Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 1, pp. 195 ~ 210

Nonparametric Granger Causality Test[1)]

Kiho Jeong[2)] · Yoshihiko Nishiyama[3)]

Abstract

This paper develops a consistent nonparametric test for Granger causality in the context of strong-mixing process, which covers a large class of stationary processes including ARMA and ARCH models. The previously proposed tests require absolute regularity (β-mixing) more stringent than the strong-mixing condition. We prove the consistency of the test under a high level assumption on the approximation error of U statistic by its projection. Due to the sample splitting, the test statistic we propose is asymptotically normally distributed under the null.

Keywords : Granger Causality Test, Nonparametric Kernel Method, Strong-mixing, U-statistic

1. Introduction

Whether movements in one economic variable anticipate movements in another variable is an important question to the formulation of a policy. Most empirical researches of such relations have been done in the context of Granger-causality in a multivariate framework, the mean version of which is defined as follows:
(i) M does not cause Y in mean with respect to U_{t-1} if

$$E\,(y_t\,|\,U_{t-1}\,) = E(y_t\,|\,U_{t-1} - M_{t-1}\,) \tag{1}$$

and
(ii) M is a prima facie cause in mean of Y with respect to U_{t-1} if

$$E\,(y_t\,|\,U_{t-1}\,) \neq E(y_t\,|\,U_{t-1} - M_{t-1}\,), \tag{2}$$

---

1) This work was supported by Korea Research Foundation Grant (KRF-2003-041-B00071). The second author's research is partially supported by Japan Society for the Promotion of Science Grant 17653021.

2) Professor, School of Economics and Trade, Kyungpook National University, Daegu, 702-701, Korea
E-mail : khjeong@knu.ac.kr

3) Professor, Institute of Economic Research, Kyoto University, Sakyo Kyoto 606-8501, Japan
E-mail : nishiyama@kier.kyoto-u.ac.jp

where M and Y are two time series, M_t is M's entire histories up to and including time t with $M_t = \{m_t, s \geq 0\}$ and U_t is the information set available at time t. $U_t - M_t$ indicates the information set excluding M_t. For further discussion, we refer to Granger(1988).

A conventional approach of testing Granger-causality is to assume a parametric linear model for the conditional mean $E(y_t|U_{t-1})$, regress the y_t on finite number of lagged values of Y and M, and test the null hypothesis that the coefficients on lagged values of M are all zero. Although conventional linear tests have high power in uncovering linear causal relations, its power can be low against nonlinear causal relations. An example of such cases is that both Y and M are Gaussian, and Y depends on M only through the squares of lagged values of M.

In order to circumvent this possible nonlinearity problem, Baek and Brock(1992) proposed a nonparametric causality test based on the correlation integral. As being widely used in nonlinear dynamics, the correlation integral is an estimator of spatial probabilities across time that can be used to detect nonlinear relations between time series. The Baek and Brock test depends on a strong assumption of mutually independent and individually iid for the time series. The strong assumption is relaxed by Hiemstra and Jones(1994) to the more general case of weakly dependent time series, absolutely regularity. The Hiemstra and Jones test seems to be most used one for nonlinear causality hypothesis testing in the literature.

It is important to note that the test proposed by Baek and Brock(1992) and modified by Hiemstra and Jones(1994) is for the distribution version of Granger-causality. The distribution version is defined by replacing expectation operator "E" in the equations (1) and (2) with probability operator "Pr". In real applications, the distribution version of Granger-causality is too strong in that the non-existence of Granger-causality in the distribution version implies the non-existence of Granger-causality in all distributional characteristics including mean. In practice, the Hiemstra and Jones test is applied to regression errors from a linear VAR model. The test looks for nonlinear predictive power between two time series after filtering out any linear predictive power of lagged values of one series for present and future values of another series. As such, applying the distribution version of Granger-causality to the regression errors of the linear VAR model seems too strong and even inconsistent applying the mean version would be suitable.

This paper proposes an alternative approach to the nonparametric Granger causality test. Based on the mean version of Granger-causality, the proposed test directly compares nonparametric kernel estimators of the two conditional means in equations (1) and (2). The preliminary estimation of a linear VAR model used in the Hiemstra and Jones test is not required in our test. We derive a normal null limiting distribution for the proposed test statistic in the context of strong-mixing process. The strong-mixing process covers a large class of stationary processes including ARMA and ARCH models and is weaker than absolute regularity assumed by Hiemstra and Jones (1994). For example, some stationary strong mixing Gaussian processes does not belong to the absolute regular process; see Doukhan(1994). The proposed test is shown to have the desired feature of consistency.

2. Nonparametric Granger Causality Test

To simplify the exposition, we assume a bivariate case, or only $\{y_t, m_t\}$ are observable. Also assuming that y is an ARMAX process written as the following regression relationship with known p and q.

$$y_t = E(y_t | y_{t-1}, \dots, y_{t-p}, m_{t-1}, \dots, m_{t-q}) + \epsilon_t. \tag{3}$$

From the definitions (1) and (2), the hypotheses of causality test are specified as

$$H_o: E(y_t | y_{t-1}, \dots, y_{t-p}, m_{t-1}, \dots, m_{t-q}) = E(y_t | y_{t-1}, \dots, y_{t-p}),$$
$$H_1: E(y_t | y_{t-1}, \dots, y_{t-p}, m_{t-1}, \dots, m_{t-q}) \neq E(y_t | y_{t-1}, \dots, y_{t-p}),$$

and that under the null hypothesis of non-causality, the regression (3) reduces to

$$y_t = E(y_t | y_{t-1}, \dots, y_{t-p}) + \epsilon_t. \tag{4}$$

Thus, L_2-measure comparison of nonparametric estimators of equations (3) and (4) can be one way to test (4) against (3).

For notational simplicity, let $z_t = (y_{t-1}, \dots, y_{t-p}, m_{t-1}, \dots, m_{t-q})$, $x_t = (y_{t-1}, \dots, y_{t-p})$ and $k = p + q$. To measure the discrepancy between (3) and (4), we consider the following L_2-measure:

$$\begin{aligned} M_C &= E[\{E(y|z) - E(y|x)\}^2] \\ &= E[\{y - E(y|x)\}^2] - E[\{y - E(y|z)\}^2] \end{aligned} \tag{5}$$

Apparently, H_o and H_1 are equivalent to $M_C = 0$ and $M_C > 0$ respectively.

Using the techniques of sample-splitting of Robinson(1989) and data-trimming of Hardle and Stoker(1989), our test is based on the kernel estimate of $M_C/2$ denoted as

$$\begin{aligned} \widetilde{M}_C = {} & T^{-1} \sum w_t \{y_t - \hat{E}(y_t|x_t)\}^2 1(\hat{f}(z_t) > b_T) \\ & - T^{-1} \sum (1 - w_t) \{y_t - \hat{E}(y_t|z_t)\}^2 1(\hat{f}(z_t) > b_T), \end{aligned} \tag{6}$$

where $\hat{E}(y_t | \cdot)$ denotes the Nadaraya-Watson kernel estimator defined as

$$\widehat{E}(y_t|z_t) = \frac{T^{-1}\sum_s h_T^{-k} K\left(\frac{z_t - z_s}{h_T}\right) y_s}{T^{-1}\sum_s h_T^{-k} K\left(\frac{z_t - z_s}{h_T}\right)} \equiv \frac{\widehat{D}(z_t)}{\hat{f}(z_t)}, \qquad (7)$$

$K(\cdot)$ is a kernel function, h_T is a smoothing parameter that depends on the sample size T, $\hat{f}(z)$ is a nonparametric kernel estimator of the marginal density function of z, b_T is trimming bound converging to zero at some rate, $w_t = 0$ if the t-th observation belongs to the first half of the sample and $w_t = 1$ otherwise.

We establish the asymptotic distribution of the statistic $\widetilde{M}_C$ under the condition of the strong mixing processes, one of the weak dependent processes, including stationary finite order ARMA processes with innovations satisfying some general conditions (Athreya and Pantula(1986) and nonlinear AR model with ARCH term satisfying some conditions (Marsry and Tjostheim(1995)).

We give the formal definition of strong mixing. For a<b, denote μ_a^b as the s-algebra of events generated by $\xi_a, \ldots, \xi_b$.

Definition 2.1 $\{\xi_t\}$ is strong-mixing if

$$\alpha(n) = sup_l \, sup_{A \in \mu_{-\infty}^l, B \in \mu_{l+n}^{\infty}} |P(A \cap B) - P(A)P(B)| \to 0 \text{ as } n \to \infty.$$

In order to prove Theorem, we assume the following conditions.

Assumption 1:
(a) $\{(y_t, m_t)\}$ is a 2-dimensional strictly stationary, strong-mixing sequence of stochastic vectors with mixing coefficient $\alpha(n)$.
(b) The distribution of $z_t = (y_{t-1}, \ldots, y_{t-p}, m_{t-1}, \ldots, m_{t-q})$ is absolutely continuous with respect to the Lebesque measure. The density function, denoted by f_z, is bounded, strictly positive, and λ-times differentiable with bounded λ^{th} derivatives, where $\lambda > 2k$.
(c) $E(y|z)$ is λ-times differentiable with bounded λ^{th} derivatives.
(d) $E|y|^3 < \infty$.

Assumption 2: The kernel K satisfies
(a) K is bounded.
(b) $uK(u) \to 0$ as $|u| \to \infty$.

(c) $\int K(u)\,du = 1$.

(d) $\int u_1^{r_1} \ldots u_k^{r_k} K(u)\,du \quad \begin{matrix} = 0, \text{ for } 0 < r_1 + \ldots + r_k \\ < C \neq 0, \text{ for } \lambda \leq r_1 + \ldots + r_k \end{matrix}$

where $k = p + q$.

Assumption 3: Let $\rho_T \equiv T^{-1} \sum_{n=1}^{T} \{\alpha(n)\}^{1/3}$. The bandwidth sequence h_T, the trimming bound b_T, and mixing coefficient $\alpha(n)$ satisfy

(a) $b_T^{-1} T^{1/2} h_T^{\lambda} = o(1)$.

(b) $b_T^{-1} \rho_T T^{1/2} h_T^{-2k} = o(1)$.

(c) $b_T^{2+\delta} T h_T^{k(1+\delta)} \to \infty$ for some $\delta > 0$.

Assumption 4:

(a) Let $A_T = \{x \mid f_x(x) \leq b_T\}$,

$B_T = \{z \mid f_z(z) \leq b_T\}$, and

$C_T = \{z \mid f_x(x) > b_T \text{ and } f_z(z) \leq b_T\}$.

As $T \to \infty$, $\int_{A_T} V(y|x) f_x(x)\,dx = o_p(T^{-1/2})$,

$\int_{B_T} V(y|z) f_z(z)\,dz = o_p(T^{-1/2})$, and

$\int_{C_T} \{E(y|z) - E(y|x)\}^2 f_z(z)\,dz = o_p(T^{-1/2})$.

(b) Let $d_{ts} \equiv (2h_T^k)^{-1} K_{ts} \cdot \left\{ \frac{(1-w_t)\epsilon_t (g_s - g_t) I_t}{f_t} + \frac{(1-w_s)\epsilon_s (g_t - g_s) I_s}{f_s} \right\}$

$\mu_T \equiv E(d_{ts}) = E[E(d_{ts}|\xi_t)]$,

$U_T \equiv \binom{T}{2}^{-1} \sum_{t<s}^{T} d_{ts}$, and

$\hat{U}_T \equiv \mu_T + 2T^{-1} \sum_{t=1}^{T} (E(d_{ts}|\xi_t) - \mu_T) = 2T^{-1} \sum_{t=1}^{T} E(d_{ts}|\xi_t) - \mu_T$,

where $\xi \equiv (z, \epsilon)$, $K_{ts} \equiv K\left(\frac{z_t - z_s}{h_T}\right)$, $g_t \equiv E(y_t|z_t)$, $\epsilon_t \equiv y_t - g_t$, $f_t \equiv f_z(z_t)$ and $I_t \equiv 1(f_z(z_t) > b_T)$. U_T and $\hat{U}_T$ satisfy $U_T - \hat{U}_T = o_p(T^{-1/2})$.

Assumption 1 formulates conditions on the smoothness of the density and conditional mean functions with bounded moment conditions. Together with Assumption 2, the smoothness conditions are assumed for the asymptotic bias of the nonparametric kernel estimators to vanish quickly. The kernel function satisfying Assumption 2-(d) is called high order kernel. The way of constructing a high order kernel function from density function is suggested in Bierens(1987) and Haerdle(1990). Assumption 3 controls the uniform convergence rate of nonparametric kernel estimators. Note that with $0 < \delta < 1$, $\rho_T = O(T^{-1})$. Assumption 3-(a) and (b) are needed for Taylor expansion of kernel regression estimators, $\hat{E}(y_t|z_t)$ and $\hat{E}(y_t|x_t)$, and to make the bias of the statistic $\widetilde{M}_C$ to vanish at rate $\sqrt{T}$. Assumption 4-(a) formulates conditions on the conditional moments in the tails of the distribution, which is required for the consistency of the test statistic. Assumption 4-(b) assumes the asymptotic equivalence of U-statistic of order 2 and its projection holds for our test statistic under strong-mixing processes. This limiting behavior of U-statistic is well established under the absolute regularity; refer to Yoshihara(1976). To the best of our knowledge, a similar result for strong-mixing processes is not available yet. Research on this direction is an important extension in our view.

For notational simplicity, we introduce the following notation:

$$v_{Tt} \equiv \{y_{t+T/2} - E(y_{t+T/2}\,|z_{t+T/2})\}^2\, 1(f(z_{t+T/2} > b_T) - \{y_t - E(y_t|z_t)\}^2 1(f_z(z_t) > b_T) \text{ and} \tag{8}$$

$$\sigma^2 \equiv \lim_{T\to\infty} V\left((T/2)^{-1}\sum_{t=1}^{T/2} v_t\right). \tag{9}$$

Theorem 2.2 Under the Assumptions 1, 2, 3 and 4,

(a) $\widetilde{M}_C - \frac{1}{2}M_C = o_p(1)$

(b) $(2T)^{1/2}\sigma^{-1}\widetilde{M}_C \to N(0,1)$ in distribution.

Proof: see appendix.

Theorem 2.2 provides the basis of a causality test. Due to the unknown σ, however, it is infeasible. In order to construct a feasible test, we plug a consistent estimator for the asymptotic variance σ^2. We exploit existing results in the literature on heteroscedasticity and autocorrelation consistent estimation of covariance matrices, in particular the works of White(1984), Newey and West(1987), Hansen(1990) , Andrews(1991) and Jeong(1996). Rewrite σ^2 as

$$\sigma^2 = \lim_{T\to\infty} V\left((T/2)^{-1}\sum_{t=1}^{T/2} v_{Tt}\right)$$
$$= \lim_{T\to\infty}\left(E(v_{T0}^2) + 2\sum_{s=1}^{(T/2)-1} (T/2)^{-1/2}\sum_{t=s+1}^{T/2} E(v_{Tt}v_{T(t-s)})\right). \quad (10)$$

We consider the following estimator of σ^2:

$$s_{Tl}^2 = (T/2)^{-1}\sum_{t=1}^{T/2} \hat{v}_{Tt}^2 + 2\sum_{s=1}^{l} \omega\left(\frac{s}{l}\right)(T/2)^{-1}\sum_{t=s+1}^{T/2} \hat{v}_{Tt}\hat{v}_{T(t-s)}, \quad (11)$$

where

$$\hat{v}_{Tt} \equiv \{y_{t+T/2} - \hat{E}(y_{t+T/2}|z_{t+T/2})\}^2 1(\hat{f}(z_{t+T/2}) > b_T)$$
$$- \{y_t - \hat{E}(y_t|z_t)\}^2 1(\hat{f}(z_t) > b_T),$$

$\omega(\cdot)$ is a real-valued kernel defined below and $l \le T/2$ is a lag truncation parameter; the sample autocovariance functions $(T/2)^{-1}\sum_{t=s+1}^{T/2} \hat{v}_{Tt}\hat{v}_{T(t-s)}$ receive weight $\omega(\cdot)$ for lags of order $s \le l$, but zero weight for $s > l$. The motivation for this estimator is that for a stationary random sequence the long-run variance σ^2 equals 2π times the spectral density evaluated at zero and the estimators s_{Tl}^2 is the value at frequency zero of 2π times an estimator of the spectral density of v_{Tt}, where the kernel weights are used to smooth the sample autocovariance functions (Hansen(1990), Newey and West(1987)).

For the consistency of s_{Tl}^2, we need the additional conditions.

Assumption 5:
(a) $\Pr(f_z(z) \le b_T) = O(b_T)$.
(b) $\alpha(n) = O(n^{-\lambda})$, for $\lambda > \dfrac{4+2\delta}{\delta}$.
(c) $E[\{y - E(y|z)\}^{8(1+\delta)}] < \infty$.
(d) $b_T^{\delta/(1+\delta)} l^3 = o(1)$.

Assumption 6:
(a) $0 \le \omega(x) \le 1$ and $\omega(x) = \omega(-x)$ for all $x \in R$.
(b) $\omega(0) = 1$.
(c) $\omega(x)$ is continuous at zero and for almost all $x \in R$.
(d) $\int \omega(x)\,dx < \infty$.

(e) $(2\pi)^{-1}\int \omega(x)\exp(-ix\lambda)\,dx \geq 0$ for all $x \in R$.

Assumption 5-(a) is condition on the behavior of $f_z(z)$ in tails of the distribution, requiring the tails to be enough thin; a sufficient condition is a normal density. Assumption 5-(b) and (c) state mixing condition and a moment condition. Assumption 5-(d) allows the lag truncation parameter l to increase as T increases, but also controls the increasing rate. Assumption 6 specifies the class of kernels $\omega(\cdot)$. The condition of nonnegative kernel is technical. With additional notation, the condition can be replaced by the one that $|\omega(\cdot)| \leq 1$. As commented by Andrews(1991), Assumption 6-(b) and (c) reflect the fact that for lags small relative to T/2, one wants the weight given to sample autocovariance functions to be close to one. Assumption 6-(e) is required to generate nonnegativeness of estimators s^2_{Tl} in finite samples (Andrews (1991)). Examples of kernels satisfying Assumption 5 include Bartlett and Parzen kernels. Under these conditions, we can establish the consistency of s^2_{Tl} for σ^2.

Theorem 2.3 Under Assumptions 1 through 6, $s^2_{Tl} - \sigma^2 = o_p(1)$.
Proof: Since the proof is almost the same as in Jeong(1996), we omit the proof.

Replacing σ in Theorem 2.2-(b) by s_{Tl}, we complete the nonparametric test statistic for nonlinear causal relations.

Theorem 2.4 Under Assumptions 1 through 6, $(2T)^{1/2} s^{-1}_{Tl} \widetilde{M}_C \rightarrow N(0,1)$ in distribution, if the null is true.

It is obvious that the statistic explodes as $T \rightarrow \infty$ under the alternatives, or the test has the desired feature of consistency.

3. Conclusion

This paper develops a consistent nonparametric test for Granger-causality in mean by using the kernel method. Using sample-splitting and data-trimming techniques, we provide an asymptotically normally distributed test statistic under the null in the context of strong-mixing process. To the best of our knowledge, previously proposed nonparametric tests for causality normally assume stronger conditions such as absolute regularity. In addition, our test does not need any preliminary estimation of parametric models such as a linear VAR model, which is required in Hiemstra-Jones(1994) test, the most frequently applied one for nonlinear causality hypothesis testing in the literature.

We suspect that the proposed testing procedure would not have power against Pitman local alternatives though some other nonparametric tests do have in spite that they assume stronger dependence conditions. We expect it is possible to construct a test which has nontrivial power against it under the strong mixing condition. Research on these directions is currently under way.

References

1. Andrews, D.W.K. (1991), Heteroskedasticity and Autocorrelation Consistent Matrix Estimation, *Econometrica*, 59, 817-858
2. Athreya, K.B., and Pantula, S.G. (1986), A Note on Strong Mixing of ARMA Processes, *Statistics & Probability Letters*, 4, 187-190
3. Baek, E.G., and Brock, W.A. (1992), A General Test for Nonlinear Granger Causality: Bivariate model, Technical Report, Iowa State University and Wisconsin University-Madison.
4. Bierens, H.,J. (1983), Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions, *Journal of the American Statistical Association*, 78, 699-707.
5. Bierens, H. J. (1987), Kernel Estimators of Regression Functions, in: T. F. Bewley, eds, *Advances in Econometrics, New York*: Cambridge University Press, 99-144.
6. Doukhan, P. (1994), *Mixing: Properties and Examples*, Springer-Verlag
7. Granger, C. W. J (1988), Some Recent Development in a Concept of Causality, *Journal of Econometrics*, 39, 199-211.
8. Haerdle, W. (1990), *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.
9. Haerdle, W. and Stoker, T. (1989), Investigating Smooth Multiple Regression by the Method of Average Derivatives, *Journal of the American Statistical Association*, 84, 986-995
10. Hansen, B. E. (1990), Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes, Mimeographed, University of Rochester.
11. Hiemstra, C. and Jones, J.D. (1994), Testing for Linear and Nonlinear Granger Causality in the Stock Price Volume Relation, *the Journal of Finance*, 49, 1639-1664
12. Hoeffding, W. (1948), A Class of Statistics with Asymptotically Normal Distribution, *Annals of Mathematical Statistics*, 19, 293-325.
13. Ibragimov, I. A. and Linnik, Yu. V. (1971), *Independent and Stationary Sequence of Random Variables*, Groningen: Walters-Noordhoff.
14. Jeong, K. (1996), Kernel Estimators of 2π Times the Spectral Density at Frequency Zero of Nonparametric Regression Errors, *Journal of*

Economic Theory and Econometrics, 2, 125-135.
15. Marsry, E. and Tjostheim, D. (1995), Nonparametric Estimation and Identification of Nonlinear ARCH Time Series: Strong convergence and asymptotic normality, *Econometric Theory*, 11, 258-289.
16. Newey, W. K and West, K. (1987), A Simple, Positive Semidefinite, Heterosce- dasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
17. Robinson, P.M. (1989), Consistent Nonparametric Entropy-based Testing, Discussion Paper, No. 30, London School of Economics
18. White, H. (1984), *Asymptotic Theory for Econometricians*, New York: Academic Press.
19. Yoshihara, K. (1976), Limiting Behavior of U-statistics for Stationary, Absolutely Regular Processes, Z. *Wahrscheinlichkeitstheorie Verw. Gerbiete*, 35, 237-252.

Appendix

To prove Theorem 2.2, we need to establish the uniform consistency and the uniform convergence rate of nonparametric kernel estimators for regression function and marginal density function under the condition of strong-mixing processes, which are given in Lemma A. The results are well established in the literature. Therefore we omit the proof of Lemma A which follows standard lines of arguments. Details can be found in Jeong(1996), where the proof is based on the approach of Bierens(1983) in using Fourier transform of the kernel function.

Lemma A Suppose Assumption 1,2 and 3 hold. Denote Z as the domain of z and define $\theta_T = \max\left(h_T^{\lambda}, h_T^{k}\rho_T^{1/2}\right)$ and $D(z) = f_z(z)E(y|z)$. Then for every $\epsilon \in (0, sup_z f_z(z))$, we have

(a) $sup_z|\hat{f}_z(z) - f_z(z)| = O_p(\theta_T)$.

(b) $sup_z|\widehat{D}(z) - D(z)| = O_p(\theta_T)$.

(c) $sup_{\{f_z(z)\geq\epsilon\}}|\widehat{E}(y|z) - E(y|z)| = O_p(\theta_T)$.

Proof of Theorem 2.2-(a)

Recall that from equation (5) and (6)

$$M_C = E[\{y - E(y|x)\}^2] - E[\{y - E(y|z)\}^2]$$

$$\widetilde{M}_C = T^{-1}\sum w_t\left\{y_t - \widehat{E}(y_t|x_t)\right\}^2 1\left(\hat{f}_z(z_t) > b_T\right)$$

$$- T^{-1}\sum (1-w_t)\{y_t - \widehat{E}(y_t|z_t)\}^2 1(\hat{f}_z(z_t) > b_T).$$

From the intermediate results in the proof of Theorem 2.2-(b) given below, it is obvious that we have

$$\widetilde{M}_C - \frac{1}{2}E[\{(y-E(y|x))^2 - (y-E(y|z))^2\}1(f_z(z) > b_T)] = o_p(1).$$

Therefore, it is enough to show that

$$M_C - E[\{(y-E(y|x))^2 - (y-E(y|z))^2\}1(f_z(z) > b_T)] = o_p(1).$$

We have

$$\begin{aligned}
&\left| M_C - E[\{(y-E(y|x))^2 - (y-E(y|z))^2\}1(f_z(z) > b_T)]\right| \\
&= \left| E[\{(y-E(y|x))^2 - (y-E(y|z))^2\}1(f_z(z) \le b_T)]\right|
\end{aligned}$$

$$\begin{aligned}
= \Big| & E[(y-E(y|x))^2 1(f_x(x) \le b_T)] - E(y-E(y|z))^2 1(f_z(z) \le b_T) \\
&+ E[(y-E(y|x))^2\{1(f_z(z) \le b_T) - 1(f_x(x) \le b_T)\}]\Big| \\
\le\ & E[(y-E(y|x))^2 1(f_x(x) \le b_T)] && [*1] \\
&+ E[(y-E(y|z))^2 1(f_z(z) \le b_T)] && [*2] \\
&+ E[(y-E(y|x))^2 1(f_x(x) > b_T, f_z(z) \le b_T)] && [*3] \\
&+ E[(y-E(y|x))^2 1(f_x(x) \le b_T, f_z(z) > b_T)] && [*4]
\end{aligned}$$

By Assumption 4-(a), we have

$$[*1] = \int_{A_T} V(y|x) f_x(x)\,dx = o_p(T^{-1/2}),$$

$$[*2] = \int_{B_T} V(y|z) f_z(z)\,dz = o_p(T^{-1/2}),$$

$$\begin{aligned}
[*3] &= \int_{C_T} V(y|z) f_z(z)\,dz + \int_{C_T}\{E(y|z) - E(y|x)\}^2 f_z(z)\,dz \\
&\le \int_{B_T} V(y|z) f_z(z)\,dz + \int_{C_T}\{E(y|z) - E(y|x)\}^2 f_z(z)\,dz \\
&= o_p(T^{-1/2})
\end{aligned}$$

$$[*4] \le \int_{A_T} V(y|x) f_x(x)\,dx = o_p(T^{-1/2}).$$

Proof of Theorem 2.2-(b)

Denote

$$\widehat{\triangle}_x \equiv T^{-1}\sum w_t \left\{y_t - \hat{E}(y_t|x_t)\right\}^2 1(\hat{f}_z(z_t) > b_T),$$

$$\overline{\triangle}_x \equiv T^{-1}\sum w_t \left\{y_t - \hat{E}(y_t|x_t)\right\}^2 1(f_z(z_t) > b_T),$$

$$\triangle_x \equiv T^{-1}\sum w_t \{y_t - E(y_t|x_t)\}^2 1(f_z(z_t) > b_T),$$

$$\widehat{\triangle}_z \equiv T^{-1}\sum w_t \left\{y_t - \hat{E}(y_t|z_t)\right\}^2 1(\hat{f}_z(z_t) > b_T),$$

$$\overline{\triangle}_z \equiv T^{-1}\sum w_t \left\{y_t - \hat{E}(y_t|z_t)\right\}^2 1(f_z(z_t) > b_T),$$

$$\triangle_z \equiv T^{-1}\sum w_t \{y_t - E(y_t|z_t)\}^2 1(f_z(z_t) > b_T),$$

With the above notations, the proof of Theorem 2.2-(b) consists of 5 steps, summarized as

step 1: $\sqrt{T}\,(\widehat{\triangle}_x - \overline{\triangle}_x) = o_p(1)$

step 2: $\sqrt{T}\,(\overline{\triangle}_x - \triangle_x) = o_p(1)$,

step 3: $\sqrt{T}\,(\widehat{\triangle}_z - \overline{\triangle}_z) = o_p(1)$,

step 4: $\sqrt{T}\,(\overline{\triangle}_z - \triangle_z) = o_p(1)$, and

step 5: $(2T)^{1/2}\sigma^{-1}(\triangle_x - \triangle_z) \rightarrow N(0,1)$ in distribution.

Since the proofs of steps 1 and 2 have the same structure as the proofs of step 3 and 4, we omit the former. And the proof of step 3 follows the same line as step 4 of the proof of Theorem 3.1 in Heardle and Stocker(1989), so we can focus on the proof of step 4. We also omit the proof of step 5 which is simply derived using the central limit theorem for strong-mixing processes, for example, Theorem 18.5.3 of Ibragimov and Linnik(1971).

Simple algebraic manipulation gives

$$\overline{\triangle}_z - \triangle_z$$

$$= 2T^{-1}\sum (1-w_t)\{y_t - E(y_t|z_t)\}\left\{E(y_t|z_t) - \hat{E}(y_t|z_t)\right\}1(f_z(z_t) > b_T)$$

$$+ T^{-1}\sum (1-w_t)\left\{E(y_t|z_t) - \hat{E}(y_t|z_t)\right\}^2 1(f_z(z_t) > b_T). \quad \text{(A.1)}$$

Using Lemma A and Assumption 3, we have

$$\begin{aligned} sup_{\{f(z) \geq b_T\}} | \hat{E}(y|z) - E(y|z) |^2 &= O_p(\rho_T h_T^{-2k}) \\ &= O_p(T^{-1/2} T^{1/2} h_T^{-2k} \rho_T) \\ &= o_p(T^{-1/2}) . \end{aligned} \quad \text{(A.2)}$$

Thus equation (A.1) can be written as

$$\overline{\triangle}_z - \triangle_z$$

$$= 2T^{-1}\sum (1-w_t)\{y_t - E(y_t|z_t)\}\left\{E(y_t|z_t) - \hat{E}(y_t|z_t)\right\}1(f_z(z_t) > b_T)$$

$$+ o_p(T^{-1/2}). \quad \text{(A.3)}$$

Using step 1 in Haerdle and Stocker(1989), kernel estimator in equation (A.3) is asymptotically equivalent to a set of terms linear in sums over the data as follows,

$$\overline{\triangle}_z - \triangle_z$$

$$= 2T^{-1}\sum (1-w_t)\{y_t - E(y_t|z_t)\}f_z(z_t)^{-1}\left\{\hat{D}(z_t) - E(y_t|z_t)\hat{f}_z(z_t)\right\}$$

$$\cdot 1(f_z(z_t) > b_T) + o_p(T^{-1/2}) \quad \text{(A.4)}$$

where $D(z) \equiv f_z(z)E(y|z)$. In order to complete the proof, it suffices to show that the first term in the R.H.S. of equation (A.4) equals to $o_p(T^{-1/2})$. For notational simplicity, we define $K_{ts} \equiv K\left(\frac{z_t - z_s}{h_T}\right)$, $g_t \equiv E(y_t|z_t)$, $\epsilon_t \equiv y_t - g_t$, $f_t \equiv f_z(z_t)$ and $I_t \equiv 1(f_z(z_t) > b_T)$. Using these notations and the definitions of the nonparametric kernel estimators $\hat{f}_t$ and $\hat{D}_t$, the first term in the R.H.S. of equation (A.4) with "2" being omitted can be written as

$$T^{-1}\sum (1-w_t)\{y_t - E(y_t|z_t)\}f_z(z_t)^{-1}\left\{\widehat{D}(z_t) - E(y_t|z_t)\hat{f}_z(z_t)\right\}$$

$$= T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T} h_T^{-k}K_{ts}(1-w_t)\epsilon_t f_t^{-1}(g_s - g_t)I_t$$

$$+ T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T} h_T^{-k}K_{ts}(1-w_t)\epsilon_t f_t^{-1}\epsilon_s I_t. \tag{A.5}$$

Only the first term of the R.H.S. of equation (A.5) is considered, the second term being similarly analyzed. Because $g_t - g_s = 0$ for $t = s$, we have

$$T^{-2}\sum_{s=1}^{T}\sum_{t=1}^{T} h_T^{-k}K_{ts}(1-w_t)\epsilon_t f_t^{-1}(g_s - g_t)I_t$$

$$= T^{-2}\sum_{t\neq s}^{T} h_T^{-k}K_{ts}(1-w_t)\epsilon_t f_t^{-1}(g_s - g_t)I_t. \tag{A.6}$$

Defining $\xi \equiv (z, \epsilon)$, we can rewrite the R.H.S. of equation (A.6) as

$$T^{-2}\sum_{t\neq s}^{T} h_T^{-k}K_{ts}(1-w_t)\epsilon_t f_t^{-1}(g_s - g_t)I_t$$

$$= 2T^{-2}\sum_{t\neq s}^{T}(2h_T^{k})^{-1}K_{ts}\left\{\frac{(1-w_t)\epsilon_t(g_s-g_t)I_t}{f_t} + \frac{(1-w_s)\epsilon_s(g_t-g_s)I_s}{f_s}\right\}$$

$$\equiv 2T^{-2}\sum_{t<s}^{T} d_{ts} \tag{A.7}$$

Then the R.H.S. of equation (A.7) being $o_p(T^{-1/2})$ holds if the following two equalities hold;

$$T^{-1}\sum_{t=1}^{T} E(d_{ts}|\xi_t) = o_p(T^{-1/2}) \text{ and} \tag{A.8}$$

$$2T^{-2}\sum_{t<s}^{T} d_{ts} - 2T^{-1}\sum_{t=1}^{T} E(d_{ts}|\xi_t) = o_p(T^{-1/2}). \tag{A.9}$$

Define

$$\mu_T \equiv E(d_{ts}) = E[E(d_{ts}|\xi_t)],$$

$$U_T \equiv \binom{T}{2}^{-1}\sum_{t<s}^{T} d_{ts}, \text{ and}$$

$$\widehat{U}_T \equiv \mu_T + 2T^{-1}\sum_{t=1}^{T}(E(d_{ts}|\xi_t) - \mu_T) = 2T^{-1}\sum_{t=1}^{T} E(d_{ts}|\xi_t) - \mu_T.$$

Note that the R.H.S. of equation (A.7) is equivalent to U_T with the only difference in the leading factors of T^{-2} and $(T(T-1)^{-1})$ which are asymptotically negligible.

The remaining part of the proof is to show that equation (A.8) and (A.9) hold. We first consider the proof of equation (A.8). Using the definition of d_{ts} in equation (A.7), we have

$$E\left| T^{-1}\sum_{t=1}^{T} E(d_{ts}|\xi_t)\right| \le T^{-1}\sum_{t=1}^{T} E\left| h_T^{-k}E(K_{ts}(g_s-g_t)|\xi_t)\frac{(1-w_t)\epsilon_t I_t}{f_t}\right| + T^{-1}\sum_{t=1}^{T} E\left| h_T^{-k}E\left(K_{ts}\frac{(1-w_s)\epsilon_s(g_t-g_s)I_s}{f_s}|\xi_t\right)\right|. \quad \text{(A.10)}$$

We consider only the first term of the R.H.S. of inequality (A.10), the second term being similarly analyzed

$$T^{-1}\sum_{t=1}^{T} E\left| h_T^{-k}E(K_{ts}(g_s-g_t)|\xi_t)\frac{(1-w_t)\epsilon_t I_t}{f_t}\right| \le b_T^{-1}T^{-1}\sum_{t=1}^{T} E\left| E(h_T^{-k}K_{ts}(g_s-g_t)|\xi_t)\epsilon_t\right|. \quad \text{(A.11)}$$

Let us focus on the conditional expectation term in the R.H.S. of inequality (A.11);

$$E(h_T^{-k}K_{ts}(g_s-g_t)|\xi_t) \le M\left|\int K(u_T)\{g(z_t-h_Tu_T)-g(z_t)\}du_T\right|, \quad \text{(A.12)}$$

where $u_T \equiv (z_t-z_s)/h_T$ and M is an upper bound of conditional density of z_s given ξ_t. By Assumption 1, $g(\cdot)$ is λ-times differentiable with bounded λ^{th} derivatives. For j = 1,..,k, let us define

$$R_j \equiv \{(r_1,\ldots,r_k)|r_1+\ldots+r_k=j\} \text{ and}$$

$$\Pi_j(z) \equiv \sum_{R_j}\frac{\partial g(z)}{\partial z_1^{r_1}\ldots\partial z_k^{r_k}}.$$

Then using a Taylor's expansion, one can write $g(z_t-h_Tu_T)$ as follows,

$$g(z_t-h_Tu_T) = g(z_t) + A_1 + A_2, \quad \text{(A.13)}$$

where $A_1 \equiv \sum_{j=1}^{\lambda-1}(-1)^j\frac{h_T^j}{j!}\Pi_j(z_t)\sum_{R_j}u_{T_1}^{r_1}\ldots u_{T_k}^{r_k}$,

$$A_2 \equiv (-1)^\lambda\frac{h_T^\lambda}{\lambda!}\sum_{R_\lambda}u_{T_1}^{r_1}\ldots u_{T_k}^{r_k}\Pi_\lambda(z_t-\bar{h}_Tu_T),$$

and $\bar{h}_T$ lies between zero and h_T. Thus the R.H.S. of inequality (A.12) can be written as

$$
\begin{aligned}
M&\left| \int K(u_T)\{g(z_t - h_T u_T) - g(z_t)\} du_T \right| \\
&= M\left| \int K(u_T)\{A_1 + A_2\} du_T \right| \\
&= O(h_T^{\lambda}),
\end{aligned} \qquad \text{(A.14)}
$$

where the second equality follows from Assumption 2. Plugging equation (A.14) into the R.H.S. of equation (A.11) and inserting the result into the R.H.S. of equation (A.10), we have

$$
E\left| T^{-1} \sum_{t=1}^{T} E(d_{ts} | \xi_t \right| = O(b_T^{-1} h_T^{\lambda}) = o(T^{-1/2}), \qquad \text{(A.15)}
$$

where the last equality follows from Assumption 3. Thus we have the result of equation (A.8).

Now we turn to the proof of equation (A.9). By Assumption 4-(b), we have

$$
U_T - \hat{U}_T = o_p(T^{-1/2}). \qquad \text{(A.16)}
$$

Equation (A.8) implies that $\mu_T = E\{E(d_{ts} | \xi_t)\}$ is negligible asymptotically. Then the result of equation (A.9) directly follows from equation (A.16).

[received date : Nov. 2006, accepted date : Dec. 2006]