Measuring Nonlinear Granger Causality in Mean*

Xiaojun Song †
Peking University

Abderrahim Taamouti‡
Durham University Business School

February 3, 2016

*The authors thank two anonymous referees, an Associate Editor, and the Editor Rong Chen for several useful comments. We would also like to thank the seminar participants at ESSEC Paris, University of Liverpool, Maastricht University, and XXXVIII Simposio of the Spanish Economic Association in Santander, for their very useful comments. Financial support from the Spanish Ministry of Education through grants #ECO2010-19357 are also acknowledged.

†Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. E-mail: sxj@gsm.pku.edu.cn.

‡*Corresponding author*: Department of Economics and Finance, Durham University Business School. Address: Mill Hill Lane, Durham, DH1 3LB, UK. TEL: +44-1913345423. E-mail: abderrahim.taamouti@durham.ac.uk.

Measuring Nonlinear Granger Causality in Mean

ABSTRACT

We propose model-free measures for *Granger causality in mean* between random variables. Unlike the existing measures, ours are able to detect and *quantify* nonlinear causal effects. The new measures are based on nonparametric regressions and defined as logarithmic functions of restricted and unrestricted mean square forecast errors. They are easily and consistently estimated by replacing the unknown mean square forecast errors by their nonparametric kernel estimates. We derive the asymptotic normality of nonparametric estimator of causality measures, which we use to build tests for their statistical significance. We establish the validity of smoothed local bootstrap that one can use in finite sample settings to perform statistical tests. Monte Carlo simulations reveal that the proposed test has good finite sample size and power properties for a variety of data-generating processes and different sample sizes.

Finally, the empirical importance of measuring nonlinear causality in mean is also illustrated. We quantify the degree of nonlinear predictability of equity risk premium using variance risk premium. Our empirical results show that the variance risk premium is a very good predictor of risk premium at horizons less than six months. We also find that there is a high degree of predictability at horizon one-month which can be attributed to a nonlinear causal effect.

Keywords: Granger causality measures; nonlinear causality in mean; nonparametric estimation; time series; bootstrap; volatility index; realized volatility; variance risk premium; risk premium.

Journal of Economic Literature classification: C12; C14; C15; C19; G1; G12; E3; E4.

1 Introduction

The concept of causality introduced by Wiener (1956) and Granger (1969) constitutes a basic notion for analyzing dynamic relationships between time series. In studying Wiener-Granger causality, predictability is the central issue which is of great importance to economists, policymakers and investors. Much research has been devoted to building tests of non-causality *in mean*. However, once we have concluded that a "causal relation" is present, it is usually important to assess the strength of this relationship. Only a few papers have been proposed to *measure* the causality in mean between the variables of interest; see Geweke (1982, 1984) and Dufour and Taamouti (2010). Those papers consider *parametric linear* models for the conditional mean function. Thus, the proposed measures ignore nonlinear causal effects, which might lead to invalid causal analysis. Hence, we simply cannot use the existing measures to quantify the strength of nonlinear causality in mean. The present paper aims to propose model-free measures to quantify nonlinear causality in mean.

Wiener-Granger analysis distinguishes between three basic types of causality: from Y to X, from X to Y, and an instantaneous causality, where X and Y are the variables of interest. In practice, it is possible that all three causality relations coexist simultaneously. Hence the importance of providing tools that quantify and compare the degree of these causalities. Unfortunately, causality tests fail to accomplish this task, because they only provide evidence on the presence of causality. A large effect may not be statistically significant (at a given level), and a statistically significant effect may not be "large" from an economic viewpoint (or more generally from the viewpoint of the subject at hand) or relevant for decision making. As emphasized by McCloskey and Ziliak (1996), it is crucial to distinguish between the numerical value of the measure and its statistical significance. Hence, beyond accepting or rejecting non-causality hypotheses – which state that certain variables do not help forecasting other variables – we wish to assess the magnitude of the forecast improvement, where the latter is defined in terms of some loss function (mean square forecast errors). Even if the hypothesis of no improvement (non-causality) cannot be rejected from looking at the available data (for example, because the sample size or the structure of the process does allow for high test power), sizeable improvements may remain consistent with the same data. Or, by contrast, a statistically significant improvement – which may easily be produced by a large data set - may not be relevant from a practical viewpoint.

The topic of measuring Granger causality has attracted much less attention. Furthermore, most of the existing measures focus on linear causality in mean, thus, those measures can not be applied in the presence of *nonlinear* causality in mean. Geweke (1982, 1984) introduce measures of causality in mean based on *linear parametric* autoregressive models. Dufour and Taamouti (2010) extend Geweke's (1982, 1984) work to propose measures for short and long run causality in mean using *parametric* ARMA models. Gouriéroux et al. (1987) build measures of causality based on Kullback information criterion and use a *parametric* approach for the estimation of their measures. Polasek (1994, 2002) show how causality measures can be computed using Akaike Information Criterion (AIC) and a Bayesian approach. Taamouti et al. (2014) have recently proposed a nonparametric estimator and test for Granger causality measures that quantify Granger causality in *distribution*. However, the main issue of Taamouti et al.'s (2014) measures is that they are not informative about the level(s) (mean, variance, other high-order moments, quantiles) of distribution where the causality exists.

We introduce new model-free measures to quantify nonlinear Granger causality in mean. The new measures are defined in the context of nonparametric regressions as logarithmic functions of restricted and unrestricted mean square forecast errors. A consistent nonparametric estimator of these measures is defined in terms of Nadaraya-Watson kernel estimators of mean square forecast errors. We establish the asymptotic normality of this estimator that we use to build tests for statistical significance of measures. We also show the validity of smoothed local bootstrap that we apply to perform statistical tests in finite sample settings. Furthermore, a Monte Carlo simulation study is performed to investigate the finite sample properties (size and power) of the proposed nonparametric test and the results reveal that the latter behaves well for a variety of typical data generating processes.

Moreover, since testing that the value of measure is equal to zero is equivalent to testing for non-causality in mean, we consider an additional simulation exercise to compare the empirical size and power of our test with those of nonparametric test of Granger non-causality in mean introduced by Nishiyama et al. (2011). Simulation results indicate that our test has comparable size, but better power than Nishiyama et al.'s (2011) test.

Finally, we apply our nonparametric causality measures to quantify the degree of nonlinear predictability of equity risk premium using what is known as variance risk premium; see Bollerslev et al. (2009). The latter is defined as the difference between risk-neutral and objective expectations of realized variance. Our results indicate that the variance risk premium is a good predictor of risk premium at horizons less than six months. Contrary to Bollerslev et al. (2009), we find that there is a high degree of predictability at horizon one-month which can be attributed to a nonlinear causal effect.

The plan of the paper is as follows. Section 2 provides the motivation for considering measures for nonlinear Granger causality in mean. Section 3 presents the general theoretical framework which underlies the definition of causality in mean. In Section 4 and 5, we define the theoretical nonparametric measures of Granger causality in mean. In Section 6 we introduce a consistent nonparametric estimator of causality measures based on kernel estimation of mean-square forecast errors of restricted and unrestricted nonparametric regressions. We also establish the asymptotic distribution of nonparametric estimator of causality in mean measures and discuss the asymptotic validity of a smoothed local bootstrap assisted test. In Section

7 we extend our results to the case where the random variables of interest are multivariate. In Section 8 we provide a simulation exercise to investigate the finite sample properties of our test of causality measures. Section 9 is devoted to an empirical application and the conclusion relating to the results is given in Section 10. Additional data generating processes for simulations and proofs of main results appear in the Appendices A and B, respectively.

2 Motivation

The causality measures that we consider here constitute a generalization of those developed by Geweke (1982, 1984) and others. The existing measures quantify the effect of one variable Y on another variable X assuming that the regression function linking the two variables of interest is *known* and linear. The significance of such measures is limited in the presence of unknown regression functions and in the presence of nonlinear causality in mean.

We propose measures of causality between random variables based on nonparametric regression functions. Such measures detect and quantify nonlinear causality in mean. To see the importance of these causality measures, consider the following example.

Example 1 *[Brock (1991)] Consider the following nonlinear regression model*

$$X_{t+1} = \beta \; Y_t \cdot X_t + \varepsilon_{t+1}, \tag{1}$$

where $\{Y_t\}$ and $\{X_t\}$ are mutually independent and individually i.i.d. $N(0,1)$, and β denotes a parameter. Equation (1) shows that Y nonlinearly causes the conditional mean of X, since X_{t+1} depends nonlinearly on the past value of Y_t. However, since all autocorrelations and cross correlations between X and Y are zero, Y does not linearly cause the conditional mean of X. Thus, this example illustrates the case where the causality in mean does not exist linearly, but it does nonlinearly. But, how can we measure the degree of this nonlinear causality in mean? Existing measures do not answer this question.

Generally speaking, the existing measures of linear Granger causality in mean might have low power in detecting certain kinds of nonlinear causal effects. Formally, let us assume that the true conditional mean of random variable X is a nonlinear function of random variable Y

$$X_{t+1} = \mu + \beta \; f\left(Y_t\right) + \alpha X_t + u_{t+1}, \tag{2}$$

where u_{t+1} is an i.i.d. process with mean zero and variance σ^2. Let us now approximate the nonlinear relationship in (2) by a linear one, using a Taylor expansion around zero:

$$X_{t+1} \simeq \mu + \beta \; \left[f\left(0\right) + Y_t f'\left(0\right)\right] + \alpha X_t + u_{t+1} = \eta + \lambda Y_t + \alpha X_t + u_{t+1}, \tag{3}$$

where the new constant η is equal to $\mu + \beta f(0)$ and the new slope λ is equal to $\beta f'(0)$. Equation (3) shows that the impact of Y on conditional mean of X depends on the first derivative of f evaluated at zero. Thus, if $f'(0)$ is zero or close to zero, then the above linear approximation has no power or low power in detecting nonlinear relationship between X and Y. Broadly speaking, a Taylor expansion of $f(Y_t)$ around any value a such that $f'(a)$ is zero or close to zero will lead to low power in detecting nonlinear relationship between X and Y. Hence the importance of using nonparametric regression to build measures of nonlinear Granger causality in mean.

3 Framework

The notion of non-causality studied here is defined in terms of orthogonality conditions between subspaces of a Hilbert space of random variables with finite second moments. We denote $L^2 \equiv L^2(\Omega, \mathcal{A}, Q)$ a Hilbert space of real random variables with finite second moments, defined on a common probability space $(\Omega, \mathcal{A}, Q)$. If E and F are two Hilbert subspaces of L^2, we denote $E + F$ the smallest subspace of L^2 which contains both E and F.

"Information" is represented here by nondecreasing sequences of Hilbert subspaces of L^2. In particular, we consider a sequence I of "reference information sets" $I(t)$,

$$I = \{I(t) : t \in \mathbb{Z}\,,\, t > \omega\} \text{ with } t < t' \Rightarrow I(t) \subseteq I(t') \text{ for all } t > \omega\,, \tag{4}$$

where $I(t)$ is a Hilbert subspace of L^2, $\omega \in \mathbb{Z} \cup \{-\infty\}$ represents a "starting point", and $\mathbb{Z}$ is the set of the integers. The "starting point" ω is typically equal to a finite initial date (such as $\omega = -1$, 0 or 1) or to $-\infty$; in the latter case $I(t)$ is defined for all $t \in \mathbb{Z}$. We also consider two stochastic processes

$$X = \{X_t : t \in \mathbb{Z},\, t > \omega\}\ ,\ \ Y = \{Y_t : t \in \mathbb{Z},\, t > \omega\}\,,$$

where

$$\begin{aligned} X_t &= (x_{1,t}, \ldots, x_{d_1,t})',\ x_{i,t} \in L^2, \quad i = 1, \ldots, d_1\,,\ d_1 \geq 1, \\ Y_t &= (y_{1,t}, \ldots, y_{d_2,t})',\ y_{i,t} \in L^2, \quad i = 1, \ldots, d_2\,,\ d_2 \geq 1, \end{aligned}$$

and a (possibly empty) Hilbert subspace H of L^2, whose elements represent information available at any time, such as time independent variables (*e.g.*, the constant in a regression model) and deterministic processes (e.g., deterministic trends). We denote $X(\omega, t]$ the Hilbert space spanned by the components $x_{i,\tau}, i = 1, \ldots, d_1$, of X_τ, $\omega < \tau \leq t$, and similarly for $Y(\omega, t]$: $X(\omega, t]$ and $Y(\omega, t]$ represent the information contained in the history of the variables X and Y respectively up to time t. Finally, the information sets obtained by "adding" $X(\omega, t]$ to $I(t)$ and $Y(\omega, t]$ to $I_X(t)$ are defined as

$$I_X(t) = I(t) + X(\omega, t]\,,\ I_{XY}(t) = I_X(t) + Y(\omega, t]. \tag{5}$$

For any information set B_t [some Hilbert subspace of L^2], we denote $P[x_{i,t+1}|\, B_t]$ the best (nonlinear) forecast of $x_{i,t+1}$ based on the information set B_t,

$$u[x_{i,t+1}|\, B_t] = x_{i,t+1} - P[x_{i,t+1}|\, B_t]$$

the corresponding prediction error, and $\sigma^2[x_{i,t+1}\,|\, B_t] = \mathsf{E}\{u[x_{i,t+1}|\, B_t]^2\}$. Then, the best forecast of X_{t+1} is

$$P[X_{t+1}|\, B_t] = \big(P[x_{1,t+1}\,|\, B_t],\, \dots\,,\, P[x_{d_1,t+1}\,|\, B_t]\big)',$$

the corresponding vector of prediction errors is

$$U[X_{t+1}\,|\, B_t] = \big(u[x_{1,t+1}\,|\, B_t],\, \dots\,,\, u[x_{d_1,t+1}\,|\, B_t]\big)', \tag{6}$$

and the corresponding matrix of second moments is

$$\Sigma[X_{t+1}\,|\, B_t] = \mathsf{E}\big\{U[X_{t+1}\,|\, B_t]\, U[X_{t+1}\,|\, B_t]'\big\}. \tag{7}$$

Provided B_t contains a constant, $\Sigma[X_{t+1}\,|\, B_t]$ is covariance matrix of $U[X_{t+1}\,|\, B_t]$. Each component $P[x_{i,t+1}\,|\, B_t]$ of $P[X_{t+1}\,|\, B_t]$ is the orthogonal projection of $x_{i,t+1}$ on the subspace B_t.

Following the definitions in Dufour and Taamouti (2010), characterization of non-causality between X and Y can be expressed in terms of the variance of the forecast errors.

Definition 1 *(*Covariance Characterization of Non-causality*). Y does not cause X given I iff*

$$\det \Sigma[X_{t+1}\,|\, I_X(t)] = \det \Sigma[X_{t+1}\,|\, I_{XY}(t)]\,,\ \forall t > \omega,$$

where $\Sigma[X_{t+1}\,|\,\cdot\,]$ is defined in (7). Similarly, X does not cause Y given I iff

$$\det \Sigma[Y_{t+1}\,|\, I_Y(t)] = \det \Sigma[Y_{t+1}\,|\, I_{XY}(t)]\,,\ \forall t > \omega,$$

where $\Sigma[Y_{t+1}\,|\,\cdot\,] = \mathsf{E}\{U[Y_{t+1}\,|\,\cdot]\, U[Y_{t+1}\,|\,\cdot]'\}$.

This definition means that Y causes X (resp. X causes Y) if the past of Y (resp. X) improves the forecast of X_{t+1} (resp. Y_{t+1}) based on the information in $I(t)$ and $X(\omega,\, t]$ (resp. $Y(\omega,\, t]$).

4 Causality measures

The causality measures that we propose here are defined using similar measure functions as in Geweke (1982, 1984) and Dufour and Taamouti (2010). Important properties of these measures include: (i) they are non-negative, and (ii) they cancel only when there is no causality. Specifically, we propose the following causality measures where by convention $\ln(0/0) = 0$ and $\ln(x/0) = +\infty$ for $x > 0$.

Definition 2 *(*Mean-Square Causality Measures*). The function*

$$C(Y \to X \mid I) = \ln\left[\frac{\det \Sigma[X_{t+1} \mid I_X(t)]}{\det \Sigma[X_{t+1} \mid I_{XY}(t)]}\right] \tag{8}$$

defines the mean-square causality measure *from Y to X, given I. Similarly, the function*

$$C(X \to Y \mid I) = \ln\left[\frac{\det \Sigma[Y_{t+1} \mid I_Y(t)]}{\det \Sigma[Y_{t+1} \mid I_{XY}(t)]}\right]$$

defines the mean-square causality measure *from X to Y, given I.*

Since we only consider mean-square measures, the term "mean square causality measure" will be abbreviated to "causality measure". Clearly, $C(Y \to X \mid I) = 0$ (resp. $C(X \to Y \mid I) = 0$) if $Y(\omega, t] \subseteq I_X(t)$ (resp. $X(\omega, t] \subseteq I_Y(t)$), so $C(Y \to X \mid I)$ (resp. $C(X \to Y \mid I)$) provides useful information mainly when $Y(\omega, t] \nsubseteq I_X(t)$ (resp. $X(\omega, t] \nsubseteq I_Y(t)$). For $d_1 = d_2 = 1$, Definition 2 reduces to

$$C(Y \to X \mid I) = \ln\left[\frac{\sigma^2[X_{t+1} \mid I_X(t)]}{\sigma^2[X_{t+1} \mid I_{XY}(t)]}\right], \; C(X \to Y \mid I) = \ln\left[\frac{\sigma^2[Y_{t+1} \mid I_Y(t)]}{\sigma^2[Y_{t+1} \mid I_{XY}(t)]}\right].$$

$C(Y \to X \mid I)$ (resp. $C(X \to Y \mid I)$) measures the degree of causal effect from Y to X (resp. X to Y) given I and the past of Y (resp. X). In terms of predictability, this can be viewed as the amount of information brought by the past of Y (resp. X) which can improve the forecast of X_{t+1} (resp. Y_{t+1}). We now define an instantaneous causality measure between X and Y as follows.

Definition 3 *(*Mean-Square Instantaneous Causality Measure*). The function*

$$C(X - Y \mid I) = \ln\left[\frac{\det \Sigma[X_{t+1} \mid I_{XY}(t)] \; \det \Sigma[Y_{t+1} \mid I_{XY}(t)]}{\det \Sigma[\big(X_{t+1},\, Y_{t+1}\big) \mid I_{XY}(t)]}\right],$$

where $\Sigma[(X_{t+1},\, Y_{t+1}) \mid I_{XY}(t)] = \mathsf{E}\big\{U[Z_{t+1} \mid I_{XY}(t)]U[Z_{t+1} \mid I_{XY}(t)]'\big\}$ *and* $Z_t = \big(X_t',\, Y_t'\big)'$, *defines the* mean-square instantaneous causality measure *between X and Y.*

For $d_1 = d_2 = 1$ and provided $I(t)$ includes a constant variable, we have:

$$\begin{aligned}\det \Sigma[\big(X_{t+1},\, Y_{t+1}\big) \mid I_{XY}(t)] &= \sigma^2[X_{t+1} \mid I_{XY}(t)]\, \sigma^2[Y_{t+1} \mid I_{XY}(t)] \\ &\quad -\big(\mathsf{cov}[X_{t+1},\, Y_{t+1} \mid I_{XY}(t)]\big)^2,\end{aligned}$$

so that

$$C(X - Y \mid I) = \ln\left[\frac{1}{1 - \rho[X_{t+1},\, Y_{t+1} \mid I_{XY}(t)]^2}\right], \tag{9}$$

where

$$\rho[X_{t+1},\, Y_{t+1} \mid I_{XY}(t)] = \frac{\mathsf{cov}[X_{t+1},\, Y_{t+1} \mid I_{XY}(t)]}{\sigma[X_{t+1} \mid I_{XY}(t)]\sigma[Y_{t+1} \mid I_{XY}(t)]}.$$

is the conditional correlation coefficient between X_{t+1} and Y_{t+1} given the information set $I_{XY}(t)$. Thus, instantaneous causality increases with the absolute value of the conditional correlation coefficient.

We also define a measure of dependence between X and Y. This will enable one to check whether the processes X and Y must be considered together or whether they can be treated separately.

Definition 4 *(*Measure of Dependence*). The function*

$$C(X,\, Y\,|\,I) = C(X \to Y\,|\,I) + C(Y \to X\,|\,I) + C(X{-}Y\,|\,I) \tag{10}$$

defines the intensity *of the dependence between X and Y, given I.*

It is easy to see that the intensity of the dependence between X and Y can be written in the alternative form:

$$C(X, Y\,|\,I) = \ln\left[\frac{\det \Sigma[X_{t+1}\,|\,I_X(t)]\ \det \Sigma[Y_{t+1}\,|\,I_Y(t)]}{\det \Sigma[\big(X_{t+1},\, Y_{t+1}\big)\,|\,I_{XY}(t)]}\right],$$

where $I_Y(t)$ represents the Hilbert subspace spanned by the components of Y_t and similarly for $I_X(t)$.

5 Causality measures for nonparametric regression models

Let $\big\{(X_t, Y_t) \in \mathbb{R}\times\mathbb{R} \equiv \mathbb{R}^2,\ t = 0, ..., T\big\}$ be a sample of stationary stochastic process in $\mathbb{R}^2$. For simplicity of exposition, we consider univariate Markov processes of order one. Later, see Section 7, we will extend the results to the case where the variables X and Y can be multivariate Markov processes of any order p, for $p \geq 1$.

We now focus on the following bivariate nonparametric regression

$$Z_{t+1} = \Phi\,(Z_t) + u_{t+1}, \tag{11}$$

where $Z_{t+1} = \big(X_{t+1},\, Y_{t+1}\big)'$, $\Phi\,(Z_t)$ is an unknown function of Z_t such that $\Phi\,(Z_t) = E\,[Z_{t+1}|\, Z_t]$, and $u_{t+1} = \big(u^X_{t+1},\, u^Y_{t+1}\big)'$ is an error term with $E\,[u_{t+1}|\, Z_t] = 0$. From the bivariate nonparametric regression in (11), we obtain the following marginal regressions for X_{t+1} and Y_{t+1} :

$$X_{t+1} = \Phi_1\,(Z_t) + u^X_{t+1} \tag{12}$$

and

$$Y_{t+1} = \Phi_2\,(Z_t) + u^Y_{t+1},$$

where $\Phi_1\,(Z_t) = E\,[X_{t+1}|\, X_t, Y_t]$ and $\Phi_2\,(Z_t) = E\,[Y_{t+1}|\, X_t, Y_t]$. In addition, we have

$$\sigma^2[X_{t+1}\,|\,X_t, Y_t] = Var\,\big[u^X_{t+1}\big] = Var\,[(X_{t+1} - \Phi_1\,(Z_t))]\,.$$

Similarly, we have

$$\sigma^2[Y_{t+1}\,|\,X_t, Y_t] = Var\,\big[u^Y_{t+1}\big] = Var\,[(Y_{t+1} - \Phi_2\,(Z_t))]\,.$$

To quantify the degree of causality from Y to X and from X to Y we also need to consider the following constrained nonparametric regressions of X and Y, respectively,

$$X_{t+1} = \bar{\Phi}_1\,(X_t) + \bar{u}^X_{t+1} \tag{13}$$

and

$$Y_{t+1} = \bar{\Phi}_2 (Y_t) + \bar{u}^Y_{t+1},$$

where $\bar{\Phi}_1 (X_t)$ and $\bar{\Phi}_2 (Y_t)$ are unknown functions of X_t and Y_t such that $\bar{\Phi}_1 (X_t) = E[X_{t+1}| X_t]$ and $\bar{\Phi}_2 (Y_t) = E[Y_{t+1}| Y_t]$, respectively, and $\bar{u}^X_{t+1}$and $\bar{u}^Y_{t+1}$ are error terms such that $E\left[\bar{u}^X_{t+1}\middle| X_t\right] = 0$ and $E\left[\bar{u}^Y_{t+1}\middle| Y_t\right] = 0$, respectively. We have

$$\bar{\sigma}^2[X_{t+1} \,|\, X_t] = Var\left[\bar{u}^X_{t+1}\right] = Var\left[\left(X_{t+1} - \bar{\Phi}_1 (X_t)\right)\right].$$

Similarly, we have

$$\bar{\sigma}^2[Y_{t+1} \,|\, Y_t] = Var\left[\bar{u}^Y_{t+1}\right] = Var\left[\left(Y_{t+1} - \bar{\Phi}_2 (Y_t)\right)\right].$$

We can now immediately deduce the following result by using the definitions of causality measures from Y to X and from X to Y [Definition 2].

Proposition 1 *(*Measures of Nonlinear Granger Causality in Mean*). Under assumption (11), the measure of Granger causality in mean from Y to X is given by:*

$$C(Y \to X| \, I\,) = \ln \left[\frac{Var\left[\left(X_{t+1} - \bar{\Phi}_1 (X_t)\right)\right]}{Var\left[(X_{t+1} - \Phi_1 (Z_t))\right]}\right], \tag{14}$$

where $\bar{\Phi}_1 (X_t)$ and $\Phi_1 (Z_t)$ are the restricted and unrestricted nonparametric regression functions of X_t, respectively. Similarly, the measure of Granger causality in mean from X to Y is given by:

$$C(X\!\to Y| \, I) = \ln \left[\frac{Var\left[\left(Y_{t+1} - \bar{\Phi}_2 (Y_t)\right)\right]}{Var\left[(Y_{t+1} - \Phi_2 (Z_t))\right]}\right],$$

where $\bar{\Phi}_2 (Y_t)$ and $\Phi_2 (Z_t)$ are the restricted and unrestricted nonparametric regression functions of Y_t, respectively.

Using Equation (9), a nonparametric regression-based measure of the instantaneous causality between X and Y is given by the following proposition.

Proposition 2 *(*Measure of Instantaneous Granger Causality in Mean*). Under assumption (11), the measure of the instantaneous Granger causality in mean between X and Y is given by:*

$$C(X\!-\!Y \,|\, I)\!=\!\ln \left[\frac{1}{1 - \rho[(X_{t+1} - \Phi_1 (Z_t))\,,\, (Y_{t+1} - \Phi_2 (Z_t))\,]^2}\right], \tag{15}$$

where $\Phi_1 (Z_t)$ and $\Phi_2 (Z_t)$ are the unrestricted nonparametric regression functions of X_{t+1} and Y_{t+1}, respectively.

Finally, the nonparametric regression-based measure of dependence between X and Y can be deduced from its decomposition in (10).

In the next section, we propose nonparametric kernel estimators for the previous Granger causality measures, and we derive their asymptotic distributions. The basic idea is to consider a nonparametric estimation for the restricted and unrestricted forecast errors: $X_{t+1} - \bar{\Phi}_1(X_t)$, $Y_{t+1} - \bar{\Phi}_2(Y_t)$, $X_{t+1} - \Phi_1(Z_t)$, and $Y_{t+1} - \Phi_2(Z_t)$. The causality measures can be simply and consistently estimated by replacing the unknown mean square forecast errors by their nonparametric kernel estimates.

6 Estimation and inference

6.1 Estimation

We have shown, see Section 5, that Granger causality measures can be written in terms of variances of the restricted and unrestricted forecast errors. Thus, these measures can be estimated by replacing the unknown variances by their nonparametric estimates from a finite sample. Hereafter, we focus on the estimation of Granger causality measures from Y to X, $C(Y \to X|\, I)$, which is defined in (14). We can similarly propose estimators for the measures of Granger causality from X to Y and of the instantaneous causality between X and Y. For simplicity of exposition, we will omit the conditioning set I in the Granger causality measures defined in the previous sections.

To consistently estimate $C(Y \to X)$, we need to provide consistent estimates of the restricted and unrestricted mean square forecast errors $\bar{\sigma}^2[X_{t+1} \,|\, X_t]$ and $\sigma^2[X_{t+1} \,|\, X_t, Y_t]$, respectively. The latter depend on the restricted and unrestricted conditional regression functions $\bar{\Phi}_1(X_t)$ and $\Phi_1(Z_t)$. Thus, $\bar{\sigma}^2[X_{t+1} \,|\, X_t]$ and $\sigma^2[X_{t+1} \,|\, X_t, Y_t]$, consequently $C(Y \to X)$, can be consistently estimated using nonparametric estimators of the functions $\bar{\Phi}_1(X_t)$ and $\Phi_1(Z_t)$. The well studied nonparametric estimators for the regression functions are given by the Nadaraya-Watson kernel estimators; see Nadaraya (1964) and Watson (1964). To this end, we define the weights

$$W_{1,t+1}(x, \bar{h}) = \frac{K\left(\frac{x - X_t}{\bar{h}}\right)}{\sum_{s=0}^{T-1} K\left(\frac{x - X_s}{\bar{h}}\right)} \tag{16}$$

that we use to estimate the restricted conditional regression function $\bar{\Phi}_1(\cdot) = E(X_{t+1}|X_t = \cdot)$, where $\bar{h} = \bar{h}_T \in \mathbb{R}^+$ is a sequence of smoothing parameters (i.e. bandwidths) and K is a univariate kernel function. We obtain the following Nadaraya-Watson estimator of $\bar{\Phi}_1(\cdot)$

$$\hat{\bar{\Phi}}_1(x) = \sum_{t=0}^{T-1} W_{1,t+1}(x, \bar{h}) X_{t+1},$$

from which we can obtain the nonparametric residual-based estimator of the restricted mean square forecast error

$$\hat{\bar{\sigma}}^2[X_{t+1} \,|\, X_t] := \frac{1}{T} \sum_{t=0}^{T-1} (\hat{\bar{u}}_{t+1}^X)^2 = \frac{1}{T} \sum_{t=0}^{T-1} (X_{t+1} - \hat{\bar{\Phi}}_1(X_t))^2. \tag{17}$$

Similarly, for $z = (x, y)'$ and a vector of smoothing parameters $h = (h_1, h_2)'$, we define

$$W_{2,t+1}(z, h) = \frac{K_2\left(\frac{z-Z_t}{h}\right)}{\sum_{s=0}^{T-1} K_2\left(\frac{z-Z_s}{h}\right)} \tag{18}$$

as the Nadaraya-Watson weights that we will use to estimate the unrestricted conditional regression function $\Phi_1(\cdot) = E(X_{t+1}|Z_t = \cdot)$, where $K_2\left(\frac{z-Z_t}{h}\right) = K\left(\frac{x-X_t}{h_1}\right) K\left(\frac{y-Y_t}{h_2}\right)$ is a two-product kernel function and K is a univariate kernel function. Likewise, we have

$$\hat{\Phi}_1(z) = \sum_{t=0}^{T-1} W_{2,t+1}(z, h) X_{t+1}$$

and

$$\hat{\sigma}^2\left[X_{t+1} \,|\, X_t, Y_t\right] := \frac{1}{T}\sum_{t=0}^{T-1}(\hat{u}_{t+1}^X)^2 = \frac{1}{T}\sum_{t=0}^{T-1}(X_{t+1} - \hat{\Phi}_1(Z_t))^2 \tag{19}$$

as the Nadaraya-Watson estimator of $\Phi_1(\cdot)$ and nonparametric residual-based estimator of $\sigma^2[X_{t+1} \,|\, X_t, Y_t]$, respectively.

Notice that we have adopted three different bandwidths $\bar{h}$, h_1 and h_2 to take into account the possible data heterogeneity among X_t and Y_t in the nonparametric estimation of restricted and unrestricted conditional regression functions and mean square forecast errors. Furthermore, as is well known in the nonparametric estimation literature, the choices of kernel functions are not important compared to the choices of bandwidths. Therefore, broadly speaking, we could consider the same univariate kernel function $K(\cdot)$ in the above kernel estimators (Assumption **A.2.1**). To further simplify the notation and our asymptotic analysis, in the unrestricted estimation outlined above, we let $h_1 = h_2 = h$. However, different bandwidths could be considered and, with little more complexity, the asymptotic theory developed in this paper will still be valid when $h_1 \neq h_2$.

Using the previous nonparametric estimators $\hat{\bar{\sigma}}^2\left[X_{t+1} \,|\, X_t\right]$ and $\hat{\sigma}^2\left[X_{t+1} \,|\, X_t, Y_t\right]$, an estimator of measure of Granger causality in mean $C(Y \to X)$ is given by

$$\hat{C}(Y \to X) := \ln\left(\frac{\hat{\bar{\sigma}}^2(X_{t+1} \,|\, X_t)}{\hat{\sigma}^2(X_{t+1} \,|\, X_t, Y_t)}\right) = \ln\left(\frac{\frac{1}{T}\sum_{t=0}^{T-1}(X_{t+1} - \sum_{s=0}^{T-1} W_{1,s+1}(X_t, \bar{h})X_{s+1})^2}{\frac{1}{T}\sum_{t=0}^{T-1}(X_{t+1} - \sum_{s=0}^{T-1} W_{2,s+1}(Z_t, h)X_{s+1})^2}\right), \tag{20}$$

where the weights $\bar{W}_{t+1}(x, \bar{h})$ and $W_{t+1}(z, h)$ are defined in (16) and (18), respectively. The most basic property that the above estimator should have is consistency. To prove consistency, some regularity assumptions are needed. We consider a set of standard assumptions on the stochastic processes, bandwidth parameters, and kernel choices in the Nadaraya-Watson estimators.

Assumption A.1 on the stochastic process

(A.1.1) $\{Z_t = (X_t, Y_t)' \in \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2, t \geq 0\}$ is a strictly stationary and ergodic bivariate Markov process of order 1. Furthermore, $E|X_t| < \infty$ and $E|Y_t| < \infty$.

(A.1.2) The marginal density $f_X(x)$ of X_t and the joint density $f_Z(z)$ of Z_t are bounded away from zero and bounded above.

Assumption A.2 on the bandwidth parameters and kernel function

(A.2.1) The kernel functions $K(u)$ and $K_2(u,v) = K(u)K(v)$ are univariate and two-product kernel functions, respectively. Specifically, the univariate kernel $K(u)$ satisfies $K(-u) = K(u)$, $\int K(u)\,du = 1$, $\int uK(u)\,du = 0$ and $\int u^2K(u)\,du < \infty$.

(A.2.2) The univariate bandwidth parameter $\bar{h}$ and the bivariate bandwidth parameter h satisfy $\bar{h} \to 0$, $h \to 0$ and $h^2 = o(\bar{h})$ as $T \to \infty$. Further, $T\bar{h} \to \infty$ and $Th^2 \to \infty$, as $T \to \infty$.

Assumption (**A.1.1**) is standard in asymptotic theory on nonparametric regression for dependent data. It is satisfied by many processes such as ARMA and ARCH processes. This assumption implies that $\bar{u}_t^X$ and u_t^X are strictly stationary and ergodic. Observe that Assumption (**A.1.2**)**,** which requires that the densities $f_X(x)$ and $f_Z(z)$ have to be bounded away from zero, is for convenient purposes only. This assumption eases greatly the derivation of the asymptotic theory. It can be weakened by employing, for example, indicator function as a trimming function like in Robinson (1988) to trim out near zero densities in the nonparametric estimation procedure. We can also use a general weighting function (like densities) to circumvent the problem of random denominator often encountered in the nonparametric estimation and testing literature. We need Assumptions (**A.2.1**) and (**A.2.2**) to show the asymptotic normality of nonparametric estimators. In particular, Assumption (**A.2.1**) is needed to alleviate the bias terms in the nonparametric estimators of variances constructed from the restricted and unrestricted nonparametric residuals. Assumption (**A.2.2**) is a common and minimal assumption in the nonparametric regression literature.

We now state the consistency of the nonparametric estimator $\hat{C}(Y \to X)$ in (20).

Proposition 3 *Under Assumptions (**A.1.1**)-(**A.2.2**), the estimator $\hat{C}(Y \to X)$ in (20) converges in probability to the true Granger causality measure $C(Y \to X)$.*

The proof of Proposition 3 can be found in Appendix B. In the next section we establish the asymptotic normality of $\hat{C}(Y \to X)$ in (20). This will enable us to build tests and confidence intervals for the proposed Granger causality measures.

6.2 Inference

6.2.1 Asymptotic distribution

The causality measures defined in the previous sections can also be used to test for Granger non-causality in mean between X and Y. If there is no causality from Y to X, then the restricted and unrestricted mean square forecast errors of X will be equal: $\bar{\sigma}^2\left[X_{t+1} \mid X_t\right] = \sigma^2[X_{t+1} \mid X_t, Y_t] = \sigma^2$. Hereafter, we wish to test the null hypothesis

$$H_0 : C(Y \to X) = 0. \tag{21}$$

Again, we focus on Granger causality from Y to X, but similar results can be obtained for Granger causality from X to Y and instantaneous causality between X and Y. We now derive the asymptotic normality of nonparametric estimator defined in Equation (20), and we establish the consistency of the test statistic (hereafter $\hat{\Gamma}_T$) which will be used to test H_0.

The following theorem provides the asymptotic normality of nonparametric estimator of our Granger causality measure [see the proof of Theorem 1 in Appendix B].

Theorem 1 *Suppose Assumptions* ***(A.1.1)-(A.2.2)*** *are satisfied. Then under* H_0*, we have*

$$Th\hat{C}(Y \to X) \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where

$$\Omega = \frac{2}{\sigma^4} E \frac{\sigma^4(Z_t)}{f_Z(Z_t)} \int (2K(u) - \overline{K}(u))^2 \, du,$$

with $\overline{K}(u) = \int K(v)K(u+v)\, dv$ *is the convolution of kernels.*

The variance Ω in Theorem 1 can be consistently estimated as follows:

$$\hat{\Omega} = \frac{2}{(\frac{1}{T}\sum_{t=0}^{T-1} \hat{\hat{u}}_{t+1}^{X2})^2} \frac{1}{T(T-1)} \sum_{t=0}^{T-1} \sum_{s=0, s\neq t}^{T-1} K_2\left(\frac{Z_t - Z_s}{h}\right) \frac{\hat{\hat{u}}_{t+1}^{X2} \hat{\hat{u}}_{s+1}^{X2}}{h^2 \hat{f}_Z^2(Z_t)} \int (2K(u) - \overline{K}(u))^2 \, du, \tag{22}$$

where $\hat{\hat{u}}_{t+1}^{X} = X_{t+1} - \hat{\hat{\Phi}}_1(X_t)$ denotes the nonparametric residuals obtained from the restricted nonparametric regression and $\hat{f}_Z(Z_t) = \frac{1}{(T-1)h^2}\sum_{s=0, s\neq t}^{T-1} K_2\left(\frac{Z_t - Z_s}{h}\right)$ is the leave-one-out Nadaraya-Watson estimator of density $f_Z(Z_t)$.

We now define the following test statistic

$$\hat{\Gamma}_T = \frac{Th\hat{C}(Y \to X)}{\sqrt{\hat{\Omega}}}.$$

Theorem 1 implies that under H_0, the test statistic $\hat{\Gamma}_T$ is asymptotically distributed as $N(0, 1)$. This forms the basis for the following one-sided asymptotic test for H_0: for a given significance level α, we reject H_0 if $\hat{\Gamma}_T > z_\alpha$, where z_α is the upper α-percentile of the standard normal distribution.

The following proposition establishes the consistency of the above test [see the proof of Proposition 4 in Appendix B].

Proposition 4 *Under Assumptions* ***(A.1.1)-(A.2.2)****, the test defined by the statistic* $\hat{\Gamma}_T$ *and Theorem 1 is consistent for any loss functions* $\bar{\sigma}^2\left[X_{t+1} \mid X_t\right]$ *and* $\sigma^2[X_{t+1} \mid X_t, Y_t]$ *such that:*

$$\bar{\sigma}^2\left[X_{t+1} \mid X_t\right] - \sigma^2[X_{t+1} \mid X_t, Y_t] > 0,$$

where $\sigma^2[X_{t+1} \mid X_t, Y_t]$ and $\bar{\sigma}^2[X_{t+1} \mid X_t]$ are the mean square forecast errors from the unrestricted and restricted nonparametric regressions in (12) and (13), respectively.

We now examine the asymptotic local power property of the above test. Define the sequence of local alternatives

$$H_{1T} : C(Y \to X) = \frac{1}{Th}\mu, \tag{23}$$

where μ is a finite positive constant, indicating the deviation of $C(Y \to X)$ (degree of causality in mean from Y to X) from zero. The following proposition states that our test has non-trivial asymptotic power against local alternatives converging to the null H_0 at the rate of $(Th)^{-1}$.

Proposition 5 *Under Assumptions **(A.1.1)-(A.2.2)** and H_{1T}, we have*

$$\hat{\Gamma}_T \xrightarrow{d} \mathcal{N}(\mu, 1).$$

We do not provide the proof of Proposition 5 since it is obvious from the proof of Proposition 4. We can immediately conclude that the limiting distribution of nonparametric estimator $\hat{C}(Y \to X)$ is non-trivially shifted whenever $\mu > 0$, and therefore the proposed test is able to detect local alternatives that converge to the null H_0 at the rate of $(Th)^{-1}$. The local power of the test increases with the deviation of μ, and thus our test has non-trivial power against the local alternatives in (23), which are arbitrarily close to the parametric rate $T^{-1/2}$ by taking a large bandwidth h.

6.2.2 Smoothed local bootstrap

The result in Theorem 1 is valid only asymptotically, and the asymptotic normal distribution might not work well in the finite samples. Particularly, for high dimensional random variables the asymptotic test is subject to size distortion because of possible finite sample bias in the nonparametric estimation due to curse of dimensionality. One way to improve the size performance of the asymptotic test is to use the smoothed local bootstrap introduced in Paparoditis and Politis (2000). One major advantage of smoothed local bootstrap procedure is it can preserve the unknown dependence structure in the data, thus it can "mimic" the finite sample distribution of our test statistic.

In the sequel, $X \sim f_X$ means that the random variable X is generated from a density function f_X. Let $L_1(\cdot)$, $L_2(\cdot)$ $L_3(\cdot)$ be three univariate kernels that satisfy Assumption (**A.2.1**) and h^* be a smoothing parameter satisfying Assumption **A.3** below. Hereafter, we discuss the implementation of local smoothed bootstrap. The method is easy to implement in the following four steps:

(1) We draw a bootstrap sample $\{(X_t^*, Y_t^*)\}_{t=1}^T$. We first draw X_{t-1}^* using the nonparametric kernel density of X

$$X_{t-1}^* \sim \frac{1}{Th^*} \sum_{s=1}^{T} L_1 \left(\frac{X_{s-1} - x}{h^*} \right);$$

then conditional on X^*_{t-1}, we draw X^*_t and Y^*_{t-1} independently from the following nonparametric conditional densities

$$X^*_t \sim \frac{1}{h^*}\sum_{s=1}^{T} L_1\left(\frac{X_{s-1}-X^*_{t-1}}{h^*}\right) L_2\left(\frac{X_s - y}{h^*}\right) / \sum_{s=1}^{T} L_1\left(\frac{X_{s-1}-X^*_{t-1}}{h^*}\right)$$

and

$$Y^*_{t-1} \sim \frac{1}{h^*}\sum_{s=1}^{T} L_1\left(\frac{X_{s-1}-X^*_{t-1}}{h^*}\right) L_3\left(\frac{Y_{s-1} - z}{h^*}\right) / \sum_{s=1}^{T} L_1\left(\frac{X_{s-1}-X^*_{t-1}}{h^*}\right);$$

(2) Based on the bootstrap sample, we compute the bootstrapped version of the test statistic: $\hat{\Gamma}^*_T = Th\hat{C}^*(Y \to X)/\sqrt{\hat{\Omega}^*}$;

(3) Repeat the steps (1)-(2) B times so that we get $\hat{\Gamma}^*_{j,T}$, for $j = 1, \ldots, B$;

(4) We compute the bootstrapped *p-value* using $p^* = B^{-1}\sum_{j=1}^{B} 1(\hat{\Gamma}^*_{j,T} > \hat{\Gamma}_T)$, where $\hat{\Gamma}_T = Th\hat{C}(Y \to X)/\sqrt{\hat{\Omega}}$ is the test statistic based on the original sample, and for a given significance level α, we reject the null hypothesis if $p^* < \alpha$.

Notice that in the above bootstrap procedure, we have taken the same bandwidth h^* in the nonparametric kernel estimators of the marginal density of X_{t-1} and the conditional densities of X_t and Y_{t-1}. However, using different bandwidth parameters will not invalidate the local bootstrap. In order to validate the above smoothed local bootstrap, we need to impose an additional assumption concerning the bandwidth parameter h^*.

Assumption A.3 on the bootstrap bandwidth parameter: The bootstrap bandwidth parameter h^* satisfies $h^* \to 0$ and $Th^{*5}/(\ln T)^{\gamma} \to C$, for some $\gamma > 0$ and $0 < C < \infty$, as $T \to \infty$.

Theorem 2 *Suppose Assumptions **(A.1.1)-(A.2.2)** and **A.3** are satisfied. Then under H_0, we have*

$$\hat{\Gamma}^*_T := \frac{Th^*\hat{C}^*(Y \to X)}{\sqrt{\Omega}} \xrightarrow{d} \mathcal{N}(0,1),$$

where Ω is defined in Theorem 1.

7 Measuring causality between high dimensional variables

Due to the prevalent curse of dimensionality, it has been well established that the convergence rate of nonparametric estimator of regression function is slow and decreasing with the dimension of the covariates. This result is also generally true for the estimation of the variance in the context of nonparametric regression. More details on the estimation of variance in the presence of high-dimensional covariates can be found in Spokoiny (2002). In this section, we extend our previous analysis to high dimensional variables.

We consider the following set of standard assumptions. They are only mild modification of Assumptions **(A.1.1)-(A.2.2)**.

Assumption A.1′ on the stochastic process

(A.1.1′) $\{Z_t = (X_t', Y_t')' \in \mathbb{R}^{d_1 p} \times \mathbb{R}^{d_2 p} \equiv \mathbb{R}^d,\, t \geq 0\}$ is a strictly stationary, ergodic vector Markov process of order p.

(A.1.2′) The marginal density $f_X(x)$ of X_t and joint density $f_Z(z)$ of Z_t are bounded away from zero and bounded above. We also assume that both $f_X(x)$ and $f_Z(z)$ are $r+1$-times continuously differentiable on their supports $\mathcal{X}$ and $\mathcal{Z}$, respectively.

Assumption A.2′ on the bandwidth parameters and kernel function

(A.2.1′) The kernel functions $K_{d_1 p}(\cdot)$ and $K_{dp}(\cdot)$ are d_1p-product and dp-product kernel functions, respectively and they are symmetric and bounded. That is, $K_{d_1 p}(u) = \prod_{j=1}^{d_1 p} K(u_j)$ and $K_{dp}(u) = \prod_{j=1}^{dp} K(u_j)$, where $K(u)$ is a univariate kernel that satisfies $\int K(u)\, du = 1$ and $\int u^i K(u)\, du = 0$ for $1 \leq i \leq r-1$ and $\int u^r K(u)\, du < \infty$ with $r \geq 2$.

(A.2.2′) The bandwidth parameters $\bar{h}$ and h satisfy $\bar{h} \to 0$, $h \to 0$ and $h^{d_1+d_2} = o(\bar{h}^{d_1})$ as $T \to \infty$. Further, $T\bar{h}^{d_1 p} \to \infty$, $Th^{(d_1+d_2)p} \to \infty$, as $T \to \infty$.

As we saw in the previous sections, for univariate and first order Markov (i.e. $d_1 = d_2 = 1$ and $p = 1$) processes, a standard Gaussian kernel function ($r = 2$) suffices. However, for high dimensional and high order Markov processes, a higher order kernel is needed. In the following proposition we state the consistency of nonparametric estimator of Granger causality measure $C\,(Y \to X)$.

Proposition 6 *Under Assumptions **(A.1.1′)-(A.2.2′)**, the nonparametric estimator $\hat{C}(Y \to X)$ of measure of Granger causality from Y to X converges in probability to the true Granger causality measure $C(Y \to X)$.*

We now derive the asymptotic normality of $\hat{C}(Y \to X)$ under the null H_0 in (21).

Theorem 3 *Under Assumptions **(A.1.1′)-(A.2.2′)** and H_0, we have*

$$Th^{\frac{(d_1+d_2)p}{2}} \hat{C}(Y \to X) \xrightarrow{d} \mathcal{N}(0, \Omega),$$

where

$$\Omega = \frac{2}{\sigma^4} E \frac{\sigma^4(Z_t)}{f_Z(Z_t)} \int (2K_{d_1 p}(u) - \overline{K}(u))^2\, du,$$

with $\overline{K}(u) = \int K_{d_1 p}(v) K_{d_1 p}(u+v)\, dv$ is the convolution of kernels.

A consistent estimator of the variance Ω in Theorem 3 can be obtained using a similar formula as the one in Equation (22). Furthermore, note that the proofs of Proposition 6 and Theorem 3 are similar to those of Proposition 3 and Theorem 1, respectively, hence we omit them. However, it is important to mention that because of the curse of dimensionality that affects the nonparametric estimation of conditional regression functions, for small samples we suggest to use the bootstrap-assisted test, see Section 6.2.2, to alleviate or eliminate the bias term in the test statistic.

8 Monte Carlo simulations

In this section, we conduct a Monte Carlo simulation study to investigate the performance of the bootstrap-based test that we proposed previously. Our primary interest is to evaluate the empirical size and power of the test in Theorem 2. We will also compare with size and power of Nishiyama et al.'s (2011) nonparametric test for testing the Granger non-causality in mean.

Throughout this section, we consider two univariate time series processes X_t and Y_t. The null hypothesis of interest corresponds to Granger non-causality in mean from Y to X, i.e. $H_0 : C(Y \to X) = 0$. In the sequel, η_t and ε_t are two independent sequences of independently and identically distributed (i.i.d.) standard normal random variables.

8.1 Bootstrap-based test

Though the asymptotic-based test is not time consuming and easy to implement, in small samples the size of the test statistic $\hat{\Gamma}_T$ may differ significantly from the significance level. The size distortion is almost unavoidable for small samples. However, it is also known that some types of bootstrap such as smoothed local bootstrap or moving block bootstrap can help to eliminate or mitigate the asymptotically negligible higher order terms that may have substantial adverse effect on the size of $\hat{\Gamma}_T$. Additional benefits of using smoothed local bootstrap-based test are: (i) it can handle an unknown form of dependence in the data and (ii) it is not very sensitive to changes in the bandwidth parameter δ.

In this section, we investigate the performance of nonparametric test statistic $\hat{\Gamma}^*_T$ in Theorem 2 using the local smoothed bootstrap of Paparoditis and Politis (2000). In particular, we examine its size and power properties using the data generating processes (DGPs) presented in Table 1. Last column of Table 1 summarizes different directions of causality and non-causality in those DGPs. The first four DGPs (DGP S1 to DGP S4) of Y and X are used to investigate the size property, since in those DGPs the null hypothesis is satisfied. However, in DGP P1 to DGP P5 of X the null hypothesis is not satisfied, and therefore those GDPs serve to illustrate the power of our test. Furthermore, notice that DGP S1 and DGP P1 correspond to linear processes, while the other DGPs are highly nonlinear. All DGPs under consideration are strictly stationary and ergodic.

The local bootstrap-based test depends on the bandwidth parameters $\bar{h}$, h, and h^*. In our simulations, we adopt two different bandwidths. For the restricted regression model we use a univariate bandwidth

Table 1: Data-generating processes

DGPs	Variables of Interest		Direction of Causality
	Y_t	X_t	
DGP S1	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP S2	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = \lvert X_{t-1}\rvert^{0.8} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP S3	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1}\exp\{-0.5X_{t-1}^2\} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP S4	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = \sin(X_{t-1}) + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP P1	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + 0.5Y_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P2	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + 0.5Y_{t-1} + 0.5\sin(-2Y_{t-1}) + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P3	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + 0.5Y_{t-1}^2 + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P4	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1}Y_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP P5	$Y_t = 0.5Y_{t-1} + \varepsilon_t$	$X_t = \sin(2(X_{t-1} + Y_{t-1})) + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$

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

$\bar{h} = T^{-1/(2+\delta)}$, and for the unrestricted regression model we use a bivariate bandwidth $(h_1, h_2)'$, with $h_1 = h_2 = h = T^{-1/5}$. To meet the Assumption **(A.2.2)**, $\delta > 0.5$ suffices. We have experimented various δ's ranging from 0.6 to 1 and the choice of $\delta = 0.6$ or 0.8 seems to produce the best overall results. For simplicity, the bandwidth h^* takes the same values as the above univariate bandwidth $\bar{h}$. How to optimally choose the bandwidths $\bar{h}$ and h in order to maximize our test's performance is not yet investigated and calls for more attention in the future work. At least three different ways can be used to choose the bandwidth in practice. The first one is the cross-validation bandwidth proposed by Li et al. (2013). However, strictly speaking, since the cross-validated bandwidth is random, the asymptotic theory can be justified only through certain stochastic equicontinuity argument. The cross-validation technique is used in Li et al. (2009) for testing the equality of two unconditional and conditional functions in the context of mixed categorical and continuous data. This approach, which is optimal for the estimation, loses the optimality for nonparametric kernel testing. The second way is given by an adaptive-rate-optimal rule proposed by Horowitz and Spokoiny (2001) for testing a parametric model for conditional mean function against a nonparametric alternative. The third way for selecting a practical bandwidth is introduced by Gao and Gijbels (2008). The latter propose, using the Edgeworth expansion of the asymptotic distribution of the test, to choose the bandwidth such that the power function of the test is maximized while the size function is controlled. Finally, to estimate the conditional restricted and unrestricted regression functions, we take the univariate kernel function $K(\cdot)$ equal to the standard normal density.

Table 2: Empirical size of local bootstrap-based test

Bandwidth Parameter	DGPs			
$\bar{h} = T^{-1/(2+\delta)}$, $h = T^{-1/5}$	DGP S1	DGP S2	DGP S3	DGP S4
		T = 50		
$\delta = 0.6$	0.046	0.048	0.064	0.048
$\delta = 0.8$	0.048	0.052	0.062	0.050
		T = 75		
$\delta = 0.6$	0.048	0.052	0.060	0.038
$\delta = 0.8$	0.052	0.053	0.064	0.032
		T = 200		
$\delta = 0.6$	0.052	0.048	0.055	0.045
$\delta = 0.8$	0.050	0.046	0.054	0.046

Note: This table reports the empirical size of local bootrsap-based test in Theorem 2 for testing the non-causality in mean from Y to X at $\alpha = 5\%$ significance level. The number of simulations is equal to 1000 and the number of bootstrap resamples is $B = 199$.

Three sample sizes $T = 50$, 75 and 200 are considered. For each DGP, we first generate $T + 200$ observations and then discard the first 200 observations to minimize the effect of the initial values. Other sample sizes ($T = 100$, 300, 400, 500, and 1000) for different DGPs are considered in Section 8.2. We use 1000 simulations to compute the empirical size and power. For each simulation we use $B = 199$ bootstrap replications. Finally, we focus on the nominal size 5%.

Tables 2 and 3 report the empirical size and power of the test statistic Γ_T^* in Theorem 2, respectively. As expected, the local bootstrap-based test performs well, in terms of size, in small samples as small as $T = 50$. This test also has reasonable power against various alternatives. Thus, given the small and moderate samples under consideration, the performance of local bootstrap-based test is satisfactory.

8.2 Comparison with Nishiyama et al.'s (2011) test

In Section 6.2 we saw that testing that the causality measure is equal to zero is equivalent to testing for Granger non-causality in mean. Thus, the tests in Theorems 1 and 2 can be viewed as tests of Granger non-causality in mean. Nishiyama et al. (2011) have recently proposed a nonparametric test for testing nonlinear Granger causality in mean. Under the null hypothesis of non-causality in mean, the asymptotic distribution of Nishiyama et al.'s (2011) test statistic is not normal, but its critical regions can be computed using simulation.

In this section, we consider an additional simulation exercise to compare the empirical size and power of our bootstrap-based test with those of Nishiyama et al.'s (2011) nonparametric test. We use the same

Table 3: Empirical power of local bootstrap-based test

Bandwidth Parameter	DGPs				
$\bar{h} = T^{-1/(2+\delta)}$, $h = T^{-1/5}$	DGP P1	DGP P2	DGP P3	DGP P4	DGP P5
			T = 50		
$\delta = 0.6$	0.232	0.210	0.398	0.254	0.284
$\delta = 0.8$	0.234	0.212	0.390	0.242	0.292
			T = 75		
$\delta = 0.6$	0.394	0.326	0.700	0.494	0.488
$\delta = 0.8$	0.406	0.400	0.624	0.504	0.494
			T = 200		
$\delta = 0.6$	0.630	0.752	0.982	0.870	0.852
$\delta = 0.8$	0.664	0.776	0.950	0.882	0.844

Note: This table reports the empirical power of local bootrsap-based test in Theorem 2 for testing the non-causality in mean from Y to X at $\alpha = 5\%$ significance level. The number of simulations is equal to 1000 and the number of bootstrap resamples is $B = 199$.

simulation settings as in Nishiyama et al. (2011) and consider their DGPs that we report in Table 6 of Appendix A. Since $\delta = 0.6$ and $\delta = 0.8$ lead to quite similar results in terms of size and power, see Section 8.1, in this section we focus on $\delta = 0.8$. We also use the following bandwidths: univariate bandwidth $\bar{h} = T^{-1/(2.8)}$ and bivariate bandwidth $h_1 = h_2 = h = T^{-1/5}$. To facilitate the comparison between the two tests, sample sizes $T = 100$, 200, 300, 400, 500, and 1000 are considered.

Table 4 compares the empirical size and power of our bootstrap-based test in Theorem 2 with those of Nishiyama et al.'s (2011) test for testing the non-causality in mean from Y to X at $\alpha = 5\%$ significance level. From this, we see that the empirical size and power of the two tests are quite reasonable. Both tests control the size and have non trivial power. We see that our test, particularly for DGP3, has better power for relatively small samples. This might be relevant when data scarcity happens, like in our empirical application. For example, for $T = 100$ and for DGP 1 to DGP3, the power of Nishiyama et al.'s (2011) test is 0.092, 0.115 and 0.109, respectively, while the power of our test is equal to 0.158, 0.341 and 0.343, respectively. However, for large sample sizes, the empirical power catches up to one quickly for both tests so that there is in fact no distinguishable differences for the power performance of the two tests.

9 Nonlinear predictability of risk premium

This section aims to join the recent vast literature and study the predictive power of *variance risk premium* for the expected stock excess returns. The variance risk premium is defined as the difference between the

Table 4: Empirical size and power of our local bootstrap-based test and Nishiyama et al. (2011)'s test

Bandwidth Parameters: $\bar{h} = T^{-1/(2.8)}$, $h = T^{-1/5}$								
Sample size	DGPs							
	DGP 0		DGP 1		DGP 2		DGP 3	
	ST test	NHKJ test	ST test	NHKJ test	ST test	NHKJ test	ST test	NHKJ test
$T = 100$	0.058	0.049	0.158	0.092	0.341	0.115	0.343	0.109
$T = 200$	0.051	0.047	0.237	0.236	0.415	0.335	0.503	0.286
$T = 300$	0.050	0.049	0.356	0.455	0.588	0.614	0.635	0.565
$T = 400$	0.053	0.053	0.605	0.614	0.763	0.778	0.822	0.781
$T = 500$	0.048	0.054	0.757	0.760	0.902	0.911	0.905	0.899
$T = 1000$	0.051	0.049	0.995	0.989	1.000	1.000	1.000	0.998

Note: This table reports and compares the empirical size and power of our local bootstrap-based test (ST test) and Nishiyama et al.'s (2011) test (NHKJ test) for testing Granger non-causality in mean from Y to X at $\alpha = 5\%$ significance level based on 1000 replications. The data generating processes used in the simulation study are reported in Table 6. We simulate (Y_t, X_t), for $t = 1, \ldots, T$, under the assumption that $(\varepsilon_t, \eta_t)'$ are i.i.d. from $N(0, I_2)$.

risk-neutral and objective expectations of realized variance, where the risk-neutral expectation of variance is measured as the end-of-month Volatility Index-squared de-annualized, and the realized variance is the sum of squared 5-minute log returns of the S&P 500 index over the month.

Recently, many papers have shown the importance of using variance risk premium for predicting expected stock and bond returns and exchange rates; see Bollerslev et al. (2009), Wang et al. (2013), Bollerslev et al. (2014), and Della Corte et al. (2014). For the post-1990 period, Bollerslev et al. (2009) find that the variance risk premium is able to explain a non-trivial fraction of the time series variation in aggregate stock market returns, with high (low) premia predicting high (low) future returns.

Most existing works, however, focus on linear predictability. The econometric methodology used in this context is an ordinary least squares regression of returns onto the past of variance risk premium. In this section, we examine *nonlinear* predictability of expected stock excess returns (risk premium) using variance risk premium. Nonparametric Granger causality measures proposed in the previous sections do not impose any restriction on the model linking the dependent variable (stock excess return) to the independent variable (variance risk premium).

9.1 Variance risk premium

In this subsection, we define the variance risk premium (VRP) that we use as a predictor of risk premium. To do so, we need to define the model-free realized and implied variances.

Let us first set some notations. We denote by p_t the logarithmic price of risky asset at time t, and by $r_{t+1} = p_{t+1} - p_t$ the continuously compounded return from time t to $t+1$. We implicitly assume that the price process could belong to the class of continuous-time jump diffusion processes,

$$dp_t = \mu_t dt + \sigma_t dW_t + \kappa_t dq_t, \;\; 0 \leq t \leq T, \tag{24}$$

where μ_t is a continuous and locally bounded variation process, σ_t is the stochastic volatility process, W_t denotes a standard Brownian motion, dq_t is a counting process such that $dq_t = 1$ represents a jump at time t (and $dq_t = 0$ no jump) with jump intensity λ_t. The parameter κ_t refers to the size of the jumps. Further, we normalize the time-interval to unity and we divide it into h periods. Each period has length $\Delta = 1/h$. Let the discretely sampled Δ-period returns be denoted by $r_{(t,\,\Delta)} = p_t - p_{t-\Delta}$. The realized variance over the discrete t to $t+1$ time interval is defined as the summation of the h high-frequency intradaily squared returns:

$$RV_{t,t+1} \equiv \sum_{j=1}^{h} r^2_{(t+j\Delta,\,\Delta)}$$

and it satisfies

$$\lim_{\Delta \to 0} RV_{t,t+1} = Var_{t,t+1}, \tag{25}$$

where $Var_{t,t+1}$ is the variance of stock excess return between time t and $t+1$. Equation (25) indicates that the realized variance is a consistent estimator of the true variance of stock excess return; see Andersen and Bollerslev (1998), Andersen et al. (2001, 2010), Barndorff-Nielsen and Shephard (2002a), Barndorff-Nielsen and Shephard (2002b), and Comte and Renault (1998).

We now define the model-free implied variance. Let $C_t(T, K)$ denote the price of a European call option with time to maturity T and strike price K, and $B(t, T)$ denotes the price of a time t zero-coupon bond maturing at time T. Carr and Madan (1998), Demeterfi et al. (1999) and Britten-Jones and Neuberger (2000), have shown that implied variance between time t and $t+1$, say $IV_{t,t+1}$, can be replicated by a portfolio of European calls as follows:

$$IV_{t,t+1} \equiv E_t^{\mathbb{Q}}\left(Var_{t,t+1}\right) = 2\int_0^{\infty} \frac{1}{K^2}\left[C_t(t+1, \frac{K}{B(t,t+1)}) - C_t(t,K)\right] dK, \tag{26}$$

where "$E_t^{\mathbb{Q}}$" denotes the conditional expectation with respect to risk-neutral probability, see also Bakshi and Madan (2000). Equation (26) depends on an increasing number of calls with strikes spanning zero to infinity. In practice $IV_{t,t+1}$ must be constructed on the basis of a finite number of strikes. Several recent works have argued that even with relatively few different option strikes this tends to provide a fairly accurate approximation to the true risk-neutral expectation of the future market variance; for review see Jiang and Tian (2005), Carr and Wu (2009), and Bollerslev et al. (2011).

We now use the above model-free realized and implied variances to define the variance risk premium, which is theoretically equal to the difference between the ex-ante risk neutral expectation of the future stock return variance and the expectation of stock return variance between time t and $t+1$:

$$VRP_t \equiv E_t^{\mathbb{Q}}\left(Var_{t,t+1}\right) - E_t^{\mathbb{P}}\left(Var_{t,t+1}\right), \tag{27}$$

where "$E_t^{\mathbb{P}}$" denotes the conditional expectation with respect to physical probability. VRP_t in Equation (27) is unobservable, since the quantities $E_t^{\mathbb{Q}}\left(Var_{t,t+1}\right)$ and $E_t^{\mathbb{P}}\left(Var_{t,t+1}\right)$ are unobservable. Estimating VRP_t depends on the estimation of risk neutral and physical expectations:

$$\widehat{VRP}_t \equiv \hat{E}_t^{\mathbb{Q}}\left(Var_{t,t+1}\right) - \hat{E}_t^{\mathbb{P}}\left(Var_{t,t+1}\right).$$

In practice, $\hat{E}_t^{\mathbb{Q}}\left(Var_{t,t+1}\right)$ and $\hat{E}_t^{\mathbb{P}}\left(Var_{t,t+1}\right)$ are commonly replaced by the squared-Volatility Index (VIX) and the realized variance $RV_{t,t+1}$, respectively. 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.

In the literature there is no unique approach for constructing the physical expectation $\hat{E}_t^{\mathbb{P}}(.)$. Bollerslev et al. (2009) and Zhou (2010) have estimated a reduced-form multi-frequency autoregression with potentially multiple lags for $\hat{E}_t^{\mathbb{P}}\left(Var_{t,t+1}\right)$. Following Bollerslev et al. (2009) and Zhou (2010), we use time-t realized variance $RV_{t,t-1}$, which ensures that the variance risk premium proxy for predicting various risk premia is in the time t information set and would be a correct choice if the realized variance process were unit-root.

9.2 Empirical results

We use monthly aggregate S&P 500 composite index over the period January 1996 to September 2008. Our empirical analysis is based on the logarithmic return on the S&P 500 in excess of 3-month T-bill rate. We also consider monthly realized variance, implied variance, and variance risk premium which can be downloaded from Hao Zhou's website.

Table 5 presents the results of estimating the measures of Granger causality from variance risk premium to risk premium at horizons one month to 9 months. It also reports the information about the statistical significance of the estimates of the measures using the bootstrap-based test proposed in Section 6.2.2, as well as the NHKJ's nonparametric test for testing the null hypothesis of Granger non-causality in mean. From this, we see that the degree of predictability (causality) starts out fairly high at the monthly horizon with an estimate around 1.30 and it is significant at 1% level. This result is not in line with what have been found in Bollerslev et al. (2009), who, using linear regression models, find that there is a very weak predictability from VRP to risk premium at one month horizon. Thus, our result can be interpreted as nonlinear predictability of risk premium. On the other hand, NHKJ's test indicates that there is non-causality from VRP to risk premium at all horizons. This might be due to the relatively large sample sizes needed for the NHKJ test to have power.

Table 5: Measures of causality (predictability) from variance risk premium to risk premium

Direction of Causality	Bandwidth $n^{-1/(2+\delta)}$	Estimate of Causality Measure	NHKJ Test
	Horizon: One Month		
VRP→RP			
	$\delta = 0.6$	1.298***	
	$\delta = 0.8$	1.307***	1.41
	$\delta = 1.0$	1.316***	
	Horizon: Three Months		
VRP→RP			
	$\delta = 0.6$	0.436***	
	$\delta = 0.8$	0.468***	0.85
	$\delta = 1.0$	0.500***	
	Horizon: Six Months		
VRP→RP			
	$\delta = 0.6$	0.170	
	$\delta = 0.8$	0.239	1.82
	$\delta = 1.0$	0.309*	
	Horizon: Nine Months		
VRP→RP			
	$\delta = 0.6$	0.000	
	$\delta = 0.8$	0.016	0.33
	$\delta = 1.0$	0.128	

Note: This table reports the results of the estimation and inference for measures of Granger causality (predictability) from variance risk premium (VRP) to risk premium (RP), at different horizons. "***" and "*" mean the statistical significance at 1% and 10% significance levels, respectively. The 5% critical value for Nishiyama et al.'s (2011) (NHKJ) test is 14.38.

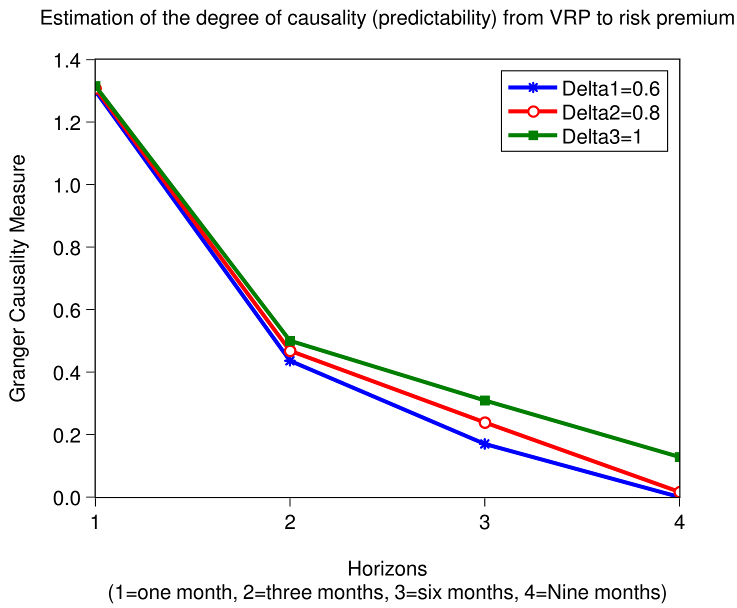

Figure 1: This figure plots the estimates of Granger causality measures from variance risk premium (VRP) to risk premium at horizons that go from one to 9 months, and using different bandwidth parameter δ (in the figure delta). The data on the S&P 500 market index goes from January 1996 to September 2008.

The degree of predictability is still higher (measure around 0.468) and statistically significant at three-month horizon, which is in line with the findings of Bollerslev et al. (2009). However, the degree of predictability decreases and becomes statistically insignificant after three-month horizon. Figure 1 shows that the predictive power of VRP is a decreasing function of time horizon. In contrast, Bollerslev et al. (2009) find that the degree of predictability starts out fairly low at the monthly horizon, rising to its maximum around a quarter, gradually tapering off thereafter for longer return horizons. Hence, unlike Bollerslev et al. (2009), we find a high degree of predictability at one-month horizon which can be attributed to a nonlinear causal effect from VRP to risk premium.

10 Conclusion

In this paper, we extended the existing *parametric* measures of Granger causality in mean [see Geweke (1982, 1984) and Dufour and Taamouti (2010)] by proposing *nonparametric* measures that are able to detect and quantify nonlinear Granger causality in mean between univariate and multivariate random variables. The new measures are model-free, therefore they do not require the specification of model that links the variables of interest.

The proposed causality measures are defined as a logarithmic function of restricted and unrestricted mean square forecast errors. To consistently estimate these measures, it suffices to find consistent estimates of the above mean square forecast errors. The latter are defined using Nadaraya-Watson kernel estimators. Since testing for the statistical significance of measures is also indispensable in time series analysis, we derived the asymptotic normality of nonparametric estimator of measures that we used to build valid tests. We also established the validity of smoothed local bootstrap that one can use in finite sample settings to perform statistical tests. Monte Carlo simulation study has been performed to investigate the finite sample properties (size and power) of the proposed test and the results reveal that the latter behaves well for a variety of typical data generating processes.

Using the above nonparametric test for testing the null hypothesis that the true value of measure is equal to zero is equivalent to testing for non-causality in mean. Thus, our test can be viewed as a competitor of the exiting nonparametric tests of Granger causality in mean. There is only one nonparametric test of Granger causality in mean which is proposed by Nishiyama et al. (2011). We considered an additional simulation exercise to compare the empirical size and power of our test with those of Nishiyama et al.'s (2011) test. Simulation results indicate that our test has comparable size, but better power than Nishiyama et al.'s (2011) test.

Finally, we applied our nonparametric measures to quantify the degree of nonlinear predictability of risk premium using variance risk premium. The latter is defined as difference between risk-neutral and objective expectations of realized variance, where the risk-neutral expectation of variance is measured as the end-of-month volatility index-squared de-annualized and the realized variance is the sum of squared 5-minute log returns of the S&P 500 index over the month. Our results showed that the variance risk premium is a good predictor of risk premium at horizons less than six months. Unlike Bollerslev et al. (2009), we found that there is a high degree of predictability at horizon one-month which can be attributed to a nonlinear causal effect.

References

[1] Andersen, T. G., and Bollerslev, T. (1998), "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review,* 39, 885–905.

[2] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association,* 96, 42–55.

[3] Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2010), "Parametric and Nonparametric Volatility Measurement," In Y. Aït-Sahalia and L. P. Hansen (eds.), Handbook of Financial Econometrics. Volume 1-Tools and Techniques, chapter 2, 67–138. Amsterdam: Elsevier.

[4] Bakshi, G., and Madan, D. (2000), "Spanning and Derivative-Security Valuation," *Journal of Financial Economics*, 55, 205–238.

[5] Barndorff-Nielsen, O. E., and Shephard, N. (2002a), "Econometric Analysis of Realised Volatility and Its Use in Stochastic Volatility Models," *Journal of the Royal Statistical Society*, Series B, 64, 253–280.

[6] Barndorff-Nielsen, O. E., and Shephard, N. (2002b), "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics,* 17, 457–478.

[7] Bollerslev, T., Gibson, M., and Zhou, H. (2011), "Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion From Option-Implied and Realized Volatilities," *Journal of Econometrics*, 160, 235–245.

[8] Bollerslev, T., Marrone, J., Xu, L., and Zhou, H. (2014), "Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence," *Journal of Financial and Quantitative Analysis,* 49, 633–661.

[9] Bollerslev, T., Tauchen, G., and Zhou, H. (2009), "Expected Stock Returns and Variance Risk Premia," *Review of Financial Studies*, 22, 4463–4492.

[10] Britten-Jones, M., and Neuberger, A. (2000), "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 55, 839–866.

[11] Brock, W. (1991), "Causality, Chaos, Explanation and Prediction in Economics and Finance," in J. Casti and A. Karlqvist, Eds.: Beyond Belief: Randomness, Prediction and Explanation in Science (CRC Press, Boca Raton, Fla.)

[12] Carr, P., and Madan, D. (1998), "Towards a Theory of Volatility Trading," in Volatility: New Estimation Techniques for Pricing Derivatives, chap.29, 417–427, Robert Jarrow (ed.). London: Risk Books.

[13] Carr, P., and Wu, L. (2009), "Variance Risk Premiums," *Review of Financial Studies*, 22, 1311–1341.

[14] Comte, F., and Renault, E. (1998), "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance,* 8, 291–323.

[15] Della Corte, P., Ramadorai, T., and Sarno, L. (2014), "Volatility Premia and Exchange Rate Predictability," *Journal of Financial Economics*, *forthcoming.*

[16] Demeterfi, K., Derman, E., Kamal, M., and Zou, J. (1999), "A Guide to Volatility and Variance Swaps," *Journal of Derivatives,* 6, 9–32.

[17] Dufour, J.-M., and Taamouti, A. (2010), "Short and Long Run Causality Measures: Theory and Inference," *Journal of Econometrics*, 154, 42–58.

[18] Gao, J., and Gijbels, I. (2008), "Bandwidth Selection in Nonparametric Kernel Testing," *Journal of the American Statistical Association*, 103, 1584–1594.

[19] Geweke, J. (1982), "Measurement of Linear Dependence and Feedback Between Multiple Time Series," *Journal of the American Statistical Association*, 77, 304–313.

[20] Geweke, J. (1984), "Measures of Conditional Linear Dependence and Feedback Between Time Series," *Journal of the American Statistical Association*, 79, 907–915.

[21] Gouriéroux, C., Monfort, A., and Renault, E. (1987), "Kullback Causality Measures," *Annales d'Économie et de Statistique,* 6/7, 369–410.

[22] Granger, C. W. J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–459.

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

[24] Härdle, W. (1992), "Applied Nonparametric Regression (Econometric Society Monographs)," Cambridge University Press.

[25] Horowitz, J. L., and Spokoiny, V. G. (2001), "Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternatives," *Econometrica*, 69, 599–631.

[26] Jiang, G., and Tian, Y. (2005), "Model-Free Implied Volatility and Its Information Content," *Review of Financial Studies*, 18, 1305–1342.

[27] Li, Q.,Lin, J., and Racine, J. S. (2013), "Optimal Bandwidth Selection for Nonparametric Conditional Distribution and Quantile Functions," *Journal of Business and Economic Statistics*, 31, 57–65.

[28] Li, Q., Maasoumi, E., and Racine, J. S. (2009), "A Nonparametric Test for Equality of Distributions with Mixed Categorical and Continuous Data," *Journal of Econometrics*, 148, 186–200.

[29] McCloskey, D. N., and Ziliak, S. (1996), "The Standard Error of Regressions," *The Journal of Economic Literature*, XXXIV, 97–114.

[30] Müller, U., Schick, A., and Wefelmeyer, W. (2003), "Estimating the Error Variance in Nonparametric Regression by a Covariate-Matched U-Statistic," *Statistics*, 37, 179–188.

[31] Nadaraya, E. A. (1964), "On Estimating Regression," *Theory of Probability and its Applications*, 9, 141–142.

[32] Nishiyama, Y., Hitomi, K., Kawasaki, Y., and K. Jeong. (2011), "A Consistent Nonparametric Test for Nonlinear Causality-Specification in Time Series Regression," *Journal of Econometrics*, 165, 112–127.

[33] Paparoditis, E., and Politis, D. (2000), "The Local Bootstrap for Kernel Estimators Under General Dependence Conditions," *Annals of the Institute of Statistical Mathematics*, 52, 139–159.

[34] Polasek, W. (1994), "Temporal Causality Measures Based on AIC," *in H. Bozdogan, ed., Proceedings of the Frontier of Statistical Modeling: An Informal Approach*, Kluwer, Dordrecht, The Netherlands, 159–168.

[35] Polasek, W. (2002), "Bayesian Causality Measures for Multiple ARCH Models Using Marginal Likelihoods," *Technical report*, Institute of Statistics and Econometrics, Univeristy of Basel., Switzerland.

[36] Robinson, P. M. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

[37] Spokoiny, V. (2002), "Variance Estimation for High-Dimensional Regression Models," *Journal of Multivariate Analysis*, 82, 111–133.

[38] Taamouti, A., Bouezmarni, T., and El Ghouch, A. (2014), "Nonparametric Estimation and Inference for Granger Causality Measures," *Journal of Econometrics,* 180, 251–264.

[39] Tenreiro, C. (1997), "Loi Asymptotique des Erreurs Quadratiques Intégrées des Estimateurs a Noyau de la Densité et de la Régression Sous des Conditions de Dépendance," *Portugaliae Mathematica*, 54, 197–213.

[40] Wang, H., Zhou, H., and Zhou, Y. (2013), "Credit Default Swap Spreads and Variance Risk Premia," *Journal of Banking and Finance,* 37, 3733–3746.

[41] Watson, G. S. (1964), "Smooth Regression Analysis," *Sankya, Series A*, 26, 359–372.

[42] White, H. (2000), "Asymptotic Theory for Econometricians," 2nd Edition, New York: Academic Press.

[43] Wiener, N. (1956), "The Theory of Prediction," *in E. F. Beckenback, ed., The Theory of Prediction*, McGraw-Hill, New York, chapter 8.

[44] Zhou, H. (2010), "Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty," *Federal Reserve Board Working Paper.*

Table 6: Nishiyama et al. (2011)'s data-generating processes

DGPs	Variables of Interest		Direction of Causality
	Y_t	X_t	
DGP 0	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \nrightarrow X$
DGP 1	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + 0.2Y_{t-1} + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP 2	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + 0.2Y_{t-1} + 0.4\sin(-2Y_{t-1}) + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$
DGP 3	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + 0.2Y_{t-1}^2 + \eta_t$	$X \nrightarrow Y$, $Y \rightarrow X$

Note: This table summarizes the data generating processes that we use in the simulation study to compare the empirical size and power of our local bootstrap-based test in Theorem 2 with those of Nishiyama et al.'s (2011) test. We simulate (Y_t, X_t), for $t = 1, \ldots, T$, under the assumption that $(\varepsilon_t, \eta_t)'$ are i.i.d. from $N(0, I_2)$. The last column of the table summarizes the directions of Granger causality and non-causality in each DGP. "$\rightarrow$" and "$\nrightarrow$" refer to Granger causality and Granger non-causality, respectively.

A Appendix: Additional DGPs

B Appendix: Proofs

This appendix provides the proofs of the main theoretical results developed in Sections 6 and 7.

Proof of Proposition 3. First, observe that

$$\begin{aligned}\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] =&\frac{1}{T}\sum_{t=0}^{T-1}(\hat{\bar{u}}_{t+1}^X)^2 = \frac{1}{T}\sum_{t=0}^{T-1}(\bar{u}_{t+1}^X + (\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t)))^2 \\ =&\frac{1}{T}\sum_{t=0}^{T-1}(\bar{u}_{t+1}^X)^2 + 2\frac{1}{T}\sum_{t=0}^{T-1}\bar{u}_{t+1}^X\left(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t)\right) + \frac{1}{T}\sum_{t=0}^{T-1}\left(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t)\right)^2.\end{aligned}$$

Thus, we have

$$\begin{aligned}\left|\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] - \bar{\sigma}^2[X_{t+1}\,|\,X_t]\right| \leq& \left|\frac{1}{T}\sum_{t=0}^{T-1}(\bar{u}_{t+1}^X)^2 - \bar{\sigma}^2[X_{t+1}\,|\,X_t]\right| \\ &+ 2\sup_{x\in S_X}\left|\bar{\Phi}_1(x) - \hat{\bar{\Phi}}_1(x)\right|\left|\frac{1}{T}\sum_{t=0}^{T-1}\left|\bar{u}_{t+1}^X\right|\right| \\ &+ \frac{1}{T}\sum_{t=0}^{T-1}\left(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t)\right)^2 = o_p(1),\end{aligned}$$

where the first term is an $o_p(1)$ by the strict stationarity, ergodicity, and (weak) law of large numbers. The second term is an $o_p(1)$ by utilizing the fact that $\sup_{x\in S_X}\left|\bar{\Phi}_1(x) - \hat{\bar{\Phi}}_1(x)\right| = o_p(1)$ and $E|\bar{u}_{t+1}^X|$ is bounded according to Assumption **(A.1.1)**. The last term is also an $o_p(1)$ by noticing that it converges to $E\left(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t)\right)^2$; the mean squared errors (MSE) of Nadaraya-Watson kernel estimator $\hat{\bar{\Phi}}_1(\cdot)$, which is asymptotically negligible, see Härdle (1992). Consequently, we have shown that $\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] = \bar{\sigma}^2[X_{t+1}\,|\,X_t] + o_p(1)$. Similarly, we can show that $\hat{\sigma}^2[X_{t+1}\,|\,X_t, Y_t] = \sigma^2[X_{t+1}\,|\,X_t, Y_t] + o_p(1)$. Moreover, under some appropriate assumptions, we could even show stronger results, that is; $\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] = \sum_{t=0}^{T-1}(\bar{u}_{t+1}^X)^2/T + o_p(T^{-1/2})$ and $\hat{\sigma}^2[X_{t+1}\,|\,X_t, Y_t] = \sum_{t=0}^{T-1}(u_{t+1}^X)^2/T + o_p(T^{-1/2})$. For more details, see for instance, Spokoiny (2002) and Müller et al. (2003).

Finally, by the first order Taylor expansion of $\ln x = \ln x_0 + (x - x_0)/x_0 +$ higher order terms, and because $\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] = \bar{\sigma}^2[X_{t+1}\,|\,X_t] + o_p(1)$ and $\hat{\sigma}^2[X_{t+1}\,|\,X_t, Y_t] = \sigma^2[X_{t+1}\,|\,X_t, Y_t] + o_p(1)$, we get

$$\ln\left(\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t]\right) = \ln\left(\bar{\sigma}^2[X_{t+1}\,|\,X_t]\right) + o_p(1),$$

$$\ln\left(\hat{\sigma}^2[X_{t+1}\,|\,X_t, Y_t]\right) = \ln\left(\sigma^2[X_{t+1}\,|\,X_t, Y_t]\right) + o_p(1).$$

Therefore,

$$\hat{C}(Y \to X) := \ln\left(\frac{\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t]}{\hat{\sigma}^2[X_{t+1}\,|\,X_t, Y_t]}\right) = \ln\left(\frac{\bar{\sigma}^2[X_{t+1}\,|\,X_t]}{\sigma^2[X_{t+1}\,|\,X_t, Y_t]}\right) + o_p(1).$$

Hence, $\hat{C}(Y \to X)$ converges in probability to $C(Y \to X)$. ■

Proof of Theorem 1. Let $Z_t = (X_t, Y_t)'$. First of all, recall that

$$\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] = \frac{1}{T}\sum_{t=0}^{T-1}(X_{t+1} - \hat{\bar{\Phi}}_1(X_t))^2, \quad \hat{\sigma}^2[X_{t+1}\,|\,Z_t] = \frac{1}{T}\sum_{t=0}^{T-1}(X_{t+1} - \hat{\Phi}_1(Z_t))^2,$$

and

$$|\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] - \bar{\sigma}^2[X_{t+1}\,|\,X_t]| = o_p(1), \quad |\hat{\sigma}^2[X_{t+1}\,|\,Z_t] - \sigma^2[X_{t+1}\,|\,Z_t]| = o_p(1).$$

Now, since $\ln x \approx (x-1)$, by the first order Taylor expansion around $x = 1$, we obtain

$$\hat{C}(Y \to X) = \ln\left(\frac{\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t]}{\hat{\sigma}^2[X_{t+1}\,|\,Z_t]}\right) \approx \left(\frac{\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t]}{\hat{\sigma}^2[X_{t+1}\,|\,Z_t]} - 1\right) := A_T.$$

We shall show that under the null hypothesis of Granger non-causality in mean (i.e. $\bar{\sigma}^2[X_{t+1}\,|\,X_t] = \sigma^2[X_{t+1}\,|\,Z_t] = \sigma^2$), and under assumptions of Theorem 1, we have $ThA_T \xrightarrow{d} \mathcal{N}(0,\Omega)$, where the asymptotic variance Ω is specified as in Theorem 1.

Notice that $\bar{u}^X_{t+1} = u^X_{t+1}$ a.s. under the null (for simplicity, we will omit the superscript X from now on), we have that

$$\begin{aligned}
A_T =& \frac{\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] - \hat{\sigma}^2[X_{t+1}\,|\,Z_t]}{\hat{\sigma}^2[X_{t+1}\,|\,Z_t]} = \frac{\hat{\bar{\sigma}}^2[X_{t+1}\,|\,X_t] - \hat{\sigma}^2[X_{t+1}\,|\,Z_t]}{\sigma^2}[1+o_p(1)] \\
=& \frac{\frac{1}{T}\sum_{t=0}^{T-1}(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t) + u_{t+1})^2 - \frac{1}{T}\sum_{t=0}^{T-1}(\bar{\Phi}_1(Z_t) - \hat{\bar{\Phi}}_1(Z_t) + u_{t+1})^2}{\sigma^2}[1+o_p(1)] \\
=& \frac{1}{\sigma^2}\left[\frac{2}{T}\sum_{t=0}^{T-1}(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t))u_{t+1} - \frac{2}{T}\sum_{t=0}^{T-1}(\Phi_1(Z_t) - \hat{\Phi}_1(X_t))u_{t+1}\right][1+o_p(1)] \\
&+ \frac{1}{\sigma^2}\left[\frac{1}{T}\sum_{t=0}^{T-1}(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t))^2 - \frac{1}{T}\sum_{t=0}^{T-1}(\Phi_1(Z_t) - \hat{\Phi}_1(Z_t))^2\right][1+o_p(1)] \\
=& \frac{1}{\sigma^2}(A_{1T} - A_{2T} + A_{3T})[1+o_p(1)],
\end{aligned}$$

where

$$A_{1T} := \frac{2}{T}\sum_{t=0}^{T-1}(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t))u_{t+1}, \quad A_{2T} := \frac{2}{T}\sum_{t=0}^{T-1}(\Phi_1(Z_t) - \hat{\Phi}_1(X_t))u_{t+1},$$

$$A_{3T} := \frac{1}{T}\sum_{t=0}^{T-1}(\bar{\Phi}_1(X_t) - \hat{\bar{\Phi}}_1(X_t))^2 - \frac{1}{T}\sum_{t=0}^{T-1}(\Phi_1(Z_t) - \hat{\Phi}_1(Z_t))^2.$$

We can show that $Th(A_{1T} - A_{2T})$ and ThA_{3T} are both asymptotically normal and they are asymptotically independent.

Since $\sup_{x\in\mathcal{X}}|\hat{f}_X(x) - f_X(x)| = o_p(1)$, we can write

$$\begin{aligned}
A_{1T} =& \frac{2}{T(T-1)}\sum_{t\neq s}\frac{1}{\bar{h}f_X(X_t)}K\left(\frac{X_t - X_s}{\bar{h}}\right)(\bar{\Phi}_1(X_t) - \bar{\Phi}_1(X_s))u_{t+1}[1+o_p(1)] \\
&- \frac{2}{T(T-1)}\sum_{t\neq s}\frac{1}{\bar{h}f_X(X_t)}K\left(\frac{X_t - X_s}{\bar{h}}\right)u_{t+1}u_{s+1}[1+o_p(1)].
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
A_{2T} =& \frac{2}{T(T-1)}\sum_{t\neq s}\frac{1}{h^2 f_Z(Z_t)}K_2\left(\frac{Z_t - Z_s}{h}\right)(\Phi_1(Z_t) - \Phi_1(Z_s))u_{t+1}[1+o_p(1)] \\
&- \frac{2}{T(T-1)}\sum_{t\neq s}\frac{1}{h^2 f_Z(Z_t)}K_2\left(\frac{Z_t - Z_s}{h}\right)u_{t+1}u_{s+1}[1+o_p(1)].
\end{aligned}$$

We can show later that both first terms of A_{1T} and A_{2T} are $o_p((Th)^{-1})$ under the conditions assumed in Theorem 1. Thus,

$$\begin{aligned}
Th(A_{1T}-A_{2T}) =& Th\frac{2}{T(T-1)}\sum_{t\neq s}\left(\frac{1}{h^2 f_Z(Z_t)}K_2\left(\frac{Z_t-Z_s}{h}\right)u_{t+1}u_{s+1}\right.\\
&\left.-\frac{1}{\bar{h}f_X(X_t)}K\left(\frac{X_t-X_s}{\bar{h}}\right)u_{t+1}u_{s+1}\right)+o_p(1)\\
=&\frac{2}{T-1}\sum_{t\neq s}h\left(\frac{1}{h^2 f_Z(Z_t)}K_2\left(\frac{Z_t-Z_s}{h}\right)\right.\\
&\left.-\frac{1}{\bar{h}f_X(X_t)}K\left(\frac{X_t-X_s}{\bar{h}}\right)\right)u_{t+1}u_{s+1}+o_p(1).
\end{aligned}$$

Let $W_t=(Z_t',u_{t+1})'$ and

$$u_T(W_t,W_s)=\left(\frac{1}{h^2 f_Z(Z_t)}K_2\left(\frac{Z_t-Z_s}{h}\right)-\frac{1}{\bar{h}f_X(X_t)}K\left(\frac{X_t-X_s}{\bar{h}}\right)\right)u_{t+1}u_{s+1},$$

so that we can rewrite

$$\begin{aligned}
Th(A_{1T}-A_{2T}) =&\frac{1}{T-1}\sum_{t\neq s}h(u_T(W_t,W_s)+u_T(W_s,W_t))+o_p(1)\\
=&2\frac{1}{T-1}\sum_{t<s}H_T(W_t,W_s)+o_p(1):=2T_{11}+o_p(1),
\end{aligned}$$

where $H_T(W_t,W_s)=h(u_T(W_t,W_s)+u_T(W_s,W_t))$. Notice that $E[H_T(W_t,W_s)]=0$. Therefore, T_{11} is a degenerate second order U-statistic. We shall show that $T_{11}\xrightarrow{d}\mathcal{N}(0,\Omega_1)$ with properly defined asymptotic variance Ω_1. To achieve so, we shall verify the conditions of Theorem 1 in Tenreiro (1997). Let $\tilde{W}_0$ be an independent copy of W_0. So, $E[H_T(W_0,\tilde{W}_0)]^2=2\Omega_1+o(1)$.

Notice that

$$H_T(W_t,W_s)=h\left[\frac{1}{h^2}\left(\frac{1}{f_Z(Z_t)}+\frac{1}{f_Z(Z_s)}\right)K_2\left(\frac{Z_t-Z_s}{h}\right)-\frac{1}{\bar{h}}\left(\frac{1}{f_X(X_t)}+\frac{1}{f_X(X_s)}\right)K\left(\frac{X_t-X_s}{\bar{h}}\right)\right]u_{t+1}u_{s+1}.$$

In what follows, the conditional variance of u_{t+1}, conditional on Z_t, will be denoted by $\sigma^2(Z_t)$, i.e. $\sigma^2(Z_t)=E(u_{t+1}^2|Z_t)$. The dominant term of $E[H_T(W_0,\tilde{W}_0)]^2$ is

$$\begin{aligned}
&E\left[h^2\left(\frac{1}{h^4}\left(\frac{1}{f_Z(Z_0)}+\frac{1}{f_Z(\tilde{Z}_0)}\right)^2K_2^2\left(\frac{Z_0-\tilde{Z}_0}{h}\right)u_1^2\tilde{u}_1^2\right)\right]\\
=&\frac{1}{h^2}\int_{z_0,\tilde{z}_0}\left(\frac{1}{f_Z(z_0)}+\frac{1}{f_Z(\tilde{z}_0)}\right)^2K_2^2\left(\frac{z_0-\tilde{z}_0}{h}\right)\sigma^2(z_0)\sigma^2(\tilde{z}_0)f_Z(z_0)f_Z(\tilde{z}_0)\,dz_0\,d\tilde{z}_0\\
=&\int_a\int_{z_0}\left(\frac{1}{f_Z(z_0)}+\frac{1}{f_Z(z_0+ha)}\right)^2K^2(a)\sigma^2(z_0)\sigma^2(z_0+ha)f_Z(z_0)f_Z(z_0+ha)\,dz_0\,da\\
=&\int_{z_0}\frac{4}{f_Z^2(z_0)}\sigma^4(z_0)f_Z^2(z_0)\,dz_0\int K^2(a)\,da+o(1)\\
=&4\int_{z_0}\sigma^4(z_0)\,dz_0\int K^2(a)\,da+o(1)\\
=&4E\left(\frac{\sigma^4(Z_t)}{f_Z(Z_t)}\right)\int K^2(a)\,da+o(1):=2\Omega_1+o(1).
\end{aligned}$$

Thus, $Th(A_{1T} - A_{2T}) \xrightarrow{d} \mathcal{N}(0, 4\Omega_1)$.

To show that the first term of A_{1T} is $o_p((Th)^{-1})$, we follow the similar proof as above. Since

$$\begin{aligned}
&Th\frac{2}{T(T-1)}\sum_{t\neq s}\frac{1}{\bar{h}f_X(X_t)}K\left(\frac{X_t-X_s}{\bar{h}}\right)(\bar{\Phi}_1(X_t)-\bar{\Phi}_1(X_s))u_{t+1}\\
=&\frac{h}{\bar{h}^{1/2}}\frac{2}{(T-1)}\sum_{t<s}\frac{1}{\bar{h}^{1/2}}\left(\frac{u_{t+1}}{f_X(X_t)}-\frac{u_{s+1}}{f_X(X_s)}\right)K\left(\frac{X_t-X_s}{\bar{h}}\right)(\bar{\Phi}_1(X_t)-\bar{\Phi}_1(X_s))\\
=&\frac{h}{\bar{h}^{1/2}}\frac{2}{(T-1)}\sum_{t<s}H_{1T}(W_t,W_s).
\end{aligned}$$

Following the same arguments as for T_{11}, one can show that $\frac{2}{(T-1)}\sum_{t<s}H_{1T}(W_t, W_s)$ is asymptotically negligible because by straightforward calculation $E[H_{1T}(W_0, \tilde{W}_0)]^2 = O(\bar{h}^2)$. By $h^2 = o(\bar{h})$, we conclude that the above term is $o_p(1)$. The first term of ThA_{2T} is also $o_p(1)$ and can be proved analogously as that of ThA_{1T} and hence is omitted.

We now proceed to show that ThA_{3T} is asymptotically normal. Because

$$\begin{aligned}
A_{3T} =&\frac{1}{T}\sum_t(\hat{\Phi}_1(Z_t)-\hat{\bar{\Phi}}_1(X_t))(\bar{\Phi}_1(X_t)-\hat{\bar{\Phi}}_1(X_t))+\frac{1}{T}\sum_t(\hat{\Phi}_1(Z_t)-\hat{\bar{\Phi}}_1(X_t))(\Phi_1(Z_t)-\hat{\Phi}_1(Z_t))\\
=&A_{3T}^1+A_{3T}^2,
\end{aligned}$$

we shall prove A_{3T}^1 and A_{3T}^2 converge in distribution to normal random variables, respectively. Concerning the first term A_{3T}^1, using $X_{t+1} = \bar{\Phi}(X_t) + u_{t+1}$, we have

$$\begin{aligned}
A_{3T}^1 =&\frac{1}{T}\sum_t\left(\frac{((T-1)h^2)^{-1}\sum_{s\neq t}K_2\left(\frac{Z_t-Z_s}{h}\right)X_{s+1}}{f_Z(Z_t)}-\frac{((T-1)\bar{h})^{-1}\sum_{s\neq t}K\left(\frac{X_t-X_s}{\bar{h}}\right)X_{s+1}}{f_X(X_t)}\right)\\
&\times\left(\frac{((T-1)\bar{h})^{-1}\sum_{j\neq t}K\left(\frac{X_t-X_j}{\bar{h}}\right)(\bar{\Phi}(X_t)-X_{j+1})}{f_X(X_t)}\right)[1+o_p(1)]\\
=&-\frac{1}{T(T-1)^2}\sum_{s\neq t,j\neq t}\frac{1}{h^2\bar{h}}K_2\left(\frac{Z_t-Z_s}{h}\right)K\left(\frac{X_t-X_j}{\bar{h}}\right)\frac{u_{s+1}u_{j+1}}{f_Z(Z_t)f_X(X_t)}\\
&+\frac{1}{T(T-1)^2}\sum_{s\neq t,j\neq t}\frac{1}{\bar{h}^2}K\left(\frac{X_t-X_s}{\bar{h}}\right)K\left(\frac{X_t-X_j}{\bar{h}}\right)\frac{u_{s+1}u_{j+1}}{f_X^2(X_t)}+\text{higher order terms}\\
=&A_{3T}^{11}+A_{3T}^{12}+\text{higher order terms}.
\end{aligned}$$

For the the first term, define

$$\psi_T(W_t, W_s, W_j) = \frac{1}{h^2\bar{h}}K_2\left(\frac{Z_t-Z_s}{h}\right)K\left(\frac{X_t-X_j}{\bar{h}}\right)\frac{u_{s+1}u_{j+1}}{f_Z(Z_t)f_X(X_t)},$$

and $\phi_T(W_t, W_s, W_j) = \psi_T(W_t, W_s, W_j) + \psi_T(W_s, W_t, W_j) + \psi_T(W_j, W_s, W_t)$. We have

$$-A_{3T}^{11} = \frac{6}{T(T-1)(T-2)}\sum_{t<s<j}\phi_T(W_t, W_s, W_j),$$

a third order U-statistic. We use Hoeffding decomposition to study A_{3T}^{11}. Notice that $E[\phi_T(W_t, W_s, W_j)] = 0$, $E[\phi_T(W_t, W_s, W_j)|W_t] = 0$, and

$$E[\phi_T(W_t, W_s, W_j)|W_t, W_s] = \frac{u_{t+1}u_{s+1}}{h^2\bar{h}} E\left[\frac{1}{f_Z(Z_j)f_X(X_j)} K_2\left(\frac{Z_t - Z_j}{h}\right) K\left(\frac{X_s - X_j}{\bar{h}}\right) |W_t, W_s\right] = \phi_{2T}(W_t, W_s).$$

Thus, we have

$$-A_{3T}^{11} \approx \frac{6}{T(T-1)} \sum_{t<s} \phi_{2T}(W_t, W_s).$$

One can further establish that $A_{3T}^{12} = o_p((Th)^{-1})$ by first decomposing A_{3T}^{12} (also a third order U-statistic) into a second order U-statistic and using $h^2 = o(\bar{h})$.

Now, regarding the term A_{3T}^2, we follow the same arguments as above to obtain $A_{3T}^2 \approx A_{3T}^{21} + A_{3T}^{22}$, where

$$A_{3T}^{21} \approx \frac{6}{T(T-1)} \sum_{t<s} \phi'_{2T}(W_t, W_s), \quad A_{3T}^{22} \approx \frac{6}{T(T-1)} \sum_{t<s} \phi''_{2T}(W_t, W_s),$$

with

$$\phi'_{2T}(W_t, W_s) = \frac{u_{t+1}u_{s+1}}{h^2\bar{h}} E\left[\frac{1}{f_Z(Z_j)f_X(X_j)} K_2\left(\frac{Z_s - Z_j}{h}\right) K\left(\frac{X_t - X_j}{\bar{h}}\right) |W_t, W_s\right],$$

and

$$\phi''_{2T}(W_t, W_s) = \frac{u_{t+1}u_{s+1}}{h^4} E\left[\frac{1}{f_Z^2(Z_j)} K_2\left(\frac{Z_t - Z_j}{h}\right) K_2\left(\frac{Z_s - X_j}{h}\right) |W_t, W_s\right].$$

After some tedious calculation, we finally show that the term ThA_{3T} converges to a normal distribution with variance $2E\left(\frac{\sigma^4(Z_t)}{f_Z(Z_t)}\right)(\int \overline{K}^2(u)\, da - 4\int K(u)\overline{K}(u)\, du)$, where $\overline{K}(u) = \int K(v)K(u+v)\, dv$ is the convolution of kernels. Hence we have proved Theorem 1. ■

Proof of Proposition 4. To prove the consistency of our test, we only have to follow the same steps as in the proof of Theorem 1. Notice that, under the global alternative hypothesis $\bar{\sigma}^2[X_{t+1} \mid X_t] - \sigma^2[X_{t+1} \mid Z_t] > 0$, we have

$$\begin{aligned} ThA_T =& Th\frac{\frac{1}{T}\sum_{t=0}^{T-1}\{(\hat{\bar{u}}_{t+1}^2 - \bar{\sigma}^2[X_{t+1} \mid X_t]) - (\hat{u}_{t+1}^2 - \sigma^2[X_{t+1} \mid Z_t])\}}{\sigma^2[X_{t+1} \mid Z_t]}[1 + o_p(1)] \\ &+ \frac{Th(\bar{\sigma}^2[X_{t+1} \mid X_t] - \sigma^2[X_{t+1} \mid Z_t])}{\sigma^2[X_{t+1} \mid Z_t]}[1 + o_p(1)] \\ =& O_p(1) + \frac{Th(\bar{\sigma}^2[X_{t+1} \mid X_t] - \sigma^2[X_{t+1} \mid Z_t])}{\sigma^2[X_{t+1} \mid Z_t]}. \end{aligned}$$

The second equality follows because the first term converges in distribution to a normal random variable following the same arguments as in proof of Theorem 1. Therefore, the statistic $Th\hat{C}(Y \to X)$ diverges to infinity. Hence, we have shown that the proposed test is consistent. ■

Proof of Theorem 2. Theorem 2 can be proved using similar arguments to the ones we used in the proof of Theorem 1, with the term A_T replaced by its bootstrapped versions A_T^* using the bootstrap data $\{X_t^*, Y_t^*, \}_{t=1}^T$. Conditionally on $\{X_t, Y_t, \}_{t=1}^T$ and using Theorem 1 of Hall (1984), we obtain the result in Theorem 2. ■