

UC San Diego

Recent Work

Title

Testing Conditional Independence Via Empirical Likelihood

Permalink

<https://escholarship.org/uc/item/35v8g0fm>

Authors

Su, Liangjun
White, Halbert

Publication Date

2003-10-01

Testing Conditional Independence via Empirical Likelihood

Liangjun Su and Halbert White
Department of Economics, UCSD
La Jolla, CA 92093-0508

October 1st, 2003

Abstract

Let $f(y|x, z)$ (resp. $f(y|x)$) be the conditional density of Y given (X, Z) (resp. X). We construct a class of smoothed empirical-likelihood-based tests for the conditional independence hypothesis: $\Pr[f(Y|X, Z) = f(Y|X)] = 1$. We show that the test statistics are asymptotically normal under the null hypothesis and derive their asymptotic distributions under a sequence of local alternatives. The tests are shown to possess a weak optimality property in large samples. Simulation results suggest that the tests behave well 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: β -mixing, Conditional independence, Empirical likelihood, Exchange rates, Granger causality, Local bootstrap, Money and income, Nonlinear dependence, Nonparametric regression, Stock returns, U -statistics.

JEL Classification: C12, C14, C22.

1 Introduction

In this paper, we investigate a class of nonparametric tests of conditional independence. Let X , Y and Z be random variables. As in Su and White (2002, 2003), we write

$$Y \perp Z \mid X \tag{1.1}$$

to denote that Y is independent of Z given X , i.e., $\Pr[f(Y|X, Z) = f(Y|X)] = 1$, where $f(y|x, z)$ is the conditional density¹ of Y given (X, Z) and $f(y|x)$ is that of Y given X .

There are many nonparametric tests of independence or serial independence in the literature², starting with Hoeffding (1948), and followed by those based on empirical distribution functions such as Blum et al. (1961), Skaug and Tjostheim (1993) and Delgado (1996), those based on smoothing methods like Robinson (1991), Skaug and Tjostheim (1996), Zheng (1997) and Hong and White (2000), and other approaches, such as that of Brock et al. (1987). Nevertheless, there are still relatively few nonparametric tests for conditional independence of continuous variables.³ Linton and Gozalo (1997) propose two nonparametric tests of conditional independence for i.i.d. variables based on a generalization of the empirical distribution function. Since the asymptotic null distributions of their test statistics are complicated functionals of a Gaussian process and depend on the underlying distributions, i.e., neither test is distribution free, a bootstrap procedure is needed for calculating the critical values. This hinders its potential application; to date no applications have appeared that we are aware of. Fernandes and Flores (1999) employ a generalized entropy measure to test conditional independence but the asymptotic normal null distribution relies heavily on the choice of suitable weighting functions. Simulation results indicate that their test has poor size properties and low or no power against Granger causality in variance. Recently, Su and White (2002) have proposed a test for conditional independence based on a weighted version of Hellinger distance between the two conditional densities $f(y|x, z)$ and $f(y|x)$, and they show that the asymptotic null distribution of their test statistic is normal. Although their test is easy to implement, it has some limitations in that it uses the same bandwidth sequence in estimating all required joint and marginal densities nonparametrically, and such a procedure is less than satisfactory when the dimension of (X, Y, Z) is above three. More recently, Su and White (2003) have proposed a new test for conditional independence which is based upon the properties of the conditional characteristic functions. The latter test is shown to have better finite sample performance than Su and White (2002) for all processes they examine except for certain GARCH-type processes. Further it has yet to be shown whether it is optimal in some sense.

In this paper, we propose a new class of tests for conditional independence based on the nonparametric likelihood ratio. 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. Third, the Neyman Pearson lemma tells us that the likelihood ratio tests or their asymptotic equivalents possess certain optimality properties. In a series of papers, Owen (1988, 1990, 1991) studies the use of inference based on the nonparametric likelihood ratio, which is particularly useful when testing hypotheses that can be expressed as 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 empirical likelihood for testing such unconditional moment restrictions. Tripathi and Kitamura (2002) extend the empirical likelihood paradigm to handle the testing of a finite

¹We assume throughout that the joint distribution of (X, Y, Z) admits a joint density, $f(x, y, z)$, say.

²For an excellent survey, see Tjostheim (1996).

³For categorical data there are also numerous tests of independence and conditional independence, see Rosenbaum (1984), Agresti (1990) and Yao and Tritchler (1993), among others.

number of conditional moment restrictions. They 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 the application of empirical likelihood 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. We examine these issues in this paper.

As a motivating example, suppose that $\{U_t, t = 1, \dots, n\}$ is a random sample in \mathbb{R}^d and we want to test the null hypothesis $E[U_1] = 0$. Owen's empirical likelihood ratio testing procedure goes as follows. First, maximize the log likelihood under the null hypothesis of a discrete distribution that has support on the data and satisfies the moment condition; i.e., obtain the restricted empirical log likelihood

$$EL^r = \left\{ \max_{p_1, \dots, p_n} \sum_{t=1}^n \log p_t \mid p_t \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t U_t = 0 \right\}.$$

Next, obtain the unrestricted empirical log likelihood

$$EL^u = \left\{ \max_{p_1, \dots, p_n} \sum_{t=1}^n \log p_t \mid p_t \geq 0, \sum_{t=1}^n p_t = 1 \right\}.$$

Finally, construct the empirical (log) likelihood ratio $ELR = 2\{EL^r - EL^u\}$, and reject the null hypothesis if ELR is large. Owen demonstrates that under the null, $ELR \xrightarrow{d} \chi_d^2$. When conditional moment restrictions are used, Tripathi and Kitamura (2002) demonstrate that the above procedure fails and we need to use a smoothed version of the empirical likelihood.

The contributions of this paper lie primarily in three directions. First, we show that a smoothed version of the empirical likelihood 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 smoothed empirical likelihood ratio, $SELR(\tau)$, say. Then one obtains a test statistic by integrating τ out of a weighted version of $SELR(\tau)$. After being appropriately centered and rescaled, the resulting test statistic is shown to be asymptotically distributed as $N(0, 1)$. Second, we study the asymptotic distribution of the test statistic under a sequence of local alternatives and show that our test is weakly optimal in that it attains maximum average local power with respect to a certain space of functions for the local alternatives. Third, unlike most work in the empirical likelihood literature, including that of Tripathi and Kitamura (2002), our tests allow for data dependence and thus is applicable to time series data. This generalization is due to the use of certain kernel weights in the formation of the empirical likelihood function.

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 method of empirical likelihood. A variety of interesting and important hypotheses other than conditional independence in economics and finance, including conditional goodness-of-fit, conditional homogeneity, conditional quantiles and conditional symmetry, can also be studied using our approach.⁴ These tests are naturally suited to helping answer such questions as “Are the distributions of assets, consumption or income implied by a particular dynamic macroeconomic model close to the actual distributions in the data?” “Is there any significant difference in wage distributions between, say, blacks and whites conditional on their

⁴For a different approach, see Inoue (1998) who builds on ideas from Bierens (1990) and de Jong (1996) and proposes a unified approach for consistent testing of linear restrictions on the conditional distribution function of a time series. As in Bierens (1990) and de Jong (1996), the asymptotic null distribution is not standard and the proposed test is conservative for small and moderate sample sizes.

characteristics such as age, education and experience?” or “Does the stock market react symmetrically to positive and negative shocks after taking into account the influence of all fundamentals?”

It is well known that distributional Granger non-causality [Granger (1980)] is a particular case of conditional independence [see Florens and Mouchart (1982), and Florens and Fougere (1996)]. Our tests can be directly applied to test for Granger non-causality with no need to specify a particular linear or nonlinear model.⁵ Further, using the same techniques as in Su and White (2002), it is easy to show that our tests can be applied to the situation where not all variables of interest are continuously valued. In particular, our tests apply to situations where limited dependent variables or discrete conditioning variables are involved. Also, it is common in econometrics that a conditional independence test would naturally be conducted using estimated residuals or other estimated random variables, which are a function of the observed data and some parameter estimators. Although for brevity we do not prove this here, it is highly plausible that our tests easily extend to handle these cases as well, analogous to the results of Su and White (2002).

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 nonparametric tests for conditional independence when there is no parameter estimation and all random variables are continuously valued. We study the asymptotic distribution of the test statistic under both the null hypothesis and a sequence of local alternatives in Section 3, and we discuss a version of our tests based on smoother moment conditions that has better power properties in Section 4. In Section 5, we study the optimality of our tests in terms of average local power. We examine the finite sample performance of our smoother-moment-conditions-based test via Monte Carlo simulation in Section 6, and we apply it to some economic and financial time series data in Section 7. Final remarks are contained in Section 8. All technical details are relegated to Appendices *A* through *C*.

2 Test Statistic Based On The CDFs

In this paper, we are interested in the question of 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) , $t = 1, \dots, n$. For notational simplicity, we assume that $d_2 = 1$ throughout the paper; the generalization to generic d_2 is trivial theoretically.

The joint density (resp. cumulative distribution function) of (X_t, Y_t, Z_t) is denoted by f (resp. F). Below we make reference to several marginal densities of $f(x, y, z)$ 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 $f(x) = \int \int f(x, y, z) dy dz$, where \int denotes integration on the full range of the argument of integration. This notation is compact, and, we hope, sufficiently unambiguous.

Further, 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) . The null of interest is that conditional on X , the random vectors Y and Z are independent, i.e.,

$$H_0 : \Pr [f(Y|X, Z) = f(Y|X)] = 1.$$

The alternative hypothesis is that $f(y|x, z) \neq f(y|x)$ over a non-trivial volume of the support of the joint

⁵In the same spirit, Baek and Brock (1992) propose a nonparametric test for causality based on the so called correlation integral, an estimator of spatial probabilities across time. Hiemstra and Jones (1994) generalize their approach to allow for data dependence and apply the test to aggregate daily stock prices and trading volume data, revealing significant nonlinear causal relations between them.

density f , i.e.,

$$H_1 : \Pr [f(Y|X, Z) = f(Y|X)] < 1.$$

Let $1(\cdot)$ be an indicator function, $F(\tau|x, z) \equiv E[1(Y \leq \tau)|X = x, Z = z]$ and $F(\tau|x) \equiv E[1(Y \leq \tau)|X = x]$. One way to test H_0 is to test the equivalent hypothesis

$$H'_0 : \Pr [F(\tau|X, Z) = F(\tau|X)] = 1 \text{ for all } \tau \in \mathbb{R}.$$

In Section 4, we consider another approach based on a related condition involving the characteristic function. We treat H'_0 first because of its intuitive appeal. Proceeding, fix a point $\tau \in \mathbb{R}$ and for the moment consider testing the hypothesis

$$H_0(\tau) : \Pr [F(\tau|X, Z) = F(\tau|X)] = 1.$$

First we ignore data dependence, the smoothness of the conditional density $f(y|x, z)$ and the fact that $F(\tau|x)$ is unknown. Let $v_{Y,n}$ and $v_{(X,Z),n}$ denote the counting measures on $\{Y_t, t = 1, \dots, n\}$ and $\{(X_t, Z_t), t = 1, \dots, n\}$, respectively. Consider the $n + 1$ sets of probability measures $\mathcal{P}_{Y|(X,Z)=(X_t,Z_t)}^{(n)} \equiv \{P_{Y|(X,Z)=(X_t,Z_t)} \ll v_{Y,n} : \int dP_{Y|(X,Z)=(X_t,Z_t)} = 1, \int [1(Y \leq \tau) - F(\tau|X_t)] dP_{Y|(X,Z)=(X_t,Z_t)} = 0\}$ for $t = 1, \dots, n$, and $\mathcal{P}_{(X,Z)}^{(n)} \equiv \{P_{(X,Z)} \ll v_{(X,Z),n} : \int dP_{(X,Z)} = 1\}$. Let $p_{Y_s|(X_t,Z_t)}$ be the Radon-Nikodym derivative of $P_{Y|(X,Z)=(X_t,Z_t)} \in \mathcal{P}_{Y|(X,Z)=(X_t,Z_t)}^{(n)}$ with respect to $v_{Y,n}$, evaluated at $(Y_s; X_t, Z_t)$, $t, s = 1, \dots, n$. Similarly, let $p_{(X_t,Z_t)}$ denote the Radon-Nikodym derivative of $P_{(X,Z)} \in \mathcal{P}_{(X,Z)}^{(n)}$ with respect to $v_{(X,Z),n}$, evaluated at (X_t, Z_t) , $t = 1, \dots, n$. Define $p_{(Y_t;X_t,Z_t)} = p_{(Y_t|X_t,Z_t)}p_{(X_t,Z_t)}$. The conventional empirical likelihood under the i.i.d. assumption is simply the multinomial likelihood $\prod_{t=1}^n p_{(Y_t;X_t,Z_t)} = \prod_{t=1}^n p_{(Y_t|X_t,Z_t)}p_{(X_t,Z_t)}$ maximized over the Radon-Nikodym derivatives of $P_{Y|(X,Z)=(X_t,Z_t)} \in \mathcal{P}_{Y|(X,Z)=(X_t,Z_t)}^{(n)}$ and $P_{(X,Z)} \in \mathcal{P}_{(X,Z)}^{(n)}$, which, unfortunately, does not yield any meaningful results. This problem is analogous to the failure of likelihood-based function estimation described in Hastie and Tibshirani (1986). The remedy they suggest is to maximize the expected log likelihood instead; see also Tripathi and Kitamura (2002). We apply this idea to our problem and consider maximizing the empirical analog of

$$E\{\log f(X, Y, Z)\} = E\{E[\log f(Y|X, Z)|(X, Z)]\} + E\{\log f(X, Z)\}$$

subject to $p_{(Y_s|X_t,Z_t)} \geq 0$, $p_{(X_t,Z_t)} \geq 0$, $\sum_{s=1}^n p_{(Y_s|X_t,Z_t)} = 1$, $\sum_{t=1}^n p_{(X_t,Z_t)} = 1$, and $\sum_{s=1}^n [1(Y_s \leq \tau) - F(\tau|X_s)]p_{(Y_s|X_t,Z_t)} = 0$. Here, we restrict the empirical analog of the conditional distribution $F(\tau|X_t, Z_t)$ to be the same as a smoothed version of $F(\tau|X_t)$, i.e., $\sum_{s=1}^n F(\tau|X_s)p_{(Y_s|X_t,Z_t)}$. Alternatively, one can use $F(\tau|X_t)$ instead.⁶

Since $F(\tau|x)$ is unknown, we estimate it nonparametrically. For a kernel function L , and bandwidth h_2 , we define⁷

$$L_{h_2}(u) \equiv h_2^{-d_1} L(u/h_2).$$

We estimate $F(\tau|x) \equiv E[1(Y \leq \tau)|X = x]$ by the standard Nadaraya-Watson (NW) kernel regression estimator,

$$\hat{F}_{h_2}(\tau|x) \equiv \frac{\sum_{t=1}^n L_{h_2}(x - X_t) 1(Y_t \leq \tau)}{\sum_{t=1}^n L_{h_2}(x - X_t)}. \quad (2.1)$$

⁶In an early version of this paper, we use $F(\tau|X_t)$ in place of $\sum_{s=1}^n F(\tau|X_s)p_{(Y_s|X_t,Z_t)}$. Using the former requires correcting more bias terms for the resulting test statistic than using the latter.

⁷For simplicity only, we take the multivariate kernel functions L and K below to be a product of the univariate kernel functions l and k , respectively. To keep the notation simple, we do not explicitly indicate the dependence of the bandwidth parameters on the sample size n . We also adopt the same notational convention for kernels L and K as for density f , namely, to indicate which kernel by the list of its arguments or by specifying the dimension of its arguments.

The exact conditions on the choice of the kernel L and bandwidth h_2 are specified in Assumption A2 in the next section.

Thus, the maximization problem we are interested in is:

$$\max_{\{p_{(Y_s|X_t, Z_t)}, p_{(X_t, Z_t)}, t, s = 1, \dots, n\}} n^{-1} \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log p_{(Y_s|X_t, Z_t)} + n^{-1} \sum_{t=1}^n \log p_{(X_t, Z_t)} \quad (2.2)$$

$$\begin{aligned} \text{s.t. } p_{(Y_s|X_t, Z_t)} &\geq 0, p_{(X_t, Z_t)} \geq 0, \sum_{s=1}^n p_{(Y_s|X_t, Z_t)} = 1, \sum_{t=1}^n p_{(X_t, Z_t)} = 1, \text{ and} \\ \sum_{s=1}^n [1(Y_s \leq \tau) - \widehat{F}_{h_2}(\tau|X_s)] p_{(Y_s|X_t, Z_t)} &= 0, \end{aligned} \quad (2.3)$$

where $w_{ts} \equiv K_{h_1}(X_t - X_s, Z_t - Z_s) / \sum_{s=1}^n K_{h_1}(X_t - X_s, Z_t - Z_s)$, $K_{h_1}(u) \equiv h_1^{-(d_1+d_3)} K(u/h_1)$. The w'_{ts} 's are kernel weights familiar from the nonparametric regression literature, and are mathematically quite tractable. Note that we use different kernels and bandwidth sequences in (2.1) and (2.2). Intuitively speaking, using higher order kernel in (2.1) helps to reduce the bias in estimating $F(\tau|x)$, whereas a second order positive kernel is needed in (2.2) to keep the estimators $p_{Y_s|(X_t, Z_t)}$ nonnegative almost surely when the sample size n goes to ∞ . For future use, we denote $\widehat{F}_{h_1}(\tau|x, z) \equiv \sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t) 1(Y_t \leq \tau) / \sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t)$ as the kernel estimator of $F(\tau|x, z)$.

To solve the above maximization problem, let us first rewrite it using joint probabilities. This will greatly simplify the treatment later on. Thus define $p_{ts} = p_{(X_t, Z_t)} p_{(Y_s|X_t, Z_t)}$ to be the probability mass placed at (Y_s, X_t, Z_t) by the joint distribution $P_{(X_t, Z_t)} P_{(Y_s|X_t, Z_t)}$. Since $\sum_{s=1}^n w_{ts} = 1$ for each t , after dropping the inessential factor n^{-1} in (2.2), we can rewrite the maximization problem as

$$\max_{\{p_{ts}, t, s = 1, \dots, n\}} \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log p_{ts}, \text{ s.t. } p_{ts} \geq 0, \sum_{t=1}^n \sum_{s=1}^n p_{ts} = 1, \sum_{s=1}^n [1(Y_s \leq \tau) - \widehat{F}_{h_2}(\tau|X_s)] p_{ts} = 0. \quad (2.4)$$

(2.4) is solved by maximizing the Lagrangian $\mathcal{L} \equiv \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log p_{ts} - \mu (\sum_{t=1}^n \sum_{s=1}^n p_{ts} - 1) - \sum_{t=1}^n \lambda_t \sum_{s=1}^n [1(Y_s \leq \tau) - \widehat{F}_{h_2}(\tau|X_s)] p_{ts}$, where μ is the Lagrange multiplier for the second constraint and $\{\lambda_t \in \mathbb{R}, t = 1, \dots, n\}$ the set of multipliers for the third constraint.⁸

It is easy to verify that the solution to this problem is given by $\widehat{p}_{ts} = w_{ts} / [n + \lambda_t \widehat{g}_s(\tau)]$, where $\widehat{g}_s(\tau) \equiv 1(Y_s \leq \tau) - \widehat{F}_{h_2}(\tau|X_s)$ and each λ_t solves

$$\sum_{s=1}^n \frac{w_{ts} \widehat{g}_s(\tau)}{n + \lambda_t \widehat{g}_s(\tau)} = 0, \quad t = 1, \dots, n. \quad (2.5)$$

Hence we can rewrite the restricted [i.e. under $H_0(\tau)$] smoothed empirical likelihood (*SEL*) as

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

The use of the weights, $\{w_{ts}, t, s = 1, \dots, n\}$, makes the empirical likelihood a smooth function of the data; analyzing such a procedure does not require any extra effort even if the data are weakly dependent.

⁸Both μ and the λ 's are functions of τ . For notational simplicity, we frequently suppress their dependence on τ .

Next we look at the unrestricted problem. For this we solve

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

This can also be solved by the Lagrange multiplier technique to give the solution $\tilde{p}_{ts} = w_{ts}/n$, and we can write the unrestricted *SEL* as

$$SEL^u(\tau) = \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \tilde{p}_{ts} = \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \left\{ \frac{w_{ts}}{n} \right\}.$$

An analog of the parametric likelihood ratio test statistic would then be

$$2[SEL^u(\tau) - SEL^r(\tau)] = 2 \sum_{t=1}^n \sum_{s=1}^n w_{ts} \log \left\{ 1 + \frac{\lambda_t \hat{g}_s(\tau)}{n} \right\}, \quad (2.6)$$

where λ_t solves (2.5). Heuristically speaking, $SEL^u(\tau) - SEL^r(\tau)$ will be small if $H_0(\tau)$ holds and large otherwise. Therefore, it seems sensible to base the test for $H_0(\tau)$ upon (2.6). Nevertheless, for technical reason, we follow Tripathi and Kitamura (2002) and use a modified version of (2.6) for our test because we now restrict ourselves to a situation where we are interested in the behavior of $F(\tau|X, Z) - F(\tau|X)$ only on a certain fixed subset⁹ (S , say) of the support of (X, Z) . Moreover, to facilitate our analysis of optimality, we use a weight function, \hat{V} , in the following definition of the smoothed empirical likelihood ratio (*SELR*),

$$SELR(\tau) = 2 \sum_{t=1}^n I_t \hat{V}(X_t, Z_t; \tau) \sum_{s=1}^n w_{ts} \log \left\{ 1 + \frac{\lambda_t \hat{g}_s(\tau)}{n} \right\},$$

where $I_t \equiv 1\{(X_t, Z_t) \in S\}$, $\hat{V}(X_t, Z_t; \tau) \equiv n^{-1} \sum_{s=1}^n w_{ts} [\hat{g}_s(\tau)]^2$, and each λ_t solves (2.5). One rejects $H_0(\tau)$ for large values of $SELR(\tau)$.

Finally, note that our real interest is to test for H_0 , or equivalently, H'_0 . To do so, we integrate τ out of a weighted version of $SELR(\tau)$. Specifically, our test statistic has the form

$$ISELR_n = \int SELR(\tau) dG(\tau) = 2 \sum_{t=1}^n \int I_t \hat{V}(X_t, Z_t; \tau) \sum_{s=1}^n w_{ts} \log \left\{ 1 + \frac{\lambda_t \hat{g}_s(\tau)}{n} \right\} dG(\tau), \quad (2.7)$$

where $dG(\tau) = g(\tau)d\tau$ is a probability measure with full support on \mathbb{R} .

Before proceeding further, we introduce two density estimates for future use:

$$\hat{f}_{h_1}(x, z) \equiv \frac{1}{n} \sum_{t=1}^n K_{h_1}(x - X_t, z - Z_t),$$

$$\hat{f}_{h_2}(x) \equiv \frac{1}{n} \sum_{t=1}^n L_{h_2}(x - X_t).$$

Note that we have used different kernels and bandwidths in estimating $f(x, z)$ and $f(x)$, paralleling the case for estimating $F(\tau|x, z)$ and $F(\tau|x)$.

⁹Note that S plays the similar role as the fixed trimming set used in Su and White (2002). Technically, it lets us avoid the usual edge effects associated with kernel estimators.

3 The Asymptotic Distribution of the Test Statistic

In this section we focus on the case for which our conditional independence test is based on continuously valued random variables.

3.1 Asymptotic null distribution

We work with the dependence notion of β -mixing. See Appendix A for its definition and other technical material. Our assumptions are as follows.

Assumption A1 (Stochastic Process)

(i) $\{W_t \equiv (X'_t, Y'_t, Z'_t)' \in \mathbb{R}^{d_1+1+d_3} \equiv \mathbb{R}^d, t \geq 0\}$ is a strictly stationary absolutely regular process with mixing coefficients β_m that satisfy $\beta_m = O(\rho^m)$ for some $0 < \rho < 1$.

(ii) $W_t \equiv (X'_t, Y'_t, Z'_t)'$ has a joint distribution F and joint density f such that $f \in \mathcal{W}^d(r)$, i.e., f 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. Furthermore, $\inf_{(x,z) \in S^\epsilon} f(x,z) = b > 0$, 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 probability density function (*p.d.f.*) f_{t_1, \dots, t_j} of $(W_0, W_{t_1}, \dots, W_{t_j})$ ($1 \leq j \leq 5$) is uniformly bounded.

(iv) The function defined by $F(\tau|x)$ 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 (Kernel and bandwidth)

(i) The kernel K is a product kernel of $k : K(u_1, \dots, u_{d_1+d_3}) = \prod_{i=1}^{d_1+d_3} k(u_i)$, where $k : \mathbb{R} \rightarrow \mathbb{R}$ is 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 $l : L(u_1, \dots, u_{d_1}) = \prod_{i=1}^{d_1} l(u_i)$, where $l : \mathbb{R} \rightarrow \mathbb{R}$ is r times continuously differentiable and satisfies

$$\begin{aligned} \int_{\mathbb{R}} u^i l(u) du &= \delta_{i0} \quad (i = 0, 1, \dots, r-1), \\ C_0 \equiv \int_{\mathbb{R}} u^r l(u) du &< \infty, \int_{\mathbb{R}} u^2 l(u)^2 du < \infty, \text{ and} \\ l(u) &= O((1 + |u|^{r+1+\delta})^{-1}) \text{ for some } \delta > 0, \end{aligned}$$

where δ_{ij} is the Kronecker's delta.

(iii) The bandwidth sequences $h_1 = O(n^{-1/\delta_1})$ and $h_2 = O(n^{-1/\delta_2})$ are such that $\delta_1 > 2(d_1 + d_3)$, $\max\{2\delta_1(d_1 - 2)/(2\delta_1 + d_1 + d_3), \delta_1 d_1/(d_1 + d_3)\} < \delta_2 < \delta_1 r \max\{1, 4/d_1 + d_3\}$, $(d_1 + d_3)/\delta_1 + d_1/\delta_2 < 1$ and $(d_1 + d_3)/2\delta_1 + 2r/\delta_2 > 1$.

Assumption A3 (Weight function)

Suppose $dG(\tau) = g(\tau)d\tau$. The weight function $g(\cdot)$ is uniformly bounded, integrable and nonnegative everywhere on its support \mathbb{R} .

Remarks.

Assumption A1(i) requires that $\{W_t\}$ be a stationary absolutely regular process with geometric decay rate. This is standard for application of a central limit theorem for U -statistics for weakly dependent data [e.g., Fan and Li (1999a)]. This condition is not stringent because it is weaker than ϕ -mixing

and many well-known processes are absolutely regular with geometric decay rate.¹⁰ For example, linear stationary ARMA processes satisfy this condition provided the innovation process $\{\varepsilon_t\}$ satisfies certain conditions (e.g. one sufficient condition is that $\{\varepsilon_t\}$ has absolutely continuous distribution with respect to the Lebesgue measure). Moreover, under certain conditions, a large class of processes implied by numerous nonlinear models such as bilinear models, NLAR models and ARCH models satisfy absolute regularity with geometric decay rate [see Fan and Li (1999b)]. 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 A2(iii) specifies conditions on the choice of bandwidth sequences. Under the assumptions made on the bandwidth sequences, we have in particular that $nh_1^{2(d_1+d_3)}/(\ln n)^3 \rightarrow \infty$, $nh_1^{d_1+d_3}h_2^{d_1} \rightarrow \infty$, $nh_1^{-(d_1+d_3)/2}h_2^{d_1-2} \rightarrow \infty$, $h_1^{-(d_1+d_3)/2}h_2^{2r} \rightarrow 0$, $h_1^{(d_1+d_3)/2}h_2^{-(d_1+d_3)/2} \rightarrow 0$, and $nh_1^{(d_1+d_3)}h_2^{2r} \rightarrow 0$. 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 δ_1 and δ_2 .

We will construct a test statistic that is asymptotically normally distributed. To state the first main result, we define the following notation:

$$\begin{aligned} V(x, z; \tau) &\equiv E\{[1(Y \leq \tau) - F(\tau|X, Z)]^2 | X = x, Z = z\} = F(\tau|x, z)[1 - F(\tau|x, z)], \\ V(x; \tau) &\equiv E\{[1(Y \leq \tau) - F(\tau|X)]^2 | X = x\} = F(\tau|x)[1 - F(\tau|x)], \\ V(x, z; \tau, \tau') &\equiv \text{cov}(1(Y \leq \tau), 1(Y \leq \tau') | X = x, Z = z) = F(\tau \wedge \tau'|x, z) - F(\tau|x, z)F(\tau'|x, z), \\ B &\equiv C_1^{(d_1+d_3)} \int_{\mathbb{R}} \int_S V(x, z; \tau) d(x, z) dG(\tau), \end{aligned}$$

where

$$C_1 \equiv \int k(u)^2 du \text{ and } \tau \wedge \tau' = \min(\tau, \tau').$$

For simplicity, we will often omit the integration ranges when there is no confusion.

Further, define

$$\sigma^2 \equiv 2C_3^{(d_1+d_3)} \int_S \int V^2(x, z; \tau, \tau') dG(\tau) dG(\tau') d(x, z),$$

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), the C_i 's can be calculated explicitly. If we use the Gaussian kernel¹¹ for $k(\cdot)$, i.e., $k(u) = \varphi(u)$, the standard normal density function, the C_i 's can be obtained as follows: $C_1 = 1/(2\sqrt{\pi})$, and $C_3 = 1/(2\sqrt{2\pi})$. If we use the Epanechnikov kernel instead, i.e., $k(u) = 0.75(1 - u^2)1(|u| \leq 1)$, then $C_1 = 0.6$, and $C_3 = 0.4338$.

Further, let $\hat{B}_n \equiv \sum_{t=1}^n I_t \sum_{s=1}^n \int [w_{ts} \hat{g}_s(\tau)]^2 dG(\tau)$, $\hat{V}(x, z; \tau, \tau') \equiv \hat{F}_{h_1}(\tau \wedge \tau'|x, z) - \hat{F}_{h_1}(\tau|x, z)\hat{F}_{h_1}(\tau'|x, z)$, $\hat{\sigma}^2 \equiv 2n^{-1}C_3^{(d_1+d_3)} \sum_{t=1}^n \int \int \left[\hat{V}(X_t, Z_t; \tau, \tau') \right]^2 \hat{f}_{h_1}^{-1}(X_t, Z_t) dG(\tau) dG(\tau')$, and

$$\hat{T}_n \equiv h_1^{(d_1+d_3)/2} \{ISELR_n - \hat{B}_n\} / \hat{\sigma}.$$

We can now state our first result.

Theorem 3.1 *Under Assumptions A1 – A3, $\hat{T}_n \xrightarrow{d} N(0, 1)$ under H_0 .*

The proof of this theorem relies on the use of some preliminary results in Tenreiro (1997). Tenreiro uses U -statistic theory to study the asymptotics for the integrated squared error of kernel density estimators

¹⁰It is well known that the ϕ -mixing condition is highly restrictive; for example, an ARMA process is never ϕ -mixing but generally geometrically absolutely regular. See Harel and Puri (1996).

¹¹While 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).

[see also Tenreiro (1995) and Gourieroux and Tenreiro (2001)], and his result can be adapted to our framework. A size- α test for H_0 can be obtained by comparing \widehat{T}_n with the one-sided critical value z_α from the standard normal distribution, i.e., $z_{0.01} = 2.327$, $z_{0.05} = 1.645$ and $z_{0.10} = 1.282$, and we reject the null when $\widehat{T}_n > z_\alpha$.

When we have at most three conditioning variables, i.e., $d_1 + d_3 \leq 3$, let $\widehat{B}_{n,1} \equiv C_1^{(d_1+d_3)} n^{-1} \sum_{t=1}^n \int I_t \times \widehat{V}(X_t, Z_t; \tau) \widehat{f}_{h_1}^{-1}(X_t, Z_t) dG(\tau)$, and

$$\widehat{T}_{n,1} \equiv h_1^{(d_1+d_3)/2} \{ISELR_n - h_1^{-(d_1+d_3)} \widehat{B}_{n,1}\} / \widehat{\sigma}.$$

We have the following corollary.

Corollary 3.2 *Under Assumptions A1 – A3, if $d_1 + d_3 \leq 3$, then $\widehat{T}_{n,1} \xrightarrow{d} N(0, 1)$ under H_0 .*

When $d_1 + d_3 > 3$, $\widehat{B}_{n,1}$ is only a consistent estimator of $B_1 \equiv C_1^{(d_1+d_3)} \int \int_S V(x, z; \tau) d(x, z) dG(\tau)$. The latter is the leading bias term for $ISELR_n$ to be corrected. In both cases, dispensing with smaller order terms, one can show that the global power of the above test is a monotone function of $Q_n \equiv h_1^{(d_1+d_3)/2} \{ISELR_n - h_1^{-(d_1+d_3)} B_1\}$. Thus in principle one can take Q_n as an objective function and choose bandwidths (h_1, h_2) to maximize it.

3.2 Limiting behavior under local alternatives

We now derive the asymptotic power function of \widehat{T}_n under a sequence of local alternatives that approach the null hypothesis as $n \rightarrow \infty$. To generate the local alternatives, we follow the approach of Su and White (2002); namely, the local alternatives are defined by a sequence of densities $f^{[n]}(x, y, z)$ such that, for $f^{[n]}(x, y) \equiv \int f^{[n]}(x, y, z) dz$, $f^{[n]}(x, z) \equiv \int f^{[n]}(x, y, z) dy$, and $f^{[n]}(x) \equiv \int f^{[n]}(x, y, z) dy dz$, we have

$$\|f^{[n]}(x, y, z) - f(x, y, z)\|_\infty = o(n^{-1/2} h_1^{-(d_1+d_2)/4}).$$

Let $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. Let E_n denote expectation under the probability law associated with $f^{[n]}$. Define $F^{[n]}(\tau|x, z) \equiv E_n[1(Y \leq \tau)|X = x, Z = z]$ and $F^{[n]}(\tau|x) \equiv E_n[1(Y \leq \tau)|X = x]$. Given our setup, the local alternative can be specified as¹²

$$H_1(\alpha_n) : \sup \left\{ \left| F^{[n]}(\tau|x, z) - F^{[n]}(\tau|x) - \alpha_n \Delta(x, z; \tau) \right| : (x, z) \in \mathbb{R}^{d_1+d_3}, \tau \in \mathbb{R} \right\} = o(\alpha_n), \quad (3.1)$$

where $\Delta(x, z; \tau)$ satisfies

$$\delta \equiv \int \int_S \Delta^2(x, z; \tau) dF(x, z) dG(\tau) < \infty.$$

The following proposition shows that our test can distinguish local alternatives $H_1(\alpha_n)$ at rate $\alpha_n = n^{-1/2} h_1^{-(d_1+d_3)/4}$ while maintaining a constant level of asymptotic power.

Proposition 3.3 *Under Assumptions A1–A3, suppose that $\alpha_n = n^{-1/2} h_1^{-(d_1+d_3)/4}$ in $H_1(\alpha_n)$. Then, $\Pr(\widehat{T}_n \geq z_\alpha | H_1(\alpha_n)) \rightarrow 1 - \Phi(z_\alpha - \delta/\sigma)$.*

When $d_1 + d_3 \leq 3$, a similar result holds for $\widehat{T}_{n,1}$.

¹²Alternatively, one can specify local alternatives in terms of densities as done in Su and White (2002): $f^{[n]}(y|x, z) = f^{[n]}(y|x)[1 + a_n \Delta(x, y, z) + o(a_n) \widetilde{\Delta}_n(x, y, z)]$. Then $\Delta(x, z; \tau) = \lim_{n \rightarrow \infty} \int H(y + \tau) \widetilde{\Delta}(x, y, z) f^{[n]}(y|x) dy$ in (3.1).

Corollary 3.4 *Under Assumptions A1–A3, suppose that $\alpha_n = n^{-1/2}h_1^{-(d_1+d_3)/4}$ in $H_1(\alpha_n)$. If $d_1 + d_3 \leq 3$, then $\Pr(\widehat{T}_{n,1} \geq z_\alpha | H_1(\alpha_n)) \rightarrow 1 - \Phi(z_\alpha - \delta/\sigma)$.*

The above proposition and corollary say that our test statistics, \widehat{T}_n and $\widehat{T}_{n,1}$, have nontrivial power against $H_1(\alpha_n)$ with $\alpha_n = n^{-1/2}h_1^{-(d_1+d_2)/4}$ whenever $\delta > 0$. The rate $\alpha_n = n^{-1/2}h_1^{-(d_1+d_2)/4}$ is slower than the parametric rate $n^{-1/2}$, as $h_1 \rightarrow 0$, but is usually faster than $n^{-1/4}$. For example, when $d_1 = d_3 = 1$, one can choose $h_1 \propto n^{-1/6}$, $h_2 \propto n^{-1/5}$, and have $n^{-1/2}h_1^{-(d_1+d_2)/4} \propto n^{-5/12}$, which converges to zero faster than $n^{-1/3}$. The rate α_n could be made even closer to $n^{-1/2}$ but is always slower than $n^{-1/2}$. In practice, we need to choose h_1 and h_2 to balance the level and power in finite samples and data-driven methods will be desirable in choosing the h_i 's in simulation and empirical applications.

4 Smoother Moment Conditions

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

4.1 Choice of conditional moment restrictions

Above, we considered a testing procedure based on an empirical analogue of the infinite sequence of expected log likelihood ratios $SELR(\tau)$, built upon the infinite sequence of conditional moment restrictions $F(\tau|x, z) = F(\tau|x)$, indexed by the nuisance parameter¹³ 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 used in forming the sample analogue of these conditions; see (2.3) and (2.4). Motivated by the equivalence of conditional distributions and conditional characteristic functions, we now follow Su and White (2003) [see also Bierens (1982)] and consider a smoother class of conditional moment conditions. For this, we define $H(y) = \int e^{iu'y} dG_0(u)$, the characteristic function of a well-chosen probability measure $dG_0(u)$. Let

$$\psi(u; x, z) = E[\exp(iu'Y)|X = x, Z = z] - E[\exp(iu'Y)|X = x].$$

Then $\int \psi(u; x, z) e^{i\tau'u} dG_0(u) = E[H(Y + \tau)|X = x, Z = z] - E[H(Y + \tau)|X = x] \equiv m(x, z; \tau) - m(x; \tau)$. Under a mild assumption (see Assumption A4 below), the null hypothesis can be expressed as

$$H_0^* : Pr [m(X, Z; \tau) = m(X; \tau)] = 1 \text{ for all } \tau \in \mathbb{R}.$$

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

$$\widehat{ISELR}_n = 2 \sum_{t=1}^n \int I_t \widetilde{V}(X_t, Z_t; \tau) \sum_{s=1}^n w_{ts} \log \left\{ 1 + \frac{\widetilde{\lambda}_t \widetilde{g}_s(\tau)}{n} \right\} dG(\tau), \quad (4.1)$$

where $\widetilde{g}_s(\tau) \equiv H(Y_s + \tau) - \widetilde{m}(X_s; \tau)$, $\widetilde{m}(x; \tau) \equiv \sum_{s=1}^n L_{h_2}(x - X_s) H(Y_s + \tau) / \sum_{s=1}^n L_{h_2}(x - X_s)$, $\widetilde{V}(X_t, Z_t; \tau) \equiv n^{-1} \sum_{s=1}^n w_{ts} [\widetilde{g}_s(\tau)]^2$, and each $\widetilde{\lambda}_t = \widetilde{\lambda}_t(\tau)$ solves

$$\sum_{s=1}^n \frac{w_{ts} \widetilde{g}_s(\tau)}{n + \widetilde{\lambda}_t \widetilde{g}_s(\tau)} = 0, \quad t = 1, \dots, n. \quad (4.2)$$

¹³Strictly speaking, x and z are also nuisance parameters.

Further, let $\tilde{B}_n \equiv \sum_{t=1}^n I_t \sum_{s=1}^n \int [w_{ts} \tilde{g}_s(\tau)]^2 dG(\tau)$ and

$$\tilde{T}_n \equiv h_1^{(d_1+d_3)/2} \{ \widetilde{ISELR}_n - \tilde{B}_n \} / \tilde{\sigma},$$

where $\tilde{\sigma}^2 = 2n^{-1} C_3^{(d_1+d_3)} \sum_{t=1}^n \int \int \tilde{V}^2(X_t, Z_t, \tau, \tau') \hat{f}_{h_1}^{-1}(X_t, Z_t) dG(\tau) dG(\tau')$ and

$$\begin{aligned} \tilde{V}(x, z; \tau, \tau') &\equiv \frac{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s) H(Y_s + \tau) H(Y_s + \tau')}{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s)} \\ &\quad - \frac{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s) H(Y_s + \tau)}{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s)} \frac{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s) H(Y_s + \tau')}{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s)}. \end{aligned}$$

Assumption A4 (Fourier transform)

$dG_0(u) = g_0(u) du$ is such that $g_0(u)$ has full support on \mathbb{R} , is bounded, even, integrable and everywhere positive, and is chosen such that its corresponding characteristic function H is real-valued.

Further we need to replace Assumption A1(iv) by:

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, $|m(x'; \tau) - m(x; \tau)| \leq \alpha(\tau) \|x' - x\|$, where $\int \alpha^2(\tau) dG(\tau) < \infty$.

Theorem 4.1 Under Assumptions A1(i) – (iii), A1(iv*), and A2–A4, $\tilde{T}_n \xrightarrow{d} N(0, 1)$ under H_0 .

In the case when $d_1 + d_3 \leq 3$, define

$$\begin{aligned} \tilde{V}(x, z; \tau) &\equiv \frac{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s) H^2(Y_s + \tau)}{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s)} - \left[\frac{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s) H(Y_s + \tau)}{\sum_{s=1}^n K_{h_1}(x - X_s, z - Z_s)} \right]^2, \\ \tilde{B}_{n,1} &\equiv C_1^{(d_1+d_3)} n^{-1} \sum_{t=1}^n \int I_t \tilde{V}(X_t, Z_t; \tau) / \hat{f}_{h_1}(X_t, Z_t) dG(\tau), \text{ and} \\ \tilde{T}_{n,1} &\equiv h_1^{(d_1+d_3)/2} \{ \widetilde{ISELR}_n - h_1^{-(d_1+d_3)} \tilde{B}_{n,1} \} / \tilde{\sigma}. \end{aligned}$$

We have the following result.

Corollary 4.2 Under Assumptions A1(i) – (iii), A1(iv*), and A2–A4, if $d_1 + d_3 \leq 3$, $\tilde{T}_{n,1} \xrightarrow{d} N(0, 1)$ under H_0 .

Our simulations are conducted by using this specific test statistic, $\tilde{T}_{n,1}$, because we mainly consider the cases $d_1 + d_3 \leq 3$ and when $d_1 + d_3 = 4$, we also find that using $\tilde{B}_{n,1}$ as a bias correction term has better small sample performance than using \tilde{B}_n .

4.2 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, Y, Z) are continuously valued. For example, when Y is discretely valued, our testing procedure can be modified easily to accommodate this case by replacing the integration by summation over possible values of Y . This is more than a superficial change, as it allows the applications of our test to any situations involving limited dependent variables. For example, Y may be a discrete response, or a more complicated censored or truncated version of a continuous (latent) variable. Also, one can allow a mixture of continuous and discrete conditioning variables. The modification can be done by following the approach in Racine and Li (2000).

4.3 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

$$H_0^* : \Pr [f(Y|Z) = f(Y)] = 1.$$

The alternative is that $f(y|z) \neq f(y)$ over a significant range of the support of the joint density $f(y, z)$. One can modify our previous procedure by replacing $\hat{F}_{h_2}(\tau|X_s)$ in (2.4) by $m(\tau) \equiv n^{-1} \sum_{t=1}^n 1(Y_t \leq \tau)$ or $\tilde{m}(X_s; \tau)$ in (4.1) 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.

4.4 The bootstrap and subsampling

One can develop suitable versions of bootstrap or subsampling methods that may improve the small sample performance of our tests. It is routine to justify that subsampling works in our context (Politis et al. 1999). Simulation results suggest that subsampling produces correct critical values but results in significant loss of power despite its high computational cost. Our focus here is thus on the bootstrap.

The basic problem for the bootstrap is how to impose the null hypothesis in the resampling scheme. Simple resampling from the empirical joint distribution of $W_t = (X_t', Y_t', Z_t)'$ will not impose the null restriction. Paparoditis and Politis (2000, PP hereafter) propose a local bootstrap procedure for non-parametric kernel estimators under general dependence conditions. We essentially do the same thing here except that our conditioning variables are not necessarily lagged dependent variables. Let $P_n^{Y|Z|X}$ be the empirical distribution of (Y, Z) conditional on X . $P_n^{Y|X}$ and $P_n^{Z|X}$ are analogously defined. Our proposal consists of drawing resamples $\{X_t^*, Y_t^*, Z_t^*\}_{t=1}^n$, where $X_t^* = X_t$, from the conditional distribution $\tilde{P}_n^{Y_t Z_t | X_t}$ in which we impose the null hypothesis of independence of Y_t and Z_t conditional on X_t . That is,

$$\tilde{P}_n^{Y_t Z_t | X_t} = \tilde{P}_n^{Y_t | X_t} \cdot \tilde{P}_n^{Z_t | X_t},$$

where $\tilde{P}_n^{Y_t | X_t}$ and $\tilde{P}_n^{Z_t | X_t}$ denote the bootstrap conditional distributions of Y_t and Z_t , respectively. We explain only the procedure for computing $\tilde{P}_n^{Y_t | X_t}$ since $\tilde{P}_n^{Z_t | X_t}$ is constructed in the same manner. As in PP, the local bootstrap procedure proposed here is based on the simple idea of obtaining bootstrap replicates of the sequence of observed pairs $\{X_t, Y_t\}_{t=1}^n$ by resampling the observation Y_t given X_t and using a simple and consistent estimator of the conditional distribution function $F(Y_t|X_t)$. In particular, for $t = 1, \dots, n$, denote by (X_t, Y_t^*) the bootstrap replicates of the pair (X_t, Y_t) , where

$$Y_t^* \sim \tilde{P}_n^{Y_t | X_t}(\cdot | X_t), \tag{4.3}$$

and for any set $A \subset \mathbb{R}^{d_2}$, $\tilde{P}_n^{Y_t | X_t}(\cdot | X_t)$ is a version of the empirical conditional distribution given by

$$\tilde{P}_n^{Y_t | X_t}(A | X_t) = \frac{\sum_{s=1}^n \kappa_b(X_t - X_s) 1(Y_s \in A)}{\sum_{s=1}^n \kappa_b(X_t - X_s)}, \tag{4.4}$$

where $\kappa_b(\cdot) = b^{-d_1} \kappa(\cdot/b)$ with κ being a symmetric and square integrable density on \mathbb{R}^{d_1} and satisfying Assumption A4 in PP, and $b > 0$ is called the resampling bandwidth satisfying Assumption A3 in PP.

Note that by (4.3) the distribution of Y_t^* varies with the index t and that conditionally on the observed sequence $\{X_t, Y_t\}_{t=1}^n$, the bootstrap replicates (X_t, Y_t^*) and (X_s, Y_s^*) are independent for $t \neq s$. Thus for

every index t , Y_t^* is a random variable taking values in the set $\{Y_1, \dots, Y_n\}$ with probability mass function given by

$$P(Y_t^* = Y_s | X_t) = \kappa_b(X_t - X_s) / \sum_{r=1}^n \kappa_b(X_t - X_r), \quad s = 1, \dots, n.$$

Similarly, one can obtain the bootstrap replicates (X_t, Z_t^*) for $t = 1, \dots, n$ and hence $\{(X_t, Y_t^*, Z_t^*)\}_{t=1}^n$. Repeating this procedure B times, we obtain B bootstrap resamples and with each sample we compute the test statistic T_n^* in analogous fashion to \widehat{T}_n (or \widetilde{T}_n). The level α critical values \widetilde{c}_α are computed as an approximate solution to

$$\Pr^*[T_n^* > \widetilde{c}_\alpha] = \alpha,$$

where \Pr^* denotes probability conditional on the sample. The consistency of this procedure can be established in exactly the same way as in PP. See PP for the reason why the above procedure “works”.

5 Weak Asymptotic Optimality Property of the ISELR Test

As noted in the introduction, there exist a number of alternative conditional-moment-based tests for conditional independence. Su and White (2003) use a property of conditional characteristic functions to show that their test is consistent and behaves well in finite samples. This section identifies an asymptotically optimal test among a class of conditional moment restrictions that are formulated in terms of conditional distributions. The asymptotic optimality is weak in the sense that we are able to treat only a restricted class of local alternatives. Nevertheless, this is the first non-trivial result available for asymptotic optimality for testing distributional hypotheses in terms of local average power.

Following the approach of Su and White (2003), one can consider a sequence of test statistics that are based upon $\widetilde{\Gamma}(a) \equiv \frac{1}{n} \sum_{s=1}^n \int [\widetilde{F}_{b_1}(\tau | X_s, Z_s) - \widetilde{F}_{b_2}(\tau | X_s)]^2 a(X_s, Z_s) dG(\tau)$, indexed by the weight function, a . Here $\widetilde{F}_{b_1}(\tau | x, z)$ and $\widetilde{F}_{b_2}(\tau | x)$ are nonparametric kernel density estimators of $F(\tau | x, z)$ and $F(\tau | x)$, respectively. They are based upon the univariate kernels k and l and bandwidth sequences b_1 and b_2 , respectively. Note that we could have allowed a to depend on the integrand τ , but this would make the analysis terribly complicated. The test statistics are based upon

$$\eta(a) = \frac{nb_1^{(d_1+d_3)/2} \{\widetilde{\Gamma}(a) - n^{-1}b_1^{-(d_1+d_3)} B_{11} - n^{-1}b_1^{-d_1} B_{12} - n^{-1}b_2^{-d_1} B_{13}\}}{\sqrt{\sigma_1^2}}, \quad (5.1)$$

where

$$\begin{aligned} B_{11} &\equiv C_1^{(d_1+d_3)} \int \int V(x, z; \tau) a(x, z) d(x, z) dG(\tau), \\ B_{12} &\equiv -2C_2^{d_1} \int \int V(x, z; \tau) [f(x, z)/f(x)] a(x, z) d(x, z) dG(\tau), \\ B_{13} &\equiv C_1^{d_1} \int \int V(x; \tau) [f(x, z)/f(x)] a(x, z) d(x, z) dG(\tau), \\ \sigma_1^2 &\equiv 2C_3^{(d_1+d_3)} \int \int V^2(x, z; \tau, \tau') a^2(x, z) d(x, z) dG(\tau) dG(\tau'), \text{ and} \\ C_2 &\equiv k(0). \end{aligned}$$

Since $\eta(ca) = \eta(a)$ for any $c \neq 0$, without loss of generality, we assume that $\int_S a^2(x, z) d(x, z) = 1$.

Now let

$$M(a, \Delta) = \frac{\int_S \int \Delta^2(x, z; \tau) f(x, z) a(x, z) dG(\tau) d(x, z)}{\sqrt{2C_3^{(d_1+d_3)} \int_S \int \int V^2(x, z; \tau, \tau') a^2(x, z) dG(\tau) dG(\tau') d(x, z)}}. \quad (5.2)$$

Under $H_1 \left(n^{-1/2} b_1^{-(d_1+d_3)/4} \right)$ it follows that

$$\eta(a) \xrightarrow{d} N(M(a, \Delta), 1).$$

The asymptotic power of the test with critical value c_α is thus given by

$$\pi(a, \Delta) = 1 - \Phi(c_\alpha - M(a, \Delta)). \quad (5.3)$$

Comparing (5.3) and Theorem 3.1, we can see that our ISELR test is asymptotically equivalent to the $\eta(a)$ test with the weighting function

$$a_{ISELR}(x, z) = 1\{(x, z) \in S\} \|S\|^{-1/2},$$

where $\|S\|$ is the Lebesgue measure of S . We shall demonstrate that this choice of weighting function, which is implicitly achieved by our ISELR test, is optimal in a certain sense.

If Δ were known, it would be easy to derive the optimal weighting function that maximizes (5.2) and thus (5.3). For this, an application of the Cauchy-Schwarz inequality to (5.2) shows that (5.3) is maximized by choosing a weighting function that is proportional to

$$a_\Delta(x, z) = \frac{\int \Delta^2(x, z; \tau) f(x, z) dG(\tau)}{\int \int V^2(x, z; \tau, \tau') dG(\tau) dG(\tau')}. \quad (5.4)$$

The notation $a_\Delta(x, z)$ indicates that the optimal choice of a depends on Δ . This result is not useful since Δ is unknown in practice. It is also clear from (5.4) that there is no uniformly (in Δ) optimal test. This resembles the multi-parameter optimal testing problem considered in the seminal paper of Wald (1943).

In a parametric framework, Wald shows that the likelihood ratio test, and their asymptotic equivalents, for a hypothesis about finite dimensional parameters is optimal in terms of an average power criterion. Loosely put, he considers a weighted average of the power function where uniform weights are given along each probability contour of the distribution of the MLE estimator. Similarly, Andrews and Ploberger (1994) consider optimal inference in a nonstandard testing problem. They derive a test that is optimal with respect to a Wald-type average power criterion. Their optimal test performs well in finite samples, indicating the practical relevance of Wald's approach.

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., $\Delta(x, z, \tau)$ in the above notation). A natural extension of Wald's approach is to consider a probability measure on an appropriate space of functions and let the measure mimic the distribution of the estimator¹⁴ $\tilde{F}_{b_1}(\tau|x, z)$, either explicitly or implicitly. Therefore, we propose to use a probability measure that approximates the asymptotic distribution of the sample path of $\tilde{F}_{b_1}(\tau|x, z)$.

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space. Let $\tilde{\Delta}(x, z, \tau) \equiv \Delta((x, z), \tau; \omega) : S \times \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ be a random function,¹⁵ 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 $\tilde{\Delta}(x, z, \tau) = f^{-1/2}(x, z) V^{1/2}(x, z; \tau) \beta^{-1/2}(x, z) \psi(x, z)$, where $\beta(x, z) \equiv \int V(x, z; \tau) dG(\tau)$ and for $v \equiv (x, z) \in \mathbb{R}^{d_1+d_3}$, $\psi(v) \equiv \prod_{i=1}^{d_1+d_3} \int_0^{1/\gamma_i} \kappa_i(v_i/\gamma_i - z) dU_i(z)$, κ_i are arbitrarily 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 diameter¹⁶ of S restricted in the direction of v_i . We further require $0 < 1/\gamma_i \leq l_i$. This implies that the joint distribution of the bivariate vector $(\int_S f(\psi(v)) dv, \psi(v_0))$

¹⁴It is unnecessary to mimic the distribution of $\tilde{F}_{b_2}(\tau|x)$ because this has no impact on the power function.

¹⁵Alternatively, a random function can be defined by specifying a certain measure on a certain function space whose elements are functions on $S \times R$. See, for example, Gihman and Skorohod (1974, p.44).

¹⁶Without 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 .

does not depend on the location $v_0 \in S$. Other properties of Δ , such as its Gaussianity, are not important in our argument below. Moreover, our optimality result does not depend on the choice of v_i and γ_i .

We are now ready to define our average power concept. Let Q be the probability measure induced by $\tilde{\Delta}$ on continuous functions defined on $S \times \mathbb{R}$. Using the definition of $\tilde{\Delta}$, rewrite the random variable $M(a, \tilde{\Delta})$ as

$$M(a, \tilde{\Delta}) = \frac{1}{\sqrt{2C_3^{(d_1+d_3)}}} \int_S a(x, z) \psi^2(x, z) d(x, z), \quad (5.5)$$

where we have imposed the restriction $\int_S a^2(x, z) d(x, z) = 1$. Let F_a be the cdf of $M(a, \tilde{\Delta})$. The average asymptotic power of the test is the following functional of a :

$$\bar{\pi}(a) = \int \pi(a, \tilde{\Delta}) dQ(\tilde{\Delta}) = \int_0^\infty [1 - \Phi(c_\alpha - m)] F_a(dm). \quad (5.6)$$

Observe that the integrand in (5.6) is strictly increasing in m . So if there exists a smooth, bounded, square integrable function $a^* : S \rightarrow \mathbb{R}_+$ such that $\int_S [a^*(x, z)]^2 d(x, z) = 1$ and for all a the cdf F_{a^*} first order stochastically dominates F_a , then a^* maximizes $\bar{\pi}(a)$. The following proposition delivers the desired result.

Proposition 5.1 *Let $a^* = 1\{(x, z) \in S\} / \|S\|^{-1/2}$, where $\|S\|$ is the Lebesgue measure of S . Then $a^* = \arg \max_{a \in C_b(S)} \bar{\pi}(a)$, where $C_b(S)$ is the space of continuously bounded functions on S .*

This result shows that the ISELR test attains the maximum average local power when the sequence of alternatives are restricted to the space of functions generated by $\tilde{\Delta}$. An alternative way of achieving this optimality is to use $a(x, z) = 1\{(x, z) \in S\}$ in (5.1).

We emphasize that such optimality is weak compared to the asymptotic optimality result in the GMM framework; see Sowell (1996). Like the discussion in Section 4.1, if we replace the conditional-distribution-based moment conditions with their corresponding characteristic-function-based conditions, we would expect that the resulting test in Theorem 4.1 or Corollary 4.2 is optimal with respect to different sequences of local alternatives which are restricted on the space of functions generated by $\tilde{\Delta}^*(x, z, \tau) = f^{-1/2}(x, z) V^{*1/2}(x, z; \tau) \beta^{-1/2}(x, z) \psi(x, z)$, where $V^*(x, z; \tau) = \text{var}(H(Y + \tau) | X = x, Z = z)$.

6 Monte Carlo Results

In this section we report the results of some Monte Carlo simulation experiments designed to examine the finite sample performance of our nonparametric conditional independence test based on \widehat{ISELR}_n . Our simulation covers three cases. We set $d_1 = d_2 = d_3 = 1$ in the first case, $d_1 = 2$ and $d_2 = d_3 = 1$ in the second case, and $d_1 = 3$ and $d_2 = d_3 = 1$ in the third case. For each DGP under study, we standardize the data $\{(X'_t, Y'_t, Z'_t)'\}$, $t = 1, \dots, n$ before implementing our test so that each variable has mean zero and variance one. Further, we let the compact set S expand slowly as the sample size n grows: $S = \left\{ u = (x, z) : |u_i| \leq 0.9\sqrt{\ln n}, i = 1, \dots, d_1 + d_3 \right\}$.

We use the following data generating processes (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 *i.i.d.* $N(0, I_3)$.

For DGP2 through DGP7, $W_t = (Y_{t-1}, Y_t, Z_{t-1})'$, where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, and

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

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

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

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

DGP6: $Y_t = 0.5Y_{t-1} + (0.3 + 0.5\alpha)Z_{t-1}\varepsilon_{1,t}$;

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

DGP8: $W_t = (Y_{t-1}, Y_t, Z_{t-1})'$, where $Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}$, $Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}$, $h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + \alpha Z_{t-1}^2$, $h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2$, where $\{\varepsilon_{1,t}, \varepsilon_{2,t}\}$ are *i.i.d.* $N(0, I_2)$ in DGPs 2-8 and $\alpha = 0.5$ in DGPs 3-8.

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, where these processes are commonly thought to apply. These DGPs are identical to those used in Su and White (2002) and Su and White (2003).

We follow Corollary 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 *p.d.f.* of the standard normal distribution. To save on computation, we avoid using the double integration routine in Matlab. Instead we choose both the second order kernel¹⁷ $k(\cdot)$ and weighting functions $g(\cdot)$ and $g_0(\cdot)$ to be standard normal density functions and work out explicit expressions for the variance estimator, $\tilde{\sigma}^2$. The bias correction estimator $\tilde{B}_{n,1}$ also has a simple analytical expression which does not involve numerical integration. Hence only one numerical integration¹⁸ is needed in each repetition to calculate \widehat{ISELR}_n .

In principle, we can choose the two bandwidth sequences, h_1 and h_2 , to maximize the global power associated with our test. Nevertheless, this approach does not work well in our simulations in that we find the resulting level of our test to be inflated. Instead, we find it useful to choose bandwidths of the form $h_1 = 0.85n^{-1/6}$ and $h_2 = cn^{-1/5}$, where c varies over a compact set on the real line. In some preliminary simulations, we find that the averages of c chosen by leave-one-out least squares cross validation for the marginal density $f(x)$ range between 0.7 to 1.8 across different DGPs.

For DGPs 1 and 2, we first conduct 1000 repetitions for each sample size and each value of c under study. Specifically, we choose n to be 100, 200 and 500 with $c = [0.5 \ 1 \ 2 \ 2.5]$, which includes the range of the averages of our cross-validated values for c . Table 1 reports the empirical rejection frequency of our tests $\tilde{T}_{n,1}$ as a function of c for the sample sizes under investigation. For notational convenience, we shall denote $\tilde{T}_{n,1}$ by SEL_n in Tables 1-4. From the table it appears that the level of SEL_n is well behaved over a large range of values for c . The test is undersized for small values of c . When c increases, the level increases as well. This is true across all samples and for both the 5% and 10% tests. Further, as sample size increases, the level of the test tends to decrease for fixed value of c .

We now compare our test with some previous tests proposed by Linton and Gozalo (1997) and by Su and White (2002, 2003). Linton and Gozalo (1997) base their nonparametric tests of conditional independence on the functional $A_n(w) = \{n^{-1} \sum_{t=1}^n 1(W_t \leq w)\} \{n^{-1} \sum_{t=1}^n 1(X_t \leq x)\} - \{n^{-1} \sum_{t=1}^n 1(X_t \leq x) \times 1(Y_t \leq y)\} \{n^{-1} \sum_{t=1}^n 1(X_t \leq x) 1(Z_t \leq z)\}$, where $w = (x', y', z)'$. Specifically, their test statis-

¹⁷Other kernels have been tried and qualitatively similar results were obtained.

¹⁸See Tripathi and Kitamura (2002) for other practical considerations in implementing a smoothed empirical likelihood ratio based test.

tics are of the Cramér von-Mises and Kolmogorov-Smirnov types: $CM_n = n^{-1} \sum_{t=1}^n A_n^2(W_t)$, $KS_n = \max_{1 \leq t \leq n} |A_n(W_t)|$. The asymptotic null distribution of both test statistics is non-standard so that a local bootstrap procedure is needed to obtain the critical values.¹⁹ To implement the tests, we set the number of bootstrap resamples to be 100, use the product kernel of k as before, and choose the bandwidth parameter, b_n , for the local bootstrap procedure to be $b_n = n^{-1/5}$.

Su and White (2002) base a test for conditional independence 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 $f(x)$. Let $\hat{f}_h(x, y, z)$, $\hat{f}_h(x, y)$, $\hat{f}_h(x, z)$, and $\hat{f}_h(x)$ denote the estimates. Further, define $\Gamma_1 \equiv \frac{1}{n} \sum_{t=1}^n \left[1 - \sqrt{\hat{f}_h(X_t, Y_t) \hat{f}_h(X_t, Z_t) / \hat{f}_h(X_t, Y_t, Z_t) \hat{f}_h(X_t)} \right]^2 \times a(X_t, Y_t, Z_t)$, where a is a weighting function that is compactly supported. Their test statistic, HEL_n , is based on the functional Γ_1 and is asymptotically distributed as $N(0, 1)$ under the null. Su and White (2002) conduct simulations for a variety of bandwidth sequences: $h = n^{-1/\delta}$, where $\delta = 8, 8.5$ and 9 . Nevertheless, we only report here the case $\delta = 8.5$ because the resulting level and power tend to behave better than the other two cases.

Su and White's (2003) 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. Define $\Gamma_2 = \frac{1}{n} \sum_{t=1}^n \int |\hat{m}_{b_1}(X_t, Z_t; \tau) - \hat{m}_{b_2}(X_t; \tau)|^2 I_t dG(\tau)$. Their test statistic, CHF_n , is based on the functional Γ_2 , and is also asymptotically distributed as $N(0, 1)$ under the null. We choose the bandwidth sequences b_1 and b_2 according to Su and White (2003).

Table 2 reports the empirical rejection frequency of the five tests, namely, CM_n , KS_n , HEL_n , CHF_n and SEL_n , for nominal sizes 5% and 10%. In obtaining $\tilde{T}_{n,1}$, we use $h_1 = 0.85n^{-1/6}$ and $h_2 = 2.4n^{-1/5}$. Given the computational burden of our experiments, for all tests there are 1000 Monte Carlo replications in the experiments for $n = 100, 200$ and 500 , and 500 repetitions for $n = 1000$ when the null is true. The number of repetitions is 250 when the null is false.

From Table 2, we see that all five tests have reasonably good size properties for all sample sizes under investigation except that the level of our test $\tilde{T}_{n,1}$ tends to decrease as the sample size increases.²⁰ Even so, we can tell that our test behaves better than or as well as all previous tests in terms of power. From the preceding section, we know that both CHF_n and SEL_n are partially asymptotically optimal with respect to alternatives generated by the same random functions, and it is thus interesting that one test behaves better than the other in finite samples. CHF_n exhibits significantly greater empirical power in detecting conditional dependence (Granger-causality) implied by DGPs 3 through 6 than HEL_n but it is the other way around for DGPs 7 and 8. CM_n and KS_n are dominated by CHF_n in all DGPs but DGP 3 and by HEL_n in DGPs 4 and 6-8. We emphasize that Su and White's HEL_n test has as great power as our test in detecting Granger-causality in GARCH-type processes, but this does not necessarily hold for cases other than $d_1 = d_2 = d_3 = 1$ because of the disadvantage of using the same bandwidth sequence h in estimating all required densities in their approach.²¹

To see how the above tests are sensitive to the pseudo-true parameter α that controls the degree of conditional dependence in DGPs 3-8, we choose 40 different α 's, equally spaced values on the compact

¹⁹The setup of Linton and Gozalo (1997) is for i.i.d. data. One can replace their bootstrap procedure by a local bootstrap (see Section 4.4) to account for data dependence. See Paparoditis and Politis (2000) for more about the local bootstrap.

²⁰This suggests that a larger coefficient c can be used for the bandwidth $h_2 = cn^{-1/5}$ for larger n . For example, a small number of simulations indicate that when $n = 2000$, $c = 3$ can be used. In our applications, however, we conservatively set $c = 2.5$.

²¹The choice of bandwidth for these other cases becomes an extremely difficult task for the test of Su and White (2002).

interval $[0, 0.7]$, in all the above DGPs. For each value of α , we conduct 2000 repetitions and calculate the empirical rejection frequency for the tests CM_n , KS_n , HEL_n , CHF_n and SEL_n .

Figure 1 displays the results for the above six sets of DGPs, with cases (a) through (f) corresponding to DGPs 3-8 when α varies over $[0, 0.7]$. Also reported in Figure 1 is the empirical power function for the conventional linear Granger causality test. In the graphs, Lin stands for the linear causality test, and CM, KS, Hel, Chf and Sel stand for the tests CM_n , KS_n , HEL_n , CHF_n and SEL_n , respectively. When the processes are truly linear, one expects the linear Granger test to be most powerful. This is verified in Figure 1 (a). Except for this case, one can see that the linear Causality test performs worst. From Figure 1, we also see that SEL_n outperforms all the other tests for DGPs 4-7. Like the linear Granger causality test, both CM_n and KS_n have little power in detecting conditional dependence for GARCH-type processes.

In the second case ($d_1 = 2$, $d_2 = d_3 = 1$), we use the following DGP's in our study:

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 *i.i.d.* $N(0, I_2)$.

For DGP2' through DGP7', $W_t = ((Y_{t-1}, Y_{t-2}), Y_t, Z_{t-1})'$, where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, and

DGP2': $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \varepsilon_{1,t}$;

DGP3': $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + \alpha Z_{t-1} + \varepsilon_{1,t}$;

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': $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + (0.3 + 0.5\alpha Z_{t-1})\varepsilon_{1,t}$;

DGP7': $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Y_{t-2}^2 + 0.5\alpha Z_{t-1}^2$; where $\alpha = 0.5$, $\{\varepsilon_{1,t}, \varepsilon_{2,t}\}$ is *i.i.d.* $N(0, I_2)$.

DGP8': same as DGP8.

Since the implementation of HEL_n becomes difficult here because of the previously mentioned bandwidth selection problem, we only study the finite sample behavior of the tests CM_n , KS_n , CHF_n and SEL_n . We use the same kernel and weighting functions and number of bootstrap resamples as in the first case. The only difference is that now we choose the bandwidth sequences differently. Specifically, we set $h_1 = n^{-1/7.5}$ and $h_2 = 1.8n^{-1/6}$ for the test²² SEL_n , and $b_n = n^{-1/6}$ for the CM_n and KS_n tests. Sample sizes $n = 100, 200, 500$ and 1000 are studied. When the null is true, there are 1000 Monte Carlo replications in the experiments for $n = 100, 200$ and 500 , and 500 repetitions for $n = 1000$. The number of repetitions is 250 when the null is false.

Table 3 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 our test dominates all others in terms of power. The CHF_n test dominates CM_n and KS_n for all nonlinear DGPs under investigation. As the dimension of the conditioning variable increases, one might expect that the power of the tests would be adversely affected. Table 3 suggests this conjecture is valid for small sample sizes but the effect of dimensionality is not severe. For both SEL_n and CHF_n , the power is 1 or close to 1 for all DGPs under study when n is 500 for the 10% test, whereas for the tests CM_n and KS_n the power is less than 1 for GARCH-type processes even with $n = 1000$.

To see how the above tests are sensitive to the pseudo-true parameter α that controls the degree of conditional dependence in DGPs 3-8, we choose 40 different α 's, equally spaced values on the compact interval $[0, 0.7]$, in all the above DGPs. For each value of α , we conduct 2000 repetitions and calculate the empirical rejection frequency for the tests CM_n , KS_n , CHF_n and SEL_n . Figure 2 displays the results

²²For $h_1 = n^{-1/7.5}$, we find through preliminary simulations that $h_2 = cn^{-1/6}$ works reasonably well for $c \in [1, 2]$. Like the first case, the test is undersized for smaller values of c and oversized for larger values of c .

for the above six sets of DGPs, with cases (a) through (f) corresponding to DGPs 3-8 when α varies over $[0, 0.7]$. As before, also reported in Figure 2 is the empirical power function for the conventional linear Granger causality test. In the graphs, Lin stands for the linear causality test, and CM, KS, Chf, and Sel stand for the tests CM_n , KS_n , CHF_n and SEL_n , respectively. The results are largely the same as the case where $d_1 = 1$.

In the third case ($d_1 = 3$, $d_2 = d_3 = 1$), we use the following DGP's in our study:

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

For DGP2" through DGP7", $W_t = ((Y_{t-1}, Y_{t-2}, Y_{t-3}), Y_t, Z_{t-1})'$, where $Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}$, and

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

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

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

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

DGP6": $Y_t = 0.5Y_{t-1} + 0.25Y_{t-2} + 0.125Y_{t-3} + (0.3 + \alpha Z_{t-1})\varepsilon_{1,t}$;

DGP7": $Y_t = \sqrt{h_t}\varepsilon_{1,t}$, $h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Y_{t-2}^2 + 0.125Y_{t-3}^2 + \alpha Z_{t-1}^2$; where $\alpha = 0.5$ and $\{\varepsilon_{1,t}, \varepsilon_{2,t}\}$ is *i.i.d.* $N(0, I_2)$.

DGP8": same as DGP8.

We use the same kernel and weighting functions as in the first case. The only difference is that we choose the bandwidth sequences differently and only consider sample sizes²³ $n = 200, 500$ and 1000 . Specifically, we set $h_1 = n^{-\frac{1}{9}}$, $h_2 = 1.4n^{-\frac{1}{7}}$ for SEL_n , and $b_n = n^{-1/7}$ for the CM_n and KS_n tests. The number of repetitions is set as in the above two cases.

Table 4 reports the empirical size and power behavior of our tests. As we can see, the results are similar to the second case above. The test SEL_n outperforms all other tests significantly and the CHF_n test dominates the CM_n and KS_n tests for all nonlinear DGPs in terms of empirical power. The curse of dimensionality also exerts its expected effect.

7 Applications to Economic 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. As documented by Hiemstra and Jones (1994), all previous studies of Granger causal relationship rely exclusively on the traditional linear Granger causality test, which unfortunately has little power in detecting nonlinear relationships.

In this section, we first study the dynamic linkage between pairwise daily exchange rates across four industrialized countries by using both our new empirical likelihood test for conditional independence $\tilde{T}_{n,1}$ and the traditional linear Granger causality test.²⁴ Then with the same technique, we study the dynamic linkage between three US stock market price indices (Dow Jones 65 components, Nasdaq, and S&P 500) and the trading volume in the New York Stock Exchange (NYSE), Nasdaq, and NYSE markets,

²³We don't consider the $n < 200$ case because we need to estimate nonparametrically a 4-dimensional density ($d_1 + d_3 = 4$) and this cannot be done with desirable accuracy with less than 200 observations. Also, when the dimension of the conditioning variables increases, the feasible range for the bandwidth sequences become narrower. For example, if we set $h_1 = n^{-\frac{1}{9}}$, then $h_2 = cn^{-\frac{1}{7}}$ works reasonably well for $c \in [1, 1.6]$ for the sample sizes under investigation. When c is smaller than 1, the level degenerates to 0 at a rapid speed. When c is above 1.6, the level inflates fast too.

²⁴For either test, we only consider two variables at a time, thus omitting possible influences from other variables lurking in the background.

respectively. We then revisit the Granger causal link between exchange rates and stock prices in some developed countries. Finally we use our tool to investigate the relationship between money supply, output, and prices in macroeconomics.

7.1 Application 1: exchange rates

Over the last two decades much research has focused on the nonlinear dependence exhibited by foreign exchange rates, but there is not much research that examines 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—Deutschemark and 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 four industrialized countries, namely, Canada, France, Italy, and the UK and, compare it with the conventional linear test for Granger causality. The data are obtained from Datastream with the sample period from January 2nd, 1995 to December 17th, 2002 with 2077 observations total. 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 in Datastream are missing but entered with the realizations from previous trading days. Moreover, different nations have different national holidays and thus different missing observations. Because we do causality tests with exchange rates from pairwise countries, 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 all time series involved 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 exchange rates (E_t) for all four countries under investigation. The test results indicate that there is a unit root in all level series but not in the first differenced series. Therefore, both Granger causality tests will be conducted on the first differenced data, which we denote as ΔE_t in the following text. Next, since the appropriate formulation of a linear Granger causality analysis may need to incorporate an error correction term into the test if the underlying variables (pairwise E_t here) are cointegrated, we employ Johansen’s likelihood ratio method to examine whether or not exchange rates for pairwise countries are cointegrated. The conclusion is that there is no cointegration between any pair of exchange rates. Consequently, no error correction terms need to be included in the linear Granger causality test.

7.1.1 Linear Granger causality test results

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

$$H_{0,L} : \beta_1 = \dots = \beta_{L_x} = 0 \tag{7.1}$$

holds in

$$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + \dots + \alpha_{L_y} DY_{t-L_y} + \beta_1 DX_{t-1} + \dots + \beta_{L_x} DX_{t-L_x} + \epsilon_t, \tag{7.2}$$

where $\epsilon_t \sim i.i.d.(0, \sigma^2)$ under $H_{0,L}$. An F -statistic can be constructed to check whether the null $H_{0,L}$ 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 the test for a variant²⁵ of $H_{0,L}$:

$$H_{0,L}^* : \beta = 0 \quad (7.3)$$

in

$$DY_t = \alpha_0 + \alpha_1 DY_{t-1} + \dots + \alpha_{L_y} DY_{t-L_y} + \beta DX_{t-i} + \epsilon_t, \quad i = 1, \dots, L_x. \quad (7.4)$$

The results of linear Granger causality tests between pairwise exchange rates are given in Panel A of Table 5, where we choose L_y to be 1, 2 or 3. When it is 1, we also choose L_x to be 1 so that we only check whether DX_{t-1} should enter (7.4) or not. This corresponds to the first row for each country in Panel A. 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 (7.4) or not, which corresponds to the second and third rows for each country in Panel A. The case for $L_y = 3$ is done analogously, corresponding to the fourth to sixth rows.

To summarize the results in Panel A of Table 5, we focus on the 5% test only. First, the test reveals only two Granger causal links. One is from the exchange rate of Canada to that of Italy and the other is from the exchange rate of France to that of the UK. Secondly, there is no bidirectional causal link that is detected by the linear Granger causality test.²⁶ The findings here are intriguing, and they motivate us to ask whether there are some causal links that the linear causality test fails to detect and others that cannot be detected by our nonparametric test for nonlinear Granger causality.

7.1.2 Nonlinear Granger causality test results

To implement our test, we set all smoothing parameters according to those used in the simulations done for Tables 2-4. The null of interest is now

$$H_{0,NL} : \Pr [f(DY_t|DY_{t-1}, \dots, DY_{t-L_y}; DX_{t-1}, \dots, DX_{t-L_x}) = f(DY_t|DY_{t-1}, \dots, DY_{t-L_y})] = 1. \quad (7.5)$$

Due to the ‘‘curse of dimensionality’’, we must choose L_y to be small. Specifically, we study the cases in which $L_y = 1, 2$ and 3 , respectively. Further, for each test we only include one lagged DX_t in the conditioning set. So we actually test a variant²⁷ of $H_{0,NL}$:

$$H_{0,NL}^* : \Pr [f(DY_t|DY_{t-1}, \dots, DY_{t-L_y}; DX_{t-i},) = f(DY_t|DY_{t-1}, \dots, DY_{t-L_y})] = 1, \quad i = 1, \dots, L_x. \quad (7.6)$$

When L_y is 1, we also choose L_x to be 1 so that we only check whether DX_{t-1} should enter (7.6) or not. This corresponds to the first row for each country in Panel B of Table 5. 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 (7.6) or not, which corresponds to the second and third rows for each country in Panel B of Table 5. The case for $L_y = 3$ is done analogously, corresponding to the fourth to sixth rows.

The results in Panel B of Table 5 are interesting. First, unlike the case for the linear Granger causality test, our nonparametric test reveals causal links between all 6 pairs of exchange rates at the 5% significance level. Second, most of the causal links are bidirectional. The only exception is that the French exchange

²⁵Clearly, the null $H_{0,L}^*$ is nested in the null $H_{0,NL}$. The rejection of $H_{0,NL}^*$ indicates the rejection of $H_{0,NL}$ but not the other way around.

²⁶We also conduct the linear Granger causality test for the null (7.1). Applying either the Bayesian information criterion (BIC) or the Akaike information criterion (AIC) to choose the numbers of lags, we find that the exchange rate in Italy is led by that in Canada and there is no other causal link at the 5% significance level.

²⁷The null $H_{0,NL}^*$ is nested in the null $H_{0,NL}$. The rejection of $H_{0,NL}^*$ indicates the rejection of $H_{0,NL}$ but not the other way around. In this sense, our test is conservative.

rate is led by the Italian rate but not the other way around. Third, most of the causal links are robust in that they don't vanish at two- to three-day lag. This suggests that at a one- to three-day lags, the exchange rates across the four countries interact strongly with each other. One obvious reason for the failure of the linear Granger causality test in detecting such causal linkages is that exchange rates exhibit unambiguously nonlinear dependence across markets. The volatility spillover between exchange rates [see Hong (2001) and the reference therein] is a special case of such nonlinear dependence.

To facilitate further comparison between the linear and nonlinear test results, we use boldfaced numbers in Panel A of Table 5 to denote the causal relations which are revealed by the linear Granger causality test but not by our nonparametric test at the 5% significance level. Similarly, the boldfaced numbers in Panel B of Table 5 denote the causal relations which are revealed by our nonparametric test but not by the linear Granger causality test at the 5% significance level. From Table 5, one can see that our nonparametric test can reveal 22 additional causal relations at various lags for all pairs of exchange rates under consideration besides those revealed by the linear causality test, strong evidence in favor of nonlinear dependence between exchange rates. Also, as expected, one can tell from Table 5 that our nonparametric test cannot reveal all linear causal relations. Specifically, it fails to detect two causal relations indicated by the linear causality test.²⁸

7.2 Application 2: stock prices 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 causal 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

²⁸One explanation is that although the nonparametric test has power against all causal relations, linear or nonlinear, causality in the conditional mean or in the conditional variance, it may have low power in some particular direction. If there exists some linear causal relation that is weak, the nonparametric test may fail to detect whereas the linear Granger causality test can pick it up easily. On the other hand, even though the latter test has high power against linear casual relations, it has little or no power against some nonlinear causal relations, which explains why it can't detect some causal relations revealed by the nonparametric test.

nonlinear causality between them. Here we reinvestigate this issue using the latest daily data for the U.S. three major stock market price indices and trading volume. The data are obtained from Yahoo Finance with the sample period from January 2nd, 1995 to January 10th, 2003. After excluding weekends and holidays, the total numbers of observations are 2022 for the Dow Jones 65 composite and Nasdaq series and 2021 for the S&P 500 series. Following the literature, we let P_t and V_t stand for the natural logarithm of stock price indices and volumes multiplied by 100, respectively.

We first employ the augmented Dickey-Fuller test to check for stationarity of $\{P_t\}$ and $\{V_t\}$. The test results indicate that there is a unit root in all level series but not in the first differenced series. Therefore, both Granger causality tests will be conducted on the first differenced data, which we denote as ΔP_t and ΔV_t in the following text. Next, Johansen’s likelihood ratio cointegration tests suggest there is no cointegration between P_t and V_t for all three cases. Consequently, no error correction term needs to be included in the linear Granger causality test. We focus on the causal links between stock returns (ΔP_t) and percentage volume changes (ΔV_t).²⁹

7.2.1 Linear Granger causality test results

We first let ΔP_t and ΔV_t play the roles of DX_t and DY_t in (7.4) and test the null that stock price does not Granger cause trading volume linearly. Then we reverse their roles to test the null that trading volume does not linearly Granger cause stock price. The results of the linear causality test between stock prices and volumes are given in Panel *A* of Table 6. At all levels of L_y , we find causal links from stock prices to trading volumes for the Nasdaq and S&P 500 data but not for the Dow Jones at the 5% significance level. Unambiguously, no causality from trading volume to stock price is revealed by the linear causality test.³⁰

7.2.2 Nonlinear Granger causality test results

We first let ΔP_t and ΔV_t play the roles of DX_t and DY_t in (7.6) and test the null that stock price does not Granger cause trading volume. Then we reverse their roles to test the null that trading volume does not Granger cause stock price. The results for our nonparametric test are reported in Panel *B* of Table 6. From Panel *B*, we find that stock prices lead trading volumes for all three datasets and this is true at all lags of our study. Further, our nonparametric test reveals bidirectional causal relations between stock prices and trading volumes at the one day lag only for the Nasdaq and S&P 500 data, in strong contrast with the results of Hiemstra and Jones (1994) who find bidirectional causal relations for the Dow Jones stock price and trading volume up to a 7-day lag. So like the linear Granger causality test results, our nonparametric test results lend little support to the two theories articulated above.

To facilitate the comparison between the linear and nonlinear test results, we use boldfaced numbers in Panel *A* of Table 6 to denote the causal relations which are revealed by the linear Granger causality test but not by our nonparametric test at the 5% significance level. Similarly, the boldfaced numbers in Panel *B* of Table 6 denote the causal relations which are revealed by our nonparametric test but not

²⁹Alternatively, one can follow Gallant et al. (1992) and Hiemstra and Jones (1994) and remove systematic day-of-the-week and month-of-the year calendar effects from stock returns (ΔP_t) and percentage volume changes (ΔV_t) before conducting the causality tests. Similar results are found with this approach.

³⁰As is done for the case of exchange rates, we also conduct the linear Granger causality test for the null (7.1) by using the BIC and AIC to choose the numbers of lags, L_x and L_y , the maxima of which are set to be 10. According to the BIC, only one linear causal relation is found at the 5% significance level. That is, the S&P 500 stock price index tends to lead the NYSE volume. According to the AIC, all the three stock price indices tend to lead the corresponding trading volumes at the 5% significance level. No other causal links are found.

by the linear Granger causality test at the 5% significance level. From Panel *A*, one can see that our nonparametric test detects all causal relations revealed by the linear causality test. From Panel *B*, one can see that our nonparametric test can reveal 12 extra causal relations at various lags between stock prices and trading volumes besides those revealed by the linear causality test, strong evidence in favor of the nonlinear dependence between the two variables. One obvious reason for the failure of the linear Granger causality test in detecting such causal links is that trading volumes may only have nonlinear predictive power for stock returns.

7.3 Application 3: stock prices and exchange rates

A number of hypotheses support the existence of a causal relation between stock prices and exchange rates. The goods market approach [e.g., Aggarwal (1981)] suggests that changes in exchange rates affect the competitiveness of a multinational firm directly (and that of the domestic firms indirectly), which in turn influence the firm's earnings or its cost of funds and hence its stock prices. The portfolio balance approach [e.g., Kang and Stulz (1997)] stresses the role of capital account transaction. Like all commodities, exchange rates are determined by market mechanisms, i.e., the market demand and supply conditions. A growing domestic stock market would attract capital flows from foreign investors, which may cause an increase in the demand for a country's currency. As a result, rising (declining) stock prices would lead to an appreciation (depreciation) in domestic currency.

Although such theories suggest causal relations between stock prices and exchange rates, existing evidence indicates a weak link between them on a micro level. On a macro level, however, the evidence is mixed. Ma and Kao (1990) find that a currency appreciation negatively affects the domestic stock market for an export-dominant country and positively affects the domestic stock market for an import-dominant country, which seems to be consistent with the goods market theory. Ajayi and Mougoue (1996) find evidence in favor of a dynamic effect of stock prices on exchange rates for eight industrialized countries and Kanas (2002) finds stock return volatility is a significant determinant of exchange rate volatility for the US, the UK and Japan, both lending support to the portfolio balance approach.

Here, we reinvestigate the dynamic linkage between stock prices and exchange rates using daily data for six industrialized countries, namely, Canada, France, Germany, Italy, Japan and the UK. The data are obtained from Datastream with the sample period from December 18th, 1992 to December 17th, 2002 with 2608 observations total. The exchange rates are the local currency against the US dollar. The stock market indices consist of the principal market index for each country, namely, S&P/TSX Composite for Canada, CAC 40 for France, DAX 100 for Germany, BCI Global for Italy, Nikkei 225 for Japan, and FTSE 100 for the UK. Nevertheless, due to national holidays or certain other reasons, some observations for stock prices and exchange rates in Datastream are missing but entered with the realizations from previous trading days. We delete these, and this results in varying number of observations for each country, ranging from 2361 for Japan to 2423 for Canada. As before, we let S_t and E_t stand for the natural logarithms of stock prices and exchange rates multiplied by 100, respectively; and both Granger causality tests will be conducted on the first differenced data, ΔS_t and ΔE_t .

The test results for linear Granger causality between stock prices and exchange rates are given in Panel A of Table 7. To summarize the results, we focus on the case of 5% significance level only. First of all, linear causal links between stock prices and exchange rates are revealed in all countries but the UK. In Japan, it is unidirectional from exchange rate to stock price whereas in other countries, it is bidirectional. Secondly, the magnitudes of the F -statistics show that the causal effect from exchange rates to stock prices is much greater than that from stock prices to exchange rates in all countries but

Canada. Thirdly, when exchange rate causes stock price, it is only a one day effect in all countries but Canada. In comparison, when stock price causes exchange rate, its effect lasts for at least three days in most cases. In short, the linear Granger causality test results indicate that significant causal relations between exchange rates and stock prices exist although there does not exist a universal causal direction.

The nonlinear Granger causality test results are reported in Panel B of Table 7. They suggests that a bidirectional causal relation between exchange rate and stock price exists for all countries but the UK. In the UK, the stock price Granger causes the exchange rate but not the other way around. To compare with the results in Panel A of Table 7, we can find that our nonparametric test complements the linear causality test. There are 17 causal links that are revealed by the nonparametric test only and another 5 revealed by the linear Granger causality test only. The comparison reveals that in modeling the dynamics of exchange rate (resp. stock price) in Canada, France, Germany and Italy, both linear and nonlinear terms for stock price (resp. exchange rate) should appear.

In short, our empirical results show the causal structure is more complex than that implied by either the goods market approach or the portfolio approach alone. A variety of mechanisms are at work. On the one hand, the stock market consistently Granger causes the foreign exchange market in all six countries, which indicates that the focus should be on stabilizing the stock market by means of domestic policy measures. On the other hand, the exchange market has unambiguous impact on the stock market in all countries but the UK, so that the policy-makers in these countries should well be cautious in their implementation of exchange rate policies, since they have significant ramifications for the stock markets.

7.4 Application 4: 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, he developed a test for linear Granger causality and applied it to the U.S. data to examine the causal relationship between money and income, finding the evidence of unidirectional causality from money to income as claimed by the Monetarists. However, his results were not supported by subsequent studies, which indicates that the empirical evidence regarding causal relations between money and the other two variables, income and price, remain inconclusive. 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) are obtained from the Federal Reserve Bank of St. Louis with a sample period from January, 1959 to June, 2003. The total number of observations is 534. 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. Johansen cointegration tests indicate there are three cointegrating pairs, namely, M1 and DPI, M1 and RDPI, and M2 and DPI for the log level data.

Panel A of Table 8 reports the linear Granger causality test results. Two results stand out. First, there is strong evidence of unidirectional causality from the three income variables, DPI, RDPI and IP, to money if M1 is employed as the monetary aggregate. When M2 is employed, however, there is weak evidence of the feedback from money to income. Second, the causal links between money and prices are weak. In particular, if M1 is used, the left part of panel A indicates there is no linkage between the two variables at all. In sum, the linear test lends more support to the Keynesian camp than to the Monetarist camp.

Panel B of Table 8 reports the nonparametric Granger causality test results. They show strong evidence of bidirectional causality between money and income, and between money and prices when M1 is employed. When M2 is used instead, some causal links vanish but others survive. 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.

8 Concluding Remarks

We construct a class of empirical-likelihood-based tests for the null of conditional independence and extend the applicability of empirical likelihood 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 conditional-distribution-based moment restrictions and employing the idea of “smoothed” empirical likelihood, we construct an intuitively appealing test statistic and show that it is asymptotically normal under the null. We also derive its asymptotic distribution under a sequence of local alternatives. Although this test statistic has intuitive appeal, it delivers poor power in small samples because of the discrete nature of the indicator functions used in forming the sample analogue of the moment restrictions. Thus we build on Su and White (2003) and consider a class of smoother moment conditions to construct a new empirical-likelihood-based test. We show that in large samples both tests are weakly optimal in that they attain maximum average local power with respect to different spaces of functions for the local alternatives. Simulations suggest that the smoother-moment-conditions-based test outperforms all previous tests in small 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 the traditional linear Granger causality test fails to detect.

Appendix

A Some Useful Definitions, Lemmas and Theorems

In this appendix, we introduce some useful definitions, lemmas and some theorems which are used in the proofs of the main theorems and propositions in the text.

Definition A.1 Let $\{U_t, t \geq 0\}$ be a d -dimensional strictly stationary stochastic process, and let \mathcal{F}_s^t denote the sigma algebra generated by (U_s, \dots, U_t) for $s \leq t$. The process is called β -mixing or absolutely regular, if as $m \rightarrow \infty$,

$$\beta_m \equiv \sup_{s \in \mathbb{N}} E \left[\sup_{A \in \mathcal{F}_{s+m}^\infty} \{|P(A|\mathcal{F}_{-s}^\infty) - P(A)|\} \right] \rightarrow 0.$$

The following Lemma is due to Yoshihara (1976); see also Li (1999).

Lemma A.2 Let $\{U_t, t \geq 0\}$ be a d -dimensional stochastic process satisfying Assumption A.1(i) in the text. Let $h(v_1, \dots, v_k)$ be a Borel measurable function on \mathbb{R}^{kd} such that for some $\delta > 0$ and given j , $M \equiv \max \left\{ \int_{\mathbb{R}^{kd}} |h(v_1, \dots, v_k)|^{1+\delta} dF(v_1, \dots, v_k), \int_{\mathbb{R}^{kd}} |h(v_1, \dots, v_k)|^{1+\delta} dF^{(1)}(v_1, \dots, v_j) dF^{(2)}(v_{j+1}, \dots, v_k) \right\}$ exists. Then $\left| \int_{\mathbb{R}^{kd}} h(v_1, \dots, v_k) dF(v_1, \dots, v_k) - \int_{\mathbb{R}^{kd}} h(v_1, \dots, v_k) dF^{(1)}(v_1, \dots, v_j) dF^{(2)}(v_{j+1}, \dots, v_k) \right| \leq 4M^{1/(1+\delta)} \beta_m^{\delta/(1+\delta)}$, where $m \equiv i_{j+1} - i_j$, $F, F^{(1)}$ and $F^{(2)}$ are distributions of random vectors $(U_{i_1}, \dots, U_{i_k})$, $V_1 \equiv (U_{i_1}, \dots, U_{i_j})$ and $V_2 \equiv (U_{i_{j+1}}, \dots, U_{i_k})$, respectively; and $i_1 < i_2 < \dots < i_k$.

The next lemma is due to Yoshihara (1989).

Lemma A.3 Let h be defined as above; then $E|E[h(V_1, V_2)|V_1] - E_{V_1}h(V_1, V_2)| \leq 4M^{1/(1+\delta)} \beta_m^{\delta/(1+\delta)}$, where $E_{V_1}h(V_1, V_2) \equiv H(V_1)$ with $H(v_1) \equiv E[h(v_1, V_2)]$.

Now, let $h_n(\cdot, \cdot)$, $n = 1, 2, \dots$, be Borel measurable functions on $\mathbb{R}^d \times \mathbb{R}^d$. Suppose $E[h_n(U_0, v)] = 0$ and $h_n(u, v) = h_n(v, u)$ for all $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$. Define $\mathcal{H}_n \equiv n^{-1} \sum_{1 \leq i < j \leq n} [h_n(U_i, U_j) - E h_n(U_i, U_j)]$, a degenerate U -statistic of order 2. Let $p > 0$ and let $\{\bar{U}_t, t \geq 0\}$ be an *i.i.d.* sequence where \bar{U}_0 is an independent copy of U_0 . Further, define

$$\begin{aligned} u_n(p) &\equiv \max \left\{ \max_{1 \leq i \leq n} \|h_n(U_i, U_0)\|_p, \|h_n(U_0, \bar{U}_0)\|_p \right\}, \\ v_n(p) &\equiv \max \left\{ \max_{1 \leq i \leq n} \|G_{n0}(U_i, U_0)\|_p, \|G_{n0}(U_0, \bar{U}_0)\|_p \right\}, \\ w_n(p) &\equiv \|G_{n0}(U_0, U_0)\|_p, \\ z_n(p) &\equiv \max_{0 \leq i \leq n} \max_{1 \leq j \leq n} \left\{ \|G_{nj}(U_i, U_0)\|_p, \|G_{nj}(U_0, U_i)\|_p, \|G_{nj}(U_0, \bar{U}_0)\|_p \right\}, \end{aligned}$$

where $G_{ni}(u, v) \equiv E[h_n(U_i, u)h_n(U_0, v)]$, and $\|\cdot\|_p \equiv \{E|\cdot|^p\}^{1/p}$.

Theorem A.4 (Tenreiro 1997). 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) $z_n(2)n^{\gamma_1} = O(1)$; and (v) $E[h_n(U_0, \bar{U}_0)]^2 = 2\tilde{\sigma}^2 + o(1)$. Then $\mathcal{H}_n \xrightarrow{d} N(0, \tilde{\sigma}^2)$.

Lemma A.5 Under Assumptions A1-A2,

$$\begin{aligned} (i) \sup_{\tau \in \mathbb{R}} \sup_{(x,z) \in S} |\widehat{F}_{h_1}(\tau|x, z) - F(\tau|x, z)| &= O_p(\mu_{1n}), \\ (ii) \sup_{\tau \in \mathbb{R}} \sup_{x \in S_1} |\widehat{F}_{h_2}(\tau|x) - F(\tau|x)| &= O_p(\mu_{2n}), \end{aligned}$$

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 $\mu_{2n} \equiv n^{-1/2} h_2^{-d_1/2} \sqrt{\ln n} + h_2^2$.

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: $\sup_{\tau \in \mathbb{R}} \sup_{(x,z) \in S} \left| \widehat{F}_{h_1}(\tau|x,z) - F(\tau|x,z) \right| = O_{a.s.} \left(n^{-1/2} h_1^{-(d_1+d_3+\epsilon)/2} (\ln n)^2 + h_1^2 \right)$, 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: $S^\epsilon \equiv \{u \in \mathbb{R}^{d_1+d_3} : \|u - v\| \leq \epsilon \text{ for some } v \in S\}$.

Lemma A.6 Under Assumptions A1-A2 and H_0 ,

- (i) $\sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} |I_t \sum_{s=1}^n w_{ts} \widehat{g}_s(\tau)| = O_p(\mu_n)$,
- (ii) $\sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} \left| I_t [\widehat{V}(X_t, Z_t; \tau) - V(X_t, Z_t; \tau)] \right| = O_p(\mu_{1,n})$,

where $\mu_n \equiv n^{-1/2} h_1^{-(d_1+d_3)/2} \sqrt{\ln n}$.

Proof. Denote $K_{ts} = K_{h_1}(X_t - X_s, Z_t - Z_s)$, $K_{(x,z),s} = K_{h_1}(x - X_s, z - Z_s)$, $L_{ts} = L_{h_2}(X_t - X_s)$, $L_{x,s} = L_{h_2}(x - X_s)$, $f_{1t} = f(X_t, Z_t)$, $\widehat{f}_{1t} = \widehat{f}_{h_1}(X_t, Z_t)$, and $\widehat{f}_{2t} = \widehat{f}_{h_2}(X_t)$.

(i) Under H_0 , write $\sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} |I_t \sum_{s=1}^n w_{ts} \widehat{g}_s(\tau)| \leq \sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} |I_t \sum_{s=1}^n w_{ts} [1(Y_s \leq \tau) - F(\tau|X_s, Z_s)]| + \sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} \left| I_t \sum_{s=1}^n w_{ts} [F(\tau|X_s) - \widehat{F}_{h_2}(\tau|X_s)] \right| \equiv A_{1n} + A_{2n}$. By Newey (1994, Lemma B.1), $\sup_{1 \leq t \leq n} I_t \left| \widehat{f}_{1t} - E\widehat{f}_{1t} \right| = O_p(\mu_n)$ and $\sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} \left| n^{-1} I_t \sum_{s=1}^n K_{ts} [1(Y_s \leq \tau) - F(\tau|X_s, Z_s)] \right| = O_p(\mu_n)$. Therefore, $A_{1n} = O_p(\mu_n)$ by Tripathi and Kitamura (2002, Lemma C.4). Noting that $\sum_{s=1}^n w_{ts} = 1$, $A_{2n} \leq \sup_{\tau \in \mathbb{R}} \sup_{x \in S_1} \left| \widehat{F}_{h_2}(\tau|x) - F(\tau|x) \right| = O_p(\mu_{2n}) = o_p(\mu_n)$ by Assumption A.2(iii). The desired result follows.

(ii) Recall that $\widehat{V}(X_t, Z_t; \tau) = \widehat{f}_{1t}^{-1} n^{-1} \sum_{s=1}^n K_{ts} \left[1(Y_s \leq \tau) - \widehat{F}_{h_2}(\tau|X_s) \right]^2$. So under H_0 , by the triangle inequality, we have

$$\begin{aligned} & \sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} \left| I_t [\widehat{V}(X_t, Z_t; \tau) - V(X_t, Z_t; \tau)] \right| \\ & \leq \sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} \left| I_t \widehat{f}_{1t}^{-1} n^{-1} \sum_{s=1}^n K_{ts} [1(Y_s \leq \tau) - F(\tau|X_s, Z_s)]^2 - V(X_t, Z_t; \tau) \right| \\ & \quad + \sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} \left| 2 I_t \widehat{f}_{1t}^{-1} n^{-1} \sum_{s=1}^n K_{ts} [1(Y_s \leq \tau) - F(\tau|X_s, Z_s)] [F(\tau|X_s) - \widehat{F}_{h_2}(\tau|X_s)] \right| \\ & \quad + \sup_{\tau \in \mathbb{R}} \sup_{1 \leq t \leq n} \left| I_t \widehat{f}_{1t}^{-1} n^{-1} \sum_{s=1}^n K_{ts} [F(\tau|X_s) - \widehat{F}_{h_2}(\tau|X_s)]^2 \right| \\ & \equiv \xi_{1,n} + \xi_{2,n} + \xi_{3,n}. \end{aligned}$$

First,

$$\begin{aligned} \xi_{1,n} & \leq \sup_{\tau \in \mathbb{R}} \sup_{(x,z) \in S} \left| \widehat{f}_{h_1}^{-1}(x,z) n^{-1} \sum_{s=1}^n K_{(x,z),s} [1(Y_s \leq \tau) - F(\tau|X_s, Z_s)]^2 - V(x,z;\tau) \right| \\ & \leq \sup_{\tau \in \mathbb{R}} \sup_{(x,z) \in S} \left| \int K(u,v) [1(y \leq \tau) - F(\tau|x,z)]^2 f(y;x,z) / f(x,z) dy dudv (1 + O_p(\mu_{1,n})) - V(x,z;\tau) \right| \\ & = O_p(\mu_{1,n}), \end{aligned}$$

where $\mu_{1,n}$ is defined in Lemma A.5. Now, for large enough n , take $\epsilon = h_1$. Since K has compact support on $[-1,1]^{d_1+d_3}$ by Assumption A2(i), we have, by Assumption A2(iii),

$$\xi_{2,n} \leq 2 \sup_{\tau \in \mathbb{R}} \sup_{x \in S_1^\epsilon} \left| F(\tau|x) - \widehat{F}_{h_2}(\tau|x) \right| = O_p(\mu_{2,n}) = o_p(\mu_n); \text{ and}$$

$$\xi_{3,n} \leq \sup_{\tau \in \mathbb{R}} \sup_{x \in S_1^\epsilon} \left| F(\tau|x) - \widehat{F}_{h_2}(\tau|x) \right|^2 = O_p\left((\mu_{2,n})^2\right) = o_p(\mu_{1,n});$$

where $S_1^\epsilon = S^\epsilon \cap \mathbb{R}^{d_1}$ and $\mu_{2,n}$ is defined in Lemma A.5. This completes the proof. ■

B Proof Theorem 3.1

In this appendix, C is a generic constant which may vary from case to case. Denote $W_t = (X_t', Y_t, Z_t')$, $f_{1t} = f(X_t, Z_t)$, $\widehat{f}_{1t} = \widehat{f}_{h_1}(X_t, Z_t)$, $f_{2t} = f(X_t)$, $\widehat{f}_{2t} = \widehat{f}_{h_2}(X_t)$, $K_{ts} = K_{h_1}(X_t - X_s, Z_t - Z_s)$, $L_{ts} = L_{h_2}(X_t - X_s)$, $K_{(x,z),t} = K_{h_1}(x - X_t, z - Z_t)$, $L_{x,t} = L_{h_1}(x - X_t)$, and $r_n(W_t; x, z, \tau) = K_{(x,z),t}[1(Y_t \leq \tau) - F(\tau|X_t, Z_t)]$. The bar notation denotes an *i.i.d.* process. For example, $\{\overline{W}_t, t \geq 0\}$ is an *i.i.d.* sequence having the same marginal distributions as $\{W_t, t \geq 0\}$. See Lemma B.3 for details.

Lemma B.1 *Let Assumptions A1 – A3 hold. Then, under H_0 ,*

$$ISELR_n = \widehat{B}_n + \widehat{R}_n + o_p(h_1^{-(d_1+d_3)/2}),$$

where $\widehat{B}_n = \sum_{t=1}^n I_t \int \sum_{s=1}^n w_{ts}^2 \widehat{g}_s(\tau)^2 dG(\tau)$, and $\widehat{R}_n = \sum_{t=1}^n I_t \sum_{s=1}^n \sum_{j=1, j \neq s}^n \int w_{ts} \widehat{g}_s(\tau) w_{tj} \widehat{g}_j(\tau) dG(\tau)$.

Proof. From (2.5), we have

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

where

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

Consequently (recall $\lambda_t = \lambda_t(\tau)$),

$$I_t \widehat{V}(X_t, Z_t; \tau) \lambda_t(\tau) = n I_t \sum_{s=1}^n w_{ts} \widehat{g}_s(\tau) + I_t r_{1t}(\tau). \quad (\text{B.1})$$

Eq. (2.5) also implies

$$\sum_{s=1}^n \frac{w_{ts} [\lambda_t \widehat{g}_s(\tau)]^2}{n + \lambda_t \widehat{g}_s(\tau)} = \sum_{s=1}^n w_{ts} \widehat{g}_s(\tau) \lambda_t.$$

Hence, as $n + \lambda_t \hat{g}_s(\tau) > 0$ (because $\hat{p}_{ts} \geq 0$, $w_{ts} \geq 0$ and $\hat{p}_{ts} = w_{ts}/[n + \lambda_t \hat{g}_s(\tau)]$),

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

Thus by Lemma A.6 (i),

$$\sup_{\tau \in \mathbb{R}} \max_{1 \leq t \leq n} I_t |r_{1t}(\tau)| = O_p(\mu_n) \sup_{\tau \in \mathbb{R}} \max_{1 \leq t \leq n} |\lambda_t(\tau)|$$

and

$$I_t \hat{V}(X_t, Z_t; \tau) \lambda_t(\tau) = O_p(n\mu_n) + O_p(\mu_n) |\lambda_t(\tau)|.$$

Consequently,

$$\sup_{\tau \in \mathbb{R}} \max_{1 \leq t \leq n} I_t \lambda_t(\tau) = O_p(n\mu_n) \text{ and } \sup_{\tau \in \mathbb{R}} \max_{1 \leq t \leq n} I_t |r_{1t}(\tau)| = O_p(n\mu_n^2).$$

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

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

where the remainder term $|I_t \bar{\eta}_{ts}(\tau)| \leq C |I_t \lambda_t \hat{g}_s(\tau)/n|^3 = O_p(\mu_n^3)$ uniformly in t, s and τ because $\sup_{\tau \in \mathbb{R}} \max_{1 \leq s, t \leq n} |\lambda_t \hat{g}_s(\tau)/n| \leq 1$ for large enough n .

Using (2.7), (B.1), and (B.2), a little algebra shows that w.p.a.1,

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

Noting that $n^{-2} \sum_{t=1}^n \int I_t r_{1t}^2(\tau) dG(\tau) \leq n \sup_{\tau \in \mathbb{R}} \max_{1 \leq s, t \leq n} [r_{1t}(\tau)/n]^2 = O_p(n\mu_n^4) = o_p(h_1^{-(d_1+d_3)/2})$, and $\sum_{t=1}^n I_t \int \sum_{s=1}^n w_{ts} \hat{V}(X_t, Z_t; \tau) \bar{\eta}_{ts}(\tau) dG(\tau) \leq n \sup_{\tau \in \mathbb{R}} \max_{1 \leq s, t \leq n} |\hat{V}(X_t, Z_t; \tau) \bar{\eta}_{ts}(\tau)| = O_p(n\mu_n^3) = o_p(h_1^{-(d_1+d_3)/2})$ by Assumption A2(iii), the conclusion of the lemma follows. ■

Lemma B.2 *Let Assumptions A1 – A3 hold. Then $h_1^{(d_1+d_3)/2} \hat{R}_n \xrightarrow{d} N(0, \sigma^2)$ under H_0 .*

Proof. Under H_0 , write

$$\hat{R}_n = n^{-2} \sum_{t=1}^n I_t \hat{f}_{1t}^{-2} \sum_{s=1}^n \sum_{j=1, j \neq s}^n \int K_{ts} K_{tj} \hat{g}_s(\tau) \hat{g}_j(\tau) dG(\tau) = R_{n,1} + R_{n,2} + 2R_{n,3}, \quad (\text{B.3})$$

where

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

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

$$R_{n,3} \equiv n^{-2} \sum_{t=1}^n I_t \sum_{s=1}^n \sum_{j=1, j \neq s}^n \widehat{f}_{1t}^{-2} K_{ts} K_{tj} \int [1(Y_s \leq \tau) - F(\tau|X_s)] [F(\tau|X_j) - \widehat{F}_{h_2}(\tau|X_j)] dG(\tau).$$

It suffices to show

$$h_1^{(d_1+d_3)/2} R_{n,1} \xrightarrow{d} N(0, \sigma^2), \quad (\text{B.4})$$

$$h_1^{(d_1+d_3)/2} R_{n,2} = o_p(1), \text{ and} \quad (\text{B.5})$$

$$h_1^{(d_1+d_3)/2} R_{n,3} = o_p(1). \quad (\text{B.6})$$

Let

$$\widetilde{R}_{n,1} = n^{-1} \sum_{s=1}^n \sum_{t=1, t \neq s}^n \int \int_S \widehat{f}_{h_1}^{-2}(x, z) r_n(W_t; x, z, \tau) r_n(W_s; x, z, \tau) dF(x, z) dG(\tau). \quad (\text{B.7})$$

By Lemma B.3, $h_1^{(d_1+d_3)/2} R_{n,1} = h_1^{(d_1+d_3)/2} \widetilde{R}_{n,1} + o_p(1)$; and by Lemma B.4, $h_1^{(d_1+d_3)/2} \widetilde{R}_{n,1} \xrightarrow{d} N(0, \sigma^2)$. Thus (B.4) follows. Next, write

$$\begin{aligned} R_{n,2} &= \sum_{t=1}^n I_t \int \left\{ \sum_{s=1}^n w_{ts} [\widehat{F}_{h_2}(\tau|X_s) - F(\tau|X_s)] \right\}^2 dG(\tau) \\ &\quad - \sum_{t=1}^n I_t \sum_{s=1}^n w_{ts}^2 \int [\widehat{F}_{h_2}(\tau|X_s) - F(\tau|X_s)]^2 dG(\tau) \\ &\equiv \widetilde{R}_{n,2} - \widetilde{R}_n. \end{aligned} \quad (\text{B.8})$$

Let $\epsilon = h_1$ and $I_s^\epsilon = 1 \{(X_s, Z_s) \in S^\epsilon\}$. Since $\sum_{s=1}^n \int I_s^\epsilon [\widehat{F}_{h_2}(\tau|X_s) - F(\tau|X_s)]^2 dG(\tau) = O_p(h_2^{-d_1})$ (see Tenreiro (1997) and Su and White (2003) in a similar context), we have, w.p.a.1, $h_1^{(d_1+d_3)/2} \widetilde{R}_n \leq C n^{-1} h_1^{-(d_1+d_3)/2} \sum_{s=1}^n \int I_s^\epsilon [\widehat{F}_{h_2}(\tau|X_s) - F(\tau|X_s)]^2 dG(\tau) = O_p(n^{-1} h_1^{-(d_1+d_3)/2} h_2^{-d_1}) = o_p(1)$ by Assumption A2(iii). By Lemma B.6, $h_1^{(d_1+d_3)/2} \widetilde{R}_{n,2} = o_p(1)$. So (B.5) holds. Finally, (B.6) holds by Lemma B.7. The proof is complete. ■

Lemma B.3 *Let Assumptions A1–A3 hold. Then $h_1^{(d_1+d_3)/2} R_{n,1} = h_1^{(d_1+d_3)/2} \widetilde{R}_{n,1} + o_p(1)$, where $R_{n,1}$ and $\widetilde{R}_{n,1}$ are defined by (B.3) and (B.7), respectively.*

Proof. Let $\widehat{F}(x, z)$ denote the empirical distribution function of the random sample $\{X_t, Z_t\}_{t=1}^n$. Then we can write

$$\begin{aligned} & h_1^{(d_1+d_3)/2} \{R_{n,1} - \widetilde{R}_{n,1}\} \\ &= h_1^{(d_1+d_3)/2} n^{-1} \sum_{s=1}^n \sum_{t=1, t \neq s}^n \int_S \int r_n(W_s; x, z, \tau) r_n(W_t; x, z, \tau) \widehat{f}_{h_1}^{-2}(x, z) dG(\tau) d[\widehat{F}(x, z) - F(x, z)] \\ &= \widetilde{\Delta}_{n,1} + \widetilde{\Delta}_{n,2} + \widetilde{\Delta}_{n,3}, \end{aligned}$$

where $\widetilde{\Delta}_{n,1} = h_1^{(d_1+d_3)/2} n^{-2} \sum_{j \neq t, j \neq s, t \neq s} \{ \int I_j r_n(W_s; X_j, Z_j, \tau) r_n(W_t; X_j, Z_j, \tau) \widehat{f}_{h_1}^{-2}(X_j, Z_j) dG(\tau) - \int_S \int r_n(W_s; x, z, \tau) r_n(W_t; x, z, \tau) \widehat{f}_{h_1}^{-2}(x, z) dG(\tau) dF(x, z) \}$

is the summation of the centered terms with $j \neq s$, $j \neq t$, and $t \neq s$;

$$\tilde{\Delta}_{n,2} = 2h_1^{(d_1+d_3)/2} n^{-2} \sum_{s=1, t \neq s} \int I_s r_n(W_s; X_s, Z_s, \tau) r_n(W_t; X_s, Z_s, \tau) \hat{f}_{h_1}^{-2}(X_s, Z_s) dG(\tau)$$

is the summation of the terms for $j = s$ or $j = t$; and

$$\tilde{\Delta}_{n,3} = -2h_1^{(d_1+d_3)/2} n^{-2} \sum_{s=1, t \neq s} \int_S \int r_n(W_s; x, z, \tau) r_n(W_t; x, z, \tau) \hat{f}_{h_1}^{-2}(x, z) dG(\tau) dF(x, z)$$

is the summation of the centering terms for $\tilde{\Delta}_{n,2}$.

Dispensing with the simplest term first, we have $-\tilde{\Delta}_{n,3} = 2n^{-1} \left\{ h_1^{(d_1+d_3)/2} \tilde{R}_{n,1}^{(1)} \right\} = n^{-1} O_p(1) = o_p(1)$ by Lemma B.4. It is difficult to show that the other two terms are small. Our strategy is to use Lemmas A.2–A.3 repeatedly and show asymptotically negligibility in that $\tilde{\Delta}_{n,i} = o_p(1)$, for $i = 1$ and 2 .

Write $\tilde{\Delta}_{n,2} = 2h_1^{(d_1+d_3)/2} n^{-2} \sum_{s=1, t \neq s} \int I_s r_n(W_s; X_s, Z_s, \tau) r_n(W_t; X_s, Z_s, \tau) f^{-2}(X_s, Z_s) dG(\tau) \{1 + o_p(1)\} \equiv \Delta_{n,2} \{1 + o_p(1)\}$; so it suffices to show $\Delta_{n,2} = o_p(1)$. For $s \neq t$, by Assumptions A1(ii)–(iii), one can show that uniformly (recall that the bar notation means *i.i.d.* sequence and $E[r_n(W_s; X_s, Z_s, \tau) | X_s, Z_s] = 0$)

$$E \left[\int I_s r_n(W_s; X_s, Z_s, \tau) r_n(W_t; X_s, Z_s) f^{-2}(X_s, Z_s) dG(\tau) \right] = O(h_1^{-(d_1+d_3)}), \text{ and} \quad (\text{B.9})$$

$$E \left[\int I_s r_n(\bar{W}_s; \bar{X}_s, \bar{Z}_s, \tau) r_n(\bar{W}_t; \bar{X}_s, \bar{Z}_s) f^{-2}(\bar{X}_s, \bar{Z}_s) dG(\tau) \right] = 0. \quad (\text{B.10})$$

To bound $D_n \equiv E(\Delta_{n,2})$, let $m = \lceil B \log n \rceil$ (the integer part of $B \log n$), where B is a large positive constant so that $n^5 \beta_m^{\delta/(1+\delta)} = o(1)$ for some $\delta > 0$ by Assumption A.1(i).³¹ We consider two different cases for D_n : (a) $|s - t| > m$ and (b) $|s - t| \leq m$. We use $D_{n,a}$ and $D_{n,b}$ to denote these two cases. For case (a), we use Lemma A.2 to obtain

$$\begin{aligned} D_{n,a} &= 2n^{-2} h_1^{(d_1+d_3)/2} \sum_{|s-t| > m} E \left[\int I_s r_n(W_s; X_s, Z_s, \tau) r_n(W_t; X_s, Z_s) f^{-2}(X_s, Z_s) dG(\tau) \right] \\ &\leq C n^{-2} h_1^{(d_1+d_3)/2} \left\{ n^2 \left(h_1^{-(d_1+d_3)} \right)^{(1+2\delta)/(1+\delta)} \beta_m^{\delta/(1+\delta)} \right\} \\ &= o \left(h_1^{-2(d_1+d_3)} \beta_m^{\delta/(1+\delta)} \right) = o \left(n \beta_m^{\delta/(1+\delta)} \right) = o(1). \end{aligned}$$

For case (b),

$$\begin{aligned} D_{n,b} &= 2h_1^{(d_1+d_3)/2} n^{-2} \sum_{|s-t| \leq m} E \left[\int I_s r_n(W_s; X_s, Z_s, \tau) r_n(W_t; X_s, Z_s) f^{-2}(X_s, Z_s) dG(\tau) \right] \\ &\leq C h_1^{(d_1+d_3)/2} n^{-2} n m h_1^{-(d_1+d_3)} = O \left(n^{-1} h_1^{-2(d_1+d_3)} \right) O \left(m h_1^{3(d_1+d_3)/2} \right) = o(1). \end{aligned}$$

In consequence, $E(\Delta_{n,2}) = o(1)$. Next, we want to show

$$\begin{aligned} E_n &\equiv E(\Delta_{n,2})^2 \\ &= h_1^{(d_1+d_3)} n^{-4} \sum_{t_1 \neq t_2} \sum_{t_3 \neq t_4} E \left\{ \int \int I_{t_1} r_n(W_{t_1}; X_{t_1}, Z_{t_1}, \tau) r_n(W_{t_2}; X_{t_1}, Z_{t_1}, \tau) f^{-2}(X_{t_1}, Z_{t_1}) \right. \\ &\quad \times I_{t_3} r_n(W_{t_3}; X_{t_3}, Z_{t_3}, \tau') r_n(W_{t_4}; X_{t_3}, Z_{t_3}, \tau') f^{-2}(X_{t_3}, Z_{t_3}) dG(\tau) dG(\tau') \left. \right\} \\ &= o(1). \end{aligned}$$

We consider two cases: (a) for each $i \in \{1, 2, 3, 4\}$, $|t_i - t_j| > m$ for all $j \neq i$; and (b) all the other remaining cases. We will use $E_{n,s}$ to denote these cases ($s = a, b$). Using Lemma A.2 three times (e.g., the first time is to separate (t_1, t_2) -indexed random variables from (t_3, t_4) -indexed random variables),

³¹For example, for fixed $\delta > 0$, if $\rho < 1/2.71828$ in Assumption A.1(i), $B = 5(1 + \delta)/\delta$ would suffice.

(B.10), Assumptions A1(ii)-(iii), and A3, we have:

$$\begin{aligned}
E_{n,a} &\leq h_1^{(d_1+d_3)} n^{-4} \sum_{t_1 \neq t_2} \sum_{t_3 \neq t_4} E \left\{ \int I_{t_1} r_n(\bar{W}_{t_1}; \bar{X}_{t_1}, \bar{Z}_{t_1}, \tau) r_n(\bar{W}_{t_2}; \bar{X}_{t_1}, \bar{Z}_{t_1}, \tau) f^{-2}(\bar{X}_{t_1}, \bar{Z}_{t_1}) dG(\tau) \right\} \\
&\quad \times E \left\{ \int I_{t_3} r_n(\bar{W}_{t_3}; \bar{X}_{t_3}, \bar{Z}_{t_3}, \tau') r_n(\bar{W}_{t_3}; \bar{X}_{t_3}, \bar{Z}_{t_3}, \tau') f^{-2}(\bar{X}_{t_3}, \bar{Z}_{t_3}) dG(\tau') \right\} \\
&\quad + Ch_1^{(d_1+d_3)} \left(h_1^{-(d_1+d_3)} \right)^{(2+4\delta)/(1+\delta)} \beta_m^{\delta/(1+\delta)} + Ch_1^{(d_1+d_3)} \left\{ \left(h_1^{-(d_1+d_3)} \right)^{(1+2\delta)/(1+\delta)} \beta_m^{\delta/(1+\delta)} \right\}^2 \\
&= o \left(h_1^{-3(d_1+d_3)} \beta_m^{\delta/(1+\delta)} \right) = o \left(n^2 \beta_m^{\delta/(1+\delta)} \right) = o(1).
\end{aligned}$$

For all the other remaining cases, there exists at least one $i \in \{1, 2, 3, 4\}$, such that $|t_i - t_j| \leq m$ for some $j \neq i$. The number of such terms is of the order $O(n^3 m)$. For $t_1 \neq t_2$ and $t_3 \neq t_4$, one can bound $E \left| \int I_{t_1} r_n(W_{t_1}; X_{t_1}, Z_{t_1}, \tau) r_n(W_{t_2}; X_{t_1}, Z_{t_1}, \tau) f^{-2}(X_{t_1}, Z_{t_1}) I_{t_3} r_n(W_{t_3}; X_{t_3}, Z_{t_3}, \tau') r_n(W_{t_4}; X_{t_3}, Z_{t_3}, \tau') \times f^{-2}(X_{t_3}, Z_{t_3}) dG(\tau) dG(\tau') \right|$ by $Ch_1^{-2(d_1+d_3)}$ if $\{t_1, t_2\} \cap \{t_3, t_4\} \neq \{t_1, t_2\}$ and by $Ch_1^{-3(d_1+d_3)}$ otherwise. Consequently, $E_{n,b} \leq Ch_1^{(d_1+d_3)} n^{-4} \left(n^3 m h_1^{-2(d_1+d_3)} + n^2 h_1^{-3(d_1+d_3)} \right) = O \left(n^{-1} m h_1^{-(d_1+d_3)} + n^{-2} h_1^{-2(d_1+d_3)} \right) = o(1)$. In sum, $E(\Delta_{n,2})^2 = o(1)$, and by the Chebyshev inequality, we have $\Delta_{n,2} = o_p(1)$.

Now, we want to show $\tilde{\Delta}_{n,1} = o_p(1)$. Write

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

and decompose $\Delta_{n,1}$ as

$$\begin{aligned}
&h_1^{(d_1+d_3)/2} n^{-2} \sum_{j \neq t, j \neq s, t \neq s} \left\{ \int I_j r_n(W_s; X_j, Z_j, \tau) r_n(W_t; X_j, Z_j, \tau) f^{-2}(X_j, Z_j) dG(\tau) \right. \\
&\quad \left. - E \left[\int I_j r_n(W_s; X_j, Z_j, \tau) r_n(W_t; X_j, Z_j, \tau) f^{-2}(X_j, Z_j) dG(\tau) | W_s, W_t \right] \right\} \\
&+ h_1^{(d_1+d_3)/2} n^{-2} \sum_{j \neq t, j \neq s, t \neq s} \left\{ E \left[\int I_j r_n(W_s; X_j, Z_j, \tau) r_n(W_t; X_j, Z_j, \tau) f^{-2}(X_j, Z_j) dG(\tau) | W_s, W_t \right] \right. \\
&\quad \left. - \int \int r_n(W_s; x, z, \tau) r_n(W_t; x, z, \tau) f^{-2}(x, z) dF(x, z) dG(\tau) \right\} \\
&\equiv \Delta_{n,1}^{(1)} + \Delta_{n,1}^{(2)}.
\end{aligned}$$

It suffices to show each term in the last expression is $o_p(1)$. By the triangle inequality, Lemma A.3, Assumptions A1(ii)-(iii) and A3,

$$\begin{aligned}
& E \left| \Delta_{n,1}^{(2)} \right| \\
& \leq h_1^{(d_1+d_3)/2} n^{-2} \sum_{j \neq t, j \neq s, t \neq s} E |E \left[\int I_j r_n(W_s; X_j, Z_j, \tau) r_n(W_t; X_j, Z_j, \tau) f^{-2}(X_j, Z_j) dG(\tau) | W_s, W_t \right] \\
& \quad - \int_S \int r_n(W_s; x, z, \tau) r_n(W_t; x, z, \tau) dG(\tau) f^{-2}(x, z) dF(x, z) | \\
& \leq C h_1^{(d_1+d_3)/2} n^{-2} \left(n^3 \left(h_1^{-(d_1+d_3)} \right)^{2\delta/(1+\delta)} \beta_m^{\delta/(1+\delta)} + n^2 m \right) \\
& = o \left(n h_1^{-3(d_1+d_3)/2} \beta_m^{\delta/(1+\delta)} + m h_1^{(d_1+d_3)/2} \right) = o(1),
\end{aligned}$$

implying $\Delta_{n,1}^{(2)} = o_p(1)$ by the Markov inequality. Let $\sum_{j \neq s, t}$ denote $\sum_{j \neq s, j \neq t, s \neq t}$. Further, denote

$$\begin{aligned}
S_{s,t,j} & \equiv \int I_j r_n(W_s; X_j, Z_j, \tau) r_n(W_t; X_j, Z_j, \tau) f^{-2}(X_j, Z_j) dG(\tau) \\
& \quad - E \left[\int I_j r_n(W_s; X_j, Z_j, \tau) r_n(W_t; X_j, Z_j, \tau) f^{-2}(X_j, Z_j) dG(\tau) | W_s, W_t \right].
\end{aligned}$$

Then $\Delta_{n,1}^{(1)} = h_1^{(d_1+d_3)/2} n^{-2} \sum_{j \neq s, t} S_{s,t,j}$ with $E(\Delta_{n,1}^{(1)}) = 0$ because $E(S_{s,t,j}) = 0$ by the law of iterated expectation. We shall show

$$F_n \equiv E \left(\Delta_{n,1}^{(1)} \right)^2 = h_1^{(d_1+d_3)} n^{-4} \sum_{t_1 \neq t_2, t_3} \sum_{t_4 \neq t_5, t_6} E \{ S_{t_1, t_2, t_3} S_{t_4, t_5, t_6} \} = o(1).$$

We consider four different cases: (a) for each $i \in \{1, 2, 3, 4, 5, 6\}$, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly four different i 's, $|t_i - t_j| > m$ for all $j \neq i$; (c) for exactly three different i 's, $|t_i - t_j| > m$ for all $j \neq i$; (d) all the other remaining cases. We will use $F_{n,s}$ to denote these cases ($s = a, b, c, d$). For each case, one can use Lemma A.2 to show $F_{n,s} = o(1)$. For example, for case (a), noticing that $E(S_{t_1, t_2, t_3}) = E(S_{t_4, t_5, t_6}) = 0$, we have

$$\begin{aligned}
|F_{n,a}| & \leq C h_1^{(d_1+d_3)} n^{-4} n^6 \left(h_1^{-(d_1+d_3)} \right)^{4\delta/(1+\delta)} \beta_m^{\delta/(1+\delta)} \\
& = o \left(n^2 h_1^{-3(d_1+d_3)} \beta_m^{\delta/(1+\delta)} \right) = o \left(n^4 \beta_m^{\delta/(1+\delta)} \right) = o(1).
\end{aligned}$$

For case (d), the number of terms in the summation is of order $O(n^3 m^3)$, and each term can be bounded by $C h_1^{-i(d_1+d_3)}$ for some finite positive constant C if there are at least $(6 - i)$ distinct elements in $\{t_1, t_2, t_3, t_4, t_5, t_6\}$, where $i = 2, 3$ and 4. So

$$\begin{aligned}
F_{n,d} & = h_1^{(d_1+d_3)} n^{-4} O \left(n^3 m^3 h_1^{-2(d_1+d_3)} + n^3 h_1^{-3(d_1+d_3)} + n^2 h_1^{-4(d_1+d_3)} \right) \\
& = O \left(n^{-1} m^3 h_1^{-(d_1+d_3)} + n^{-1} h_1^{-2(d_1+d_3)} + n^{-2} h_1^{-3(d_1+d_3)} \right) \\
& = o(1) \text{ by Assumption A2(iii)}.
\end{aligned}$$

In sum, $F_n = o(1)$ and thus $\Delta_{n,1}^{(1)} = o_p(1)$ by the Chebyshev inequality. The conclusion thus follows. ■

Lemma B.4 *Let Assumptions A1 – A3 hold. Then $h_1^{(d_1+d_3)/2} \tilde{R}_{n,1} \xrightarrow{d} N(0, \sigma^2)$, where $\tilde{R}_{n,1}$ is defined by (B.7).*

Proof.

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

where

$$H_n(W_s, W_t) = h_1^{(d_1+d_3)/2} \int \int_S r(W_s; x, z, \tau) r(W_t; x, z, \tau) f^{-2}(x, z) dF(x, z) dG(\tau).$$

We now verify the conditions in Theorem A.4 hold for $S_{n,1}$ with $h_n(u, v)$ in the theorem replaced by $H_n(u, v)$. First, by construction, $H_n(w, v) = H_n(v, w)$, and $EH_n(W_0, v) = 0$.

$$\begin{aligned}
& E |H_n(W_t, W_0)|^p \\
&= h_1^{p(d_1+d_3)/2} \int \int | \int \int_S K_{h_1}(x - x_t, z - z_t) [1(y_t \leq \tau) - F(\tau|x_t, z_t)] K_{h_1}(x - X_0, z - Z_0) \\
&\quad \times [1(Y_0 \leq \tau) - F(\tau|X_0, Z_0)] f^{-2}(x, z) dF(x, z) dG(\tau) |^p f_{0t}(w_0, w_t) dw_0 dw_t \\
&\leq C h_1^{p(d_1+d_3)/2} h_1^{-(d_1+d_3)(p-1)} \int_{R^{d_1+d_3}} \int_{R^{d_1+d_3}} |K(u_1)K(u_1 + u_2)|^p du_1 du_2 \\
&= O\left(h_1^{(d_1+d_3)(1-p/2)}\right) \text{ by Assumptions A1(ii)-(iii), A2(i) and A3,}
\end{aligned}$$

so we have $\|H_n(W_t, W_0)\|_p \leq C(h_1^{(d_1+d_3)}(1/p-1/2))$. Let \bar{W}_0 be an independent copy of W_0 ; one can show by similar argument that $\|H_n(W_0, \bar{W}_0)\|_p \leq C(h_1^{d_1+d_3})^{(1/p-1/2)}$. Consequently, one obtains $u_n(p) \leq C(h_1^{d_1+d_3})^{(1/p-1/2)}$ for some $C > 0$.

Now we show $v_n(p) \leq C(h_1^{d_1+d_3})^{1/p}$. By Assumptions A1(ii)-(iii), A2(i) and A3, we have

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

so $\|G_{n0}(W_t, W_0)\|_p \leq C(h_1^{d_1+d_3})^{1/p}$. Similarly, one can show $\|G_{n0}(W_0, \bar{W}_0)\|_p \leq C(h_1^{d_1+d_3})^{1/p}$, and thus $v_n(p) \leq C(h_1^{d_1+d_3})^{1/p}$. By the same argument, we have: $w_n(p) \equiv \|G_{n0}(W_0, W_0)\|_p \leq C$ and $z_n(p) \leq$

$C(h_1^{d_1+d_3})$. For some fixed $\delta_0 > 0$, take $\gamma_0 = (2+\delta_0)(d_1+d_3)/[(8+2\delta_0)\delta_1] \leq (2+\delta_0)/(8+2\delta_0) \in (0, 1/2)$ and $\gamma_1 \in (0, \gamma_0]$. Then it is easy to see that Conditions (i)- (iv) of Theorem A.4 are satisfied. Finally,

$$\begin{aligned} & E[H_n(W_0, \overline{W}_0)^2] \\ &= h_1^{d_1+d_3} E\left\{ \int \int \int_S \int_S r_n(W_0; x, z, \tau) r_n(\overline{W}_0; x, z, \tau) r_n(W_0; x', z', \tau') r_n(\overline{W}_0; x', z', \tau') \right. \\ &\quad \left. \times dF(x, z) dF(x', z') dG(\tau) dG(\tau') \right\} \\ &= C_3^{(d_1+d_3)} \int \int \int_S V^2(x, z; \tau, \tau') d(x, z) dG(\tau) dG(\tau') + o(1), \end{aligned}$$

where C_3 and $V(x, z; \tau, \tau')$ are defined in the main text (Section 3.1). It follows that $S_{n,1} \xrightarrow{d} N(0, \sigma^2)$, where $\sigma^2 = 2C_3^{(d_1+d_3)} \int \int \int_S V^2(x, z, \tau, \tau') d(x, z) dG(\tau) dG(\tau')$. And by Lemma B.5, $S_{n,2} = o(1)$. Consequently, $h_1^{(d_1+d_3)/2} \tilde{R}_{n,1} \xrightarrow{d} N(0, \sigma^2)$. ■

Lemma B.5 *Let Assumptions A1 – A3 hold. Then under H_0 , $S_{n,2} \equiv 2n^{-1} \sum_{1 \leq s < t \leq n} E[H_n(W_s, W_t)] = o(1)$.*

Proof. Let m be defined as in the proof of Lemma B.3. We consider two different cases for $S_{n,2}$: (a) $t - s > m$ and (b) $0 < t - s \leq m$. We use $S_{n,2a}$ and $S_{n,2b}$ to denote these two cases. For case (a), we use Lemma A.2 and the bound $u_n(p) \leq C(h_1^{(d_1+d_3)})^{1/p-1/2}$ with $p = 1 + \delta$ (see Theorem A.4 for the definition of $u_n(p)$) to obtain $S_{n,2a} = 2n^{-1} \sum_{s-t > m} EH_n(W_t, W_s) \leq Cn^{-1}n^2 \left(h_1^{(d_1+d_3)}\right)^{1/(1+\delta)-1/2} \beta_m^{\delta/(1+\delta)} = o\left(nh_1^{-(d_1+d_3)/2} \beta_m^{\delta/(1+\delta)}\right) = o(1)$. For case (b), using the bound $u_n(1) \leq Ch_1^{(d_1+d_3)/2}$, we have $S_{n,2b} = n^{-1} \sum_{s-t \leq m} EH_n(W_t, W_s) \leq Cn^{-1}nmh_1^{(d_1+d_3)/2} = O\left(mh_1^{(d_1+d_3)/2}\right) = o(1)$. The proof is complete. ■

Lemma B.6 *Let Assumptions A1 – A3 and the null hypothesis hold. Then $h_1^{(d_1+d_3)/2} \tilde{R}_{n,2} = o_p(1)$, where $\tilde{R}_{n,2}$ is defined in (B.8).*

Proof. Let $\varepsilon_t(\tau) = 1(Y_t \leq \tau) - F(\tau|X_t)$ and $M_{ts}(\tau) = F(\tau|X_t) - F(\tau|X_s)$. We can write

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

Noting that $G_{n1} + G_{n2} + G_{n3}$ is nonnegative, it suffices to show that $E[G_{ni}] = o(1)$, $i = 1, 2$, and 3. To show $EG_{n,1} \equiv E[G_{n1}] = o(1)$, we let $S_{t_0, t_1, t_2, t_3, t_4}^{(1)} \equiv E[\int I_{t_0} f_{1t_0}^{-2} f_{2t_1}^{-1} f_{2t_3}^{-1} K_{t_0t_1} K_{t_0t_3} L_{t_1t_2} L_{t_3t_4} \varepsilon_{t_2}(\tau) \varepsilon_{t_4}(\tau)$

$dG(\tau)$] and consider three different cases for $EG_{n,1}$: (a) for each $i \in \{0, 1, 2, 3, 4\}$, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly three different i 's, $|t_i - t_j| > m$ 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 the null, we can apply Lemma A.2 immediately to get $EG_{n,1a} \leq Ch_1^{(d_1+d_3)/2} n^{-4} n^5 \left(h_1^{-(d_1+d_3)}\right)^{4\delta/(1+\delta)} \beta_m^{\delta/(1+\delta)} = o\left(nh_1^{-4(d_1+d_3)} \beta_m^{\delta/(1+\delta)}\right) = o\left(n^3 \beta_m^{\delta/(1+\delta)}\right) = o(1)$. For case (b), if either t_2 or t_4 is among the three elements that lie at least at m -distance from all the other elements, one can bound the term $S_{t_0, t_1, t_2, t_3, t_4}^{(1)}$ as in case (a). Otherwise, bound the term by C . Consequently, $EG_{n,1b} = O\left(n^3 \beta_m^{\delta/(1+\delta)} + h_1^{(d_1+d_3)/2} n^{-4} n^4 m\right) = o(1)$. For case (c), the total number of terms in the summation is of order $O(n^3 m^2)$ and one can readily obtain

$$\begin{aligned} & EG_{n,1c} \\ &= h_1^{(d_1+d_3)/2} n^{-4} O\left(n^3 m^2 + n^3 m h_1^{-(d_1+d_3)} + n^3 h_1^{-2(d_1+d_3)} + n^2 h_1^{-2(d_1+d_3)} h_2^{-d_1} + n h_1^{-2(d_1+d_3)} h_2^{-2d_1}\right) \\ &= o(1). \end{aligned}$$

Next, let $S_{t_0, t_1, t_2, t_3, t_4}^{(2)} \equiv E\left[\int I_{t_0} f_{1t_0}^{-2} f_{2t_1}^{-1} f_{2t_3}^{-1} K_{t_0 t_1} K_{t_0 t_3} L_{t_1 t_2} L_{t_3 t_4} M_{t_2 t_1}(\tau) M_{t_4 t_3}(\tau) dG(\tau)\right]$. 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,

$$E[G_{n2}] = h_1^{(d_1+d_3)/2} n^{-4} O\left(n^5 h_2^{2r} + n^3 h_2^{2-d_1} + n^4 h_1^{-(d_1+d_3)} h_2^{2r}\right) = o(1).$$

Similarly, one can show that $E[G_{n3}] = o(1)$, and the proof is complete. ■

Lemma B.7 *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.6).*

Proof. Using the notation introduced before, we can write

$$\begin{aligned} & - h_1^{(d_1+d_3)/2} R_{n,3} \\ &= h_1^{(d_1+d_3)/2} n^{-3} \sum_{t_1, t_2, t_3, t_4; t_2 \neq t_3} \int I_{t_1} \widehat{f}_{1t_1}^{-1} \widehat{f}_{2t_3}^{-1} K_{t_1 t_2} K_{t_1 t_3} L_{t_3 t_4} \varepsilon_{t_2}(\tau) [\varepsilon_{t_4}(\tau) + M_{t_4 t_3}(\tau)] dG(\tau) \\ &= \{h_1^{(d_1+d_3)/2} n^{-3} \sum_{t_1, t_2, t_3, t_4; t_2 \neq t_3} \int I_{t_1} f_{1t_1}^{-2} f_{2t_3}^{-1} K_{t_1 t_2} K_{t_1 t_3} L_{t_3 t_4} \varepsilon_{t_2}(\tau) \varepsilon_{t_4}(\tau) dG(\tau) \\ & \quad + h_1^{(d_1+d_3)/2} n^{-3} \sum_{t_1, t_2, t_3, t_4; t_2 \neq t_3} \int I_{t_1} f_{1t_1}^{-2} f_{2t_3}^{-1} K_{t_1 t_2} K_{t_1 t_3} L_{t_3 t_4} \varepsilon_{t_2}(\tau) M_{t_4 t_3}(\tau) dG(\tau)\} \{1 + o_p(1)\} \\ &\equiv \{I_{n1} + I_{n2}\} \{1 + o_p(1)\}. \end{aligned}$$

Let $T_{t_1, t_2, t_3, t_4}^{(1)}(\tau) \equiv I_{t_1} f_{1t_1}^{-2} f_{2t_3}^{-1} K_{t_1 t_2} K_{t_1 t_3} L_{t_3 t_4} \varepsilon_{t_2}(\tau) \varepsilon_{t_4}(\tau)$ and consider three different cases for $EI_{n,1} \equiv E[I_{n1}]$: (a) for each $i \in \{1, 2, 3, 4\}$, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly two different i 's, $|t_i - t_j| > m$ for all $j \neq i$; (c) all the other remaining cases. We use $EI_{n,1s}$ to denote these cases ($s = a, b, c$). For case (a), we apply Lemma A.2 immediately to get $EI_{n,1a} \leq Ch_1^{(d_1+d_3)/2} n^{-3} n^4 \left(h_1^{-(d_1+d_3)}\right)^{3\delta/(1+\delta)} \beta_m^{\delta/(1+\delta)} = o\left(nh_1^{-3(d_1+d_3)} \beta_m^{\delta/(1+\delta)}\right) = o\left(n^3 \beta_m^{\delta/(1+\delta)}\right) = o(1)$. For case (b), if either t_2 or t_4 is among the two

elements that lie at least at m -distance from all the other elements, one can bound the term $T_{t_1, t_2, t_3, t_4}^{(1)}(\tau)$ as in case (a). Otherwise, bound the term by C . Consequently, $EG_{1b} = O\left(n^3 \beta_m^{\delta/(1+\delta)} + h_1^{(d_1+d_3)/2} n^{-3} n^3 m\right) = o(1)$. For case (c), the total number of terms in the summation is of order $n^2 m^2$ and one can readily obtain $EI_{n,1c} = h_1^{(d_1+d_3)/2} n^{-3} O\left(n^2 m^2 + n^2 m h_1^{-(d_1+d_3)} + n^2 h_1^{-(d_1+d_3)} h_2^{-d_1}\right) = o(1)$. So $EI_{n,1} = o(1)$.

Next, we want to show $EI_{n,2} \equiv E[I_{n1}]^2 = o(1)$. By the Jensen inequality,

$$EI_{n,2} \leq h_1^{(d_1+d_3)} n^{-6} \sum_{t_1, t_2, t_3, t_4; t_2 \neq t_3} \sum_{t_5, t_6, t_7, t_8; t_6 \neq t_7} \int T_{t_1, t_2, t_3, t_4}^{(1)}(\tau) T_{t_5, t_6, t_7, t_8}^{(1)}(\tau) dG(\tau).$$

We consider three different cases for $EI_{n,2}$: (a) for at least five i 's in $\{1, 2, 3, 4, 6, 7, 8\}$, $|t_i - t_j| > m$ for all $j \neq i$; (b) for exactly four different i 's, $|t_i - t_j| > m$ 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.2 immediately to get $EI_{n,2a} \leq C h_1^{(d_1+d_3)} n^{-6} n^8 \left(h_1^{-(d_1+d_3)}\right)^{6\delta/(1+\delta)} \beta_m^{\delta/(1+\delta)} = o\left(n^2 h_1^{-5(d_1+d_3)} \beta_m^{\delta/(1+\delta)}\right) = o\left(n^5 \beta_m^{\delta/(1+\delta)}\right) = o(1)$. For case (b), the number of terms in the summation is of order $O(n^5 m^3)$. If either t_2, t_4, t_6 , or t_8 , is among the four elements that lie at least at m -distance from all the other elements, one can bound the term $T_{t_1, t_2, t_3, t_4}^{(1)}(\tau)$ as in case (a). Otherwise, bound the term by $C h_1^{-(d_1+d_3)}$. Consequently, $EI_{n,2b} = O\left(n^5 \beta_m^{\delta/(1+\delta)} + h_1^{(d_1+d_3)} n^{-6} n^5 m^3 h_1^{-(d_1+d_3)}\right) = o(1)$. For case (c), the total number of terms in the summation is of order $O(n^4 m^4)$ and one can readily obtain $EI_{n,2c} = h_1^{(d_1+d_3)} n^{-6} O\left(n^4 m^4 + n^4 m^3 h_1^{-(d_1+d_3)}\right) = o(1)$. Thus $I_{n1} = o_p(1)$ by the Chebyshev inequality.

Similarly, one can show that $E[I_{n2}] = O\left(n^3 \beta_m^{\delta/(1+\delta)} + h_1^{(d_1+d_3)/2} n^{-3} n^3 m h_2^r\right) = o(1)$ and $E[I_{n2}]^2 = O\left(n^5 \beta_m^{\delta/(1+\delta)} + h_1^{(d_1+d_3)} n^{-6} n^7 m h_2^{2r}\right) = o(1)$, implying $I_{n2} = o_p(1)$ by the Chebyshev inequality. The proof is complete. ■

Putting Lemmas B.1-B.7 together, we have proved Theorem 3.1 in the main text.

C Proof of Lemmas, Propositions and Corollaries

Proof of Corollary 3.2. It suffices to show that $h_1^{(d_1+d_3)} \left(\widehat{B}_n - \widehat{B}_{n,1}\right) = o_p(1)$ when $d_1 + d_3 \leq 3$.

Under the null,

$$\begin{aligned} \widehat{B}_n &= \sum_{t=1}^n I_t \sum_{s=1}^n \int [w_{t,s} \widehat{g}_s(\tau)]^2 dG(\tau) \\ &= n^{-2} h_1^{-2(d_1+d_3)} \sum_{t=1}^n I_t \widehat{f}_{1t}^{-2} \sum_{s=1}^n K_{ts}^2 \int [1(Y_s \leq \tau) - F(\tau|X_s, Z_s)]^2 dG(\tau) \\ &\quad + 2n^{-2} h_1^{-2(d_1+d_3)} \sum_{t=1}^n I_t \widehat{f}_{1t}^{-2} \sum_{s=1}^n K_{ts}^2 \int [1(Y_s \leq \tau) - F(\tau|X_s, Z_s)] \left[F(\tau|X_s) - \widehat{F}_{h_2}(\tau|X_s)\right] dG(\tau) \\ &\quad + n^{-2} h_1^{-2(d_1+d_3)} \sum_{t=1}^n I_t \widehat{f}_{1t}^{-2} \sum_{s=1}^n K_{ts}^2 \int \left[F(\tau|X_s) - \widehat{F}_{h_2}(\tau|X_s)\right]^2 dG(\tau) \\ &\equiv B_{n,1} + B_{n,2} + B_{n,3}. \end{aligned}$$

Suppose Assumptions A1-A2 hold and $d_1 + d_3 \leq 3$; then by Lemma A.5, one can show that $h_1^{(d_1+d_3)/2} B_{n,1} = h_1^{-(d_1+d_3)/2} C_1^{(d_1+d_3)} \int_{\mathcal{S}} V(x, z, \tau) d(x, z) dG(\tau) + O_p\left(h_1^{2-(d_1+d_3)/2} + n^{-1/2} h_1^{-(d_1+d_3)} \sqrt{\ln n}\right) = h_1^{(d_1+d_3)/2} \widehat{B}_{n,1} + o_p(1)$; and $B_{n,3} = O_p\left(n h_1^{-(d_1+d_3)} h_2^{-d_1}\right) = o_p(1)$. The result then follows by the Cauchy-Schwarz inequality. ■

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 $h_1^{(d_1+d_3)/2} ISELTR_n$ that were not present under the null, among

which only one term is asymptotically non-negligible under $H_1(\alpha_n)$:

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

where the second line follows from dominated convergence arguments. Consequently,

$$\Pr(T_n \geq z_\alpha | H_1(\alpha_n)) \rightarrow 1 - \Phi(z_\alpha - \delta/\sigma). \quad \blacksquare$$

Proof of Corollary 3.4. This follows from Corollary 3.2 and Proposition 3.3. \blacksquare

Proof of Theorem 4.1. Note that the characteristic function H is uniformly bounded on its support, 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:

$$\begin{aligned}
& \sup_{\tau \in \mathbb{R}} \sup_{(x,z) \in S} |\widehat{m}_{h_1}(x, z; \tau) - m(x, z; \tau)| = O_p \left(n^{-1/2} h_1^{-(d_1+d_3)/2} \sqrt{\ln n} + h_1^2 \right), \text{ and} \\
& \sup_{\tau \in \mathbb{R}} \sup_{x \in S_1} |\widehat{m}_{h_2}(x; \tau) - m(x; \tau)| = O_p \left(n^{-1/2} h_2^{-d_1/2} \sqrt{\ln n} + h_2^r \right).
\end{aligned}$$

The above uniform consistency results can be established in the exact same fashion as done in Lemma A.5. \blacksquare

Proof of Corollary 4.2. The argument is identical to the proof of Corollary 3.2. \blacksquare

Proof of Proposition 5.1. To find a^* , fix $m \in \mathbb{R}$ arbitrarily and consider solving the following variational problem over all piecewise smooth, bounded, square integrable functions from $S \rightarrow \mathbb{R}_+$:

$$\min_a F_a(m) \text{ s.t. } \int_S a^2(x, z) d(x, z) = 1. \quad (\text{C.1})$$

For any $(x_0, z_0) \in S$, let $F_a(m|\psi(x_0, z_0))$ be the conditional cdf of $M(a, \widetilde{\Delta})$ given $\psi(x_0, z_0)$. Let $f_a(m|\psi(x_0, z_0))$ denote the conditional *p.d.f.* corresponding to $F_a(m|\psi(x_0, z_0))$. It is clear that $F_a(m) = E_{\psi(x_0, z_0)}[F_a(m|\psi(x_0, z_0))]$, where the symbol $E_{\psi(x_0, z_0)}$ indicates that the expectation is over $\psi(x_0, z_0)$. Furthermore,

$$\frac{\partial F_a(m|\psi(x_0, z_0))}{\partial a(x_0, z_0)} = \frac{\partial E \left\{ \int_S a(x, z) \psi(x, z)^2 d(x, z) | \psi(x_0, z_0) \right\}}{\partial a(x_0, z_0)} = \psi(x_0, z_0)^2 f_a(m|\psi(x_0, z_0)).$$

This implies that

$$\frac{\partial F_a(m)}{\partial a(x_0, z_0)} = E_{\psi(x_0, z_0)} [\psi(x_0, z_0)^2 f_a(m|\psi(x_0, z_0))] \text{ for all } (x_0, z_0) \in S.$$

Thus the Euler-Lagrange equation for the variational problem (C.1) is

$$E_{\psi(x_0, z_0)} [\psi(x_0, z_0)^2 f_a(m|\psi(x_0, z_0))] = 2\lambda a^*(x_0, z_0) \text{ for all } (x_0, z_0) \in S, \quad (\text{C.2})$$

where λ is the Lagrange multiplier for the constraint in (C.1) and a^* is the solution. To solve the problem, we use a guess and verify approach. So suppose that $a^*(x, z) = 1\{(x, z) \in S\} / |S|^{-1/2}$, which

clearly satisfies the constraint in (C.1). As noted in the text regarding the nature of the random process ψ , the joint distribution of $M(a^*, \tilde{\Delta}) = \int_S \psi^2(x, z) d(x, z) / \left\{ \|S\| \sqrt{2C_3^{d_1+d_3}} \right\}$ and $\psi(x_0, z_0)$ does not depend on $(x_0, z_0) \in S$. Therefore $C \equiv E_{\psi(x_0, z_0)} [\psi(x_0, z_0)^2 f_a(m|\psi(x_0, z_0))]$ does not depend on $(x_0, z_0) \in S$ and (C.2) is satisfied with $a^*(x, z) = 1\{(x, z) \in S\} \|S\|^{-1/2}$ and $\lambda = C \|S\|^{1/2} / 2$. That is, $a^*(x, z) = 1\{(x, z) \in S\} \|S\|^{-1/2}$ solves the variational problem. ■

References

- [1] Aggarwal, R. (1981) Exchange Rates and Stock Prices: A Study of the U.S. Capital Markets under Floating Exchange Rates, *Akron Business and Economic Review* 12, 7-12.
- [2] Agresti, A. (1990), *Categorical Data Analysis*, John Wiley, New York.
- [3] Ahn, H. (1997), Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors, *Econometric Theory* 13, 3-31.
- [4] Ajayi, R. A. and M. Mougoue (1996), On the Dynamic Relation between Stock Prices and Exchange Rates, *Journal of Financial Research*, 19, 193-207.
- [5] Andrews, D. W. K. and W. Ploberger (1994), Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative, *Econometrica* 62, 1383-1414.
- [6] Baek, E., and W. Brock (1992), A General Test for Nonlinear Granger Causality: Bivariate model, *Discussion Paper*, Iowa State University and University of Wisconsin, Madison.
- [7] Bierens H. J. (1982), Consistent Model Specification Test, *Journal of Econometrics* 20, 105-134.
- [8] Bierens, H. J. and W. Ploberger (1997) Asymptotic Theory of Integrated Conditional Moment Tests, *Econometrica* 65, 1129-1152.
- [9] Black, D. C., P. R. Corrigan, and M. R. Dowd (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.
- [10] Blum, J. R., J. Kiefer, and M. Rosenblatt (1961), Distribution Free Tests of Independence Based on the Sample Distribution Function, *Annals of Mathematical Statistics* 32, 485-498.
- [11] Boning, W. B. and F. Sowell (1999), Optimality for the Integrated Conditional Moment Test, *Econometric Theory* 15, 710-718.
- [12] Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes*, Springer-Verlag, New York.
- [13] Brock, W., Dechert, W. and Scheinkman, J. (1987), A Test for Independence Based on the Correlation Dimension, *Working Paper*, University of Wisconsin, Madison.
- [14] Collomb, G. and W. Härdle (1986), Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction: Kernel Regression Estimation from Dependent Observations, *Stochastic Processes and Their Applications* 23, 77-89.

- [15] Copeland, T. (1976), A Model of Asset Trading under the Assumption of Sequential Information Arrival, *Journal of Finance* 45, 379-395.
- [16] Delgado, M. (1996), Testing Serial Independence Using the Sample Distribution Function, *Journal of Time Series Analysis* 17, 271-287.
- [17] DeLong, J., A. Shleifer, L. Summers, and B. Waldmann (1990), Positive Feedback Investment Strategies and Destabilizing Speculation, *Journal of Finance* 31, 135-155.
- [18] Fan Y., and R. Gencay (1993), Hypothesis Testing Based on Modified Nonparametric Estimation of an Affine Measure between Two Distributions, *Journal of Nonparametric Statistics* 2, 389-403.
- [19] Fan Y. and Q. Li (1999a), Central Limit Theorem for Degenerate U-statistics of Absolutely Regular Processes with Applications to Model Specification Testing, *Journal of Nonparametric Statistics* 10, 245-271.
- [20] Fan Y. and Q. Li (1999b), Root-N-consistent Estimation of Partially Linear Time Series Models, *Journal of Nonparametric Statistics* 11, 251-269.
- [21] Fernandes, M. (2000), Nonparametric Entropy-Based Tests of Independence between Stochastic Processes, *Discussion Paper*, European University Instit.
- [22] Fernandes, M. and R. G. Flores (2000), Tests for Conditional Independence, Markovian Dynamics and Noncausality, *Discussion Paper*, European University Instit.
- [23] Feuerverger, A. (1987), On Some ECF Procedures for Testing Independence, in *Time Series and Econometric Modelling*, eds. I. B. MacNeill and G. J. Umphrey, Boston Reidel, pp. 189-206.
- [24] Florens, J. P., and D. Fougere (1996), Noncausality in Continuous Time, *Econometrica* 64, 1195-1212.
- [25] Florens, J. P., and M. Mouchart (1982), A Note on Causality, *Econometrica* 50, 583-591.
- [26] Friedman, B. M. and K. N. Kuttner (1992), Money, Income, Prices, and Interest Rates, *American Economic Review* 82, 472-492.
- [27] Friedman, B. M. and K. N. Kuttner (1993), Another Look at the Evidence on Money-Income Causality, *Journal of Econometrics* 57, 189-203.
- [28] Gallant, R., P. Rossi, and G. Tauchen (1992), Stock Prices and Volume, *Review of Financial Studies* 5, 199-242.
- [29] Gihman, I. I. and A. V. Skorohod (1974), *The Theory of Stochastic Processes*. Springer-Verlag, New York.
- [30] Gouriéroux, C., and C. Tenreiro (2001), Local Power Properties of Kernel Based Goodness of Fit Tests, *Journal of Multivariate Analysis* 78, 161-190.
- [31] Granger, C. W. J. (1980), Testing for Causality: A Personal Viewpoint, *Journal of Economic Dynamics and Control* 2, 329-352.
- [32] Gurtler, N. and N. Henze (2000), Goodness-of-fit Tests for the Cauchy Distribution Based on the Empirical Characteristic Function, *Annals of the Institute of Statistical Mathematics* 52, 267-286.

- [33] Härdle, W. (1990), *Applied Nonparametric Regression*. Cambridge University Press. New York.
- [34] Härdle, W., and E. Mammen (1993), Comparing Nonparametric versus Parametric Regression Fits, *Annals of Statistics* 21, 1926-1947.
- [35] Harel, M. and M. L. Puri (1996), Conditional U -statistics for Dependent Random Variables, *Journal of Multivariate Analysis* 57, 84-100.
- [36] Hastie, T. and R. Tibshirani (1986), Generalized Additive Models, *Statistical Science* 1, 297-318.
- [37] Hiemstra, C., and J. D. Jones (1994), Testing for Linear and Nonlinear Granger Causality in the Stock Price-Volume Relation, *Journal of Finance* 49, 1639-1664.
- [38] Hoeffding, W. (1948), A Non-parametric Test of Independence, *Annals of Mathematical Statistics* 58, 546-557.
- [39] Hong, Y. (2001), A Test for Volatility Spillover with Application to Exchange Rates, *Journal of Econometrics* 103, 183-224.
- [40] Hong, Y. and H. White (2000), Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence, *Discussion Paper*, Department of Economics, Cornell University and UCSD.
- [41] Hsieh, D. A. (1989), Testing for Nonlinear Dependence in Daily Foreign Exchange Rates. *Journal of Business* 62, 339-368.
- [42] Hsieh, D. A. (1991), Chaos and Nonlinear Dynamics: Application to Financial Markets. *Journal of Finance* 5, 1839-1877.
- [43] Inoue, A. (1998), A Conditional Goodness-of-fit Test for Time Series, Discussion Paper, Department of Agricultural and Resource Economics, North Carolina State University.
- [44] Inoue, A. (2001), Testing for Distributional Change in Time Series, *Econometric Theory* 17, 156-187.
- [45] Kanas, A. (2002), Is Exchange Rate Volatility Influenced by Stock Return Volatility? Evidence from the US, the UK and Japan, *Applied Economics Letters* 9, 501-503.
- [46] Kang, J. K. and R. M. Stulz (1997), Why is There a Home Bias? An Analysis of Foreign Portfolio Equity Ownership in Japan, *Journal of Financial Economics* 46, 3-28.
- [47] Khashimov, S. A. (1992), Limiting Behavior of Generalized U -statistics of Weakly Dependent Stationary Processes, *Theory of Probability and its Applications* 37, 148-150.
- [48] Kitamura, Y. (2001), Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions, *Econometrica* 69, 1661-1672.
- [49] Li, Q. (1999), Consistent Model Specification Tests for Time Series Econometric Models, *Journal of Econometrics* 92, 101-147.
- [50] Liescher, E. (1996), Strong Convergence of Sums of α -mixing Random Variables with Applications to Density Estimation, *Stochastic Process and their Applications* 65, 69-80.

- [51] Linton, O. (2002), Edgeworth Approximations for Semiparametric Instrumental Variable Estimators and Test Statistics, *Journal of Econometrics* 106, 325-368.
- [52] Linton, O. and P. Gozalo (1997), Conditional Independence Restrictions: Testing and Estimation, *Discussion Paper*, Cowles Foundation for Research in Economics, Yale University.
- [53] Ma, C. K. and G. W. Kao (1990), On Exchange Rate Changes and Stock Price Reactions, *Journal of Business, Finance & Accounting* 17, 441-449.
- [54] Newey, K. N. (1994), Kernel Estimation of Partial Means and a General Variance Estimator, *Econometric Theory* 10, 233-253.
- [55] Owen, A. (1988), Empirical Likelihood Ratio Confidence Intervals for a Single Functional, *Biometrika* 75, 237-249.
- [56] Owen, A. (1990), Empirical Likelihood Ratio Confidence Regions, *Annals of Statistics* 18, 90-120.
- [57] Owen, A. (1991), Empirical Likelihood for Linear Models, *Annals of Statistics* 19, 1725-1747.
- [58] Pagan, A., and A. Ullah (1999), *Nonparametric Econometrics*. Cambridge University Press, Cambridge.
- [59] Paparoditis, E. and D. N. Politis (2000), The Local Bootstrap for Kernel Estimators under General Dependence Conditions, *Annals of the Institute of Statistical Mathematics* 52, 139-159.
- [60] Racine, J. and Q. Li (2000), Nonparametric Estimation of Conditional Distributions with Mixed Categorical and Continuous Data, *Discussion Paper*, Dept. of Economics, Univ. of South Florida and Texas A & M Univ.
- [61] Robinson, P. M. (1988), Root- N -consistent Semiparametric Regression, *Econometrica* 56, 931-954.
- [62] Robinson P. M. (1991), Consistent Nonparametric Entropy-based Testing, *Review of Economic Studies* 58, 437-453.
- [63] Rosenbaum, P. R. (1984), Testing the Conditional Independence and Monotonicity Assumptions of Item Response Theory, *Psychometrika* 49, 425-435.
- [64] Sheedy, E. (1998), Correlation in Currency Markets: A Risk-Adjusted Perspective, *Journal of International Financial Markets, Institutions & Money* 8, 59-82.
- [65] Sims, C. A. (1972), Money, Income, and Causality, *American Economic Review* 62, 540-552.
- [66] Skaug, H. J. and D. Tjøstheim (1993), A Nonparametric Test of Serial Independence Based on the Empirical Distribution Function, *Biometrika* 80, 591-602.
- [67] Skaug, H. J. and D. Tjøstheim (1996), Measure of Distance between Densities with Application to Testing for Serial Independence, In P. M. Robinson and M. Rosenblatt (eds.), *Time Series Analysis in Memory of E. J. Hannan*, 363-377. Springer : New York.
- [68] Sowell, F. (1996), Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework, *Econometrica* 64, 1085-1107.

- [69] Stinchcombe, M. B., and H. White (1998), Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative, *Econometric Theory* 14, 295-325.
- [70] Su, L., and H. White (2002), A Nonparametric Hellinger Metric Test for Conditional Independence, *Working Paper*, Dept. of Economics, UCSD.
- [71] Su, L., and H. White (2003), A Characteristic-Function-Based Test for Conditional Independence, *Discussion Paper 2003-11*, Dept. of Economics, UCSD.
- [72] Swanson, N. R. (1998), Money and Output Viewed Through a Rolling Window, *Journal of Monetary Economics* 41, 455-473.
- [73] Tenreiro, C (1995), Theoremes Limites pour les Erreurs Quadratiques Integrees des Estimateurs a Noyau de la Densite et de la Regression sous des Conditions de Dependance, *C. R. Acad. Sci. Paris* 320, 1535-1538.
- [74] 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.
- [75] Tjostheim, D (1996), Measures and Tests of Independence: a Survey, *Statistics* 28, 249-284.
- [76] Tripathi, G. and Y. Kitamura (2002), Testing Conditional Moment Restrictions, *Discussion Paper*, Dept. of Economics, University of Wisconsin-Madison.
- [77] 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.
- [78] White, H. (2000), *Asymptotic Theory for Econometricians*. Academic Press, San Diego.
- [79] Yao, Q. and D. Tritchler (1993), An Exact Analysis of Conditional Independence in Several 2×2 Contingency Tables, *Biometrics* 49, 233-236.
- [80] Yoshihara, K (1976), Limiting Behavior of U-statistics for Stationary, Absolutely Regular Processes, *Z. Wahrsch. Verw. Gebiete* 35, 237-252.
- [81] Yoshihara, K (1989), Limiting Behavior of Generalized Quadratic Forms Generated by Absolutely Regular Processes, In *Proceedings of the Fourth Prague Conference on Asymptotic Statistics*, 539-547.
- [82] Zheng, J. Z. (1997), A Consistent Specification Test of Independence, *Nonparametric Statistics* 7, 297-306.

Table 1: Empirical Rejection Frequency of the Test $\check{T}_{n,1}$ ($d_1=d_2=d_3=1$)

		5%							
		DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
n=100									
c=0.5		0.009	0.016	0.788	0.844	0.596	0.892	0.420	0.244
c=1		0.022	0.014	0.872	0.888	0.716	0.956	0.504	0.392
c=2		0.049	0.055	0.928	0.960	0.812	0.996	0.792	0.532
c=2.5		0.071	0.099	0.964	0.992	0.820	0.996	0.800	0.600
n=200									
c=0.5		0.014	0.011	0.996	0.996	0.936	1	0.78	0.544
c=1		0.015	0.015	1	1	0.98	0.996	0.86	0.668
c=2		0.04	0.033	0.996	1	0.968	1	0.944	0.828
c=2.5		0.043	0.06	1	1	0.98	1	0.968	0.844
n=500									
c=0.5		0.008	0.002	1	1	1	1	1	0.98
c=1		0.022	0.014	1	1	1	1	1	0.996
c=2		0.022	0.016	1	1	1	1	1	1
c=2.5		0.038	0.048	1	1	1	1	1	1
		10%							
		DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
n=100									
c=0.5		0.021	0.026	0.856	0.892	0.644	0.928	0.524	0.316
c=1		0.034	0.028	0.908	0.904	0.78	0.976	0.628	0.476
c=2		0.077	0.09	0.952	0.96	0.86	0.996	0.844	0.604
c=2.5		0.11	0.15	0.972	0.992	0.86	1	0.868	0.7
n=200									
c=0.5		0.025	0.017	1	1	0.964	1	0.856	0.644
c=1		0.025	0.023	1	1	0.984	1	0.912	0.756
c=2		0.066	0.067	0.996	1	0.988	1	0.968	0.884
c=2.5		0.073	0.107	1	1	0.992	1	0.988	0.896
n=500									
c=0.5		0.012	0.006	1	1	1	1	1	0.992
c=1		0.032	0.026	1	1	1	1	1	0.996
c=2		0.06	0.032	1	1	1	1	1	1
c=2.5		0.07	0.09	1	1	1	1	1	1

Table 2: Comparison of Tests for Causality ($d_1=d_2=d_3=1$)

	5%							
	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
n=100								
CM _n	0.054	0.058	0.920	0.548	0.504	0.412	0.384	0.188
KS _n	0.042	0.056	0.780	0.404	0.380	0.288	0.292	0.156
HEL _n	0.072	0.055	0.412	0.658	0.454	0.920	0.728	0.560
CHF _n	0.078	0.079	0.792	0.844	0.592	0.944	0.656	0.420
SEL _n	0.078	0.066	0.952	0.948	0.852	1	0.832	0.564
n=200								
CM _n	0.045	0.056	0.992	0.748	0.788	0.680	0.476	0.360
KS _n	0.067	0.053	0.952	0.552	0.660	0.532	0.336	0.284
HEL _n	0.045	0.025	0.829	0.870	0.485	1	0.984	0.852
CHF _n	0.063	0.062	0.940	0.968	0.848	1	0.864	0.604
SEL _n	0.044	0.044	1	1	0.98	1	0.976	0.864
n=500								
CM _n	0.025	0.046	1	0.984	0.992	0.984	0.824	0.728
KS _n	0.042	0.044	1	0.884	0.976	0.912	0.76	0.592
HEL _n	0.034	0.03	0.994	0.997	0.909	1	1	1
CHF _n	0.048	0.032	1	1	1	1	1	0.876
SEL _n	0.048	0.038	1	1	1	1	1	1
n=1000								
CM _n	0.048	0.042	1	1	1	1	0.976	0.904
KS _n	0.042	0.044	1	0.992	1	1	0.952	0.804
HEL _n	0.052	0.026	1	1	1	1	1	1
CHF _n	0.036	0.024	1	1	1	1	1	1
SEL _n	0.012	0.036	1	1	1	1	1	1

Table 2: Comparison of Tests for Causality ($d_1=d_2=d_3=1$, cont.)

	10%							
	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
n=100								
CM _n	0.094	0.101	0.964	0.652	0.644	0.480	0.472	0.304
KS _n	0.100	0.111	0.868	0.492	0.496	0.428	0.408	0.232
HEL _n	0.106	0.085	0.468	0.718	0.529	0.928	0.772	0.636
CHF _n	0.122	0.115	0.844	0.888	0.688	0.976	0.728	0.552
SEL _n	0.117	0.109	0.972	0.968	0.916	1	0.892	0.648
n=200								
CM _n	0.095	0.100	1	0.856	0.904	0.752	0.592	0.508
KS _n	0.096	0.089	0.988	0.676	0.756	0.676	0.484	0.404
HEL _n	0.08	0.05	0.889	0.912	0.602	1	1	0.904
CHF _n	0.100	0.101	0.960	0.988	0.908	1	0.924	0.708
SEL _n	0.08	0.068	1	1	0.992	1	0.992	0.9
n=500								
CM _n	0.052	0.065	1	0.992	1	0.992	0.896	0.792
KS _n	0.066	0.068	1	0.944	0.992	0.960	0.836	0.696
HEL _n	0.075	0.055	0.996	0.997	0.951	0.998	1	1
CHF _n	0.068	0.052	1	1	1	1	1	0.924
SEL _n	0.08	0.068	1	1	1	1	1	1
n=1000								
CM _n	0.112	0.076	1	1	1	1	0.992	0.936
KS _n	0.088	0.074	1	1	1	1	0.968	0.872
HEL _n	0.078	0.070	1	1	1	1	1	1
CHF _n	0.080	0.076	1	1	1	1	1	1
SEL _n	0.042	0.06	1	1	1	1	1	1

Table 3: Comparison of Tests for Causality ($d_1=2, d_2=d_3=1$)

	5%							
	DGP1'	DGP2'	DGP3'	DGP4'	DGP5'	DGP6'	DGP7'	DGP8'
n=100								
CM _n	0.029	0.040	0.684	0.368	0.204	0.608	0.248	0.344
KS _n	0.032	0.045	0.440	0.304	0.108	0.512	0.200	0.312
CHF _n	0.071	0.072	0.708	0.812	0.532	0.800	0.396	0.396
SEL _n	0.067	0.053	0.924	0.92	0.788	0.968	0.504	0.496
n=200								
CM _n	0.053	0.028	0.964	0.656	0.352	0.872	0.336	0.432
KS _n	0.040	0.025	0.792	0.480	0.196	0.780	0.292	0.388
CHF _n	0.049	0.051	0.920	0.972	0.800	0.988	0.588	0.640
SEL _n	0.044	0.036	1	1	0.984	1	0.772	0.84
n=500								
CM _n	0.068	0.028	1	0.936	0.692	0.992	0.632	0.764
KS _n	0.056	0.04	0.996	0.784	0.456	0.988	0.552	0.656
CHF _n	0.018	0.052	1	1	0.996	1	0.872	0.944
SEL _n	0.024	0.064	1	1	1	1	1	0.996
n=1000								
CM _n	0.076	0.033	1	1	0.952	1	0.912	0.968
KS _n	0.062	0.048	1	0.952	0.856	1	0.832	0.944
HEL _n	0.014	0.076	1	1	1	1	1	1
CHF _n	0.008	0.088	1	1	1	1	1	1
	10%							
	DGP1'	DGP2'	DGP3'	DGP4'	DGP5'	DGP6'	DGP7'	DGP8'
n=100								
CM _n	0.068	0.088	0.788	0.472	0.332	0.712	0.344	0.428
KS _n	0.064	0.076	0.632	0.408	0.224	0.624	0.296	0.400
CHF _n	0.106	0.114	0.756	0.880	0.624	0.892	0.524	0.504
SEL _n	0.11	0.087	0.944	0.96	0.836	0.988	0.616	0.624
n=200								
CM _n	0.104	0.060	0.984	0.756	0.468	0.908	0.456	0.528
KS _n	0.104	0.064	0.896	0.588	0.332	0.848	0.408	0.468
CHF _n	0.084	0.105	0.956	0.992	0.872	1	0.716	0.792
SEL _n	0.075	0.058	1	1	0.996	1	0.836	0.872
n=500								
CM _n	0.127	0.056	1	0.952	0.824	1	0.732	0.824
KS _n	0.104	0.061	1	0.876	0.62	1	0.644	0.772
CHF _n	0.042	0.082	1	1	0.998	1	0.904	0.968
SEL _n	0.048	0.096	1	1	1	1	1	1
n=1000								
CM _n	0.112	0.066	1	1	0.992	1	0.964	0.984
KS _n	0.092	0.065	1	0.992	0.924	1	0.920	0.960
CHF _n	0.016	0.104	1	1	1	1	1	1
SEL _n	0.048	0.132	1	1	1	1	1	1

Table 4: Comparison of Tests for Causality ($d_1=3, d_2=d_3=1$)

5%								
	DGP1"	DGP2"	DGP3"	DGP4"	DGP5"	DGP6"	DGP7"	DGP8"
n=200								
CM _n	0.036	0.044	0.724	0.368	0.236	0.772	0.320	0.440
KS _n	0.052	0.04	0.344	0.276	0.128	0.684	0.256	0.408
CHF _n	0.039	0.021	0.872	0.960	0.612	0.904	0.408	0.480
SEL _n	0.071	0.05	1	1	0.904	0.992	0.636	0.728
n=500								
CM _n	0.044	0.055	1.000	0.852	0.464	0.996	0.544	0.808
KS _n	0.047	0.042	0.872	0.656	0.280	0.976	0.472	0.740
CHF _n	0.044	0.020	1	1	0.964	1	0.792	0.884
SEL _n	0.066	0.046	1	1	0.992	1	0.936	1
n=1000								
CM _n	0.056	0.032	1	0.976	0.804	1	0.776	0.960
KS _n	0.078	0.016	1	0.864	0.576	1	0.752	0.944
CHF _n	0.040	0.088	1	1	1	1	0.884	1
SEL _n	0.074	0.082	1	1	1	1	1	1
10%								
	DGP1"	DGP2"	DGP3"	DGP4"	DGP5"	DGP6"	DGP7"	DGP8"
n=200								
CM _n	0.084	0.068	0.832	0.532	0.380	0.832	0.416	0.540
KS _n	0.116	0.06	0.572	0.420	0.248	0.788	0.376	0.480
CHF _n	0.061	0.042	0.888	0.976	0.696	0.952	0.492	0.636
SEL _n	0.128	0.077	1	1	0.94	0.996	0.716	0.796
n=500								
CM _n	0.081	0.068	1.000	0.892	0.656	0.996	0.640	0.856
KS _n	0.096	0.057	0.944	0.784	0.444	0.988	0.588	0.824
CHF _n	0.074	0.038	1	1	0.976	1	0.816	0.960
SEL _n	0.114	0.062	1	1	0.996	1	0.964	1
n=1000								
CM _n	0.08	0.072	1	0.992	0.888	1	0.846	0.976
KS _n	0.112	0.044	1	0.920	0.704	1	0.800	0.966
CHF _n	0.100	0.144	1	1	1	1	0.892	1
SEL _n	0.114	0.102	1	1	1	1	1	1

Table 5: Bivariate linear Granger causality test between exchange rates

H ₀ : Row doesn't cause column		Panel A: Linear Granger causality test				Panel B: Nonlinear Granger causality test			
		Canada	France	Italy	UK	Canada	France	Italy	UK
Canada	L _y =1, DX _{t-1} used	-	2.65	7.36 ^a	3.47 ^c	-	1.63 ^c	2.88 ^a	-0.70
	L _y =2, DX _{t-1} used	-	2.65	7.56 ^a	3.47 ^c	-	2.65^a	3.65 ^a	0.87
	L _y =2, DX _{t-2} used	-	0.22	2.77	0.66	-	0.63	1.52 ^c	0.68
	L _y =3, DX _{t-1} used	-	2.59	7.56 ^a	3.52 ^c	-	2.52^a	4.43 ^a	1.45 ^c
	L _y =3, DX _{t-2} used	-	0.22	2.68	0.62	-	0.17	0.63	0.52
	L _y =3, DX _{t-3} used	-	0.15	0.00	0.40	-	2.34^a	2.05^b	1.79^b
France	L _y =1, DX _{t-1} used	2.24	-	3.19 ^c	0.00	0.00	-	0.49	0.74
	L _y =2, DX _{t-1} used	2.37	-	3.57 ^c	0.00	0.93	-	-1.14	0.77
	L _y =2, DX _{t-2} used	2.67	-	3.16 ^c	3.88^b	1.13	-	0.12	0.69
	L _y =3, DX _{t-1} used	2.40	-	3.59 ^c	0.00	1.68^b	-	-4.44	2.01^b
	L _y =3, DX _{t-2} used	2.59	-	3.27 ^c	3.89^b	2.69^a	-	-3.85	0.99
	L _y =3, DX _{t-3} used	0.00	-	0.06	2.16	3.05^a	-	-4.95	-0.52
Italy	L _y =1, DX _{t-1} used	2.18	2.75 ^c	-	1.84	-0.26	3.77^a	-	2.27^b
	L _y =2, DX _{t-1} used	2.32	2.76 ^c	-	1.90	2.37^a	1.07	-	1.55 ^c
	L _y =2, DX _{t-2} used	0.16	0.52	-	0.17	1.84^b	2.13^b	-	-0.20
	L _y =3, DX _{t-1} used	2.48	2.76 ^c	-	1.80	2.15^a	-2.33	-	2.91^a
	L _y =3, DX _{t-2} used	0.11	0.54	-	0.15	2.43^b	-2.49	-	1.45 ^c
	L _y =3, DX _{t-3} used	0.00	0.49	-	0.69	1.34 ^c	-3.06	-	0.45
UK	L _y =1, DX _{t-1} used	0.48	0.20	0.16	-	-0.47	0.59	0.98	-
	L _y =2, DX _{t-1} used	0.50	0.19	0.12	-	1.51 ^c	2.44^a	2.41^a	-
	L _y =2, DX _{t-2} used	0.64	0.45	0.18	-	0.64	0.83	0.63	-
	L _y =3, DX _{t-1} used	0.51	0.21	0.10	-	1.85^b	2.23^b	2.25^b	-
	L _y =3, DX _{t-2} used	0.62	0.46	0.16	-	0.72	-0.10	-1.07	-
	L _y =3, DX _{t-3} used	0.00	0.70	1.12	-	0.55	-0.59	-1.62	-

The superscripts a, b and c denote rejection of the noncausality hypothesis at 1%, 5% and 10% significance levels, respectively. The bold elements in Panel A indicate the linear causal links that our nonlinear Granger causality test fails to detect at the 5% significance level and those in Panel B indicate the nonlinear causal links where the linear Granger causality test fails at the 5% significance level.

Table 6: Granger causality tests between stock prices and trading volumes

Panel A: Linear Granger causality test between ΔP and ΔV						
	$H_0: \Delta P$ doesn't cause ΔV			$H_0: \Delta V$ doesn't cause ΔP		
	Dow Jones	Nasdaq	S&P 500	Dow Jones	Nasdaq	S&P 500
$L_y=1$, DX_{t-1} used	0.11	8.28 ^a	8.63 ^a	0.66	0.32	0.02
$L_y=2$, DX_{t-1} used	0	6.85 ^a	9.84 ^a	0.65	0.15	0.04
$L_y=2$, DX_{t-2} used	1.22	3.63 ^c	3.69 ^c	0.1	0.11	0.03
$L_y=3$, DX_{t-1} used	0.05	5.78 ^b	10.82 ^a	0.65	0.15	0.05
$L_y=3$, DX_{t-2} used	1.85	3.21 ^c	4.88 ^b	0.1	0.11	0.07
$L_y=3$, DX_{t-3} used	3.83 ^c	0.02	6.49 ^b	0.07	0	0.02

Panel B: Nonlinear Granger causality test between ΔP and ΔV						
	$H_0: \Delta P$ doesn't cause ΔV			$H_0: \Delta V$ doesn't cause ΔP		
	Dow Jones	Nasdaq	S&P 500	Dow Jones	Nasdaq	S&P 500
$L_y=1$, DX_{t-1} used	18.38^a	2.56 ^a	8.17 ^a	1.00	0.82	2.11^b
$L_y=2$, DX_{t-1} used	7.08^a	6.23 ^a	7.31 ^a	-0.57	3.14^a	-0.36
$L_y=2$, DX_{t-2} used	6.70^a	3.66^a	5.75^a	-1.04	1.55 ^c	1.39 ^c
$L_y=3$, DX_{t-1} used	4.13^a	1.70 ^b	2.79 ^a	-2.82	-2.05	0.94
$L_y=3$, DX_{t-2} used	3.40^a	3.36^a	3.67 ^a	-2.80	-1.13	-0.42
$L_y=3$, DX_{t-3} used	5.86^a	2.08^b	3.01 ^a	-1.28	0.17	-0.40

The superscripts a, b and c denote rejection of the noncausality hypothesis at 1%, 5% and 10% significance levels, respectively. The bold elements in Panel A indicate the linear causal links that our nonlinear Granger causality test fails to detect at the 5% significance level and those in Panel B indicate the nonlinear causal links where the linear Granger causality test fails at the 5% significance level.

Table 7: Granger causality test between exchange rates (ΔE) and stock prices (ΔS)

Panel A: Linear Granger causality test between (ΔE) and stock prices (ΔS)												
	H₀: ΔE doesn't cause ΔS						H₀: ΔS doesn't cause ΔE					
	Canada	France	Germany	Italy	Japan	UK	Canada	France	Germany	Italy	Japan	UK
$L_y=1$, DX_{t-1} used	0.07	20.96 ^a	34.38 ^a	14.82^a	12.05 ^a	3.13 ^c	2.73 ^c	1.47	4.99^b	0.76	1.57	0.58
$L_y=2$, DX_{t-1} used	0.08	20.35 ^a	33.71 ^a	14.63 ^a	11.82 ^a	2.87 ^c	2.88 ^c	1.81	5.68^b	0.70	1.63	0.67
$L_y=2$, DX_{t-2} used	2.38	0.23	0.07	1.07	0.49	0.32	5.19^b	0.87	1.43	2.09	2.90 ^c	1.09
$L_y=3$, DX_{t-1} used	0.04	21.23 ^a	34.02 ^a	14.10 ^a	12.20 ^a	3.45	2.86	1.86	5.68 ^b	0.68	1.63	0.70
$L_y=3$, DX_{t-2} used	2.32	0.16	0.05	0.99	0.51	0.44	5.20^b	0.69	1.35	2.19	2.79 ^c	1.11
$L_y=3$, DX_{t-3} used	4.27 ^b	0.03	0.10	0.38	1.82	0.15	3.05 ^c	7.02 ^a	3.85 ^b	4.06 ^b	0.89	2.83 ^c

Panel B: Nonlinear Granger causality test between (ΔE) and stock prices (ΔS)												
	H₀: ΔE doesn't cause ΔS						H₀: ΔS doesn't cause ΔE					
	Canada	France	Germany	Italy	Japan	UK	Canada	France	Germany	Italy	Japan	UK
$L_y=1$, DX_{t-1} used	1.29	2.53 ^a	4.16 ^a	0.44	1.65 ^b	-0.40	1.54 ^c	-1.20	1.15	-1.71	-0.47	0.64
$L_y=2$, DX_{t-1} used	2.14^b	3.14 ^a	4.89 ^a	2.91 ^a	3.30 ^a	0.34	1.71 ^b	-0.08	1.29 ^c	-0.92	1.51 ^c	1.72^b
$L_y=2$, DX_{t-2} used	1.14	1.82^b	2.78^a	-0.25	1.56 ^c	1.26	0.71	0.72	2.45^a	1.58 ^c	3.54^a	2.31^b
$L_y=3$, DX_{t-1} used	2.17^b	4.29 ^a	5.87 ^a	1.65 ^b	4.03 ^a	-0.06	1.98^b	1.33 ^c	2.43 ^a	0.94	0.83	0.90
$L_y=3$, DX_{t-2} used	2.46^a	0.70	1.66^b	-0.11	0.00	1.17	0.79	1.24	3.23^a	1.12	3.29^a	2.89^a
$L_y=3$, DX_{t-3} used	3.77 ^a	0.83	1.95^b	2.60^a	0.41	1.01	1.77 ^b	1.96 ^b	2.08 ^b	2.00 ^b	1.58 ^c	2.70^a

The superscripts a, b and c denote rejection of the noncausality hypothesis at 1%, 5% and 10% significance levels, respectively. The bold elements in Panel A indicate the linear causal links that our nonlinear Granger causality test fails to detect at the 5% significance level and those in Panel B indicate the nonlinear causal links where the linear Granger causality test fails at the 5% significance level.

Table 8: Granger causality test between money, output and prices

Panel A: Linear Granger causality test between money, output and prices										
H ₀ :	DPI	RDPI	IP	CPI	PPI	DPI	RDPI	IP	CPI	PPI
	does not cause M1					does not cause M2				
L _y =1, DX _{t-1} used	9.74 ^a	5.43 ^b	0.10	0.22	1.18	11.71 ^a	8.74 ^a	0.24	1.75	5.40 ^b
L _y =2, DX _{t-1} used	10.36^a	6.99 ^a	0.02	0.55	1.02	11.61 ^a	9.30 ^a	0.43	2.10	5.85^b
L _y =2, DX _{t-2} used	0.03	0.34	2.16	0.01	2.41	0.02	0.10	6.70^a	0.33	0.00
L _y =3, DX _{t-1} used	10.45^a	7.74^a	0.00	1.31	1.22	11.04 ^a	9.39 ^a	0.86	3.18 ^c	6.40 ^b
L _y =3, DX _{t-2} used	0.00	0.04	2.75 ^c	0.06	2.27	0.02	0.02	7.69^a	0.11	0.03
L _y =3, DX _{t-3} used	4.88^b	7.35^a	8.89 ^a	1.18	0.43	0.55	1.73	4.53^b	3.78 ^c	0.69
H ₀ :	M1 does not cause					M2 does not cause				
	DPI	RDPI	IP	CPI	PPI	DPI	RDPI	IP	CPI	PPI
L _y =1, DX _{t-1} used	0.01	1.91	1.96	1.54	0.13	1.33	0.01	1.15	0.87	0.66
L _y =2, DX _{t-1} used	0.08	1.21	1.88	2.28	0.06	1.91	0.22	1.25	1.67	1.06
L _y =2, DX _{t-2} used	0.19	0.87	0.89	3.38 ^c	0.14	2.55	0.16	3.62 ^c	3.00 ^c	0.35
L _y =3, DX _{t-1} used	0.09	1.07	2.26	2.34	0.00	1.99	0.42	1.64	1.61	1.27
L _y =3, DX _{t-2} used	0.22	0.56	0.86	3.87 ^b	0.20	2.73	0.43	3.91^b	3.69^b	0.70
L _y =3, DX _{t-3} used	0.20	0.86	0.03	1.67	0.04	4.27 ^b	0.67	4.52^b	4.04^b	0.02

Panel B: Nonlinear Granger causality test between money, output and prices										
H ₀ :	DPI	RDPI	IP	CPI	PPI	DPI	RDPI	IP	CPI	PPI
	does not cause M1					does not cause M2				
L _y =1, DX _{t-1} used	2.42 ^a	3.24 ^a	0.95	2.50^a	1.82^b	5.12 ^a	2.96 ^a	2.36^a	1.12	1.89 ^b
L _y =2, DX _{t-1} used	1.17	1.88 ^b	0.80	3.48^a	2.37^a	4.11 ^a	3.85 ^a	1.28	1.49 ^c	1.44 ^c
L _y =2, DX _{t-2} used	1.37 ^c	1.47 ^c	2.17^b	2.85^a	1.23	3.48^a	4.26^a	0.08	1.57 ^c	2.17 ^b
L _y =3, DX _{t-1} used	0.24	0.82	0.50	2.84^a	2.34^a	4.56 ^a	6.24 ^a	-0.20	-0.16	0.07
L _y =3, DX _{t-2} used	0.15	0.35	1.10	2.17^b	-0.28	-0.67	1.32 ^c	-0.35	0.17	0.74
L _y =3, DX _{t-3} used	-0.58	-0.83	4.57 ^a	2.89^a	1.03	0.08	0.36	-0.46	1.19	-0.15
H ₀ :	M1 does not cause					M2 does not cause				
	DPI	RDPI	IP	CPI	PPI	DPI	RDPI	IP	CPI	PPI
L _y =1, DX _{t-1} used	2.46^a	2.44^a	2.26^b	1.83^b	1.26	2.71^a	2.27^b	0.55	1.07	1.92^b
L _y =2, DX _{t-1} used	1.26	1.29 ^c	0.80	2.02^b	4.64^a	1.09	3.96^a	0.20	2.34^a	4.52^a
L _y =2, DX _{t-2} used	0.63	3.26^a	-0.33	5.75^a	4.55^a	1.09	4.57^a	1.28	0.85	7.75^a
L _y =3, DX _{t-1} used	0.11	2.96^a	0.36	0.97	2.62^a	-0.50	4.89^a	-0.70	0.92	-0.85
L _y =3, DX _{t-2} used	-0.50	3.49^a	-1.33	3.31 ^a	2.05^b	-0.01	5.95^a	-0.07	-0.28	0.53
L _y =3, DX _{t-3} used	2.25^b	4.32^a	0.65	2.99^a	0.55	3.18 ^a	7.94^a	-1.20	0.29	0.10

The superscripts a, b and c denote rejection of the noncausality hypothesis at 1%, 5% and 10% significance levels, respectively. The bold elements in Panel A indicate the linear causal links that our nonlinear Granger causality test fails to detect at the 5% significance level and those in Panel B indicate the nonlinear causal links where the linear Granger causality test fails at the 5% significance level.

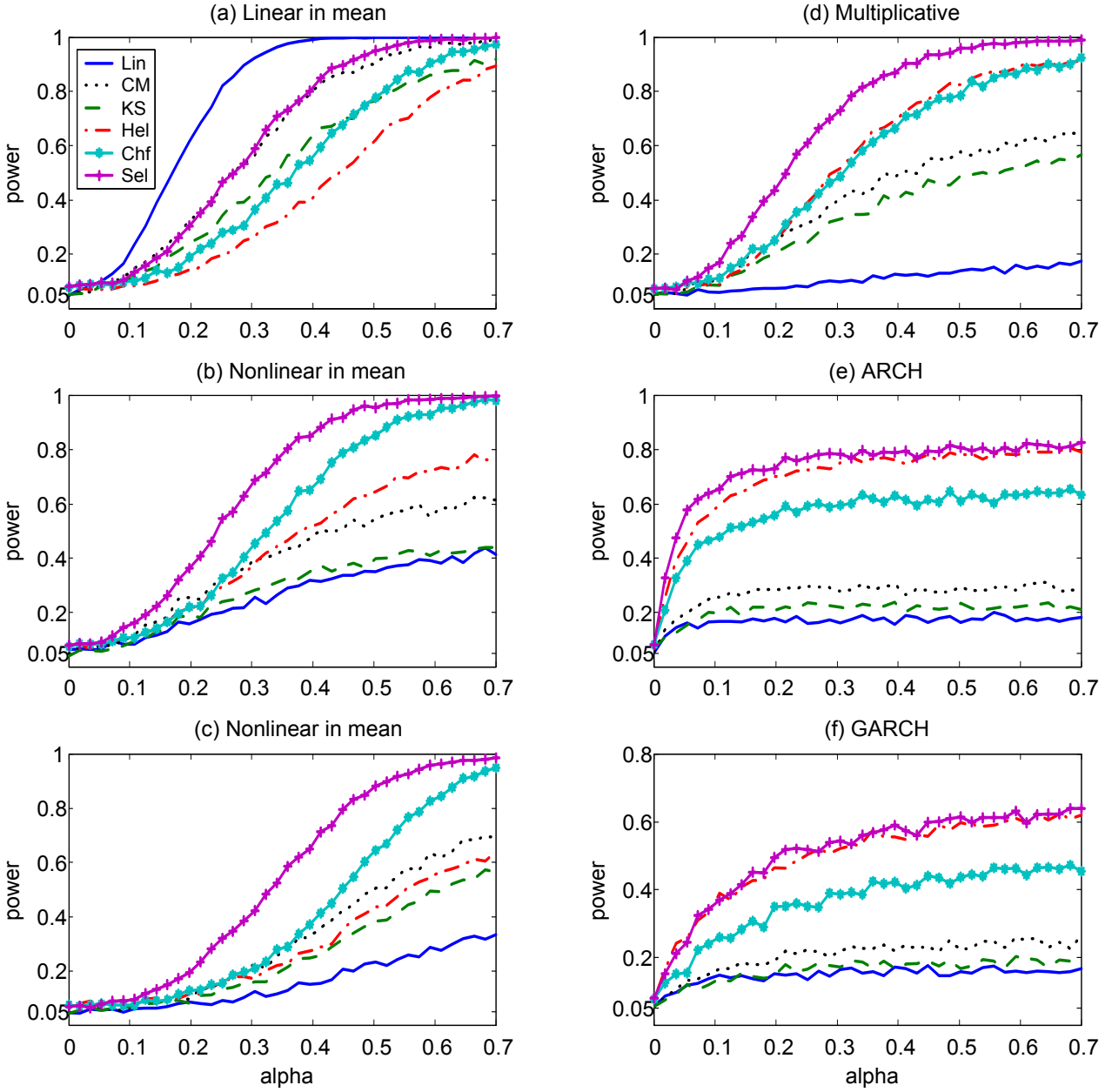


Figure 1: Power Comparison for Linear and Nonlinear Causality Tests ($d_1 = 1$, $n = 100$, 5%)

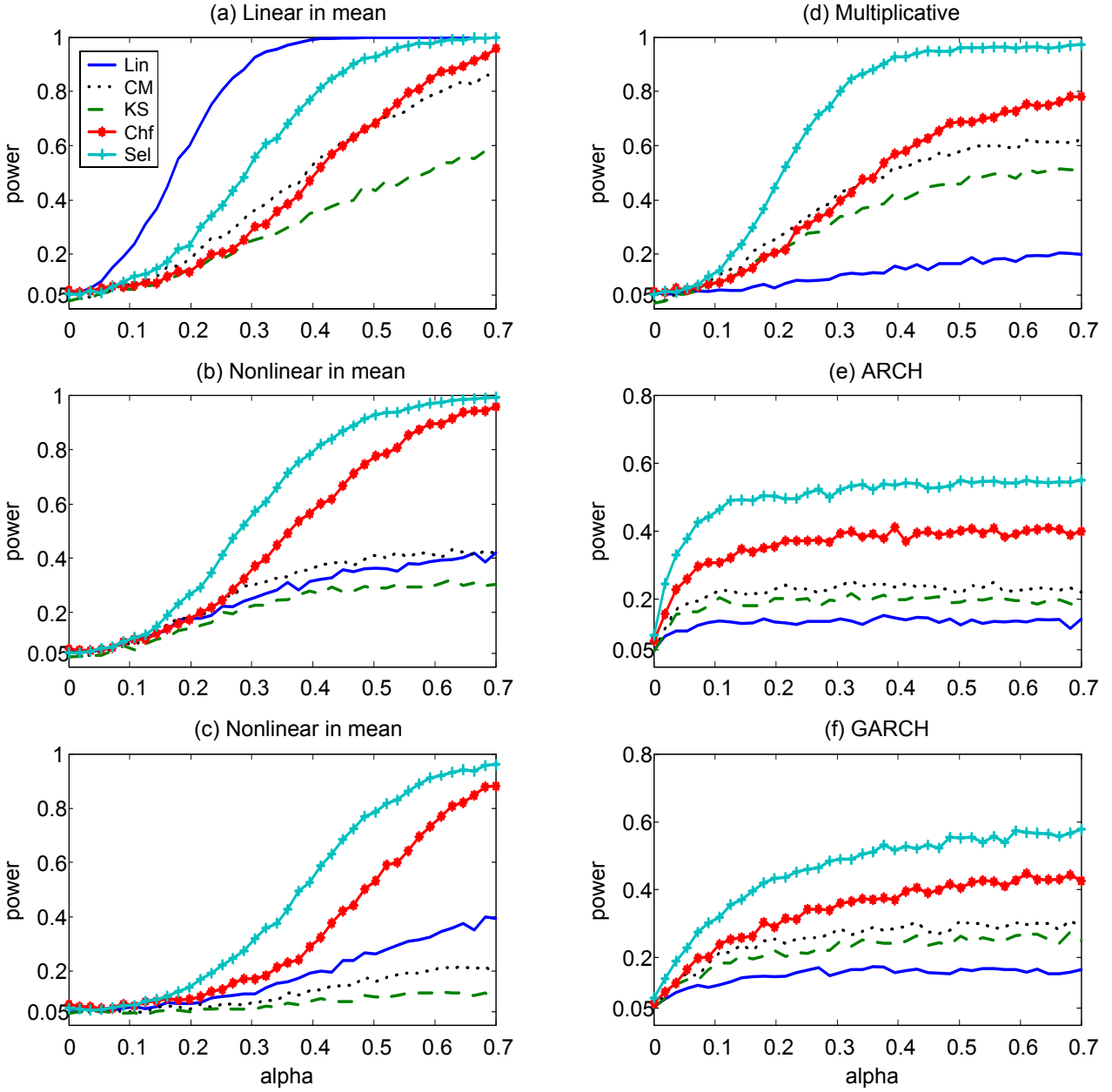


Figure 2: Power Comparison for Linear and Nonlinear Causality Tests ($d_1 = 2$, $n = 100$, 5%)