

Nonparametric Tests for Conditional Independence Using Conditional Distributions*

Taoufik Bouezmarni[†]
Université de Sherbrooke

Roch Roy[‡]
Université de Montréal

Abderrahim Taamouti[§]
Universidad Carlos III de Madrid

September 13, 2010

*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 the second author. We 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.

[‡]Department of Mathematics and Statistics, Université de Montréal. Address: Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, succursale Centre-ville Montréal, Canada, H3C 3J7; e-mail: royr@dms.UMontreal.ca. TEL: +1-514 3437977; FAX: +1-514 3435700.

[§]*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.

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. The power of the proposed test is better than that of Su and White (2008)'s test and it has the same power compared to the characteristic function-based test of Su and White (2007), and it is simple to implement. 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 many other financial variables. Contrary to the conventional t-test based on a linear mean-regression model, we find that dividend-price ratio predicts excess returns at long horizons but not at short horizons.

Key words: Nonparametric tests; time series; conditional independence; Granger non-causality; Nadaraya-Watson estimator; conditional distribution function.

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, generally speaking, it is possible to have situations where the causality in low moments (like mean) doesn't exist, but it does exist in high moments. Consequently, non-causality tests should be defined based on distribution functions.

Several nonparametric tests are available to test unconditional independence, 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 (2001) compare the power properties of various kernel based nonparametric tests with the integrated conditional moment tests of Bierens and Ploberger (1997), and Delgado and González Manteiga (2001) propose a test for selecting explanatory variables in nonparametric regression based on the bootstrap. However, the literature on nonparametric conditional independence tests for continuous variables is rather recent. 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 (2008) propose a nonparametric test based on the weighted Hellinger distance which 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. Linton and Gozalo (1997) develop a non-pivotal nonparametric empirical distribution function based test of conditional independence, the asymptotic null distribution of which is a functional of a Gaussian process. Critical values are computed using the bootstrap and the test has power against alternatives at distance $T^{-1/2}$. Recently, Bouezmarni, Rombouts, and Taamouti (2009) 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. Li, Maasoumi, and Racine (2009), propose a nonparametric test for the equality of conditional density functions with mixed categorical and continuous data using an L_2 distance. Finally, Song (2009) proposes Rosenblatt-transform based test of conditional independence between two random variables given a single-index involving an unknown finite dimensional parameter. He also 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 (ND) 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. 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 many empirical applications where we examine the Granger non-causality between S&P 500 Index returns and many other financial variables (dividend-price ratio, volatility index (VIX) and liquidity factors of Pastor and Stambaugh (2003)). Contrary to the conventional t-test based on a linear mean-regression model, we find that the dividend-price ratio can predict excess returns at long horizons, but not at short horizons. We also find that VIX can predict stock excess returns both at short and long-run horizons. Finally, it seems that the traded and non-traded liquidity factors can not predictor the stock excess returns.

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 applications using financial data. Section 6 concludes. The proofs of the asymptotic results are presented in the 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 to test 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

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

against the alternative hypothesis

$$H_1 : \Pr \{F(y | (X, Z)) = F(y | X)\} < 1. \tag{2}$$

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

$$\hat{F}_{h_1}(y|\bar{v}) = \frac{\sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t) \mathbf{I}_{A_{Y_t}}(y)}{\sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t)}, \quad (3)$$

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 the 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 X is defined by:

$$\hat{F}_{h_2}(y|x) = \frac{\sum_{t=1}^T K_{h_2}^*(x - X_t) \mathbf{I}_{A_{Y_t}}(y)}{\sum_{t=1}^T K_{h_2}^*(x - X_t)}, \quad (4)$$

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

$$\hat{\Gamma} = \frac{1}{T} \sum_{t=1}^T \left\{ \hat{F}_{h_1}(Y_t|\bar{V}_t) - \hat{F}_{h_2}(Y_t|X_t) \right\}^2 w(\bar{V}_t), \quad (5)$$

where $w(\cdot)$ is a nonnegative weighting function of the data \bar{V}_t , for $1 \leq t \leq T$. In simulation and application sections, and because the data are standardized, we consider a bounded support for the weight $w(\cdot)$, however in theory unbounded support can be allowed. 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 $w(\cdot)$ 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. 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 (5) 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 give its power function against local alternatives. We also establish the asymptotic validity of the bootstrapped version of the test.

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 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

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

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 estimator of the conditional distribution function.

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 $\beta(n)$, such that $\beta(n) = O(\nu^n)$, 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 $r \geq 2$. The marginal densities of X_t and $\bar{V}_t = (X_t, Z_t)$, denoted by g^* and g respectively, are twice differentiable.

Assumption A.2 (Kernel and Bandwidth)

A2.1 The kernels K and K^* are the product of a univariate symmetric and bounded kernel $k : \mathbb{R} \rightarrow \mathbb{R}$, 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 $1 \leq i \leq r - 1$.

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 $T \rightarrow \infty$, $Th_1^{(d_1+d_3)} \rightarrow \infty$ but $Th_1^{(d_1+d_3)/2+2r} \rightarrow 0$.

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 Carasco 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 estimator of the conditional distribution function. 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 $h_1 = \text{constant } T^{-1/\psi}$ 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:

$$\begin{aligned}
D_1 &= C_1 h_1^{-(d_1+d_3)} \int_{v_t} \{w(\bar{v}_t)(1 - F(y_t|\bar{v}_t))/g(\bar{v}_t)\} f(v_t) dv_t, \\
D_2 &= C_2 h_2^{-d_1} \int_{v_t} \{w^*(x_t)(1 - F(y_t|x_t))/g^*(x_t)\} f(x_t, y_t) dx_t dy_t, \\
D_3 &= -2C_3 h_1^{-d_1} \int_{v_t} \{w(\bar{v}_t)(1 - F(y_t|\bar{v}_t))/g^*(x_t)\} f(v_t) dv_t, \\
D &= (D_1 + D_2 + D_3)/T,
\end{aligned} \tag{6}$$

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

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

and

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

Further, we denote

$$\sigma^2 = \frac{C}{6} \int_{v_t} \frac{w^2(\bar{v}_t)}{g(\bar{v}_t)} \{1 - F(y_t|\bar{v}_t)\}^2 (1 + 2F(y_t|\bar{v}_t)) f(v_t) dv_t, \tag{7}$$

where

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

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 (5). 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*

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

where $\hat{\Gamma}$ is given by (5) and D and σ^2 are defined in Equations (6) and (7), 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:

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

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

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

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

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

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

where the kernels $K(\cdot)$ and $K^*(\cdot)$ 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 (7) is needed and we propose the following:

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

where $\hat{F}_{h_1}(Y_t | \bar{V}_t)$ and $\hat{g}(\cdot)$ 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, then the test based on $\hat{\Gamma}$ defined by (5) 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. First, we denote by $f^{[T]}(x, y, z)$ a sequence of densities and we suppose that $\{(X_{Tt}, Y_{Tt}, Z_{Tt}), t = 1, \dots, T\}$ is a strictly stationary β -mixing process with coefficients $\beta^{[T]}(h)$ such that $\sup_T \beta^{[T]}(h) = O(\nu^h)$, for some $0 < \nu < 1$ and $\|f^{[T]}(x, y, z) - f(x, y, z)\|_\infty = o(T^{-1}h_1^{-(d_1+d_3)/2})$. We also consider the following sequence of local alternatives

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

where $\Delta(x, y, z)$ satisfies

$$\int \Delta^2(x, y, z)w(x, z)f(x, y, z)dxdydz = \gamma < \infty \quad (8)$$

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

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

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

where D , σ^2 , and γ are defined by (6), (7), and (8), 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.

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.

3.3 Local bootstrap

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. 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^* ,

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

and

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

(2) based on the bootstrap sample, we compute the bootstrap test statistic $\Gamma^* = Th_1^{\frac{1}{2}(d_1+d_3)} (\hat{\Gamma}^* - T^{-1} \hat{D}^*) / (\hat{\sigma}^* \sqrt{2})$;

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

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

In this section, 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 $T \rightarrow \infty$, $h \rightarrow 0$ and $Th^{d+2r}/(\ln T)^\gamma \rightarrow C > 0$, 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 and under the null hypothesis H_0 , we have*

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

■

A proof is 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 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 $h_1 = c_1T^{-1/4.75}$ and $h_2 = c_2T^{-1/4.25}$ 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 $(X_t, Y_t, Z_{t-1})'$, $t =$

$1, \dots, T + 200$. 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 their impact, the first 200 observations were discarded.

Our test is valid for testing both linear and nonlinear Granger causality 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).

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
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.

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

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	
	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.

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 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 Empirical application

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 two financial or economic variables. We particularly examine the linear and nonlinear relationship between stock market excess return and many financial variables: dividend-price ratio, volatility index (VIX) and liquidity factors of Pastor and Stambaugh (2003). We test whether stock market excess returns can be predictable at short and long-run horizons using the above financial variables. We compare the results using the conventional t -test and the new nonparametric test and we show that the latter can be used to overcome incorrect inference in the standard predictive regression.

5.1 Stock return predictability using Dividend-Price Ratio, VIX and Liquidity

Many empirical studies have investigated whether stock excess returns can be predictable. The econometric method used in this context is an ordinary least squares regression of stock returns onto the past of some financial variables.¹ Fama and French (1988) argue that using the lagged dividend-price ratio as a predictor variable has a significant effect on stock returns. Campbell and Shiller (1988) find that the lagged dividend-price ratio together with the lagged dividend growth rate have a significant predictive power on stock returns. Since the publication of Fama and French (1988) and Campbell and Shiller (1988), the question of whether stock returns are predictable or not has attracted much more attention from economists. Kothari and Shanken (1997) add to the lagged dividend-price ratio the lagged book-to-market ratio as a predictor and find a reliable evidence that both ratios track time-series variation in expected real stock returns. Lewellen (2004) adds the earnings-price ratio to dividend-price ratio and the lagged book-to-market ratio and finds that those ratios have stronger forecasting power than previously thought. Consequently, the finding of Campbell and Shiller (1988) and Fama and French (1988) was confirmed by subsequent studies and considered to be a new stylized fact by Cochrane (1999) and Campbell (1999). Thus,

¹Previous 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.

new theoretical models, such as those of Campbell and Cochrane (1999), Cecchetti, Lam, and Mark (2000), and Bansal and Yaron (2004), have been proposed to explain stock return predictability.

However, the simple regression of stock returns on the dividend-price ratio, which is often called the predictive regression, raises a problem of estimation. This problem is a consequence of persistence in the dividend-price ratio and the strong correlation between residuals in the predictive regression and innovations in the process of the dividend-price ratio; see Hodrick (1992), Nelson and Kim (1993), Stambaugh (1999) among others. Consequently, using the standard asymptotic theory leads to incorrect inference for samples of typical size: the asymptotic test of the null hypothesis of the no predictability leads too often to rejection; see Mankiw and Shapiro (1986). Under the same properties of predictor variables as in Mankiw and Shapiro (1986), Stambaugh (1999) shows that in finite samples, the estimated predictive slope coefficient is biased. In a linear return-dividend yield system, he finds that there is an upward bias in the predictive coefficient on the dividend yield, deriving from the negative correlation between return and dividend yield innovations and the persistence of dividend yields. Thus, new approaches to correct such econometric problems have been proposed to test for the stock return predictability; see Wolf (2000), Ang (2002), Ang and Bekaert (2007), Lanne (2002), Valkanov (2003), Lewellen (2004), Torous, Valkanov, and Yan (2005), and Campbell and Yogo (2006).

In this section, we use a nonparametric approach to overcome the incorrect inference in the standard predictive regression. The nonparametric approach doesn't impose any restriction on the model linking the dependent variable to the independent variables, thus the persistence in the predictor variables and the correlation between residuals in the predictive regression and innovations in the process of the predictor variables will not affect the nonparametric test statistic. Consequently, the inference will remain valid as long as the stationarity assumption is satisfied. We also consider VIX and liquidity factors of Pastor and Stambaugh (2003) for testing stock return predictability. Recent works use those variables to predict excess stock 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. Pastor and Stambaugh (2003) find that expected stock returns are related cross-sectionally to the sensitivities of returns to fluctuations in aggregate liquidity. They find that over a 34-year period, the average return on stocks with high sensitivities to liquidity exceeds that for stocks with low sensitivities by 7.5% annually, adjusted for exposures to the market return as well as size, value, and momentum factors. In what follows, we use VIX and liquidity factors together with nonparametric tests to check whether the excess returns on S&P 500 Index are predictable. We compare our results to those obtained with 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 the following monthly financial variables: dividend-price ratio, VIX and liquidity factors of Pastor and Stambaugh (2003). The monthly dividend-price ratio is computed from the Center for Research in Security Prices (CRSP) indices for the S&P 500 universe which contains monthly index files with value-weighted returns, with and without dividends. We also consider monthly data for VIX index. The VIX volatility index is an indication of the expected volatility of the S&P 500 stock index for the next thirty days. The VIX is provided by the Chicago Board Options Exchange (CBOE) in the US, and is calculated using the near term S&P 500 options markets. It is based on the highly liquid S&P 500 index options along with the “model-free” approach. Finally, we consider the traded and non-traded liquidity factors of Pastor and Stambaugh (2003). These factors can be downloaded from Stambaugh’s website. More details about how those liquidity factors are computed can be found in Section II-A of Pastor and Stambaugh (2003). To illustrate the data, Figure 1 exhibits the monthly time series of dividend-price ratio, VIX and traded and non-traded liquidity factors of Pastor and Stambaugh (2003). All the series cover the period from January 1996 to September 2008 for a total of 153 observations.

We performed an Augmented Dickey-Fuller test (hereafter ADF-test) for nonstationarity of the above financial variables (dividend-price ratio, VIX and liquidity factors) and the stationarity hypothesis was not rejected. We also estimated an autoregressive AR(1) model for dividend-price ratio and the results show that this ratio is persistent: the coefficient in the AR(1) is equal to 0.885.

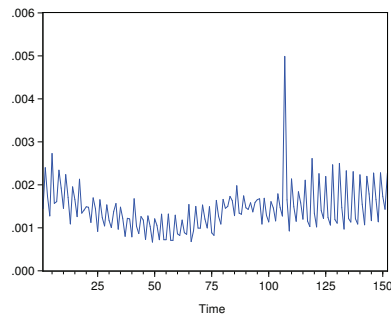
5.3 Causality tests

To test linear causality between S&P 500 excess return and the above financial variables, we consider the following linear regression model

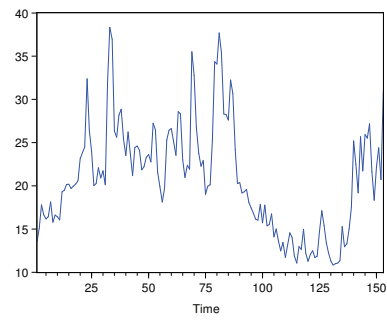
$$exr_{t+\tau} = \mu_\tau + \beta_\tau exr_t + \alpha_\tau Z_t + \varepsilon_{t+\tau},$$

where $exr_{t+\tau}$ is the excess return τ months ahead and Z_t represents one of the financial variables: dividend-price ratio, VIX, non-traded liquidity factor, and traded liquidity factor. In the empirical application, we take $\tau = 1, 2, 3, 6,$ and 9 months. In this model, Z_t 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.

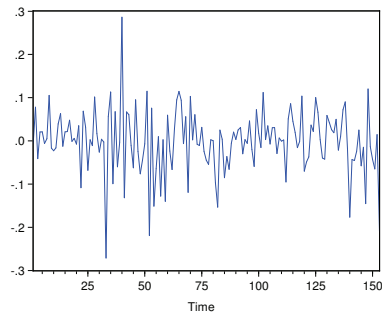
The results of linear causality (predictability) tests between stock excess returns and dividend-price ratio, VIX and non-traded and traded liquidity factors of Pastor and Stambaugh (2003) are



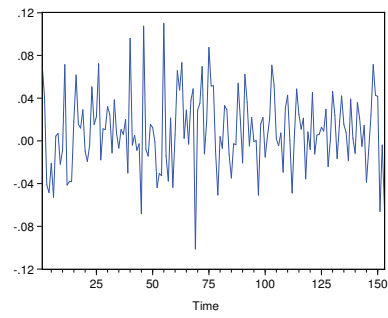
(a) Dividend-Price Ratio



(b) Volatility Index (VIX)



(c) Non-Traded liquidity factor



(d) Traded liquidity factor

Figure 1: Financial variables: dividend-price ratio, volatility index (VIX) and traded and non-traded liquidity factors of Pastor and Stambaugh (JPE 2003). The sample covers the period from January 1996 to September 2008 for a total of 153 observations.

presented in Tables 4, 5, 6, and 7, respectively. At the 5% significance level, we find evidence that excess return can not be predicted at short and long-run horizons using VIX and traded liquidity factors. However, dividend-price ratio predicts excess returns one month and 3 months ahead and non-traded liquidity factor predicts excess returns only one month ahead.

Now, to test for the presence of nonlinear predictability using the above variables, we consider the following null hypotheses:

Table 4: P-values for linear and nonlinear causality tests between Return at different horizons and Dividend-Price Ratio.

Test statistic / Horizon Return	1 Month	2 Months	3 Months	6 Months	9 Months
LIN	0.024	0.743	0.046	0.252	0.330
BRT $c_1 = c_2 = 1.5$	0.075	0.540	0.300	0.015	0.025
BRT $c_1 = c_2 = 1.2$	0.120	0.400	0.270	0.005	0.032
BRT $c_1 = c_2 = 1$	0.160	0.260	0.250	0.030	0.039
BRT $c_1 = 0.85, c_2 = 0.7$	0.175	0.200	0.280	0.047	0.040
BRT $c_1 = 0.75, c_2 = 0.6$	0.195	0.150	0.445	0.080	0.061

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.

Table 5: 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.

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

against the alternative hypothesis

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

Table 6: P-values for linear and nonlinear causality tests between Return at different horizons and innovations in Aggregate Liquidity.

Test statistic / Horizon Return	1 Month	2 Months	3 Months	6 Months	9 Months
LIN	0.039	0.578	0.743	0.536	0.111
BRT $c_1 = c_2 = 1.5$	0.215	0.590	0.115	0.160	0.105
BRT $c_1 = c_2 = 1.2$	0.240	0.500	0.110	0.215	0.220
BRT $c_1 = c_2 = 1$	0.230	0.365	0.090	0.245	0.330
BRT $c_1 = 0.85, c_2 = 0.7$	0.225	0.280	0.130	0.260	0.500
BRT $c_1 = 0.75, c_2 = 0.6$	0.260	0.205	0.230	0.315	0.535

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.

Table 7: P-values for linear and nonlinear causality tests between Return at different horizons and Traded liquidity factor.

Test statistic / Horizon Return	1 Month	2 Months	3 Months	6 Months	9 Months
LIN	0.715	0.653	0.595	0.616	0.839
BRT $c_1 = c_2 = 1.5$	0.115	0.425	0.125	0.130	0.065
BRT $c_1 = c_2 = 1.2$	0.100	0.480	0.115	0.115	0.070
BRT $c_1 = c_2 = 1$	0.085	0.555	0.100	0.100	0.115
BRT $c_1 = 0.85, c_2 = 0.7$	0.071	0.700	0.100	0.090	0.255
BRT $c_1 = 0.75, c_2 = 0.6$	0.065	0.690	0.110	0.090	0.335

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 dividend-price ratio, VIX and non-traded and traded liquidity factors of Pastor and Stambaugh (2003) are presented in the tables 4, 5, 6, and 7, respectively. Before we start discussing our empirical 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 first paragraph of Section 4]. Also, 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 $c_1 = c_2 = 1$. Contrary to the conventional t -test, at the 5% significance level, our nonparametric test show that dividend-price ratio is a good predictor of excess returns at long horizons (6 and 9 months ahead), but not at short horizons (1, 2, 3 months ahead). These results are in line with those found in Campbell and Yogo (2006) and many other works. We also find that VIX can predict stock excess returns both at short and long-run horizons, which is not the case when we use the aforementioned linear model and the standard t -test.² The latter shows that VIX cannot predict stock excess returns at all horizons considered. Finally, the traded and non-traded liquidity factors of Pastor and Stambaugh (2003) can not cause stock excess returns. This means that using those liquidity factors will not help to predict the time series of stock market excess returns. Using cross-section data, Pastor and Stambaugh (2003) argue that the liquidity factors help to explain the cross-section of individual stock returns: they find that over a 34-year period, the average return on stocks with high sensitivities to liquidity exceeds that for stocks with low sensitivities by 7.5% annually. To conclude, liquidity factors explains the variation in the cross-section of individual stock returns, but not the variation in the market stock returns (S&P 500).

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 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

²Other 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.

(size and power) of the test.

We illustrate the practical relevance of our nonparametric test by considering many empirical applications where we examine Granger non-causality between S& P500 Index returns and many other financial variables (dividend-price ratio, volatility index (VIX) and liquidity factors of Pastor and Stambaugh (2003)). Contrary to the linear causality analysis based on the conventional t -test, we find that the dividend-price ratio can predict excess returns at long horizons, but not at short horizons. We also find that VIX can predict stock excess returns both at short and long-run horizons. Further, it seems that the traded and non-traded liquidity factors of Pastor and Stambaugh (2003) can not predict the stock excess returns.

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 Stoke (2001) and Tenreiro (1997), of course with adapted calculations for our test.

We first recall Theorem 1 of Tenreiro (1997). Let $g_T(\cdot)$ and $h_T(\cdot, \cdot)$ 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

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

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

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

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

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

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

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

Theorem (Tenreiro, 1997) *Suppose that there exist $\delta_0, \gamma_1 > 0$ 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 $t \geq 0$; (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 $c_t = \mathbb{E}(g_T(U_0)g_T(U_t))$, $t \geq 0$. ■*

Now, we establish the asymptotic normality of the test statistic $\hat{\Gamma}$ defined in (5). 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

$$J_t(v) = \frac{K_{h_1}(\bar{v} - \bar{V}_t) \mathbb{I}_{A_{Y_t}}(y)}{\frac{1}{T} \sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t)} - \frac{K_{h_2}^*(x - X_t) \mathbb{I}_{A_{Y_t}}(y)}{\frac{1}{T} \sum_{t=1}^T K_{h_2}^*(x - X_t)}$$

and

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

where \mathbb{I}_{A_t} is an indicator function defined on the set A_t . The pseudo-statistic Γ can be written as

follows

$$\begin{aligned}
\Gamma &= \frac{1}{T^2} \int \left(\sum_{t=1}^T J_t(v) \right)^2 w(x, z) dF(v) \\
&\quad + \frac{1}{T^2} \left\{ \sum_{t=1}^T \int J_t^2(v) w(x, z) dF(v) \right\} + \frac{1}{T^2} \left\{ T(T-1) \int \mathbb{E}^2(J_t(v)) w(x, z) dF(v) \right\} \\
&= 2T^{-1} h_1^{-\frac{(d_1+d_3)}{2}} \left\{ T^{-1} \sum_{t < s} H_T(V_t, V_s) \right\} + 2T^{-1/2} (1 - T^{-1}) h_1^r \left\{ T^{-1/2} \sum_{t=1}^T G_T(V_t) \right\} + T^{-1} B_T + N_T \\
&\equiv 2T^{-1} h_1^{-\frac{(d_1+d_3)}{2}} T_{11} + 2T^{-1/2} (1 - T^{-1}) h_1^r T_{12} + T^{-1} B_T + N_T
\end{aligned} \tag{9}$$

where

$$\begin{aligned}
B_T &= \frac{1}{T} \left\{ \sum_{t=1}^T \int J_t^2(v) w(x, z) dF(v) \right\}, \quad N_T = \frac{1}{T^2} \left\{ T(T-1) \int \mathbb{E}^2(J_t(v)) w(x, z) dF(v) \right\} \\
T_{11} &= T^{-1} \sum_{t < s} H_T(V_t, V_s), \quad T_{12} = T^{-1/2} \sum_{t=1}^T G_T(V_t),
\end{aligned} \tag{10}$$

with

$$H_T(V_t, V_s) = h_1^{-\frac{(d_1+d_3)}{2}} \int J_t^*(v) J_s^*(v) w(x, z) dF(v) \text{ and } G_T(V_t) = h_1^{-r} \int J_t^*(v) \mathbb{E}(J_t(v)) w(x, z) dF(v).$$

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 $\sigma^2 = \frac{1}{2} \mathbb{E} \left[H_T(\tilde{V}_0, \tilde{V}_0) \right]^2$, 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

$$\sigma^2 = \frac{C}{6} \int_{v_t} \frac{w^2(\bar{v}_t)}{g(\bar{v}_t)} \{1 - F(y_t|\bar{v}_t)\}^2 (1 + 2F(y_t|\bar{v}_t)) f(v_t) dv_t. \tag{11}$$

■

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, since $\sup_v |G_T(v)| < \infty$, we have $\|G_T(V_0)\|_4 = O(1)$. Second, let's calculate the covariance between $G_T(V_t)$ and $G_T(V_0)$. We have

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

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 $\gamma(v) = \frac{1}{r!} \mu_r F^{(r)}(y|\bar{v})$, we have $G_T(V_t) = \int \gamma(v) J_t^*(v) w(\bar{v}) f(v) dv + o_p(1)$. Therefore,

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

where $\xi(v) = \gamma(v) w(\bar{v}) f(v)$. Under some regularity conditions on the density function g and the bandwidth parameter h_1 (resp. g^* and the bandwidth parameter h_2), we have

$$\frac{1}{T} \sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t) = g(\bar{v}) + O_p(T^{-1/2} h_1^{-\frac{d_1+d_3}{2}}), \quad (12)$$

resp.

$$\frac{1}{T} \sum_{t=1}^T K_{h_2}^*(x - X_t) = g^*(x) + O_p(T^{-\frac{1}{2}} h_2^{-\frac{d_1}{2}}),$$

where g (resp. g^*) is the density function of the vector \bar{V}_t (resp. X_t). Then,

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

The change of variable $\bar{v} - \bar{v}_t/h_1 = \bar{a}; (a_2 = y)$ and $\bar{v}' - \bar{v}_s = b, b_2 = y'$ leads to

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

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

$$\begin{aligned} \int J_t(v)J_s(v')\xi(v)\xi(v')dv_tdv_sdv'dv' &= \int \left(\frac{K(\bar{a}) I_{A_t}(a_2)}{g(\bar{v}_t)} \right) \left(\frac{K(\bar{b}) I_{A_s}(b_2)}{g(\bar{v}_s)} \right) \\ &\quad \xi(x_t, a_2, z_t)\xi(x'_t, b_2, z'_t)f(v_t, v_s)dv_tdv_sda db \\ &= \int_{v_t, v_s} \zeta(v_t)\zeta(v_s)f(v_t, v_s)dv_tdv_s + o_p(1), \end{aligned}$$

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

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

Consequently,

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

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

$$\frac{K_{h_1}(\bar{v} - \bar{V}_t) I_{A_{Y_t}}(y)}{\frac{1}{T} \sum_{t=1}^T K_{h_1}(\bar{v} - \bar{V}_t)} \times \frac{K_{h_1}(\bar{v} - \bar{V}_s) I_{A_{Y_s}}(y)}{\frac{1}{T} \sum_{s=1}^T K_{h_1}(\bar{v} - \bar{V}_s)}.$$

By equation (12), we have

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

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

Now, 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 $a = (a_1, a_2, a_3)$ ($a_2 = \tilde{y}_0$) 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,

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

We apply Taylor expansion to deduce that

$$\mathbb{E} \left[H_T(V_0, \tilde{V}_0) \right]^2 = C \int_{v_0} \varphi^2(\bar{v}_0) f(v_0) \int_{a_2} \left\{ \int_{b_2} I_{A_{(y_0, a_2)}}(b_2) f(x_0, b_2, z_0) db_2 \right\}^2 f(x_0, a_2, z_0) da_2 dv_0 + o(1)$$

where $C = \int_{a_1, a_3} \left(\int_{b_1, b_3} K(\bar{b} + \bar{a}) K(\bar{b}) db_1 db_3 \right)^2 da_1 da_3$ and $\varphi(\bar{v}_0) = w(\bar{v}_0)/g^2(\bar{v}_0)$.

Let's calculate the integration over a_2 and b_2 . In fact,

$$\int_{a_2} \left\{ \int_{b_2} I_{A_{(y_0, a_2)}}(b_2) f(x_0, b_2, z_0) db_2 \right\}^2 f(x_0, a_2, z_0) da_2 dv_0 = L_1 + L_2,$$

where

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

and

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

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

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

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 $G_T(u, v) = \mathbb{E}(H_T(V_0, u)H_T(V_0, v))$.

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

By change of variables, as for $\mathbb{E} \left[H_T(V_0, \tilde{V}_0) \right]^2$, 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

$$\begin{aligned} G_T(u, v) &= \mathbb{E}(H_T(V_0, u)H_T(V_0, v)) \\ &\approx h_1^{(d_1+d_3)} \mathbb{E} \left(\int \int \{K_{h_1}(\bar{\xi} - \bar{V}_0) I_{A_{y_0}}(\xi_2)\} \{K_{h_1}(\bar{\xi} - \bar{u}) I_{A_{v_2}}(\xi_2)\} \right. \\ &\quad \left. \{K_{h_1}(\bar{\xi} - \bar{V}_0) I_{A_{y_0}}(\tilde{\xi}_2)\} \{K_{h_1}(\bar{\xi} - \bar{v}) I_{A_{v_2}}(\tilde{\xi}_2)\} \alpha_{\bar{u}}(\xi) \alpha_{\bar{v}}(\tilde{\xi}) d\xi d\tilde{\xi} \right) \\ &\leq C h_1^{-3(d_1+d_3)} \int \int \int K((\bar{\xi} - \bar{\xi}_0)/h_1) K((\bar{\xi} - \bar{u})/h_1) K((\bar{\xi} - \bar{\xi}_0)/h_1) \\ &\quad K((\bar{\xi}^+ - v^+)/h_1) d\xi d\tilde{\xi} d\xi_0, \end{aligned}$$

where $\alpha_X(\cdot) = \frac{w(\cdot)f(\cdot)}{g_{V_0}(\cdot)g_X(\cdot)}$. 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*

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

where the terms D and B_T are defined in (6) and (10), respectively. ■

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

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

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

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

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

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

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

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

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 $x = x_t + h_2x'$ and $z = z_t + h_1z'$, we obtain

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

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

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

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

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

Thus,

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

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 (10). ■

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

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 (5). ■

Proof. This result follows using the same argument as in Ait-Sahalia, Bickel, and Stoke (2001). ■

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 (10), is now given by

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

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 observe that

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

Second, under the alternative hypothesis, we have

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

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

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

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 $G_T(V_t^*)$ 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 (9) 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 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.
- ANG, A. (2002): “Characterizing the Ability of Dividend Yields to Predict Future Dividends in Log-Linear Present Value Models,” Discussion paper, Columbia University.
- ANG, A., AND G. BEKAERT (2007): “Stock Return Predictability: Is It There?,” *Review of Financial Studies*, 20(3), 651–707.
- BANSAL, R., AND A. YARON (2004): “Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles,” *Journal of Finance*, 59, 1481–1509.
- 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 (2009): “A Nonparametric Copula Based Test for Conditional Independence with Applications to Granger Causality,” Discussion paper, Université de Montréal and Universidad Carlos III de Madrid.
- CAI, Z. (2002): “Regression Quantiles for Time Series,” *Econometric Theory*, 18, 169–192.
- CAMPBELL, J. Y. (1999): “Asset Prices, Consumption, and the Business Cycle,” *Handbook of Macroeconomics*, 1, 1231–1303.
- CAMPBELL, J. Y., AND J. COCHRANE (1999): “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy*, 107, 205–251.

- 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.
- CAMPBELL, J. Y., AND M. YOGO (2006): “Efficient Tests of Stock Return Predictability,” *Journal of Financial Economics*, 81, 27–60.
- CARASCO, M., AND X. CHEN (2002): “Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models,” *Econometric Theory*, 18, 17–39.
- CECCHETTI, S. G., P. LAM, AND N. MARK (2000): “Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True?,” *American Economic Review*, 90, 787–805.
- CENESIZOGLU, T., AND A. TIMMERMANN (2008): “Is the Distribution of Stock Returns Predictable?,” Discussion paper, HEC Montreal and UCSD.
- COCHRANE, J. (1999): “New Facts in Finance,” *Economic Perspectives*, XXIII(3), 36–58.
- 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.
- (2001): “Consistent Model Specification Tests: Nonparametric 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.
- HODRICK, R. (1992): “Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement,” *Review of Financial Studies*, 5, 357–386.
- 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.

- LANNE, M. (2002): “Testing the Predictability of Stock Returns,” *Review of Economics and Statistics*, 84, 407–415.
- 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., E. MAASOUMI, AND J. 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. 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.
- MANKIW, N., AND M. SHAPIRO (1986): “Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models,” *Economic Letters*, 20, 139–145.
- 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.
- NELSON, C., AND M. KIM (1993): “Predictable Stock Returns: The Role of Small Sample Bias,” *Journal of Finance*, 48, 641–661.
- NEWBOLD, P. (1982): *Causality Testing in Economics. in Time Series Analysis. Theory and Practice 1*. ed. by O. D. Anderson. Amsterdam, North-Holland.

- NEWKEY, 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.
- PASTOR, L., AND R. STAMBAUGH (2003): “Liquidity Risk and Expected Stock Returns,” *Journal of Political Economy*, 111, 642–685.
- 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, 40114045.
- STAMBAUGH, R. (1999): “Predictive Regressions,” *Journal of Financial Economics*, 54, 375–421.
- 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.
- TOROUS, W., R. VALKANOV, AND S. YAN (2005): “On Predicting Stock Returns with Nearly Integrated Explanatory Variables,” *Journal of Business*, 77, 937–966.
- TROUNG, Y., AND C. STONE (1992): “Nonparametric Function Estimation Involving Time Series,” *Annals of Statistics*, 20, 77–97.
- VALKANOV, R. (2003): “Long-Horizon Regressions: Theoretical Results and Applications,” *Journal of Financial Economics*, 68, 201–232.
- WATSON, G. (1964): “Smooth regression analysis,” *Sankhya*, 26, 359–372.
- WIENER, N. (1956): *The Theory of Prediction*, chap. 8. McGraw-Hill, New York.
- WOLF, M. (2000): “Stock Returns and Dividend Yields Revisited: A New Way to Look at an Old Problem,” *Journal of Business and Economic Statistics*, 18, 18–30.
- YOSHIHARA, K. (1976): “Limiting behavior of U-statistics for stationary, absolutely regular processes,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 35, 237–252.