UNIVERSIDAD CARLOS III DE MADRID
working papers

Working Paper 12-17
Economic Series
January, 2012

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Nonparametric Tests for Conditional Independence Using Conditional Distributions*

Taoufik Bouezmarni† — Université de Sherbrooke
Abderrahim Taamouti‡ — Universidad Carlos III de Madrid

January 6, 2012

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 aim 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. Further, 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 model, 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.

*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 to write this paper. We thank Mohamed Ouzineb for his programming assistance. We also thank seminar participants at the Institute of Actuarial Sciences, Université Catholique de Louvain, and the participants at the 2010 Joint Statistical Meetings in Vancouver, 2011 AEA Annual meeting in Denver, 25th Congress of the European Economic Association in Glasgow, and 18th Annual Symposium of the Society for Nonlinear Dynamics and Econometrics in Novara. 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.

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 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) [see also 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 (2006), 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. Further, it is possible to have situations where the causality in low moments (mean, variance) does not exist, but it does exist in high moments. Hence, non-causality tests should be defined based on distribution functions.

Several nonparametric tests are available to test for independence between random variables, starting with the rank-based test of Hoeffding (1948), including empirical distribution-based methods such as Blum, Kiefer, and Rosenblatt (1961) or Skaug and Tjøstheim (1993), smoothing-based methods like Rosenblatt (1975), Robinson (1991), and Hong and White (2005). Further, nonparametric regression tests are also introduced by Fan and Li (1996) who develop tests for the significance of a subset of regressors and tests for the specification of the semiparametric functional form of the regression function. Fan and Li (2000) compare the power properties of various kernel based nonparametric tests with the integrated conditional moment tests of Bierens and Ploberger (1997). Delgado and González Manteiga (2001) propose a test for selecting explanatory variables in nonparametric regression. The asymptotic null distribution of the test depends on certain features of the data generating process. To estimate the critical values, they use the wild bootstrap based on nonparametric residuals. Delgado and González Manteiga (2001) [see their section 5] also propose an omnibus test of conditional independence using the weighted difference of the estimated conditional distributions under the null and the alternative. With respect to nonparametric conditional independence tests, Linton and Gozalo (1997) develop a non-pivotal nonparametric empirical distribution function based test of conditional independence. The asymptotic null distribution of the test statistic is a functional of a Gaussian process and the critical values are computed using the bootstrap. Finally, Lee and Whang (2009) provide a nonparametric test for the treatment effects conditional on covariates.

The above nonparametric independence tests are derived under an i.i.d. assumption. Only a few recent papers have been proposed to test nonparametrically for conditional independence using time series data. Su and White (2003) construct a class of smoothed empirical likelihood-based tests which are asymptotically normal under the null hypothesis and they derive their asymptotic distributions under a sequence of local alternatives. The tests are shown to possess a weak optimality property in large samples and simulation results suggest that these tests behave well in finite samples. Su and White (2007) propose a nonparametric test based on the conditional characteristic function. They work with the squared Euclidean distance and need to specify two weighting functions in the test statistic. Su and White (2008) propose a nonparametric test based on density functions and the weighted Hellinger distance. Their test is consistent, asymptotically normal under β-mixing conditions, and has power against alternatives at distance $T^{-1/2}h^{-d/4}$ where T denotes the sample size, h the bandwidth parameter and d the dimension of the vector of all variables in the study. Recently, Bouezmarni, Rombouts, and Taamouti (2011) provide a nonparametric test for conditional independence based on comparison of Bernstein copula densities using the Hellinger distance. Their test statistic does not involve a weighting function and it is asymptotically pivotal under the null hypothesis. Finally, 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. He 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.

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 statistics and we motivate the validity of the local bootstrap. We show that our conditional distribution-based test is more powerful than Su and White (2008)'s 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 model, we find that VIX index 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 hypothesis of conditional independence, the alternative hypothesis 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 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 reminder 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 to capture the dependence due to both low and high-order moments. Thus, testing the conditional independence between Y

and Z conditionally on X, corresponds to test the null hypothesis

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

against the alternative hypothesis

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

Since the conditional distribution functions $F(y \mid (X,Z))$ and $F(y \mid X)$ 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 $v = (x,y,z) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}$, $\overline{V} = (X,Z)$ and $\bar{v} = (x,z)$, then the Nadaraya-Watson estimator of the conditional distribution function of Y given X and Z is defined by

$$\hat{F}_{h_1}(y|\bar{v}) = \frac{\sum_{t=1}^{T} K_{h_1}(\bar{v} - \overline{V_t})\ \mathrm{I}_{A_{Y_t}}(y)}{\sum_{t=1}^{T} K_{h_1}(\bar{v} - \overline{V_t})}, \tag{2}$$

where $K_{h_1}(.) = h_1^{-(d_1+d_3)} K(./h)$, for $K(.)$ a kernel function, $h_1 = h_{1,n}$ is a bandwidth parameter, and $\mathrm{I}_{A_{Y_t}}(.)$ 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:

$$\hat{F}_{h_2}(y|x) = \frac{\sum_{t=1}^{T} K^*_{h_2}(x - X_t)\ \mathrm{I}_{A_{Y_t}}(y)}{\sum_{t=1}^{T} K^*_{h_2}(x - X_t)}, \tag{3}$$

where $K^*_{h_2}(.) = h_2^{-d_1} K^*(./h)$, for $K^*(.)$ 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 (1) against the alternative hypothesis (1), we propose the following test statistic which is based on the conditional distribution function estimators

$$\hat{\Gamma} = \frac{1}{T} \sum_{t=1}^{T} \left\{ \hat{F}_{h_1}(Y_t|\overline{V_t}) - \hat{F}_{h_2}(Y_t|X_t) \right\}^2 w(\overline{V_t}), \tag{4}$$

where $w(.)$ is a nonnegative weighting function of the data $\overline{V_t}$, 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 $w(.)$. In the latter case we suggest to use a large bandwidth parameter for the estimation of the conditional distribution function in the tails. The weight $w(.)$ 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; gross domestic product,...) to positive income. Further, 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 (4) depends obviously on the sample size 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 $\hat{\Gamma}$ and we derive its power function against local alternatives. We also establish the asymptotic 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 distributions. We follow the literature on U-statistics and assume $\beta-$mixing dependent variables; see Tenreiro (1997) and Fan and Li (1999) among others. To recall the definition of a $\beta-$mixing process, let's consider $\{V_t; t \in \mathbb{Z}\}$ a strictly stationary stochastic process and denote $\mathcal{F}_s^t$ the $\sigma-$algebra generated by the observations $(V_s, ..., V_t)$, for $s \leq t$. The process $\{V_t\}$ is called β-mixing or absolutely regular if

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

For more details about mixing processes, the reader can consult Doukhan (1994). 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 $\left\{V_t = (X_t, Y_t, Z_t) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3},\ t \in \mathbb{Z}\right\}$ 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 $\overline{V_t} = (X_t, Z_t)$, denoted by g^* and g respectively, are twice differentiable and bounded away from zero on the compact support of $w(.)$.

Assumption A.2 (Kernel and Bandwidth)

A2.1 The kernels K and K^* are the product of a univariate symmetric and bounded kernel $k : \mathbb{R} \to \mathbb{R}$, i.e. $K(\eta_1, ..., \eta_{d_1+d_3}) = \prod_{j=1}^{d_1+d_3} k(\eta_j)$ and $K^*(\eta_1, ..., \eta_{d_1}) = \prod_{j=1}^{d_1} k(\eta_j)$, 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 $T \to \infty$, the bandwidth parameters h_1 and h_2 are such that $h_1, h_2 \to 0$, $h_2 = o(h_1)$ and $h_1^{d_1+d_3} = o(h_2^{d_1})$. Further, as $T \to \infty$, $Th_1^{(d_1+d_3)} \to \infty$ and $Th_1^{(d_1+d_3)/2+2r} \to 0$.

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 (2002) among others]. 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. According to Assumption **A2.2,** if $h_1 = constant\, T^{-1/\psi}$ is considered, then $d_1 + d_3 < \psi < (d_1 + d_3)/2 + 2r$.

3.1 Asymptotic distribution of the test statistic

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

$$\begin{aligned}
&D_1 = C_1 h_1^{-(d_1+d_3)} \int_{v_t} \{w(\bar{v}_t)(1 - F(y_t|\bar{v}_t))/g(\overline{v_t})\} f(v_t)dv_t, \\
&D_2 = C_2 h_2^{-d_1} \int_{v_t} \{w^*(x_t)(1 - F(y_t|x_t))/g^*(x_t)\} f(x_t, y_t)dx_t dy_t, \\
&D_3 = -2C_3 h_1^{-d_1} \int_{v_t} \{w(\bar{v}_t)(1 - F(y_t|\bar{v}_t))/g^*(x_t)\} f(v_t)dv_t, \\
&D = (D_1 + D_2 + D_3)/T,
\end{aligned} \tag{5}$$

where $f(x_t, y_t) = \int f(v_t)dz_t$,

$$w^*(\bar{v}) = \int_z \frac{w(\bar{v})g(\bar{v})}{g^*(x)} dz,$$

$$C_1 = \int K^2(x, z)dxdz, \quad C_2 = \int K^{*2}(x)dx, \; C_3 = K^*(0).$$

Further, we denote

$$\sigma^2 = \frac{C}{6} \int_{v_t} \frac{w^2(\bar{v}_t)}{g(\bar{v}_t)} \{1 - F(y_t|\bar{v}_t)\}^2 (1 + 2F(y_t|\bar{v}_t)) f(v_t)dv_t, \tag{6}$$

where

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

for $\bar{a} = (a_1, a_3)$ and $\bar{b} = (b_1, b_3)$ in $\mathbb{R}^{d_1+d_3}$. The following theorem establishes the asymptotic normality of the test statistic $\hat{\Gamma}$ defined in (4). In the sequel, "$\xrightarrow{d}$" stands for convergence in distribution.

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

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

where $\hat{\Gamma}$ is given by (4) and D and σ^2 are defined in Equations (5) and (6), 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 consider the normal density function as a kernel. However, if the test is for higher dimensions, a higher order kernel is required. Now, to implement our test statistic, we have to estimate the bias terms, D_1, D_2 and D_3 and we consider the following consistent estimators:

$$\hat{D}_1 = \tfrac{C_1 h_1^{-(d_1+d_3)}}{T} \textstyle\sum_{t=1}^{T} \left\{ w(\bar{V}_t)(1 - \hat{F}_{h_1}(Y_t|\bar{V}_t))/\hat{g}(\bar{V}_t) \right\},$$

$$\hat{D}_2 = \tfrac{C_2 h_2^{-d_1}}{T} \textstyle\sum_{t=1}^{T} \left\{ \widehat{w^*}(X_t)(1 - \hat{F}_{h_2}(Y_t|X_t))/\hat{g}^*(X_t) \right\},$$

$$\hat{D}_3 = -\tfrac{2C_3 h_1^{-d_1}}{T} \textstyle\sum_{t=1}^{T} \left\{ w(\bar{V}_t)(1 - \hat{F}_{h_1}(Y_t|\bar{V}_t))/\hat{g}^*(X_t) \right\},$$

$$\hat{D} = (\hat{D}_1 + \hat{D}_2 + \hat{D}_3)/T$$

where

$$\widehat{w^*}(X_t) = \frac{\sum_{s=1}^{T} w(\bar{V}_s) K_{h_2}^*(X_t - X_s)}{\sum_{s=1}^{T} K_{h_2}^*(X_t - X_s)},$$

and $\hat{F}_{h_1}(Y_t|\overline{V_t})$, $\hat{F}_{h_2}(Y_t|X_t)$ are the Nadaraya-Watson estimators of the conditional distribution functions $F(y \mid (x, z))$ and $F(y \mid x)$, respectively. The functions $\hat{g}(.)$ and $\hat{g}^*(.)$ are consistent estimators for the density functions $g(.)$ and $g^*(.)$, respectively. Here we consider nonparametric kernel estimators of g and g^*:

$$\hat{g}(x, z) = \frac{1}{T} \sum_{t=1}^{T} h_1^{-(d_1+d_3)} K(\bar{v} - \overline{V_t}), \ \hat{g}^*(x) = \frac{1}{T} \sum_{t=1}^{T} h_2^{-d_1} K^*(x - X_t)$$

where the kernels $K(.)$ and $K^*(.)$ are defined in Assumption **A.2.1** and the bandwidth parameters h_1 and h_2 satisfy **A.2.2**. Further, a consistent estimator of the variance σ^2 in (6) is needed and we propose the following estimator:

$$\hat{\sigma}^2 = \frac{C}{6T} \sum_{t=1}^{T} \frac{w^2(\overline{V_t})}{\hat{g}(\overline{V_t})} \left\{ 1 - \hat{F}_{h_1}(Y_t|\bar{V}_t) \right\}^2 \left(1 + 2\hat{F}_{h_1}(Y_t|\bar{V}_t) \right),$$

where $\hat{F}_{h_1}(Y_t|\overline{V_t})$ and $\hat{g}(.)$ are defined above. Finally, we reject the null hypothesis when $Th^{\frac{1}{2}(d_1+d_3)}(\hat{\Gamma} - \hat{D})/(\hat{\sigma}\sqrt{2}) > z_\alpha$, 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 against fixed or local alternatives. The following proposition states the consistency of the test for a fixed alternative.

Proposition 1 *If Assumptions* ***A.1*** *and* ***A.2*** *hold, than the test based on* $\hat{\Gamma}$ *in (4) is consistent for any distributions* $F(y \mid (x,z))$ *and* $F(y \mid x)$ *such that* $\int (F(y|x,z) - F(y|x))^2 w(x,z) dxdydz > 0$. ■

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

$$H_1(\xi_T): \quad F^{[T]}(y \mid (x,z)) = F^{[T]}(y \mid x) + \xi_T \Delta(x,y,z),$$

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 $X_{T,t}$). 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 $\{(X_{Tt}, Y_{Tt}, Z_{Tt}),\ t = 1,..,T\}$ is a strictly stationary β-mixing process with coefficients $\beta^{[T]}(l)$ such that $\sup_T \beta^{[T]}(l) = O(\nu^l)$, for some $0 < \nu < 1$ and $||f^{[T]}(x,y,z) - f(x,y,z)||_\infty = o(T^{-1} h_1^{-(d_1+d_3)/2})$. The function $\Delta(x,y,z)$ satisfies

$$\int \Delta^2(x,y,z) w(x,z) f(x,y,z) dxdydz = \gamma < \infty, \tag{7}$$

and $\xi_T \to 0$ as $T \to \infty$.

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

$$T h_1^{\frac{1}{2}(d_1+d_3)} (\hat{\Gamma} - D) \xrightarrow{d} N(\gamma, 2\sigma^2), \quad as\ T \to \infty,$$

where D, σ^2, *and* γ *are defined by (5), (6), and (7), respectively.*

■

Notice that our test has power against alternatives at distance $T^{-1/2} h_1^{-(d_1+d_3)/4}$ compared to that of Su and White (2008) which has power only against alternatives at distance $T^{-1/2} h_1^{-d/4}$. Further, our test has an asymptotic power at the same distance as the characteristic function-based test of Su and White (2007) and it is very simple to implement.

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 local bootstrap. Thus, in practice we recommend to employ the bootstrap version of our test.

Here, we are in a time series context and we cannot use the simple bootstrap for independent and identically distributed observations because the conditional dependence structure in the data would not be preserved. In our context, 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 $\{(X_t^*, Y_t^*, Z_t^*),\ t = 1, ..., T\}$ 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^*,

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

and

$$Z_t^* \sim h^{-d_2} \sum_{s=1}^{T} L_1\left((X_s - X_t^*)/h\right) L_3\left((Z_s - y)/h\right) / \sum_{s=1}^{T} L_1\left((X_s - X_t^*)/h\right);$$

(2) based on the bootstrap sample, we compute the bootstrap test statistic $\Gamma^* = T h_1^{\frac{1}{2}(d_1+d_3)}(\hat{\Gamma}^* - T^{-1}\hat{D}^*)/(\hat{\sigma}^* \sqrt{2})$;

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

(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 in our context.

Assumption A.3 (Bootstrap Bandwidth)

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

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* $\mathcal{V}_T = \{V_t \equiv (X_t, Y_t, Z_t)\}_{t=1}^T$ *and under the null hypothesis* H_0*, we have*

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

■

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}	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$	$Z_t = \sqrt{h_{2,t}}\varepsilon_{2t}$
		$h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05Y_{t-1}^2$	$h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$
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}	$Y_t = \sqrt{h_{1,t}}\varepsilon_{1t}$	$Z_t = 0.5Z_{t-1} + \varepsilon_{2t}$
		$h_{1,t} = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2$	

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 correspond to 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. In the last five DGPs , Y and Z are by construction dependent and have served to evaluate the power. We have considered two different sample sizes, $T = 200$ and $T = 300$. 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 $C = 1/4\pi$. Since optimal bandwidths are not available in the present paper, we have considered $h_1 = c_1 T^{-1/4.75}$ and $h_2 = c_2 T^{-1/4.25}$ for various values of c_1 and c_2, which corresponds to the most practical . Finally, for generating the bootstrap replications, we have also used the normal kernel with a different bandwidth, the one provided by the rule of thumb proposed in Silverman (1986). Because the data are standardized, the weighting function here is given by the indicator function defined on the set $A = \{(x, z), -2 \leq x, z \leq 2\}$.

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 $(X_t, Y_t, Z_{t-1})'$, $t = 1, \ldots, T + 200$. 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 their impact, 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 regression model:

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

The null hypothesis of Granger non-causality corresponds to $H_0 : \alpha = 0$ 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 presence of possibly dependent errors ε_t's, $\hat{\sigma}_{\hat{\alpha}}$ was computed using the commonly used heteroscedasticity autocorrelation consistent (HAC) estimator suggested by Newey and West (1987).

The empirical sizes of the linear causality test (LIN) and of the distribution-based test (BRT) 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 $T = 200$. With LIN, all rejection frequencies are within 2 standard errors from the nominal levels 5% and 10%. With BRT, 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, with BRT, the empirical sizes seem slightly closer to the corresponding nominal sizes when $c_1 = c_2 = 1$.

The empirical powers of both tests are given in Table 3. As expected, with the linear DGP5, LIN performs extremely well but the nonparametric test BRT performs almost as well. With the

Table 2: Empirical size of the bootstraped 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
BRT, c_1=1, c_2=1	0.050	0.056	0.044	0.038	0.096	0.104	0.098	0.098
BRT, c_1=0.85, c_2=0.7	0.048	0.044	0.064	0.056	0.104	0.128	0.132	0.100
BRT, 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
BRT, c_1=1, c_2=1	0.053	0.043	0.068	0.040	0.120	0.097	0.110	0.100
BRT, c_1=0.85, c_2=0.7	0.060	0.036	0.068	0.060	0.120	0.084	0.108	0.130
BRT, c_1=0.75, c_2=0.6	0.044	0.032	0.060	0.056	0.108	0.076	0.096	0.112

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

four nonlinear models considered, BRT clearly outperforms LIN. In most cases, BRT 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 300.

5 Application: Stock return predictability using VIX

We use real data to illustrate the practical importance of the proposed nonparametric test. We show that using tests based on linear models may lead to wrong conclusions about the existence of a relationship between financial variables. We particularly 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.

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.[1] Here we examine the short and long-run stock return predictability using VIX volatility index in a broader framework that allows us to leave free the specification of the underlying model. Nonparametric tests are well suited for that since

[1]Previous 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.

Table 3: Empirical power of the bootstraped 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
	BRT, c_1=1, c_2=1	0.996	0.812	0.852	1.000	0.936
	BRT, c_1=0.85, c_2=0.7	0.988	0.728	0.792	1.000	0.908
	BRT, 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
	BRT, c_1=1, c_2=1	1.000	0.976	0.966	1.000	1.000
	BRT, c_1=0.85, c_2=0.7	1.000	0.884	0.908	1.000	0.984
	BRT, c_1=0.75, c_2=0.6	1.000	0.784	0.868	1.000	0.960
$\alpha = 10\%$	$T = 200$					
	LIN	1.000	0.410	0.211	0.134	0.161
	BRT, c_1=1, c_2=1	0.992	0.916	0.916	0.984	0.980
	BRT, c_1=0.85, c_2=0.7	0.996	0.844	0.868	1.000	0.960
	BRT, 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
	BRT, c_1=1, c_2=1	1.000	1.000	0.951	1.000	1.000
	BRT, c_1=0.85, c_2=0.7	1.000	0.948	0.964	1.000	1.000
	BRT, c_1=0.75, c_2=0.6	1.000	0.912	0.924	1.000	0.984

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

they do not impose any restriction on the model linking the dependent variable to the independent variables.

Recent works use VIX index 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 index together with nonparametric tests 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.

5.1 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 monthly data for VIX index. 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 the highly liquid S& P500 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.

5.2 Causality tests

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

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

where $exr_{t+\tau}$ is the excess return τ months ahead and VIX_t represents VIX index at time t. In the empirical application, we take $\tau = 1$, 2, 3, 6, and 9 months. VIX index does not 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 index 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 predictability we consider the following null hypotheses:

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

against the alternative hypothesis

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

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
BRT $c_1 = c_2 = 1.5$	0.000	0.000	0.010	0.000	0.000
BRT $c_1 = c_2 = 1.2$	0.000	0.000	0.015	0.000	0.000
BRT $c_1 = c_2 = 1$	0.000	0.005	0.025	0.010	0.000
BRT $c_1 = 0.85$, $c_2 = 0.7$	0.000	0.010	0.035	0.036	0.000
BRT $c_1 = 0.75$, $c_2 = 0.6$	0.000	0.045	0.085	0.061	0.005

LIN and BRT 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 BRT of Table 4]. Before we start discussing our empirical results, we have to mention that the data are standardized and the weighting function $w(.)$ is the same like the one used in the simulation study [see first paragraph of 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.[2]

6 Conclusion

We propose a new statistic to test the conditional independence and Granger non-causality between two 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 has power against alternatives

[2]Other 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.

at distance $T^{-1/2}h^{-(d_1+d_3)/4}$ compared to that of Su and White (2008), which has power only for alternatives at distance $T^{-1/2}h^{-d/4}$, where $d = d_1 + d_2 + d_3$. Further, in term of power against local alternatives, our test has the same performance compared to the test of Su and White (2007) and it is very simple to implement. We ran a simulation study to investigate the finite sample properties (size and power) of the test 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 VIX volatility index. Contrary to the linear causality analysis based on the conventional t-test, we find that VIX index 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 (2009). 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 that in Ait-Sahalia, Bickel, and Stoker (2001) and Tenreiro (1997), of course with adapted calculations for our test.

We first recall Theorem 1 of Tenreiro (1997). Let $\{U_t,\ t \in \mathbb{Z}\}$ be a strictly stationary and absolutely regular process. Let $g_T(.)$ and $h_T(.,.)$ two Boreal measurable functions on $\mathbb{R}^d$ and $\mathbb{R}^d \times \mathbb{R}^d$, respectively. Assume that $\mathbb{E}[g_T(U_0)] = \mathbb{E}[h_T(U_0, u)] = 0$ and $h_T(u_1, u_2) = h_T(u_2, u_1)$ for all $(u_1, u_2) \in \mathbb{R}^d \times \mathbb{R}^d$ and define

$$\mathcal{G}_T \equiv T^{-1/2} \sum_{t=1}^{T} g_T(U_t),$$

and

$$\mathcal{H}_T \equiv T^{-1} \sum_{1 \le t_1 < t_2 \le T} [h_T(U_{t_1}, U_{t_2}) - \mathbb{E}(h_T(U_{t_1}, U_{t_2}))].$$

Observe that $\mathcal{G}_T$ and $\mathcal{H}_T$ are degenerate U-statistics of orders 1 and 2, respectively. Let p be a positive constant and $\tilde{U}_t$, for $t \geq 0$, be an $i.i.d.$ sequence, with $\tilde{U}_0$ being an independent copy of U_0.

Further, define the following terms

$$u_T(p) \equiv \max\{\max_{1\le t\le T} ||h_T(U_t, U_0)||_p, ||h_T(U_t, \tilde{U}_0)||_p\},$$

$$v_T(p) \equiv \max\{\max_{1\le t\le T} ||G_{T0}(U_t, U_0)||_p, ||G_{T0}(U_0, \tilde{U}_0)||_p\},$$

$$w_T(p) \equiv ||G_{T0}(U_0, U_0)||_p,$$

$$z_T(p) \equiv \max_{\substack{1\le t_1\le T \\ 1\le t_2\le T}} \max\{||G_{Tt_2}(U_{t_1}, U_0)||_p, ||G_{Tt_2}(U_0, U_{t_1})||_p, ||G_{Tt_2}(U_0, \tilde{U}_0)||_p,$$

where $G_{Tt}(u_1, u_2) \equiv \mathbb{E}\left[h_T(U_t, u_1)h_T(U_0, u_2)\right]$ and $||.||_p \equiv \{\mathbb{E}|.|_p\}^{1/p}$. Here is Theorem 1 of Tenreiro (1997).

Theorem *(Tenreiro, 1997) Suppose that there exist δ_0, $\gamma_1 > 0$ and $\gamma_0 < 1/2$ such that* **(i)** $||g_T(U_0)||_4 = O(1)$; **(ii)** $\mathbb{E}[g_T(U_t)g_T(U_0)] = c_t + o(1)$, *for* $t\ge 0$; **(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)** $z_T(2)T^{\gamma_1} = O(1)$; **(vii)** $\mathbb{E}[h_T(U_0, \tilde{U}_0)]^2 = 2\tilde{\sigma}_2^2 + o(1)$. *Then $(\mathcal{G}_T, \mathcal{H}_T)'$ is asymptotically normally distributed with mean zero and variance-covariance matrix* $\begin{bmatrix} \tilde{\sigma}_1^2 & 0 \\ 0 & \tilde{\sigma}_2^2 \end{bmatrix}$, *where* $\tilde{\sigma}_1^2 \equiv c_0 + 2\sum_{t=1}^{\infty} c_t$, *with* $c_t = \mathbb{E}(g_T(U_0)g_T(U_t)), t\ge 0$. ■

Now, we establish the asymptotic normality of the test statistic $\hat{\Gamma}$ defined in (4). The 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's define the following pseudo-statistic

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

where the empirical distribution function $F_T(v)$ in $\hat{\Gamma}$ is replaced by the true distribution function $F(v)$. We begin by studying the asymptotic distribution of Γ. We show, see Lemma 4, that replacing $F_T(v)$ by $F(v)$ will not affect the asymptotic normality of the test statistics $\hat{\Gamma}$.

Let's denote by

$$J(v_t, v) = \frac{K_{h_1}(\overline{v} - \overline{v_t})\, \mathrm{I}_{A_{y_t}}(y)}{\frac{1}{T}\sum_{t=1}^T K_{h_1}(\overline{v} - \overline{v_t})} - \frac{K^*_{h_2}(x - x_t)\, \mathrm{I}_{A_{y_t}}(y)}{\frac{1}{T}\sum_{t=1}^T K^*_{h_2}(x - x_t)}$$

and

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

where I_{A_t} is an indicator function defined on the set A_t. The pseudo-statistic Γ can be written as follows

$$
\begin{aligned}
\Gamma &= \frac{1}{T^2}\int\left(\sum_{t=1}^{T} J(V_t,v)\right)^2 w(x,z)dF(v) \\
&= \frac{2}{T^2}\sum_{t<s}^{T}\int J(V_t,v)J(V_s,v)w(x,z)dF(v) + \frac{1}{T^2}\left\{\sum_{t=1}^{T}\int J^2(V_t,v)w(x,z)dF(v)\right\} \\
&= \frac{2}{T^2}\sum_{t<s}^{T}\int J^*(V_t,v)J^*(V_s,v)w(x,z)dF(v) + \frac{2}{T^2}\left\{(T-1)\sum_{t=1}^{T}\int J^*(V_t,v)\mathbb{E}^2(J(V_1,v))w(x,z)dF(v)\right\} \\
&\quad + \frac{1}{T^2}\left\{T(T-1)\int \mathbb{E}^2(J(V_1,v))w(x,z)dF(v)\right\} + \frac{1}{T}\left\{\sum_{t=1}^{T}\int J^2(V_t,v)w(x,z)dF(v)\right\} \\
&= 2\,T^{-1}h_1^{-\frac{(d_1+d_3)}{2}}\left\{T^{-1}\sum_{t<s}^{T}H_T(V_t,V_s)\right\} + 2T^{-1/2}(1-T^{-1})h_1^r\left\{T^{-1/2}\sum_{t=1}^{T}G_T(V_t)\right\} + T^{-1}B_T + N_T \\
&\equiv 2\,T^{-1}h_1^{-\frac{(d_1+d_3)}{2}}T_{11} + 2T^{-1/2}(1-T^{-1})h_1^r T_{12} + T^{-1}B_T + N_T \qquad (8)
\end{aligned}
$$

where

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

$$
T_{11} = T^{-1}\textstyle\sum_{t<s}^{T} H_T(V_t,V_s), \qquad T_{12} = T^{-1/2}\textstyle\sum_{t=1}^{T} G_T(V_t), \qquad (9)
$$

with

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

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

Now, let's show the asymptotic independence and normality of T_{11} and T_{12}. 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*

$$
\begin{pmatrix} T_{11} \\ T_{12} \end{pmatrix} \xrightarrow{d} \mathcal{N}\begin{pmatrix} \tilde{\sigma}^2 & 0 \\ 0 & \sigma^2 \end{pmatrix},
$$

where $\tilde{\sigma}^2 < \infty$ *and*

$$\sigma^2 = \frac{C}{6}\int_{v_t} \frac{w^2(\bar{v}_t)}{g(\bar{v}_t)} \left\{1 - F(y_t|\bar{v}_t)\right\}^2 (1 + 2F(y_t|\bar{v}_t)) f(v_t)dv_t. \tag{10}$$

■

Proof. Observe that by construction we have $\mathbb{E}(G_T(V_t)) = 0$. We can show that conditions **(i)** and **(ii)** are fulfilled. First, we show that $\sup_v \mid G_T(v) \mid < C$, where C is a constant. We have

$$\begin{aligned}
\mathbb{E}(J(V_t, v)) &= \mathbb{E}\left(\frac{K_{h_1}(\overline{v} - \overline{V_t})\, \mathrm{I}_{A_t}(y)}{\frac{1}{T}\sum_{t=1}^T K_{h_1}(\overline{v} - \overline{V_t})} - \frac{K^*_{h_2}(x - X_t)\, \mathrm{I}_{A_t}(y)}{\frac{1}{T}\sum_{t=1}^T K^*_{h_2}(x - X_t)}\right) \\
&= \frac{1}{r!}\mu_r \left\{h_1^r F^{(r)}(y|\overline{v}) - h_2^r F^{(r)}(y|x)\right\} + o(h_1^r + h_2^r) \\
&= \frac{1}{r!}\mu_r h_1^r F^{(r)}(y|\overline{v}) + o(h_1^r),
\end{aligned}$$

under the assumption $h_2 = o(h_1)$, where $F^{(r)}$ is the r^{th} derivative of F and $\mu_r = \int s^r K(s)ds$. Hence, for $\gamma(v) = \frac{1}{r!}\mu_r F^{(r)}(y|\overline{v})$, we have $G_T(V_t) = \int \gamma(v)J^*(V_t, v)w(\overline{v})f(v)dv + o_p(1)$. Then, using assumptions A1.2, A2.1 and a change of variables, we obtain that $\sup_v \mid G_T(v) \mid < C$. Therefore, $||G_T(V_0)||_4 = O(1)$. Second, let's calculate the covariance between $G_T(V_t)$ and $G_T(V_0)$.

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

where $\xi(v) = \gamma(v)w(\overline{v})f(v)$. Under Assumption A1 and A2, we have

$$\left\|\frac{1}{T}\sum_{t=1}^T K_{h_1}(\overline{v} - \overline{V_t}) - g(\overline{v})\right\|_\infty = \sup_{v\in\mathcal{V}}\left|\frac{1}{T}\sum_{t=1}^T K_{h_1}(\overline{v} - \overline{V_t}) - g(\overline{v})\right| = o_p(1), \tag{11}$$

and

$$\left\|\frac{1}{T}\sum_{t=1}^T K^*_{h_2}(x - X_t) - g^*(x)\right\|_\infty = \sup_{x\in\mathcal{X}}\left|\frac{1}{T}\sum_{t=1}^T K^*_{h_2}(x - X_t) - g^*(x)\right| = o_p(1),$$

where g (resp. g^*) is the density function of the vector $\overline{V_t}$ (resp. X_t). Then,

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

The change of variables $\overline{v}-\overline{v_t}/h_1=\bar{a};(a_2=y)$ and $\overline{v'}-\overline{v_s}=b, b_2=y'$ leads to

$$\begin{aligned}\int J(v_t,v)J(v_s,v')\xi(v)\xi(v')f(v_t,v_s)dv_tdv_sdvdv' &= \int\left\{\frac{K(\bar{a})\,\mathrm{I}_{A_t}(a_2)}{g(\bar{v}_t)}-\frac{h_1^{d_1+d_3}}{h_2^{d_1}}\frac{K^*(h_1x_t/h_2)\,\mathrm{I}_{A_t}(a_2)}{g^*(x_t)}\right\}\\ &\quad\left\{\frac{K(\bar{b})\,\mathrm{I}_{A_s}(b_2)}{g(\bar{v}_s)}-\frac{h_1^{d_1+d_3}}{h_2^{d_1}}\frac{K^*(h_1x_s/h_2)\,\mathrm{I}_{A_t}(b_2)}{g^*(x_s)}\right\}\\ &\quad\xi(x_t,a_2,z_t)\xi(x_t',b_2,z_t')f(v_t,v_s)dv_tdv_sdadb+o_p(1).\end{aligned}$$

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

$$\begin{aligned}\int J(v_t,v)J(v_s,v')\xi(v)\xi(v')f(v_t,v_s)dv_tdv_sdvdv' &= \int\left(\frac{K(\bar{a})\,\mathrm{I}_{A_t}(a_2)}{g(\bar{v}_t)}\right)\left(\frac{K(\bar{b})\,\mathrm{I}_{A_s}(b_2)}{g(\bar{v}_s)}\right)\\ &\quad\xi(x_t,a_2,z_t)\xi(x_t',b_2,z_t')f(v_t,v_s)dv_tdv_sda\,db\\ &= \int_{v_t,v_s}\zeta(v_t)\zeta(v_s)f(v_t,v_s)dv_tdv_s+o_p(1),\end{aligned}$$

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 $\delta(v_t)=\int_{a_2}\mathrm{I}_{A_t}(a_2)\xi(x_t,a_2,z_t)da_2$. Using similar arguments, we show that

$$\int\mathbb{E}(J(V_s,v))\xi(v)dv=\int_{v_t}\zeta(v_t)f(v_t)dv_t+o_p(1).$$

Consequently,

$$\tilde{\sigma}^2=Var(\zeta(V_0))+2\sum_{i\geq1}Cov(\zeta(V_1),\zeta(V_{1+i}))<\infty.$$

Now, let us check the conditions **(iii)-(vi)**. Observe that the product $J(V_t,v)\times J(V_s,v)$ is composed of four terms and that the dominant one is

$$\frac{K_{h_1}(\overline{v}-\overline{V_t})\;I_{A_{Y_t}}(y)}{\frac{1}{T}\sum_{t=1}^T K_{h_1}(\overline{v}-\overline{V_t})}\times\frac{K_{h_1}(\overline{v}-\overline{V_s})\;I_{A_{Y_s}}(y)}{\frac{1}{T}\sum_{s=1}^T K_{h_1}(\overline{v}-\overline{V_s})}.$$

By equation (11), we have

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

where $A_{(y_0,\tilde{y}_0)}=\{v=(x,y,z),\ \max(y_0,\tilde{y}_0)\leq y\}$ and $\varphi(\overline{v})=w(\overline{v})/g^2(\overline{v})$.

Now, two changes of variables are needed. The first one is $\bar{\tilde{v}}_0=(\tilde{x}_0,\tilde{z}_0)=\bar{v}_0+h_1\,\bar{a}$, $(d\tilde{v}_0=h_1^{d_1+d_3}da)$ with $a=(a_1,a_2,a_3)(a_2=\tilde{y}_0)$ and the second one is $\bar{v}=\bar{v}_0+h_1(\bar{b}+\bar{a})$, $(dv=h_1^{d_1+d_3}db$ with $b=(b_1,b_2,b_3)(b_2=y))$. We obtain,

$$\begin{aligned}\mathbb{E}\left[H_T(V_0,\tilde{V}_0)\right]^2 &= \int_{v_0,a}\left\{\int_b K\left(\bar{b}+\bar{a}\right)\;K\left(\bar{b}\right)\varphi\left(\bar{v}_0+h_1(\bar{b}+\bar{a})\right)I_{A_{(y_0,a_2)}}(b_2)\right.\\ &\quad\left.f(x_0+h_1(a_1+b_1),b_2,z_0+h_1(a_3+b_3))db\right\}^2f(v_0)\\ &\quad f((x_0+h_1a_1,a_2,z_0+h_1a_3)dv_0\,da+o(1).\end{aligned}$$

We apply Taylor expansion to deduce that

$$\mathbb{E}\left[H_T(V_0,\tilde{V_0})\right]^2 = C\int_{v_0}\varphi^2(\bar{v}_0)f(v_0)\int_{a_2}\left\{\int_{b_2} I_{A_{(y_0,a_2)}}(b_2)f(x_0,b_2,z_0)db_2\right\}^2 f(x_0,a_2,z_0)da_2dv_0 + o(1)$$

where $C=\int_{a_1,a_3}\left(\int_{b_1,b_3} K\left(\bar{b}+\bar{a}\right)\ K\left(\bar{b}\right)db_1db_3\right)^2 da_1da_3$ and $\varphi(\bar{v}_0)=w(\bar{v}_0)/g^2(\bar{v}_0)$.
Let's calculate the integration over a_2 and b_2. In fact,

$$\int_{a_2}\left\{\int_{b_2} I_{A_{(y_0,a_2)}}(b_2)f(x_0,b_2,z_0)db_2\right\}^2 f(x_0,a_2,z_0)da_2dv_0 = L_1+L_2,$$

where

$$\begin{aligned} L_1 &= g^3(\bar{v}_0)\int_{a_2>y_0}\left\{\int_{b_2>a_2} f(b_2|\bar{v}_0)db_2\right\}^2 f(a_2|\bar{v}_0)da_2 \\ &= \frac{1}{3}g^3(\bar{v}_0)\left\{1-F(y_0|\bar{v}_0)\right\}^3 \end{aligned}$$

and

$$\begin{aligned} L_2 &= g^3(\bar{v}_0)\int_{a_2<y_0}\left\{\int_{b_2>y_0} f(b_2|\bar{v}_0)db_2\right\}^2 f(a_2|\bar{v}_0)da_2 \\ &= g^3(\bar{v}_0)\left\{1-F(y_0|\bar{v}_0)\right\}^2 F(y_0|\bar{v}_0). \end{aligned}$$

Therefore, $2\sigma^2$ is given by

$$\mathbb{E}\left[H_T(V_0,\tilde{V_0})\right]^2 = \frac{C}{3}\int_{v_0}\frac{w^2(\bar{v}_0)}{g(\bar{v}_0)}\left\{1-F(y_0|\bar{v}_0)\right\}^2\left(1+2F(y_0|\bar{v}_0)\right)f(v_0)dv_0 + o(1).$$

Now, we check the conditions **(iii)-(iv)** of Tenreiro (1997). To do that we 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 $G_T(u,v)=\mathbb{E}(H_T(V_0,u)H_T(V_0,v))$.

$$\begin{aligned} \mathbb{E}(|H_T(V_t,V_0)|^p) &\approx h_1^{\frac{p(d_1+d_3)}{2}}\int\int\left|\int \frac{K_{h_1}(\bar{v}-\bar{v}_t)1\,I_{A_{y_t}}(y)}{\frac{1}{T}\sum_{t=1}^T K_{h_1}(\bar{v}-\bar{v}_t)}\frac{K_{h_1}(\bar{v}-\bar{v}_0)1\,I_{A_{Y_0}}(y)}{\frac{1}{T}\sum_{s=1}^T K_{h_1}(\bar{v}-\bar{v}_s)}\right. \\ &\quad w(x,z)dF(v)\Big|^p f(v_t,v_0)dv_t\,dv_0 \\ &= h_1^{\frac{-p(d_1+d_3)}{2}}\int\int\left|\int \frac{K((\bar{v}-\bar{v}_t)/h_1)\mathbb{I}(y_t\le y)}{\frac{1}{T}\sum_{t=1}^T K_{h_1}(u-u_t)}\frac{K((\bar{v}-\bar{v}_0)/h_1)\mathbb{I}(y_0\le y)}{\frac{1}{T}\sum_{s=1}^T K_{h_1}(u-u_s)}\right. \\ &\quad w(x,z)dF(v)\Big|^p f(v_t,v_0)dv_t\,dv_0. \end{aligned}$$

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

Let's now calculate the following term

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

where $\alpha_X(.) = \frac{w(.)f(.)}{g_{U_0}(.)g_X(.)}$. By the change of variables, $\xi = \xi_0 + h_1\tau$, $\tilde{\xi} = \xi_0 + h_1(\tau+\tilde{\tau})$ and $\xi_0 = u + h_1(\tau_0 - \tau)$, we obtain

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

Hence

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

Then, $v_T(p) = O(h^{d/p})$. 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.

■

The following lemma provides the asymptotic bias of the pseudo-statistic Γ.

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

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

where the terms D and B_T are defined in (5) and (9

■

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

$$
\begin{aligned}
\mathbb{E}(B_T) \equiv \int \mathbb{E}(J(V_t,v))^2 w(\overline{v})f(v)dv &= \int \mathbb{E}\left(\frac{K_{h_1}(\overline{v}-\overline{V_t})\ I_{A_t}(y)}{g(\overline{v})} - \frac{K^*_{h_2}(x-X_t)\ I_{A_t}(y)}{g(x)}\right)^2 w(\overline{v})f(v)dv \\
&= \int \mathbb{E}\left(\frac{K_{h_1}(\overline{v}-\overline{V_t})\ I_{A_t}(y)}{g(\overline{v})}\right)^2 w(\overline{v})f(v)dv \\
&\quad + \int I\!\!E\left(\frac{K^*_{h_2}(x-X_t)\ I_{A_t}(y)}{g(x)}\right)^2 w(\overline{v})f(v)dv \\
&\quad -2\int I\!\!E\left(\frac{K_{h_1}(\overline{v}-\overline{V_t})\ I_{A_t}(y)}{g(\overline{v})}\right)\left(\frac{K^*_{h_2}(x-X_t)\ I_{A_t}(y)}{g(x)}\right) w(\overline{v})f(v)dv \\
&= D_1 + D_2 + D_3.
\end{aligned}
$$

First, the change of variables, $\bar{v}' = (\bar{v} - \bar{v}_t)/h_1$ and $v' = (v_1', v_2', v_3')$ with $v_2' = y$, yields

$$
\begin{aligned}
D_1 &= \int\int \frac{K_{h_1}^2(\overline{v} - \overline{v_t})\; I_{A_t}(v_2')}{g(\overline{v})^2} w(\overline{v}) f(v) f(v_t) dv\, dv_t \\
&= h_1^{-(d_1+d_3)} \int\int \frac{K^2(\bar{v}')\; I_{A_t}(v_2')}{g(\overline{v_t})^2} w(\bar{v}_t) f(x_t, v_2', z_t) f(v_t)\, dv_t\, dv' + o(1) \\
&= h_1^{-(d_1+d_3)} \int K^2(\bar{v}') d\bar{v}' \int_{v_t} \frac{w(\bar{v}_t) f(v_t)}{g(\overline{v_t})^2} \int_{v_2'} I_{A_t}(v_2') f(x_t, v_2', z_t) dv_2'\, dv_t.
\end{aligned}
$$

Since

$$
\begin{aligned}
\int_{v_2'} I_{A_t}(v_2') f(x_t, v_2', z_t) dv_2' &= g(\bar{v}_t) \int_{v_2' \geq y_t} f(v_2' | \bar{v}_t) dv_2' \\
&= g(\bar{v}_t)(1 - F(y_t | \bar{v}_t)),
\end{aligned}
$$

we get

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

where $C_1 = \int K^2(\bar{v}') d\bar{v}'$. Second, by the change of variable $(x - x_t)/h_2 = x'$ and Taylor expansion, we have

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

Under H_0, we get

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

and hence

$$
D_2 = h_2^{-d_1} C_2 \int_{x_t, y_t} \frac{w^*(x_t)(1 - F(y_t|x_t))}{g^2(x_t)} f(x_t, y_t) dx_t dy_t,
$$

where $C_2 = \int K^2(x) dx$. Finally, again using the following change of variables $x = x_t + h_2 x'$ and $z = z_t + h_1 z'$, we obtain

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

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

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

where $C_3 = K(0)$. Also, note that

$$
Var\left(T^{-1} B_T\right) \equiv \frac{1}{T^2} \sum_{t=1}^{T} \int \mathbb{E}(J_t^2) w(\bar{v}) f(v) dv = O(T^{-3} h^{-2(d_1+d_3)}).
$$

Thus,

$$Var\left(Th_1^{\frac{d_1+d_3}{2}}(T^{-1}B_T - D)\right) \equiv \frac{1}{T^2}\sum_{t=1}^{T}\int \mathbb{E}(J_t^2)w(\bar{v})f(v)dv = o(1),$$

and this concludes the proof.

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

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

where the term N_T is defined in (9).

■

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

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

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

where $\hat{\Gamma}$ is defined in (4).

■

Proof. This result follows using the same argument as 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 (9), is now given by

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

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

■

Proof of Proposition 2. First observe that

$$\begin{aligned} \Gamma &= \int \left\{\hat{F}_{h_1}(y|x,z) - \hat{F}_{h_2}(y|x)\right\}^2 w(x,z)\, dF(v) \\ &= \int \{F(y|x,z) - F(y|x)\}^2 w(x,z)\, dF(v) \\ &\quad + \int \left\{\hat{F}_{h_1}(y|x,z) - \hat{F}_{h_2}(y|x) - (F(y|x,z) - F(y|x))\right\}^2 w(x,z)\, dF(v) \\ &\quad +2\int \left\{(F(y|x,z) - F(y|x))(\hat{F}_{h_1}(y|x,z) - \hat{F}_{h_2}(y|x) - (F(y|x,z) - F(y|x))\right\} w(x,z)\, dF(v). \end{aligned}$$

Second, under the alternative hypothesis, we have

$$\int \{F(y|x,z) - F(y|x)\}^2 \, w(x,z) \, dF(v) = T^{-1} h_1^{-(d_1+d_3)/2} \int \Delta^2(x,y,z) w(x,z) \, dF(v).$$

Finally, following the same argument as in the proof of Theorem 1, we obtain

$$Th_1^{(d_1+d_3)/2} \left(\int \left\{ \hat{F}_{h_1}(y|x,z) - \hat{F}_{h_2}(y|x) - (F(y|x,z) - F(y|x)) \right\}^2 w(x,z) \, dF(v) - D \right) \xrightarrow{d} N\left(0, \sigma^2/2\right).$$

■

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 the similar argument as in the proof of Theorem 1, with the terms, T_{11}, T_{12}, B_T and N_T in (8) are replaced by their boostraped 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. Stoker (2001): "Goodness-of-fit tests for Kernel Regression with an Application to Option Implied Volatilities," *Journal of Econometrics*, 105, 363–412.

Bierens, H., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

Blum, J., J. Kiefer, and M. Rosenblatt (1961): "Distribution Free Tests of Independence Based on the Sample Distribution Function," *Annals of Mathematical Statistics*, 32, 485–498.

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. 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. Rombouts, and A. Taamouti (2011): "A Nonparametric Copula Based Test for Conditional Independence with Applications to Granger Causality," *forthcoming in Journal of Business Economic Statistics.*

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.

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. (1994): *Mixing: Properties and Examples*, vol. 85. Springer-Verlag, New York.

Dufour, J.-M., and E. Renault (1998): "Short-Run and Long-Run Causality in Time Series," *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. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

Fan, Y., and Q. Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semi-Parametric Functional Forms," *Econometrica*, 64, 865–890.

——— (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.

——— (2000): "Consistent Model Specification Tests: Kernel-Based Tests Versus Bierens ICM Tests," *Econometric Theory*, 16, 1016–1041.

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., and R. Roll (1986): "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17, 5–26.

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, 424–459.

——— (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*. Academic Press., 2ed. edn.

Hall, P. (1984): "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis*, 14, 1–16.

Hall, P., C. Wolff, and Q. Yao (1999): "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association*, 94, 154–163.

Hoeffding, W. (1948): "A Non-parametric Test of Independence," *The Annals of Mathematical Statistics*, 19, 546–557.

Hong, Y., and H. White (2005): "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence," *Econometrica*, 73, 837–901.

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, S., and Y.-J. Whang (2009): "Nonparametric Tests of Conditional Treatment Effects," Discussion paper, University College London and Seoul National University.

Lee, T.-H., and W. Yang (2006): "Money-Income Granger-Causality in Quantiles," Discussion paper, University of California, Riverside.

Lewellen, J. (2004): "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, 74, 209–235.

Li, Q., and J. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

——— (2009): "Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data," *Journal of Business and Economic Statistics, forthcoming.*

Linton, O., and P. Gozalo (1997): "Conditional Independence Restrictions: Testing and Estimation," *Discussion paper, Cowles Foundation Discussion Paper 1140.*

Lo, A., and 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 (2002): "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-Hollanded.

NEWEY, W., AND 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.

ROBINSON, P. (1991): "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, 58, 437–453.

ROSENBLATT, M. (1975): "A Quadratic Measure of Deviation of two-Dimensional Density Estimates and a Test of Independence," *The Annals of Statistics*, 3, 1–14.

SAIDI, A., AND R. ROY (2008): "Robust Optimal Tests for Causality in Multivariate Time Series," *Econometric Theory*, 24, 948–987.

SHILLER, R. (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. (1996): *Smoothing Methods in Statistics.* Springer.

SKAUG, H. J., AND D. TJØSTHEIM (1993): "A Nonparametric Test of Serial Independence Based on the Empirical Distribution Function," *Biometrika*, 80, 591–602.

SONG, K. (2009): "Testing Conditional Independence via Rosenblatt Transforms," *Annals of Statistics*, 37, 4011–4045.

Su, L., and H. White (2003): "Testing Conditional Independence via Empirical Likelihood," Discussion Paper, University of California San Diego.

——— (2007): "A Consistent Characteristic Function-Based Test for Conditional Dependence," *Journal of Econometrics*, 141, 807–837.

——— (2008): "A Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory*, 24, 829–864.

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.

Watson, G. (1964): "Smooth regression analysis," *Sankhya*, 26, 359–372.

Wiener, N. (1956): *The Theory of Prediction,*chap. 8. 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.