9-2014

Testing conditional independence via empirical likelihood

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Halbert WHITE
University of Southern California

Citation

SU, Liangjun and WHITE, Halbert. Testing conditional independence via empirical likelihood. (2014). *Journal of Econometrics*. 182, (1), 27-44.
Available at: https://ink.library.smu.edu.sg/soe_research/1570

Published in Journal of Econometrics, 2014 September, 182 (1) , 27-44.
http://dx.doi.org/10.1016/j.jeconom.2014.04.006

Testing Conditional Independence via Empirical Likelihood*

Liangjun Su[a] and Halbert White[b]
[a] School of Economics, Singapore Management University, Singapore
[b] Department of Economics, UCSD, La Jolla, CA 92093-0508

September 30, 2011

Abstract

We construct two classes of smoothed empirical likelihood ratio tests for the conditional independence hypothesis by writing the null hypothesis as an infinite collection of conditional moment restrictions indexed by a nuisance parameter. One class is based on the CDF; another is based on smoother functions. We show that the test statistics are asymptotically normal under the null hypothesis and a sequence of Pitman local alternatives. We also show that the tests possess an asymptotic optimality property in terms of average power. Simulations suggest that the tests are well behaved in finite samples. Applications to some economic and financial time series indicate that our tests reveal some interesting nonlinear causal relations which the traditional linear Granger causality test fails to detect.

Key words: Conditional independence, Empirical likelihood, Granger causality, Local bootstrap, Nonlinear dependence, Nonparametric regression, U-statistics.
JEL Classification: C12, C14, C22.

1 Introduction

Recently there has been a growing interest in testing the conditional independence (CI) of two random vectors Y and Z given a third random vector $X : Y \perp Z \mid X$. Linton and Gozalo (1997) propose two nonparametric tests of CI for independent and identically distributed (IID) variables based on generalized empirical distribution functions. Fernandes and Flores (1999) employ a generalized entropy measure to test CI but rely heavily on the choice of suitable weighting functions to avoid distributional degeneracy. Delgado and González-Manteiga (2001) propose an omnibus test of CI using the weighted difference of the estimated conditional distributions under the null and the alternative. Su and White (2007, 2008) consider testing CI by comparing conditional densities and conditional-characteristic-function-based moment conditions. de Maros and Fernandes (2007) and Chen and Hong (2010) propose nonparametric tests for the Markov property (a special case of CI) based on the comparison of densities and generalized

*We would like to express our appreciation to Qihui Chen for his outstanding research assistance. Address correspondence to: Halbert White, Department of Economics, UCSD, La Jolla, CA 92093-0508, USA. Phone: +1 858 534-3502; e-mail: hwhite@weber.ucsd.edu.

cross spectrums, respectively. Song (2009) studies an asymptotically pivotal test of CI via the probability integral transform. Huang (2010) proposes a test of CI based on the estimation of the maximal nonlinear conditional correlation. Huang and White (2010) develop a flexible test for CI based on the generically comprehensively revealing functions of Stinchcombe and White (1998). Spindler and Su (2010) consider testing for asymmetric information (a special case of conditional dependence) by comparing conditional distributions with both continuous and discrete variables. Bouezmarni, Rombouts and Taamouti (2010) and Bouezmarni, Roy and Taamouti (2011) propose tests for CI by comparing Bernstein copulas using the Hellinger distance and conditional distributions using the L_2-distance, respectively. Bergsma (2011) proposes a test for CI by means of the partial copula.[1]

In this paper, we propose two new classes of tests for CI based on empirical likelihood (EL). The motivation is as follows. First, the equality of two conditional distributions can be expressed in terms of an infinite sequence of conditional moment restrictions. Second, there are many powerful tests available in the literature to test for conditional moment restrictions, including EL-based tests. Third, EL has been shown to share some key properties with parametric likelihood such as Wilks' theorem and Bartlett correctability. Owen (1988, 1990, 1991) studies inference based on the nonparametric likelihood ratio, which is particularly useful in testing moment restrictions. Kitamura (2001) investigates the asymptotic efficiency of moment restriction tests for a finite number of unconditional moments in terms of large deviations and demonstrates the optimality of EL for testing such unconditional moment restrictions. Tripathi and Kitamura (2003, **TK** hereafter) extend the EL paradigm to test for a finite number of conditional moment restrictions and show that their test possesses an optimality property in large samples and behaves well in small samples. As yet, it remains unknown whether one can extend EL methods to test for an infinite collection of conditional moment restrictions, and, if so, whether the test continues to possess some optimality property and behaves reasonably well in finite samples. These issues are the focus of this paper.

The contributions of this paper lie primarily in four directions. First, we show that a smoothed empirical likelihood ratio (SELR) can be used to test hypotheses that can be expressed in terms of an infinite collection of conditional moment restrictions, indexed by a nuisance parameter, τ, say. Corresponding to each τ, one can construct a SELR. Then one obtains a test statistic by integrating τ out. After being appropriately centered and rescaled, the resulting test statistic is shown to be asymptotically distributed as $N(0,1)$ under the null. Second, we study the asymptotic distribution of the test statistic under a sequence of local alternatives and show that our test is asymptotically optimal in that it attains the maximum average local power with respect to a certain space of functions for the local alternatives. Third, unlike most work in the EL literature, including that of **TK**, our tests allow for data dependence and thus are applicable to time series data.[2] Fourth, our paper offers a convenient approach to testing distributional hypotheses via an infinite collection of conditional moment restrictions. It further extends the applicability of the EL method. A variety of interesting and important hypotheses other than CI in economics and finance, including conditional goodness-of-fit, conditional homogeneity, conditional

[1] Most of the aforementioned papers came out after the first version of this paper in 2003. Note that for categorical data, the literature traces back to Rosenbaum (1984) and Yao and Tritchler (1993).

[2] Chen, Härdle, and Li (2003) consider an EL goodness-of-fit test for time series. They integrate out the conditioning variable and employ only finite-dimensional parameter estimates in the constraints. This is quite different from our approach.

quantile restrictions, and conditional symmetry, can also be studied using our approach.

It is well known that distributional Granger non-causality (Granger, 1980) is a particular case of CI. Our tests can be directly applied to test for Granger non-causality with no need to specify a particular linear or nonlinear model. Using the same techniques as in Su and White (2008), it is also easy to show that our tests can be applied to the situation where not all variables of interest are continuously valued and some have to be estimated from the data. In particular, our tests apply to situations where limited dependent variables or discrete conditioning variables are involved, and to parametrically or nonparametrically generated regressors/residuals. For brevity, however, we only focus on the case where all random vectors are observed and continuously valued.

The remainder of this paper is organized as follows. In Section 2, we treat a simple version of our tests based on CDF's in order to lay out the basic framework for our SELR tests for CI. In Section 3, we study the asymptotic distributions of the test statistics under both the null hypothesis and a sequence of local alternatives, and show the asymptotic optimality of our tests in terms of average local power. We discuss a version of our SELR tests based on smoother moment conditions that has better finite sample power properties in Section 4. We examine the finite sample performance of our smoother SELR test via Monte Carlo simulations in Section 5, and we apply it to some macroeconomic and financial time series data in Section 6. Final remarks are contained in Section 7. All technical details are relegated to the Appendix.

2 Test statistic based on the CDFs

In this paper, we are interested in testing whether Y and Z are independent conditional on X, where X, Y and Z are vectors of dimension d_1, d_2 and d_3, respectively. The data consist of n identically distributed but weakly dependent observations $\{X_t, Y_t, Z_t\}_{t=1}^n$. For notational simplicity, we assume that $d_2 = 1$ throughout the paper.

2.1 Hypotheses

Let $f(x, y, z)$ and $F(x, y, z)$ denote the joint probability density function (PDF) and cumulative distribution function (CDF) of (X_t, Y_t, Z_t), respectively. Below we make reference to several marginal densities of f which we denote simply using the list of their arguments – for example $f(x, y) = \int f(x, y, z)dz$, $f(x, z) = \int f(x, y, z)dy$, and $f(x) = \int\int f(x, y, z)dydz$. This notation is compact, and, we hope, sufficiently unambiguous. Let $f(\cdot|\cdot)$ denote the conditional density of one random vector given another. We assume that $f(y|x, z)$ is smooth in (x, z). Let $1(\cdot)$ be the usual indicator function, $F(\tau|x, z) \equiv E[1(Y_t \leq \tau)|X_t = x, Z_t = z]$ and $F(\tau|x) \equiv E[1(Y_t \leq \tau)|X_t = x]$. The null of interest is that conditional on X, the random vectors Y and Z are independent, i.e.,

$$\mathbb{H}_0 : \Pr\left[F(\tau|X_t, Z_t) = F(\tau|X_t)\right] = 1 \text{ for all } \tau \in \mathbb{R}. \tag{2.1}$$

The alternative hypothesis is that for τ with a nontrivial volume of the support of Y_t,

$$\mathbb{H}_1 : \Pr\left[F(\tau|X_t, Z_t) = F(\tau|X_t)\right] < 1. \tag{2.2}$$

In Section 4, we consider another approach based on a related condition involving the characteristic function. We treat $\mathbb{H}_0$ first because of its intuitive appeal.

2.2 Test statistics

Noting that $\mathbb{H}_0$ specifies an infinite collection of conditional moment restrictions that are indexed by τ : $E[\varepsilon_t(\tau)|X_t, Z_t] = 0$ a.s. for all $\tau \in \mathbb{R}$ where $\varepsilon_t(\tau) = 1(Y_t \le \tau) - F(\tau|X_t)$, we can test $\mathbb{H}_0$ by testing a single conditional moment restriction for given τ,

$$\mathbb{H}_0(\tau): \ \Pr[F(\tau|X_t, Z_t) = F(\tau|X_t)] = 1 \tag{2.3}$$

based on the EL principle, and obtain the final test statistic by integrating out τ. Due to the use of integration, it is computationally expensive to calculate the statistic. Therefore we also consider a weaker version of the CI hypothesis:

$$\mathbb{H}_0': \ \Pr[F(Y_t|X_t, Z_t) = F(Y_t|X_t)] = 1, \tag{2.4}$$

which is implied by (2.1). Spindler and Su (2010) and Bouezmarni, Roy and Taamouti (2011) independently propose a L_2-distance-based test for (2.4) by comparing the weighted difference between the nonparametric kernel estimates of $F(Y_t|X_t, Z_t)$ and $F(Y_t|X_t)$.

To proceed, we first consider a SELR test statistic for $\mathbb{H}_0(\tau)$. Let $p_{ts} \equiv p_{(Y_s;X_t,Z_t)}$ denote the probability mass placed at $(Y_s; X_t, Z_t)$ by a discrete distribution with support $\{Y_s\}_{s=1}^n \times \{(X_t, Z_t)\}_{t=1}^n$. Let $\hat{\varepsilon}_s(\tau) \equiv 1(Y_s \le \tau) - \hat{F}_{h_2}(\tau|X_s)$, where

$$\hat{F}_{h_2}(\tau|x) \equiv n^{-1} \sum_{t=1}^{n} L_{h_2}(x - X_t) 1(Y_t \le \tau)/\hat{f}_{h_2}(x), \tag{2.5}$$

$\hat{f}_{h_2}(x) \equiv n^{-1} \sum_{t=1}^n L_{h_2}(x - X_t)$, $L_{h_2}(u) \equiv h_2^{-d_1} L(u/h_2)$, L is a kernel function defined on $\mathbb{R}^{d_2}$, and $h_2 \equiv h_2(n)$ is a bandwidth sequence. We consider the following restricted (i.e., under $\mathbb{H}_0(\tau)$) maximization problem:

$$\max_{\{p_{ts}, t,s=1,\dots,n\}} \sum_{t=1}^{n} \sum_{s=1}^{n} w_{ts} \log p_{ts}, \ \text{s.t.} \ p_{ts} \ge 0, \sum_{t=1}^{n} \sum_{s=1}^{n} p_{ts} = 1, \sum_{s=1}^{n} \hat{\varepsilon}_s(\tau) p_{ts} = 0, \tag{2.6}$$

where $w_{ts} \equiv K_{h_1}(X_t - X_s, Z_t - Z_s)/\sum_{s=1}^n K_{h_1}(X_t - X_s, Z_t - Z_s)$ is a kernel weight such that $\sum_{s=1}^n w_{ts} = 1$ for each t, $K_{h_1}(u) \equiv h_1^{-(d_1+d_3)} K(u/h_1)$, and the last constraint in (2.6) imposes a sample analog of the constraint in $\mathbb{H}_0(\tau)$. Note that we use different kernels and bandwidth sequences in (2.5) and (2.6). Intuitively speaking, using a higher order kernel in (2.5) helps to reduce the bias in estimating $F(\tau|x)$, whereas a second order positive kernel is needed in (2.6) to keep the estimator of p_{ts} nonnegative almost surely when the sample size n goes to ∞. For future use, we denote the Nadaraya-Watson (NW) kernel estimator of $F(\tau|x, z)$ by $\hat{F}_{h_1}(\tau|x, z) \equiv n^{-1} \sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t) 1(Y_t \le \tau)/\hat{f}_{h_1}(x, z)$, where $\hat{f}_{h_1}(x, z) \equiv n^{-1} \sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t)$. (2.6) is solved by maximizing the Lagrangian $\mathcal{L} \equiv \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log p_{ts} - \mu(\sum_{t=1}^n \sum_{s=1}^n p_{ts} - 1) - \sum_{t=1}^n \lambda_t \sum_{s=1}^n [1(Y_s \le \tau) - \hat{F}_{h_2}(\tau|X_s)] p_{ts}$, where μ and $\{\lambda_t \in \mathbb{R}, \ t = 1, ..., n\}$ are the Lagrange multipliers for the second and third constraints, respectively. One can verify that the solution to this problem is given by $\hat{p}_{ts} = w_{ts}/[n + \lambda_t(\tau) \hat{\varepsilon}_s(\tau)]$, where $\lambda_t(\tau)$ solves

$$\sum_{s=1}^{n} \frac{w_{ts} \hat{\varepsilon}_s(\tau)}{n + \lambda_t(\tau) \hat{\varepsilon}_s(\tau)} = 0, \ \ t = 1, ..., n. \tag{2.7}$$

Under $\mathbb{H}_0(\tau)$, we have the following restricted smoothed empirical likelihood (SEL)

$$SEL^r(\tau) = \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \hat{p}_{ts} = \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \left\{ \frac{w_{ts}}{n + \lambda_t(\tau)\hat{\varepsilon}_s(\tau)} \right\}.$$

For the unrestricted problem we solve

$$\max_{\{p_{ts}, t,s=1,\ldots,n\}} \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log p_{ts}, \text{ s.t. } p_{ts} \geq 0, \sum_{t=1}^n \sum_{s=1}^n p_{ts} = 1.$$

It is well known that the solution to this problem is $\tilde{p}_{ts} = w_{ts}/n$, and the unrestricted SEL is $SEL^u(\tau) \equiv \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \tilde{p}_{ts} = \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \left\{ \frac{w_{ts}}{n} \right\}$. Thus, we obtain a SELR test statistic

$$2[SEL^u(\tau) - SEL^r(\tau)] = 2\sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \{1 + \lambda_t(\tau)\hat{\varepsilon}_s(\tau)/n\}. \tag{2.8}$$

Clearly, $SEL^u(\tau) - SEL^r(\tau)$ is small if $\mathbb{H}_0(\tau)$ holds and large otherwise, and one can test $\mathbb{H}_0(\tau)$ based on (2.8). Nevertheless, for technical reasons, we follow **TK** and use a modified version of (2.8) for our test to accommodate the fact that conditional distributions cannot be estimated as precisely in the tails as desired. We thus define the SELR as

$$SELR(\tau) = 2\sum_{t=1}^n I_t \sum_{s=1}^n w_{ts} \log \{1 + \lambda_t(\tau)\hat{\varepsilon}_s(\tau)/n\},$$

where $I_t \equiv 1\{(X_t, Z_t) \in S\}$ and S is fixed subset of the support of (X_t, Z_t). One rejects $\mathbb{H}_0(\tau)$ for large values of $SELR(\tau)$.

To test $\mathbb{H}_0$, we integrate out τ using researcher-specified weights. Specifically, our test statistic is

$$ISELR_n = \int SELR(\tau)dG(\tau) = 2\sum_{t=1}^n I_t \sum_{s=1}^n \int w_{ts} \log \{1 + \lambda_t(\tau)\hat{\varepsilon}_s(\tau)/n\} dG(\tau), \tag{2.9}$$

where $G(\cdot)$ is a CDF with support on $\mathbb{R}$.

Remark 1. The idea for the above testing procedure is intuitively clear. Because we have an infinite number of conditional moment conditions, it seems impossible to handle them once and for all in a single maximization problem without resorting to some regularization techniques of the sort used by Carrasco (2010). We thus consider one conditional moment restriction at a time to obtain the SELR, and then aggregate them to obtain a single test statistics. As mentioned above, **TK** demonstrated that the SELR-based test statistic possesses an asymptotic optimality when it is used in testing a finite number of conditional moment restrictions. As we show, a similar property is inherited by our integrated SELR-based test.

Remark 2. Because of the computational burden of the $ISELR_n$ test, we also consider a SELR-based test for $\mathbb{H}_0'$ in (2.4). Simulations in Spindler and Su (2010) and Bouezmarni, Roy and Taamouti (2011) demonstrate that their L_2-distance-based tests for $\mathbb{H}_0'$ possess reasonable asymptotic power in finite samples. In this case, the restricted (i.e., under $\mathbb{H}_0'$) maximization problem becomes

$$\max_{\{p_{ts}, t,s=1,\ldots,n\}} \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log p_{ts}, \text{ s.t. } p_{ts} \geq 0, \sum_{t=1}^n \sum_{s=1}^n p_{ts} = 1, \sum_{s=1}^n \hat{\varepsilon}_s(Y_t) p_{ts} = 0. \tag{2.10}$$

Using the notation defined above, we define the following SELR test statistic

$$SELR_{n1} = 2\sum_{t=1}^{n} I_t I_{1t} \sum_{s=1}^{n} w_{ts} \log\left\{1 + \lambda_t\left(Y_t\right)\hat{\varepsilon}_s(Y_t)/n\right\}, \tag{2.11}$$

where $I_{1t} \equiv 1\left\{Y_t \in S_Y\right\}$, and S_Y is a fixed set contained in the interior of the support of Y_t. The use of I_{1t} in (2.11) helps to trim out some extreme values Y_t such that $F\left(Y_t|x,z\right)$ is close to 0 or 1 for certain $(x,z) \in S$. See Remark 5 below for more detail. We will study the asymptotic properties of $SELR_{n1}$ and demonstrate that it also possesses an asymptotic optimality property for testing $\mathbb{H}_0'$.

3 The asymptotic distributions of the test statistics

3.1 Assumptions

To derive the asymptotic distribution of our test statistic, we impose the following assumptions.

Assumption A1 (Stochastic Process). (i) $\{W_t \equiv (X_t', Y_t, Z_t')' \in \mathbb{R}^{d_1+1+d_3}, t \geq 0\}$ is a strictly stationary strong mixing process with mixing coefficients $\alpha(\cdot)$ such that $\sum_{s=0}^{\infty} s^4 \alpha(s)^{\delta/(1+\delta)} \leq C$ for some $\delta > 0$ with $\delta/(1+\delta) \leq 1/2$, and $\alpha(s)^{(2+\tilde{\delta})/[3(4+\tilde{\delta})]} = O\left(s^{-1}\right)$ and $\alpha(s)^{\tilde{\delta}/(2+\tilde{\delta})} = O\left(s^{-2+\epsilon}\right)$ for some $\tilde{\delta} \in (0,\delta)$ and sufficiently small $\epsilon > 0$.

(ii) The PDF f of W_t has continuous partial derivatives up to order $r \geq 2$ which are bounded and integrable on $\mathbb{R}^d$. f satisfies a Lipschitz condition: $|f(w+u) - f(w)| \leq D(w)||u||$ where D has finite $(2+\eta)th$ moment for some $\eta > 0$ and $||\cdot||$ is the usual Euclidean norm. $\inf_{(x,z)\in S^{\epsilon}} f(x,z) = b > 0$, where $S^{\epsilon} \equiv \{u \in \mathbb{R}^{d_1+d_3} : \|u - v\| \leq \epsilon \text{ for some } v \in S\}$ for some small positive ϵ.

(iii) The joint PDF $f_{t_1,\ldots,t_j}$ of $(W_0, W_{t_1}, ..., W_{t_j})$ $(1 \leq j \leq 5)$ is uniformly bounded.

(iv) $F(\tau|x)$ is $(r+1)$ times partially continuously differentiable with respect to x for each $\tau \in \mathbb{R}$ and the partial derivatives up to the $(r+1)$th are bounded on $S_1^{\epsilon} \equiv S^{\epsilon} \cap \mathbb{R}^{d_1}$. Furthermore, $|F(\tau|x') - F(\tau|x)| \leq \alpha(\tau)\,||x' - x||$, where $\int \alpha^2(\tau)dG(\tau) < \infty$.

Assumption A2 (Kernels). (i) The kernel K is a product kernel of $k : K(u_1, .., u_{d_1+d_3}) = \Pi_{i=1}^{d_1+d_3} k(u_i)$, where $k : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable PDF that is symmetric about the origin and has compact support $[-1, 1]$.

(ii) The kernel L is a product kernel of $l : L(u_1, .., u_{d_1}) = \Pi_{i=1}^{d_1} l(u_i)$, where $l : \mathbb{R} \to \mathbb{R}$ is r times continuously differentiable such that $\int_{\mathbb{R}} u^i l(u)du = \delta_{i0}$ $(i = 0, 1, ..., \ r-1)$, $C_0 \equiv \int_{\mathbb{R}} u^r l(u)du < \infty$, $\int_{\mathbb{R}} u^2 l(u)^2 du < \infty$, and $l(u) = O((1 + |u|^{r+1+\eta})^{-1})$ for some $\eta > 0$, where δ_{ij} is Kronecker's delta.

Assumption A3 (Bandwidths). (i) The bandwidth sequences $h_1 = O(n^{-1/\alpha_1})$ and $h_2 = O(n^{-1/\alpha_2})$ are such that $\alpha_1 > 2(d_1 + d_3)$, $\max\{2\alpha_1(d_1-2)/(2\alpha_1 + d_1 + d_3), \alpha_1 d_1/(d_1 + d_3)\} < \alpha_2 < \alpha_1 r \max\{1, 4/(d_1 + d_3)\}$, $(d_1 + d_3)/\alpha_1 + d_1/\alpha_2 < 1$, and $(d_1 + d_3)/(2\alpha_1) + 2r/\alpha_2 > 1$.

(ii) $(n^{1+4\gamma_0+3\delta_1} + n^{2+\delta_1})\alpha(n^{\delta_2})^{\delta_0/(4+\delta_0)} = o(1)$ for some δ_0, δ_1 and δ_2 such that $0 < \delta_2 < \delta_1 < (1 + \delta_0/6)/(4 + \delta_0)$ where $\gamma_0 = (2 + \delta_0)/(16 + 4\delta_0)$ and $\delta_0 > 0$.

(iii) There exists a diverging sequence $\{\iota_n\}$ such that as $n \to \infty$, $\iota_n^3 n^{-1} \to 0$, $\iota_n h_1^{(d_1+d_3)/2} \to 0$ and $\max\{nh_1^{-(1-3\delta)(d_1+d_3)/[2(1+\delta)]}, n^2\}h_1^{(1-3\delta)(d_1+d_3)/(1+\delta)} h_2^{-2d_1\delta/(1+\delta)} \alpha\left(\iota_n\right)^{\delta/(1+\delta)} \to 0$.

Assumption A4 (Weight function). The CDF G satisfies $\int_S \int V(x,z;\tau)^{-1} f(x,z)\, dG(\tau)\, d(x,z) < \infty$, where $V(x,z;\tau) \equiv F(\tau|x,z)[1 - F(\tau|x,z)]$.

Assumption A1(i) requires that $\{W_t\}$ be a strong mixing process with algebraic decay rate. It is weaker than the absolute regularity assumed in Su and White (2007, 2008), Bouezmarni, Rombouts and Taamouti (2010), and Bouezmarni, Roy and Taamouti (2011), who further assume a geometric decay rate. Assumptions A1(ii)-(iv) are primarily smoothness conditions, some of which can be relaxed at the cost of additional technicalities. Assumption A2(i) requires that the kernel K be of second order and compactly supported, whereas Assumption A2(ii) requires that the kernel L be of r-th order. The compact support of K can be relaxed with some additional technicalities. Assumption A3 specifies conditions on the choice of bandwidth sequences. Under A3(i), we have in particular that $nh_1^{2(d_1+d_3)}/(\ln n)^3 \to \infty$, $nh_1^{d_1+d_3} h_2^{d_1} \to \infty$, $nh_1^{-(d_1+d_3)/2} h_2^{d_1-2} \to \infty$, $h_1^{-(d_1+d_3)/2} h_2^{2r} \to 0$, $h_1 h_2^{-1} \to 0$, and $nh_1^{(d_1+d_3)} h_2^{2r} \to 0$. When the dimension of (X_t, Z_t) is low, e.g., $d_1 + d_3 \leq 4$, $r = 2$ will suffice for well chosen α_1 and α_2. The conditions in A3(ii)-(iii) are automatically satisfied if one assumes that the mixing coefficient $\alpha(\cdot)$ has geometric or faster decay rate. Assumption A4 can be met if one restricts the density of G to have compact support, or its tails to decay to zero sufficiently fast.

3.2 Asymptotic null distribution

Let $V(x,z;\tau,\tau') \equiv F(\tau \wedge \tau'|x,z) - F(\tau|x,z)F(\tau'|x,z)$ where $\tau \wedge \tau' = \min(\tau,\tau')$. Let $\hat{V}(x,z;\tau,\tau') \equiv \hat{F}_{h_1}(\tau \wedge \tau'|x,z) - \hat{F}_{h_1}(\tau|x,z)\hat{F}_{h_1}(\tau'|x,z)$, $\hat{V}(x,z;\tau) \equiv \hat{F}_{h_1}(\tau|x,z)[1 - \hat{F}_{h_1}(\tau|x,z)]$, $\hat{V}_1(X_t, Z_t;\tau) = \sum_{s=1}^n w_{ts} \times [\hat{\varepsilon}_s(\tau)]^2$, and $\hat{f}_{1t} = \hat{f}_{h_1}(X_t, Z_t)$. Define

$$\begin{aligned}
\hat{B}_n &\equiv \sum_{t=1}^n I_t \sum_{s=1}^n \int \hat{V}_1(X_t, Z_t;\tau)^{-1} [w_{ts}\hat{\varepsilon}_s(\tau)]^2 dG(\tau), \\
\hat{\sigma}_n^2 &\equiv 2n^{-1} C_3^{d_1+d_3} \sum_{t=1}^n I_t \int\int \hat{V}(X_t, Z_t;\tau)^{-1} \hat{V}(X_t, Z_t;\tau')^{-1} \hat{V}(X_t, Z_t;\tau,\tau')^2 \hat{f}_{1t}^{-1} dG(\tau) dG(\tau'), \\
\hat{T}_n &\equiv h_1^{(d_1+d_3)/2} \{ISELR_n - \hat{B}_n\}/\hat{\sigma}_n,
\end{aligned}$$

where $C_3 \equiv \int \left[\int k(u+v)k(u)du\right]^2 dv$. For any given univariate kernel satisfying Assumption A2(i), C_3 can be calculated explicitly. If we use the Gaussian kernel for $k(\cdot)$,[3] then $C_3 = 1/(2\sqrt{2\pi})$. If we use the Epanechnikov kernel instead, i.e., $k(u) = 0.75(1-u^2)1(|u| \leq 1)$, then $C_3 = 0.4338$.

We can now state our first main result.

Theorem 3.1 *Under Assumptions A1-A4 and* $\mathbb{H}_0$, $\hat{T}_n \xrightarrow{d} N(0,1)$.

Remark 3. Theorem 3.1 relies on a central limit theorem (CLT) for second order U-statistics with strong mixing observations; this is adapted from Tenreiro (1997), who proved a CLT for second order U-statistics with β-mixing observations. When proving the CLT, Tenreiro (1997) relies on certain inequalities for β-mixing processes. It turns out that similar inequalities also hold for α-mixing processes. See Lemma A.3 in the appendix. Noting that the $\hat{T}_n$ test is one-sided, we reject the null for large values of $\hat{T}_n$.

[3] While the Gaussian kernel does not have compact support, it can be approximated arbitrarily well by kernels that satisfy all the conditions in Assumption A2(i). See Ahn (1997, p.13).

Remark 4. Define $B \equiv C_1^{d_1+d_3} \int_S 1d(x,z) = C_1^{d_1+d_3}\text{vol}(S)$ and $\sigma^2 \equiv 2C_3^{d_1+d_3} \int_S \int \int V(x,z;\tau)^{-1} V(x,z;\tau)^{-1} V^2(x,z;\tau,\tau')dG(\tau)dG(\tau')d(x,z)$, where $C_1 \equiv \int k(u)^2 du$. The dominant term of $\hat{B}_n$ is $h_1^{-(d_1+d_3)}B$, which implies that the asymptotic bias of $ISELR_n$ shares the same structure as that in **TK**. But similar structure is not shared by the asymptotic variance, because we cannot write σ^2 as something like $2C_3^{d_1+d_3}\text{vol}(S)$, as in **TK,** due to the presence of an infinite number of moment conditions in (2.1).

Next, let

$$\begin{aligned}
\hat{B}_{n1} &\equiv \sum_{t=1}^n I_t I_{1t} \hat{V}_{1t}^{-1} \sum_{s=1}^n [w_{ts}\hat{\varepsilon}_s(Y_t)]^2, \\
\hat{\sigma}_{n1}^2 &\equiv 2n^{-2}C_3^{d_1+d_3} \sum_{t=1}^n \sum_{s=1}^n K_{ts} I_t I_{1t} I_{1s} \hat{V}_t^{-1} \hat{V}_s^{-1} \hat{V}(X_t,Z_t;Y_t,Y_s)^2 \hat{f}_{1t}^{-1} \hat{f}_{1s}^{-1}, \\
\hat{T}_{n1} &\equiv h_1^{(d_1+d_3)/2}\{SELR_{n1} - \hat{B}_{n1}\}/\hat{\sigma}_{n1},
\end{aligned}$$

where $\hat{V}_{1t} \equiv \hat{V}_1(X_t,Z_t;Y_t)$, $\hat{V}_t \equiv \hat{V}(X_t,Z_t;Y_t)$, and $K_{ts} \equiv K_{h_1}(X_s - X_t, Z_s - Z_t)$. The next result provides the asymptotic null distribution of $\hat{T}_{n1}$ for testing $\mathbb{H}_0'$ in (2.4).

Theorem 3.2 *Suppose that* $\int_S \int_{S_Y} V(x,z;y)^{-1} f(x,z)\, dF(y|x,z)\, d(x,z) < \infty$. *Then under Assumptions A1-A3 and* $\mathbb{H}_0'$, $\hat{T}_{n1} \xrightarrow{d} N(0,1)$.

Remark 5. Note that Assumption A4 is now replaced by the requirement that $\int_S \int_{S_Y} V(x,z;y)^{-1} \times f(x,z)\, dF(y|x,z)\, d(x,z) < \infty$ in Theorem 3.2, and one cannot take $S_Y = \mathbb{R}$. To see why, using a change of variable gives

$$\begin{aligned}
\int_S \int_{\mathbb{R}} V(x,z;y)^{-1} f(x,z)\, dF(y|x,z)\, d(x,z) &= \int_S \int_0^1 \frac{1}{u(1-u)} du f(x,z)\, d(x,z) \\
&= c_f \int_0^1 \frac{1}{u(1-u)} du = \infty
\end{aligned}$$

where $c_f \equiv \int_S f(x,z)\, d(x,z)$ and the last line follows from the fact that the antiderivative of $1/[u(1-u)]$ is given by $\ln[u/(1-u)]$.

Remark 6. Like $ISELR_n$, Theorem 3.2 indicates that $SELR_{n1}$ shares the same bias structure as the test statistic in **TK.** The leading term in the asymptotic expansion of $\hat{B}_{n1}$ is given by $h_1^{-(d_1+d_3)}B_1$ where $B_1 = C_1^{d_1+d_3}\text{vol}(S)\text{vol}(S_Y)$, compared to $C_1^{d_1+d_3}\text{vol}(S)$ in **TK**. If we could take $S_Y = \mathbb{R}$, then Theorem 3.2 also implies that $SELR_{n1}$ shares the same asymptotic variance structure as in **TK.** It is easy to see that the probability limit of $\hat{\sigma}_{n1}^2$ is given by

$$\sigma_1^2 \equiv 2C_3^{d_1+d_3} \int_S \int_{S_Y} \int_{S_Y} V(x,z;y)^{-1} V(x,z;y')^{-1} V(x,z;y,y')^2 dF(y|x,z)\, dF(y'|x,z)\, d(x,z).$$

In the case where $S_Y = \mathbb{R}$, we have by the change of variables that

$$\begin{aligned}
\sigma_1^2 &= 4C_3^{d_1+d_3} \int_S \int \int_{y>y'} \frac{[1-F(y|x,z)]\, F(y'|x,z)}{F(y|x,z)[1-F(y'|x,z)]} dF\,(y|x,z)\, dF\,(y'|x,z)\, d(x,z) \\
&= 4C_3^{d_1+d_3} \int_S \int_0^1 \frac{u'}{1-u'} \int_{u'}^1 \frac{1-u}{u} du du' d(x,z) \\
&= 4C_3^{d_1+d_3} \mathrm{vol}\,(S) \int_0^1 \frac{u'}{1-u'} \left(u' - 1 - \ln u'\right) du' \\
&= 4C_3^{d_1+d_3} \mathrm{vol}\,(S) \left[-\frac{1}{2} - \int_0^1 \frac{u \ln u}{1-u} du \right] = 2C_3^{d_1+d_3} \mathrm{vol}\,(S) \left(\frac{\pi^2}{3} - 3 \right)
\end{aligned}$$

where the last equality follows from the fact that $\int_0^1 \frac{u \ln u}{1-u} du = 1 - \frac{\pi^2}{6}$. Thus, σ_1^2 is distinct from **TK**'s asymptotic variance only in the scaling factor $(\frac{\pi^2}{3} - 3)$.

3.3 Asymptotic local power properties

To derive the asymptotic power function of $\hat{T}_n$ under a sequence of local alternatives, we consider the triangular array process $\{W_{nt} \equiv (X_{nt}', Y_{nt}, Z_{nt}')' \in \mathbb{R}^{d_1+1+d_3},\ t = 1, ..., n,\ n = 1, 2, ...\}$. Let $f^{[n]}(x, y, z)$ and $F^{[n]}(x, y, z)$ denote the PDF and CDF of (X_{nt}, Y_{nt}, Z_{nt}), respectively. Let E_n denote expectation under the probability law associated with $f^{[n]}$. Define $F^{[n]}(\tau|x, z) \equiv E_n[1(Y_{nt} \leq \tau)|X_{nt} = x, Z_{nt} = z]$ and $F^{[n]}(\tau|x) \equiv E_n[1(Y_{nt} \leq \tau)\ |X_{nt} = x]$. We consider the following sequence of Pitman local alternatives:

$$\mathbb{H}_1(\gamma_n) : \sup_{\tau \in \mathbb{R}} \sup_{(x,z) \in \mathbb{R}^{d_1+d_3}} \left\{ |F^{[n]}(\tau|x,z) - F^{[n]}(\tau|x) - \gamma_n \Delta(x,z;\tau)| \right\} = o(\gamma_n), \tag{3.1}$$

where $\gamma_n \to 0$ as $n \to \infty$, $\Delta(x, z; \tau)$ satisfies $\mu \equiv \lim_{n\to\infty} \int_S \int V^{[n]}\,(x, z; \tau)^{-1}\, \Delta(x, z; \tau)^2 dG(\tau) dF^{[n]}(x, z) < \infty$ for the test of $\mathbb{H}_0$ in (2.1), and $\mu_1 \equiv \lim_{n\to\infty} \int_S \int V^{[n]}\,(x, z; y)^{-1}\, \Delta(x, z; y)^2 dF^{[n]}(y|x, z) dF^{[n]}(x, z) < \infty$ for the test of $\mathbb{H}_0'$ in (2.4); $V^{[n]}\,(x, z; \tau) \equiv F^{[n]}(\tau|x, z)[1 - F^{[n]}(\tau|x, z)]$; and $F^{[n]}(x, z) = F^{[n]}(x, \infty, z)$. For simplicity, we assume that $\lim_{n\to\infty} F^{[n]}(x, y, z) = F(x, y, z)$ for all (x, y, z).

Following Su and White (2010), we define the mixing coefficients: $\alpha_n\,(j) = \sup_{1\leq l\leq n-j}\{P\,(A \cap B)$ $-P(A)P(B)|A \in \sigma\,(W_{nt} : 1 \leq t \leq l)\,,\, B \in \sigma\,(W_{nt} : l + j \leq t \leq n)\}$ if $j \leq n - 1$, and $\alpha_n\,(j) = 0$ if $j \geq n$. Define the coefficient of strong mixing as $\alpha\,(j) = \sup_{n\in\mathbb{N}} \alpha_n\,(j)$ for $j \in \mathbb{N}$ and $\alpha\,(0) = 1$. We modify Assumption 1 as follows.

Assumption A1*. The triangular process $\{W_{nt}\}$ is a strictly stationary strong mixing process with mixing coefficients $\alpha\,(\cdot)$ satisfying the condition in Assumption A1 (i). Assumptions A1(ii)-(iv) are satisfied for $\{W_{nt}\}$ with the obvious modifications, e.g., with $f^{[n]}$ and $F^{[n]}$ replacing f and F, respectively.

The following two propositions study the asymptotic local power properties of $\hat{T}_n$ and $\hat{T}_{n1}$ for the tests of $\mathbb{H}_0$ and $\mathbb{H}_0'$, respectively.

Proposition 3.3 *Let Assumptions A1* and A2-A4 hold. Let* $\gamma_n = n^{-1/2} h_1^{-(d_1+d_3)/4}$ *in* $\mathbb{H}_1(\gamma_n)$. *Then* $\Pr(\hat{T}_n \geq z|\mathbb{H}_1(\gamma_n)) \to 1 - \Phi(z - \mu/\sigma)$.

Proposition 3.4 *Suppose that* $\int_S \int_{S_Y} V(x, z; y)^{-1} f\,(x, z)\, dF\,(y|x, z)\, d\,(x, z) < \infty$, *and Assumptions A1*, A2 and A3 hold. Let* $\gamma_n = n^{-1/2} h_1^{-(d_1+d_3)/4}$ *in* $\mathbb{H}_1(\gamma_n)$. *Then* $\Pr(\hat{T}_{n1} \geq z|\mathbb{H}_1(\gamma_n)) \to 1 - \Phi(z - \mu_1/\sigma_1)$.

3.4 An asymptotic optimality property

Motivated by **TK,** we now consider an asymptotic optimality property associated with the SELR tests. Following the approaches of Su and White (2007) and Bouezmarni, Roy and Taamouti (2011), one can consider two sequences of test statistics that are respectively based upon

$$\hat{\Gamma}(a) \equiv \frac{1}{n}\sum_{s=1}^{n}\int [\hat{F}_{h_1}(\tau|X_s,Z_s) - \hat{F}_{h_2}(\tau|X_s)]^2 a(X_s,Z_s;\tau)dG(\tau), \tag{3.2}$$

and

$$\hat{\Gamma}_1(a) \equiv \frac{1}{n}\sum_{s=1}^{n}[\hat{F}_{h_1}(Y_s|X_s,Z_s) - \hat{F}_{h_2}(Y_s|X_s)]^2 a(X_s,Z_s;Y_s), \tag{3.3}$$

indexed by the weight function a defined on $S \times \mathbb{R}$.[4] Let $\eta(a)$ and $\eta_1(a)$ denote the corresponding normalized test statistics that are asymptotically $N(0,1)$ under $\mathbb{H}_0$ and $\mathbb{H}_0'$, respectively. Then we can show that their asymptotic local power functions under $\mathbb{H}_1(\gamma_n)$ are given respectively by

$$\begin{aligned}\Pi(a,\Delta) &= \lim_{n\to\infty}\Pr\left(\eta(a) > z \mid \mathbb{H}_1(\gamma_n)\right) = 1-\Phi(z - M(a,\Delta)),\\ \Pi_1(a,\Delta) &= \lim_{n\to\infty}\Pr\left(\eta_1(a) > z \mid \mathbb{H}_1(\gamma_n)\right) = 1-\Phi(z - M_1(a,\Delta)),\end{aligned}$$

where $\gamma_n = n^{-1/2}h_1^{-(d_1+d_3)/4}$,

$$M(a,\Delta) \equiv \frac{\int_S\int \Delta(x,z;\tau)^2 f(x,z)a(x,z;\tau)dG(\tau)d(x,z)}{\sqrt{2C_3^{d_1+d_3}\int_S\int\int V(x,z;\tau,\tau')^2 a(x,z;\tau)a(x,z;\tau')dG(\tau)dG(\tau')d(x,z)}}, \text{ and} \tag{3.4}$$

$$M_1(a,\Delta) \equiv \frac{\int_S\int \Delta(x,z;y)^2 f(x,z)a(x,z;y)dF(y|x,z)d(x,z)}{\sqrt{2C_3^{d_1+d_3}\int_S\int\int V(x,z;y,y')^2 a(x,z;y)a(x,z;y')dF(y|x,z)dF(y'|x,z)d(x,z)}}. \tag{3.5}$$

Comparing the above power functions with Propositions 3.3 and 3.4, we can show that $\hat{T}_n$ is asymptotically equivalent to the $\eta(a)$ test with the weighting function $a(x,z;\tau) = 1\{(x,z) \in S\}V(x,z;\tau)^{-1}$, and $\hat{T}_{n1}$ is asymptotically equivalent to the $\eta_1(a)$ test with the weighting function $a_1(x,z;y) = 1\{(x,z) \in S\}V(x,z;y)^{-1}1(y \in S_Y)$. We will show that these choices of weighting functions, which are implicitly achieved by our SELR tests, are asymptotically optimal in a certain sense.

If Δ were known, it would be easy to derive the optimal weighting function a that maximizes Π or Π_1. Clearly, such a weighting function would depend on the unknown object Δ, and no uniformly (in Δ) optimal test exists. This resembles the multi-parameter optimal testing problem considered by Wald (1943), who shows that the likelihood ratio test for a hypothesis about finite-dimensional parameters is optimal in terms of an average power criterion. Similarly, Andrews and Ploberger (1994) consider optimal inference in a nonstandard testing problem where a nuisance parameter is present only under the alternative. Our testing problem is a nonparametric analogue of Andrews and Ploberger's (1994). In their case, the parameter of interest in the sequence of local alternatives is of finite dimension (h in their notation), whereas the parameter of interest in our local alternatives is an unknown function (i.e., $\Delta(\cdot,\cdot;\cdot)$ in the above notation). A natural extension of Wald's approach is to consider a probability measure on an

[4] Note that both Su and White (2007) and Bouezmarni, Roy and Taamouti (2011) only allow the weight function a to depend on (x,z).

appropriate space of functions and let the measure mimic the distribution of the estimator $\tilde{F}_{b_1}(\tau|x,z)$.[5] Therefore, we follow the lead of **TK** and propose to use a probability measure that approximates the asymptotic distribution of the sample path of $\tilde{F}_{b_1}(\tau|x,z)$.

Let $\{\Omega, \mathcal{F}, \text{P}\}$ be a probability space. Let $\tilde{\Delta}(x,z,\tau) \equiv \Delta((x,z);\tau;\omega) : S \times \mathbb{R} \times \Omega \to \mathbb{R}$ be a random function,[6] i.e., for arbitrary and fixed (x,z,τ), $\Delta((x,z);\tau;\,\cdot)$ is a measurable mapping of $\{\Omega,\mathcal{F}\}$ into $\{\mathbb{R},\mathcal{B}\}$ where $\mathcal{B}$ is the Borel sigma-field on $\mathbb{R}$ and for fixed ω, $\Delta(\cdot;\cdot;\omega)$ is a function. Next let $\tilde{\Delta}(x,z,\tau) = f^{-1/2}(x,z)V^{1/2}(x,z;\tau)\Psi(x,z)$, where for $v \equiv (x,z) \in \mathbb{R}^{d_1+d_3}$, $\Psi(v) \equiv \Pi_{i=1}^{d_1+d_3} \int_0^{1/\gamma_i} \kappa_i(v_i/\gamma_i - z)dU_i(z)$, the κ_i are arbitrary cyclical univariate kernel functions on $\mathbb{R}$ with period $1/\gamma_i$, and the U_i are mutually independent Brownian motions on $[0,1/\gamma_i]$ starting at the origin such that $E[U_i(1/\gamma_i)]^2 < \infty$ for each i. Let l_i be the diameter of S restricted in the direction of v_i.[7] We further require $0 < 1/\gamma_i \leq l_i$. As in **TK**, this implies that the joint distribution of the bivariate vector $(\int_S s(\Psi(v))dv,\, \Psi(v_0))$ does not depend on the location $v_0 \in S$ for any function s such that $\int_S s(\Psi(v))dv$ is well defined, the Gaussianity of $\tilde{\Delta}$ is not important, and our optimality result does not depend on the choice of v_i and γ_i.

For simplicity, we now explicitly study Π_1. To define our average power, let Q be the probability measure induced by $\tilde{\Delta}$ on continuous functions defined on $S \times \mathbb{R}$. Noting that $\Pi_1(a,\tilde{\Delta}) = \Pi_1(ca,\tilde{\Delta})$ for any $c > 0$, we choose a such that

$$\int_S \int \int V^2(x,z;y,y')a(x,z;y)a(x,z;y')dF(y|x,z)dF(y'|x,z)d(x,z) = 1. \tag{3.6}$$

Using the definition of $\tilde{\Delta}$, we can then rewrite the random variable $M_1(a,\tilde{\Delta})$ as

$$M_1(a,\tilde{\Delta}) = \left(2C_3^{d_1+d_3}\right)^{-1/2} \int_S V\,(x,z;y)\,a(x,z;y)\Psi^2(x,z)dF(y|x,z)d(x,z). \tag{3.7}$$

Let F_a be the CDF of $M_1(a,\tilde{\Delta})$. The average asymptotic power of η_1 is then given by

$$\bar{\Pi}_1(a) = \int \Pi_1(a,\tilde{\Delta})dQ(\tilde{\Delta}) = \int_0^{\infty} [1-\Phi(z-e)]F_a(de). \tag{3.8}$$

Observe that the integrand in (3.8) is strictly increasing in e. So if there exists a smooth, bounded, square integrable function $a^* : S \times \mathbb{R} \to \mathbb{R}_+$ such that (3.6) is satisfied for a^* and for all a the CDF F_{a^*} first order stochastically dominates F_a, then a^* maximizes $\bar{\Pi}_1\,(a)$. Following **TK**, we have:

Proposition 3.5 *Let* $a_1^*\,(x,z;y) \equiv 1\{(x,z) \in S\}V\,(x,z;y)^{-1}\,[vol(S)\,(\pi^2/3-3)]^{-1/2}$. *Then* $a_1^* = \arg \max_{a \in C_b(S\times\mathbb{R})} \bar{\Pi}_1\,(a)$, *where* $C_b(S \times \mathbb{R})$ *is the space of continuously bounded functions on* $S \times \mathbb{R}$.

Remark 7. Proposition 3.5 shows that the $\hat{T}_{n1}$ test attains the maximum average local power when the sequence of alternatives are restricted to the space of functions generated by $\tilde{\Delta}$, provided one ignores the restriction that $y \in S_Y$. The last restriction seems necessary in order for $V\,(x,z;y)^{-1}$ to be well defined for all $(x,z) \in S$. An alternative way of achieving this optimality is to use $a(x,z;y) = 1\{(x,z) \in S\}V\,(x,z;y)^{-1}\,1\,\{y \in S_Y\}$ in (3.3), and in practice one has to replace $V\,(x,z;y)$ by its consistent estimate. Similarly, we can establish the asymptotic optimality of the $\hat{T}_n$ test for testing $\mathbb{H}_0$.

---

[5] It is unnecessary to mimic the distribution of $\tilde{F}_{b_2}(\tau|x)$ because this has no impact on the asymptotic power function.

[6] Alternatively, a random function can be defined by specifying a suitable measure on a certain function space whose elements are functions on $S \times \mathbb{R}$. See, for example, Gihman and Skorohod (1974, p.44).

[7] Without loss of generality, one can assume $S = [-e,e]^{d_1+d_3}$, where e is a positive real number. In this case, $l_i = 2e$ for each i.

4 Smoother moment conditions

In this section we extend our testing procedure to permit a smoother family of conditional moment restrictions.

4.1 Characteristic function-based conditional moment restrictions

Above, we considered a SELR test for CI based on the infinite sequence of conditional moment restrictions: $E[1(Y_t \leq \tau) - F(\tau|X_t)|X_t, Z_t] = 0$ a.s., indexed by the nuisance parameter τ. This choice for the conditional moment restrictions is intuitive but typically delivers poor power in finite samples because of the discrete nature of the indicator functions. Motivated by the equivalence of conditional distributions and conditional characteristic functions, we now follow Bierens (1982) and Su and White (2007) and consider a smoother class of conditional moment restrictions. For this, let $H(y) \equiv \int e^{iu'y} dG_0(u)$, the characteristic function of a well-chosen probability measure $dG_0(u)$. For example, if $G_0(\cdot)$ is the standard normal CDF, then $H(y) = e^{-y^2/2}$. Let $\psi(u;x,z) \equiv E[\exp(iu'Y_t)|X_t = x, Z_t = z] - E[\exp(iu'Y_t)|X_t = x]$. Then $\int \psi(u;x,z)e^{i\tau' u} dG_0(u) = E[H(Y_t+\tau)|X_t = x, Z_t = z] - E[H(Y_t+\tau)|X_t = x] \equiv m(x,z;\tau) - m(x;\tau)$. Under a mild assumption, $\mathbb{H}_0$ and $\mathbb{H}_0'$ can be respectively expressed as

$$\begin{aligned} \mathbb{H}_{0,\mathrm{chf}} &: \ \Pr[m(X_t, Z_t; \tau) = m(X_t; \tau)] = 1 \text{ for all } \tau \in \mathbb{R}, \text{ and} \\ \mathbb{H}_{0,\mathrm{chf}}' &: \ \Pr[m(X_t, Z_t; Y_t) = m(X_t; Y_t)] = 1. \end{aligned}$$

Therefore we can formulate a variant of our preceding test statistics based upon

$$\widetilde{ISELR}_n = 2\sum_{t=1}^n I_t \sum_{s=1}^n \int w_{ts} \log\left\{1 + \tilde{\lambda}_t(\tau)\tilde{\varepsilon}_s(\tau)/n\right\} dG(\tau), \tag{4.1}$$

$$\widetilde{SELR}_{n1} = 2\sum_{t=1}^n I_t I_{1t} \sum_{s=1}^n w_{ts} \log\left\{1 + \tilde{\lambda}_t(Y_t)\tilde{\varepsilon}_s(Y_t)/n\right\}, \tag{4.2}$$

where $\tilde{\varepsilon}_s(\tau) \equiv H(Y_s + \tau) - \tilde{m}(X_s;\tau)$, $\tilde{m}(x;\tau) \equiv \sum_{s=1}^n L_{h_2}(x - X_s)H(Y_s+\tau)/\sum_{s=1}^n L_{h_2}(x - X_s)$, and each $\tilde{\lambda}_t(\tau)$ solves

$$\sum_{s=1}^n \frac{w_{ts}\tilde{\varepsilon}_s(\tau)}{n + \tilde{\lambda}_t(\tau)\tilde{\varepsilon}_s(\tau)} = 0, \ t = 1, ..., n. \tag{4.3}$$

Let $\tilde{V}_1(X_t, Z_t; \tau) \equiv n^{-1}\sum_{s=1}^n w_{ts}[\tilde{\varepsilon}_s(\tau)]^2$, $\tilde{V}_{1t} \equiv \tilde{V}_1(X_t, Z_t; Y_t)$, and $K_{(x,z),t} \equiv K_{h_1}(x - X_t, z - Z_t)$. Let $\tilde{V}(x,z;\tau) \equiv n^{-1}\sum_{s=1}^n K_{(x,z),s}H(Y_s+\tau)^2/\hat{f}_{h_1}(x,z) - \hat{H}(x,z;\tau)^2$, $\tilde{V}_t \equiv \tilde{V}(X_t, Z_t; Y_t)$, $\tilde{V}(x,z;\tau,\tau') \equiv \hat{H}(x,z;\tau,\tau') - \hat{H}(x,z;\tau)\hat{H}(x,z;\tau')$, $\hat{H}(x,z;\tau) = n^{-1}\sum_{s=1}^n K_{(x,z),s}H(Y_s+\tau)/\hat{f}_{h_1}(x,z)$, and $\hat{H}(x,z;\tau,\tau')$ $= n^{-1}\sum_{s=1}^n K_{(x,z),s}H(Y_s+\tau)H(Y_s+\tau')/\hat{f}_{h_1}(x,z)$. Define

$$\begin{aligned} \tilde{B}_n &\equiv \sum_{t=1}^n I_t \sum_{s=1}^n \int \tilde{V}_1(X_t, Z_t; \tau)^{-1}[w_{ts}\tilde{\varepsilon}_s(\tau)]^2 dG(\tau), \ \tilde{B}_{n1} \equiv \sum_{t=1}^n I_t I_{1t} \sum_{s=1}^n \tilde{V}_{1t}^{-1}[w_{ts}\tilde{\varepsilon}_s(Y_t)]^2, \\ \tilde{\sigma}_n^2 &\equiv 2n^{-1}C_3^{d_1+d_3} \sum_{t=1}^n I_t \int\int \tilde{V}(X_t, Z_t; \tau)^{-1}\tilde{V}(X_t, Z_t; \tau')^{-1}\tilde{V}^2(X_t, Z_t, \tau, \tau')\hat{f}_{1t}^{-1} dG(\tau)dG(\tau'), \\ \tilde{\sigma}_{n1}^2 &\equiv 2n^{-2}C_3^{d_1+d_3} \sum_{t=1}^n \sum_{s=1}^n I_t I_{1t} I_{1s} K_{ts} \tilde{V}_t^{-1}\tilde{V}_s^{-1}\tilde{V}(X_t, Z_t; Y_t, Y_s)^2 \hat{f}_{1t}^{-1}\hat{f}_{1s}^{-1}. \end{aligned}$$

Further, define

$$\tilde{T}_n \equiv h_1^{(d_1+d_3)/2}\{\widetilde{ISELR}_n - \tilde{B}_n\}/\tilde{\sigma}_n \text{ and } \tilde{T}_{n1} \equiv h_1^{(d_1+d_3)/2}\{\widetilde{SELR}_{n1} - \tilde{B}_{n1}\}/\tilde{\sigma}_{n1}. \tag{4.4}$$

We add the following assumptions.

Assumption A5 (Fourier transform) Suppose $dG_0(u) = g_0(u)du$. $g_0(u)$ is a symmetric and uniformly bounded PDF on $\mathbb{R}$.

Assumption A1(iv^*) The function $m(x;\tau)$ is $(r+1)$ times partially continuously differentiable with respect to x for each $\tau \in \mathbb{R}$, and the partial derivatives up to the $(r+1)$th order are bounded on $S_1^\epsilon \equiv S^\epsilon \cap \mathbb{R}^{d_1}$. Furthermore, $|m(x';\tau) - m(x;\tau)| \leq D(\tau)\,||x' - x||$, where $\int D(\tau)^2 dG(\tau) < \infty$.

The following two theorems establish the asymptotic distributions of $\tilde{T}_n$ and $\tilde{T}_{n,1}$ under $\mathbb{H}_{0,\text{chf}}$ and $\mathbb{H}'_{0,\text{chf}}$, respectively.

Theorem 4.1 *Suppose Assumption A4 holds with V being replaced by $\bar{V}(x,z;\tau) \equiv Var[H\,(Y_t+\tau)\,|\,(X_t,Z_t) = (x,z)]$. Then under Assumptions A1(i)-(iii), A1(iv*), A2, A3 and A5, and $\mathbb{H}_{0,chf}$, $\tilde{T}_n \xrightarrow{d} N(0,1)$.*

Theorem 4.2 *Suppose that $\int_S \int_{S_Y} \bar{V}(x,z;y)^{-1} f\,(x,z)\, dF\,(y|x,z)\, d\,(x,z) < \infty$. Then under Assumptions A1(i)-(iii), A1(iv*), A2, A3 and A5, and $\mathbb{H}'_{0,chf}$, $\tilde{T}_{n1} \xrightarrow{d} N(0,1)$.*

Remark 8. Our simulations indicate that the above smoother SELR tests generally outperform the CDF-based SELR tests studied in Section 3 for a variety of data generating processes (DGPs), and the $\tilde{T}_{n1}$ test is much easier to implement than $\tilde{T}_n$. So we will report simulation and application results based on $\tilde{T}_{n1}$. In addition, as in Sections 3.3 and 3.4, we can also derive the asymptotic local power properties of $\tilde{T}_n$ and $\tilde{T}_{n1}$, and establish their asymptotic optimality for testing $\mathbb{H}_{0,\text{chf}}$ and $\mathbb{H}'_{0,\text{chf}}$, respectively. The procedure is analogous and thus omitted.

4.2 Remarks

The above results apply when all variables in (X_t, Y_t, Z_t) are continuously valued. While this is sufficient for many empirical applications, it is worth mentioning that our testing procedure can be easily modified to allow a much wider range of situations. Also, it is well known many nonparametric tests based on asymptotic normal critical values behave poorly in finite samples, and a bootstrap approximation improves matters. We now discuss these issues.

1. *Limited dependent variables and discrete conditioning variables*: As mentioned in the introduction, our tests are also applicable to situations in which not all variables in (X_t, Y_t, Z_t) are continuously valued. For example, when Y_t is discretely valued, $\hat{T}_n$ and $\tilde{T}_n$ can be easily modified by replacing the integration with summation over the possible values of Y_t;there is no need to modify either $\hat{T}_{n1}$ or $\tilde{T}_{n1}$. Also, one can allow a mixture of continuous and discrete conditioning variables. The modification can be done by following the approaches in Li and Racine (2008) and Spindler and Su (2010).

2. *Testing for independence*: It is possible to extend our procedure to the case where $d_1 = 0$, i.e., testing for independence between Y and Z. In this case, the null hypothesis reduces to $\mathbb{H}_0^* : \Pr\left[F(\tau|Z_t) = F(\tau)\right] = 1$, where $F(\tau|z)$ and $F(\tau)$ denote the conditional and unconditional CDFs of Y_t given $Z_t = z$, respectively.

One can modify our previous procedure by replacing $\hat{F}_{h_2}(\tau|X_s)$ in the definition of $\hat{\varepsilon}_s(\tau)$ by $m(\tau) \equiv n^{-1}\sum_{t=1}^{n} 1(Y_t \leq \tau)$ or $\tilde{m}(X_s;\tau)$ in the definition of $\tilde{\varepsilon}_s(Y_t)$ by $\tilde{m}(\tau) \equiv n^{-1}\sum_{t=1}^{n} H(Y_t+\tau)$ and making corresponding changes. For brevity, we don't repeat the argument.

3. *Smoothed local bootstrap*: The key issue for the bootstrap is how to impose the null hypothesis of CI in the resampling scheme. Motivated by Paparoditis and Politis (2000), Su and White (2008) propose a smoothed local bootstrap procedure for testing CI. Simply put, we obtain the bootstrap resamples $\{X_t^*, Y_t^*, Z_t^*\}_{t=1}^n$ in two steps: (i) Draw a bootstrap sample $\{X_t^*\}_{t=1}^n$ from the estimated kernel density $\tilde{f}_b(x) = n^{-1}b^{-d_1}\sum_{t=1}^{n}\kappa((X_t - x)/b)$, where κ is kernel function defined on $\mathbb{R}^{d_1}$ ($\mathbb{R}$, or $\mathbb{R}^{d_3}$ below as indicated by its argument) and $b = b(n)$ is a bandwidth; (ii) For $t = 1, ..., n$, given X_t^*, draw Y_t^* and Z_t^* independently from the estimated conditional density $\tilde{f}(y|X_t^*) = n^{-1}b^{-d_1-1}\sum_{s=1}^{n}\kappa((Y_s - y)/b)\,\kappa((X_s - X_t^*)/b)/\tilde{f}_b(X_t^*)$ and $\tilde{f}(z|X_t^*) = n^{-1}b^{-d_1-d_3}\sum_{s=1}^{n}\kappa((Z_s - z)/b)$ $\kappa((X_s - X_t^*)/b)/\tilde{f}_b(X_t^*)$, respectively. Let $W_t^* \equiv (X_t^{*\prime}, Y_t^*, Z_t^{*\prime})'$. Then we can calculate the bootstrap statistic $\tilde{T}_{n1}^*$ in analogous fashion to $\tilde{T}_{n1}$, with $\{W_t^*\}_{t=1}^n$ replacing $\{W_t\}_{t=1}^n$ and reject the null when the bootstrap p-value is smaller than the prescribed level of significance α. One can follow Su and White (2008) to establish the asymptotic validity of this bootstrap procedure.

5 Simulations

In this section we conduct some Monte Carlo simulations to examine the finite sample performance of our nonparametric test based on $\tilde{T}_{n1}$ in (4.4). We consider three cases, where $d_2 = d_3 = 1$, and $d_1 = 1$, 2, 3 in the first, second and third cases, respectively. For each data generating process (DGP) under study, we standardize the data $\{(X_t', Y_t, Z_t')',\ t = 1, ..., n\}$ before implementing our test, so that each variable has mean zero and variance one. Throughout, we take $S \equiv \{u = (x', z')' : |u_i| \leq 2.33,\ i = 1, ..., d_1 + d_3\}$ and $S_Y \equiv \{y : |y| \leq 2.33\}$, where u_i denotes the ith element of u. For each variable, this trims out about 2% of the tail observations, given a normal distribution.

We use the following DGPs for the first case:

DGP1: $W_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})'$, where $\{\varepsilon_{1,t},\ \varepsilon_{2,t},\ \varepsilon_{3,t}\}$ are IID $N(0, I_3)$.

For DGP2 through DGP8, $W_t = (Y_{t-1}, Y_t, Z_{t-1})'$, with

DGP2: $Y_t = 0.5Y_{t-1} + \varepsilon_{1,t}$;

DGP3: $Y_t = 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{1,t}$;

DGP4: $Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}^2 + \varepsilon_{1,t}$;

DGP5: $Y_t = 0.5Y_{t-1}Z_{t-1} + \varepsilon_{1,t}$;

DGP6: $Y_t = 0.5Y_{t-1} + 0.55Z_{t-1}\varepsilon_{1,t}$;

DGP7: $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2$; and

DGP8: $Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}$, $h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + 0.5Z_{t-1}^2$;

where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$ in DGPs 2-7, $Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}$ with $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$ in DGP8, and $\{\varepsilon_{1,t},\ \varepsilon_{2,t}\}$ are IID $N(0, I_2)$ in DGPs 2-8.

DGP1 and DGP2 allow us to examine the level of the test. DGPs 3-8 cover a variety of linear and nonlinear time series processes commonly used in time series analysis. Of these, DGPs 3-5 (resp. DGPs 6-8) are alternatives that allow us to study the power properties of our test for Granger-causality in the mean (resp. variance). DGP3 studies Granger linear causality in the mean, whereas DGPs 4-5 study Granger nonlinear causality in the mean. In DGPs 6-8, $\{Z_t\}$ Granger-causes $\{Y_t\}$ only through the variance. A conditional mean-based Granger causality test, linear or nonlinear, may fail to detect such causality. Note that DGP7 is an ARCH-type specification, and DGP8 specifies a bivariate GARCH process. Consequently, the study of such processes indicates whether our test may be applicable to financial time series.

To implement our test, we also need to choose the kernels, weight functions, and bandwidths. We follow Theorem 4.2 and use a fourth order kernel in estimating $f(x)$ and $m(x;\tau)$: $l(u) = (3-u^2)\varphi(u)/2$, where $\varphi(u)$ is the standard normal PDF. We choose both the second order kernel $k(\cdot)$ and weighting function $g_0(\cdot)$ to be $\varphi(u)$. For the two bandwidth sequences, h_1 and h_2, we follow Su and White (2007), set $h_1 = c_1 n^{-1/(4+d_1+d_3)}$ and $h_2 = c_2 n^{-1/(8+d_1)}$ and choose c_1 and c_2 by using the leave-one-out least squares cross validation (LSCV) for estimating the conditional expectations of Y_t given (X_t, Z_t) and X_t respectively. We denote our test statistic as SEL_n below.

We compare our test with some previous tests proposed by Linton and Gozalo (1997, **LG** hereafter) and by Su and White (2007, 2008). **LG** base their nonparametric tests of CI on the functional $A_n(w) = \{n^{-1}\sum_{t=1}^n 1(W_t \le w)\}\{n^{-1}\sum_{t=1}^n 1(X_t \le x)\} - \{n^{-1}\sum_{t=1}^n 1(X_t \le x)1(Y_t \le y)\}\{n^{-1}\sum_{t=1}^n 1(X_t \le x)1(Z_t \le z)\}$, where $w = (x', y, z')'$. Specifically, their test statistics are of the Cramér von-Mises and Kolmogorov-Smirnov types: $CM_n = n^{-1}\sum_{t=1}^n A_n^2(W_t)$, $KS_n = \max_{1\le t\le n}|A_n(W_t)|$. The asymptotic null distributions of both test statistics are non-standard, so that one needs to use bootstrap to obtain the critical values. Su and White (2008) base a test for CI on the Hellinger distance between the two conditional densities $f(y|x,z)$ and $f(y|x)$. They use the same bandwidth sequence h in estimating all the required densities, namely, $f(x,y,z)$, $f(x,y)$, $f(x,z)$ and $f(x)$, by $\hat f_h(x,y,z)$, $\hat f_h(x,y)$, $\hat f_h(x,z)$ and $\hat f_h(x)$. Define $\Gamma_1 \equiv \frac{1}{n}\sum_{t=1}^n [1-\sqrt{\hat f_h(X_t,Y_t)\hat f_h(X_t,Z_t)/\hat f_h(X_t,Y_t,Z_t)\hat f_h(X_t)}]^2 a(X_t,Y_t,Z_t)$, where a is a weighting function that is compactly supported. Their test statistic, HEL_n, is a normalized version of Γ_1 and is asymptotically normally distributed under the null. Following Su and White (2008), we set $h = n^{-1/8.5}$. Su and White's (2007) test is based upon a property of the conditional characteristic function. Let $\hat m_{b_1}(x,z;\tau)$ and $\hat m_{b_2}(x;\tau)$ be nonparametric kernel estimates for $m(x,z;\tau) \equiv E[H(Y+\tau)|X=x, Z=z]$ and $m(x;\tau) \equiv E[H(Y+\tau)|X=x]$ with bandwidth sequences b_1 and b_2, respectively. Let $\Gamma_2 = \frac{1}{n}\sum_{t=1}^n \int |\hat m_{b_1}(X_t,Z_t;\tau) - \hat m_{b_2}(X_t;\tau)|^2 I_t dG(\tau)$. Their test statistic, CHF_n, is based on Γ_2 and is also asymptotically normally distributed under the null. Here we use $b_1 = h_1$ and $b_2 = h_2$.

Table 1 reports the empirical rejection frequency of the five tests, namely, CM_n, KS_n, HEL_n, CHF_n and SEL_n, for nominal size 5%. Given the computational burden, we consider three sample sizes: n = 100, 200 and 400. We use 500 replications for all tests when the null is true and 250 replications when the null is false. We apply 200 bootstrap resamples in each replication. From Table 1, we see that the sizes of all five tests are reasonably well-behaved, despite the fact that the HEL_n test is moderately oversized for small sample sizes. In terms of power, we observe that the three kernel-based tests HEL_n, CHF_n and SEL_n tend to be more powerful than **LG**'s tests CM_n and KS_n. Interestingly, HEL_n has the largest power in detecting alternatives in DGPs 7-8. SEL_n outperforms the other tests in terms of power in DGPs 3-6 but is slightly outperformed by CHF_n and HEL_n in DGPs 7-8.

For the second case ($d_1 = 2,\ d_2 = d_3 = 1$), we use the following DGPs:

Table 1: Comparison of tests for causality ($d_1 = d_2 = d_3 = 1$), nominal level: 0.05

n	Tests	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
100	CM_n	0.042	0.036	0.872	0.440	0.444	0.468	0.216	0.172
	KS_n	0.036	0.042	0.692	0.276	0.356	0.348	0.152	0.140
	HEL_n	0.094	0.090	0.624	0.752	0.400	0.908	0.836	0.608
	CHF_n	0.034	0.058	0.780	0.792	0.520	0.780	0.728	0.580
	SEL_n	0.054	0.038	0.840	0.856	0.760	0.904	0.716	0.556
200	CM_n	0.052	0.050	1.000	0.716	0.744	0.700	0.428	0.332
	KS_n	0.048	0.048	0.956	0.572	0.608	0.580	0.268	0.204
	HEL_n	0.074	0.060	0.928	0.964	0.572	0.992	0.992	0.940
	CHF_n	0.046	0.042	0.976	0.988	0.820	0.952	0.944	0.864
	SEL_n	0.052	0.032	0.992	1.000	0.972	1.000	0.884	0.864
400	CM_n	0.054	0.042	1.000	0.936	0.972	0.904	0.728	0.592
	KS_n	0.064	0.044	1.000	0.784	0.892	0.860	0.592	0.460
	HEL_n	0.054	0.052	1.000	1.000	0.848	1.000	0.988	0.992
	CHF_n	0.044	0.044	1.000	1.000	0.984	1.000	0.988	0.984
	SEL_n	0.056	0.032	1.000	1.000	1.000	1.000	0.940	0.948

DGP1′: $W_t = (\varepsilon_{1,t}', \varepsilon_{2,t}, \varepsilon_{3,t})'$, where both $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}, \varepsilon_{3,t}\}$ are IID $N(0, I_2)$.

For DGP2′ through DGP8′, $W_t = ((Y_{t-1}, Y_{t-2}), Y_t, Z_{t-1})'$,

DGP2′: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \varepsilon_{1,t}$;

DGP3′: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.5Z_{t-1} + \varepsilon_{1,t}$;

DGP4′: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \alpha Z_{t-1}^2 + \varepsilon_{1,t}$;

DGP5′: $Y_t = \alpha Y_{t-1}Z_{t-1} + 0.25Y_{t-2} + \varepsilon_{1,t}$;

DGP6′: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.55Z_{t-1}\varepsilon_{1,t}$;

DGP7′: $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Y_{t-2}^2 + 0.25Z_{t-1}^2$;

DGP8′: same as DGP8;

where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$ and $\{\varepsilon_{1,t}, \varepsilon_{2,t}\}$ is IID $N(0, I_2)$ in DGPs 2′-7′.

In view of the difficulty in implementing the HEL_n test because of bandwidth selection, we only study the finite sample behavior of the other four tests. We use the same kernels, weighting functions, number of replications, and number of bootstrap resamples as before. The bandwidth selection rule for the CHF_n and SEL_n tests is also the same as in case 1.

Table 2 reports the empirical size and power properties of the four tests. As in the first case, the sizes are reasonably well behaved for all tests and the SEL_n and CHF_n tests tend to dominate the **LG** tests in terms of power. The CHF_n test dominates CM_n and KS_n for all nonlinear DGPs under investigation, SEL_n exhibits significantly greater empirical power than CHF_n in DGPs 3′-6′, but it is the other way around in DGPs 7′-8′. As expected, in comparison with DGP8, DGP8′ suggests that the power of our tests would be adversely affected as the dimension of the conditioning variable increases.

For the third case ($d_1 = 3,\ d_2 = d_3 = 1$), we use the following DGPs:

DGP1″: $W_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})'$,where $\{\varepsilon_{1,t}\}$ is IID $N(0, I_3)$ and $\{\varepsilon_{2,t}, \varepsilon_{3,t}\}$ is IID $N(0, I_2)$.

For DGP2″ through DGP7″, $W_t = ((Y_{t-1}, Y_{t-2}, Y_{t-3}), Y_t, Z_{t-1})'$,

DGP2″: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \varepsilon_{1,t}$;

DGP3″: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + 0.5Z_{t-1} + \varepsilon_{1,t}$;

Table 2: Comparison of tests for causality ($d_1 = 2, d_2 = d_3 = 1$), nominal level: 0.05

n	Tests	DGP1′	DGP2′	DGP3′	DGP4′	DGP5′	DGP6′	DGP7′	DGP8′
100	CM_n	0.028	0.016	0.656	0.360	0.108	0.512	0.164	0.208
	KS_n	0.040	0.020	0.400	0.284	0.056	0.380	0.124	0.176
	CHF_n	0.028	0.042	0.720	0.704	0.412	0.564	0.460	0.556
	SEL_n	0.052	0.040	0.844	0.828	0.620	0.568	0.440	0.528
200	CM_n	0.050	0.032	0.940	0.588	0.304	0.792	0.304	0.364
	KS_n	0.046	0.024	0.776	0.432	0.168	0.696	0.216	0.284
	CHF_n	0.030	0.040	0.948	0.944	0.748	0.828	0.724	0.860
	SEL_n	0.058	0.028	0.972	0.988	0.932	0.832	0.684	0.832
400	CM_n	0.056	0.024	1.000	0.884	0.552	0.980	0.556	0.668
	KS_n	0.060	0.024	1.000	0.732	0.324	0.952	0.384	0.524
	CHF_n	0.040	0.036	1.000	0.984	0.972	0.996	0.920	0.984
	SEL_n	0.040	0.030	1.000	1.000	1.000	1.000	0.836	0.884

DGP4″: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + 0.5Z_{t-1}^2 + \varepsilon_{1,t}$;

DGP5″: $Y_t = 0.5Y_{t-1}Z_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + \varepsilon_{1,t}$;

DGP6″: $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + 0.55Z_{t-1}\varepsilon_{1,t}$;

DGP7″: $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Y_{t-2}^2 + 0.125Y_{t-3}^2 + 0.5Z_{t-1}^2$;

DGP8″: same as DGP8;

where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, and $\{\varepsilon_{1,t}, \varepsilon_{2,t}\}$ is IID $N(0, I_2)$ in DGPs 2″-7″.

We use the same kernels, weighting functions, number of replications, and number of bootstrap resamples as in the first case. The main difference is that we need to adjust the LSCV bandwidths slightly to meet the conditions in Assumption A2(iii). We first set $h_1 = c_1 n^{-1/(4+d_1+d_3)}$ and $h_2 = c_2 n^{-1/(8+d_1)}$ and choose c_1 and c_2 by using the LSCV. Then we adjust h_1 to be $h_1^* n^{1/(4+d_1+d_3)} n^{-1/8.5}$ where h_1^* is the LSCV bandwidth for h_1. Consequently, the resulting bandwidths satisfy $h_1 \propto n^{-1/8.5}$ and $h_2 \propto n^{-1/(8+d_1)}$. In view of $d_1 + d_3 = 4$, $n = 100$ is too small for any nonparametric test to be well behaved. Thus, we only consider $n = 200$ and 400 in Table 3. Table 3 reports the empirical size and power behavior of our tests. Apparently, the results are similar to the second case above.

Table 3: Comparison of tests for causality ($d_1 = 3, d_2 = d_3 = 1$), nominal level: 0.05

n	Tests	DGP1″	DGP2″	DGP3″	DGP4″	DGP5″	DGP6″	DGP7″	DGP8″
200	CM_n	0.050	0.028	0.756	0.484	0.192	0.788	0.260	0.448
	KS_n	0.048	0.032	0.500	0.380	0.096	0.660	0.212	0.372
	CHF_n	0.028	0.022	0.964	0.952	0.668	0.852	0.552	0.856
	SEL_n	0.056	0.026	0.996	0.980	0.860	0.816	0.344	0.680
400	CM_n	0.050	0.032	0.992	0.728	0.400	0.912	0.390	0.620
	KS_n	0.044	0.036	0.840	0.552	0.220	0.880	0.306	0.568
	CHF_n	0.032	0.034	1.000	0.972	0.928	0.884	0.792	0.972
	SEL_n	0.056	0.038	1.000	1.000	1.000	0.888	0.616	0.876

6 Applications to macroeconomic and financial time series

Although many studies conducted during the 1980s and 1990s report that economic and financial time series, such as exchange rates and stock prices, exhibit nonlinear dependence (e.g., Hsieh 1989, 1991; Sheedy, 1998), researchers often neglect this when they test for Granger causal relationships. In this section, we first study the dynamic linkage between pairwise daily exchange rates across some industrialized countries by using both our SELR test ($\tilde{T}_{n1}$) for CI and the traditional linear Granger causality test. Then with the same techniques, we study the dynamic linkage between three US stock market price indices and their trading volumes. Finally, we investigate the relationship between money supply, output, and prices for the U.S. economy.

6.1 Application 1: exchange rates

Over the last two decades much research has focused on the nonlinear dependence exhibited by foreign exchange rates, but few studies have examined nonlinear Granger causal links between intra-market exchange rates. One exception is Hong (2001) who proposes a test for volatility spillover and applies it to study the volatility spillover between two weekly nominal U.S. dollar exchange rates, the Deutschemark and the Japanese Yen. He finds a change in past Deutschemark volatility Granger-causes a change in current Japanese Yen volatility but a change in past Japanese Yen volatility does not Granger-cause a change in current Deutschemark volatility.

In this application, we apply our nonparametric test to data for the daily exchange rates for three industrialized countries (Canada, Japan, and the UK) and the European Union (EU), and compare the results to those using a conventional linear test for Granger causality. The data are from Datastream, with the sample period running from January 2nd, 2002 to April 5th, 2011, a total of 2415 observations. The exchange rates are the local currency against the US dollar. Nevertheless, due to national holidays or certain other reasons, some observations for exchange rates are missing. Also, different nations have different national holidays and thus different missing observations. Because we do causality tests with exchange rates from countries pairwise , if the observation for one country is missing, we also delete that for the other country of the pair. Following the literature, we let E_t stand for the natural logarithm of exchange rates multiplied by 100. Since both the linear Granger causality test and our nonparametric test require that the time series be stationary and we are interested in the relation between the changes in the exchange rates, we first employ the augmented Dickey-Fuller test to check for stationarity for all four exchange rates. The test results indicate that there is a unit root in all level series E_t but not in the first differenced series ΔE_t. Therefore, both Granger causality tests will be conducted on the first differenced data.

Let DX be the first differenced exchange rate in Country X and DY the first differenced exchange rate in Country Y. The time series $\{DX_t\}$ does not linearly Granger cause the time series $\{DY_t\}$ if the null hypothesis $\mathbb{H}_{0,L} : \beta_1 = \cdots = \beta_{L_x} = 0$ holds in

$$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + \cdots + \alpha_{L_y} DY_{t-L_y} + \beta_1 DX_{t-1} + \cdots + \beta_{L_x} DX_{t-L_x} + \epsilon_t, \tag{6.1}$$

where ϵ_t is a white noise under $\mathbb{H}_{0,L}$. An F-statistic can be constructed to check whether the null $\mathbb{H}_{0,L}$ is true or not. Nevertheless, in order to make a direct comparison with our nonparametric test for nonlinear

Granger causality in the next subsection, we focus on testing $\mathbb{H}^*_{0,L}: \beta = 0$ in

$$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + \cdots + \alpha_{L_y} DY_{t-L_y} + \beta DX_{t-i} + \epsilon_t,\ i = 1, ..., L_x. \quad (6.2)$$

Apparently, $\mathbb{H}^*_{0,L}$ is nested in $\mathbb{H}_{0,NL}$. The rejection of $\mathbb{H}^*_{0,NL}$ indicates the rejection of $\mathbb{H}_{0,NL}$ but not the other way around.

To implement our test, we set all parameters according to those used in our simulations. The null of interest is now

$$\mathbb{H}_{0,NL}: \Pr\left[F(DY_t|DY_{t-1}, ..., DY_{t-L_y}; DX_{t-1}, ..., DX_{t-L_x}) = F(DY_t|DY_{t-1}, ..., DY_{t-L_y})\right] = 1. \quad (6.3)$$

Due to the "curse of dimensionality," we allow only L_y =1, 2, or 3. Further, for each test we only include one lagged DX_t in the conditioning set. Thus, we actually test a variant of $\mathbb{H}_{0,NL}$:

$$\mathbb{H}^*_{0,NL}: \Pr\left[F(DY_t|DY_{t-1}, ..., DY_{t-L_y}; DX_{t-i},) = F(DY_t|DY_{t-1}, ..., DY_{t-L_y})\right] = 1,\ i = 1, ..., L_x. \quad (6.4)$$

Again, $\mathbb{H}^*_{0,NL}$ is nested in $\mathbb{H}_{0,NL}$. The rejection of $\mathbb{H}^*_{0,NL}$ indicates the rejection of $\mathbb{H}_{0,NL}$ but not the other way around. In this sense, our test is conservative.

For both tests, when L_y is 1, we also choose L_x to be 1 so that we only check whether DX_{t-1} should enter (6.2) or (6.4) or not. When L_y is 2, we choose L_x to be 2. In this case, we check whether DX_{t-1} or DX_{t-2} (but not both) should enter (6.2) or (6.4) or not. The case for $L_y = 3$ is analogous. Consequently, for each test, we have six possible results. If any of these results suggest that we should reject the null at the 1% nominal level,[8] then we indicate that there is linear or nonlinear Granger causality from DX to DY.

Table 4: Bivariate linear Granger causality test between exchange rates (nominal level: 1%)

	Linear Granger causality				Nonlinear Granger causality			
	Canada	EU	Japan	UK	Canada	EU	Japan	UK
Canada	-	$\nRightarrow$	$\nRightarrow$	$\nRightarrow$	-	$\nRightarrow$	$\Rightarrow$	$\Rightarrow$
EU	$\nRightarrow$	-	$\nRightarrow$	$\nRightarrow$	$\nRightarrow$	-	$\nRightarrow$	$\Rightarrow$
Japan	$\Rightarrow$	$\Rightarrow$	-	$\nRightarrow$	$\Rightarrow$	$\nRightarrow$	-	$\nRightarrow$
UK	$\nRightarrow$	$\nRightarrow$	$\nRightarrow$	-	$\nRightarrow$	$\nRightarrow$	$\nRightarrow$	-

We summarize the results in Table 4, where $\Rightarrow$ and $\nRightarrow$ signify the presence and absence of Granger-causality from the column variable to the row variable, respectively, and the nonlinear tests are based on the bootstrap p-values obtained from 400 bootstrap resamples. First, the linear Granger test reveals only two unidirectional Granger causal links among the four exchange rate series: the exchange rate for Japan Granger-causes that for Canada and EU. Second, our nonparametric test fails to detect the linear Granger-causality from Japan's exchange rate to the EU's but can reveal another three more causal links among the four exchange rate series, some of which are bidirectional (Canada $\Leftrightarrow$ Japan). One obvious reason for the failure of the linear Granger causality test to detect any bidirectional causal linkages is

[8]This gives the Bonferroni bound 6%, comparable with the widely used nominal significance level 5%. Such a multiple procedure also applies to Applications 2 and 3 below.

that exchange rates exhibit unambiguously nonlinear dependence across markets. The volatility spillover between exchange rates (see Hong (2001) and the reference there) is a special case of such nonlinear dependence. Third and interestingly, neither test suggests a Granger-causal relationship between the exchanges rates of Canada or Japan and the EU.

6.2 Application 2: stock price and trading volume

There are several explanations for the presence of a bidirectional Granger causal relation between stock prices and trading volume. For brevity, we only mention two of them. The first one is the sequential information arrival model (e.g., Copeland, 1976) in which new information flows into the market and is disseminated to investors one at a time. This pattern of information arrival produces a sequence of momentary equilibria consisting of various stock price-volume combinations before a final, complete information equilibrium is achieved. Due to the sequential information flow, lagged trading volume could have predictive power for current absolute stock returns and lagged absolute stock returns could have predictive power for current trading volume. The other is the noise trader model (e.g., DeLong, 1990) that reconciles the difference between the short- and long-run autocorrelation properties of aggregate stock returns. Aggregate stock returns are positively autocorrelated in the short run, but negatively autocorrelated in the long run. Since noise traders do not trade on the basis of economic fundamentals, they impart a transitory mispricing component to stock prices in the short run. The temporary component disappears in the long run, producing mean reversion in stock returns. A positive causal relation from volume to stock returns is consistent with the assumption made in these models that the trading strategies pursued by noise traders cause stock prices to move. A positive predictive relation from stock returns to volume is consistent with the positive-feedback trading strategies of noise traders, for which the decision to trade is conditioned on past stock price movements.

Gallant et al. (1992) argue that more can be learned about the stock market by studying the joint dynamics of stock prices and trading volume than by focusing on the univariate dynamics of stock returns. Using daily data for the Dow Jones price index for the periods 1915-1990, Hiemstra and Jones (1994) study the dynamic relation between stock prices and trading volume and find significant bidirectional nonlinear predictability between them. Here we reinvestigate this issue using the latest daily data for the three major U.S. stock market price indices, i.e., the Dow Jones, the Nasdaq, and the S&P 500, and their associated trading volumes in the NYSE, Nasdaq and NYSE markets, respectively. The data are obtained from Yahoo Finance with the sample period running from January 2nd, 2002 to January 4th, 2011. Following the literature, we let P_t and V_t stand for the natural logarithm of stock price index and volume multiplied by 100, respectively. Both Granger causality tests will be conducted on the first differenced data ΔP_t and ΔV_t.

The implementation of both tests is similar to the previous application. As for the case of exchange rates, we conduct both tests for the three major market indices and summarize the results in Table 5, where $\Leftrightarrow$ signifies the presence of bidirectional Granger causality and $\not\Leftrightarrow$ its absence in both directions, and the nonlinear tests are based on the bootstrap p-values obtained from 400 bootstrap resamples. First, the linear Granger causality test suggests there is no Granger causal link between trading volumes and stock prices for all three datasets. Second, our nonparametric test reveals bidirectional Granger causal links between stock prices and trading volumes for all three datasets at the 1% level, in strong support of the results of Hiemstra and Jones (1994), who find bidirectional links for the Dow Jones stock price and trading volume up to a 7-day lag. So, unlike the linear Granger causality test results, our nonparametric test results lend some partial support to the two theories articulated above. One obvious reason for the failure of the linear Granger causality test in detecting such causal links is that stock prices may only have nonlinear predictive power for trading volumes.

Table 5: Granger causality tests between stock prices and trading volumes (nominal level: 1%)

	Linear Granger Causality	Nonlinear Granger Causality
$\Delta P \Leftrightarrow \Delta V$	-	Dow Jones, Nasdaq, S&P 500
$\Delta P \not\Leftrightarrow \Delta V$	Dow Jones, Nasdaq, S&P 500	-

6.3 Application 3: money, income, and prices

There has been a long debate in macroeconomics regarding the role of money in an economy, particularly in the determination of income and prices. Monetarists claim that money plays an active role and leads to changes in income and prices. In other words, changes in income and prices in an economy are mainly caused by changes in money stocks. Hence, the direction of causation runs from money to income and prices without any feedback, i.e., unidirectional causation. Keynesians, on the other hand, argue that money does not play an active role in changing income and prices. In fact, changes in income cause changes in money stocks via demand for money implying that the direction of causation runs from income to money without any feedback. Similarly, changes in prices are mainly caused by structural factors.

The empirical race took an interesting turn with the famous tests of Sims (1972). Specifically, Sims developed a test for linear Granger causality and applied it to U.S. data to examine the relationship between money and income, finding the evidence of unidirectional Granger causality from money to income, as claimed by the Monetarists. However, his results were not supported by subsequent studies, indicating that the empirical evidence regarding links between money and the other two variables, income and price, remain elusive. Here we re-examine the Granger causal relationships using a longer horizon of U.S. data.

Seasonally adjusted monthly data for monetary aggregates M1 and M2, disposable personal income (DPI), real disposable personal income (RDPI), industrial output (IP), consumer price index (CPI), and producer price index (PPI) were obtained from the Federal Reserve Bank of St. Louis with a sample period running from January, 1959 to February, 2011. The total number of observations is 626. As in Friedman and Kuttner (1992, 1993), Swanson (1998), and Black et al. (2000), the analysis below uses log-differences of all the series. Dickey-Fuller tests suggest that the transformed series are stationary.

The implementation of both the linear and nonparametric Granger causality is similar to Application 1. The results are summarized in Table 6 where the nonlinear tests are based on the bootstrap p-values obtained from 400 bootstrap resamples. First, the linear Granger causality test indicates that there is mixed evidence of uni- or bi-directional causality between the three income variables (DPI, RDPI and

IP) and money (M1 or M2); there is uni- (bi-) directional causality between M1 (M2) and CPI and no causal relationship between M1 (M2) and PPI. Second, the nonparametric Granger causality test results show strong evidence of uni- or bi- directional causality between money and income, and between money and prices. We thus conclude that monetary aggregates still provide predictive information for income and prices, which is largely consistent with the findings of Swanson (1998), who uses a rolling window approach to study the predictive power of monetary aggregates on output.

Table 6: Granger causality test between money, output and prices (nominal level: 1%)

	Linear Granger Causality					Nonlinear Granger Causality				
	DPI	RDPI	IP	CPI	PPI	DPI	RDPI	IP	CPI	PPI
M1	$\nLeftrightarrow$	$\Leftarrow$	$\nLeftrightarrow$	$\Leftarrow$	$\nLeftrightarrow$	$\Leftrightarrow$	$\Leftarrow$	$\Leftarrow$	$\Rightarrow$	$\Leftrightarrow$
M2	$\Leftrightarrow$	$\Leftarrow$	$\nLeftrightarrow$	$\Leftrightarrow$	$\nLeftrightarrow$	$\Leftrightarrow$	$\Leftarrow$	$\Leftrightarrow$	$\Rightarrow$	$\Leftrightarrow$

Note: At the 5% nominal level, the nonlinear tests suggest CPI $\Rightarrow$ M1 and M2.

7 Concluding remarks

We construct two classes of SELR tests for conditional independence and extend the applicability of EL from testing a finite number of moment or conditional moment restrictions to testing an infinite collection of conditional moment restrictions. Writing the null hypothesis in terms of CDF-based conditional moment restrictions and employing the SELR approach, we construct some intuitively appealing test statistics and show that they are asymptotically normal under the null and a sequence of local alternatives. Although these test statistics have intuitive appeal, they tend to deliver poor power in small samples, due to the discrete nature of the indicator functions. Thus we build on Su and White (2007) and consider a class of smoother moment conditions to construct a new class of SELR tests. We show that in large samples both tests are asymptotically optimal in that they attain maximum average local power with respect to different spaces of functions for the corresponding local alternatives. Simulations suggest that the smoother SELR test performs well in finite samples. We apply this latter test to some economic and financial time series and find that the test reveals some interesting nonlinear Granger causal relations that traditional linear Granger causality tests fail to detect.

Appendix

A Some useful lemmas

In this appendix, we introduce some lemmas that are used in the proofs of the main results in the text.

Lemma A.1 *Let* $\{\xi_t,\, t \geq 0\}$ *be a d-dimensional strong mixing process with mixing coefficient* $\alpha(\cdot)$. *Let* $F_{i_1,\ldots,i_k}$ *denote the distribution function of* $(\xi_{i_1},\ldots,\xi_{i_k})$. *For any integer* $k > 1$ *and integers* $i_1,\ldots,i_k$ *such that* $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, *let* φ_n *be a Borel measurable function on* $\mathbb{R}^{kd}$ *such that* $\int |\varphi_n(\xi_{i_1},\ldots,\xi_{i_k})|^{1+\tilde{\delta}} dP(\xi_{i_1},\ldots,\xi_{i_j})\,(v_1,\ldots,v_j)\,dP(\xi_{i_{j+1}},\ldots,\xi_{i_k}) \leq M_n$ *for some* $\tilde{\delta} > 0$. *Then* $|\int \varphi_n(\xi_{i_1},\ldots,\xi_{i_k})dP(\xi_{i_1},\ldots,\xi_{i_k}) - \int \varphi_n(\xi_{i_1},\ldots,\xi_{i_k})\,dP(\xi_{i_1},\ldots,\xi_{i_j})dP(\xi_{i_{j+1}},\ldots,\xi_{i_k})| \leq 4M_n^{1/(1+\tilde{\delta})}\alpha(i_{j+1}-i_j)^{\tilde{\delta}/(1+\tilde{\delta})}$, *where, e.g.,* $P(\xi_{i_1},\ldots,\xi_{i_k})$ *denotes the probability measure of* $(\xi_{i_1},\ldots,\xi_{i_k})$.

Proof. See Lemma 2.1 of Sun and Chiang (1997). ■

Lemma A.2 *Let* φ_n *and* M_n *be defined as in Lemma A.1. Let* $V_1 \equiv (\xi_{i_1},\ldots,\xi_{i_j})$, *and* $V_2 \equiv (\xi_{i_{j+1}},\ldots,\xi_{i_k})$. *Then* $E\,|E[\varphi_n(V_1,V_2)|V_1] - E_{V_1}\varphi_n(V_1,V_2)| \leq 4M_n^{1/(1+\tilde{\delta})}\alpha(i_{j+1}-i_j)^{\tilde{\delta}/(1+\tilde{\delta})}$, *where* $E_{V_1}\varphi_n(V_1,V_2) \equiv \Phi_n(V_1)$ *with* $\Phi_n(v_1) \equiv E[\varphi_n(v_1,V_2)]$.

Proof. Yoshihara (1989) proved the above lemma for the β-mixing case. A close examination of his proof reveals that the lemma also holds for the α-mixing case by an application of Lemma A.1. ■

Let $\{\xi_t,\, t \geq 0\}$ be as defined in Lemma A.1. Let $g_n(\cdot,\cdot)$ be Borel measurable functions on $\mathbb{R}^d \times \mathbb{R}^d$ such that $E[g_n(\xi_0, u)] = 0$ and $g_n(u,v) = g_n(v,u)$ for all $(u,v) \in \mathbb{R}^d \times \mathbb{R}^d$. Define $\mathcal{G}_n \equiv n^{-1}\sum_{1\leq t<s\leq n}[g_n(\xi_t,\xi_s) - Eg_n(\xi_t,\xi_s)]$. Let $\{\bar{\xi}_t,\, t \geq 0\}$ be an IID sequence where $\bar{\xi}_0$ is an independent copy of ξ_0. Further, define

$$\begin{aligned} u_n(p) &\equiv \max\{\max_{1\leq t\leq n} ||g_n(\xi_t,\xi_0)||_p, ||g_n(\xi_0,\bar{\xi}_0)||_p\},\ v_n(p) \equiv \max\{\max_{1\leq t\leq n} ||G_{n0}(\xi_t,\xi_0)||_p, ||G_{n0}(\xi_0,\bar{\xi}_0)||_p\}, \\ w_n(p) &\equiv ||G_{n0}(\xi_0,\xi_0)||_p,\ z_n(p) \equiv \max_{1\leq t\leq n}\max_{1\leq s\leq n}\max\left\{||G_{ns}(\xi_t,\xi_0)||_p,\ ||G_{ns}(\xi_0,\xi_t)||_p,\ ||G_{ns}(\xi_0,\bar{\xi}_0)||_p\right\}, \end{aligned}$$

where $G_{ns}(u,v) \equiv E[g_n(\xi_s,u)g_n(\xi_0,v)]$ and $||\cdot||_p \equiv \{E|\cdot|^p\}^{1/p}$ for $p \geq 1$.

Lemma A.3 *Let* $\{\xi_t,\, t \geq 0\}$ *be a d-dimensional strictly stationary strong mixing process with mixing coefficient* $\alpha(\cdot)$. *Given the above notation, suppose there exists* $\delta_0 > 0$, $\gamma_0 < 1/2$, *and* $\gamma_1 > 0$ *such that (i)* $u_n(4+\delta_0) = O(n^{\gamma_0})$; *(ii)* $v_n(2) = o(1)$; *(iii)* $w_n(2+\delta_0/2) = o(n^{1/2})$; *(iv)* $z_n(2)n^{\gamma_1} = O(1)$; *(v)* $E[g_n(\xi_0,\bar{\xi}_0)]^2 = 2\tilde{\sigma}^2 + o(1)$; *(vi)* $(n^{1+4\gamma_0+3\delta_1} + n^{2+\delta_1})\alpha(n^{\delta_2})^{\delta_0/(4+\delta_0)} = o(1)$ *for some* δ_1 *and* δ_2 *such that* $0 < \delta_2 < \delta_1 < \min(\gamma_1/2, (1-2\gamma_0)/3)$. *Then* $\mathcal{G}_n \xrightarrow{d} N(0,\tilde{\sigma}^2)$.

Proof. Tenreiro (1997) proved the above lemma under conditions (i)-(v) and the assumption that $\{\xi_t,\, t \geq 0\}$ is a strictly stationary absolutely regular process with geometric mixing coefficient $\beta(n)$ (i.e., $\beta(n) \leq C\rho^n$ for some $C > 0$ and $\rho \in (0,1)$). By examining his proof closely, we find that the assumption of absolute regularity is only used for applying two inequalities of Yoshihara (1976) and another one of Yoshihara (1989) (see Corollaries 1 and 2 and Lemma 1 in Tenreiro, 1997). The first two inequalities can be replaced by the Davydov inequality for strong mixing processes (e.g., Bosq, 1996, p.19; Hall and Heyde, 1980, p.278) and the inequality in Lemma A.1. The third inequality also has a strong mixing analog, as detailed in Lemma A.2. The geometric mixing rate was assumed only for the convenience of the determination of asymptotically negligible terms. A close read of the proof of Lemma 3 in Tenreiro (1997) shows that under conditions (i)-(v), the requirements on the strong mixing rate are given in condition (vi). ■

Remark. Fan and Li (1999) prove a CLT for second-order degenerate U-statistics of absolutely regular processes under different sets of regularity conditions. One can also relax their assumption of absolute regularity to that of strong mixing. Gao (2007, Theorem A.1) also proves a CLT for second-order degenerate U-statistics of strong mixing processes with geometric decay rate.

Lemma A.4 *Under Assumptions A1-A3, (i)* $\sup_{\tau\in\mathbb{R}}\sup_{(x,z)\in S}|\hat{F}_{h_1}(\tau|x,z)-F(\tau|x,z)|=O_p(\mu_{1n})$, *(ii)* $\sup_{\tau\in\mathbb{R}}\sup_{x\in S_1}|\hat{F}_{h_2}(\tau|x)-F(\tau|x)|=O_p(\mu_{2n})$, *where* $S_1\equiv S\cap\mathbb{R}^{d_1}$, $\mu_{1n}\equiv n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n}+h_1^2$ *and* $\mu_{2n}\equiv n^{-1/2}h_2^{-d_1/2}\sqrt{\ln n}+h_2^r$.

Proof. The proof is a modification of the proof of Lemma B.3 in Newey (1994). ■

Remark. For part (i) of the above lemma, Boente and Fraiman (1991) prove a slightly different result: $\sup_{\tau\in\mathbb{R}}\sup_{(x,z)\in S}|\hat{F}_{h_1}(\tau|x,z)-F(\tau|x,z)|=O_{a.s.}(n^{-1/2}h_1^{-(d_1+d_3+\epsilon)/2}(\ln n)^2+h_1^2)$, where ϵ is an arbitrarily small positive number. The above lemma continues to hold if we replace the compact set S by its ϵ-extension: $S^{\epsilon}\equiv\{u\in\mathbb{R}^{d_1+d_3}:\|u-v\|\leq\epsilon \text{ for some } v\in S\}$.

To proceed, let $f_{1t}=f(X_t,Z_t)$, $\hat{f}_{1t}=\hat{f}_{h_1}(X_t,Z_t)$, $f_{2t}=f(X_t)$, $\hat{f}_{2t}=\hat{f}_{h_2}(X_t)$, $K_{ts}=K_{h_1}(X_t-X_s,Z_t-Z_s)$, $L_{ts}=L_{h_2}(X_t-X_s)$, $K_{(x,z),t}=K_{h_1}(x-X_t,z-Z_t)$, and $L_{x,t}=L_{h_1}(x-X_t)$.

Lemma A.5 *Under Assumptions A1-A3 and* $\mathbb{H}_0$, *(i)* $\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|I_t\sum_{s=1}^n w_{ts}\hat{\varepsilon}_s(\tau)|=O_p(\mu_n)$, *(ii)* $\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|I_t[\hat{V}_1(X_t,Z_t;\tau)-V(X_t,Z_t;\tau)]|=O_p\left(\mu_{1,n}+\mu_{2,n}\right)$, *where* $\mu_n\equiv\mu_{0n}+\mu_{2n}$, *and* $\mu_{0n}\equiv n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n}$.

Proof. (i) Under $\mathbb{H}_0$, $\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|I_t\sum_{s=1}^n w_{ts}\hat{\varepsilon}_s(\tau)|\leq\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|I_t\sum_{s=1}^n w_{ts}[1(Y_s\leq\tau)-F(\tau|X_s,Z_s)]|+\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|I_t\sum_{s=1}^n w_{ts}[F(\tau|X_s)-\hat{F}_{h_2}(\tau|X_s)]|\equiv A_{1n}+A_{2n}$. By Newey (1994, Lemma B.1), $\max_{1\leq t\leq n}I_t|\hat{f}_{1t}-E(\hat{f}_{1t})|=O_p(\mu_n)$ and $\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|n^{-1}I_t\sum_{s=1}^n K_{ts}[1(Y_s\leq\tau)-F(\tau|X_s,Z_s)]|=O_p(\mu_{0n})$. Therefore, $A_{1n}=O_p(\mu_n)$. Noting that $\sum_{s=1}^n w_{ts}=1$, $A_{2n}\leq\sup_{\tau\in\mathbb{R}}\sup_{x\in S_1}|\hat{F}_{h_2}(\tau|x)-F(\tau|x)|=O_p(\mu_{2n})$ by Assumption A.3(i). The desired result follows.

(ii) Note that $\hat{V}_1(X_t,Z_t;\tau)=\sum_{s=1}^n w_{ts}[\hat{\varepsilon}_s(\tau)]^2=\hat{f}_{1t}^{-1}n^{-1}\sum_{s=1}^n K_{ts}[1(Y_s\leq\tau)-\hat{F}_{h_2}(\tau|X_s)]^2$. By the triangle inequality, we have $\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|I_t[\hat{V}_1(X_t,Z_t;\tau)-V(X_t,Z_t;\tau)]|\leq\xi_{1,n}+2\xi_{2,n}+\xi_{3,n}$ under $\mathbb{H}_0$, where

$$\xi_{1,n}=\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}\left|I_t\hat{f}_{1t}^{-1}n^{-1}\sum_{s=1}^n K_{ts}[1(Y_s\leq\tau)-F(\tau|X_s,Z_s)]^2-V(X_t,Z_t;\tau)\right|,$$

$$\xi_{2,n}=\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}\left|I_t\hat{f}_{1t}^{-1}n^{-1}\sum_{s=1}^n K_{ts}\left[1(Y_s\leq\tau)-F(\tau|X_s,Z_s)\right]\left[F(\tau|X_s)-\hat{F}_{h_2}(\tau|X_s)\right]\right|,\text{ and}$$

$$\xi_{3,n}=\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}\left|I_t\hat{f}_{1t}^{-1}n^{-1}\sum_{s=1}^n K_{ts}\left[F(\tau|X_s)|-\hat{F}_{h_2}(\tau|X_s)\right]^2\right|.$$

First,

$$\begin{aligned}
\xi_{1,n} &\leq \sup_{\tau\in\mathbb{R}}\sup_{(x,z)\in S}\left|\hat{f}_{h_1}^{-1}(x,z)n^{-1}\sum_{s=1}^{n}K_{(x,z),s}\{[1(Y_s\leq\tau)-F(\tau|X_s,Z_s)]^2-V(X_s,Z_s;\tau)]\}\right| \\
&\quad+\sup_{\tau\in\mathbb{R}}\sup_{(x,z)\in S}\left|\hat{f}_{h_1}^{-1}(x,z)n^{-1}\sum_{s=1}^{n}K_{(x,z),s}[V(X_s,Z_s;\tau)-V(x,z;\tau)]\right| \\
&= O_p\left(n^{-1/2}h_1^{-(d_1+d_3)/2}\sqrt{\ln n}\right)+O_p\left(h_1^2\right)=O_p\left(\mu_{1,n}\right).
\end{aligned}$$

Let $\epsilon = h_1$. Since K has compact support on [-1,1]$^{d_1+d_3}$ by Assumption A2(i), we have, by Assumption A3(i) and for sufficiently large n, $\xi_{2,n} \leq \sup_{\tau\in\mathbb{R}}\sup_{x\in S_1^\epsilon}|F(\tau|x)- \hat{F}_{h_2}(\tau|x)| = O_p\left(\mu_{2,n}\right)$, $\xi_{3,n} \leq \sup_{\tau\in\mathbb{R}}\sup_{x\in S_1^\epsilon}|F(\tau|x)-\hat{F}_{h_2}(\tau|x)|^2 = O_p((\mu_{2,n})^2)$, where $S_1^\epsilon = S^\epsilon\cap\mathbb{R}^{d_1}$. This completes the proof. ■

B Proof of Theorem 3.1

Let C denote a generic constant which may vary from case to case. Let $\{\bar{W}_t = \left(\bar{X}_t',\bar{Y}_t,\bar{Z}_t'\right)',\ t\geq 0\}$ denote an IID sequence where $\bar{W}_t$ is an independent copy of W_t.

Lemma B.1 *Let Assumptions A1-A4 hold. Then under* $\mathbb{H}_0$, $ISELR_n = \hat{B}_n+\hat{R}_n+o_p(h_1^{-(d_1+d_3)/2})$, *where* $\hat{B}_n = \sum_{t=1}^n I_t\int\hat{V}_{1t}(\tau)^{-1}\sum_{s=1}^n w_{ts}^2\hat{\varepsilon}_s(\tau)^2dG(\tau)$, $\hat{R}_n = \sum_{t=1}^n I_t\sum_{s=1}^n\sum_{j=1,j\neq s}^n\int\hat{V}_{1t}(\tau)^{-1}w_{ts}\hat{\varepsilon}_s(\tau)w_{tj}\hat{\varepsilon}_j(\tau)$ $dG(\tau)$, *and* $\hat{V}_{1t}(\tau)\equiv\hat{V}_1(X_t,Z_t;\tau)\equiv\sum_{s=1}^n w_{ts}\hat{\varepsilon}_s(\tau)^2$.

Proof. From (2.7), we have

$$\begin{aligned}
0 &= \sum_{s=1}^{n}\frac{w_{ts}\hat{\varepsilon}_s(\tau)}{n+\lambda_t(\tau)\hat{\varepsilon}_s(\tau)}=\frac{1}{n}\sum_{s=1}^{n}w_{ts}\hat{\varepsilon}_s(\tau)\left\{1-\frac{\lambda_t(\tau)\hat{\varepsilon}_s(\tau)}{n}+\frac{[\lambda_t(\tau)\hat{\varepsilon}_s(\tau)/n]^2}{1+\lambda_t(\tau)\hat{\varepsilon}_s(\tau)/n}\right\} \\
&= \frac{1}{n}\sum_{s=1}^{n}w_{ts}\hat{\varepsilon}_s(\tau)-\frac{1}{n^2}\hat{V}_{1t}(\tau)\lambda_t(\tau)+\frac{r_{1t}(\tau)}{n^2},
\end{aligned}$$

where $r_{1t}(\tau)=\sum_{s=1}^n\frac{w_{ts}\hat{\varepsilon}_s(\tau)[\lambda_t(\tau)\hat{\varepsilon}_s(\tau)]^2}{n+\lambda_t(\tau)\hat{\varepsilon}_s(\tau)}$. Consequently

$$I_t\hat{V}_{1t}(\tau)\lambda_t(\tau)=nI_t\sum_{s=1}^{n}w_{ts}\hat{\varepsilon}_s(\tau)+I_tr_{1t}(\tau). \tag{B.1}$$

Eq.(2.7) also implies $\sum_{s=1}^n\frac{w_{ts}[\lambda_t(\tau)\hat{\varepsilon}_s(\tau)]^2}{n+\lambda_t(\tau)\hat{\varepsilon}_s(\tau)} = \sum_{s=1}^n w_{ts}\hat{\varepsilon}_s(\tau)\lambda_t(\tau)$. Hence, as $n+\lambda_t(\tau)\hat{\varepsilon}_s(\tau) > 0$ (because $\hat{p}_{ts}\geq 0$, $w_{ts}\geq 0$ and $\hat{p}_{ts}=w_{ts}/[n+\lambda_t(\tau)\hat{\varepsilon}_s(\tau)]$),

$$\begin{aligned}
\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|r_{1t}(\tau)| &\leq \sup_{\tau\in\mathbb{R}}\max_{1\leq s\leq n}|\hat{\varepsilon}_s(\tau)|\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}\left|\sum_{s=1}^{n}\frac{w_{ts}[\lambda_t(\tau)\hat{\varepsilon}_s(\tau)]^2}{n+\lambda_t(\tau)\hat{\varepsilon}_s(\tau)}\right| \\
&= \sup_{\tau\in\mathbb{R}}\max_{1\leq s\leq n}|\hat{\varepsilon}_s(\tau)|\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}\left|\sum_{s=1}^{n}w_{ts}\hat{\varepsilon}_s(\tau)\lambda_t(\tau)\right| \\
&\leq C\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}\left|\sum_{s=1}^{n}w_{ts}\hat{\varepsilon}_s(\tau)\right|\sup_{\tau\in\mathbb{R}}\max_{1\leq t\leq n}|\lambda_t(\tau)|.
\end{aligned}$$

Thus by Lemma A.5(i), $\sup_{\tau\in\mathbb{R}}\max_{1\le t\le n} I_t|r_{1t}(\tau)| = O_p(\mu_n)\sup_{\tau\in\mathbb{R}}\max_{1\le t\le n} I_t|\lambda_t(\tau)|$ and $I_t\hat{V}_{1t}(\tau)$ $\lambda_t(\tau) = O_p(n\mu_n)+O_p(\mu_n)I_t|\lambda_t(\tau)|$. Consequently, $\sup_{\tau\in\mathbb{R}}\max_{1\le t\le n} I_t|\lambda_t(\tau)| = O_p(n\mu_n)$ and $\sup_{\tau\in\mathbb{R}}\max_{1\le t\le n}$ $I_t|r_{1t}(\tau)| = O_p(n\mu_n^2)$.

Now by a second order Taylor expansion, with probability approaching 1 as $n\to\infty$ (w.p.a.1),

$$I_t \log\left(1+\frac{\lambda_t(\tau)\,\hat{\varepsilon}_s(\tau)}{n}\right) = I_t\left\{\frac{\lambda_t(\tau)\,\hat{\varepsilon}_s(\tau)}{n} - \frac{1}{2}\left[\frac{\lambda_t(\tau)\,\hat{\varepsilon}_s(\tau)}{n}\right]^2 + \bar{\eta}_{ts}(\tau)\right\}, \tag{B.2}$$

where the remainder term $|I_t\bar{\eta}_{ts}(\tau)| \le C\,|I_t\lambda_t(\tau)\,\hat{\varepsilon}_s(\tau)/n|^3 = O_p(\mu_n^3)$ uniformly in (t,s,τ). By (2.9), (B.1) and (B.2), we can show that w.p.a.1,

$$\begin{aligned} ISELR_n &= \sum_{t=1}^n I_t\int \hat{V}_{1t}(\tau)^{-1}\sum_{s=1}^n w_{ts}^2\hat{\varepsilon}_s(\tau)^2 dG(\tau) + \sum_{t=1}^n I_t\sum_{s=1}^n\sum_{j=1,j\neq s}^n\int \hat{V}_{1t}(\tau)^{-1}w_{ts}\hat{\varepsilon}_s(\tau)w_{tj}\hat{\varepsilon}_j(\tau)dG(\tau) \\ &\quad -n^{-2}\sum_{t=1}^n\int I_t\hat{V}_{1t}(\tau)^{-1}r_{1t}^2(\tau)dG(\tau) + 2\sum_{t=1}^n I_t\int\sum_{s=1}^n w_{ts}\bar{\eta}_{ts}(\tau)dG(\tau). \end{aligned}$$

Noting that $n^{-1}\sum_{t=1}^n I_t\int\hat{V}_{1t}(\tau)^{-1}dG(\tau) = n^{-1}\sum_{t=1}^n I_t\int V(X_t,Z_t;\tau)^{-1}dG(\tau) + o_p(1) = O_p(1)$ by Lemma A.5, Assumption A4 and the Markov inequality, we have $n^{-2}\sum_{t=1}^n\int I_t\hat{V}_{1t}(\tau)^{-1}r_{1t}^2(\tau)dG(\tau) \le$ $O_p(1)\,n\sup_{\tau\in\mathbb{R}}\max_{1\le s,t\le n}[r_{1t}(\tau)/n]^2 = O_p\left(n\mu_n^4\right) = o_p(h_1^{-(d_1+d_3)/2})$, and $\sum_{t=1}^n I_t\int\sum_{s=1}^n w_{ts}\bar{\eta}_{ts}(\tau)dG(\tau)$ $\le n\sup_{\tau\in\mathbb{R}}\max_{1\le s,t\le n}|\bar{\eta}_{ts}(\tau)| = O_p(n\mu_n^3) = o_p(h_1^{-(d_1+d_3)/2})$ by Assumption A3(i), the conclusion of the lemma follows. ■

Lemma B.2 *Let Assumptions A1-A4 hold. Then* $h_1^{(d_1+d_3)/2}\hat{R}_n \xrightarrow{d} N(0,\sigma^2)$ *under* $\mathbb{H}_0$.

Proof. Under $\mathbb{H}_0$, write $\hat{R}_n = n^{-2}\sum_{t=1}^n I_t\hat{f}_{1t}^{-2}\sum_{s=1}^n\sum_{j=1,j\neq s}^n\int\hat{V}_{1t}(\tau)^{-1}K_{ts}K_{tj}\hat{\varepsilon}_s(\tau)\hat{\varepsilon}_j(\tau)dG(\tau) =$ $R_{n,1}+R_{n,2}+2R_{n,3}$, where

$$R_{n,1} \equiv n^{-2}\sum_{t=1}^n I_t\sum_{j\neq s}\hat{f}_{1t}^{-2}\int\hat{V}_{1t}(\tau)^{-1}r_n(W_s;X_t,Z_t,\tau)r_n(W_j;X_t,Z_t,\tau)dG(\tau), \tag{B.3}$$

$$R_{n,2} \equiv n^{-2}\sum_{t=1}^n I_t\sum_{j\neq s}\hat{f}_{1t}^{-2}K_{ts}K_{tj}\int\hat{V}_{1t}(\tau)^{-1}[F(\tau|X_s)-\hat{F}_{h_2}(\tau|X_s)][F(\tau|X_j)-\hat{F}_{h_2}(\tau|X_j)]dG(\tau), \tag{B.4}$$

$$R_{n,3} \equiv n^{-2}\sum_{t=1}^n I_t\sum_{j\neq s}\hat{f}_{1t}^{-2}K_{ts}K_{tj}\int\hat{V}_{1t}(\tau)^{-1}\varepsilon_s(\tau)[F(\tau|X_j)-\hat{F}_{h_2}(\tau|X_j)]dG(\tau), \tag{B.5}$$

$\sum_{j\neq s}\equiv\sum_{s=1}^n\sum_{j=1,j\neq s}^n$, and $r_n(W_t;x,z,\tau)\equiv K_{(x,z),t}[1(Y_t\le\tau)-F(\tau|X_t,Z_t)]$. Let

$$\tilde{R}_{n,1} \equiv n^{-1}\sum_{s=1}^n\sum_{t=1,t\neq s}^n\int\int_S\hat{f}_{h_1}^{-2}(x,z)\hat{V}_1(x,z;\tau)^{-1}r_n(W_t;x,z,\tau)r_n(W_s;x,z,\tau)dF(x,z)dG(\tau). \tag{B.6}$$

By Lemmas B.3 and B.4 below, $h_1^{(d_1+d_3)/2}R_{n,1} = h_1^{(d_1+d_3)/2}\tilde{R}_{n,1}+o_p(1) \xrightarrow{d} N(0,\sigma^2)$. By Lemmas B.5 and B.6, $h_1^{(d_1+d_3)/2}R_{n,l} = o_p(1)$ for $l=2,3$. Consequently $h_1^{(d_1+d_3)/2}\hat{R}_n \xrightarrow{d} N(0,\sigma^2)$ under $\mathbb{H}_0$. ■

Lemma B.3 *Let Assumptions A1-A4 hold. Then* $h_1^{(d_1+d_3)/2}R_{n,1} = h_1^{(d_1+d_3)/2}\tilde{R}_{n,1}+o_p(1)$, *where* $R_{n,1}$ *and* $\tilde{R}_{n,1}$ *are defined in (B.3) and (B.6), respectively.*

Proof. Let $\hat{F}(x,z)$ denote the empirical distribution function of $\{(X_t,Z_t)\}_{t=1}^n$. Then $h_1^{(d_1+d_3)/2}(R_{n,1}-\tilde{R}_{n,1}) = h_1^{(d_1+d_3)/2} n^{-1}\sum_{s=1}^n \sum_{t=1,t\neq s}^n \int_S \int r_n(W_s;x,z,\tau) r_n(W_t;x,z,\tau)\hat{f}_{h_1}^{-2}(x,z)\hat{V}(x,z;\tau)^{-1}dG(\tau)d[\hat{F}(x,z) -F(x,z)] = \tilde{\Delta}_{n,1}+2\tilde{\Delta}_{n,2}-2\tilde{\Delta}_{n,3}$, where

$$\begin{aligned}\tilde{\Delta}_{n,1} &\equiv n^{-2}h_1^{(d_1+d_3)/2}\sum_{t\neq s\neq j}\left\{\int I_j r_n(W_s;X_j,Z_j,\tau)r_n(W_t;X_j,Z_j,\tau)\hat{f}_{h_1}^{-2}(X_j,Z_j)\hat{V}_1(X_j,Z_j;\tau)^{-1}dG(\tau)\right.\\ &\qquad \left. -\int_S\int r_n(W_s;x,z,\tau)r_n(W_t;x,z,\tau)\hat{f}_{h_1}^{-2}(x,z)\hat{V}_1(x,z;\tau)^{-1}dG(\tau)dF(x,z)\right\},\\ \tilde{\Delta}_{n,2} &\equiv n^{-2}h_1^{(d_1+d_3)/2}\sum_{t\neq s}\int I_s r_n(W_s;X_s,Z_s,\tau)r_n(W_t;X_s,Z_s,\tau)\hat{f}_{h_1}^{-2}(X_s,Z_s)\hat{V}_1(X_s,Z_s;\tau)^{-1}dG(\tau),\\ \tilde{\Delta}_{n,3} &\equiv n^{-2}h_1^{(d_1+d_3)/2}\sum_{t\neq s}\int_S\int r_n(W_s;x,z,\tau)r_n(W_t;x,z,\tau)\hat{f}_{h_1}^{-2}(x,z)\hat{V}_1(x,z;\tau)^{-1}dG(\tau)dF(x,z),\end{aligned}$$

where $\sum_{t\neq s}\equiv\sum_{t=1}^n\sum_{s=1,s\neq t}^n$ and $\sum_{t\neq s\neq j}\equiv\sum_{t=1}^n\sum_{s=1,s\neq t}^n\sum_{j=1,j\neq s,t}^n$. Dispensing with the simplest term first, we have $\tilde{\Delta}_{n,3}=2n^{-1}\{h_1^{(d_1+d_3)/2}\tilde{R}_{n,1}\}=n^{-1}O_p(1)=o_p(1)$ by Lemma B.4 below.

For $\tilde{\Delta}_{n,2}$, we have $\tilde{\Delta}_{n,2}=\Delta_{n,2}\{1+o_p(1)\}$ where $\Delta_{n,2}=n^{-2}h_1^{(d_1+d_3)/2}\sum_{t\neq s}\vartheta^0(W_t,W_s)$, $\vartheta^0(W_t,W_s) \equiv\int I_s r_n(W_s;X_s,Z_s,\tau)r_n(W_t;X_s,Z_s,\tau)\ f_{1s}^{-2}V_s(\tau)^{-1}dG(\tau)$, and $V_s(\tau)\equiv V(X_s,Z_s;\tau)$. We want to show $\Delta_{n,2}=o_p(1)$. Define the symmetric version of ϑ^0 as $\vartheta(W_t,W_s)=[\vartheta^0(W_t,W_s)+\vartheta^0(W_s,W_t)]/2$. Then we have $\Delta_{n,2}=n^{-2}h_1^{(d_1+d_3)/2}\sum_{1\leq t<s\leq n}\vartheta(W_t,W_s)$. By construction, $\vartheta(u,v)=\vartheta(v,u)$. By the fact that $E[r_n(W_s;X_s,Z_s,\tau)|\ X_s,Z_s]=0$ and $E[r_n(\bar{W}_t;\bar{X}_s,\bar{Z}_s,\tau)|\bar{X}_t,\bar{Z}_t]=0$ for $t\neq s$, we have $E[\vartheta(W_t,v)]=0$. Further, it is easy to verify that

$$\max_{1<t\leq n}E\left|\vartheta(W_1,W_t)\right|^{2(1+\delta)}\leq Ch_1^{-(3+4\delta)(d_1+d_3)} \text{ and } \max_{1<t\leq n}E\left|\vartheta\left(\bar{W}_1,\bar{W}_t\right)\right|^{2(1+\delta)}\leq Ch_1^{-(3+4\delta)(d_1+d_3)}.$$

Then by Lemma A.2(ii) of Gao (2007), $E(\Delta_{n,2})^2=n^{-4}h_1^{d_1+d_3}O(n^2h_1^{-(3+4\delta)(d_1+d_3)/(1+\delta)})=O(n^{-2} h_1^{-(2+3\delta)(d_1+d_3)/(1+\delta)})=o(1)$. Hence $\Delta_{n,2}=o_P(1)$ by the Chebyshev inequality.

Now, we show $\tilde{\Delta}_{n,1}=o_p(1)$. Note that $\tilde{\Delta}_{n,1}=\Delta_{n,1}\{1+o_p(1)\}$ where $\Delta_{n,1}=n^{-2}h_1^{(d_1+d_3)/2}\sum_{t\neq s\neq j}\phi^0 (W_t,W_s,W_j)$, and $\phi^0(W_t,W_s,W_j)\equiv\int I_j r_n(W_s;X_j,Z_j,\tau)r_n(W_t;X_j,Z_j,\tau)f_{1j}^{-2}V_j(\tau)^{-1}dG(\tau)-\int_S\int r_n(W_s; x,z,\tau)r_n(W_t;x,z,\tau)f^{-2}(x,z)V(x,z;\tau)^{-1}dG(\tau)dF(x,z)$. Define the symmetric version of ϕ^0 as $\phi(W_t,W_s, W_j)=[\phi^0(W_t,W_s,W_j)+\phi^0(W_t,W_j,W_s)+\phi^0(W_s,W_j,W_t)]/3$. Then $\Delta_{n,1}=n^{-2}h_1^{(d_1+d_3)/2}\sum_{1\leq t<s<j\leq n}\phi (W_t,W_s,W_j)$. By construction, $\phi(u,v,w)$ is symmetric in its arguments and $E[\vartheta(W_t,u,v)]=0$. Further, it is easy to verify that

$$\max_{1<t<s\leq n}\max\{E\left|\phi(W_1,W_t,W_s)\right|^{2(1+\delta)},E\left|\phi\left(\bar{W}_1,\bar{W}_t,W_s\right)\right|^{2(1+\delta)}\}\leq Ch_1^{-(2+4\delta)(d_1+d_3)}.$$

and $\max_{1<t<s\leq n}\max\{\int|\phi(W_1,W_t,W_s)|\,dF(W_1)\,dF(W_t,W_s)^{2(1+\delta)},\int|\phi(W_1,W_t,W_s)|^{2(1+\delta)}dF(W_s)\,dF (W_1,W_t)\}\leq Ch_1^{-(2+4\delta)(d_1+d_3)}$, where $F(W_t,W_s)$ denotes the joint CDF of W_t and W_s. Then by Lemma A.2(i) of Gao (2007),[9]

$$E(\Delta_{n,1})^2=n^{-4}h_1^{d_1+d_3}O(n^3h_1^{-(2+4\delta)(d_1+d_3)/(1+\delta)})=O(n^{-1}h_1^{-(1+3\delta)(d_1+d_3)/(1+\delta)})=o(1).$$

Hence $\Delta_{n,1}=o_P(1)$ by the Chebyshev inequality. It follows that $\tilde{\Delta}_{n,1}=o_p(1)$. ■

[9] Lemma A.2 in Gao (2007) implicitly requires that $\sum_{s=1}^\infty s^4\alpha(s)^{\delta/(1+\delta)}\leq C<\infty$.

Lemma B.4 *Let Assumptions A1-A3 hold. Then* $h_1^{(d_1+d_3)/2}\tilde{R}_{n,1} \xrightarrow{d} N(0,\sigma^2)$ *under* $\mathbb{H}_0$*, where* $\tilde{R}_{n,1}$ *is defined in (B.6) and* $\sigma^2 \equiv 2C_3^{d_1+d_3}\int_S\int\int V(x,z;\tau)^{-1}V(x,z;\tau')^{-1}V(x,z;\tau,\tau')^2 dG(\tau)dG(\tau')d(x,z)$*.*

Proof.

$$
\begin{aligned}
& h_1^{(d_1+d_3)/2}\tilde{R}_{n,1} \\
= \; & h_1^{(d_1+d_3)/2} n^{-1}\sum_{t\neq s}\int\int_S r_n(W_s;x,z,\tau)r_n(W_t;x,z,\tau)f^{-2}(x,z)V(x,z;\tau)^{-1}dF(x,z)dG(\tau)\left\{1+o_p(1)\right\} \\
= \; & \left\{2n^{-1}\sum_{1\leq s<t\leq n}\left\{H_n(W_s,W_t)-E[H_n(W_s,W_t)]\right\}+2n^{-1}\sum_{1\leq s<t\leq n}E[H_n(W_s,W_t)]\right\}\left\{1+o_p(1)\right\} \\
\equiv \; & \left\{S_{n,1}+S_{n,2}\right\}\left\{1+o_p(1)\right\},
\end{aligned}
$$

where $H_n(W_s,W_t) = h_1^{(d_1+d_3)/2}\int\int_S r_n(W_s;x,z,\tau)r_n(W_t;x,z,\tau)f^{-2}(x,z)V(x,z;\tau)^{-1}dF(x,z)dG(\tau)$. We now verify the conditions in Lemma A.3 hold for $S_{n,1}$ with $H_n(u,v)$ replacing $g_n(u,v)$ in the lemma. First, by construction, $H_n(u,v)=H_n(v,u)$ and $EH_n(W_0,v)=0$. Then

$$
\begin{aligned}
E\left|H_n(W_t,W_0)\right|^p \; = \; & h_1^{p(d_1+d_3)/2}\int\int\left|\int\int_S K_{h_1}(x-x_t,z-z_t)[1(y_t\leq\tau)-F(\tau|x_t,z_t)]K_{h_1}(x-x_0,z-z_0)\right. \\
& \left.\times[1(Y_0\leq\tau)-F(\tau|z_0,z_0)]f^{-2}(x,z)V(x,z;\tau)^{-1}dF(x,z)dG(\tau)\right|^p f_t(w_0,w_t)dw_0dw_t \\
\leq \; & Ch_1^{p(d_1+d_3)/2}h_1^{-(d_1+d_3)(p-1)}\int_{\mathbb{R}^{d_1+d_3}}\int_{\mathbb{R}^{d_1+d_3}}|K(u_1)K(u_1+u_2)|^p du_1du_2 \\
= \; & O\left(h_1^{(d_1+d_3)(1-p/2)}\right) \text{ by Assumptions A1(ii)-(iii), A2(i) and A4.}
\end{aligned}
$$

Hence $||H_n(W_t,W_0)||_p \leq Ch_1^{(d_1+d_3)(1/p-1/2)}$. Analogously, we have $||H_n(W_0,\bar{W}_0)||_p \leq h_1^{(d_1+d_3)(1/p-1/2)}$. Consequently, one obtains $u_n(p)\leq Ch_1^{(d_1+d_3)(1/p-1/2)}$ for some $C>0$.

Now we show $v_n(p)\leq Ch_1^{(d_1+d_3)/p}$. By Assumptions A1(ii)-(iii), A2(i) and A4, we have

$$
\begin{aligned}
G_{n0}(w_t,w_0) \; \equiv \; & E[H_n(W_0,w_t)H_n(W_0,w_0)] \\
= \; & h^{d_1+d_3}\int\int\int_S\int_S E\left\{r_n(W_0;x,z,\tau)r_n(w_t;x,z,\tau)r_n(W_0;x',z',\tau')r_n(w_0;x',z',\tau')\right. \\
& \left.\times V(x,z;\tau)^{-1}V(x',z';\tau')^{-1}dF(x,z)dF(x',z')dG(\tau)dG(\tau')\right\} \\
\leq \; & C\int_{\mathbb{R}^{d_1+d_3}}\int_{\mathbb{R}^{d_1+d_3}}\int_{\mathbb{R}^{d_1+d_3}}K(u)K(u+u')K(\tilde{u})K(\tilde{u}+u'+(w_t-w_0)/h_1)dudu'd\tilde{u} \\
\leq \; & C\int_{\mathbb{R}^{d_1+d_3}}K(u+(w_t-w_0)/h_1)du,
\end{aligned}
$$

so $||G_{n0}(W_t,W_0)||_p \leq Ch_1^{(d_1+d_3)/p}$. Similarly, one can show $||G_{n0}(W_0,\bar{W}_0)||_p \leq Ch_1^{(d_1+d_3)/p}$ and thus $v_n(p)\leq Ch_1^{(d_1+d_3)/p}$. By the same argument, we can show that $w_n(p)\equiv||G_{n0}(W_0,W_0)||_p\leq C$ and $z_n(p)\leq Ch_1^{d_1+d_3}$. It follows that $v_n(2)=o(1)$ and $w_n(2+\delta_0)=o(n^{1/2})$ for any $\delta_0>0$. That is, Conditions (ii)-(iii) in Lemma A.3 are satisfied. Fix $\delta_0>0$. Take $\gamma_0=(2+\delta_0)/(16+4\delta_0)\in(0,1/4)$ and $\gamma_1=1/2-\epsilon_1$ for some $\epsilon_1\leq\delta_0/(24+6\delta_0)$. Then Conditions (i) and (iv) in Lemma A.3 are satisfied. Then $\min(\gamma_1/2,(1-2\gamma_0)/3)=(6+\delta_0)/(24+6\delta_0)$, and Condition (vi) in Lemma A.3 is ensured by

Assumption A3(ii). Finally, straightforward calculations yield

$$E[H_n(W_0,\bar{W}_0)^2] = h_1^{d_1+d_3}\int\int\int_S\int_S E\{r_n(W_0;x,z,\tau)r_n(\bar{W}_0;x.z,\tau)r_n(W_0;x',z',\tau')r_n(\bar{W}_0;x',z',\tau') \times V(x,z;\tau)^{-1}V(x',z';\tau')^{-1}dF(x,z)dF(x',z')dG(\tau)dG(\tau')\} = \sigma^2/2+o(1).$$

It follows that $S_{n,1}\xrightarrow{d}N(0,\sigma^2)$.

Now we show that $S_{n,2}=o(1)$. By the triangle inequality, Lemma A.1, and using the bound $u_n(p)\leq Ch_1^{(d_1+d_3)(1/p-1/2)}$ with $p=1+\delta$, we have

$$\begin{aligned} S_{n,2} &\leq 2n^{-1}\sum_{s=1}^{n-1}\sum_{t=s+1}^{n}|E[H_n(W_s,W_t)]| \leq Cn^{-1}\sum_{s=1}^{n-1}\sum_{t=s+1}^{n}h_1^{(d_1+d_3)[1/(1+\delta)-1/2]}\alpha(t-s)^{\delta/(1+\delta)} \\ &\leq Ch_1^{(1-\delta)(d_1+d_3)/[2(1+\delta)]}\sum_{s=1}^{\infty}\alpha(s)^{\delta/(1+\delta)} = O\left(h_1^{(1-\delta)(d_1+d_3)/[2(1+\delta)]}\right)=o(1). \end{aligned}$$

This completes the proof. ■

Lemma B.5 *Let Assumptions A1-A3 and the null hypothesis hold. Then* $h_1^{(d_1+d_3)/2}R_{n,2}=o_p(1)$, *where* $R_{n,2}$ *is defined in (B.4).*

Proof. Note that $R_{n,2}=R_{n,21}-R_{n,22}$, where $R_{n,21}=\sum_{t=1}^n I_t\int\hat{V}_{1t}(\tau)^{-1}\{\sum_{s=1}^n w_{ts}[\hat{F}_{h_2}(\tau|X_s)-F(\tau|X_s)]\}^2dG(\tau)$, and $R_{n,22}=\sum_{t=1}^n I_t\sum_{s=1}^n w_{ts}^2\int\hat{V}_{1t}(\tau)^{-1}[\hat{F}_{h_2}(\tau|X_s)-F(\tau|X_s)]^2dG(\tau)$. Using Lemma A.4, we can readily show that $h_1^{(d_1+d_3)/2}R_{n,22}=O_p(h_1^{-(d_1+d_3)/2}\mu_{2n}^2)=o_p(1)$ by Assumption A2(iii). It suffices to show that $h_1^{(d_1+d_3)/2}R_{n,22}=o_p(1)$.

Let $\hat{g}_s(\tau)\equiv[\hat{F}_{h_2}(\tau|X_s)-F(\tau|X_s)]\hat{f}_{2s}$. By the Cauchy-Schwarz inequality, $R_{n,21}=n^{-2}\sum_{t=1}^n I_t\hat{f}_{1t}^{-2}\int[\sum_{s=1}^n K_{ts}\hat{f}_{2s}^{-1}\hat{g}_s(\tau)]^2dG(\tau)\leq 2R_{n,211}+2R_{n,212}$, where $R_{n,211}=n^{-2}\sum_{t=1}^n I_t\hat{f}_{1t}^{-2}\int\hat{V}_{1t}(\tau)^{-1}[\sum_{s=1}^n K_{ts}f_{2s}^{-1}\hat{g}_s(\tau)]^2dG(\tau)$ and $R_{n,212}=n^{-2}\sum_{t=1}^n I_t\hat{f}_{1t}^{-2}\int\hat{V}_{1t}(\tau)^{-1}[\sum_{s=1}^n K_{ts}(\hat{f}_{2s}^{-1}-f_{2s}^{-1})\,\hat{g}_s(\tau)]^2dG(\tau)$. By Lemma A.4, it is easy to show that $h_1^{(d_1+d_3)/2}R_{n,212}=nh_1^{(d_1+d_3)/2}(\mu_{2n})^4=o_p(1)$ and $h_1^{(d_1+d_3)/2}R_{n,211}=\bar{R}_{n,2}+nh_1^{(d_1+d_3)/2}O_p\left((\mu_{2n})^3\right)=\bar{R}_{n,2}+o_p(1)$, where $\bar{R}_{n,2}=n^{-2}h_1^{(d_1+d_3)/2}\sum_{t=1}^n I_t f_{1t}^{-2}V_t(\tau)^{-1}\int[\sum_{s=1}^n K_{ts}f_{2s}^{-1}\hat{g}_s(\tau)]^2 dG(\tau)$ and $V_t(\tau)\equiv V(X_t,Z_t;\tau)$. Let $\varsigma_{js}(\tau)\equiv F(\tau|X_j)-F(\tau|X_s)$. Then $\hat{g}_s(\tau)=n^{-1}\sum_{j=1}^n L_{js}[\varepsilon_j(\tau)+\varsigma_{js}(\tau)]$ and

$$\begin{aligned} \bar{R}_{n,2} &= n^{-4}h_1^{(d_1+d_3)/2}\sum_{t_0,t_1,t_2,t_3,t_4}\int I_{t_0}V_{t_0}(\tau)^{-1}f_{1t_0}^{-2}f_{2t_1}^{-1}f_{2t_3}^{-1}K_{t_0t_1}K_{t_0t_3}L_{t_1t_2}L_{t_3t_4}[\varepsilon_{t_2}(\tau)+\varsigma_{t_2t_1}(\tau)] \\ &\quad\times[\varepsilon_{t_4}(\tau)+\varsigma_{t_4t_3}(\tau)]dG(\tau) \\ &= n^{-4}h_1^{(d_1+d_3)/2}\sum_{t_0,t_1,t_2,t_3,t_4}\int I_{t_0}V_{t_0}(\tau)^{-1}f_{1t_0}^{-2}f_{2t_1}^{-1}f_{2t_3}^{-1}K_{t_0t_1}K_{t_0t_3}L_{t_1t_2}L_{t_3t_4}\varepsilon_{t_2}(\tau)\varepsilon_{t_4}(\tau)dG(\tau) \\ &\quad+n^{-4}h_1^{(d_1+d_3)/2}\sum_{t_0,t_1,t_2,t_3,t_4}\int I_{t_0}V_{t_0}(\tau)^{-1}f_{1t_0}^{-2}f_{2t_1}^{-1}f_{2t_3}^{-1}K_{t_0t_1}K_{t_0t_3}L_{t_1t_2}L_{t_3t_4}\varsigma_{t_2t_1}(\tau)\varsigma_{t_4t_3}(\tau)dG(\tau) \\ &\quad+2n^{-4}h_1^{(d_1+d_3)/2}\sum_{t_0,t_1,t_2,t_3,t_4}\int I_{t_0}V_{t_0}(\tau)^{-1}f_{1t_0}^{-2}f_{2t_1}^{-1}f_{2t_3}^{-1}K_{t_0t_1}K_{t_0t_3}L_{t_1t_2}L_{t_3t_4}\varepsilon_{t_2}(\tau)\varsigma_{t_4t_3}(\tau)dG(\tau) \\ &\equiv G_{n1}+G_{n2}+2G_{n3},\ \text{say.} \end{aligned}$$

Noting that $G_{n1}+G_{n2}+G_{n3}$ is nonnegative, by the Markov inequality it suffices to show that $E(G_{ni})=o(1)$ for $i=1$, 2, and 3.

To show $EG_{n,1} \equiv E(G_{n1}) = o(1)$, let $S^{(1)}_{t_0,t_1,t_2,t_3,t_4} \equiv E[\int I_{t_0} V_{t_0}(\tau)^{-1} f^{-2}_{1t_0} f^{-1}_{2t_1} f^{-1}_{2t_3} K_{t_0t_1} K_{t_0t_3} L_{t_1t_2} L_{t_3t_4} \varepsilon_{t_2}(\tau)\varepsilon_{t_4}(\tau) dG(\tau)]$. Let ι_n be as defined in Assumption A3(*iii*). We consider three different cases for $EG_{n,1}$: (*a*) for each $i \in \{0,1,2,3,4\}$, $|t_i - t_j| > \iota_n$ for all $j \neq i$; (*b*) all time indices are distinct and for exactly three different $i's$, $|t_i - t_j| > \iota_n$ for all $j \neq i$; (*c*) all the other remaining cases. We use $EG_{n,1s}$ to denote these cases ($s = a, b, c$). For case (*a*), noting that $E[\varepsilon_t(\tau)|X_t, Z_t] = 0$ under $\mathbb{H}_0$, we can apply Lemma A.1 directly to obtain $EG_{n,1a} \leq Cn^{-4} h_1^{(d_1+d_3)/2} n^5 h_1^{-2(d_1+d_3)\delta/(1+\delta)} h_2^{-2d_1\delta/(1+\delta)} \alpha(\iota_n)^{\delta/(1+\delta)} = O(nh_1^{(1-3\delta)(d_1+d_3)/[2(1+\delta)]} h_2^{-2d_1\delta/(1+\delta)} \alpha(\iota_n)^{\delta/(1+\delta)}) = o(1)$ by Assumption A3(*iii*). For case (*b*), if either t_2 or t_4 is among the three elements that lie at least at ι_n-distance from all the other elements, one can bound the term $S^{(1)}_{t_0,t_1,t_2,t_3,t_4}$ as in case (*a*). Otherwise, bound it by C and the total number of such terms is $O(n^4\iota_n)$. Consequently, $EG_{n,1b} = o(1) + n^{-4} h_1^{(d_1+d_3)/2} O(n^4 \iota_n) = o(1)$ by Assumption A3(*iii*). For case (*c*), the total number of terms in the summation is of order $O(n^3\iota_n^2)$ and one can readily obtain $EG_{n,1c} = n^{-4} h_1^{(d_1+d_3)/2} O(n^3\iota_n^2 + n^3\iota_n h_1^{-(d_1+d_3)} + n^3 h_1^{-2(d_1+d_3)} + n^2 h_1^{-2(d_1+d_3)} h_2^{-d_1} + nh_1^{-2(d_1+d_3)} h_2^{-2d_1}) = o(1)$. It follows that $EG_{n,1} = o(1)$.

Next, let $S^{(2)}_{t_0,t_1,t_2,t_3,t_4} \equiv E[\int I_{t_0} V_{t_0}(\tau)^{-1} f^{-2}_{1t_0} f^{-1}_{2t_1} f^{-1}_{2t_3} K_{t_0t_1} K_{t_0t_3} L_{t_1t_2} L_{t_3t_4} \varsigma_{t_2t_1}(\tau)\varsigma_{t_4t_3}(\tau) dG(\tau)]$. By Assumption A1 and dominated convergence arguments, for $t_1 \neq t_2$ and $t_3 \neq t_4$, this term is bounded by Ch_2^{2r} if $\{t_1,t_2\} \cap \{t_3,t_4\} \neq \{t_1,t_2\}$ and $t_1 \neq t_0 \neq t_3$, by $Ch_2^{2-d_1}$ if $\{t_1,t_2\} \cap \{t_3,t_4\} = \{t_1,t_2\}$ and $t_1 \neq t_0 \neq t_3$, by $Ch_1^{-(d_1+d_3)} h_2^{2r}$ if $\{t_1,t_2\} \cap \{t_3,t_4\} \neq \{t_1,t_2\}$ and either t_1 or t_3 (but not both) equals t_0. The other cases are of smaller orders after summation. Consequently, $E(G_{n2}) = n^{-4} h_1^{(d_1+d_3)/2} O(n^5 h_2^{2r} + n^3 h_2^{2-d_1} + n^4 h_1^{-(d_1+d_3)} h_2^{2r}) = o(1)$. Similarly, we have $E(G_{n3}) = o(1)$ and the proof is complete. ■

Lemma B.6 *Let Assumptions A1-A3 hold. Then* $h_1^{(d_1+d_3)/2} R_{n,3} = o_p(1)$, *where* $R_{n,3}$ *is defined in (B.5).*

Proof. As in the proof of Lemma B.5, we can show that $-h_1^{(d_1+d_3)/2} R_{n,3} = n^{-3} h_1^{(d_1+d_3)/2} \sum_{t_1,t_2,t_3,t_4; t_2 \neq t_3} \int I_{t_1} \hat{V}_{t_1}(\tau)^{-1} \hat{f}^{-1}_{1t_1} \hat{f}^{-1}_{2t_3} K_{t_1t_2} K_{t_1t_3} L_{t_3t_4} \varepsilon_{t_2}(\tau)[\varepsilon_{t_4}(\tau) + M_{t_4t_3}(\tau)] dG(\tau) = \bar{R}_{n,31} + \bar{R}_{n,32} + o_p(1)$, where

$$\bar{R}_{n,31} = n^{-3} h_1^{(d_1+d_3)/2} \sum_{t_1,t_2,t_3,t_4; t_2 \neq t_3} \int I_{t_1} V_{t_1}(\tau)^{-1} f^{-2}_{1t_1} f^{-1}_{2t_3} K_{t_1t_2} K_{t_1t_3} L_{t_3t_4} \varepsilon_{t_2}(\tau)\varepsilon_{t_4}(\tau) dG(\tau), \text{ and}$$

$$\bar{R}_{n,32} = n^{-3} h_1^{(d_1+d_3)/2} \sum_{t_1,t_2,t_3,t_4; t_2 \neq t_3} \int I_{t_1} \hat{V}_{t_1}(\tau)^{-1} \hat{f}^{-1}_{1t_1} \hat{f}^{-1}_{2t_3} K_{t_1t_2} K_{t_1t_3} L_{t_3t_4} \varepsilon_{t_2}(\tau) M_{t_4t_3}(\tau) dG(\tau).$$

Let $T^{(1)}_{t_1,t_2,t_3,t_4}(\tau) \equiv E[I_{t_1} V_{t_1}(\tau)^{-1} f^{-2}_{1t_1} f^{-1}_{2t_3} K_{t_1t_2} K_{t_1t_3} L_{t_3t_4} \varepsilon_{t_2}(\tau)\varepsilon_{t_4}(\tau)]$ and consider three different cases for $ER_{n,1} \equiv E(\bar{R}_{n,31})$: (*a*) for each $i \in \{1,2,3,4\}$, $|t_i - t_j| > \iota_n$ for all $j \neq i$; (*b*) all time indices are distinct and for exactly two different $i's$, $|t_i - t_j| > \iota_n$ for all $j \neq i$; (*c*) all the other remaining cases. We use $ER_{n,1s}$ to denote these cases ($s = a, b, c$). For case (*a*), we apply Lemma A.1 immediately to get $ER_{n,1a} \leq Ch_1^{(d_1+d_3)/2} n^{-3} n^4 h_1^{-2(d_1+d_3)\delta/(1+\delta)} h_2^{-d_1\delta/(1+\delta)} \alpha(\iota_n)^{\delta/(1+\delta)} = O(nh_1^{(1-3\delta)(d_1+d_3)/(1+\delta)} h_2^{-d_1\delta/(1+\delta)} \alpha(\iota_n)^{\delta/(1+\delta)})$ $= o(1)$ by Assumption A3(*iii*). For case (*b*), if either t_2 or t_4 is among the two elements that lie at least at ι_n-distance from all the other elements, one can bound the term $T^{(1)}_{t_1,t_2,t_3,t_4}(\tau)$ as in case (*a*). Otherwise, bound it by C and the total number of such terms is $O(n^3\iota_n)$. Consequently, $EG_{1b} = o(1) + n^{-3} h_1^{(d_1+d_3)/2} O(n^3\iota_n) = o(1)$. For case (*c*), the total number of terms in the summation is of order $n^2\iota_n^2$ and one can readily obtain $ER_{n,1c} = n^{-3} h_1^{(d_1+d_3)/2} O(n^2\iota_n^2 + n^2 m h_1^{-(d_1+d_3)} + n^2 h_1^{-(d_1+d_3)} h_2^{-d_1}) = o(1)$. So $EI_{n,1} = o(1)$.

Next, we want to show $ER_{n,2} \equiv E(\bar{R}_{n,31})^2 = n^{-6} h_1^{d_1+d_3} \sum_{t_1,t_2,t_3,t_4; t_2 \neq t_3} \sum_{t_5,t_6,t_7,t_8; t_6 \neq t_7} E[\int\int T^{(1)}_{t_1,t_2,t_3,t_4}(\tau)$ $T^{(1)}_{t_5,t_6,t_7,t_8}(\tau') dG(\tau) G(\tau')] = o(1)$. We consider three different cases for $ER_{n,2}$: (*a*) for at least five $i's$ in

$\{1,2,3,4,6,7,8\}$, $|t_i - t_j| > \iota_n$ for all $j \neq i$; (b) all time indices are distinct and for exactly four different $i's$, $|t_i - t_j| > \iota_n$ for all $j \neq i$; (c) all the other remaining cases. We use $EI_{n,2s}$ to denote these cases $(s = a, b, c)$. For case (a), we apply Lemma A.1 to obtain $EI_{n,2a} \leq Cn^{-6}h_1^{d_1+d_3}n^8 h_1^{-4(d_1+d_3)\delta/(1+\delta)}h_2^{-2d_1\delta/(1+\delta)}$ $\alpha(\iota_n)^{\delta/(1+\delta)} = O(n^2 h_1^{(1-3\delta)(d_1+d_3)/(1+\delta)}h_2^{-2d_1\delta/(1+\delta)}\alpha(\iota_n)^{\delta/(1+\delta)}) = o(1)$ by Assumption A3(iii). For case (b), the number of terms in the summation is of order $O(n^5\iota_n^3)$. If either t_2, t_4, t_6, or t_8, is among the four elements that lie at least at ι_n-distance from all the other elements, one can bound the term $T^{(1)}_{t_1,t_2,t_3,t_4}(\tau)$ as in case (a). Otherwise, bound the term by $Ch_1^{-(d_1+d_3)}$ and the total number of such terms is $O(n^5\iota_n^3)$. Consequently, $EI_{n,2b} = o(1) + n^{-6}h_1^{d_1+d_3}O(n^5\iota_n^3 h_1^{-(d_1+d_3)}) = o(1)$. For case (c), the total number of terms in the summation is of order $O(n^4\iota_n^4)$ and one can readily obtain $EI_{n,2c} = n^{-6}h_1^{d_1+d_3}O(n^4\iota_n^4 + n^4\iota_n^3 h_1^{-(d_1+d_3)}) = o(1)$. Thus $I_{n1} = o_p(1)$ by the Chebyshev inequality.

Similarly, one can show that $E\left(\bar{R}_{n,32}\right) = o(1) + n^{-3}h_1^{(d_1+d_3)/2}O(n^3\iota_n h_2^r) = o(1)$ and $E\left(\bar{R}_{n,32}\right)^2 = o(1) + n^{-6}h_1^{d_1+d_3}O(n^7 h_2^{2r}) = o(1)$. Hence $\bar{R}_{n,32} = o_p(1)$ by the Chebyshev inequality. ■

Putting Lemmas B.1-B.6 together with the fact that $\hat{\sigma}_n^2 = \sigma^2 + o_p(1)$, we have proved Theorem 3.1.

C Proof of other results

Proof of Theorem 3.2. The proof is similar to that of Theorem 3.1 and thus omitted. ■

Proof of Proposition 3.3. The analysis is similar to the proof of Theorem 3.1, now keeping the additional terms in the expansion of $h_1^{(d_1+d_3)/2}ISELR_n$ that were not present under the null, among which only one term is asymptotically non-negligible under $\mathbb{H}_1(\gamma_n)$:

$$
\begin{aligned}
& h_1^{(d_1+d_3)/2}\sum_{t=1}^n I_t \hat{f}_{1t}^{-2}\int \hat{V}_{1t}(\tau)^{-1}\left\{n^{-1}\sum_{s=1}^n K_{ts}\left[F^{[n]}(\tau|X_s,Z_s) - F^{[n]}(\tau|X_s)\right]\right\}^2 dG(\tau) \\
= \; & n^{-1}\sum_{t=1}^n I_t \int V^{[n]}(X_t,Z_t;\tau)^{-1}\triangle(X_t,Z_t;\tau)^2 dG(\tau)\{1+o_p(1)\} \\
= \; & \int_S \int V^{[n]}(x,z;\tau)^{-1}\triangle(x,z;\tau)^2 dG(\tau)dF^{[n]}(x,z) + o_p(1) = \mu + o_p(1).
\end{aligned}
$$

Consequently, $\Pr(\hat{T}_n \geq z|\mathbb{H}_1(\gamma_n)) \to 1 - \Phi(z - \mu/\sigma)$. ■

Proof of Proposition 3.4. The proof is analogous to that of Proposition 3.3 and thus omitted. ■

Proof of Proposition 3.5. The proof follows closely from **TK**. The problem is equivalent to the following variational problem over all piecewise smooth, bounded, square integrable functions from $S \times \mathbb{R} \to \mathbb{R}_+$:

$$
\min_a F_a(e) \text{ s.t. } \int_S \int \int V^2(x,z;y,y')a(x,z;y)a(x,z;y')dF(y|x,z)dF(y'|x,z)d(x,z) = 1, \tag{C.1}
$$

where $e \in \mathbb{R}$ is arbitrarily chosen. Following **TK**, we can show that the Euler-Lagrange equation for the variational problem (C.1) is

$$
\begin{aligned}
& E_{\Psi(x_0,z_0)}[\Psi(x_0,z_0)^2 f_a(e|\Psi(x_0,z_0))]V(x_0,z_0;y_0) \\
= \; & \lambda \int [V^2(x_0,z_0;y_0,y) + V^2(x_0,z_0;y,y_0)]a^*(x_0,z_0;y_0)dF(y|x_0,z_0)
\end{aligned}
$$

for any $(x_0, z_0, y_0) \in S \times \mathbb{R}$, where $E_{\Psi(x_0,z_0)}$ indicates the expectation is over $\Psi(x_0, z_0)$, λ is the Lagrange multiplier for the constraint in (C.1), and a^* is the solution. Then one can guess and verify that $a^*(x, z; y) = 1\{(x, z) \in S\} V(x, z; y)^{-1} [\text{vol}(S)(\pi^2/3 - 3)]^{-1/2}$ solves the variational problem. ■

Proof of Theorem 4.1. Note that the characteristic function $\sup_{y \in \mathbb{R}} |H(y)| \leq 1$, and the proof is analogous to that of Theorem 3.1. The main difference is that we need the following results in place of Lemma A.5: $\sup_{\tau \in \mathbb{R}} \sup_{(x,z) \in S} |\hat{m}_{h_1}(x, z; \tau) - m(x, z; \tau)| = O_p(n^{-1/2} h_1^{-(d_1+d_3)/2} \sqrt{\ln n} + h_1^2)$, and $\sup_{\tau \in \mathbb{R}} \sup_{x \in S_1} |\hat{m}_{h_2}(x; \tau) - m(x; \tau)| = O_p(n^{-1/2} h_2^{-d_1/2} \sqrt{\ln n} + h_2^r)$. The above uniform consistency results can be established in the exact same fashion as done in Lemma A.4. ■

Proof of Theorem 4.2. The argument is identical to the proof of Theorem 3.2. ■

References

Ahn, H., 1997. Semiparametric estimation of a single-index model with nonparametrically generated regressors, Econometric Theory 13, 3-31.

Andrews, D. W. K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383-1414.

Bergsma, W. P., 2011. Nonparametric testing of conditional independence by means of the partial copula. Working paper, London School of Economics and Political Science.

Boente, G., Fraiman R., 1991. Strong uniform convergence rates for some robust equivalent nonparametric regression estimates for mixing processes. International Statistical Review 59, 355-372.

Bouezmarni, T., Rombouts, J. V. K., Taamouti, A. 2010. A nonparametric copula based test for conditional independence with applications to Granger causality. Journal of Econometrics, forthcoming.

Bouezmarni, T., Roy R., Taamouti, A., 2011. Nonparametric tests for conditional independence using conditional distributions. Working paper, Dept. of Economics, Universidad Carlos III de Madrid.

Bierens, H. J., 1982. Consistent model specification test. Journal of Econometrics 20, 105-134.

Black, D. C., Corrigan, P. R., Dowd, M. R., 2000. New dogs and old tricks: do money and interest rates still provide information content for forecasts of output and prices? International Journal of Forecasting 16, 191-205.

Bosq, D., 1996. Nonparametric Statistics for Stochastic Processes. Springer-Verlag, New York.

Carrasco, M., 2010. A regularization approach to the many instruments problem. Working paper, Université de Montréal.

Chen, S. X., Härdle, W., Li., M., 2003. An empirical likelihood goodness-of-fit test for time series. Journal of the Royal Statistical Society, Series B. 65, 663-678.

Chen, B., Hong, Y., 2010. Testing for the Markov property in time series. Econometric Theory, forthcoming.

Copeland, T., 1976. A model of asset trading under the assumption of sequential information arrival. Journal of Finance 45, 379-395.

DeLong, J., Shleifer, A., Summers, L., Waldmann, B. 1990. Positive feedback investment strategies and destabilizing speculation. Journal of Finance 31, 135-155.

de Maros, J. A., Fernandes, M., 2007. Testing the Markov property with high frequency data. Journal of Econometrics 141, 44-64.

Fan Y., Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. Journal of Nonparametric Statistics 10, 245-271.

Fernandes, M., Flores, R. G., 1999. Tests for conditional independence, Markovian dynamics and noncausality. Discussion Paper, European University Instit.

Friedman, B. M., Kuttner, K. N., 1992. Money, income, prices, and interest rates. American Economic Review 82, 472-492.

Friedman, B. M., Kuttner, K. N., 1993. Another look at the evidence on money-income causality. Journal of Econometrics 57, 189-203.

Gallant, R., Rossi, P., Tauchen, G., 1992. Stock prices and volume. Review of Financial Studies 5, 199-242.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall, New York.

Gihman, I. I., Skorohod, A. V., 1974. The Theory of Stochastic Processes. Springer-Verlag, New York.

Granger, C. W. J., 1980. Testing for causality: a personal viewpoint. Journal of Economic Dynamics and Control 2, 329-352.

Hall, P., Heyde C. C., 1980. Martingale Central Limit Theory and Its Applications. Academic Precess, New York.

Hastie, T., Tibshirani R., 1986. Generalized Additive Models, Statistical Science 1, 297-318.

Hiemstra, C., Jones J. D., 1994. Testing for linear and nonlinear Granger causality in the stock price-volume relation. Journal of Finance 49, 1639-1664.

Hong, Y., 2001. A test for volatility spillover with application to exchange rates. Journal of Econometrics 103, 183-224.

Hsieh, D. A., 1989. Testing for nonlinear dependence in daily foreign exchange rates. Journal of Business 62, 339-368.

Hsieh, D. A. 1991. Chaos and nonlinear dynamics: application to financial markets. Journal of Finance 5, 1839-1877.

Huang, M., White, H., 2010. A flexible test for conditional independence. Working Paper, Dept. of Economics, UCSD.

Huang, T.-M., 2010. Testing conditional independence using maximal nonlinear conditional independence. Annals of Statistics 38, 2047-2091.

Kitamura, Y., 2001. Asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica 69, 1661-1672.

Kitamura, Y. Tripathi, G., Ahn, H., 2004. Empirical likelihood-based inference in conditional moment models. Econometrica 72, 1667-1714.

Li, Q.,1999. Consistent model specification tests for time series econometric models. Journal of Econometrics 92, 101-147.

Li, Q., Racine, J. S., 2008. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. Journal of Business & Economic Statistics 26, 423-434.

Linton, O., Gozalo, P., 1997. Conditional independence restrictions: testing and estimation. Discussion Paper, Cowles Foundation for Research in Economics, Yale University.

Newey, K. N., 1994. Kernel estimation of partial means and a general variance estimator. Econometric Theory 10, 233-253.

Owen, A.,1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237-249.

Owen, A., 1990. Empirical likelihood ratio confidence regions. Annals of Statistics 18, 90-120.

Owen, A., 1991. Empirical likelihood for linear models. Annals of Statistics 19, 1725-1747.

Paparoditis, E., Politis, D. N., 2000. The local bootstrap for kernel estimators under general dependence conditions. Annals of the Institute of Statistical Mathematics 52, 139-159.

Rosenbaum, P. R., 1984. Testing the conditional independence and monotonicity assumptions of item response theory. Psychometrika 49, 425-435.

Sheedy, E. 1998. Correlation in currency markets: A risk-adjusted perspective. Journal of International Financial Markets. Institutions & Money 8, 59-82.

Sims, C. A., 1972. Money, income, and causality. American Economic Review 62, 540-552.

Song, K., 2009. Testing conditional independence via Rosenblatt transform. Annals of Statistics 37, 4011-4045.

Spindler, M., Su, L., 2010. Nonparametric testing for asymmetric information. Working paper, Singapore Management University.

Stinchcombe, M. B., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295-325.

Su, L., White, H., 2007. A consistent characteristic-function-based test for conditional independence. Journal of Econometrics 141, 807-834.

Su, L., White, H., 2008. A nonparametric Hellinger metric test for conditional independence. Econometric Theory 24, 829-864.

Su, L., White, H., 2010. Testing structural change in partially linear models. Econometric Theory 26, 1761-1806.

Sun, S., Chiang, C. Y., 1997. Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. Journal of Applied Mathematics and Stochastic Analysis 10, 3-20.

Swanson, N. R., 1998. Money and output viewed through a rolling window. Journal of Monetary Economics 41, 455-473.

Tenreiro, C., 1997. Loi asymptotique des erreurs quadratiques integrees des estimateurs a noyau de la densite et de la regression sous des conditions de dependance. Portugaliae Mathematica 54, 197-213.

Tripathi, G., Kitamura, Y., 2003. Testing conditional moment restrictions. Annals of Statistics 31, 2059-2095.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Statistical Society 54, 426-482.

Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary absolutely regular processes. Z. Wahrsch. Verw. Gebiete 35, 237-252.

Yoshihara, K., 1989. Limiting behavior of generalized quadratic forms generated by absolutely regular processes. In Mandel, P. and Husková, M. (eds.), Proceedings of the Fourth Prague Conference on Asymptotic Statistics, pp. 539-547, Charles University, Prague.