

# Measuring Granger Causality in Quantiles\*

Xiaojun Song<sup>†</sup>

Peking University

Abderrahim Taamouti<sup>‡</sup>

Durham University Business School

February 22, 2020

---

\*The authors thank Todd Clark (Editor), an Associate Editor and three anonymous referees for several useful comments. They also thank the seminar participants from Aix-Marseille University, University of Bristol, and Dresden University, and the participants from the 5th Annual Econometrics Workshop in Liverpool, the Southampton Finance and Econometrics Workshop in Southampton, CIREQ Montreal Econometrics Workshop, and the 6th Shanghai Econometrics Workshop for giving very useful comments. Special thanks to Marine Carrasco, Jean-Marie Dufour, Christian Gouriéroux, Nour Meddahi, Marcelo Moreira, Benoît Perron, and Victoria Zinde-Walsh for their valuable remarks and comments. An earlier version of this paper circulated under the title “Measuring Nonlinear Granger Causality in Quantiles”.

<sup>†</sup>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](mailto:sxj@gsm.pku.edu.cn). The author acknowledges the financial support from China’s National Key Research Special Program Grants 2016YFC0207705, the National Natural Science Foundation of China (Grant No. 71532001, 71973005), and Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University), Ministry of Education, China.

<sup>‡</sup>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](mailto:abderrahim.taamouti@durham.ac.uk).

# Measuring Granger Causality in Quantiles

## ABSTRACT

We consider measures of Granger causality in quantiles, which *detect* and *quantify* both linear and nonlinear causal effects between random variables. The measures are based on nonparametric quantile regressions and defined as logarithmic functions of restricted and unrestricted expectations of quantile check loss functions. They can consistently be estimated by replacing the unknown expectations of check loss functions by their nonparametric kernel estimates. We derive a Bahadur-type representation for the nonparametric estimator of the measures. We establish the asymptotic distribution of this estimator, which can be used to build tests for the statistical significance of the measures. Thereafter, we show the validity of a smoothed local bootstrap that can be used in finite-sample settings to perform statistical tests. A Monte Carlo simulation study reveals that the bootstrap-based test has a good finite-sample size and power properties for a variety of data generating processes and different sample sizes. Finally, we provide an empirical application to illustrate the usefulness of measuring Granger causality in quantiles. We quantify the degree of predictability of the quantiles of equity risk premium using the variance risk premium, unemployment rate, inflation, and the effective federal funds rate. The empirical results show that the variance risk premium and effective federal funds rate have a strong predictive power for predicting the risk premium when compared to that of the predictive power of the other two macro variables. In particular, the variance risk premium is able to predict the center, lower and upper quantiles of the distribution of the risk premium; however, the effective federal funds rate predicts only the lower and upper quantiles. Nevertheless, unemployment and inflation rates have no effect on the risk premium.

**Keywords:** Measures of Granger causality; conditional quantiles; local linear estimator; smoothed local bootstrap.

**JEL Classification Number:** 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. An examination of Wiener-Granger causality reveals that predictability is the central issue and is of great importance to economists, policymakers, and investors. Many studies have investigated the building of the tests of Granger non-causality. However, once it has been established that a “causal relationship” exists, it is usually important to assess its strength. Few studies have proposed to measure Granger causality in mean between the variables of interest; refer to Geweke (1982, 1984), Dufour and Taamouti (2010), and Song and Taamouti (2018). However, these studies ignore or pay less attention to the causality that can occur at other levels or aspects of the conditional distribution, such as conditional quantiles. This study considers the measures of Granger causality in quantiles and, particularly, proposes model-free measures that can quantify both linear and nonlinear causal effects in quantiles.

Causality tests fail to accomplish the task of quantifying the degree of Granger causality between the variables of interest because they only provide the evidence on the presence of causality. A large effect may not, at a given level, be statistically significant. Further, a statistically significant effect may be neither “large” from an economic or a subject at hand viewpoint nor is it relevant for decision making. In this study, beyond the acceptance or rejection of the non-causality hypothesis, which state that certain variables do not help forecast other variables, we wish to assess the magnitude of the forecast improvement, where the latter is defined in terms of some loss function (expectation of quantile check loss function). Even if the hypothesis of no improvement cannot be rejected by looking at the available data (for example, because the sample size or the structure of the process does not allow for high test power), sizeable improvements may remain consistent with the same data. By contrast, a statistically significant improvement may easily be produced by a large data set, which may not be relevant from a practical viewpoint.

The topic of measuring Granger causality has not attracted much attention. Further, most of the existing measures focus on Granger causality in mean; thus, they cannot be used in the presence of causality in quantiles. Geweke (1982, 1984) introduce measures of causality in mean based on vector autoregressive models. Dufour and Taamouti (2010) extend Geweke’s work to propose measures for short and long run causality in mean using vector autoregressive moving average 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 and a Bayesian approach. Song and Taamouti (2018) propose nonparametric measures of Granger causality in mean, extending the parametric approach of Geweke (1982, 1984) and Dufour and Taamouti (2010). Taamouti et al. (2014) propose a nonparametric estimator and test for measures of Granger causality in *distribution*. However, the latter measures are not informative enough regarding the level(s) (quantiles)

of the conditional distribution wherein the causality exists; hence, the importance of providing measures of Granger causality in quantiles is relevant. There are also a number of papers dealing with Granger causality in fields different from economics; e.g. communications in complex systems and neuroscience. To study Granger causality between individual components in complex systems, Marinazzo, Pellicoro, and Stramaglia (2008) use measures of nonlinear causality based on the theory of reproducing kernel Hilbert spaces. To examine the causal relations in neural systems, Li, Wei, Billings, and Liao (2012) consider nonlinear time-varying parametric models to build Granger causality measures that detect temporal causal interactions. Furthermore, Hu and Liang (2014) suggest a model-free copula-based Granger causality measure to examine nonlinear and high-order moment causality in neural data.

This study considers measures of Granger causality in quantiles that are able to *detect* and *quantify* both linear and nonlinear causal effects, which can happen at any quantile of the conditional distribution of the variable of interest. The measures are based on nonparametric quantile regressions and defined as logarithmic function of restricted and unrestricted expectations of quantile check loss functions. A consistent estimator of these measures is defined in terms of local polynomial estimators of the expectations of check loss functions. Further, we derive a Bahadur-type representation for this estimator and establish its asymptotic distribution, which we use to build tests for statistical significance of measures. Thereafter, we show the validity of smoothed local bootstrap that we apply to perform statistical tests in finite-sample settings. A Monte Carlo simulation study reveals that the bootstrap-based test has a good finite-sample size and power properties for a variety of data generating processes as well as different sample sizes. Moreover, since testing that the value of measure is equal to zero is equivalent to testing for the non-causality in quantile, we consider an additional simulation exercise to compare the empirical size and power of our test with those of the nonparametric test of non-causality in quantile, introduced by Jeong, Härdle and Song (2012).<sup>1</sup> For the finite samples, the simulation results indicate that our test controls the size and has better power compared to that of Jeong et al.'s (2012) test. We have also suggested a bootstrap bias-corrected estimator for our causality measure and investigated its finite sample performance using Monte Carlo simulations.

Finally, the empirical importance of measuring causality in quantiles is illustrated. We quantify the degree of predictability of equity risk premium using the variance risk premium, unemployment rate, inflation, and the effective federal funds rate. The empirical results show that the variance risk premium and effective federal funds rate have a strong predictive power for predicting the risk premium compared to that of the predictive power of the other two macro variables. In particular, the variance risk premium is able to predict the center, the lower and upper quantiles of the distribution of risk premium; however, the effective federal funds rate only predicts the lower and upper quantiles. Nevertheless, unemployment rate and inflation have no effect on the quantiles (distribution) of risk premium.

---

<sup>1</sup>For the parametric tests of Granger causality in quantiles; refer to Lee and Yang (2014) and Troster (2018), among others.

The plan of the study is as follows. Section 2 provides the motivation for considering measures for Granger causality in quantiles. Section 3 presents the theoretical framework that underlies the definition of measures of causality in quantiles. In Section 4, we provide an estimator of the measures using nonparametric quantile regressions and based on local polynomial approach. We establish the Bahadur representation of this estimator in Section 5.1. In Section 5.2, we derive the asymptotic distribution of the nonparametric estimator, which can be used to build tests for the statistical significance of the measures; we further study their properties under some local alternatives. In Section 6, we show the validity of smoothed local bootstrap, which can offer more accurate approximation to the finite-sample distribution of our test statistic. In Section 7, we extend our results to the case wherein additional control variables are considered. Section 8 presents a Monte Carlo simulation exercise to investigate the finite-sample properties of the estimates and tests of causality measures. Section 9 is devoted to an empirical application and the conclusion of the results is given in Section 10. The key assumptions of the study and the proofs of the theoretical results are presented in the Technical Appendix, which is available online.

## 2 Motivation

This study considers measures of Granger causality in *quantiles*. Most of the existing measures focus on quantifying Granger causality in *mean*; refer to Geweke (1982, 1984), Dufour and Taamouti (2010), and Song and Taamouti (2018). Other measures have been proposed to quantify Granger causality in *distribution*; refer to Taamouti et al. (2014). The latter measures of causality in distribution, however, are not informative enough of the level(s) (quantiles) of distribution wherein the causality exists. Thus, the significance of such measures is limited in the presence of Granger causality in quantiles. A recent paper by Adrian, Boyarchenko, and Giannone (2018) illustrates the importance of studying the causality at different levels (quantiles) of the conditional distribution of GDP growth using economic and financial conditions. Using parametric quantile regressions, they found that the lower quantiles of GDP growth exhibit strong variation as a function of financial conditions, whereas the upper quantiles do not change over time. Hereafter, we consider measures of causality in quantiles based on nonparametric quantile regressions. Such measures can quantify both linear and nonlinear causality in quantiles. To motivate such measures, consider the following example.

**Example 1** Consider the following mean regression model:

$$X_t = \mu + \zeta X_{t-1} + (\gamma + \delta Y_{t-1}^2)^{1/2} \varepsilon_t, \text{ for some } \gamma, \delta \geq 0, \quad (1)$$

where  $\varepsilon_t$  is normally (elliptically) distributed with mean zero and variance one. Since  $\varepsilon_t$  is normally distributed,  $X_t$  is also conditionally normally distributed:

$$X_t | I_{XY}(t-1) \sim \mathcal{N}(\mu + \zeta X_{t-1}, \gamma + \delta Y_{t-1}^2),$$

where  $I_{XY}(t-1)$  will be formally defined in Section 3, but it represents the available information at time  $(t-1)$  that contains both past of  $X$  and  $Y$ . Consequently, for  $\tau \in (0, 1)$ , the  $\tau$ -th quantile of  $X_t$  conditional on  $I_{XY}(t-1)$  is given by:

$$Q^\tau(X_t | I_{XY}(t-1)) = \mu + \zeta X_{t-1} + (\gamma + \delta Y_{t-1}^2)^{1/2} \Phi^{-1}(\tau), \quad (2)$$

where  $\Phi^{-1}(\tau)$  is the  $\tau$ -th quantile of  $\mathcal{N}(0, 1)$ . Equation (1) shows that  $Y$  does not cause  $X$  through its mean. However, from Equation (2) we observe that  $Y$  does cause  $X$  through its  $\tau$ -th quantile. For example, for  $\tau = 0.10$ ,  $Y$  does cause  $X$  through its 10-th quantile even if there is no causality through its mean. This example illustrates the case where the causality in mean does not exist, but it does exist in quantiles. However, how can we measure the degree of this causality in quantile?

It should also be noted that the functional form of Granger causality in quantiles from  $Y$  to  $X$  can be unknown. In other words, the mean regression model can generally be defined as follows:

$$X_t = m(X_{t-1}, Y_{t-1}) + \sigma(X_{t-1}, Y_{t-1}) \varepsilon_t,$$

where  $\varepsilon_t$  is normally (elliptically) distributed with mean zero and variance one. This implies that the  $\tau$ -th quantile of  $X_t$  conditional on  $I_{XY}(t-1)$  is given by:

$$Q^\tau(X_t | I_{XY}(t-1)) = m(X_{t-1}, Y_{t-1}) + \sigma(X_{t-1}, Y_{t-1}) \Phi^{-1}(\tau),$$

where the functional forms of  $m(X_{t-1}, Y_{t-1})$  and  $\sigma(X_{t-1}, Y_{t-1})$  are unknown. This illustrates the importance of developing model-free measures that can quantify the degree of both linear and nonlinear Granger causality in quantiles, which can be achieved using nonparametric quantile regressions. As pointed out in Cai and Xu (2008), for many empirical applications, using linear quantile regression model might not be “rich” enough to capture the underlying relationship between the variables of interest. However, there is an active and growing literature on nonparametric quantile regressions that can accommodate nonlinear dependencies. In particular, many smoothing methods (e.g. kernel methods) have been used to estimate the nonparametric quantile regression for both independent and time series data; see Chaudhuri (1991), Yu and Jones (1998), Honda (2000), Cai and Xu (2008), Jeong et al. (2012) and references therein.

### 3 Framework

We consider two variables of interest  $X$  and  $Y$ . For the simplicity of exposition, we assume that  $X$  and  $Y$  are univariate processes; however, in Section 7 we include a third variable  $W$ . In other words, in Section 7 we will still focus on the univariate quantiles; we will nevertheless extend the information set to include a third variable that can play the role of an auxiliary variable.

We wish to measure the Granger causality in quantiles between  $X$  and  $Y$ . When it comes to causality analysis, defining the information sets is crucial. Hereafter, we consider a sequence  $I$  of “reference information sets”  $I(t-1)$  such that:

$$I = \{I(t) : t \in \mathbb{Z}, t > \omega\} \text{ with } t < t' \Rightarrow I(t) \subseteq I(t') \text{ for all } t > \omega, \quad (3)$$

where  $I(t)$  is an information set,  $\omega \in \mathbb{Z} \cup \{-\infty\}$  represents a “starting point”, and  $\mathbb{Z}$  is the set of integers. The “starting point”  $\omega$  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  $t \in \mathbb{Z}$ . The information set  $I(t)$  could correspond to a (possibly empty) set, whose elements represent the information available at any point of time, such as time independent variables (e.g., the constant in a regression model) and deterministic processes (e.g., deterministic trends).

We denote  $X(\omega, t-1]$  the information set spanned by  $X_s$  for  $\omega < s \leq t-1$ , and similarly for  $Y(\omega, t-1]$ . That is,  $X(\omega, t-1]$  and  $Y(\omega, t-1]$  represent the information contained in the history of the variables  $X$  and  $Y$ , respectively, up to time  $t-1$ . Furthermore, the information sets obtained by “adding”  $X(\omega, t-1]$  to  $I(t-1)$  and  $Y(\omega, t-1]$  to  $I_X(t-1)$  are defined as

$$I_X(t-1) = I(t-1) + X(\omega, t-1], \quad I_{XY}(t-1) = I_X(t-1) + Y(\omega, t-1]. \quad (4)$$

We now remind the reader how Granger non-causality in quantiles can be characterized in terms of restricted and unrestricted expectations of quantile check loss functions. First of all, to avoid useless repetitions, in the following sections we treat only the causality from  $Y$  to  $X$ . For any information set  $B_{t-1}$  [ $B_{t-1}$  might represents  $I_X(t-1)$  or  $I_{XY}(t-1)$ ], we denote  $Q^\tau(X_t | B_{t-1})$  the best linear/nonlinear quantile forecast of  $X_t$  based on the information set  $B_{t-1}$ . Hereafter, the best forecast of  $X_t$  will be defined based on the quantile check loss function:

$$\rho_\tau(\varepsilon) \equiv (\tau - 0.5)\varepsilon + 0.5|\varepsilon| = \varepsilon(\tau - 1(\varepsilon < 0)), \quad (5)$$

where  $1(A)$  is an indicator function of an event  $A$  and  $\tau$  is the quantile of interest. In other words, the best quantile forecast of  $X_t$  based on the information set  $B_{t-1}$  is

$$Q^\tau(X_t | B_{t-1}) = \operatorname{argmin}_\theta E[\rho_\tau(X_t - \theta) | B_{t-1}].$$

Using the above notations the following definition provides a characterization of Granger non-causality in quantiles from  $Y$  to  $X$ .

**Definition 1** (Characterization of Non-Causality in Quantiles). *For  $\tau \in (0, 1)$ ,  $Y$  does not cause  $X$  through its  $\tau$ -th quantile given  $I$ , iff*

$$E[\rho_\tau(X_t - Q^\tau(X_t | I_X(t-1)))] = E[\rho_\tau(X_t - Q^\tau(X_t | I_{XY}(t-1)))]], \quad \forall t > \omega,$$

where the sets  $I_X(t-1)$  and  $I_{XY}(t-1)$  are defined in (4).

Using Definition 1, the measures of Granger causality in quantiles that we consider are based on the following simple idea. For  $\tau \in (0, 1)$ , we state that “ $Y$  causes  $X$  through its  $\tau$ -th quantile” if

$$E[\rho_\tau(X_t - Q^\tau(X_t | I_X(t-1)))] > E[\rho_\tau(X_t - Q^\tau(X_t | I_{XY}(t-1)))] . \quad (6)$$

Consequently, measuring the difference between the two loss functions on the left and right-hand side of the inequality (6) is the same as measuring the strength of causality in the  $\tau$ -th quantile from  $Y$  to  $X$ . The causality measures that we use here are suitable functions of the “distance” between the loss functions on the left and right-hand side of the inequality in (6). These measures are defined using similar measure functions as in Geweke (1982, 1984) as well as Dufour and Taamouti (2010). Important properties of these measure functions are that they are non-negative and cancel only when there is no causality. Specifically, we consider the following causality measures where by convention  $\ln(0/0) = 0$  and  $\ln(x/0) = +\infty$  for  $x > 0$ .

**Definition 2** (Quantile Causality Measure). *For  $\tau \in (0, 1)$ , the function*

$$C_\tau(Y \rightarrow X | I) = \ln \left[ \frac{E[\rho_\tau(X_t - Q^\tau(X_t | I_X(t-1)))]}{E[\rho_\tau(X_t - Q^\tau(X_t | I_{XY}(t-1)))]} \right] \quad (7)$$

*defines the  $\tau$ -th quantile causality measure from  $Y$  to  $X$ , given  $I$ .*

$C_\tau(Y \rightarrow X | I)$  measures the degree of causal effect from  $Y$  to the  $\tau$ -th quantile of  $X$  given the past of  $Y$ . With relation to predictability, it can be viewed as a measure of the amount of information brought by the past of  $Y$ , which can in turn improve the forecast of the  $\tau$ -th quantile of  $X_t$ .

## 4 Estimation

Let  $\{(X_t, Y_t) \in \mathbb{R} \times \mathbb{R} \equiv \mathbb{R}^2, t = 0, \dots, T\}$  be a sample of strictly stationary stochastic process in  $\mathbb{R}^2$ . Denote  $\underline{X}_{t-1} = (X_{t-1}, \dots, X_{t-d_1})'$  and  $\underline{Y}_{t-1} = (Y_{t-1}, \dots, Y_{t-d_2})'$  for some fixed known integers  $d_1 \geq 1$  and  $d_2 \geq 1$ . We assume that the following nonparametric unconstrained quantile regression hold,

$$X_t = q_\tau(\underline{Z}_{t-1}) + \varepsilon_t, \text{ for a fixed quantile } \tau \in (0, 1), \quad (8)$$

where  $\underline{Z}_{t-1} = (\underline{X}'_{t-1}, \underline{Y}'_{t-1})'$  is a  $d \equiv d_1 + d_2$  dimensional random vector,  $q_\tau(\underline{Z}_{t-1}) = Q^\tau(X_t | \underline{Z}_{t-1})$ , and  $\varepsilon_t = X_t - Q^\tau(X_t | \underline{Z}_{t-1})$ . Thus,  $\varepsilon_t$  is an error term with its  $\tau$ -th conditional quantile given  $\underline{Z}_{t-1}$  equal to zero, i.e.  $Pr(\varepsilon_t \leq 0 | \underline{Z}_{t-1} = \underline{z}) = \tau$  for almost all  $\underline{z}$ . To quantify the degree of Granger causality in quantile from  $Y$  to  $X$ , we further consider the following nonparametric constrained quantile regression:

$$X_t = \bar{q}_\tau(\underline{X}_{t-1}) + \bar{\varepsilon}_t, \quad (9)$$

for  $\bar{q}_\tau(\underline{X}_{t-1}) = Q^\tau(X_t | \underline{X}_{t-1})$ ,  $\bar{\varepsilon}_t = X_t - Q^\tau(X_t | \underline{X}_{t-1})$ , with  $Pr(\bar{\varepsilon}_t \leq 0 | \underline{X}_{t-1} = \underline{x}) = \tau$  for almost all  $\underline{x}$ .



From Equation (7) and after replacing the unconstrained and constrained conditional quantiles of  $X_t$  by the corresponding regression functions  $q_\tau(\underline{Z}_{t-1})$  and  $\bar{q}_\tau(\underline{X}_{t-1})$ , respectively, we obtain the following nonparametric regression-based measures of Granger causality in quantiles from  $Y$  to  $X$ .

**Proposition 1** *Under the quantile regressions (8) and (9), for non-negative weighting function  $w(\cdot)$  and a fixed quantile  $\tau$ , for  $0 < \tau < 1$ , the function*

$$C_\tau(Y \rightarrow X | I) = \ln \left[ \frac{E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})]}{E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})]} \right] \quad (10)$$

is the  $\tau$ -th quantile causality measure from  $Y$  to  $X$ , given  $I$ .

The inclusion of non-negative weighting function  $w(\cdot)$  in the expression of the above measure is to avoid its highly uncertain estimation especially in regions with sparse or noisy data. In practice,  $w(\cdot)$  will typically take the value that is one in the center of the support of  $\underline{Z}$  and zero near the boundary. The introduction of the weight  $w(\cdot)$  also permits the users to focus their analysis on a given range of interest of the variables. Furthermore, notice that when  $q_\tau(\underline{Z}_{t-1}) = \bar{q}_\tau(\underline{X}_{t-1})$  with probability one (i.e., when there is no causality from  $Y$  to  $X$  through its  $\tau$ -th quantile), the measure  $C_\tau(Y \rightarrow X | I)$  becomes equal to zero.

Next, we provide a consistent nonparametric estimator for the above measure by replacing the unknown expectations of check loss functions by their consistent nonparametric estimates. In particular, we need nonparametric estimates for the following restricted and unrestricted quantile regression errors:  $X_t - \bar{q}_\tau(\underline{X}_{t-1})$  and  $X_t - q_\tau(\underline{Z}_{t-1})$ , respectively. Due to its well-known advantages, we use the local polynomial approach to construct our nonparametric estimates as discussed in Fan and Gijbels (1996). For simplicity of exposition and to economize space, hereafter, we omit the information sets  $\underline{Z}_{t-1}$  and  $\underline{X}_{t-1}$  in the notations of subsequent Granger causality measures and their corresponding estimates.

First of all, assume that the constrained [resp. unconstrained]  $\tau$ -th conditional quantile function  $\bar{q}_\tau(\underline{x})$  [resp.  $q_\tau(\underline{z})$ ] is differentiable continuously at  $\underline{x} = (x_1, \dots, x_{d_1})'$  [resp.  $\underline{z} = (\underline{x}', \underline{y}')$ , with  $\underline{y} = (y_1, \dots, y_{d_2})'$ ] up to order  $p + 1$  [resp.  $q + 1$ ]. Then, the multivariate  $p$ -th order local polynomial approximation of the constrained  $\tau$ -th conditional quantile function  $\bar{q}_\tau(\underline{a})$ , for any  $\underline{a}$  close to  $\underline{x}$ , is given by

$$\bar{q}_\tau(\underline{a}) \approx \sum_{0 \leq |\underline{r}| \leq p} \frac{1}{\underline{r}!} D^{\underline{r}} \bar{q}_\tau(\underline{x}) (\underline{a} - \underline{x})^{\underline{r}},$$

where  $\underline{r} = (r_1, \dots, r_{d_1})$ ,  $|\underline{r}| = \sum_{i=1}^{d_1} r_i$ ,  $\underline{r}! = r_1! \times \dots \times r_{d_1}!$ , and

$$D^{\underline{r}} \bar{q}_\tau(\underline{x}) = \frac{\partial^{\underline{r}} \bar{q}_\tau(\underline{x})}{\partial x_1^{r_1} \dots \partial x_{d_1}^{r_{d_1}}}$$

for  $\underline{x}^{\underline{r}} = x_1^{r_1} \times \dots \times x_{d_1}^{r_{d_1}}$  and  $\sum_{0 \leq |\underline{r}| \leq p} = \sum_{j=0}^p \underbrace{\sum_{r_1=0}^j \dots \sum_{r_{d_1}=0}^j}_{r_1 + \dots + r_{d_1} = j}$ .

Now, let  $K(\underline{u})$  be a multivariate product density function on  $\mathbb{R}^{d_1}$  and  $h_1 \equiv h_{1T} \in \mathbb{R}^+$  a bandwidth parameter converging to zero at appropriate rates specified in Assumption **A.9** of Appendix A of the Online Technical Appendix. Using the sample  $\{(X_t, Y_t)\}_{t=0}^T$ , we consider minimizing the following quantity with respect to  $\beta_{\underline{r}}$ ,  $0 \leq |\underline{r}| \leq p$ , to derive the constrained estimator  $\widehat{\bar{q}}_{\tau}(\underline{x})$  evaluated at the data point  $\underline{x} = \underline{X}_{t-1}$ :

$$\min_{\beta_{\underline{r}}} \sum_{s=1, s \neq t}^T K_{h_1}(\underline{X}_{s-1} - \underline{X}_{t-1}) \rho_{\tau} \left( X_s - \sum_{0 \leq |\underline{r}| \leq p} \beta_{\underline{r}} (\underline{X}_{s-1} - \underline{X}_{t-1})^{\underline{r}} \right), \quad (11)$$

where  $K_{h_1}(\underline{u}) = K(\underline{u}/h_1)/h_1^{d_1}$ . Let  $\hat{\beta}_{\underline{r}}(\underline{X}_{t-1})$ , for  $0 \leq |\underline{r}| \leq p$ , be the minimizer of the above optimization problem. The leave-one-out estimator for the  $\tau$ -th constrained conditional quantile function  $\bar{q}_{\tau}(\underline{x})$  and its partial derivatives  $D^{\underline{r}} \bar{q}_{\tau}(\underline{x})$  evaluated at  $\underline{x} = \underline{X}_{t-1}$  are then given by

$$\widehat{\bar{q}}_{-t, \tau}(\underline{X}_{t-1}) = \hat{\beta}_0 \quad \text{and} \quad \widehat{D}^{\underline{r}} \bar{q}_{-t, \tau}(\underline{X}_{t-1}) = \underline{r}! \hat{\beta}_{\underline{r}}, \quad \text{for } 1 \leq |\underline{r}| \leq p.$$

Similarly, we denote the multivariate  $q$ -th order local polynomial leave-one-out estimator of the unconstrained  $\tau$ -th conditional quantile function  $q_{\tau}(\underline{z})$  and its derivatives  $D^{\underline{r}} q_{\tau}(\underline{z})$  evaluated at the data point  $\underline{z} = \underline{Z}_{t-1}$  by  $\widehat{q}_{-t, \tau}(\underline{Z}_{t-1})$  and  $\widehat{D}^{\underline{r}} q_{-t, \tau}(\underline{Z}_{t-1})$ , for  $1 \leq |\underline{r}| \leq q$ , respectively. Therefore, based on the above local polynomial leave-one-out estimators  $\widehat{\bar{q}}_{-t, \tau}(\underline{X}_{t-1})$  and  $\widehat{q}_{-t, \tau}(\underline{Z}_{t-1})$ , the term

$$C_{\tau}(\widehat{Y \rightarrow X}) = \ln \left( \frac{T^{-1} \sum_{t=1}^T \rho_{\tau} \left( X_t - \widehat{\bar{q}}_{-t, \tau}(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1})}{T^{-1} \sum_{t=1}^T \rho_{\tau} \left( X_t - \widehat{q}_{-t, \tau}(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1})} \right) \quad (12)$$

is a suitable nonparametric estimator of the causality measure  $C_{\tau}(Y \rightarrow X)$  in (10). As we will show in the proof of Theorem 2, using leave-one-out estimators helps to reduce the bias in estimating the measure  $C_{\tau}(Y \rightarrow X)$ . The consistency of the estimator in (12) will be established for any continuous weight function  $w(\cdot)$  defined on a compact and non-empty interior support. Finally, the nonparametric estimator of the measure of Granger causality from  $X$  to  $Y$  through its  $\tau$ -th quantile can be obtained following the same steps. Hereafter, we will omit the subscript “ $-t$ ” of the leave-one-out estimators  $\widehat{\bar{q}}_{-t, \tau}(\underline{X}_{t-1})$  and  $\widehat{q}_{-t, \tau}(\underline{Z}_{t-1})$  in the expression (12) to minimize the notation.

## 5 Asymptotic properties

This section aims to establish the asymptotic properties of nonparametric estimator  $C_{\tau}(\widehat{Y \rightarrow X})$  defined in Equation (12). Before proceeding with this, we need to introduce some notations to facilitate the study of local polynomial estimators. Following Kong et al. (2010), let  $N_i = \binom{i+d_1-1}{d_1-1}$  be the number of distinct  $d_1$ -tuples  $\underline{r}$  with  $|\underline{r}| = i$ . Arrange these  $N_i$   $d_1$ -tuples as a sequence in a lexicographical order, with the highest priority given to the last position so that  $(0, \dots, 0, i)$  is the first element in the sequence and  $(i, 0, \dots, 0)$  is the last element. Let  $\pi_i$  denote the 1-to-1 mapping defined by  $\pi_i(1) = (0, \dots, 0, i), \dots, \pi_{N_i} = (i, 0, \dots, 0)$ .

Furthermore, for each  $i = 1, \dots, p$ , define a  $N_i \times 1$  vector  $\mu_i(\underline{x})$  with its  $k$ -th element given by  $\underline{x}^{\pi_i(k)}$ , and write  $\mu(\underline{x}) = (1, \mu_1(\underline{x})', \dots, \mu_p(\underline{x})')'$ , which is a column vector of length  $N = \sum_{i=0}^p N_i$ . Similarly, define vectors  $\beta_p(\underline{x})$  and  $\underline{\beta}$  through the same lexicographical arrangement of  $D^{\underline{r}}\bar{q}_\tau(\underline{x})$  and  $\beta_{\underline{r}}$  in Equation (11) for  $0 \leq |\underline{r}| \leq p$ . Thus, Equation (11) can be rewritten as

$$\min_{\underline{\beta}} \sum_{s=1, s \neq t}^T K_{h_1}(\underline{X}_{s-1} - \underline{X}_{t-1}) \rho_\tau \left( X_s - \mu(\underline{X}_{s-1} - \underline{X}_{t-1})' \underline{\beta} \right). \quad (13)$$

Now, let us denote the minimizer of equation (13) by  $\hat{\beta}_T(\underline{X}_{t-1})$ , and let  $\hat{\beta}_p(\underline{X}_{t-1}) = W_p \hat{\beta}_T(\underline{X}_{t-1})$ , where  $W_p$  is a diagonal matrix with diagonal entries that are equal to the lexicographical arrangement of  $\underline{r}!$ , for  $0 \leq |\underline{r}| \leq p$ . In addition, let  $\varphi(u) = \tau 1(u \geq 0) + (\tau - 1)1(u < 0)$  be the piecewise constant derivative of the check loss function  $\rho_\tau(\varepsilon)$  given in (5). Define  $G(\theta, \underline{x}) = E[\varphi(X_t - \theta) | \underline{X}_{t-1} = \underline{x}]$  and  $g(\underline{x}) = \partial G(\theta, \underline{x}) / \partial \theta$ . Then, it holds true that  $g(\underline{x}) = -f_{\bar{\varepsilon} | \underline{X}}(0 | \underline{x})$ , where  $f_{\bar{\varepsilon} | \underline{X}}(\cdot | \underline{x})$  is the conditional probability density function of the constrained quantile error  $\bar{\varepsilon}_t = X_t - \bar{q}_\tau(\underline{X}_{t-1})$  given  $\underline{X}_{t-1} = \underline{x}$ . Let  $\nu_{\underline{i}} = \int K(\underline{u}) \underline{u}^{\underline{i}} d\underline{u}$  and define  $\nu_{T, \underline{i}} = \int K(\underline{u}) \underline{u}^{\underline{i}} g(\underline{x} + h_1 \underline{u}) f_{\underline{X}}(\underline{x} + h_1 \underline{u}) d\underline{u}$ , where  $f_{\underline{X}}(\cdot)$  is the marginal probability density function of  $\underline{X}_{t-1}$ . Finally, for  $0 \leq j, k \leq p$ , let  $S_{j,k}$  and  $S_{T,j,k}(\underline{x})$  be two  $N_j \times N_k$  matrices with their  $(l, m)$  elements, respectively,

$$[S_{j,k}]_{l,m} = \nu_{\pi_j(l) + \pi_k(m)} \quad \text{and} \quad [S_{T,j,k}(\underline{x})]_{l,m} = \nu_{T, \pi_j(l) + \pi_k(m)}(\underline{x}).$$

Using the above notations, we now define the  $N \times N$  matrices  $S_p$  and  $S_{T,p}(\underline{x})$ :

$$S_p = \begin{bmatrix} S_{0,0} & S_{0,1} & \cdots & S_{0,p} \\ S_{1,0} & S_{1,1} & \cdots & S_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p,0} & S_{p,1} & \cdots & S_{p,p} \end{bmatrix}, \quad S_{T,p}(\underline{x}) = \begin{bmatrix} S_{T,0,0}(\underline{x}) & S_{T,0,1}(\underline{x}) & \cdots & S_{T,0,p}(\underline{x}) \\ S_{T,1,0}(\underline{x}) & S_{T,1,1}(\underline{x}) & \cdots & S_{T,1,p}(\underline{x}) \\ \vdots & \vdots & \ddots & \vdots \\ S_{T,p,0}(\underline{x}) & S_{T,p,1}(\underline{x}) & \cdots & S_{T,p,p}(\underline{x}) \end{bmatrix}.$$

Further, for  $|S_p| \neq 0$ , define

$$\beta_T^*(\underline{X}_{t-1}) = -\frac{1}{T h_1^{d_1}} W_p S_{T,p}^{-1}(\underline{x}) H_T^{-1} \sum_{s=1, s \neq t}^T K_{h_1}(\underline{X}_{s-1} - \underline{X}_{t-1}) \varphi(X_s - \mu(\underline{X}_{s-1} - \underline{X}_{t-1})' W_p^{-1} \beta_p(\underline{X}_{t-1})) \mu(\underline{X}_{s-1} - \underline{X}_{t-1}),$$

where  $H_T$  is a diagonal matrix with diagonal entries  $h_1^{|\underline{r}|}$ , for  $0 \leq |\underline{r}| \leq p$ , in the aforementioned lexicographical order. The quantity  $\beta_T^*(\underline{X}_{t-1})$  is the leading term in the Bahadur representation of  $\hat{\beta}_p(\underline{X}_{t-1}) - \beta_p(\underline{X}_{t-1})$ , which is the sum of a bias term  $E_s[\beta_T^*(\underline{X}_{t-1})]$  (with  $E_s$  denoting expectation with respect to  $s$ ) and a stochastic term  $\beta_T^*(\underline{X}_{t-1}) - E_s[\beta_T^*(\underline{X}_{t-1})]$ ; for more details, among other studies, refer to Kong et al. (2010), Guerre and Sabbah (2012), and Noh et al. (2013).

The  $q$ -th order local polynomial leave-one-out estimator for the unconstrained conditional quantile function  $q_\tau(\underline{Z}_{t-1})$  using a second bandwidth  $h_2$ , denoted by  $\hat{q}_\tau(\underline{Z}_{t-1})$ , can accordingly be defined as above for  $\hat{\bar{q}}_T(\underline{X}_{t-1})$ . Therefore, we omit the steps for constructing  $\hat{q}_\tau(\underline{Z}_{t-1})$ .

## 5.1 Bahadur representation

In this subsection, we establish a Bahadur type representation for the nonparametric estimator in Equation (12). This representation will be used to deduce the consistency of the estimator in (12). To derive the Bahadur's representation of  $C_\tau(\widehat{X \rightarrow Y})$ , some regularity assumptions are needed. Therefore, we consider a set of standard assumptions that have been widely used in the literature on nonparametric estimation and inference; refer to, for example, Hong (2003), Guerre et al. (2012) and Noh et al. (2013) for independent and identically distributed (i.i.d.) data, and Kong et al. (2010) for the strongly mixing data. For simplicity of exposition, we relegate all these assumptions to Appendix A of the Online Technical Appendix. The following theorem establishes the Bahadur representation for the estimator  $C_\tau(\widehat{X \rightarrow Y})$  [refer to the proof of Theorem 1 in Appendix B of the Online Technical Appendix].

**Theorem 1** *Let  $d = d_1 + d_2$ . Suppose Assumptions A.1-A.9 in Appendix A of the Online Technical Appendix hold,  $p > d_1/2 - 1$ ,  $h_1 = O(T^{-\kappa_1})$  with  $1/(2p + 2 + d_1) < \kappa_1 < 1/(2d_1)$ ,  $q > d/2 - 1$ , and  $h_2 = O(T^{-\kappa_2})$  with  $1/(2p + 2 + d) < \kappa_2 < 1/(2d)$ . Then, for each given quantile  $\tau \in (0, 1)$ , we have*

$$\sqrt{T} \left( C_\tau(\widehat{Y \rightarrow X}) - C_\tau(Y \rightarrow X) \right) = (1 + C_\tau(Y \rightarrow X)) \frac{1}{\sqrt{T}} \sum_{t=1}^T (e_t - u_t) + o_p(1),$$

where

$$e_t = \frac{\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1})) w(\underline{Z}_{t-1}) - E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1})) w(\underline{Z}_{t-1})]}{E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1})) w(\underline{Z}_{t-1})]},$$

and

$$u_t = \frac{\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1})) w(\underline{Z}_{t-1}) - E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1})) w(\underline{Z}_{t-1})]}{E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1})) w(\underline{Z}_{t-1})]}.$$

Note that the conditions  $p > d_1/2 - 1$  and  $q > d/2 - 1$  are needed to obtain the asymptotic representations of the empirical analogue of the check loss functions involving  $\widehat{\bar{q}}_\tau(\cdot)$  and  $\widehat{q}_\tau(\cdot)$ , respectively; refer to Lemmas 3 and 4 in Appendix C of the Online Technical Appendix. They imply that the orders  $p$  and  $q$  of the local polynomial approximations of  $\bar{q}_\tau(x)$  and  $q_\tau(z)$  should increase as the dimensions  $d_1$  and  $d$  of  $\underline{X}_{t-1}$  and  $\underline{Z}_{t-1}$  increase, respectively. One immediate implication of Theorem 1 is that the estimator  $C_\tau(\widehat{Y \rightarrow X})$  in (12) is consistent. Thus, the following proposition can be straightforwardly deduced from the above Bahadur's representation of  $C_\tau(\widehat{Y \rightarrow X})$ , therefore its proof has been omitted.

**Proposition 2** *Under Assumptions A.1-A.9 in Appendix A of the Online Technical Appendix, for each given quantile  $\tau \in (0, 1)$ , the nonparametric estimator  $C_\tau(\widehat{Y \rightarrow X})$  in (12) converges in probability to the true Granger causality measure  $C_\tau(Y \rightarrow X)$  in (10).*

Besides the consistency property, the Bahadur representation implies  $\sqrt{T} \left( C_\tau(\widehat{Y \rightarrow X}) - C_\tau(Y \rightarrow X) \right)$  is asymptotically normal with mean zero and variance  $\sigma_{1\tau}^2 := (1 + C_\tau(Y \rightarrow X))^2 \Sigma_\tau$ , where

$$\Sigma_\tau = \lim_{T \rightarrow \infty} T^{-1} \text{Var} \left( \sum_{t=1}^T (e_t - u_t) \right) = \text{Var}(e_1 - u_1) + 2 \sum_{t>1} \text{Cov}(e_1 - u_1, e_t - u_t).$$

Specifically, the asymptotic normality of  $C_\tau(\widehat{Y \rightarrow X})$  can be established by applying the central limit theorem to the strongly mixing process  $\{(e_t - u_t)\}_{t=1}^T$ ; refer to, for example, Theorem 2.21 in Fan and Yao (2003). One direct application of this theorem is the construction of confidence intervals for the causality measure  $C_\tau(Y \rightarrow X)$  once a consistent estimator for  $\sigma_{1\tau}^2$  is available. For example,  $\hat{\sigma}_{1\tau}^2 = \left(1 + C_\tau(\widehat{Y \rightarrow X})\right)^2 \hat{\Sigma}_\tau$ , where  $\hat{\Sigma}_\tau$  is the long-run variance estimator (“Newey–West” or HAC estimator) based on sequence  $\{(\hat{e}_t - \hat{u}_t)\}_{t=1}^T$  with  $\hat{e}_t$  and  $\hat{u}_t$  respectively the sample analogues for  $e_t$  and  $u_t$  in Theorem 1. However, recall that by construction  $C_\tau(Y \rightarrow X)$  is non-negative and it is zero if and only if there is no causality from  $Y$  to  $X$  through its  $\tau$ -th quantile. It is immediate to observe that when  $C_\tau(Y \rightarrow X) = 0$ , the asymptotic variance  $\sigma_{1\tau}^2$ , given before, degenerates to zero. This means that the asymptotic normality result, which can be obtained using Theorem 1, is also a degenerate distribution and is meaningless except the consistency of  $C_\tau(\widehat{Y \rightarrow X})$  to zero. Thus, unlike the cases when the degree of Granger causality in quantiles is important (i.e. value of the causality measure is non-zero or large), we should herein investigate the next leading term in the previous Bahadur expansion in order to get a non-degenerated distributional result. We will study this important case in detail in the next section.

## 5.2 Inference

The causality measures, which we defined in the previous sections, can be used to test for Granger non-causality in quantiles between  $X$  and  $Y$ . If there is no causality from  $Y$  to  $X$  through its  $\tau$ -th quantile, then the restricted and unrestricted expectations of quantile check loss functions will be equal:  $E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))] = E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))]$ . Hereafter, we will use this equality to test the null hypothesis of Granger non-causality in certain quantiles. In other words, for a pre-specified quantile  $\tau$ , for  $0 < \tau < 1$ , we are interested in testing the null hypothesis

$$H_0 : C_\tau(Y \rightarrow X) = 0 \tag{14}$$

against the alternative hypothesis

$$H_1 : C_\tau(Y \rightarrow X) > 0. \tag{15}$$

As we had discussed at the end of the previous subsection, under  $H_0$ , the Bahadur representation in Theorem 1 tells us nothing about the asymptotic distribution of  $C_\tau(\widehat{Y \rightarrow X})$ . Therefore, to test  $H_0$  against  $H_1$  we need to derive a non-degenerated distribution of  $C_\tau(\widehat{Y \rightarrow X})$  when  $C_\tau(Y \rightarrow X) = 0$  is true. Using the theory for  $U$ -statistics, the following theorem provides non-degenerate asymptotic normality of the nonparametric estimator  $C_\tau(\widehat{Y \rightarrow X})$  at a given quantile  $\tau$ , for  $0 < \tau < 1$  [refer to the proof of Theorem 2 in Appendix B of the Online Technical Appendix]. Again, here, in this study, we focus only on Granger causality from  $Y$  to  $X$ , however, a similar result can be obtained for Granger causality from  $X$  to  $Y$ .

**Theorem 2** Let  $d = d_1 + d_2$ . Suppose Assumptions **A.1-A.9** in Appendix A of the Online Technical Appendix hold,  $p > d_1/2 - 1$  and  $h_1 = O(T^{-\kappa_1})$  with  $1/(2p + 2 + d_1) < \kappa_1 < 1/(2d_1)$ , and  $q > d/2 - 1$  and  $h_2 = O(T^{-\kappa_2})$  with  $1/(2q + 2 + d) < \kappa_2 < 1/(2d)$ . Then, under the null hypothesis  $H_0$  in (14), we have

$$Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) \xrightarrow{d} \mathcal{N}(0, \sigma_{0\tau}^2),$$

where

$$\sigma_{0\tau}^2 \equiv \frac{2\tau^2(1-\tau)^2}{\kappa(\tau)^2} \int K^2(\underline{u}) d\underline{u} \int \frac{w^2(\underline{z})}{f_{\varepsilon, \underline{Z}}^2(0, \underline{z})} f_{\underline{Z}}^2(\underline{z}) d\underline{z},$$

with  $\kappa(\tau) = E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})]$ ,  $f_{\varepsilon, \underline{Z}}(0, \underline{z})$  the joint density of  $\varepsilon_t = X_t - q_\tau(\underline{Z}_{t-1})$  and  $\underline{Z}_{t-1}$  evaluated at  $\varepsilon_t = 0$ , and  $f_{\underline{Z}}(\underline{z})$  the marginal density of  $\underline{Z}_{t-1}$ .

To implement the test in practice, we require a consistent estimator of  $\sigma_{0\tau}^2$ . In this study, we suggest to replace  $\sigma_{0\tau}^2$  by the following estimator:

$$\widehat{\sigma}_{0\tau}^2 = \frac{2\tau^2(1-\tau)^2}{\widehat{\kappa(\tau)}^2} \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \frac{w^2(\underline{Z}_{t-1})}{\widehat{f}_{\varepsilon, \underline{Z}}^2(0, \underline{Z}_{t-1})} \frac{1}{h_2^d} K^2\left(\frac{\underline{Z}_{t-1} - \underline{Z}_{s-1}}{h_2}\right), \quad (16)$$

where

$$\widehat{f}_{\varepsilon, \underline{Z}}(0, \underline{Z}_{t-1}) = \frac{1}{T-1} \sum_{s=1, s \neq t}^T \frac{1}{h_2^{d+1}} K\left(\frac{\widehat{\varepsilon}_{s-1}}{h_2}\right) K\left(\frac{\underline{Z}_{t-1} - \underline{Z}_{s-1}}{h_2}\right),$$

is the leave-one-out kernel density estimator for  $f_{\varepsilon, \underline{Z}}(0, \underline{Z}_{t-1})$ , and

$$\widehat{\kappa(\tau)} = \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \hat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})$$

is a consistent estimator of  $\kappa(\tau)$ .<sup>2</sup> It is worthwhile to remark that if we choose  $w(\underline{z}) = f_{\varepsilon, \underline{Z}}(0, \underline{z})$ , our expressions for  $\sigma_{0\tau}^2$  and  $\widehat{\sigma}_{0\tau}^2$  reduce to those for  $\sigma_0^2$  and  $\widehat{\sigma}_0^2$  in Jeong et al. (2012, Theorem 3.1(i) and (ii)) apart from the normalizing constant  $\kappa^2(\tau)$ . We now define the following feasible test statistic:

$$\widehat{\Gamma}_\tau := \frac{Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X})}{\widehat{\sigma}_{0\tau}}. \quad (17)$$

Under  $H_0$ , Theorem 2 implies that the test statistic  $\widehat{\Gamma}_\tau$  is asymptotically pivotal and asymptotically distributed as  $\mathcal{N}(0, 1)$ . This result forms the basis for the following one-sided asymptotic test for  $H_0$ : for a given significance level  $\alpha$ , we reject the null  $H_0$  if  $\widehat{\Gamma}_\tau > z_\alpha$ , where  $z_\alpha$  is the one-sided critical value, i.e. the upper  $\alpha$ -th percentile from the standard normal distribution. The following proposition establishes the consistency of the above test for the fixed alternative (15). It shows that the test statistic  $\widehat{\Gamma}_\tau = Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X})/\widehat{\sigma}_{0\tau}$  diverges to infinity under the alternative hypothesis  $H_1$  [refer to the proof of Proposition 3 in Appendix B of the Online Technical Appendix].

<sup>2</sup>The proof of the consistency of the estimator  $\widehat{\sigma}_{0\tau}^2$  for  $\sigma_{0\tau}^2$  can be found in the proof of Theorem 2 in Appendix B of the Online Technical Appendix.

**Proposition 3** *Let  $d = d_1 + d_2$ . Suppose Assumptions A.1-A.9 in Appendix A of the Online Technical Appendix hold,  $p > d_1/2 - 1$  and  $h_1 = O(T^{-\kappa_1})$  with  $1/(2p + 2 + d_1) < \kappa_1 < 1/(2d_1)$ , and  $q > d/2 - 1$  and  $h_2 = O(T^{-\kappa_2})$  with  $1/(2q + 2 + d) < \kappa_2 < 1/(2d)$ . Then, under the alternative hypothesis of causality in (15), we have*

$$Pr \left\{ Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) / \widehat{\sigma}_{0\tau} > B_T \right\} \rightarrow 1,$$

for any non-stochastic sequence  $B_T = o\left(Th_2^{d/2}\right)$ .

We also examine the asymptotic power of the test for detecting local departures, which converge to the null hypothesis (14) at a suitable rate. Specifically, we consider a sequence of Pitman-type local alternatives of the following form:

$$H_1(\delta_T) : q_\tau(\underline{z}) = \bar{q}_\tau(\underline{x}) + \delta_T \Delta_T(\underline{z}), \quad (18)$$

where  $\delta_T$  might depend on  $\tau$  and  $\delta_T \rightarrow 0$  as  $T \rightarrow \infty$ , and  $\Delta_T(\underline{z})$  is a non-constant continuous function in  $\underline{z}$ ; this indicates how large the deviation of  $q_\tau(\underline{z})$  is from  $\bar{q}_\tau(\underline{x})$ . Equivalently,  $\Delta_T(\underline{z})$  indicates how large the deviation of the measure  $C_\tau(Y \rightarrow X)$  is from zero. Further, we define the following term:

$$\gamma = \kappa(\tau)^{-1} \lim_{T \rightarrow \infty} E \left[ \Delta_T^2(\underline{Z}_{t-1}) w(\underline{Z}_{t-1}) f_{\varepsilon|\underline{Z}}(0|\underline{Z}_{t-1}) \right]. \quad (19)$$

We also assume  $\Delta_T(\underline{z})$  is a function that satisfies  $\gamma \neq 0$ . The following proposition states that our test has a non-trivial asymptotic local power against the sequence of Pitman local alternatives defined in (18), which converges to the null at the rate  $T^{-1/2}h_2^{-d/4}$  with  $d = d_1 + d_2$  [refer to the proof of Proposition 4 in Appendix B of the Online Technical Appendix].

**Proposition 4** *Let  $d = d_1 + d_2$ . Suppose Assumptions A.1-A.9 in Appendix A of the Online Technical Appendix hold,  $p > d_1/2 - 1$  and  $h_1 = O(T^{-\kappa_1})$  with  $1/(2p + 2 + d_1) < \kappa_1 < 1/(2d_1)$ , and  $q > d/2 - 1$  and  $h_2 = O(T^{-\kappa_2})$  with  $1/(2q + 2 + d) < \kappa_2 < 1/(2d)$ . Then, under the local alternatives  $H_1(\delta_T)$  in (18) with  $\delta_T = \left(Th_2^{d/2}\right)^{-1/2}$ , we have*

$$Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) \xrightarrow{d} \mathcal{N}(\gamma, \sigma_{0\tau}^2),$$

where  $\sigma_{0\tau}^2$  is defined in Theorem 2 and  $\gamma$  is defined in (19).

Note that our test can detect a class of local alternatives that converge to the null hypothesis at the same rate as Jeong et al. (2012), which is  $T^{-1/2}h_2^{-d/4}$ . This rate is a typical rate for tests based on kernel smoothing, see e.g. Härdle and Mammen (1993), Fan and Li (1996) and Zheng (1996) to mention only a few. On the other hand, Lavergne et al. (2015) propose a kernel-based test that smooths only over the covariates appearing under the null hypothesis, so that the rate of local alternatives for their test is  $T^{-1/2}h_2^{d_1/4}$ , which is faster than the rate  $T^{-1/2}h_2^{-(d_1+d_2)/4}$ . However, none of the aforementioned tests, including Lavergne et al. (2015)'s, can detect  $T^{-1/2}$ -local alternatives. Finally, to obtain tests that can detect  $T^{-1/2}$ -local alternatives,

global-type tests, which avoid nonparametric smoothing estimation under the alternative hypothesis, needs to be constructed, see e.g. Volgushev et al. (2013) for a stochastic process-based test for the significance of covariates in multivariate nonparametric quantile regression. Nonetheless, both Volgushev et al. (2013) and Lavergne et al. (2015) consider significance testing for i.i.d. data.

## 6 Bootstrap

The result in Theorem 2 is valid only asymptotically. The asymptotic normal distribution might not work well in the finite samples and our unreported simulations using asymptotic critical values also confirm this observation. Particularly, for high dimensional random variables the asymptotic test is subject to huge size distortion because of possible finite-sample bias in the nonparametric estimation due to the 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 that it can preserve the unknown dependence structure in the data; thus, it can “mimic” adequately the finite-sample distribution of our test statistic. Thus, in this section, we suggest a bootstrap-based procedure to improve the performance of our test in Theorem 2 in finite samples.

In the sequel,  $X \sim f_X$  means that the random variable  $X$  is generated from a density function  $f_X$ . Let  $L_1, L_2$  and  $L_3$  be three kernels and  $h^*$  be a bandwidth parameter. Hereafter, we discuss the implementation of the test based on local smoothed bootstrap. It is easy to implement in the following four steps<sup>3</sup>:

(1) We draw a bootstrapped sample  $\{(X_t^*, Y_t^*)\}_{t=1}^T$ . To do so, we first draw  $\underline{X}_{t-1}^*$  from its nonparametric (kernel) marginal probability density

$$\underline{X}_{t-1}^* \sim \frac{1}{Th^{*d_1}} \sum_{s=1}^T L_1 \left( \frac{X_{s-1} - x}{h^*} \right),$$

then conditional on  $\underline{X}_{t-1}^*$ , we draw  $X_t^*$  and  $\underline{Y}_{t-1}^*$  independently from the following two nonparametric (kernel) conditional probability densities:

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

and

$$\underline{Y}_{t-1}^* \sim \frac{1}{h^{*d_2}} \sum_{s=1}^T L_1 \left( \frac{X_{s-1} - X_{t-1}^*}{h^*} \right) L_3 \left( \frac{Y_{s-1} - y}{h^*} \right) / \sum_{s=1}^T L_1 \left( \frac{X_{s-1} - X_{t-1}^*}{h^*} \right);$$

(2) Based on the bootstrapped sample  $\{(X_t^*, Y_t^*)\}_{t=1}^T$ , we compute the bootstrapped version of the test statistic:  $\widehat{\Gamma}_\tau^* = \frac{Th_2^{d/2} C_\tau^*(\widehat{Y \rightarrow X})}{\widehat{\sigma}_{\partial\tau}^*}$ ;

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

<sup>3</sup>For details about how to generate the bootstrap replicates computationally, see Paparoditis and Politis (2000, Remark 2.1).



(4) Compute the bootstrapped  $p$ -value using  $p^* = B^{-1} \sum_{j=1}^B 1(\widehat{\Gamma}_{j,\tau}^* > \widehat{\Gamma}_\tau)$ , where  $\widehat{\Gamma}_\tau = \frac{Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X})}{\widehat{\sigma}_{0\tau}}$  is the test statistic based on the original sample  $\{(X_t, Y_t)\}_{t=1}^T$ , and for a given significance level  $\alpha$ , we reject the null hypothesis if  $p^* < \alpha$ .

In the above bootstrap-based procedure, we have taken the same bandwidth  $h^*$  for the nonparametric kernel estimators of the marginal density of  $\underline{X}_{t-1}$  and the conditional densities of  $X_t$  and  $\underline{Y}_{t-1}$  given  $\underline{X}_{t-1}$ . However, in principle, one can use different bandwidths without invalidating the local smoothed bootstrap. The next theorem establishes the asymptotic validity of the smoothed local bootstrap-based procedure [refer to the proof of Theorem 3 in Appendix B of the Online Technical Appendix].

**Theorem 3** *Suppose the assumptions in Theorem 2 and Assumption A.10 hold. Then, we have*

$$\widehat{\Gamma}_\tau^* := \frac{Th_2^{d/2} C_\tau^*(\widehat{Y \rightarrow X})}{\widehat{\sigma}_{0\tau}^*} \xrightarrow{d_*} \mathcal{N}(0, 1)$$

in probability induced by the bootstrap resampling conditional on the original sample  $\{(X_t, Y_t)\}_{t=1}^T$ , as  $T \rightarrow \infty$ , where  $\widehat{\sigma}_{0\tau}^{*2}$  is analogously defined as in Theorem 2.

The result in Theorem 3 provides an asymptotically valid approximation to the limiting null distribution of  $\widehat{\Gamma}_\tau$ . Regardless of whether the null hypothesis is true or not, the test statistic  $\widehat{\Gamma}_\tau^*$  will be normally distributed, but with different mean.

## 7 Extension: Additional control variables

In this section, we consider an additional control variable  $W$ , which can play the role of auxiliary variables and transmit a possible indirect causality between the variables of interest  $X$  and  $Y$ . As for the processes  $X$  and  $Y$ , let  $\{W_t \in \mathbb{R}, t = 0, \dots, T\}$  be a sample of strictly stationary stochastic process in  $\mathbb{R}$ , and denote  $\underline{W}_{t-1} = (W_{t-1}, \dots, W_{t-d_3})'$ , for some fixed known integer  $d_3 \geq 1$ . Hereafter, we focus on the estimation of the measure of Granger causality in quantiles from  $Y$  to  $X$  in the presence of  $W$ . We can, in a similar way, define an estimator of the measure of Granger causality in quantiles from  $X$  to  $Y$ , after controlling  $W$ .

To simplify the derivation of the main results in the presence of the auxiliary variable  $W$ , the following unconstrained partially linear quantile regression model is considered, i.e. for a fixed quantile  $\tau \in (0, 1)$ ,

$$X_t = \phi_\tau' \underline{W}_{t-1} + q_\tau(\underline{Z}_{t-1}) + \varepsilon_t, \quad (20)$$

where  $\underline{Z}_{t-1} = (\underline{X}'_{t-1}, \underline{Y}'_{t-1})'$  is a  $d \equiv d_1 + d_2$  dimensional vector that contains the past of  $X$  and  $Y$ ,  $\phi_\tau$  is a  $d_3$ -dimensional vector of unknown parameters, and  $\varepsilon_t$  is an error term with its  $\tau$ -th conditional quantile, conditional on  $(\underline{W}'_{t-1}, \underline{Z}'_{t-1})'$ , equal to zero. The quantile regression in (20) is a linear function of the past of  $W$  (i.e.  $\phi_\tau' \underline{W}_{t-1}$ ) and a nonlinear function of the past of  $X$  and  $Y$  (i.e.  $q_\tau(\underline{Z}_{t-1})$ ). It is assumed throughout

that the remainder of this section that  $\phi_\tau$  and  $q_\tau$  are identified uniquely. For sufficient conditions that guarantee the identification of the model (20), the readers can refer to Theorem 1 of Lee (2003). Partially linear quantile regression models are particularly useful because they compromise flexibly between fully parametric and fully nonparametric models to avoid the loss of precision due to the curse of dimensionality while making weaker assumptions about the functional form of the regression model.

Similarly, the constrained partially linear quantile regression model is given by

$$X_t = \overline{\phi}_\tau' \underline{W}_{t-1} + \overline{q}_\tau(\underline{X}_{t-1}) + \overline{\varepsilon}_t, \quad (21)$$

where  $\underline{X}_{t-1} = (X_{t-1}, \dots, X_{t-d_1})'$  is a  $d_1$ -dimensional vector that contains the past of  $X$ ,  $\overline{\phi}_\tau$  is a  $d_3$ -dimensional vector of unknown parameters, and  $\overline{\varepsilon}_t$  is an error term with its  $\tau$ -th conditional quantile, conditional on  $(\underline{W}'_{t-1}, \underline{X}'_{t-1})'$ , equal to zero. Note that the linear parts in the constrained and unconstrained partially linear quantile regressions are similar; thus, the only difference between the two quantile regressions is in the nonlinear parts that depend on different information sets.

Based on the constrained and unconstrained partially linear quantile regressions in (20) and (21), we define the following measure of Granger causality from  $Y$  to  $X$  through its  $\tau$ -th quantile, for  $\tau \in (0, 1)$ , after controlling for the presence of the auxiliary variable  $W$  that might be the source of indirect causality:

$$C_\tau^{PL}(Y \rightarrow X|W) = \ln \left( \frac{E \left[ \rho_\tau \left( X_t - \overline{\phi}_\tau' \underline{W}_{t-1} - \overline{q}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1}) \right]}{E \left[ \rho_\tau \left( X_t - \phi_\tau' \underline{W}_{t-1} - q_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) \right]} \right), \quad (22)$$

where the superscript  $PL$  stands for “partial linear”, and  $w(\cdot)$  is a user-chosen weighting function.

We here follow Lee’s (2003) and Sun’s (2005) approaches for the estimation of  $\phi_\tau$ ,  $q_\tau$ ,  $\overline{\phi}_\tau$ , and  $\overline{q}_\tau$ . For more details on the estimation of partially linear quantile regression, the reader can refer to Lee (2003) and Sun (2005). Using estimates of  $\phi_\tau$ ,  $q_\tau$ ,  $\overline{\phi}_\tau$ , and  $\overline{q}_\tau$ , the measure of causality in quantile given control variables  $W$  defined in Equation (22) can be consistently estimated using the following estimator:

$$C_\tau^{PL}(\widehat{Y} \rightarrow X|W) = \ln \left( \frac{\sum_{t=1}^T \rho_\tau \left( X_t - \widehat{\phi}_\tau' \underline{W}_{t-1} - \widehat{\overline{q}}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1})}{\sum_{t=1}^T \rho_\tau \left( X_t - \widehat{\phi}_\tau' \underline{W}_{t-1} - \widehat{q}_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1})} \right). \quad (23)$$

The following two theorems, respectively, provide the Bahadur representation and a non-degenerate asymptotic normal distribution for the nonparametric estimator  $C_\tau^{PL}(\widehat{Y} \rightarrow X|W)$  in (23), under the null hypothesis  $H_0 : C_\tau^{PL}(Y \rightarrow X|W) = 0$ . The proofs are similar to those of Theorems 1 and 2; hence, only sketched proofs of Theorems 4 and 5 below will be found in Appendix B of the Online Technical Appendix.

**Theorem 4** *Let  $d = d_1 + d_2$ . Suppose Assumptions **A.1-A.10** in Appendix A of the Online Technical Appendix hold,  $p > d_1/2 - 1$ ,  $h_1 = O(T^{-\kappa_1})$  with  $1/(2p + 2 + d_1) < \kappa_1 < 1/(2d_1)$ ,  $q > d/2 - 1$ , and*

$h_2 = O(T^{-\kappa_2})$  with  $1/(2p+2+d) < \kappa_2 < 1/(2d)$ . Then, for a given  $\tau \in (0, 1)$ , we have

$$\sqrt{T} \left( C_\tau^{PL}(\widehat{Y \rightarrow X|W}) - C_\tau^{PL}(Y \rightarrow X|W) \right) = (1 + C_\tau^{PL}(Y \rightarrow X|W)) \frac{1}{\sqrt{T}} \sum_{t=1}^T (e_t^{PL} - u_t^{PL}) + o_p(1),$$

where

$$e_t^{PL} = \frac{\rho_\tau \left( X_t - \bar{\phi}_\tau' W_{t-1} - \bar{q}_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) - E \left[ \rho_\tau \left( X_t - \bar{\phi}_\tau' W_{t-1} - \bar{q}_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) \right]}{E \left[ \rho_\tau \left( X_t - \bar{\phi}_\tau' W_{t-1} - \bar{q}_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) \right]},$$

and

$$u_t^{PL} = \frac{\rho_\tau \left( X_t - \phi_\tau' W_{t-1} - q_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) - E \left[ \rho_\tau \left( X_t - \phi_\tau' W_{t-1} - q_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) \right]}{E \left[ \rho_\tau \left( X_t - \phi_\tau' W_{t-1} - q_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) \right]}.$$

**Theorem 5** Let  $d = d_1 + d_2$ . Suppose Assumptions **A.1-A.10** in Appendix A of the Online Technical Appendix hold,  $p > d_1/2 - 1$ ,  $h_1 = O(T^{-\kappa_1})$  with  $1/(2p+2+d_1) < \kappa_1 < 1/(2d_1)$ ,  $q > d/2 - 1$ , and  $h_2 = O(T^{-\kappa_2})$  with  $1/(2p+2+d) < \kappa_2 < 1/(2d)$ . Then, under the null hypothesis  $H_0 : C_\tau^{PL}(Y \rightarrow X|W) = 0$ , for a given  $\tau \in (0, 1)$ , we have

$$Th_2^{d/2} C_\tau^{PL}(\widehat{Y \rightarrow X|W}) \xrightarrow{d} \mathcal{N}(0, \sigma_{0\tau}^{PL2}),$$

where

$$\sigma_{0\tau}^{PL2} \equiv \frac{2\tau^2(1-\tau)^2}{\kappa^{PL}(\tau)^2} \int K^2(\underline{u}) d\underline{u} \int \frac{w^2(\underline{z})}{f_{\varepsilon|\underline{Z}}^2(0|\underline{z})} d\underline{z},$$

with  $\kappa^{PL}(\tau) = E \left[ \rho_\tau \left( X_t - \phi_\tau' W_{t-1} - q_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) \right]$ ,  $f_{\varepsilon|\underline{Z}}(0|\underline{z})$  is the conditional density of  $\varepsilon_t := X_t - \phi_\tau' W_{t-1} - q_\tau(\underline{Z}_{t-1})$  given  $\underline{Z}_{t-1} = \underline{z}$  evaluated at  $\varepsilon_t = 0$ .

Note that  $\sigma_{0\tau}^{PL2}$  can be estimated consistently by  $\widehat{\sigma}_{0\tau}^{PL2}$  similar as (16). Lastly, as in the aforementioned partially linear quantile regressions, the linear parameter  $\phi_\tau$  [resp.  $\bar{\phi}_\tau$ ] can be estimated consistently at the parametric  $\sqrt{T}$ -rate, which is much faster than the convergence rate of the nonparametric part  $q_\tau$  [resp.  $\bar{q}_\tau$ ], the estimation effect of  $\widehat{\phi}_\tau$  [resp.  $\widehat{\bar{\phi}}_\tau$ ] is asymptotically negligible; hence, it is possible to derive the above results for the causality measure from  $Y$  to  $X$  replacing  $X_t$  by  $\tilde{X}_t = X_t - \widehat{\phi}_\tau' W_{t-1}$  [resp.  $X_t - \widehat{\bar{\phi}}_\tau' W_{t-1}$ ] as if  $\phi_\tau$  [resp.  $\bar{\phi}_\tau$ ] were known.

## 8 Monte Carlo simulations

In this section, we conduct a Monte Carlo simulation study to investigate the performance of the bootstrap-based test, which we suggested previously for testing the statistical significance of our measures of Granger causality in quantiles. Our primary interest is to evaluate the empirical size and power of the test in Theorem 3. Further, we will compare with size and power of Jeong et al.'s (2012) nonparametric test for testing the Granger non-causality in quantile.

All through 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 quantile from  $Y$  to  $X$ , i.e.  $H_0 : C_\tau(Y \rightarrow X) = 0$ , for a given quantile  $\tau \in (0, 1)$ . In the sequel,  $\eta_t$  and  $\varepsilon_t$  are two independent sequences of i.i.d. standard normal random variables.

## 8.1 Bootstrap-based Test

Though the asymptotic-based test  $\widehat{\Gamma}_\tau$  in (17), given by Theorem 2, is not time consuming and easy to implement, the empirical size of the test statistic  $\widehat{\Gamma}_\tau$  in small samples may differ significantly from that of the significance level. The size distortion is almost unavoidable for small samples, which has been confirmed in our unreported simulations. However, it is also known that some types of bootstrap such as smoothed local bootstrap or moving block bootstrap can help eliminate or mitigate the asymptotically negligible higher order terms, which may have a substantial and adverse effect on the size of  $\widehat{\Gamma}_\tau$  for small sample studies. Additional benefits of using smoothed local bootstrap test are that it can handle the unknown form of dependence in the data and - based on simulations - it is not very sensitive to changes in the bandwidth parameter. Thus, our primary interest is to evaluate the empirical size and power of the bootstrap-based test in Theorem 3 using the data generating processes (DGPs) presented in Table 1.

Before we discuss the DGPs in Table 1, we should recall the following. Assume that the variable  $X_t$  is a linear/nonlinear function of its own past,  $I_X(t-1)$ , and the past of  $Y$ ,  $I_Y(t-1)$ . Formally, consider the following mean regression equation:

$$X_t = \mu(I_X(t-1), I_Y(t-1)) + \varepsilon_t, \quad (24)$$

where  $\varepsilon_t$  is an error term that satisfies  $E[\varepsilon_t | I_X(t-1), I_Y(t-1)] = 0$ . The latter assumption implies that

$$E(X_t | I_X(t-1), I_Y(t-1)) = \mu(I_X(t-1), I_Y(t-1)).$$

Now, if we further assume that  $\varepsilon_t$  is a  $\mathcal{N}(0, \sigma^2)$ , then the  $\tau$ -th quantile of  $X_t$ , conditional on the past  $I_X(t-1) \cup I_Y(t-1)$ , will be given by

$$Q^\tau(X_t | I_X(t-1), I_Y(t-1)) = \mu(I_X(t-1), I_Y(t-1)) + \sigma \Phi^{-1}(\tau), \text{ for } \tau \in (0, 1), \quad (25)$$

where  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal  $\mathcal{N}(0, 1)$ .

Equations (24) and (25) show that under normality assumption of the error term  $\varepsilon_t$ , linear/nonlinear Granger (non-)causality in *mean* from  $Y$  to  $X$  is equivalent to linear/nonlinear Granger (non-)causality in *quantiles* from  $Y$  to  $X$ . Consequently, since the error terms in Table 1 are assumed to be normally distributed, (non-)causality in mean implies (non-)causality in quantiles. Thus, the eight DGPs in Table 1 will be used to evaluate the empirical size and power of the bootstrap-based test in Theorem 3. The last

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	$Y_t = 0.5Y_{t-1} + 0.5X_{t-1} + \varepsilon_t$	$X_t = 0.5X_{t-1} + \eta_t$	$X \rightarrow Y, Y \nrightarrow X$
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	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + 0.2Y_{t-1} + \sqrt{1 + Y_{t-1}^2}\eta_t$	$X \nrightarrow Y, Y \rightarrow X$
DGP P6	$Y_t = -0.3Y_{t-1} + \varepsilon_t$	$X_t = 0.65X_{t-1} + \sqrt{1 + Y_{t-1}^2}\eta_t$	$X \nrightarrow Y, Y \rightarrow X$

**Note:** This table summarizes the DGPs, which we consider in the simulation study, to investigate the properties (size and power) of nonparametric test of causality measures in quantiles. We simulate  $(Y_t, X_t)$ , for  $t = 1, \dots, T$ , under the assumption that  $(\varepsilon_t, \eta_t)'$  are i.i.d from  $\mathcal{N}(0, I_2)$  with  $I_2$  the  $2 \times 2$  identity matrix. 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.

column of this table summarizes the directions of (non-)causality in quantiles in those DGPs. Since in DGP S1 and DGP S2 the null hypothesis of  $Y$  Granger non-causing  $X$  in quantiles is satisfied, they are used to investigate the size property of our test. In particular, in DGP S1 neither  $X$  Granger causes  $Y$  nor  $Y$  Granger causes  $X$ , while in DGP S2  $X$  Granger causes  $Y$  but  $Y$  does not Granger cause  $X$ . DGP S2 is used to examine whether the size of our test for Granger causality of  $Y$  to  $X$  is affected by the Granger causality of  $X$  to  $Y$ . On the other hand, in DGP P1 to DGP P6 [the exception is when  $\tau = 0.50$ ; i.e.  $\Phi^{-1}(\tau) = 0$ , thus DGP P6 falls under the null], the null hypothesis is not satisfied, and therefore, these six DGPs serve the purpose of illustrating the power of our test. Further, notice that DGP P2 to DGP P6 are nonlinear and DGP P6 is taken from Example 1 in our motivating section. All DGPs under consideration are strictly stationary and ergodic processes.

For the estimation of nonparametric quantile regressions, consequent to the measures of Granger causality in quantiles, we use the nonparametric estimator defined in (13) with a polynomial order of  $p = 1$  (i.e. local linear estimator). The weight function  $w(\cdot)$  in the estimator of the measure in Equation (12) is set to be equal to one everywhere, i.e. the trivial weight function, given that the performance here is not depending heavily on the weight function. Further, to estimate the conditional restricted and unrestricted quantile regression functions, we take the univariate kernel function  $K(\cdot)$  equal to the standard normal density. For the multivariate case, we use the product kernel.

The bandwidths used to estimate the univariate (restricted) and bivariate (unrestricted) quantile regressions have the forms  $h_1 = cst_1 T^{-1/5}$  and  $h_2 = cst_2 T^{-1/6}$ , respectively, where  $cst_1$  and  $cst_2$  are selected using the cross-validation technique for quantile estimation, refer to Tong and Yao (2000) and Noh, et al. (2013) among others. For simplicity, the bandwidth  $h^*$ , used to generate the bootstrap samples, is assumed to take the same values as the above univariate bandwidth  $h_1$ . How to choose the bandwidths  $h_1$  and  $h_2$  to maximize our test's power is not yet investigated and requires more attention in future studies. Gao and Gijbels (2008) have proposed to use the Edgeworth expansion of the asymptotic distribution of the test in order to choose the bandwidth such that the power function of the test is maximized while the size function is controlled. Furthermore, three sample sizes  $T = 50, 100$  and  $200$  are considered. For each DGP, we first generate  $T + 300$  observations and then discard the first 300 observations to minimize the potential adverse effect of the initial values. We use 500 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% and we report the results for five different quantiles:  $\tau = 0.10, 0.25, 0.50, 0.75$  and  $0.90$ .

Table 2 reports the empirical size and power of the bootstrapped test statistic  $\widehat{\Gamma}_\tau^*$  in Theorem 3. As expected, the local bootstrap-based test controls its size for both small and moderate samples in DGP S1 in which non-causality of  $X$  and  $Y$  holds in both directions. However, the test is conservative in DGP S2 potentially due to the presence of Granger causality of  $X$  to  $Y$  while testing for the Granger causality of  $Y$  to

Table 2: Empirical size and power of our test

Quantiles	DGPs							
	DGP S1	DGP S2	DGP P1	DGP P2	DGP P3	DGP P4	DGP P5	DGP P6
$T = 50$								
$\tau = 0.10$	0.036	0.046	0.118	0.130	0.328	0.194	0.132	0.100
$\tau = 0.25$	0.036	0.024	0.118	0.130	0.328	0.194	0.132	0.100
$\tau = 0.50$	0.046	0.040	0.136	0.112	0.242	0.156	0.108	0.044
$\tau = 0.75$	0.048	0.028	0.112	0.122	0.264	0.144	0.128	0.114
$\tau = 0.90$	0.048	0.046	0.112	0.122	0.264	0.144	0.128	0.114
$T = 100$								
$\tau = 0.10$	0.044	0.030	0.334	0.332	0.672	0.416	0.282	0.230
$\tau = 0.25$	0.044	0.028	0.334	0.332	0.672	0.416	0.282	0.230
$\tau = 0.50$	0.052	0.018	0.312	0.358	0.720	0.474	0.230	0.048
$\tau = 0.75$	0.050	0.018	0.282	0.326	0.690	0.472	0.272	0.258
$\tau = 0.90$	0.050	0.018	0.282	0.326	0.690	0.472	0.272	0.258
$T = 200$								
$\tau = 0.10$	0.048	0.008	0.696	0.762	0.974	0.882	0.684	0.422
$\tau = 0.25$	0.048	0.018	0.696	0.762	0.974	0.882	0.684	0.422
$\tau = 0.50$	0.046	0.012	0.732	0.822	0.982	0.890	0.604	0.046
$\tau = 0.75$	0.054	0.022	0.670	0.708	0.980	0.876	0.658	0.454
$\tau = 0.90$	0.054	0.024	0.670	0.708	0.980	0.876	0.658	0.454

**Note:** This table reports the empirical size and the power of our local bootstrap-based test in Theorem 3 for testing the Granger non-causality in quantiles from  $Y$  to  $X$  at  $\alpha = 5\%$  significance level. Note that when  $\tau = 0.50$ , DGP P6 falls under the null. The number of simulations is equal to 500 and the number of bootstrap resamples is  $B = 199$ .

$X$ . Concerning the power, Table 2 shows that this test has a reasonable power against various alternatives. Its power is low for  $T = 50$ ; however, it rapidly increases as we increase the sample size to  $T = 100$  and  $200$ . Moreover, for extreme (upper and lower) quantiles the results are almost the same perhaps because very few observations are used given the sample sizes we considered. In general, given the small and moderate samples that are under consideration, the performance of local bootstrap-based test is satisfactory for the quantiles examined.

## 8.2 Comparison with Jeong et al.'s (2012) test

We saw in Section 5.2 that testing that the causality measure is equal to zero is equivalent to testing for Granger non-causality in quantile. Thus, the tests in Theorems 2 and 3 can be viewed as tests of Granger non-causality in quantile. Jeong et al. (2012) have recently proposed a nonparametric test for Granger causality in quantiles. For  $\beta$ -mixing processes and under the null hypothesis of non-causality in quantile, the asymptotic distribution of Jeong et al.'s (2012) test statistic is given by a standard normal distribution. However, Jeong et al. (2012) do not provide a bootstrap-based test as we have done in this study. It is worth mentioning that for finite samples, one might expect that our bootstrap-based test will perform better than Jeong et al.'s (2012) asymptotic-based test, which is justified only for the large samples only. In the following, for the sake of fair comparison, we investigate the performance of our local bootstrap-based test with Jeong et al.'s (2012) test implemented with the same local bootstrap procedure. But we have to remark that justifying validity of the local bootstrap procedure for their test statistic is not the main objective of this paper and thus we leave its formal justification for future study.

In this subsection, we consider an additional simulation exercise to compare the empirical size and power of our bootstrap-based test in Theorem 3 with those of Jeong et al.'s (2012) nonparametric test. We generate data according to the following model adapted from DGP P6 in Table 1:

$$X_t = 0.65X_{t-1} + \sqrt{1 + cY_{t-1}^2}\eta_t, \quad Y_t = -0.3Y_{t-1} + \varepsilon_t,$$

where  $c$  is a non-negative constant, and  $\varepsilon_t$  and  $\eta_t$  are independent standard normal random variables. Here,  $c = 0$  corresponds to the null hypothesis of non-causality from  $Y$  to  $X$ . Thus, the above DGP under  $c = 0$  is used to investigate the size property of the two tests. However, the null hypothesis is not satisfied whenever  $c > 0$ , and therefore the above GDP under  $c > 0$  serves to illustrate the power of the tests. It is worthwhile to notice that, when  $c > 0$ , there is no causality in the mean from  $Y$  to  $X$ , but there is causality in quantiles from  $Y$  to  $X$ . To investigate the power of both tests, we take  $c = 0.3, 0.5, 0.7, 0.9, 1.1$ , such that the higher  $c$  is the stronger the causality from  $Y$  to  $X$  is. The empirical size and power of both local bootstrap-based tests are computed using 500 replications and 199 bootstraps. To demonstrate the small or moderate sample improvement of bootstrap procedure, results for our asymptotic-based test and Jeong et al.'s (2012)



asymptotic-based test are also reported, which are based on 1000 replications and one-sided critical values from standard normal distribution. Furthermore, for the purpose of making fair comparison, for our test, we use the following univariate bandwidth  $h_1 = T^{-1/5}$  and bivariate bandwidths  $h_2 = T^{-1/6}$ , and for Jeong et al.'s (2012) test, the same univariate bandwidth  $h_1 = T^{-1/5}$  is adopted. Sample sizes  $T = 50, 100,$  and  $200$  are considered to facilitate the comparison of the finite-sample performance of the two tests. Finally, to save space, results for quantiles  $\tau = 0.25, 0.50,$  and  $0.75$  are provided.

Table 3 compares the empirical size and power of our bootstrap and asymptotic-based tests with those of bootstrap and asymptotic-based tests of Jeong et al.'s (2012) for testing the non-causality in quantiles from  $Y$  to  $X$  at  $\alpha = 5\%$  significance level. From this, we observe that under the null hypothesis [column  $c = 0$ ], there is a serious size distortion (oversized) for both our asymptotic-based test and the Jeong et al.'s (2012) asymptotic-based test (results in parentheses), while, as expected, both bootstrap-based tests are able to preserve the size well. This result is consistent across different levels of quantiles and sample sizes under examination. In addition, the power of Jeong et al.'s (2012) bootstrap-based tests is almost always dominated by the power of our bootstrap-based test, especially for slightly larger  $c$ 's and  $T = 100$  and  $200$ . For example, when  $c = 1.1, \tau = 0.75$  and  $T = 200$ , our bootstrap-based test has a power of 50%, i.e., five times higher than the power of the bootstrap-based test of Jeong et al.'s (2012). This illustrates the usefulness of using our local bootstrap-based test for finite sample studies when sample sizes are limited.

### 8.3 Bootstrap-based Bias Correction

In the previous sections, we have investigated the finite sample properties of the  $\widehat{\Gamma}_\tau$  test in Theorem 2. Here we provide additional simulation results to examine a bootstrap bias-corrected estimator of measure of Granger causality in quantile. Our estimator is motivated again from the local smoothed bootstrap proposed by Paparoditis and Politis (2000), which can conserve the conditional dependence structure in the data. The procedure is simple. We first use bootstrapped sample to compute the finite sample bias in the nonparametric estimator of quantile Granger causality measure. We then subtract the bias term to obtain a bootstrap bias-corrected estimate. The estimate is easy to obtain in the following four steps:

- (1) We draw a bootstrapped sample  $\{(X_t^*, Y_t^*)\}_{t=1}^T$  using the step (1) described in Section 6;
- (2) Based on the bootstrapped sample  $\{(X_t^*, Y_t^*)\}_{t=1}^T$ , we compute the bootstrapped version of the quantile Granger causality measure  $C_\tau^*(\widehat{Y \rightarrow X})$ ;
- (3) Repeat the steps (1)-(2)  $B$  times so that we get  $C_{j,\tau}^*(\widehat{Y \rightarrow X})$ , for  $j = 1, \dots, B$ ;
- (4) We approximate the bias term  $Bias = E[C_\tau(\widehat{Y \rightarrow X}) - C_\tau(Y \rightarrow X)]$  by the corresponding bootstrapped bias  $Bias^* = E^*[C_\tau^*(\widehat{Y \rightarrow X}) - C_\tau(\widehat{Y \rightarrow X})]$ , where  $E^*$  is the expectation based on the bootstrapped distribution of  $C_\tau^*(\widehat{Y \rightarrow X})$ , and  $C_\tau(\widehat{Y \rightarrow X})$  is the estimate of  $C_\tau(Y \rightarrow X)$  using the original sample.

Table 3: Empirical size and power of our test and Jeong et al.'s (2012) test

Quantiles	DGPs											
	$c = 0$		$c = 0.3$		$c = 0.5$		$c = 0.7$		$c = 0.9$		$c = 1.1$	
	ST	JHS	ST	JHS	ST	JHS	ST	JHS	ST	JHS	ST	JHS
$T = 50$												
$\tau = 0.25$	0.044	0.037	0.040	0.045	0.110	0.055	0.088	0.049	0.112	0.050	0.116	0.062
	(0.075)	(0.073)	(0.062)	(0.085)	(0.074)	(0.080)	(0.080)	(0.075)	(0.072)	0.080	(0.080)	(0.078)
$\tau = 0.50$	0.046	0.051	0.042	0.044	0.060	0.056	0.064	0.047	0.049	0.060	0.058	0.052
	(0.052)	(0.058)	(0.074)	(0.072)	(0.060)	(0.061)	(0.058)	(0.053)	(0.059)	(0.056)	(0.078)	(0.075)
$\tau = 0.75$	0.036	0.041	0.046	0.056	0.076	0.054	0.090	0.064	0.116	0.062	0.114	0.053
	(0.066)	(0.067)	(0.062)	(0.073)	(0.083)	(0.075)	(0.076)	(0.085)	(0.079)	(0.073)	(0.073)	(0.071)
$T = 100$												
$\tau = 0.25$	0.044	0.046	0.058	0.072	0.112	0.059	0.162	0.073	0.222	0.092	0.208	0.075
	(0.076)	(0.102)	(0.082)	(0.095)	(0.091)	(0.100)	(0.083)	(0.097)	(0.103)	(0.102)	(0.093)	(0.105)
$\tau = 0.50$	0.048	0.043	0.057	0.043	0.048	0.056	0.058	0.071	0.062	0.067	0.064	0.077
	(0.051)	(0.069)	(0.068)	(0.066)	(0.061)	(0.078)	(0.079)	(0.074)	(0.076)	(0.083)	(0.069)	(0.076)
$\tau = 0.75$	0.042	0.044	0.082	0.048	0.130	0.054	0.162	0.070	0.188	0.067	0.212	0.071
	(0.078)	(0.071)	(0.075)	(0.086)	(0.083)	(0.093)	(0.094)	(0.110)	(0.087)	(0.077)	(0.093)	(0.091)
$T = 200$												
$\tau = 0.25$	0.048	0.049	0.124	0.070	0.208	0.055	0.346	0.090	0.442	0.090	0.478	0.100
	(0.074)	(0.092)	(0.100)	(0.114)	(0.112)	(0.096)	(0.118)	(0.129)	(0.129)	(0.133)	(0.137)	(0.133)
$\tau = 0.50$	0.046	0.050	0.054	0.059	0.058	0.048	0.052	0.049	0.045	0.050	0.040	0.044
	(0.062)	(0.068)	(0.075)	(0.083)	(0.079)	(0.078)	(0.092)	(0.097)	(0.069)	(0.068)	(0.072)	(0.067)
$\tau = 0.75$	0.034	0.037	0.162	0.066	0.240	0.077	0.308	0.073	0.426	0.109	0.500	0.094
	(0.086)	(0.088)	(0.092)	(0.090)	(0.107)	(0.118)	(0.110)	(0.137)	(0.122)	(0.118)	(0.137)	(0.128)

**Note:** This table reports and compares the empirical size and power of our local bootstrap-based test [hereafter ST] with a local bootstrap-based test of Jeong et al.'s (2012) [hereafter JHS] for testing Granger non-causality in quantile from  $Y$  to  $X$  at  $\alpha = 5\%$  significance level. The table also provides the results for the two asymptotic-based tests: see the values between parentheses. Note that when  $\tau = 0.50$ , the considered DGPs fall under the null. The number of simulations is equal to 500 and the number of bootstrap resamples is  $B = 199$ .

This suggests the bias estimate

$$\widehat{Bias}^* = \frac{1}{B} \sum_{j=1}^B C_{j,\tau}^* \widehat{(Y \rightarrow X)} - C_{\tau} \widehat{(Y \rightarrow X)}.$$

Hence, a bootstrap bias-corrected estimator of measure of quantile Granger causality from  $Y$  to  $X$  at the  $\tau$ -th quantile is given by

$$C_{\tau,BC}^* \widehat{(Y \rightarrow X)} = C_{\tau} \widehat{(Y \rightarrow X)} - \widehat{Bias}^*. \quad (26)$$

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

$$C_{\tau,BC}^* \widehat{(Y \rightarrow X)} = \max \left\{ C_{\tau,BC}^* \widehat{(Y \rightarrow X)}, 0 \right\}.$$

Tables 1, 2, and 3 of the Online Technical Appendix report some additional simulation results for the bootstrap bias-corrected estimates of measures of Granger causality at quantiles  $\tau = 0.25, 0.50$  and  $0.75$ , for sample sizes  $T = 50, 100, 200$  and  $400$ . We consider the DGPs described in Table 1. The same kernel and bandwidths as in the previous sections are adopted. The bias terms and the average values of the bootstrap bias-corrected quantile causality measures are computed based on 500 simulations and  $B = 199$  bootstrap replications. From the three tables, we see that when there exist no Ganger causality from  $Y$  to  $X$  at the three considered quantiles for DGPs S1 and S2, the bootstrap bias-corrected estimates are close to zero and not statistically significant according to their standard deviations, while for DGPs P1-P6 these estimates are clearly different from zero and significantly positive judging by the associated standard deviations. Moreover, we find that the bootstrap bias-corrected estimates decrease as sample size increases for the considered DGPs, possibly due to the large finite sample bias in the nonparametric estimators.

## 9 Empirical application

This section aims to apply the causality measures, defined in the previous sections, to quantify and compare the predictability of quantiles of conditional distribution of stock returns using several economic and financial variables. The ultimate objective is to find which of these variables- financial or macro variables-help predict better the quantiles of stock returns.

The issue of predicting the conditional distribution of stock returns using quantile regressions has been the focus of many recent studies. For example, Chuang, Kuan, and Lin (2009) have investigated the causality (predictability) between stock return and volume based on parametric quantile regressions. Using a sup-Wald-type test for testing the Granger non-causality in all quantiles, they found that the causality from volume to return is usually heterogeneous across quantiles. In particular, the quantile causal effects of volume

on return exhibit a spectrum of (symmetric) V-shape relations so that the dispersion of return distribution increases with lagged volume. Cenesizoglu and Timmermann (2007) have also used a parametric quantile regression framework to look at whether a range of common predictor variables proposed in the finance and macro literature (e.g. Book to Market Ratio, Dividend Yield, Dividend Price Ratio, Inflation, T-bill, Stock Variance, Earnings Price Ratio, etc.) are helpful in predicting (Granger causing) specific quantiles of the stock return distribution. Empirically, they found little evidence to suggest that the center of the return distribution can be predicted. However, their main findings suggest that the tails of stock returns can be predicted by means of state variables proposed in the literature. Yang, Tu, and Zeng (2014) have applied parametric regression models to investigate the Granger causality in mean and quantiles between stock returns and exchange rate for nine Asian markets. Their empirical results show that there are more bi-directional causal relations based on the quantile regression than the conventional mean regression. Finally, Engle and Manganelli (2004) have used nonlinear parametric regressions to model the Value-at-Risk (VaR) of stock returns, which corresponds to predicting lower quantiles of the distribution of returns.

Thus, most existing works, focus on the predictability of quantiles of stock return distribution based on the parametric quantile regressions. In this section, we consider nonparametric predictability of quantiles of stock returns using several economic and financial variables. The nonparametric causality measures, defined in the previous sections, do not impose any restriction on the model linking the dependent variable (stock return) to the independent variable (financial/macroeconomic variables).

Our data consist of monthly aggregate S&P 500 index over the period of January 1990 to December 2014 and monthly Variance Risk Premium, Unemployment rate, Inflation, Effective Federal Funds rate, over the period January 1990 to December 2014. The variance risk premium is defined as the difference between the ex-ante risk neutral expectation of the future stock return variance and the expectation of stock return variance. Following Bollerslev et al. (2009) and Zhou (2010), we estimate the variance risk premium by the difference between the squared-Volatility Index (VIX) and lagged realized variance.

Our empirical analysis is based on the logarithmic return of the S&P 500 index and the macro variables that we transformed using the first-difference in log. Furthermore, the estimation and inference of measures of causality in quantiles are performed using the same kernel functions as in Section (8.1). Finally, again as in Section (8.1), the bandwidths used to estimate the univariate and bivariate quantile regressions have the forms  $h_1 = cst_1 T^{-1/5}$  and  $h_2 = cst_2 T^{-1/6}$ , respectively, where  $cst_1$  and  $cst_2$  are selected using the cross-validation technique for quantile estimation.

## 9.1 Empirical results

Table 4 presents the results of estimating the measures of Granger causality in quantiles from variance risk premium, unemployment rate, inflation, and effective federal funds rate to S&P 500 stock returns.

Table 4: Measures of causality (predictability) in quantiles of S&P 500 returns

Direction of Causality	Quantiles	Estimate of Causality Measure	<i>p</i> -value
<i>VRP</i> → <i>RP</i>			
	$\tau = 0.25$	3.825**	<b>0.012</b>
	$\tau = 0.50$	5.178**	<b>0.030</b>
	$\tau = 0.75$	3.717**	<b>0.038</b>
<i>UNEMP</i> → <i>RP</i>			
	$\tau = 0.25$	1.094	0.414
	$\tau = 0.50$	1.095	0.484
	$\tau = 0.75$	1.084	0.344
<i>INFL</i> → <i>RP</i>			
	$\tau = 0.25$	0.415	0.654
	$\tau = 0.50$	0.386	0.520
	$\tau = 0.75$	0.401	0.484
<i>EFF</i> → <i>RP</i>			
	$\tau = 0.25$	2.241**	<b>0.026</b>
	$\tau = 0.50$	2.175	0.106
	$\tau = 0.75$	2.125**	<b>0.038</b>

**Note:** This table reports the results of the estimation and inference for measures of Granger causality (predictability) in quantiles from variance risk premium (VRP), Unemployment rate (UNEMP), Inflation (INFL), Effective Federal Funds rate (EFF) to S&P 500 risk premium (RP). “\*\*\*” means the statistical significance at 5% significance level. The sample size is calculated from January 1990 to December 2014.

Furthermore, it reports the *p*-values for testing the statistical significance of the estimates of the measures using the bootstrap-based procedure introduced in Section 6.

Table 4 shows that the degree of predictability (causality) of quantiles of stock returns using variance risk premium is much higher compared to that of the other three variables. The degree is even higher at the median compared to the 0.25-th and 0.75-th quantiles of stock returns. Thus, the variance risk premium affects more the center of the distribution of stock returns. The estimates of the measures at different quantiles of stock returns are statistically significant at 5% significance level. Thereafter, the degree of predictability of quantiles of stock returns, using effective federal funds rate, is also high compared to those produced by unemployment rate and inflation. The estimates of the measures are statistically significant at 5% significance level for the 0.25-th and 0.75-th quantiles; however, not so for the median, which indicates

that the effective federal funds rate only affects the left and right tails of the distribution of stock returns, but not its center. Finally, it seems that the degrees of predictability of quantiles of stock returns using unemployment rate and inflation are weak and statistically insignificant. Finally, the above results show that only financial variables help predict quantiles of S&P 500 stock returns. On the contrary, the economic and statistical significances of the causal effect of macro variables on quantiles of stock returns are not significant.

## 10 Conclusion

We have considered measures of Granger causality in quantiles that are able to detect and quantify both linear and nonlinear causal effects between random variables. The measures are based on nonparametric quantile regressions and defined as logarithmic function of restricted and unrestricted expectations of quantile check loss functions. They can be easily and consistently estimated by replacing the unknown expectations of check loss functions by their nonparametric kernel estimates. We derived a Bahadur-type representation for the nonparametric estimator of the measures. We provided the asymptotic distribution of this estimator, which can be used to build tests for the statistical significance of the measures. We also examined the properties of these tests under certain local alternatives. Thereafter, we established the validity of a smoothed local bootstrap, which one can use in finite-sample settings, to perform statistical tests. A Monte Carlo simulation study revealed that the bootstrap-based test has a good finite-sample size and power properties for a variety of data-generating processes and different sample sizes.

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 quantile. Thus, our test can be viewed as a competitor of the existing nonparametric tests of Granger causality in quantile. There is only one nonparametric test of Granger causality in quantile, which is proposed by Jeong et al. (2012). We considered an additional simulation exercise to compare the empirical size and power of our test with those of Jeong et al.'s (2012) test. The simulation results indicate that our test controls the size and has a better power than Jeong et al.'s (2012) test. We have also suggested a bootstrap bias-corrected estimator for the quantile Granger causality measure and investigated its finite sample performance using Monte Carlo simulations.

Finally, the empirical importance of measuring Granger causality in quantiles was illustrated. We quantified the degree of predictability of quantiles of equity risk premium using the variance risk premium, unemployment rate, inflation, and the effective federal funds rate. The empirical results showed that the variance risk premium and effective federal funds rate have strong predictive power for predicting the quantiles of risk premium, compared to that of the predictive power of the other two macro variables. In particular, the variance risk premium is able to predict the center, the lower and upper quantiles of the distribution of risk premium, whereas the effective federal funds rate only predicts the lower and upper quantiles. However, the

unemployment rate and inflation have no effect on the quantiles of risk premium.

## References

- [1] Adrian, T., Boyarchenko, N., and Giannone, D. (2018), “Vulnerable Growth,” *American Economic Review*, forthcoming.
- [2] Bollerslev, T., Tauchen, G., and Zhou, H. (2009), “Expected Stock Returns and Variance Risk Premia,” *Review of Financial Studies*, 22, 4463–4492.
- [3] Bouezmarni, T., Rombouts, J.V.K., Taamouti, A. (2012). “Nonparametric Copula-based Test for Conditional Independence with Applications to Granger Causality”, *Journal of Business and Economic Statistics*, 30, 275–287.
- [4] Cai, Z. and Xu, X. (2008). “Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models,” *Journal of the American Statistical Association*, 103(484), 1595–1608.
- [5] Cenesizoglu, T., and Timmermann, A. (2007). “Predictability of Stock Returns: A Quantile Regression Approach, CIRANO Working paper.
- [6] Chaudhuri, P. (1991). “Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation,” *The Annals of statistics*, 19(2), 760–777.
- [7] Chuang, C.C., Kuan, C.M. and Lin, H.Y. (2009). Causality in Quantiles and Dynamic Stock Return–Volume Relations. *Journal of Banking & Finance*, 33(7), 1351–1360.
- [8] Dufour, J.-M., and Taamouti, A. (2010). “Short and Long Run Causality Measures: Theory and Inference,” *Journal of Econometrics*, 154, 42–58.
- [9] Engle, R.F. and Manganelli, S. (2004). “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles,” *Journal of Business & Economic Statistics*, 22, 367–381.
- [10] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, Chapman & Hall, London.
- [11] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.
- [12] Fan, Y. and Li, Q. (1996). “Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms,” *Econometrica*, 64, 865–890.

- [13] Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman & Hall, CRC.
- [14] Gao, J. and Gijbels, I. (2008). “Bandwidth Selection in Nonparametric Kernel Testing,” *Journal of the American Statistical Association*, 103, 1584–1594.
- [15] Geweke, J. (1982). “Measurement of Linear Dependence and Feedback Between Multiple Time Series,” *Journal of the American Statistical Association*, 77, 304–313.
- [16] Geweke, J. (1984). “Measures of Conditional Linear Dependence and Feedback Between Time Series,” *Journal of the American Statistical Association*, 79, 907–915.
- [17] Gouriéroux, C., Monfort, A. and Renault, E. (1987). “Kullback Causality Measures,” *Annales d’Économie et de Statistique*, 6, 369–410.
- [18] Granger, C. W. J. (1969). “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica*, 37, 424–459.
- [19] Guerre, E. and Sabbah, C. (2012). “Uniform Bias Study and Bahadur Representation for Local Polynomial Estimators of the Conditional Quantile Function,” *Econometric Theory*, 28, 87–129.
- [20] Hall, P. (1984). “Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators,” *Journal of Multivariate Analysis*, 14, 1–16.
- [21] Härdle, W.K. and Mammen, E. (1993). “Comparing Nonparametric Versus Parametric Regression Fits,” *The Annals of Statistics*, 21, 1926–1974.
- [22] Hong, S. (2003). “Bahadur Representation and Its Application for Local Polynomial Estimates in Nonparametric  $M$ -Regression,” *Journal of Nonparametric Statistics*, 15, 237–251.
- [23] 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.
- [24] Hu, M. and Liang, H. (2014). “A Copula Approach to Assessing Granger Causality,” *NeuroImage*, 100, 125–134.
- [25] Jeong, K., Härdle, W.K. and Song, S. (2012). “A Consistent Nonparametric Test for Causality in Quantile,” *Econometric Theory*, 28, 861–887.
- [26] Koenker, R, and Bassett, Jr G. (1978). “Regression Quantiles,” *Econometrica*, 1, 33–50.



- [27] Kong, E., Linton, O. and Xia, Y. (2010). “Uniform Bahadur Representation for Local Polynomial Estimates of  $M$ -regression and Its Application to the Additive Model”, *Econometric Theory*, 26, 1529–1564.
- [28] Lavergne, P., Maistre, S., and Patilea V. (2015). “A Significance Test for Covariates in Nonparametric Regression,” *Electronic Journal of Statistics*, 9, 643–678.
- [29] Lee, S. (2003). “Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model,” *Econometric Theory*, 19, 1–31.
- [30] Lee, T.-H. and Yang, W. (2014). “Granger-Causality in Quantiles Between Financial Markets: Using Copula Approach,” *International Review of Financial Analysis*, 33, 70–78.
- [31] 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.
- [32] 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.
- [33] Li, Y., Wei, H.L., Billings, S.A. and Liao, X.F. (2012). “Time-Varying Linear and Nonlinear Parametric Model for Granger Causality Analysis,” *Physical Review E*, 85(4), 041906.
- [34] Masry, E. (1996). “Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates”, *Journal of Time Series Analysis*, 17, 571–599.
- [35] Marinazzo, D., Pellicoro, M. and Stramaglia, S. (2008). “Kernel Method for Nonlinear Granger Causality,” *Physical review letters*, 100(14), 144103.
- [36] Noh, H., El Ghouch, A., Van Keilegom, I. (2013). “Quality of Fit Measures in the Framework of Quantile Regression”, *Scandinavian Journal of Statistics*, 40, 105–118.
- [37] 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.
- [38] 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.
- [39] Polasek, W. (2002). “Bayesian Causality Measures for Multiple ARCH Models Using Marginal Likelihoods,” *Technical report*, Institute of Statistics and Econometrics, University of Basel, Switzerland.

- [40] Song, X. and Taamouti, A. (2018). “Measuring Nonlinear Granger Causality in Mean,” *Journal of Business and Economics Statistics*, 36, 321–333.
- [41] Sun, Y. (2005). “Semiparametric Efficient Estimation of Partially Linear Quantile Regression Models,” *Annals of Economics and Finance*, 6, 105–127.
- [42] Su, L. and White, H.L. (2012). “Conditional Independence Specification Testing for Dependent Processes with Local Polynomial Quantile Regression. In Essays in Honor of Jerry Hausman (pp. 355-434). Emerald Group Publishing Limited.
- [43] Taamouti, A., Bouezmarni, T., and El Ghouch, A. (2014). “Nonparametric Estimation and Inference for Granger Causality Measures,” *Journal of Econometrics*, 180, 251–264.
- [44] Tong, H. and Yao, Q. (2000). “Nonparametric Estimation of Ratios of Noise to Signal in Stochastic Regression,” *Statistica Sinica*, 10, 751–770.
- [45] Troster, V. (2018). “Testing for Granger-Causality in Quantiles,” *Econometric Reviews*, 37(8), 850–866.
- [46] Volgushev, S., Birke, M., Dette, H., and Neumeier, N. (2013). “Significance Testing in Quantile Regression,” *Electronic Journal of Statistics*, 7, 105–145.
- [47] Wiener, N. (1956). “The Theory of Prediction,” in *E. F. Beckenback, ed., The Theory of Prediction*, McGraw-Hill, New York, chapter 8.
- [48] Yang, Z., Tu, A.H. and Zeng, Y. (2014). “Dynamic Linkages between Asian Stock Prices and Exchange Rates: New Evidence from Causality in Quantiles,” *Applied Economics*, 46, 1184–1201.
- [49] Yu, K. and Jones, M.C. (1998). “Local Linear Quantile Regression,” *Journal of the American statistical Association*, 93(441), 228–237.
- [50] Zheng, J.X. (1996). “A Consistent Test of Functional Form via Nonparametric Estimation Techniques,” *Journal of Econometrics*, 75, 263–289.
- [51] Zhou, H. (2010), “Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty,” *Federal Reserve Board Working Paper*.

# Measuring Granger Causality in Quantiles: Technical Appendix

February 22, 2020

This document provides the necessary assumptions and the proofs of the theoretical results in the main text, and some additional simulation results. Appendix A of this document reports a set of standard assumptions that have been widely used in the literature on nonparametric estimation and inference; see for example Kong et al. (2010) and Noh et al. (2013) among others. Appendix B contains the detailed proofs of the theoretical results developed in sections 5, 6 and 7 of the main text. In particular, it contains the proofs of theorems 1, 2, 3, 4, and 5 and propositions 3 and 4 of the main text. Appendix C provides four auxiliary lemmas which are useful to prove the results in sections 5-7 of the main text. Finally, Appendix D contains the tables of some additional simulation results that examine a bootstrap bias-corrected estimator of measure of Granger causality in quantiles.

## Appendix A: Assumptions

First of all, let  $\{(X_t, Y_t)\}$  be a jointly stationary process. Since we are interested in time series data, we need to specify the dependence in the processes of interest. In what follows, we define the mixing dependence that we consider in this paper. The stationary stochastic process  $\{(X_t, Y_t)\}$  is strongly mixing, with  $\gamma(k)$  its strong mixing coefficient, if

$$\gamma(k) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| \rightarrow 0 \text{ as } k \rightarrow \infty,$$

with  $\mathcal{F}_a^b = \sigma(\{(X_t, Y_t)\}_{t=a}^b)$ , where  $\sigma(\cdot)$  means the smallest sigma algebra. Furthermore, let  $V_x$  and  $V_z$  be two open convex sets in  $\mathbb{R}^{d_1}$  and  $\mathbb{R}^{d_1+d_2}$ , respectively. We now consider the following assumptions:

**A.1.** The processes  $\{(X_t, Y_t)\}$  are strongly mixing with mixing coefficients  $\gamma(k)$  satisfying

$$\sum_{k=1}^{\infty} k^\alpha [\gamma(k)]^{1-2/\nu_2} < \infty,$$

for some  $\nu_2 > 2$  and  $\alpha > \max\{(p + d_1 + 1)(1 - 2/\nu_2)/d_1, (q + d_1 + d_2 + 1)(1 - 2/\nu_2)/(d_1 + d_2)\}$ .

**A.2.** All partial derivatives of  $\bar{q}_\tau(\underline{x})$  up to order  $p + 1$  exist and are continuous for all  $\underline{x} \in V_x$ , and there exists a constant  $C_1 > 0$  such that  $|D^r \bar{q}_\tau(\underline{x})| \leq C_1$ , for all  $\underline{x} \in V_x$  and  $|r| = p + 1$ . All partial derivatives of  $q_\tau(\underline{z})$  up to order  $q + 1$  exist and are continuous for all  $\underline{z} \in V_z$ , and there exists a constant  $C_2 > 0$  such that  $|D^r q_\tau(\underline{z})| \leq C_2$ , for all  $\underline{z} \in V_z$  and  $|r| = q + 1$ .

**A.3.** The marginal density of  $\varepsilon_t = X_t - q_\tau(\underline{Z}_{t-1})$  is bounded and satisfies  $E(\varphi(\varepsilon_t) | \underline{Z}_{t-1}) = 0$ .

**A.4.** For all  $e$  in a neighbourhood of zero, the conditional density  $f_{\varepsilon | \underline{Z}_{-1}}(e | \underline{z})$  of  $\varepsilon_t = X_t - q_\tau(\underline{Z}_{t-1})$  given  $\underline{Z}_{t-1} = \underline{z}$  satisfies

$$\left| f_{\varepsilon | \underline{Z}_{-1}}(e | \underline{z}_1) - f_{\varepsilon | \underline{Z}_{-1}}(e | \underline{z}_2) \right| \leq K_e \|\underline{z}_1 - \underline{z}_2\|,$$

where  $K_e$  is a positive constant depending on  $e$ . Further, the conditional density is positive for  $e = 0$  for all values of  $\underline{z} \in V_z$ , and its first partial derivative with respect to  $e$ ,  $D^1 f_{\varepsilon | \underline{Z}_{-1}}(e | \underline{z})$ , is bounded for all  $\underline{z} \in V_z$  and  $e$  in a neighbourhood of zero.

**A.5.** The weight function  $w(\underline{z})$  is continuous, and its support  $\mathcal{D} \subset V_z$  is compact and has non-empty interior.

**A.6.** The kernel function  $K(\cdot)$  has a compact support and  $\left| H_{\underline{j}}(\underline{u}) - H_{\underline{j}}(\underline{v}) \right| \leq \|\underline{u} - \underline{v}\|$  for all  $\underline{j}$  with  $0 \leq \underline{j} \leq \max\{2p + 1, 2q + 1\}$ , where  $H_{\underline{j}}(\underline{u}) = \underline{u}^{\underline{j}} K(\underline{u})$ .

**A.7.** The probability density function of  $\underline{Z}_{t-1}$ ,  $f_{\underline{Z}}(\underline{z})$ , is positive and bounded with bounded first-order derivatives on  $V_z$ . The joint probability density of  $(\underline{Z}_0, \underline{Z}_l)$  satisfies  $f_{(\underline{Z}_0, \underline{Z}_l)}(\underline{u}, \underline{v}; l) \leq C < \infty$  for all  $l \geq 1$ .

**A.8.** The conditional density  $f_{\underline{Z}_{-1} | X}$  of  $\underline{Z}_{t-1}$  given  $X_t$  exists and is bounded. The conditional density function  $f_{(\underline{Z}_0, \underline{Z}_l) | (X_1, X_{l+1})}$  of  $(\underline{Z}_0, \underline{Z}_l)$  given  $(X_1, X_{l+1})$  exists and is bounded for all  $l \geq 1$ .

**A.9.** The bandwidth sequences  $h_1$  and  $h_2$  satisfy  $h_1 \rightarrow 0$ ,  $Th_1^{d_1+2(p+1)}/\log T = O(1)$ ,  $h_2 \rightarrow 0$ , and  $Th_2^{d_1+d_2+2(q+1)}/\log T = O(1)$  as  $T \rightarrow \infty$ . Furthermore, we assume  $Th_2^{2(d_1+d_2)}/(\log T)^3 \rightarrow \infty$ ,  $h_1 = o(h_2)$ , and  $h_2^{d_1+d_2} = o(h_1^{d_1})$ .

**A.10.** The bootstrap bandwidth  $h^*$  satisfies  $h^* \rightarrow 0$  and  $Th^*{}^{d_1+d_2+2(q+1)}/(\log T)^\lambda = O(1)$ , for some  $\lambda > 0$  as  $T \rightarrow \infty$ .

The assumptions presented here are frequently seen for nonparametric smoothing in multivariate time series analysis, see Masry (1996) and Kong et al. (2010). Assumptions A.1-A.2, A.6-A.8 and A.9 are standard. Assumptions A.4 and A.5 are required to derive the Bahadur representations in Lemmas 3-4 in Appendix C. Assumption A.10 is assumed to guarantee the consistency of the smoothed local bootstrap.

## Appendix B: Proofs of the main results

This section provides the proofs of the main theoretical results developed in sections 5, 6 and 7 of the main text.

**Proof of Theorem 1:** Theorem 1 can be proved by combing the first order Taylor expansion of  $C_\tau(\widehat{Y} \rightarrow X)$  around 1 (i.e. using  $\ln y \approx y-1$ ) and the asymptotic Bahadur representations in Lemmas 3 and 4 of Appendix C and with the equality  $\hat{a}/\hat{b} = a/b + \hat{b}^{-1}[(\hat{a} - a) - (\hat{b} - b)(a/b)]$ .  $\square$

**Proof of Theorem 2:** Note that for any  $x, y$ ,

$$\rho_\tau(x - y) - \rho_\tau(x) = (-y)\varphi(x) + 2(y - x)[1(y > x > 0) - 1(y < x < 0)].$$

Let  $\widehat{d}(\underline{x}) = \widehat{q}_\tau(\underline{x}) - \bar{q}_\tau(\underline{x})$  and  $\widehat{d}(\underline{z}) = \widehat{q}_\tau(\underline{z}) - q_\tau(\underline{z})$ . By straightforward calculation, under the null hypothesis of no causality, we obtain

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1}) - \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1}) \\ &= \frac{1}{T} \sum_{t=1}^T \left[ (\widehat{q}_\tau(\underline{Z}_{t-1}) - q_\tau(\underline{Z}_{t-1})) - (\widehat{q}_\tau(\underline{X}_{t-1}) - \bar{q}_\tau(\underline{X}_{t-1})) \right] w(\underline{Z}_{t-1})\varphi(\varepsilon_t) \\ & \quad + \frac{2}{T} \sum_{t=1}^T (X_t - \widehat{q}_\tau(\underline{Z}_{t-1})) \left\{ 1(\widehat{d}(\underline{Z}_{t-1}) > \varepsilon_{1t} > 0) - 1(\widehat{d}(\underline{Z}_{t-1}) < \varepsilon_t < 0) \right\} w(\underline{Z}_{t-1}) \\ & \quad - \frac{2}{T} \sum_{t=1}^T (X_t - \widehat{q}_\tau(\underline{X}_{t-1})) \left\{ 1(\widehat{d}(\underline{X}_{t-1}) > \varepsilon_t > 0) - 1(\widehat{d}(\underline{X}_{t-1}) < \varepsilon_t < 0) \right\} w(\underline{Z}_{t-1}) \\ & := A_T + B_T + C_T. \end{aligned}$$

From the above decomposition, we will show that under the assumed assumptions, the term  $A_T$  is asymptotically normal, and the terms  $B_T$  and  $C_T$  are asymptotically negligible.

Now, let us first show the asymptotic negligibility of term  $B_T$ . Define  $I(w) = \{t : \underline{Z}_{t-1} \in \mathcal{D}, t = 1, \dots, T\}$ . Note that  $X_t - \widehat{q}_\tau(\underline{Z}_{t-1}) = -\widehat{d}(\underline{Z}_{t-1}) + \varepsilon_t$ . Then,

$$\begin{aligned} |B_T| &\leq \frac{2}{T} \sum_{t=1}^T w(\underline{Z}_{t-1}) |X_t - \widehat{q}_\tau(\underline{Z}_{t-1})| 1(|\varepsilon_t| < |\widehat{d}(\underline{Z}_{t-1})|) \\ &\leq \frac{2}{T} \sum_{t=1}^T w(\underline{Z}_{t-1}) (|\widehat{d}(\underline{Z}_{t-1})| + |\varepsilon_t|) 1(|\varepsilon_t| < |\widehat{d}(\underline{Z}_{t-1})|) \\ &\leq \frac{4}{T} \sum_{t=1}^T w(\underline{Z}_{t-1}) |\widehat{d}(\underline{Z}_{t-1})| 1(|\varepsilon_t| < |\widehat{d}(\underline{Z}_{t-1})|) \\ &\leq 4 \max_{t \in I(w)} |\widehat{d}(\underline{Z}_{t-1})| \max_{\underline{z} \in \mathcal{D}} w(\underline{z}) \frac{1}{T} \sum_{t=1}^T 1\left(|\varepsilon_t| < \max_{s \in I(w)} |\widehat{d}(\underline{Z}_{s-1})|\right). \end{aligned}$$

From the Glivenko-Cantelli Theorem for strictly stationary sequences, we have

$$\sup_{a \in \mathbb{R}} \left| \frac{1}{T} \sum_{t=1}^T 1(|\varepsilon_t| < a) - Pr(|\varepsilon| < a) \right| = O_p\left(T^{-1/2}\right),$$

It thus follows that

$$\begin{aligned}
|B_T| &\leq 4 \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| \max_{\underline{z} \in \mathcal{D}} w(\underline{z}) \left\{ Pr \left( |\varepsilon| < \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| \right) + O_p \left( T^{-1/2} \right) \right\} \\
&= 4 \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| \max_{\underline{z} \in \mathcal{D}} w(\underline{z}) \left\{ F_\varepsilon \left( \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| \right) - F_\varepsilon \left( - \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| \right) \right\} \\
&\quad + 4 \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| \max_{\underline{z} \in \mathcal{D}} w(\underline{z}) O_p \left( T^{-1/2} \right) \\
&\leq C \left( \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| \right)^2 + CT^{-1/2} \max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right|,
\end{aligned}$$

where the third step follows from the Taylor expansion of  $F_\varepsilon$ , bounded marginal density of  $\varepsilon_t$  in Assumption A.3, and bounded weight function  $w(\cdot)$  in Assumption A.5. From Kong et al. (2010), we have

$$\max_{t \in I(w)} \left| \widehat{d}(\underline{Z}_{t-1}) \right| = O_p \left( \frac{\log T}{Th_2^d} \right)^{3/4},$$

It follows that  $B_T = O_p \left( \left( \frac{\log T}{Th_2^d} \right)^{3/2} + T^{-1/2} \left( \frac{\log T}{Th_2^d} \right)^{3/4} \right) = o_p \left( \left( Th_2^{d/2} \right)^{-1} \right)$  under Assumption A.9.

Similar to the term  $B_T$ , it can be proved that the term  $C_T = o_p \left( \left( Th_1^{d_1/2} \right)^{-1} \right) = o_p \left( \left( Th_2^{d/2} \right)^{-1} \right)$  under  $h_2^d = o(h_1^{d_1})$  in Assumption A.9. It follows that it is sufficient to establish that  $Th_2^{d/2} A_T$  converges in distribution to a normal random variable with asymptotic variance given by  $\tilde{\sigma}_{0\tau}^2 := \kappa(\tau)^2 \sigma_{0\tau}^2$ , for  $\kappa(\tau) = E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1})) w(\underline{Z}_{t-1})]$ .

Using Lemmas 1 and 2 of Appendix C, we have

$$\begin{aligned}
A_T &= - \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s=1, s \neq t}^T w(\underline{Z}_{t-1}) \underline{e}'_t \frac{H_T^{-1}}{h_2^d} S_{T,q}^{-1}(\underline{Z}_{t-1}) K_{h_2}(\underline{Z}_{s-1} - \underline{Z}_{t-1}) \mu(\underline{Z}_{s-1} - \underline{Z}_{t-1}) \varphi(\varepsilon_t) \varphi(\varepsilon_s) \\
&\quad + \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s=1, s \neq t}^T w(\underline{Z}_{t-1}) \underline{e}'_t \frac{H_T^{-1}}{h_1^{d_1}} S_{T,p}^{-1}(\underline{X}_{t-1}) K_{h_1}(\underline{X}_{s-1} - \underline{X}_{t-1}) \mu(\underline{X}_{s-1} - \underline{X}_{t-1}) \varphi(\varepsilon_t) \varphi(\varepsilon_s) \\
&\quad + o_p \left( \left( Th_2^{d/2} \right)^{-1} \right) \\
&:= A_{1T} + A_{2T} + o_p \left( \left( Th_2^{d/2} \right)^{-1} \right),
\end{aligned}$$

where the negligible terms with  $t = s$  have been dropped to apply  $U$ -statistic theory due to the leave one observation out in the estimation part. We will show that  $Th_2^{d/2} A_{1T}$  converges in distribution and  $A_{2T} = o_p \left( \left( Th_2^{d/2} \right)^{-1} \right)$  under our assumptions. First of all, to facilitate our analysis, from the notion of ‘‘equivalent kernel’’ representation for local polynomial estimator [see Fan and Gijbels, 1996, pp.63-64], we get

$$\begin{aligned}
A_{1T} &= \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \frac{w(\underline{Z}_{t-1})}{f_{\varepsilon|\underline{Z}}(0|\underline{Z}_{t-1}) f_{\underline{Z}}(\underline{Z}_{t-1})} K_{h_2}(\underline{Z}_{t-1} - \underline{Z}_{s-1}) \varphi(\varepsilon_t) \varphi(\varepsilon_s) + o_p \left( \left( Th_2^{d/2} \right)^{-1} \right) \\
&\equiv \bar{A}_{1T} + o_p \left( \left( Th_2^{d/2} \right)^{-1} \right), \quad \text{say.}
\end{aligned}$$

Note that we can rewrite  $Th_2^{d/2}\bar{A}_{1T}$  into a standard  $U$ -statistic form with a symmetrized kernel depending on the sample size  $T$ , i.e.

$$Th_2^{d/2}\bar{A}_{1T} = \frac{2}{T-1} \sum_{1 \leq t < s \leq T} U_T(\chi_t, \chi_s), \quad (1)$$

where  $\chi_t = (\underline{Z}_{t-1}, \varepsilon_t)$ ,  $U_T(\chi_t, \chi_s) = \eta_T(\chi_t, \chi_s) + \eta_T(\chi_s, \chi_t)$ , and

$$\eta_T(\chi_t, \chi_s) = \frac{w(\underline{Z}_{t-1})}{2f_{\varepsilon|\underline{Z}}(0|\underline{Z}_{t-1})f_{\underline{Z}}(\underline{Z}_{t-1})} \frac{1}{h_2^{d/2}} K\left(\frac{\underline{Z}_{t-1} - \underline{Z}_{s-1}}{h_2}\right) \varphi(\varepsilon_t)\varphi(\varepsilon_s).$$

Note that  $E[U_T(\chi_t, \chi_s)] = E[\eta_T(\chi_t, \chi_s)] = E[U_T(\chi_t, \chi_s)|\chi_t] = E[\eta_T(\chi_t, \chi_s)|\chi_t] = 0$  under Assumption A.3. So the previous  $U$ -statistic is a degenerate second order  $U$ -statistic. We can apply a central limit theorem (CLT) for second order degenerate  $U$ -statistic with strong mixing processes. Under Assumptions A.1, A.3, A.6, and A.9, one can verify that the conditions of Theorem A.1 in Gao (2007) are satisfied for kernel  $U_T(\chi_t, \chi_s)$  so that a CLT applies to the term  $Th_2^{d/2}\bar{A}_{1T}$ . Its asymptotic variance is given by

$$\begin{aligned} \tilde{\sigma}_{0\tau}^2 &= \lim_{T \rightarrow \infty} 2E_t E_s [U_T^2(\chi_t, \chi_s)] = \lim_{T \rightarrow \infty} 2E_t E_s [\eta_T(\chi_t, \chi_s)^2 + \eta_T(\chi_s, \chi_t)^2 + 2\eta_T(\chi_t, \chi_s)\eta_T(\chi_s, \chi_t)] \\ &= 2\tau^2 (1 - \tau)^2 \int K^2(\underline{u}) d\underline{u} \int \frac{w^2(\underline{z})}{f_{\varepsilon|\underline{Z}}^2(0|\underline{z})} d\underline{z} \\ &:= \kappa(\tau)^2 \sigma_{0\tau}^2, \end{aligned}$$

where  $E_t$  denotes the expectation with respect to  $\chi_t$ . For example, by straightforward calculation of conditional expectation, we have

$$\begin{aligned} &\lim_{T \rightarrow \infty} E_t E_s [\eta_T(\chi_t, \chi_s)^2] \\ &= \frac{1}{4} \tau^2 (1 - \tau)^2 \lim_{T \rightarrow \infty} \int \int \frac{w^2(\underline{z}_1)}{f_{\varepsilon|\underline{Z}}^2(0|\underline{z}_1)f_{\underline{Z}}^2(\underline{z}_1)} \frac{1}{h_2^d} K^2\left(\frac{\underline{z}_1 - \underline{z}_2}{h_2}\right) f_{\underline{Z}}(\underline{z}_1)f_{\underline{Z}}(\underline{z}_2) d\underline{z}_1 d\underline{z}_2 \\ &= \frac{1}{4} \tau^2 (1 - \tau)^2 \int K^2(\underline{u}) d\underline{u} \int \frac{w^2(\underline{z})}{f_{\varepsilon|\underline{Z}}^2(0|\underline{z})} d\underline{z} \end{aligned}$$

by standard use of change of variables and Assumptions A.7 and A.9. The  $U$ -statistic representation in (1), together with the form of asymptotic variance  $\tilde{\sigma}_{0\tau}^2$ , implies that  $Th_2^{d/2}A_{1T} = Th_2^{d/2}\bar{A}_{1T} + o_p(1) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_{0\tau}^2)$ .

Observing that by using almost the same steps as in proving the asymptotic normality of  $Th_2^{d/2}A_{1T}$ , we can prove that  $Th_1^{d_1/2}A_{2T}$  converges in distribution to a normal variable, and therefore  $Th_1^{d_1/2}A_{2T} = O_p(1)$ . Thus,  $Th_2^{d/2}A_T = Th_2^{d/2}A_{1T} + \left(h_2^{d/2}/h_1^{d_1/2}\right) \left(Th_1^{d_1/2}A_{2T}\right) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_{0\tau}^2)$  by the assumption  $h_2^d = o\left(h_1^{d_1}\right)$  in A.9.

In addition, a consistent estimator for  $\tilde{\sigma}_{0\tau}^2$  is given by

$$\begin{aligned} \hat{\sigma}_{0\tau}^2 &= 2\tau^2 (1 - \tau)^2 \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \frac{w^2(\underline{Z}_{t-1})}{\hat{f}_{\varepsilon, \underline{Z}}^2(0, \underline{Z}_{t-1})} \frac{1}{h_2^d} K^2\left(\frac{\underline{Z}_{t-1} - \underline{Z}_{s-1}}{h_2}\right) \\ &= 2\tau^2 (1 - \tau)^2 \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \frac{w^2(\underline{Z}_{t-1})}{\hat{f}_{\varepsilon, \underline{Z}}^2(0, \underline{Z}_{t-1})} \frac{1}{h_2^d} K^2\left(\frac{\underline{Z}_{t-1} - \underline{Z}_{s-1}}{h_2}\right) + o_p(1), \end{aligned}$$

where  $\widehat{f}_{\varepsilon, \underline{Z}}(0, \underline{Z}_{t-1})$  is the leave-one-out kernel density estimator defined in the main text for  $f_{\varepsilon, \underline{Z}}(0, \underline{Z}_{t-1}) \equiv f_{\varepsilon | \underline{Z}}(0 | \underline{Z}_{t-1}) f_{\underline{Z}}(\underline{Z}_{t-1})$ . As a consequence, the main term of  $\widehat{\sigma}_{0\tau}^2$  can also be written into a standard  $U$ -statistic form with a symmetrized kernel

$$H_T(\underline{Z}_{t-1}, \underline{Z}_{s-1}) = 2\tau^2 (1 - \tau)^2 \left( \frac{w^2(\underline{Z}_{t-1})}{f_{\varepsilon, \underline{Z}}^2(0, \underline{Z}_{t-1})} + \frac{w^2(\underline{Z}_{s-1})}{f_{\varepsilon, \underline{Z}}^2(0, \underline{Z}_{s-1})} \right) \frac{1}{h_2^d} K^2 \left( \frac{\underline{Z}_{t-1} - \underline{Z}_{s-1}}{h_2} \right).$$

Note that in contrast to (1),  $\widehat{\sigma}_{0\tau}^2$  is a non-degenerate second order  $U$ -statistic and by the usual Hoeffding decomposition, one can thus show that  $\widehat{\sigma}_{0\tau}^2 = \tilde{\sigma}_{0\tau}^2 + o_p(1)$ .

Finally, observing that by Taylor expansion of  $\ln y$  around 1 (i.e.  $\ln y \approx y - 1$ ) and using the asymptotic equivalence of  $\widehat{\kappa}(\tau)$  to  $\kappa(\tau) = E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})]$  stated in Lemma 4 of Appendix C, together with Slutsky's theorem, we have

$$\begin{aligned} Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) &= Th_2^{d/2} \left( \frac{T^{-1} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})}{T^{-1} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})} - 1 \right) + o_p(1) \\ &= \kappa(\tau)^{-1} Th_2^{d/2} A_T + o_p(1) \\ &\xrightarrow{d} \mathcal{N}(0, \sigma_{0\tau}^2), \end{aligned}$$

where  $\sigma_{0\tau}^2 := \tilde{\sigma}_{0\tau}^2 / \kappa(\tau)^2$ . It is straightforward to show that  $\widehat{\sigma}_{0\tau}^2 = \widehat{\sigma}_{0\tau}^2 / \widehat{\kappa}(\tau)^2$  is a consistent estimator for  $\sigma_{0\tau}^2$ . Thus, our test statistic  $\widehat{\Gamma}_\tau = Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) / \widehat{\sigma}_{0\tau} \xrightarrow{d} \mathcal{N}(0, 1)$ . This ends the proof of Theorem 2.  $\square$

**Proof of Proposition 3:** This result can be shown by following the same steps as in the proof of Theorem

2. Noting that, under the fixed alternative hypothesis  $H_1$  in Equation (15) of the main text, we have

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1}) - \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1}) \\ &= \left( \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1}) - \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1}) \right) \\ &\quad + A_T + B_T + C_T \\ &:= D_T + A_T + B_T + C_T, \end{aligned}$$

where the last three terms  $A_T$ ,  $B_T$ , and  $C_T$  are as defined before in the proof of Theorem 2. Following the same arguments as those in Theorem 2,  $Th_2^{d/2} (A_T + B_T + C_T) = O_p(1)$ . As a matter of fact, one can furthermore prove that all  $A_T$ ,  $B_T$  and  $C_T$  are of order  $o_p(T^{-1/2})$ , see the proof of lemma 3 in Noh et al. (2013). On the other hand, under  $H_1$  of causality, the weak law of large numbers yields immediately

$$\begin{aligned} D_T &= E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})] - E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] + o_p(1) \quad (2) \\ &= E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] \left( \frac{E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})]}{E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})]} - 1 \right) + o_p(1) \\ &= E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] C_\tau(Y \rightarrow X) + o_p(1) \\ &:= \kappa(\tau) \times C_\tau(Y \rightarrow X) + o_p(1), \end{aligned}$$



where the third step follows by a Taylor expansion of  $\ln y$  around 1.

Therefore, since under  $H_1$ ,  $C_\tau(Y \rightarrow X) > 0$ , or equivalently,  $\Pr [q_\tau(\underline{Z}_{t-1}) = \bar{q}_\tau(\underline{X}_{t-1})] < 1$ , we have

$$\begin{aligned}
& Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) \\
&= Th_2^{d/2} \frac{T^{-1} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1}) - T^{-1} \sum_{t=1}^T \rho_\tau(X_t - \hat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})}{T^{-1} \sum_{t=1}^T \rho_\tau(X_t - \hat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})} \times [1 + o_p(1)] \\
&= \kappa(\tau)^{-1} \left[ Th_2^{d/2} D_T + Th_2^{d/2} (A_T + B_T + C_T) \right] \times [1 + o_p(1)] \\
&= Th_2^{d/2} C_\tau(Y \rightarrow X) \rightarrow \infty.
\end{aligned}$$

Alternatively, under  $H_1$  of causality, one can simply apply the consistency result in Proposition 2 to show that  $C_\tau(\widehat{Y \rightarrow X})$  converges in probability to  $C_\tau(Y \rightarrow X) > 0$ , and consequently  $Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X})$  will diverge to infinity under our assumptions.

On the other hand, following arguments similar to those we have used in the proof of the consistency of estimator  $\widehat{\sigma}_{0\tau}^2$  to the asymptotic variance  $\sigma_{0\tau}^2$  in Theorem 2 under the null hypothesis, we can show that  $\widehat{\sigma}_{0\tau}^2 := \widehat{\sigma}_{0\tau}^2 / \widehat{C}_\tau^2 = O_p(1)$  under the alternative hypothesis of no causality. Proposition 3 follows then from  $Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) \rightarrow \infty$  and  $\widehat{\sigma}_{0\tau} = O_p(1)$  as  $T \rightarrow \infty$ . Hence, the test  $\widehat{\Gamma}_\tau = Th_2^{d/2} C_\tau(\widehat{Y \rightarrow X}) / \widehat{\sigma}_{0\tau}$  is diverging to infinity at the rate  $Th_2^{d/2}$  and is consistent.  $\square$

**Proof of Proposition 4:** First, following similar arguments as in Theorem 2 and Proposition 3, with the only exception that the term  $D_T$  defined in (2) now takes a different form. Specifically, we can show that under the local alternatives given in Equation (18) of the main text,

$$\begin{aligned}
& Th_2^{d/2} \left( \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \widehat{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1}) - \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \hat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1}) - D_T \right) \\
& := Th_2^{d/2} (A_T + B_T + C_T) \xrightarrow{d} \mathcal{N}(0, \widehat{\sigma}_{0\tau}^2),
\end{aligned}$$

with  $A_T, B_T, C_T$ , and  $\widehat{\sigma}_{0\tau}^2$  given in the proof of Theorem 2.

Second, under the local alternative hypotheses  $H_1(\delta_T)$ , using the second order Taylor expansion, one can calculate that

$$\begin{aligned}
D_T &= E [\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})] - E [\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] [1 + o_p(1)] \\
&= E [\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}) + \delta_T \Delta_T(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] - E [\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] [1 + o_p(1)] \\
&= \delta_T E [\Delta_T(\underline{Z}_{t-1})w(\underline{Z}_{t-1})\varphi(\varepsilon_t)] - \frac{\delta_T^2}{2} E [\Delta_T^2(\underline{Z}_{t-1})w(\underline{Z}_{t-1})g(\underline{Z}_{t-1})] [1 + o_p(1)] \\
&= \delta_T^2 E [\Delta_T^2(\underline{Z}_{t-1})w(\underline{Z}_{t-1})f_{\varepsilon|\underline{Z}}(0|\underline{Z}_{t-1})] [1 + o_p(1)],
\end{aligned}$$

where  $g(\underline{z}) = \partial E[\varphi(X_t - \theta) | \underline{Z}_{t-1} = \underline{z}] / \partial \theta = -f_{\varepsilon|\underline{Z}}(0|\underline{z})$ , and the fourth step follows by law of iterated

expectations and  $E[\varphi(\varepsilon_t)|\underline{Z}_{t-1}] = 0$  in Assumption A.3. Consequently, with  $\delta_T = \left(Th_2^{d/2}\right)^{-1/2}$ , we have

$$\begin{aligned} Th_2^{d/2} C_\tau \widehat{C_\tau(Y \rightarrow X)} &= \kappa(\tau)^{-1} \left[ Th_2^{d/2} D_T + Th_2^{d/2} (A_T + B_T + C_T) \right] \times [1 + o_p(1)] \\ &\xrightarrow{d} \mathcal{N}(\gamma, \sigma_{0\tau}^2) \end{aligned}$$

under the local alternatives with

$$\gamma = \kappa(\tau)^{-1} \lim_{T \rightarrow \infty} E \left[ \Delta_T^2(\underline{Z}_{t-1}) w(\underline{Z}_{t-1}) f_{\varepsilon|\underline{Z}}(0|\underline{Z}_{t-1}) \right].$$

This concludes the proof of Proposition 4.  $\square$

**Proof of Theorem 3:** The asymptotic validity of our bootstrap procedure can be proved using similar arguments to those used in the proof of Theorem 2, with the term  $A_{1T}$  replaced by its bootstrapped version  $A_{1T}^*$  using the bootstrapped sample  $\{(X_t^*, Y_t^*)\}_{t=1}^T$ . Conditionally on  $\{(X_t, Y_t)\}_{t=1}^T$  and using Theorem 1 of Hall (1984), we obtain the bootstrap validity result in Theorem 3.  $\square$

**Proof of Theorem 4:** The proof is similar to the proof of Theorem 1 and a sketched proof is provided.

Denote  $\widehat{d}_1(\underline{W}_{t-1}) = (\widehat{\phi}_\tau - \bar{\phi}_\tau)' \underline{W}_{t-1}$  and  $\widehat{d}_2(\underline{X}_{t-1}) = \widehat{q}_\tau(\underline{X}_{t-1}) - \bar{q}_\tau(\underline{X}_{t-1})$ . Note the following expansion holds,

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \widehat{\phi}_\tau' \underline{W}_{t-1} - \widehat{q}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1}) \\ &= \frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \bar{\phi}_\tau' \underline{W}_{t-1} - \bar{q}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1}) \\ & \quad - \frac{1}{T} \sum_{t=1}^T \left( \widehat{q}_\tau(\underline{X}_{t-1}) - \bar{q}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1}) \varphi(\bar{\varepsilon}_t) - \frac{1}{T} \sum_{t=1}^T \left( \widehat{\phi}_\tau - \bar{\phi}_\tau \right)' \underline{W}_{t-1} w(\underline{Z}_{t-1}) \varphi(\bar{\varepsilon}_t) \\ & \quad - \frac{2}{T} \sum_{t=1}^T \left( X_t - \bar{\phi}_\tau' \underline{W}_{t-1} - \bar{q}_\tau(\underline{X}_{t-1}) \right) \left\{ 1(\widehat{d}_1(\underline{W}_{t-1}) + \widehat{d}_2(\underline{X}_{t-1}) > \bar{\varepsilon}_t > 0) \right. \\ & \quad \left. - 1(\widehat{d}_1(\underline{W}_{t-1}) + \widehat{d}_2(\underline{X}_{t-1}) < \bar{\varepsilon}_t < 0) \right\} w(\underline{Z}_{t-1}) \\ &:= \frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \bar{\phi}_\tau' \underline{W}_{t-1} - \bar{q}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1}) + E_T + F_T + G_T \\ &= \frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \bar{\phi}_\tau' \underline{W}_{t-1} - \bar{q}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1}) + o_p(T^{-1/2}), \end{aligned}$$

where  $E_T = F_T = G_T = o_p(T^{-1/2})$  can be proved using the steps in the proof of lemma 3 in Noh et al. (2013) and by noting that  $\max_{t \in I(w)} |\widehat{d}_1(\underline{W}_{t-1})| = O_p(T^{-1/2})$  for bounded support. Moreover, the asymptotic representation for

$$\frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \widehat{\phi}_\tau' \underline{W}_{t-1} - \widehat{q}_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) = \frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \phi_\tau' \underline{W}_{t-1} - q_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) + o_p(T^{-1/2})$$

can be obtained using the same arguments as Lemmas 3 and 4 of Appendix C. The proof then follows by using the equality that  $\hat{a}\hat{b}^{-1} = ab^{-1} + \hat{b}^{-1} \left[ (\hat{a} - a) - (\hat{b} - b) ab^{-1} \right]$ .  $\square$

**Proof of Theorem 5:** Consider the following decomposition of  $\rho_\tau(\cdot)$ :

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \hat{\phi}'_\tau W_{t-1} - \hat{q}_\tau(\underline{X}_{t-1}) \right) w(\underline{Z}_{t-1}) - \frac{1}{T} \sum_{t=1}^T \rho_\tau \left( X_t - \hat{\phi}'_\tau W_{t-1} - \hat{q}_\tau(\underline{Z}_{t-1}) \right) w(\underline{Z}_{t-1}) \\ &= \frac{1}{T} \sum_{t=1}^T \left[ (\hat{q}_\tau(\underline{Z}_{t-1}) - q_\tau(\underline{Z}_{t-1})) - (\hat{q}_\tau(\underline{X}_{t-1}) - \bar{q}_\tau(\underline{X}_{t-1})) \right] w(\underline{Z}_{t-1}) \varphi(\varepsilon_t) \\ & \quad + \frac{1}{T} \sum_{t=1}^T \left( (\hat{\phi}_\tau - \phi_\tau) - (\hat{\phi}_\tau - \bar{\phi}_\tau) \right)' W_{t-1} w(\underline{Z}_{t-1}) \varphi(\varepsilon_t) + \text{higher order terms} \\ & := H_T + I_T + \text{higher order terms.} \end{aligned}$$

Following the same arguments as in the proof of Theorem 2, it can be shown that  $Th_2^{d/2} H_T \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_{0\tau}^2)$  and  $Th_2^{d/2} I_T = \left[ \sqrt{T} (\hat{\phi}_\tau - \phi_\tau) - \sqrt{T} (\hat{\phi}_\tau - \bar{\phi}_\tau) \right] h_2^{d/2} O_p(1) = o_p(1)$  by root- $T$  consistency properties of linear coefficients estimators  $\hat{\phi}_\tau$  and  $\hat{\bar{\phi}}_\tau$ . Therefore,

$$\begin{aligned} & Th_2^{d/2} C_\tau^{PL} (\widehat{Y \rightarrow X|W}) \\ &= (\kappa^{PL}(\tau))^{-1} Th_2^{d/2} H_T \times [1 + o_p(1)] \\ & \xrightarrow{d} \mathcal{N}(0, \sigma_{0\tau}^{PL2}), \end{aligned}$$

which proves that  $Th_2^{d/2} C_\tau^{PL} (\widehat{Y \rightarrow X|W})$  converges to a normal distribution under the null of no causality in the presence of control variables  $W$ .  $\square$

## Appendix C: Proofs of auxiliary results

In this section, we provide four auxiliary lemmas which are useful to prove our main results in Appendix B. In the first lemma, the uniform Bahadur representation for the estimator of the restricted conditional quantile function  $\bar{q}_\tau(\underline{x})$  based on a  $p$ -th order local polynomial approximation using bandwidth  $h_1$  is derived. Please notice that the proofs of the following Lemmas 1-4 can be obtained using similar arguments as in lemmas 2 and 3 in Noh et al. (2013) [or using results in Kong et al. (2010), see their corollary 1, lemmas 8 and 10, respectively], and they are therefore omitted.

**Lemma 1:** *Let  $\bar{\varepsilon}_1$  be an  $N_1 \times 1$  vector with its first element given by 1 and all others 0. Suppose **A.1-A.9** in Appendix A hold and  $h_1 = O(T^{-\kappa_1})$  with  $\kappa_1 > 1/(2p + 2 + d_1)$ . Then, with probability one, we have*

$$\widehat{\bar{q}}_\tau(\underline{x}) - \bar{q}_\tau(\underline{x}) = -\bar{\varepsilon}'_1 \frac{H_T^{-1}}{Th_1^{d_1}} S_{T,p}^{-1}(\underline{x}) \sum_{t=1}^T K_{h_1}(\underline{X}_{t-1} - \underline{x}) \varphi(\bar{\varepsilon}_t) \mu(\underline{X}_{t-1} - \underline{x}) + \bar{R}_T,$$

where  $\bar{\varepsilon}_t = X_t - \bar{q}_\tau(\underline{X}_{t-1})$  is the restricted error and  $\bar{R}_T = o_p\left(\left(Th_1^{d_1}\right)^{-1/2}\right)$  uniformly in  $\underline{x} \in \mathcal{D}_X$  and  $\mathcal{D}_X$  is the compact support of the weighting function  $w(\cdot)$  with respect to the part of  $X$ .

Analogously, the  $q$ -th order local polynomial estimator of the unrestricted conditional quantile function  $q_\tau(\underline{z})$  using bandwidth  $h_2$ , say  $\hat{q}_\tau(\underline{z})$ , can be defined accordingly as in Section 4 and its uniform Bahadur representation can be obtained similarly and is stated in the next lemma. Note that Lemma 1 is only a special case of Lemma 2.

**Lemma 2:** Denote  $d = d_1 + d_2$ . Let  $\underline{e}_1$  be an  $N_2 \times 1$  vector with its first element given by 1 and all others 0. Suppose Assumptions **A.1-A.9** in Appendix A hold and  $h_2 = O(T^{-\kappa_2})$  with  $\kappa_2 > 1/(2q + 2 + d)$ . Then, with probability one, we have

$$\hat{q}_\tau(\underline{z}) - q_\tau(\underline{z}) = -\underline{e}_1' \frac{H_T^{-1}}{Th_2^d} S_{T,q}^{-1}(\underline{z}) \sum_{t=1}^T K_{h_2}(\underline{Z}_{t-1} - \underline{z}) \varphi(\varepsilon_t) \mu(\underline{Z}_{t-1} - \underline{z}) + R_T,$$

where  $\varepsilon_t = X_t - q_\tau(\underline{Z}_{t-1})$  is the unrestricted error and  $R_T = o_p\left((Th_2^d)^{-1/2}\right)$  uniformly in  $\underline{z} \in \mathcal{D}$  and  $\mathcal{D}$  is the compact support of the weighting function  $w(\cdot)$ .

On the other hand, to derive the Bahadur representation of  $C_\tau(\widehat{Y} \rightarrow X)$ , we need to investigate the asymptotic behaviour of  $T^{-1} \sum_{t=1}^T \rho_\tau(X_t - \widehat{\bar{q}}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})$  [resp.  $T^{-1} \sum_{t=1}^T \rho_\tau(X_t - \hat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})$ ], which is stated in the next two lemmas. Again, the proof of Lemma 3 is similar to the one of Lemma 4.

**Lemma 3:** Suppose Assumptions **A.1-A.9** in Appendix A hold,  $p > d_1/2 - 1$  and  $h_1 = O(T^{-\kappa_1})$  with  $1/(2p + 2 + d_1) < \kappa_1 < 1/(2d_1)$ . Then,

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \widehat{\bar{q}}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1}) - E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})] \\ &= \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1}) - E[\rho_\tau(X_t - \bar{q}_\tau(\underline{X}_{t-1}))w(\underline{Z}_{t-1})] + o_p(T^{-1/2}). \end{aligned}$$

**Lemma 4:** Let  $d = d_1 + d_2$ . Suppose Assumptions **A.1-A.9** in Appendix A hold,  $q > d/2 - 1$  and  $h_2 = O(T^{-\kappa_2})$  with  $1/(2q + 2 + d) < \kappa_2 < 1/(2d)$ . Then, we have

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - \hat{q}_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1}) - E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] \\ &= \frac{1}{T} \sum_{t=1}^T \rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1}) - E[\rho_\tau(X_t - q_\tau(\underline{Z}_{t-1}))w(\underline{Z}_{t-1})] + o_p(T^{-1/2}). \end{aligned}$$

## Appendix D: Additional simulation results

Table 1: Bootstrap bias-corrected estimation of quantile Granger causality measures at  $\tau = 0.25$

Measure		DGP S1	DGP S2	DGP P1	DGP P2	DGP P3	DGP P4	DGP P5	DGP P6
$T = 50$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.1122 (0.1152)	0.1094 (0.1112)	0.7501 (0.3056)	0.7635 (0.3169)	1.0212 (0.3718)	0.7162 (0.3283)	0.8587 (0.3792)	0.8246 (0.3589)
$T = 100$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0609 (0.0565)	0.0544 (0.0534)	0.6049 (0.1826)	0.6295 (0.1980)	0.8371 (0.2283)	0.6157 (0.2096)	0.6260 (0.2026)	0.6279 (0.2219)
$T = 200$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0380 (0.0316)	0.0314 (0.0319)	0.5071 (0.1138)	0.5370 (0.1250)	0.7156 (0.1458)	0.5649 (0.1476)	0.4696 (0.1206)	0.4889 (0.1305)
$T = 400$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0230 (0.0180)	0.0180 (0.0184)	0.4050 (0.0670)	0.4303 (0.0729)	0.5851 (0.0861)	0.4762 (0.0984)	0.3129 (0.0673)	0.3282 (0.0755)

**Note:** This table shows the average values of bootstrap bias-corrected ( $C_{\tau,BC}^*(\widehat{Y \rightarrow X})$ ) estimates of causality measures from  $Y$  to  $X$  ( $C_\tau(Y \rightarrow X)$ ). “True” indicates the true value of causality measure, “Bias-Corrected” indicates the average value of the estimate of causality measure after bootstrap bias correction, and “—” means that the true value of causality measure is unknown. Equation (29) in the main text is used to calculate the bootstrap bias-correction estimates of causality measures. The number of simulations used to compute the averaged values of the estimates of causality measures and the number of bootstrap replications used to calculate the bias-corrected estimates are equal to 500 and 199, respectively. “No” indicates non-causality in the true DGP and “Yes” means that there is causality in the true GDP. The DGPs in the first row of the table are described in detail in Table 1 of the main text. In parenthesis is the standard deviation of the estimated value.

Table 2: Bootstrap bias-corrected estimation of quantile Granger causality measures at  $\tau = 0.50$

Measure		DGP S1	DGP S2	DGP P1	DGP P2	DGP P3	DGP P4	DGP P5	DGP P6
$T = 50$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.1076 (0.1050)	0.0882 (0.0963)	0.6746 (0.3028)	0.6725 (0.3058)	0.9096 (0.3623)	0.6369 (0.3324)	0.7973 (0.3655)	0.7355 (0.3671)
$T = 100$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0584 (0.0578)	0.0510 (0.0536)	0.5335 (0.1632)	0.5453 (0.1658)	0.8045 (0.2457)	0.5695 (0.2080)	0.5597 (0.2038)	0.5352 (0.1827)
$T = 200$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0332 (0.0307)	0.0294 (0.0277)	0.4361 (0.0998)	0.4661 (0.1090)	0.7082 (0.1482)	0.5130 (0.1403)	0.4201 (0.1172)	0.3848 (0.1058)
$T = 400$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0202 (0.0160)	0.0156 (0.0150)	0.3929 (0.0618)	0.4075 (0.0675)	0.6536 (0.1083)	0.4769 (0.1021)	0.3176 (0.0736)	0.2881 (0.0621)

**Note:** This table shows the average values of bootstrap bias-corrected ( $C_{\tau,BC}^*(\widehat{Y \rightarrow X})$ ) estimates of causality measures from  $Y$  to  $X$  ( $C_\tau(Y \rightarrow X)$ ). “True” indicates the true value of causality measure, “Bias-Corrected” indicates the average value of the estimate of causality measure after bootstrap bias correction, and “—” means that the true value of causality measure is unknown. Equation (29) in the main text is used to calculate the bootstrap bias-correction estimates of causality measures. The number of simulations used to compute the averaged values of the estimates of causality measures and the number of bootstrap replications used to calculate the bias-corrected estimates are equal to 500 and 199, respectively. “No” indicates non-causality in the true DGP and “Yes” means that there is causality in the true GDP. The DGPs in the first row of the table are described in detail in Table 1 of the main text. In parenthesis is the standard deviation of the estimated value.

Table 3: Bootstrap bias-corrected estimation of quantile Granger causality measures at  $\tau = 0.75$

Measure		DGP S1	DGP S2	DGP P1	DGP P2	DGP P3	DGP P4	DGP P5	DGP P6
$T = 50$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.1062 (0.1090)	0.0891 (0.0995)	0.6985 (0.2831)	0.7092 (0.3075)	0.9373 (0.3925)	0.6840 (0.3078)	0.7700 (0.3706)	0.7603 (0.3410)
$T = 100$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0633 (0.0601)	0.0624 (0.0601)	0.5595 (0.1828)	0.5699 (0.1751)	0.8810 (0.2648)	0.5672 (0.1953)	0.6344 (0.2160)	0.5550 (0.1961)
$T = 200$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0396 (0.0311)	0.0339 (0.0312)	0.4582 (0.1068)	0.4899 (0.1166)	0.8340 (0.1857)	0.5217 (0.1415)	0.5183 (0.1322)	0.4274 (0.1192)
$T = 400$									
$Y \rightarrow X$		No	No	Yes	Yes	Yes	Yes	Yes	Yes
True	$C_\tau(Y \rightarrow X)$	0.0000	0.0000	—	—	—	—	—	—
Bias-Corrected	$C_{\tau,BC}^*(\widehat{Y \rightarrow X})$	0.0239 (0.0194)	0.0164 (0.0165)	0.4045 (0.0674)	0.4253 (0.0709)	0.7746 (0.1179)	0.4768 (0.1010)	0.4275 (0.0851)	0.3286 (0.0718)

**Note:** This table shows the average values of bootstrap bias-corrected ( $C_{\tau,BC}^*(\widehat{Y \rightarrow X})$ ) estimates of causality measures from  $Y$  to  $X$  ( $C_\tau(Y \rightarrow X)$ ). “True” indicates the true value of causality measure, “Bias-Corrected” indicates the average value of the estimate of causality measure after bootstrap bias correction, and “—” means that the true value of causality measure is unknown. Equation (29) in the main text is used to calculate the bootstrap bias-correction estimates of causality measures. The number of simulations used to compute the averaged values of the estimates of causality measures and the number of bootstrap replications used to calculate the bias-corrected estimates are equal to 500 and 199, respectively. “No” indicates non-causality in the true DGP and “Yes” means that there is causality in the true GDP. The DGPs in the first row of the table are described in detail in Table 1 of the main text. In parenthesis is the standard deviation of the estimated value.

## References

- [1] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*, Chapman & Hall, London.
- [2] Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman & Hall, CRC.
- [3] Hall, P. (1984), “Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators,” *Journal of Multivariate Analysis*, 14, 1–16.
- [4] Kong, E., Linton, O., Xia, Y. (2010). “Uniform Bahadur Representation for Local Polynomial Estimates of M-regression and Its Application to the Additive Model”, *Econometric Theory*, 26, 1529–1564.
- [5] Masry, E. (1996). “Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates”, *Journal of Time Series Analysis*, 17, 571–599.
- [6] Noh, H., El Ghouch, A., Van Keilegom, I. (2013). “Quality of Fit Measures in the Framework of Quantile Regression”, *Scandinavian Journal of Statistics*, 40, 105–118.