}(y) - F(y|\bar{v}_t)\}^2 f(x_t, y, z_t) = \frac{1}{3} (1 - F(y_t|\bar{v}_t))^3 g(\bar{v}_t)$, we obtain

$$D_3 = -\frac{2}{3} C_3 h_1^{-d_1} \int \frac{w(\bar{v}_t)}{g^*(x_t)} (1 - F(y_t|\bar{v}_t))^3 f(v_t) dv_t,$$

where $C_3 = K(0)$. Also, using similar argument to the one above, we can show that

$$\text{Var} \left(T h_1^{\frac{d_1+d_3}{2}} (T^{-1} B_T - D) \right) \equiv \frac{1}{T^2} \sum_{t=1}^T \int \mathbb{E}(J_t^2) w(\bar{v}) f(v) dv = o(1),$$

and this concludes the proof.

Lemma 3 *Under assumptions A.1-A.2 and H_0 , we have*

$$T h_1^{(d_1+d_3)/2} h_1^r T_{12} = o_p(1),$$

where the term T_{12} is defined in (16). ■

Proof. Observe that by construction we have $\mathbb{E}(G_T(V_t)) = 0$. We denote by $\tilde{\sigma}_1^2$ the variance of T_{12} . To calculate this variance we need to evaluate the covariance between $G_T(V_t)$ and $G_T(V_0)$. First, under the assumption $h_2 = o(h_1)$, we have

$$\mathbb{E}(J(V_t, v)) = h_1^r \gamma(v) + o(h_1^r),$$

where $\gamma(v)$ is a function of the kernel K , of the r^{th} derivative of F and of g . Second,

$$\begin{aligned} \text{Cov}(G_T(V_t), G_T(V_s)) &= \mathbb{E}(G_T(V_t) G_T(V_s)) \\ &\approx \mathbb{E} \left(\int J(V_t, v) J(V_s, v') \xi(v) \xi(v') dv dv' \right) \\ &\quad - 2 \int J(v_t, v) \mathbb{E}(J(v_s, v')) \xi(v) \xi(v') f(v_t, v_s) dv_t dv_s dv dv' \\ &\quad + \int \mathbb{E}(J(v_t, v)) \mathbb{E}(J(v_s, v')) \xi(v) \xi(v') f(v_t, v_s) dv_t dv_s dv dv' + o(1) \\ &= \int J(v_t, v) J(v_s, v') \xi(v) \xi(v') f(v_t, v_s) dv_t dv_s dv dv' - \left(\int \mathbb{E}(J(v_s, v)) \xi(v) dv \right)^2 + o_p(1), \\ &= I - II + o(1) \end{aligned}$$

where $\xi(v) = \gamma(v)w(\bar{v})f(v)$. We next need to compute the terms I and II . Let us first calculate the term I . We have,

$$\begin{aligned} I &= \int \left\{ \frac{K_{h_1}(\bar{v} - \bar{v}_t) (\mathbb{I}_{A_t}(y) - F(y|\bar{v}))}{g(\bar{v})} - \frac{K_{h_2}^*(x - X_t) (\mathbb{I}_{A_t}(y) - F(y|x))}{g^*(x)} \right\} \\ &\quad \times \left\{ \frac{K_{h_1}(\bar{v}' - \bar{v}_s) (\mathbb{I}_{A_s}(y') - F(y'|\bar{v}'))}{g(\bar{v}')} - \frac{K_{h_2}^*(x' - x_s) (\mathbb{I}_{A_s}(y') - F(y'|x'))}{g^*(x')} \right\} \\ &\quad \times \xi(v)\xi(v')f(v_t, v_s)dv_t dv_s dv dv' + o_p(1). \end{aligned}$$

From the change of variables $(\bar{v} - \bar{v}_t)/h_1 = (a_1, a_3) \equiv \bar{a}; (a_2 = y)$ and $(\bar{v}' - \bar{v}_s)/h_1 = (b_1, b_3) \equiv \bar{b}, b_2 = y'$, we obtain

$$\begin{aligned} I &= \int \left\{ \frac{K(\bar{a}) (\mathbb{I}_{A_t}(a_2) - F(y|\bar{v}_t))}{g(\bar{v}_t)} - \frac{h_1^{d_1+d_3} K^*(h_1 x_t/h_2) (\mathbb{I}_{A_t}(a_2) - F(a_2|x_t))}{h_2^{d_1} g^*(x_t)} \right\} \\ &\quad \left\{ \frac{K(\bar{b}) (\mathbb{I}_{A_s}(b_2) - F(b_2|\bar{v}'_s))}{g(\bar{v}_s)} - \frac{h_1^{d_1+d_3} K^*(h_1 x_s/h_2) (\mathbb{I}_{A_s}(b_2) - F(b_2|x_s))}{h_2^{d_1} g^*(x_s)} \right\} \\ &\quad \xi(x_t, a_2, z_t)\xi(x'_t, b_2, z'_t)f(v_t, v_s)dv_t dv_s dadb + o_p(1). \end{aligned}$$

If we assume that $h_1^{d_1+d_3}/h_2^{d_1} = o(1)$, then

$$\begin{aligned} I &= \int \left(\frac{K(\bar{a}) (\mathbb{I}_{A_t}(a_2) - F(y|\bar{v}_t))}{g(\bar{v}_t)} \right) \left(\frac{K(\bar{b}) (\mathbb{I}_{A_s}(b_2) - F(b_2|\bar{v}'_s))}{g(\bar{v}_s)} \right) \\ &\quad \xi(x_t, a_2, z_t)\xi(x'_t, b_2, z'_t)f(v_t, v_s)dv_t dv_s dadb + o_p(1) \\ &= \int \zeta(v_t)\zeta(v_s)f(v_t, v_s)dv_t dv_s + o_p(1) \\ &= \mathbb{E}(\zeta(V_t)\zeta(V_s)) + o_p(1), \end{aligned}$$

where $\zeta(v_t) = C^*2 \frac{\delta(v_t)}{g(\bar{v}_t)}$ with $C^* = \int_{\bar{a}} K(\bar{a})d\bar{a}$ and $\delta(v_t) = \int_{a_2} (\mathbb{I}_{A_t}(a_2) - F(y|\bar{v}_t)) \xi(x_t, a_2, z_t)da_2$. Using similar arguments, we show that

$$II = \left(\int_{v_t} \zeta(v_t)f(v_t)dv_t \right)^2 + o_p(1) = \mathbb{E}(\zeta(V_t))^2 + o_p(1)$$

Consequently, using Assumption **A1.1**, we obtain

$$\tilde{\sigma}_1^2 = Var(\zeta(V_0)) + 2 \sum_{i \geq 1} Cov(\zeta(V_1), \zeta(V_{1+i})) < \infty.$$

Hence, under Assumption **A2.2**, we conclude the proof of Lemma 3.

Lemma 4 *Under Assumptions **A.1-A.2** and H_0 , we have*

$$Th_1^{(d_1+d_3)/2}N_T = o(1),$$

where the term N_T is defined in (16). ■

Proof. The proof is straightforward, since $\mathbb{E}(J(V_t, v)) = O(h_1^r)$ and $Th_1^{(d_1+d_3)/2+2r} \rightarrow 0$.

Lemma 5 Under assumptions **A.1-A.2** and H_0 , we have

$$Th_1^{(d_1+d_3)/2}(\hat{\Gamma} - \tilde{\Gamma}) = o_p(1),$$

where $\hat{\Gamma}$ and $\tilde{\Gamma}$ are defined in (6) and (13). ■

Proof. This result follows from the same arguments in Su and White (2008). ■

Proof of Proposition 1. This result can be shown by following the same steps as in the proof of Theorem 1. However, the term N_T defined in (16), is now given by

$$\begin{aligned} N_T &= \int \mathbb{E}^2(J(V_t, v))w(x, z)dF(v) + o(1) \\ &= \int (F(y|x, z) - F(y|x))^2w(x, z)dF(v) + o(1). \end{aligned}$$

Therefore, if $\int (F(y|x, z) - F(y|x))^2w(x, z)dF(v) > 0$, we have $Th_1^{(d_1+d_3)/2}N_T \rightarrow \infty$. Hence, the test is consistent. ■

Proof of Proposition 2. First, following similar arguments as in Lemma 1-5 for the stochastic process $\{(X_{T,t}, Y_{T,t}, Z_{T,t})\}$, we can show that

$$Th_1^{(d_1+d_3)/2} \left(\hat{\Gamma} - D - \int \{F^{[T]}(y|x, z) - F^{[T]}(y|x)\}^2w(x, z) dF^{[T]}(v) \right) \xrightarrow{d} N(0, \sigma^2).$$

Second, under the alternative hypothesis, we have

$$\int \{F^{[T]}(y|x, z) - F^{[T]}(y|x)\}^2 w(x, z) dF^{[T]}(v) = \xi_T^2 \int \Delta^2(x, y, z)w(x, z) dF^{[T]}(v) + o_p(\xi_T^2).$$
■

Proof of Proposition 3. Conditionally on $\mathcal{V}_T = \{V_t\}_{t=1}^T$, the observations $\{V_t^*\}_{t=1}^T$ forms a triangular array of independent random variables. Thus, conditionally on \mathcal{V}_T , $G_T(V_t^*)$ and $H_T(V_t^*, V_t^*)$ are independent. The result of this proposition is obtained using similar argument to the one in the proof of Theorem 1, with the terms, T_{11} , T_{12} , B_T and N_T in (15) are replaced by their bootstrapped versions T_{11}^* , T_{12}^* , B_T^* and N_T^* , respectively, using the bootstrap data $\mathcal{V}_T^* = \{V_t^*\}_{t=1}^T$. Thus, conditionally on \mathcal{V}_T and using Theorem 1 of Hall (1984), we get the result in Proposition 3. ■

References

- AIT-SAHALIA, Y., P. BICKEL, AND T. STOKE (2001): “Goodness-of-Fit Tests for Kernel Regression with an Application to Option Implied Volatilities,” *Journal of Econometrics*, 105, 363–412.
- BERGSMA, W. (2011): “Nonparametric Testing of Conditional Independence by Means of the Partial Copula,” Working paper, London School of Economics and Political Science.
- BOENTE, G., AND R. FRAIMAN (1995): “Asymptotic Distribution of Smoothers Based on Local Means and Local Medians Under Dependence,” *Journal of Multivariate Analysis*, 54, 77–90.
- BOLLERSLEV, T., G. TAUCHEN, AND H. ZHOU (2009): “Expected Stock Returns and Variance Risk Premia,” *Review of Financial Studies*, 22, 4463–4492.
- BOUDJELLABA, H., J. M. DUFOUR, AND R. ROY (1992): “Testing Causality Between Two Vectors in Multivariate ARMA Models,” *Journal of the American Statistical Association*, 87, 1082–1090.
- (1994): “Simplified Conditions for Non-Causality between Two Vectors in Multivariate ARMA Models,” *Journal of Econometrics*, 63, 271–287.
- BOUEZMARNI, T., J. V. K. ROMBOUTS, AND A. TAAMOUTI (2012): “A Nonparametric Copula Based Test for Conditional Independence with Applications to Granger Causality,” *Journal of Business and Economic Statistics*, 30, 275–287.
- CAI, Z. (2002): “Regression Quantiles for Time Series,” *Econometric Theory*, 18, 169–192.
- CAMPBELL, J. Y., AND R. SHILLER (1988): “The Dividend-Price Ratio and Expectation of Future Dividends and Discount Factors,” *Review of Financial Studies*, 1, 195–228.
- CARRASCO, M., AND X. CHEN (2002): “Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models,” *Econometric Theory*, 18, 17–39.
- CENESIZOGLU, T., AND A. TIMMERMANN (2008): “Is the Distribution of Stock Returns Predictable?,” Discussion paper, HEC Montreal and UCSD.
- DE MATOS, J. A., AND M. FERNANDES (2007): “Testing the Markov Property with High Frequency Data,” *Journal of Econometrics*, 141, 44–64.
- DEHLING, H., AND M. WENDLER (2010): “Central Limit Theorem and the Bootstrap for U-Statistics of Strongly Mixing Data,” *Journal of Multivariate Analysis*, 100, 126–137.
- DELGADO, M. A., AND W. GONZÁLEZ MANTEIGA (2001): “Significance Testing in Nonparametric Regression Based on the Bootstrap,” *Annals of Statistics*, 29, 1469–1507.
- DENKER, M., AND G. KELLER (1983): “On U-Statistics and Von Mises’ Statistic for Weakly Dependent Processes,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 64, 505–552.
- DOUKHAN, P. (1995): *Mixing: Properties and Examples*. Springer-Verlag, New York.

- DUFOUR, J. M., AND E. RENAULT (1998): “Short-Run and Long-Run Causality in Time Series: Theory,” *Econometrica*, 66, 1099–1125.
- DUFOUR, J. M., AND A. TAAMOUTI (2010): “Short and Long Run Causality Measures: Theory and Inference,” *Journal of Econometrics*, 154, 42–58.
- FAMA, E. F., AND K. R. FRENCH (1988): “Dividend Yields and Expected Stock Returns,” *Journal of Financial Economics*, 22, 3–25.
- FAN, Y., AND Q. LI (1999): “Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Tests,” *Journal of Nonparametric Statistics*, 10, 245–271.
- FLORENS, J. P., AND D. FOUGÈRE (1996): “Non-Causality in Continuous Time,” *Econometrica*, 64, 1195–1212.
- FLORENS, J. P., AND M. MOUCHART (1982): “A Note on Non-Causality,” *Econometrica*, 50, 583–591.
- FRENCH, K. R., AND R. ROLL (1986): “Stock Return Variances: The Arrival of Information and the Reaction of Traders,” *Journal of Financial Economics*, 17, 5–26.
- GAO, J., AND I. GIJBELS (2008): “Bandwidth Selection in Nonparametric Kernel Testing,” *Journal of the American Statistical Association*, 103, 1584–1594.
- GEWEKE, J. (1984): *Inference and Causality in Economic Time Series*. in Z. Griliches and M. D. Intriligator, eds, *Handbook of Econometrics*, Volume 2, North-Holland, Amsterdam, pp. 1102–1144.
- GOURIÉROUX, C., AND A. MONFORT (1997): *Time Series and Dynamic Models*. Cambridge University Press, Cambridge, U.K.
- GRANGER, C. W. J. (1969): “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica*, 37, 424–438.
- (1980): “Testing for Causality: A Personal Viewpoint,” *Journal of Economic Dynamics and Control*, 2, 329–352.
- GRANGER, C. W. J., AND P. NEWBOLD (1986): *Forecasting Economic Time Series. 2nd Edition*. Academic Press, New York.
- HALL, P. (1984): “Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators,” *Journal of Multivariate Analysis*, 14, 1–16.
- HALL, P., R. C. L. WOLFF, AND Q. YAO (1999): “Methods for Estimating a Conditional Distribution Function,” *Journal of the American Statistical Association*, 94, 154–163.
- HOROWITZ, J. L., AND V. G. SPOKOINY (2001): “Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternatives,” *Econometrica*, 69, 599–631.

- HUANG, M., Y. SUN, AND H. WHITE (2013): “A Flexible Nonparametric Test for Conditional Independence,” Working paper. Dept. of Economics, UCSD.
- HUANG, T. M. (2010): “Testing Conditional Independence Using Maximal Nonlinear Conditional Correlation,” *Annals of Statistics*, 38, 2047–2091.
- KOTHARI, S. P., AND J. SHANKEN (1997): “Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis,” *Journal of Financial Economics*, 44, 169–203.
- LEE, T.-H., AND W. YANG (2012): “Money-Income Granger-Causality in Quantiles,” *Advances in Econometrics*, 30, 385–409.
- LEWELLEN, J. (2004): “Predicting Returns with Financial Ratios,” *Journal of Financial Economics*, 74, 209–235.
- LI, Q., J. LIN, AND J. S. RACINE (2013): “Optimal Bandwidth Selection for Nonparametric Conditional Distribution and Quantile Functions,” *Journal of Business and Economic Statistics*, 31, 57–65.
- LI, Q., E. MAASOUMI, AND J. S. RACINE (2009): “A Nonparametric Test for Equality of Distributions with Mixed Categorical and Continuous Data,” *Journal of Econometrics*, 148, 186–200.
- LI, Q., AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- (2008): “Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data,” *Journal of Business and Economic Statistics*, 26, 423–434.
- LINTON, O., AND P. GOZALO (2014): “Testing Conditional Independence Restrictions,” *Econometric Reviews*, 33, 523–552.
- LO, A. W., AND A. C. MACKINLAY (1988): “Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test,” *Review of Financial Studies*, 1, 41–66.
- LÜTKEPOHL, H. (1991): *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.
- MEITZ, M., AND P. SAIKKONEN (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.
- NADARAYA, E. A. (1964): “On Estimating Regression,” *Theory of Probability and its Applications*, 9, 141–142.
- NEWBOLD, P. (1982): *Causality Testing in Economics. in Time Series Analysis. Theory and Practice 1*. ed. by O. D. Anderson. Amsterdam, North-Holland.
- NEWHEY, W. K., AND K. D. WEST (1987): “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.

- PAPARODITIS, E., AND D. POLITIS (2000): “The Local Bootstrap for Kernel Estimators under General Dependence Conditions,” *Annals of the Institute of Statistical Mathematics*, 52, 139–159.
- PIERCE, D. A., AND L. D. HAUGH (1977): “Causality in Temporal Systems: Characterizations and Survey,” *Journal of Econometrics*, 5, 265–293.
- SAIDI, A., AND R. ROY (2008): “Robust Optimal Tests for Causality in Multivariate Time Series,” *Econometric Theory*, 24, 948–987.
- SHILLER, R. J. (1984): “Theories of Aggregate Stock Price Movements,” *Journal of Portfolio Management*, 10(2), 28–37.
- SILVERMAN, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.
- SIMONOFF, J. S. (1996): *Smoothing Methods in Statistics*. Springer. New York.
- SONG, K. (2009): “Testing Conditional Independence via Rosenblatt Transforms,” *Annals of Statistics*, 37, 4011–4045.
- SU, L., AND M. SPINDLER (2013): “Nonparametric Testing for Asymmetric Information,” *Journal of Business and Economics Statistics*, 31, 208–225.
- SU, L., AND H. WHITE (2007): “A Consistent Characteristic Function-Based Test for Conditional Independence,” *Journal of Econometrics*, 141, 807–834.
- (2008): “A Nonparametric Hellinger Metric Test for Conditional Independence,” *Econometric Theory*, 24, 829–864.
- (2012): “Conditional Independence Specification Testing for Dependent Processes with Local Polynomial Quantile Regression,” *Advances in Econometrics*, 29, 355–434.
- (2014): “Testing Conditional Independence via Empirical Likelihood,” *Journal of Econometrics*, forthcoming.
- SUMMERS, L. (1986): “Does the Stock Market Rationally Reflect Fundamental Values?,” *Journal of Finance*, 41, 591–601.
- TENREIRO, C. (1997): “Loi Asymptotique des Erreurs Quadratiques Integrees des Estimateurs a Noyau de la Densite et de la Regression Sous des Conditions de Dependance,” *Portugaliae Mathematica*, 54, 197–213.
- TROUNG, Y., AND C. STONE (1992): “Nonparametric Function Estimation Involving Time Series,” *Annals of Statistics*, 20, 77–97.
- WANG, X., AND Y. HONG (2013): “Characteristic Function Based Testing for Conditional Independence: A Nonparametric Regression Approach,” Working paper. Dept. of Economics, Cornell University.

WATSON, G. (1964): "Smooth Regression Analysis," *Sankhya*, 26, 359–372.

WIENER, N. (1956): *The Theory of Prediction* chap. 8, pp. 165–190. McGraw-Hill, New York.

YOSHIHARA, K. (1976): "Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 35, 237–252.