Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Three Essays on Nonparametric Tests in Nonparametric Models

Autor:

<u>Xiaojun Song</u>

Director:

Miguel A. Delgado

DEPARTAMENTO DE ECONOMIA

Getafe, junio de 2014

Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Three Essays on Nonparametric Tests in Nonparametric Models

Autor: *Xiaojun Song*

Director: Miguel A. Delgado

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Secretario:

Calificación:

Getafe, de de

To my parents and younger brother

whose endless love, support and encouragement sustained me throughout.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest appreciation and my sincere gratitude to my advisor, Prof. Miguel A. Delgado, who has been instrumental in the development and completion of this work. I remembered all the individual meetings and chats with him to discuss the thesis and many other issues. His patience, motivation, stimulating suggestions, unconditional support and consistent encouragement, helped me through the many frustrating moments of this thesis. Without his valuable guidance, inspiring instruction, boundless enthusiasm and persistent help, this work would not have been possible. I am extremely lucky and honoured to be his student.

Thanks also go to the many professors, especially Prof. Juan Carlos Escanciano, Prof. Carlos Martins-Filho, Prof. Peter Robinson, Prof. Stefan Sperlich, Prof. Abderrahim Taamouti and Prof. Carlos Velasco, for their profound and carefully aimed comments and suggestions that greatly improve this work.

My gratitude is extended to all the academic and administrative staff of the Department of Economics at the Universidad Carlos III de Madrid, for their devoted instruction and selfless support and help all these five years of my postgraduate studies in UC3M.

Special thanks go to all my colleagues in UC3M, especially Emre, Hang, Lovleen, Mian, Nikolas, Omar, Pedro and Yunrong, for their warmth, companionship, care, support and help. I will always remember their many kindnesses to me, their friendliness and generosity.

Last but not least, I wish to express my love and gratitude to my parents and younger brother for their unshakably strong faiths in me, unfailing support, constant encouragement and endless love, through the long march of almost fifteen years of my undergraduate and postgraduate studies. I am grateful for everything that they have done for me.

Contents

Introduction

My thesis focuses on the testing problems in models where specific parametric structures are not imposed. Specifically, it lies at the intersection of nonparametric tests, nonparametric models, time series and bootstrap. The proposed tests can have direct applications in many different fields in economics and econometrics.

Chapter 1. We propose significance tests in nonparametric autoregression. Under the null, forecast of any nonlinear autoregression of order p is unaffected by considering any extra lagged value. A necessary and sufficient condition, which forms a basis for the tests, is that the residuals of the p-th order nonparametric autoregression are uncorrelated with any measurable function of the lagged variables. The test statistic is based on Fourier transform of the autocorrelation function of the nonparametric residuals and functions of the lagged values. The tests are implemented with the assistance of a bootstrap technique. We illustrate the practical performance of the test by means of simulations and an empirical application.

Chapter 2. We propose tests of symmetry of conditional distributions around a nonparametric location function, which are able to detect general non-parametric alternatives. The test is developed in a general serial dependence context, where innovations may exhibit an unknown higher order serial dependence structure. The test statistic is a functional of the joint empirical distribution of non-parametric residuals and explanatory variables, which is able to detect non-parametric alternatives converging to the null at the parametric rate $\sqrt{n}$ with n the sample size. Critical values are estimated with the assistance of a bootstrap technique easy to implement, and the validity of the resulting test is formally justified. A Monte Carlo studies the finite sample properties of the test. We also include an application of the proposed test to investigate whether losses are more likely than gains given the available information in stock markets.

Chapter 3. We propose consistent tests of conditional independence specifically designed for data with weak dependence which are based on nonparametric regression. Under the null, the generalized errors form a martingale process, i.e., its expectation given the conditioning variables is zero. A necessary and sufficient condition, which forms a basis for the tests, is that the generalized errors of the distribution nonparametric regression are not correlated with any measurable function of the conditioning variables. The tests are implemented with the assistance of a multiplier bootstrap technique. We perform an extensive Monte Carlo simulation to evaluate the finite sample performance of the proposed tests. An empirical application to the nonlinear predictability of equity premium using variance risk premium is conducted.

Chapter 1

Significance Testing in Nonparametric Autoregression

1.1 Introduction

Let $\{Y_t\}_{t \in \mathbb{Z}}$ be a strictly stationary and ergodic time series process. We consider the problem of testing conditional mean independence of Y_t, given $Y_{t-1}, \ldots, Y_{t-p}$, with respect to any lagged value Y_{t-p-k}, $k \geq 1$. That is, under the null hypothesis, the forecast of any nonlinear autoregression (NLAR) of order p is unaffected by considering any extra lagged value. Formally,

$$\mathrm{H}_0 : \mathrm{E}(Y_t | Y_{t-1}, \ldots, Y_{t-p}, Y_{t-p-k}) = \mathrm{E}(Y_t | Y_{t-1}, \ldots, Y_{t-p}) \quad \text{a.s.} \quad \forall k \geq 1. \tag{1.1}$$

That is, any lagged value Y_{t-p-k} is not significant in the nonparametric autoregression. The alternative hypothesis H_1 is the negation of H_0. Notice that H_0 nests the hypothesis that any NLAR($p+1$) is identical to a NLAR(p). We shall also discuss specific test statistics for this less general hypothesis. Our methodology is directly applicable to test that any set of l extra lags do not affect the forecast of any NLAR(p), i.e.

$$\mathrm{H}_0^{(l)}: \mathrm{E}(Y_t|Y_{t-1},\ldots,Y_{t-p},Y_{t-p-k_1},\ldots,Y_{t-p-k_l}) = \mathrm{E}(Y_t|Y_{t-1},\ldots,Y_{t-p}) \quad \text{a.s.}$$
$$\forall k_j \geq 1, \quad j=1,\ldots,l,$$

and for $i<j$, $k_i<k_j$. For simplicity, we focus on H_0 in this paper. Notice that H_0 (or $\mathrm{H}_0^{(l)}$, a stronger hypothesis) is a necessary condition, but not a sufficient one, for

$$\mathrm{E}(Y_t|\mathscr{F}_{t-1}) = \mathrm{E}(Y_t|Y_{t-1},\ldots,Y_{t-p}) \quad \text{a.s.} \tag{1.2}$$

where $\mathscr{F}_{t-1}$ is the sigma algebra generated by the infinite past history $\{Y_{t-j}\}_{j=1}^{\infty}$, which is much harder to test. While (1.2) states that there is no NLAR of any order able to forecast differently than a NLAR of order p, H_0 means that the forecasting performance of any NLAR of order p is unaffected by introducing an extra lag.

The proposed test is the natural generalization of the nonparametric significance test of Delgado and Gonzalez-Manteiga (2001) to a time series context. In fact, checking that any NLAR(p) is equivalent to any NLAR($p+1$) can be done by a straightforward application of their nonparametric significance test. The main challenge here is that, rather than testing the significance of an explanatory variable, possibly multivariate, we must check the significance of infinite explanatory variables, in a stochastic process context. That is, to test H_0, we must check the significance of Y_{t-p-k} for any $k \geq 1$.

There are two main approaches for testing significance of a set of regressors in an i.i.d. context, which can be formally expressed by means of the restriction $\mathrm{E}(Y|X,Z) = \mathrm{E}(Y|X)$ a.s., with Y scalar and X and Z random vectors. One consists of comparing smooth estimates of $\mathrm{E}(Y|X,Z)$ and $\mathrm{E}(Y|X)$, typically using kernels, which was proposed by Fan and Li (1996). Their test statistic is a degenerate U-statistic, which converges to a standard normal under the null. The second approach suggested by Delgado and Gonzalez-Manteiga (2001) consists of using functionals of the standard empirical process of all the regressors (X and Z) marked by the restricted nonparametric residuals under the null, i.e., the residuals based on a smooth estimate of $\mathrm{E}(Y|X)$. This second approach avoids to estimate $\mathrm{E}(Y|X,Z)$ using smoothers, and is able to detect local alternatives converging to the null at the parametric rate $n^{-1/2}$, where n is the sample size. The test proposed by Fan and Li (1996) can only detect these types of alternatives converging to the null at a slower rate, i.e. $nh^{(d_X+d_Z)/2}$ with h a proper bandwidth sequence converging to zero and d_X and d_Z the dimensions of X and Z, respectively, but it can detect other hight frequency local alternatives in the lines suggested by Horowitz and Spokoiny (2001) among others, which in turn cannot be detected by Delgado and Gonzalez-Manteiga (2001)'s test. Therefore, tests based on these two approaches should be viewed as complements to each other rather than substitutes.

Though the discussion is centered in testing conditional mean independence, the proposed test can be easily extended to testing any higher-order conditional moment independence under suitable regularity conditions, including total conditional independence (see, e.g. Dawid (1979)). We discuss some extensions in the last section. The hypothesis of total conditional independence has been recently considered in time series context by Wang and Hong (2012) using Fan and Li (1996)'s proposal. This test can be directly applied to test the hypothesis that Y_t is independent of Y_{t-p-k} given $\{Y_{t-j}\}_{j=1}^{p}$ for a fixed $k \geq 1$. To test this hypothesis for any $k \geq 1$, that is H_0, we apply Delgado and Gonzalez-Manteiga (2001)'s methodology combined with Hong (1998, 1999 and 2000)'s spectral approach for jointly testing lack of autocorrelation between functions.

Our testing procedure can be extended test the following hypothesis,

$$\mathrm{E}(1(Y_t \leq y)|Y_{t-1},\ldots,Y_{t-p},Y_{t-p-k}) = \mathrm{E}(1(Y_t \leq y)|Y_{t-1},\ldots,Y_{t-p}) \quad \text{a.s.} \quad \forall y \in \mathbb{R} \quad \text{and} \quad \forall k \geq 1,$$

For each fixed and given y, we have to check the significance of Y_{t-p-k} for any $k \geq 1$. To this end, notice that the hypothesis of Markov property of order p for a time series is expressed as

$$\mathrm{E}(1(Y_t \leq y)|\mathcal{F}_{t-1}) = \mathrm{E}(1(Y_t \leq y)|Y_{t-1},\ldots,Y_{t-p}) \quad \text{a.s.} \quad \forall y \in \mathbb{R},$$

where $\mathcal{F}_{t-1}$ is the sigma algebra generated by $\{Y_{t-j}\}_{j\geq 1}$. It is clear that the testing procedures in this paper can serve as a powerful tool to test the Markov property of a stationary time series, where instead of testing significance of lagged values in the mean, we in fact test significance of lagged values in the whole distributional aspect. More examples can be found in parametric distributional autoregressive models. For instance, an autoregressive distributional model of order p is assumed to satisfy

$$F(y|\mathcal{F}_{t-1}, \theta_0) = F(y|Y_{t-1}, \ldots, Y_{t-p}, \theta_0)$$

for some unknown distribution function F and some $\theta_0 \in \Theta \subset R^{d_\theta}$. Here, again, given that the parametric specification F is correct, the first p lags and only the first p lags are significant in explaining the distribution of Y_t given its past.

The proposed test is also related to the vast literature of testing for (possibly nonlinear) Granger causality in mean, see e.g. Nishiyama et al. (2011). For a given time series process $\{(Y_t, X_t)\}_{t\in\mathbb{N}}$, the null hypothesis of Granger no causality in mean is formally expressed as $\mathrm{E}(Y_t|Y_{t-1}, \ldots, Y_1, X_{t-1}, \ldots, X_1) = \mathrm{E}(Y_t|Y_{t-1}, \ldots, Y_1)$ a.s. We refer to the rejection of this hypothesis by saying that "X_t (possibly nonlinearly) Granger causes Y_t in mean". In particular, we can directly apply our test to check

$$\mathrm{E}[Y_t|Y_{t-1}, \ldots, Y_{t-p}, X_{t-k}] = \mathrm{E}[Y_t|Y_{t-1}, \ldots, Y_{t-p}] \quad \text{a.s.} \quad \forall k \geq 1.$$

Finally, our formulation of conditional moment restriction (CMR) is well studied in the literature of consistent specification testing for parametric models. Those testing procedures consist of first expressing the null hypothesis at hand into a proper CMR and then transforming the CMR, using a well-known equivalence principle, to an infinite number of unconditional moment conditions indexed by some parameters. See Bierens (1982), Stute (1997) and Stinchcombe and White (1998) for more details about consistent specification testing in a CMR framework. The main differences of our paper with those papers are that our innovations are nonparametric in nature rather than parametric ones and we have an infinite number of CMRs because all $k \geq 1$ has to be considered in the testing problem.

To summarise, our approach has several attractive features. First, we reformulate the classical problem of lag selection using various information criteria in time series into a standard nonparametric significance testing one. In particular, we examine the exploratory power in the mean for any additional lagged value apart from the first p lags. Under the null hypothesis, any extra lag has no predictive power at all in the mean. Moreover, we have considered all possible lags Y_{t-p-k} such that $k \geq 1$ in a pairwise fashion such that no lag truncation is necessary, which partly circumvents the difficulty of curse of dimensionality problem. This feature differs from many existing tests related to the current paper, for example, tests of total conditional independence and tests of Granger non-causality, which could only consider a fixed number of lags. Second, our test is able to detect local alternatives converging to the null at the parametric rate $n^{-1/2}$. Third, the test implemented with the assistance of a multiplicative bootstrap procedure, which is based on the first order asymptotic expansion of the test statistic, does not require to compute the restricted nonparametric residuals for each bootstrap resample.

The remaining of the paper is organized as follows. In section 2, we introduce a Cramér-von Mises-type test statistic using a proper L_2 norm. Section 3 establishes the asymptotic distribution of the test statistic under the null and under fixed alternatives and investigates the asymptotic local power of the test under local alternatives. To implement our test in practice, we suggest a bootstrapped version of the test in section 4 and prove its asymptotic validity. Sections 5 and 6 provide some empirical evidence both from an extensive Monte Carlo simulation and from an empirical application to four stock market indices representing different financial market conditions across the world. We conclude the paper in section 7 and discuss some related extensions. Mathematical proofs and auxiliary lemmas are collected in the Appendix A and B. Throughout the paper, we use the following notation: C denotes a generic positive finite constant that may change from content to content, $|\cdot|$ for the Euclidean norm, and a^c for the complex conjugate of any complex number a, $\mathrm{i} = \sqrt{-1}$ for the imaginary number, the modulus is defined as $|a| := \sqrt{aa^c}$. Unless otherwise stated, all limits are taken as the sample size $n \to \infty$.

1.2 Test statistic

In this section, we describe the idea of how to construct our test statistic. Henceforth, define

$$m(x_1, \dots, x_p) = \mathrm{E}(Y_t | Y_{t-1} = x_1, \dots, Y_{t-p} = x_p)$$

for $(x_1, \dots, x_p) \in \mathbb{R}^p$ and the corresponding innovations

$$\varepsilon_t = Y_t - m(Y_{t-1}, \dots, Y_{t-p}).$$

The null hypothesis can be equivalently expressed in terms of these innovations, i.e. (1.1) is satisfied if and only if,

$$\mathrm{H}_0 : \mathrm{E}(\varepsilon_t | Y_{t-1}, \dots, Y_{t-p}, Y_{t-p-k}) = 0 \quad \text{a.s.} \quad \forall k \geq 1.$$

Notice that H_0 implies that, by the law of iterated expectations,

$$\mathrm{E}(\varepsilon_t | Y_{t-p-k}) = 0 \quad \text{a.s.} \quad \forall k \geq 1, \tag{1.3}$$

but (1.3) is not a sufficient condition for H_0. Hence, we may not be able to reject (1.3) under infinite many departures of H_0.

We exploit the fact that a necessary and sufficient condition for H_0 is that ε_t is not correlated with any measurable functions of $\{Y_{t-1}, \dots, Y_{t-p}, Y_{t-p-k}\}$ for any $k \geq 1$. Using the equivalence principle, see e.g. Theorem 1 in Bierens (1982), this claim can be expressed as

$$\mathrm{H}_0 : \mathrm{E}\left[\varepsilon_t \mathrm{e}^{\mathrm{i}\left(\sum_{j=1}^{p} x_j Y_{t-j} + y Y_{t-p-k}\right)}\right] = 0, \quad \forall k \geq 1$$

almost everywhere (a.e.) for $(x_1, \cdots, x_p) \in \mathbb{R}^p$ and $y \in \mathbb{R}$. It is also advisable in the nonparametric testing literature to avoid the random denominators arising from the estimation of ε_t, i.e. through estimation of the density of $\underline{Y}_{t-1} = (Y_{t-1}, \dots, Y_{t-p})'$. This is why we finally characterize H_0 in terms of the following measure of (nonlinear) cross-covariance function

$$\gamma_k(\underline{x}, y) = \mathrm{E}[\zeta_t(\underline{x})\phi_{t-k}(y)]$$

with $\underline{x} = (x_1, \ldots, x_p)'$, where

$$\phi_t(y) = \mathrm{e}^{\mathrm{i}yY_{t-p}}$$

and

$$\zeta_t(\underline{x}) = \varepsilon_t f(\underline{Y}_{t-1})\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}},$$

where f is the probability density function of $\underline{Y}_{t-1}$. As mentioned before, we introduce the weight $f(\underline{Y}_{t-1})$ to avoid the random denominators appearing in the nonparametric estimation involved in ε_t. Notice that $\gamma_k(\underline{x}, y)$ is an effective nonlinear dependence measure able to detect any departures from the null hypothesis (1.1).

Therefore, H_0 is satisfied if and only if

$$\mathrm{H}_0 : \gamma_k(\underline{x}, y) = 0, \quad \forall(\underline{x}, y) \in \mathbb{R}^{p+1}, \quad \forall k \geq 1. \tag{1.4}$$

It is natural to consider the Fourier transform of $\gamma_k(\underline{x}, y)$ in order to characterize the infinite number of unconditional moment restrictions involved in (1.4) for any fixed $(\underline{x}, y)$. Define $\gamma_{-k}(\cdot,\cdot) = \gamma_k(\cdot,\cdot)$ for $k \geq 1$. We use the cross-spectrum of $\{\zeta_t(\underline{x})\}_{t\in\mathbb{Z}}$ and $\{\phi_t(y)\}_{t\in\mathbb{Z}}$, i.e.

$$f^0(\omega, \underline{x}, y) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k(\underline{x}, y)\cos(k\omega), \quad (\omega, \underline{x}, y) \in [-\pi, \pi] \times \mathbb{R}^{p+1},$$

The above Fourier transform contains the same information as in $\{\gamma_k\}_{k\in\mathbb{Z}}$ and $f^0(\omega, \underline{x}, y)$ exists if

$$\sup_{(\underline{x},y)\in\mathbb{R}^{p+1}}\left\{\sum_{k=-\infty}^{\infty}|\gamma_k(\underline{x}, y)|\right\} < \infty,$$

which holds under a proper mixing condition, see e.g. Hong (2000). This type of Fourier transform has been used extensively, for example, by Hong (1998, 1999 and 2000) and Escanciano and Velasco (2006) in different contexts for checking that an infinite number of generalized autocovariance functions is equal to zero.

Our test is based on the cumulative cross-spectra distribution function

$$\begin{aligned} F^0(\lambda, \underline{x}, y) =& 2\int_0^{\lambda\pi} f^0(\omega, \underline{x}, y)\, d\omega \\ =& \gamma_0(\underline{x}, y)\lambda + 2\sum_{k=1}^{\infty} \gamma_k(\underline{x}, y)\frac{\sin k\pi\lambda}{k\pi}, \quad (\lambda, \underline{x}, y) \in [0,1]\times\mathbb{R}^{p+1}. \end{aligned} \tag{1.5}$$

Thus, H_0 can be expressed as

$$\mathrm{H}_0 : F^0(\lambda, \underline{x}, y) = \gamma_0(\underline{x}, y)\lambda, \quad \forall(\lambda, \underline{x}, y) \in [0,1]\times\mathbb{R}^{p+1}.$$

The test is based on a proper estimator of F^0 in (1.5), which employs a Nadaraya-Watson (NW) kernel estimator [see Nadaraya (1964) and Watson (1964)] of $\zeta_t(\underline{x})$,

$$\begin{aligned} \hat{\zeta}_t(\underline{x}) &= \hat{\varepsilon}_t \hat{f}(\underline{Y}_{t-1})\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \\ &= \frac{1}{(n-1)h^p}\sum_{s=1, s\neq t}^{n} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h}\right)(Y_t - Y_s)\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}}, \end{aligned}$$

with $\hat{\varepsilon}_t = Y_t - \hat{m}(\underline{Y}_{t-1})$ the nonparametric counterpart of ε_t,

$$\hat{m}(\underline{Y}_{t-1}) = \frac{\frac{1}{(n-1)h^p}\sum_{s=1,s\neq t}^{n} K\left(\frac{\underline{Y}_{t-1}-\underline{Y}_{s-1}}{h}\right) Y_s}{\hat{f}(\underline{Y}_{t-1})}$$

the leave-one-out NW estimator of conditional mean function $m(\underline{Y}_{t-1}) = \mathrm{E}(Y_t|\underline{Y}_{t-1})$ under the null, and

$$\hat{f}(\underline{Y}_{t-1}) = \frac{1}{(n-1)h^p}\sum_{s=1, s\neq t}^{n} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h}\right)$$

the leave-one-out NW estimator of marginal density $f(\underline{Y}_{t-1})$, where $K(\underline{u}) = \prod_{j=1}^{p} k(u_j)$ is a p-product kernel with $k(\cdot)$ a univariate kernel function, typically a probability density, and $h = h_n \in \mathbb{R}^+$ a bandwidth sequence converging to zero at a suitable rate as $n \to \infty$. Notice that we have adopted the same bandwidth $h_1 = \ldots = h_p = h$ in the above nonparametric estimation problem. However, our asymptotic theory is still valid when using different bandwidths $h_1, \ldots, h_p$ with more complicated bandwidth conditions in Assumption A.4. Then, based on a sample of observations $\{Y_t\}_{t=1}^n$, the dependence measure $\gamma_k(\underline{x}, y)$ is estimated by

$$\hat{\gamma}_{nk}(\underline{x}, y) = \frac{1}{n-k} \sum_{t=k+1}^{n} \hat{\zeta}_t(\underline{x}) \phi_{t-k}(y),$$

a variant of a standard U-process (see, e.g. Stute (1994)), which resembles that used by Delgado and Gonzalez-Manteiga (2001) for significance testing in nonparametric regression in an i.i.d. context. See Arcones and Yu (1994) for U-processes of stationary mixing sequences. Therefore, a natural estimator of $F^0(\lambda, \underline{x}, y)$ is given by

$$\hat{F}_n^0(\lambda, \underline{x}, y) = \hat{\gamma}_{n0}(\underline{x}, y)\lambda + 2\sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right)^{1/2} \hat{\gamma}_{nk}(\underline{x}, y) \frac{\sin k\pi\lambda}{k\pi},$$

where the term $(1-k/n)^{1/2}$ is a finite sample correction factor used in Hong (1999, 2000). This correction factor is often important in practice. It delivers a better finite sample performance, since it puts less weight on larger lags, where we will have less sample information. Nevertheless, it can be replaced by one without affecting the asymptotic theory. The test statistic is a proper functional of the following stochastic process

$$\begin{aligned} \hat{S}_n(\lambda, \underline{x}, y) &= \left(\frac{n}{2}\right)^{1/2} \left\{ \hat{F}_n^0(\lambda, \underline{x}, y) - \hat{\gamma}_{n0}(\underline{x}, y)\lambda \right\} \\ &= \sum_{k=1}^{n-1} \sqrt{n-k}\, \hat{\gamma}_{nk}(\underline{x}, y) \frac{\sqrt{2}\sin k\pi\lambda}{k\pi}. \end{aligned}$$

For the sake of simplicity, we propose to use the Cramér-von-Mises (CvM) type of functional, i.e.

$$\begin{aligned} CvM_n &= \int_0^1 \int_{\mathbb{R}^{p+1}} |\hat{S}_n(\lambda, \underline{x}, y)|^2 \, W(d\underline{x}, dy) d\lambda \\ &= \sum_{k=1}^{n-1} \frac{n-k}{(k\pi)^2} \int_{\mathbb{R}^{p+1}} |\hat{\gamma}_{nk}(\underline{x}, y)|^2 \, W(d\underline{x}, dy), \end{aligned} \tag{1.6}$$

where W is a non-negative weighting function satisfying some mild conditions, which will be discussed in the next section. For the choice of W, from a practical point of view, it should be chosen in a way such that CvM_n has a closed form. A reasonable choice of W, which has been recommended in other circumstances - e.g., Escanciano and Velasco (2006), Kuan and Lee (2004) - is a standard multivariate normal cumulative distribution function (CDF) under independence, i.e.

$$W(\underline{x}, y) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} \int_{-\infty}^{y} \frac{1}{(2\pi)^{\frac{p+1}{2}}} \mathrm{e}^{-\frac{1}{2}(\sum_{j=1}^{p} \bar{x}_j^2 + \bar{y}^2)} \, d\bar{x}_1 \cdots d\bar{x}_p d\bar{y}.$$

Another choice is the multivariate exponential distribution function suggested by Kuan and Lee (2004) in their tests for martingale difference hypothesis. This choice may also produce a test statistic with closed form in our context, but we leave it for future study.

1.3 Asymptotic theory

We mainly work with weakly dependent data in this paper, since it is very natural in an autoregressive framework, parametric or nonparametric. Specifically, we focus on the dependence concept of β-mixing or absolutely regular used in Robinson (1989) for dealing with U-statistic in a serial dependence context. Let $\{V_t\}_{t \in \mathbb{Z}}$ be a strictly stationary stochastic process and $\mathscr{F}_s^t$ denote the σ-algebra generated by $\{V_s, \cdots, V_t\}$ for $s \leq t$. The stochastic process $\{V_t\}$ is called β-mixing or absolutely regular with mixing coefficient $\beta(j)$, if as $j \to \infty$,

$$\beta(j) = \sup_{s \in \mathbb{N}} \mathrm{E} \left[\sup_{A \in \mathscr{F}_{s+j}^{\infty}} |\mathrm{Pr}(A | \mathscr{F}_{-\infty}^{s}) - \mathrm{Pr}(A)| \right] \to 0$$

at an appropriate rate.

To derive the asymptotic null distribution of the test statistic CvM_n in (1.6), we need to impose the following assumptions.

Assumption A.1 (Data Generating Process) (a) $\{Y_t\}$ is a strictly stationary, ergodic and absolutely regular process on $\mathbb{R}$ with β-mixing coefficients $\beta(j) = O(j^{-(2+\alpha)/\alpha})$ for some constant $\alpha \in (0, 2/3)$; (b) the marginal density $g(y)$ of Y_t is positive, bounded and continuous, the joint density $f(\underline{x}) \equiv f(x_1, \ldots, x_p)$ of $\underline{Y}_{t-1} \equiv (Y_{t-1}, \ldots, Y_{t-p})'$ is bounded, and the conditional density $f_k(\underline{x}|y)$ of $\underline{Y}_{t-1}$ given Y_{t-p-k}, for all $k \geq 1$, is positive a.s., bounded and continuously differentiable with respect to $\underline{x}$ up to order l with $l \geq 2$; (c) $E|Y_1|^{2+\delta} < \infty$ for some δ satisfying $\delta > \alpha/(1-\alpha)$.

Assumption A.2 (Autoregression Function) The true (restricted) autoregression function $m(\underline{x}) := \mathrm{E}(Y_t|\underline{Y}_{t-1} = \underline{x})$ is measurable with respect to $\mathscr{F}_t$, bounded and continuously differentiable with respect to $\underline{x}$ up to order l.

Assumption A.3 (Kernel Function) The function K is a l-th order product kernel function satisfying $K(u) = \prod_{j=1}^{p} k(u_j)$, $\int u^i K(u)\, du = \delta_{0i}$ for $i = 0, 1, \cdots, l-1$ and $\int u^l K(u)\, du \neq 0$, where $k(\cdot)$ is a bounded, symmetric univariate function on R and δ_{ij} is delta function equal to one when $i = j$ and equal to zero otherwise.

Assumption A.4 (Bandwidth) The bandwidth sequence h is such that: (a) $h \to 0$; and (b) $nh^p \to \infty$ and $nh^{2l} \to 0$, as $n \to \infty$.

Assumption A.5 (Weighting Function) The function W is a probability measure on $\mathbb{R}^{p+1}$, absolutely continuous with respect to Lebesgue measure.

Some remarks are necessary with regard to the above assumptions. Assumption A.1(a) is mild. Mixing conditions are commonly imposed to restrict the amount of temporal dependence in $\{Y_t\}$. For example, Amaro de Matos and Fernandes (2007) assume a β-mixing condition with a geometric decay rate in checking conditional independence and Chen and Hong (2012) assume an algebraic decay rate of $\beta(k)$ when testing for Markov property. See also Wang and Hong (2012) for testing conditional independence using characteristic functions. Our mixing condition is weaker than those imposed in the above mentioned papers. Many dynamic processes in the statistical and econometric literature satisfy this mixing assumption, for example, linear stationary autoregressive (AR) processes, autoregressive and moving average (ARMA) processes, and important nonlinear processes including bilinear process, nonlinear autoregressive (NLAR) process. Amongst the NLAR models, our approach especially allows the exponential autoregressive (EAR) model and the smooth transition autoregressive (STAR) model.

Furthermore, the mixing condition also holds for autoregressive conditional heteroskedastic (ARCH) model in Engle (1982) and generalized autoregressive conditional heteroskedastic (GARCH) model in Bollerslev (1986). See also Fan and Li (1999) for a detailed discussion. Assumption A.1(b) is regular and imposed to facilitate the Taylor expansion. The moment condition in Assumption 1(c) is mild too and it reflects the trade-off between the mixing condition and the moment restriction. However, our assumption here doesn't allow to consider the data generating processes like threshold autoregressive (TAR) models developed in Tong (1990), which is influential in the fields of econometrics and economics. Thus, it would be interesting to extend our methodology to this important case in the future.

Assumption A.2 specifies a smoothness condition for the autoregression function m under the null and under the alternatives, which is similar to those imposed in other statistical inference procedures involving smooth estimators (e.g. Robinson (1989)). This assumption excludes the cases where there may be structural break in m not smooth enough, for example, autoregression with known or unknown thresholds (the state-dependent or regime-switching model family, e.g. TAR model).

Assumption A.3 is also common in the nonparametric literature. The higher-order kernel K is often used to make compatible the rates of convergence of the bandwidth needed for bias and variance. Specifically, Assumption A.3 is needed in the projection of $\hat{\gamma}_{nk}(x, y)$ when performing Hoeffding decomposition, see Lemma A.1 in Appendix A.

Bandwidth conditions in Assumption A.4 are standard.

Assumption A.5 specifies the possible candidates of the weighting function W, for example, $(p+1)$-variate multivariate normal distribution under independence or multivariate exponential distribution. Different choices of weighting functions are useful to check the performance of our test and can improve the power performance in certain directions. We also have to choose W from a practical perspective in order to ease the computation. However, optimal choice of W in order to maximize the performance of our test is beyond the scope of this paper. In the simulation part of this paper, for the leading case $p = 1$, we choose standard bivariate normal distribution in order to obtain a test statistic with a closed form. One nice feature of this choice is that the bootstrap assisted test statistic has a closed form too.

1.3.1 Asymptotic null distribution

In order to find the asymptotic null distribution of the test statistic CvM_n, we first need to establish the asymptotic distribution of the stochastic process $\hat{S}_n(\lambda, \underline{x}, y)$ under the null. For completeness, we need to introduce several notation for asymptotic theory in Hilbert space. See Politis and Romano (1994), Chen and White (1996, 1998), or Chen and Fan (1999) for more applications of Hilbert space theory in statistics and econometrics.

Let $\eta = (\lambda, \underline{x}, y) \in \Pi = [0,1] \times (-\infty, \infty)^p \times (-\infty, \infty) \equiv [0,1] \times \mathbb{R}^{p+1}$. Let ν be the product measure of W (with W the same function as in (1.6)) and the Lebesgue measure on [0,1], i.e., $d\nu(\eta) \equiv d\nu(\lambda, \underline{x}, y) = W(d\underline{x}, dy)d\lambda$. In this paper, we consider $\hat{S}_n(\eta) \equiv \hat{S}_n(\lambda, \underline{x}, y)$ as a random element in the Hilbert space $L_2(\Pi, \nu)$ of all square integrable functions (with respect to some measure ν) with inner product

$$\begin{aligned} \langle f, g \rangle &:= \int_\Pi f(\eta) g^c(\eta)\, d\nu(\eta) \\ &\equiv \int_\Pi f(\lambda, \underline{x}, y) g^c(\lambda, \underline{x}, y)\, W(d\underline{x}, dy) d\lambda \end{aligned}$$

for any complex random variables $f \in L_2(\Pi, \nu)$ and $g \in L_2(\Pi, \nu)$, where g^c denote the complex conjugate of g. Notice that $L_2(\Pi, \nu)$ is endowed with the natural Borel σ-field induced by the norm $\|f\| = \langle f, f \rangle^{1/2}$ for any $f \in L_2(\Pi, \nu)$. Moreover, if Z is a $L_2(\Pi, \nu)$-valued random element and has probability distribution μ_Z, we say that Z has mean m if $E[\langle Z, g \rangle] = \langle m, g \rangle$, $\forall g \in L_2(\Pi, \nu)$. If $E\|Z\|^2 < \infty$ and Z has mean zero, we define the covariance operator of Z or μ_Z to be

$$C_Z(g) = E[\langle Z, g \rangle Z].$$

The covariance operator $C_Z(\cdot)$ is a continuous, linear, symmetric, positive definite operator from $L_2(\Pi, \nu)$ to $L_2(\Pi, \nu)$.

We need to define the weak convergence concept in this paper. Let $\Rightarrow$ denote weak convergence in the Hilbert space $L_2(\Pi, \nu)$ endowed with the norm induced by the inner product $\langle \cdot, \cdot \rangle$. For any sequence of random elements $Z_n \in L_2(\Pi, \nu)$ and a random element $Z \in L_2(\Pi, \nu)$, we say that $Z_n \Rightarrow Z$ in the Hilbert space $L_2(\Pi, \nu)$, if and only if for every random element $g \in L_2(\Pi, \nu)$, the inner product $\langle Z_n, g \rangle$ converges in distribution to that of $\langle Z, g \rangle$. Recall that $\varepsilon_t = Y_t - m(\underline{Y}_{t-1})$ under the null. The next theorem establish the weak convergence of $\hat{S}_n$ in the Hilbert space $L_2(\Pi, \nu)$.

Theorem 1: *Suppose Assumptions A.1-A.5 hold. Then under the null,*

$$\hat{S}_n \Rightarrow S_\infty$$

in the Hilbert space $L_2(\Pi, \nu)$, where S_∞ is a zero mean Gaussian process with covariance operator $C_{S_\infty}(\cdot)$ satisfying $\sigma_g^2 = \langle C_{S_\infty}(g), g \rangle$ with

$$\sigma_g^2 = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} E \left[\varepsilon_t^2 \int_{\Pi \times \Pi} g(\eta_1) g^c(\eta_2) \psi_{t-p-j}^c(\underline{x}_1, y_1) \psi_{t-p-k}(\underline{x}_2, y_2) \Phi_j(\omega_1) \Phi_k(\omega_2)\, d\nu(\eta_1) d\nu(\eta_2) \right]$$

where $\psi_{t-p-k}(\underline{x}, y) = \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left(\mathrm{e}^{\mathrm{i}yY_{t-p-k}} f(\underline{Y}_{t-1}) - \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y} \right)$, $f_k(\underline{Y}_{t-1}, Y_{t-p-k})$ is the joint density of $(\underline{Y}_{t-1}, Y_{t-p-k})'$, $\Phi_j(\lambda) = \sqrt{2} \sin j\pi\lambda / j\pi$, $\eta_1 = (\lambda_1, \underline{x}_1, y_1)$, and $\eta_2 = (\lambda_2, \underline{x}_2, y_2)$.

The basic idea of the proof of Theorem 1 is simple. We first find a new process S_n and show that the norm of the original process $\hat{S}_n$ is asymptotically equivalent to the norm of this new process S_n. Then, we prove the weak convergence of S_n to S_∞ in the Hilbert space $L_2(\Pi, \nu)$. Finally, by Theorem 4.2 of Billingsley (1968), we conclude that $\hat{S}_n$ also converges weakly to the same process S_∞ in the Hilbert space $L_2(\Pi, \nu)$.

The next corollary is only a direct consequence of the Continuous Mapping Theorem (CMT) and Theorem 1 above. See e.g. Billingsley (1968) Theorem 5.1.

Corollary 1: *Under the assumptions of Theorem 1 and under H_0*

$$\begin{aligned} CvM_n \to_d CvM_\infty &:= \int_{\Pi} |S_\infty(\eta)|^2 \, d\nu(\eta) \\ &= \int_0^1 \int_{\mathbb{R}^{p+1}} |S_\infty(\lambda, \underline{x}, y)|^2 \, W(d\underline{x}, dy) d\lambda. \end{aligned}$$

The asymptotic distribution of CvM_n can be expressed as a weighted sum of independent χ_1^2 random variables with weights depending on the unknown DGP. As it can be seen from the above null limiting distribution, the critical values will depend on the underlying DGP in an unknown and complicated way. Since it is difficult to obtain or tabulate critical values in this context, we need to implement our test with the assistance of a novel bootstrap method. We shall propose and validate a multiplier bootstrap procedure in section 4.

1.3.2 Consistency

Under the alternative H_1, we can find at least one $k \geq 1$ such that $\mathrm{E}(Y_t|\underline{Y}_{t-1}, Y_{t-p-k}) \neq \mathrm{E}(Y_t|\underline{Y}_{t-1})$ with positive probability, while $\mathrm{E}(Y_t|\underline{Y}_{t-1}, Y_{t-p-k}) = \mathrm{E}(Y_t|\underline{Y}_{t-1}) := m(\underline{Y}_{t-1})$ for all $k \geq 1$ a.s. under the null. This difference is essential in showing the consistency of the test and is reflected in the nonlinear dependence measure $\{\gamma_k\}_{k\in\mathbb{N}}$. It guarantees that the test statistic CvM_n diverges to infinity as $n \to \infty$ under the alternative which it is bounded under the null. Formally, we have the following theorem about the consistency of the proposed test.

Theorem 2: *Under Assumptions A.1-A.5 and under H_1,*

$$\frac{1}{n} CvM_n \to_p \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} \int_{\mathbb{R}^{p+1}} |\gamma_k(\underline{x}, y)|^2 W(d\underline{x}, dy) > 0.$$

Since under the alternative H_1, there exits at least one $k \geq 1$ such that $\gamma_k(\underline{x}, y) \neq 0$ for some subset of $\mathbb{R}^{p+1}$ with a positive Lebesgue measure, $CvM_n \to \infty$ and the test is consistent against alternatives of the null (1.1). However, if we use CvM_n to test (1.2), or generally speaking, when testing hypotheses with infinite lags or conditioning variables, we have to pay a price. As demonstrated in Escanciano and Velasco (2006) and Chen and Hong (2012), since only an important implication of the maintained hypothesis is tested, tests designed in a pair-wise fashion will typically miss some dynamics and will not be able to detect alternatives such that $\gamma_k(\underline{x}, y) = 0, \forall k \geq 1$ but not satisfying (1.2), though hopefully these types of alternatives are not prevalent in economics and finance. On the contrary, our new formulation of the hypothesis of interest (1.1) to a significance testing one will not have this problem.

1.3.3 Local power analysis

In this section we study the local power properties of our test statistic CvM_n. Since CvM_n always goes to infinity under a fixed alternative to (1.1), it is desirable to check the local power performance of the test. First of all, we investigate the behaviour of $\hat{S}_n(\lambda, \underline{x}, y)$ under a sequence of alternative hypotheses approaching to the null at the parametric rate $n^{-1/2}$. To this end, we need to introduce the following nonparametric local alternatives,

$$\mathrm{H}_{1n} : \mathrm{E}(Y_t|Y_{t-1}, \ldots, Y_{t-p}, Y_{t-p-k}) = m(Y_{t-1}, \ldots, Y_{t-p}) + \frac{1}{\sqrt{n}} g(Y_{t-1}, \ldots, Y_{t-p}, Y_{t-p-k}), \quad \text{a.s.} \tag{1.7}$$

for some $k \geq 1$. Denote $g_{tk} = g(Y_{t-1}, \ldots, Y_{t-p}, Y_{t-p-k})$. In the sequel, we need the sequence $\{g_{tk}\}_{t=1}^n$ to satisfy the following regularity assumption.

Assumption A.6 (Local Alternatives) (a) $\{g_{tk}\}$ is measurable with respect to $\mathscr{F}_{t-1}$, zero mean, strictly stationary, ergodic, and square integrable sequence; (b) there exists at least one $k \geq 1$ such that $\mathrm{E}[g_{tk}\psi_{t-p-k}(\underline{x}, y)] \neq 0$ for some subset of $\mathbb{R}^{p+1}$ with a positive Lebesgue measure.

Theorem 3: *Under the local alternatives in* (1.7)*, suppose Assumptions A.1-A.6 hold, we have*

$$\hat{S}_n \Rightarrow S_\infty^1$$

with

$$S_\infty^1 \stackrel{d}{=} S_\infty + G_\infty$$

in the Hilbert space $L_2(\Pi, \nu)$, where S_∞ is the Gaussian process defined in Theorem 1 and G_∞

is the deterministic shift function with

$$G_\infty(\eta) \equiv G_\infty(\lambda, \underline{x}, y) = \sum_{k=1}^{\infty} L_k(\underline{x}, y) \frac{\sqrt{2}\sin k\pi\lambda}{k\pi},$$

where $L_k(\underline{x}, y) = E[g_{tk}\psi_{t-p-k}(\underline{x}, y)]$ *with* $\psi_{t-p-k}(\underline{x}, y)$ *defined in Theorem 1.*

Under the local alternatives H_{1n} and imposing Assumption A.6, there exists at least one $k \geq 1$ such that $L_k(\underline{x}, y) \neq 0$ for some subset $\mathbb{R}^{p+1}$ with a positive Lebesgue measure. We then have

$$\sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} \int_{\mathbb{R}^{p+1}} |L_k(\underline{x}, y)|^2 \, W(d\underline{x}, dy) > 0.$$

Therefore, our test is able to detect local alternatives that converge to the null at the parametric rate $n^{-1/2}$. Though we have employed kernel estimators to estimate the unknown innovations ε_t under the null hypothesis of no significance for any extra lagged value greater than p, the test is still able to detect local alternatives converging to the null at a parametric rate, see also Delgado and Gonzalez-Manteiga (2001) for a simpler case. This asymptotic power property is attractive. The reason behind this is partly due to the weighted averaging of all lag orders greater than p and partly due to the use of higher order kernel function. Notice that the asymptotic distribution of $\hat{S}_n$ is non-trivially shifted under the local alternatives (1.7). Since the shift term G_∞ is not identically zero over the support of $\eta = (\lambda, x, y)$ under H_{1n}, otherwise, because of the orthogonality of $\sin k\pi\lambda/k\pi$ we need to have $L_k(\underline{x}, y) \equiv 0$, $\forall k \geq 1$. Thus, our test has non-trivial power against the class of local alternatives in H_{ln}.

As in Corollary 1, the next corollary is also an application of Continuous Mapping Theorem and Theorem 4.

Corollary 2: *Under the local alternatives* (1.7)*, if Assumptions A.1-A.6 hold,*

$$\begin{aligned} CvM_n \to_d CvM_\infty^1 &:= \int_\Pi |S_\infty(\eta) + G_\infty(\eta)|^2 \, d\nu(\eta) \\ &= \int_0^1 \int_{\mathbb{R}^{p+1}} |S_\infty(\lambda, \underline{x}, y) + G_\infty(\lambda, \underline{x}, y)|^2 \, W(d\underline{x}, dy) d\lambda. \end{aligned}$$

From the above corollary, we see that under the local alternatives, the limiting distribution of the test statistic CvM_n converges to a different distribution asymptotically due to the presence of a deterministic shift function G_∞. This additional term will guarantee the non-zero local power property of our test.

1.4 Bootstrap

Since the asymptotic null distribution of $\hat{S}_n$ and the corresponding test statistic CvM_n depend on the underlying data generating process in a highly complicated and unknown way, it is rather difficult for us to calculate or tabulate the critical values in practice. To this end, we propose a bootstrap method to estimate the critical values of our tests. Our bootstrap procedure is somewhat related to the wild bootstrap, e.g., Wu (1986), Liu (1988). See also Neumeyer and Dette (2007) for recent applications of wild bootstrap in a hypothesis testing context. But our bootstrap is different, since rather than resampling imposing the restriction on H_0, we use the first order asymptotic representation of $\hat{S}_n(\eta)$. Specifically speaking, our bootstrap is of a multiplier bootstrap type proposed in Delgado and Gonzalez Manteiga (2001). It has nice theoretical and applied properties and it is straightforward to verify its asymptotic validity. Moreover, it is computationally easy to implement in practice.

We propose and justify the use of the multiplier bootstrap in this section. Inspired by Delgado and Gonzalez Manteiga (2001), in order to take advantage of the asymptotic equivalence of the norms of process $\hat{S}_n$ and the norm of process S_n defined in (1.8) in Hilbert space $L_2(\Pi, \mu)$ as shown in Lemma A.1 in Appendix A, we suggest to implement our test assisted by the following bootstrap method, i.e. we define the bootstrapped process

$$\hat{S}_n^*(\eta) = \sum_{k=1}^{n-1} \sqrt{n-k}\hat{\gamma}_{nk}^*(\underline{x}, y)\frac{\sqrt{2}\sin k\pi\lambda}{k\pi}$$

with

$$\hat{\gamma}^*_{nk}(\underline{x}, y) = \frac{1}{n-k} \sum_{t=k+1}^{n} \hat{\varepsilon}_t \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left[\mathrm{e}^{\mathrm{i}yY_{t-p-k}} \hat{f}(\underline{Y}_{t-1}) - \int \mathrm{e}^{\mathrm{i}y\bar{y}} \hat{f}_k(\underline{Y}_{t-1}, \bar{y}) \, d\bar{y} \right] v_t,$$

where $\hat{\varepsilon}_t$ are the restricted residuals from the nonparametric regression of m and $\{v_t\}_{t=1}^n$ is a sequence of random variables with zero mean, unit variance (and sometimes zero skewness depending upon the different use of v_t), bounded support and is independent of $\{Y_t\}_{t=1}^n$. Notice that $\hat{f}_k(\underline{Y}_{t-1}, Y_{t-p-k})$ is the NW kernel estimator of joint density function of $(\underline{Y}_{t-1}, Y_{t-p-k})$. In practice when we use standard normal density as the univariate kernel function in NW estimator, the form of $\hat{\gamma}^*_{nk}(\underline{x}, y)$ is algebraically equivalent to

$$\hat{\gamma}^*_{nk}(\underline{x}, y) = \frac{1}{n-k} \sum_{t=k+1}^{n} \hat{\zeta}^*_t(\underline{x}) \bar{\phi}_{t-k}(y),$$

where

$$\hat{\zeta}^*_t(\underline{x}) = \hat{\varepsilon}_t \hat{f}(\underline{Y}_{t-1}) \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} v_t$$

and

$$\bar{\phi}_{t-k}(y) = \mathrm{e}^{\mathrm{i}yY_{t-p-k}} - \frac{1}{n-k} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}yY_{t-p-k}}$$

is the centered version of $\mathrm{e}^{\mathrm{i}yY_{t-p-k}}$.

Intuitively, the bootstrap distribution of $\hat{S}^*_n$ is supposed to "mimic" the asymptotic distribution of $\hat{S}_n$ under the null hypothesis. The bootstrapped version of our original test statistic CvM_n is given by

$$\begin{aligned} CvM^*_n &= \int_0^1 \int_{\mathbb{R}^{p+1}} |\hat{S}^*_n(\lambda, \underline{x}, y)|^2 \, W(d\underline{x}, dy) d\lambda \\ &= \sum_{k=1}^{n-1} \frac{n-k}{(k\pi)^2} \int_{\mathbb{R}^{p+1}} |\hat{\gamma}^*_{nk}(\underline{x}, y)|^2 \, W(d\underline{x}, dy). \end{aligned}$$

We reject the null hypothesis at significance level α if $CvM_n > cv^*_\alpha$, the critical value of CvM^*_n.

A stronger condition concerning the bandwidth h is needed in order to validate the multiplier bootstrap in our context.

Assumption A.4' The bandwidth sequence h is such that: (a) $h \to 0$; and (b) $nh^{2p} \to \infty$ and $nh^{2l} \to 0$, as $n \to \infty$.

In the next theorem we formally establish the asymptotic validity of our multiplier bootstrap procedure, so that we can approximate the asymptotic distribution of the process $\hat{S}_n$ by that of $\hat{S}_n^*$.

Theorem 4: *Suppose Assumptions A.1-A.3, A.4' and A.5 hold, then under the null hypothesis* (1.1)*, under any fixed alternative hypothesis or under the local alternatives* (1.7)*,*

$$\hat{S}_n^* \Rightarrow S_\infty \quad \textit{in probability}$$

in the Hilbert space $L_2(\Pi, \nu)$, where S_∞ is the Gaussian process defined in Theorem 1 and $\Rightarrow$ in probability denotes the weak convergence in probability under the bootstrap law, i.e., conditional on the original sample $\{Y_t\}_{t=1}^n$.

1.5 Monte Carlo simulation

In this section, we report results from an extensive Monte Carlo simulation to investigate the performance of our test in small and moderately large samples. Throughout this section, $\{Y_t\}_{t=1}^n$ is the time series of our primary interest, while we want to check whether or not and/or how many lags of Y_t are significant and accounts for the forecasting performance of Y_t in the mean.

Our hypothesis of interest is $\mathrm{H}_0 : \mathrm{E}[Y_t|Y_{t-1}, \ldots, Y_{t-p}, Y_{t-p-k}] = \mathrm{E}[Y_t|Y_{t-1}, \ldots, Y_{t-p}]$ for any $k \geq 1$. Henceforth, we focus on the leading case of $p = 1$, which is also the main interest of economic and financial modelling and the most studied case in the analysis of time series data. To implement our test, we first outline the multiplier bootstrap procedure developed in section 4 in the following five steps:

Step 1: Estimate the nonparametric model $Y_t = m(Y_{t-1}) + \varepsilon_t$ using the original sample

$\{Y_t\}_{t=1}^n$ and obtain the nonparametric residuals $\hat{\varepsilon}_t = Y_t - \hat{m}(Y_{t-1})$ for $t = 2, \dots, n$.

Step 2: Compute $\hat{\gamma}^*_{nk}(x, y)$ for $k = 1, \cdots, n-2$ based on $\{v_t\}_{t=1}^n$, where v_t is a two point distribution, i.e. $\{v_t\}_{t=1}^n$ is a sequence of independent identically distributed (i.i.d.) Bernoulli random variables with $P(v_t = 0.5(1-\sqrt{5})) = (\sqrt{5}+1)/2\sqrt{5}$ and $P(v_t = 0.5(1+\sqrt{5})) = (\sqrt{5}-1)/2\sqrt{5}$ and is independent of the original sample $\{Y_t\}_{t=1}^n$[1].

Step 3: Compute $\hat{S}^*_n(\eta)$ and CvM^*_n.

Step 4: Repeat Steps 2-3 above B times to give a sample $\{CvM^*_{n,b}\}_{b=1}^B$ of the bootstrapped version of test statistic CvM_n. The distribution of this sample, which is commonly called the "bootstrap distribution" in the literature, will "mimic" the distribution of CvM_n under the null hypothesis.

Step 5: Let $c^{CvM*}_{\alpha,B}$ be the $(1-\alpha)$-th sample quantile of the "bootstrap distribution" of CvM^*_n. It is the bootstrap estimate of the α-level critical value. More formally, let $CvM^*_{n,(1)} \leq CvM^*_{n,(2)} \leq \cdots \leq CvM^*_{n,(B)}$ denote the ordered values of the B realizations of CvM^*_n, we choose $c^{CvM*}_{\alpha,B} = CvM^*_{n,([B(1-\alpha)+1])}$. For instance, in the case of $\alpha = 0.05$ and $B = 100$, we would take $c^{CvM*}_{\alpha,B} = CvM^*_{n,(96)}$. We reject the null hypothesis at the significance level α if $CvM_n > c^{CvM*}_{\alpha,B}$.

We state the Monte Carlo setup. In the sequel, let ε_t be a sequence of independently and identically distributed standard normal random variables, that is, i.i.d. $N(0,1)$. To examine the empirical size of the test under the null hypothesis, we consider the following five DGPs:

DGP S1 [AR(1)]: First-order autoregressive model, $Y_t = 0.5Y_{t-1} + \varepsilon_t$.

DGP S2 [NLAR(1)-1]: First-order nonlinear autoregressive model, $Y_t = |Y_{t-1}|^{0.8} + \varepsilon_t$.

DGP S3 [EXP(1)]: First-order exponential autoregressive model, $Y_t = 0.6Y_{t-1}\exp(-0.5Y^2_{t-1}) + \varepsilon_t$.

DGP S4 [NLAR(1)-2]: First-order nonlinear autoregressive model, $Y_t = -Y_{t-1}/(1 + Y^2_{t-1}) + \varepsilon_t$.

DGP S5 [ARCH(1)]: Autoregressive conditional heteroskedasticity model of order one,

[1]The above choice of v_t is widely used in the bootstrap literature, see e.g., Delgado and Gonzalez Manteiga (2001) and Escanciano and Velasco (2006). Another popular choice of $\{v_t\}$ is i.i.d. Bernoulli random variables with $P(v_t = 1) = P(v_t = -1) = 0.5$ (Rademacher random variables) and is applied in Liu (1988) and de Jong (1996).

$Y_t = \sigma_t \varepsilon_t$ with $\sigma_t^2 = 0.1 + 0.1Y_{t-1}^2$.

To examine the empirical power of our test, we use the following seven DGPs:

DGP P1 [AR(2)-1]: Second-order autoregressive model, $Y_t = 0.5Y_{t-1} + 0.3Y_{t-2} + \varepsilon_t$.

DGP P2 [AR(2)-2]: Second-order autoregressive model without the presence of the first-order term, $Y_t = 0.5Y_{t-2} + \varepsilon_t$.

DGP P3 [MA(1)]: First-order moving average model, $Y_t = \varepsilon_t + 0.5\varepsilon_{t-1}$.

DGP P4 [MA(2)-1]: Second-order moving average model, $Y_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.3\varepsilon_{t-2}$.

DGP P5 [MA(2)-2]: Second-order moving average model without the presence of the first-order term, $Y_t = \varepsilon_t + 0.5\varepsilon_{t-2}$.

DGP P6 [NLAR(2)]: Second-order nonlinear autoregressive model without the presence of the first-order term, $Y_t = |Y_{t-2}|^{0.8} + \varepsilon_t$.

DGP P7 [NLMA(2)]: Second-order nonlinear moving average model, $Y_t = \varepsilon_{t-1}\varepsilon_{t-2}(\varepsilon_{t-2} + \varepsilon_t + 1)$.

DGPs S3, S5 and P7 are used in Escanciano and Velasco (2006). DGPs S2, P3 and P7 are used in Chen and Hong (2012). We have also considered the DGP $Y_t = \varepsilon_t \sim$ i.i.d $N(0,1)$ which is the case of martingale difference hypothesis (MDH). The MDH case is artificial since i.id. $N(0,1)$ needs no lag to model it. We find that the performance of our test when testing for MDH is acceptable even with small samples. The price we have to pay for using kernel method to estimate the mean function, which is a constant zero in this case, is minimal. The results are not reported here.

We consider three sample sizes: n=100, 250, 500. For each DGP, we first generate $n + 200$ observations and then discard the first 200 observations to minimize the effect due to the initial values. We then standardize the observations for each DGP to have zero mean and unit variance. The number of Monte Carlo experiments is 1000 and the number of bootstrap replications is $B = 500$. We consider a nominal size of $\alpha = 5\%$. Results from the other nominal sizes are similar and are available upon request. We use standard normal density as our kernel function. We adopt bandwidth of the form $h = c \times n^{-1/5}$ derived from univariate regression estimation problem. We have experimented many c's ranged from 1.0 to 2.0. There seems to be not too much significant difference among them. Therefore, in this simulation, we only report the results from $c = 1.5$, which delivers overall reasonable size and power in small samples. However, regarding to the different choices of c's, one general pattern we observe is that for too small c, the proposed test tends to be undersized while for relatively large c it tends to be oversized. Some data-driven ways of selecting optimal bandwidth h^* are highly desirable in our significance testing context, e.g. the least squares cross-validation or plug-in method. However, how to select the optimal value c^* (or generally speaking, the optimal choice of bandwidth h) to maximize the overall performance of our test in terms of size and power is very tricky and beyond the scope of this paper.

As for the choice of weighting function W, the standard bivariate normal distribution with correlation parameter $\rho = 0$ (i.e. bivariate normal under independence) is used. In fact, when $W(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \mathrm{e}^{-(\bar{x}^2+\bar{y}^2)/2}/(2\pi)\, d\bar{x}d\bar{y}$, a standard bivariate normal CDF under independence, we can obtain the following two closed form expressions for CvM_n and its bootstrapped counterpart CvM_n^* based on $\{Y_t\}_{t=1}^n$,

$$
\begin{aligned}
CvM_n = & \sum_{k=1}^{n-2} \frac{1}{(n-k)(k\pi)^2} \sum_{t=k+2}^{n} \sum_{s=k+2}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_s \hat{f}(Y_{t-1}) \hat{f}(Y_{s-1}) \mathrm{e}^{-0.5(Y_{t-1}-Y_{s-1})^2} \\
& \times \mathrm{e}^{-0.5(Y_{t-1-k}-Y_{s-1-k})^2},
\end{aligned}
$$

and

$$
\begin{aligned}
CvM_n^* = & \sum_{k=1}^{n-2} \frac{1}{(n-k)(k\pi)^2} \sum_{t=k+2}^{n} \sum_{s=k+2}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_s \hat{f}(Y_{t-1}) \hat{f}(Y_{s-1}) \mathrm{e}^{-0.5(Y_{t-1}-Y_{s-1})^2} \\
& \times \left(\mathrm{e}^{-0.5(Y_{t-1-k}-Y_{s-1-k})^2} - \frac{2}{n-k-1} \sum_{t'=k+2}^{n} \mathrm{e}^{-0.5(Y_{t'-1-k}-Y_{t-1-k})^2} \right. \\
& \left. + \frac{1}{(n-k-1)^2} \sum_{t'=k+2}^{n} \sum_{s'=k+2}^{n} \mathrm{e}^{-0.5(Y_{t'-1-k}-Y_{s'-1-k})^2} \right) v_t v_s.
\end{aligned}
$$

Hence, this particular choice of W avoids the numerical integration of CvM_n and CvM_n^* and speeds the computation greatly. Moreover, this choice is appealing since it makes possible the

Table 1.1: Empirical size of the proposed test CvM_n

	$h = 1.5 \times n^{-0.2}$				
	DGP S1	DGP S2	DGP S3	DGP S4	DGP S5
$n = 100$	0.038	0.034	0.045	0.028	0.054
$n = 250$	0.044	0.056	0.048	0.056	0.052
$n = 500$	0.038	0.072	0.052	0.040	0.048

Table 1.2: Empirical Power of the proposed test CvM_n

	$h = 1.5 \times n^{-0.2}$						
	DGP P1	DGP P2	DGP P3	DGP P4	DGP P5	DGP P6	DGP P7
$n = 100$	0.330	0.944	0.202	0.683	0.934	0.992	0.104
$n = 250$	0.868	1.000	0.498	1.000	1.000	1.000	0.282
$n = 500$	0.992	1.000	0.880	1.000	1.000	1.000	0.503

closed form of CvM_n^* in our bootstrap. Different ρ other than $\rho = 0$ could possibly render the closed form expression difficult to derive. However, it may produce better power performance. Another example of W is the standard bivariate exponential distribution with parameter β under independence. For example, under this W, we have a closed form of CvM_n too, that is,

$$
\begin{aligned}
CvM_n = & \sum_{k=1}^{n-2} \frac{1}{(n-k)(k\pi)^2} \sum_{t=k+2}^{n} \sum_{s=k+2}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_s \hat{f}(Y_{t-1}) \hat{f}(Y_{s-1}) \\
& \times \frac{1 + \beta(Y_{t-1} - Y_{s-1})}{1 + \beta^2(Y_{t-1} - Y_{s-1})^2} \frac{1 + \beta(Y_{t-1-k} - Y_{s-1-k})}{1 + \beta^2(Y_{t-1-k} - Y_{s-1-k})^2},
\end{aligned}
$$

where β influences the power of the test. We leave the exponential case for future research.

In Table 1 we report the empirical sizes associated with the DGPs S1-S6. In Table 2 we report the empirical powers against the DGPs P1-P7. In summary, the proposed test implemented with the multiplier bootstrap procedure performs quite well in terms of empirical size and empirical power. In fact, it delivers relatively high power against many kinds of alternatives.

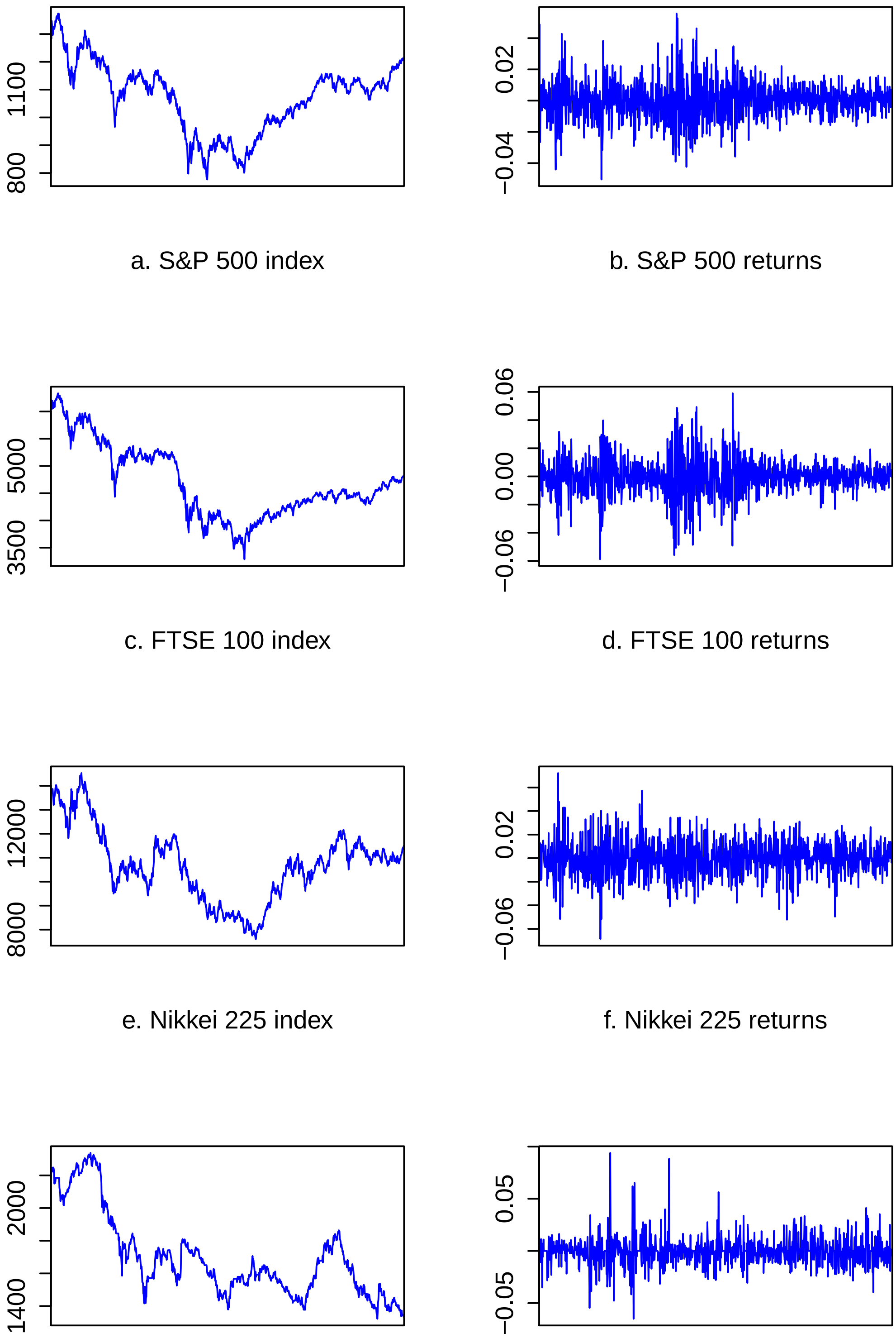
1.6 Empirical application

It is interesting to check whether some important financial time series are predictable (linearly or nonlinearly) in the mean. By contrast, Chen and Hong (2012) test the Markov property of stock prices, interest rates and foreign exchange rates. In order to confront the findings in some literature that there seems not exist nonlinear predictability in the mean of asset prices given their past history, an empirical application of our test to a number of major stock indices around the world is conducted.

In this section, we focus on four stock indices, namely S&P 500 index, FTSE 100 index, Nikkei 225 index and Shanghai A-Share index to find some evidence of nonlinear predictability of stock returns in the mean in different stock market conditions, developed and emerging markets across the world. These four indices represent very different aspects of the underlying stock returns and the corresponding market efficiency or maturity. The indices series are collected using daily data from 1 January 2001 to 31 December 2004 with a total of 1045 observations after deleting all public holidays and non-trading days. We then generate their returns series using $r_t = \log(P_t/P_{t-1}) \times 100\%$ with P_t the time series sequence for any stock index. Figure 1 plots all these indices and their returns. They exhibit the stylized facts of volatility clustering and fat tails. Standard tests indicate the returns series are stationary.

We are interested to study the nonlinear autoregressive behaviour of r_t for all four returns. The main concern here is that whether or not there exists some lagged value r_{t-k} other than r_{t-1}, which can improve the predictability of r_t in the sense of mean forecasting. Formally speaking, we aim to test $\mathrm{H}_0 : E(r_t|r_{t-1}, r_{t-k}) = E(r_t|r_{t-1})$, for any $k \geq 2$. In other words, the rejection of this hypothesis indicates that the mean of r_t given its past history could be a (nonlinear) function of lagged value r_{t-k} for some $k \geq 2$ besides r_{t-1}. The main reason we consider testing autoregressive order $p = 1$ is because it is the main interest of economic and financial modelling, e.g. AR(1)-GARCH models, widely used in financial applications. The same bandwidth $h = 1.5 \times n^{-0.2}$ as in the simulation part is adopted. For each of the four return series, we use $B = 1000$ bootstrap iterations.

The bootstrapped p-values are 0.000, 0.003, 0.012, and 0.116, respectively, for S&P 500, FTSE 100, Nikkei 225 and Shanghai A-Share. We reject the null hypothesis of no significant r_{t-k} for $k \geq 2$ at the 5% significance level for three developed stock markets, that is, it suggests that there exists some unknown form of (nonlinear) predictability in the mean for certain lagged values r_{t-k} with $k \geq 2$. We may conclude that the stock returns from these three markets do not follow an autoregressive process of order one. An unknown form of higher order (autoregressive) nonlinearity in its conditional mean is likely to be present. This phenomenon calls for more attention in modelling the (nonlinear) conditional mean before modelling higher order moments of these time series, say, conditional variance, skewness and kurtosis. For example, the widely applied GARCH-in-mean model like $r_t = \mu(r_{t-1}) + \sigma(r_{t-1})\varepsilon_t$, where $\mu(r_{t-1})$ is an known conditional mean function and $\sigma(r_{t-1})$ is the conditional variance function assumed to follow a GARCH process, may not be correctly specified and induce false conclusions. It is helpful to model the conditional mean using more lags rather than considering the first lag only. On the other hand, we can not reject the null for Shanghai A-share returns series, indicating that a (nonlinear) autoregressive order one process is sufficient to model its conditional mean.



a. S&P 500 index

b. S&P 500 returns

c. FTSE 100 index

d. FTSE 100 returns

e. Nikkei 225 index

f. Nikkei 225 returns

g. Shanghai index

h. Shanghai returns

Figure 1.1: Financial Time Series Plots

1.7 Conclusions and future research

In this paper we propose a significance testing procedure for nonparametric autoregression. Our test statistic is a functional of the generalized spectral distribution function involving smooth non-parametric autoregressive residuals. We establish asymptotic null distribution of the test statistic, prove its consistency, analyse its local power performance and provide a simple bootstrap procedure to implement our test. The practical performance of the proposed test is illustrated by means of an extensive Monte Carlo simulation and also by an investigation of the nonlinear predictability in the mean of stock returns for four indexes.

As we have mentioned in the Introduction, our test is nested in a much stronger hypothesis

$$\mathrm{H}_0^{(l)} : \mathrm{E}(Y_t|Y_{t-1},\ldots,Y_{t-p},Y_{t-p-k_1},\ldots,Y_{t-p-k_l}) = \mathrm{E}(Y_t|Y_{t-1},\ldots,Y_{t-p}) \quad \text{a.s.}$$
$$\forall k_j \geq 1, \quad j=1,\ldots,l.$$

That is, under the null hypothesis, any set of l extra lags do not affect the forecast of any nonlinear autoregression of order p. For the sake of simplicity, our main hypothesis of interest in this paper is a special case when $l = 1$. However, the methodology developed is directly applicable to test the much stronger case.

It is also theoretically and empirically relevant that we could design a test such that $l \to \infty$ at some suitable rate as the sample size $n \to \infty$ and therefore achieve a consistent test for the specification of any (nonlinear) autoregression of order p, i.e. the hypothesis (1.2). By far, to the best of our knowledge, de Jong (1996) is the only attempt towards this direction in which he considers an infinite dimensional conditioning variables. Nevertheless, his test is generally infeasible when the sample size is large, since it requires to perform a numerical integration with dimension equal to the sample size. This will be left as our future research.

1.8 Appendix A: Proofs of Main Results

Throughout the appendix, we let $C \in (0, +\infty)$ denote a generic bounded positive constant, which may be different in different places. Recall that $\varepsilon_t = Y_t - E(Y_t|Y_{t-1}, \ldots, Y_{t-p}) = Y_t - E(Y_t|\underline{Y}_{t-1})$ is the unknown regression innovations under the null. Let $f_k(\underline{Y}_{t-1}, Y_{t-p-k})$ denote the joint probability density function of $(\underline{Y}_{t-1}, Y_{t-p-k})$.

Let's denote a new stochastic process

$$S_n(\eta) = \sum_{k=1}^{n-1} \sqrt{n-k}\, \gamma_{nk}(\underline{x}, y)\, \frac{\sqrt{2}\sin k\pi\lambda}{k\pi} \tag{1.8}$$

with $\eta = (\lambda, \underline{x}, y)$, where

$$\gamma_{nk}(\underline{x}, y) = \frac{1}{n-k} \sum_{t=k+1}^{n} \varepsilon_t \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left[\mathrm{e}^{\mathrm{i}yY_{t-p-k}} f(\underline{Y}_{t-1}) - \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y} \right].$$

Notice that the quantity in brackets is similar in nature to $f(\underline{Y}_{t-1})(\mathrm{e}^{\mathrm{i}yY_{t-p-k}} - \mathrm{E}(\mathrm{e}^{\mathrm{i}yY_{t-p-k}}))$. In fact, if $\underline{Y}_{t-1} \perp Y_{t-p-k}$, the two terms are identical to each other. We can deem this quantity in brackets as a properly "weighted" version of the centered marginal characteristic function.

The following lemma states that the norm of the process $\hat{S}_n$ can be approximated by the norm of the process S_n defined in (1.8).

Lemma A.1: *Under the null and Assumptions A.1-A.5,*

$$\left| \|\hat{S}_n\|^2 - \|S_n\|^2 \right| \to_p 0.$$

Proof of Lemma A.1: Its proof consists of applying Lemma B.2 and Lemma B.3. To show the result, we first decompose $\hat{S}_n(\eta) = S_n(\eta) + R_n(\eta)$ with

$$R_n(\eta) = \sum_{k=1}^{n-1} \sqrt{n-k}\, [\hat{\gamma}_{nk}(\underline{x}, y) - \gamma_{nk}(\underline{x}, y)]\, \frac{\sqrt{2}\sin k\pi\lambda}{k\pi}.$$

So

$$\|\hat{S}_n\|^2 = \|S_n\|^2 + \|R_n\|^2 + 2\text{Re}\left\{\int_0^1 \int_{\mathbb{R}^{p+1}} S_n(\eta) R_n^c(\eta) W(d\underline{x}, dy) d\lambda\right\}, \tag{1.9}$$

where for any random element $g(\eta)$ in $L_2(\pi, \nu)$,

$$\|g\|^2 = \int_0^1 \int_{\mathbb{R}^{p+1}} |g(\eta)|^2 W(d\underline{x}, dy) d\lambda.$$

By Lemma B.3, we have $\sqrt{n-k}\left[\hat{\gamma}_{nk}(\underline{x}, y) - \gamma_{nk}(\underline{x}, y)\right] = o_p(1)$ uniformly in $(\underline{x}, y) \in \mathbb{R}^{p+1}$ under the null hypothesis. We immediately get

$$\|R_n\|^2 = o_p(1) \tag{1.10}$$

by applying Lemma B.2, where we simply take $h_{k,n}(\underline{x}, y) = \sqrt{n-k}\left[\hat{\gamma}_{nk}(\underline{x}, y) - \gamma_{nk}(\underline{x}, y)\right]$ and the two conditions are easily verified.

On the other hand, it is easy to show that under the null

$$\begin{aligned}
&\text{E}\left[\|S_n\|^2\right] = \text{E}\left\{\int_0^1 \int_{\mathbb{R}^{p+1}} |S_n(\eta)|^2 W(d\underline{x}, dy) d\lambda\right\} \\
=&\sum_{k=1}^{n-1} \frac{n-k}{(k\pi)^2} \text{E}\left\{\int_{\mathbb{R}^{p+1}} |\gamma_{nk}(\underline{x}, y)|^2 \, W(d\underline{x}, dy)\right\} \\
\leq& C \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} \\
=&O(1).
\end{aligned} \tag{1.11}$$

Combining (1.9), (1.10), (1.11) and Cauchy-Schwartz's inequality, we conclude the proof of Lemma A.1. □

Proof of Theorem 1: The weak convergence of the process $\hat{S}_n$ to S_∞ in the Hilbert space $L_2(\Pi, \nu)$ can be proved using the similar arguments in Escanciano and Velasco (2006). First, observe that by Lemma A.1, the norm of the process $\hat{S}_n$ can be approximated by the norm of S_n. So it suffices to show that the finite dimensional projections $\langle S_n(\eta), g\rangle$ are asymptotically normal $\forall g \in L_2(\Pi, \nu)$ with the appropriate asymptotic variance, and that the sequence $\{S_n(\eta)\}$ is tight, see e.g. Parthasarathy (1967). The idea is to prove these facts for a partitioned version of S_n, and then, to show that the remainder term is asymptotically negligible. We state three important steps in Theorems A.1-A.3. Combining these three theorems and Theorem 4.2 of Billingsley (1968), Theorem 1 follows.

Formally, we write for some integer L,

$$
\begin{aligned}
S_n(\eta) =& \sum_{k=1}^{L} \sqrt{n-k}\gamma_{nk}(\underline{x}, y)\frac{\sqrt{2}\sin k\pi\lambda}{k\pi} + \sum_{k=L+1}^{n-1} \sqrt{n-k}\gamma_{nk}(\underline{x}, y)\frac{\sqrt{2}\sin k\pi\lambda}{k\pi} \\
:=& S_n^L(\eta) + R_n^L(\eta).
\end{aligned}
$$

Theorem A.1: *Under the conditions of Theorem 1, for an arbitrary but fixed integer L, the finite dimensional distributions of $S_n^L(\eta)$, $\langle S_n^L(\eta), g\rangle$, converges to those of $S^L(\eta)$, $\langle S^L(\eta), g\rangle$, $\forall g \in L_2(\Pi, \nu)$, where $S^L(\eta)$ is a Gaussian process with zero mean and asymptotic projected variances*

$$
\begin{aligned}
&\sigma_{g,L}^2 := \mathit{Var}[\langle S^L, g\rangle] \\
=& \sum_{j=1}^{L}\sum_{k=1}^{L} E\left[\varepsilon_t^2 \int_{\Pi\times\Pi} g(\eta_1)g^c(\eta_2)\psi_{t-p-j}^c(\underline{x}_1, y_1)\psi_{t-p-k}(\underline{x}_2, y_2)\Phi_j(\omega_1)\Phi_k(\omega_2)\, d\nu(\eta_1)d\nu(\eta_2)\right],
\end{aligned}
$$

where $\psi_{t-p-k}(\underline{x}, y) = \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}f(\underline{Y}_{t-1}) - \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}}\int \mathrm{e}^{\mathrm{i}y\bar{y}}f_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y}$.

Theorem A.2: *Under the conditions of Theorem 1, for an arbitrary but fixed integer L, the sequence $\{S_n^L(\eta)\}$ is tight.*

Theorem A.3: *Under the conditions of Theorem 1, the process $R_n^L(\eta)$ satisfies that, for all $\epsilon > 0$,*

$$
\lim_{L\to\infty}\lim_{n\to\infty} \mathit{Pr}[\|R_n^L(\eta)\| > \epsilon] = 0.
$$

The proofs of Theorem A.1-A.3 are the same as those in Escanciano and Velasco (2006), and hence they are omitted. This finishes the proof of Theorem 1. □

Proof of Theorem 2: By Lemma B.2 and Lemma B.4, we have

$$
\begin{aligned}
\frac{1}{n}CvM_n =&\sum_{k=1}^{n-1}\frac{1}{n(k\pi)^2}\int_{\mathbb{R}^{p+1}}\left|\sqrt{n-k}\hat{\gamma}_{nk}(\underline{x},y)\right|^2 W(d\underline{x},dy)\\
=&\sum_{k=1}^{n-1}\frac{1}{n(k\pi)^2}\int_{\mathbb{R}^{p+1}}\left|\sqrt{n-k}\left[\gamma_{nk}(\underline{x},y)+\gamma^0_{nk}(\underline{x},y)\right]\right|^2 W(d\underline{x},dy)[1+o_p(1)]\\
:=&(D_{1n}+D_{2n}+D_{3n})[1+o_p(1)],
\end{aligned}
$$

where

$$
D_{1n}=\sum_{k=1}^{n-1}\frac{n-k}{n(k\pi)^2}\int_{\mathbb{R}^{p+1}}\left|\gamma_{nk}(\underline{x},y)\right|^2 W(d\underline{x},dy),
$$

$$
D_{2n}=\sum_{k=1}^{n-1}\frac{n-k}{n(k\pi)^2}\int_{\mathbb{R}^{p+1}}\left|\gamma^0_{nk}(\underline{x},y)\right|^2 W(d\underline{x},dy),
$$

and

$$
D_{3n}=2\sum_{k=1}^{n-1}\frac{(n-k)}{n(k\pi)^2}\text{Re}\left[\int_{\mathbb{R}^{p+1}}\gamma_{nk}(\underline{x},y)\gamma^{0c}_{nk}(\underline{x},y)\,W(d\underline{x},dy)\right].
$$

Under the alternative hypothesis, following the same arguments as those in Theorem 1 and continuous mapping theorem, we can conclude that nD_1 converges in distribution so that $nD_{1n}=O_p(1)$. So we have that $D_{1n}=o_p(1)$. By the fact that $\gamma^0_{nk}(\underline{x},y)\to_p\gamma_k(\underline{x},y)$ by the law of large numbers and law of iterated expectation (see Lemma B.4 for further details), we can show that

$$
D_{2n}\to_p\sum_{k=1}^{\infty}\frac{1}{(k\pi)^2}\int_{\mathbb{R}^{p+1}}\left|\gamma_k(\underline{x},y)\right|^2 W(d\underline{x},dy).
$$

and thus $D_{2n}=O_p(1)$. Finally, by Cauchy-Schwartz's inequality, $D_{1n}=o_p(1)$ and $D_{2n}=O_p(1)$ we obtain $D_{3n}=o_p(1)$. Then, $CvM_n/n=O_p(1)$. Thus, the test statistic CvM_n diverges to infinity under the alternative hypothesis as $n\to\infty$ and the test is consistent. □

Proof of Theorem 3: Denote $\varepsilon_{nt}=Y_t-m(\underline{Y}_{t-1})-n^{-1/2}g_{tk}$ under the local alternatives H_{1n} and recall $\psi_{t-p-k}(\underline{x},y)=\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}f(\underline{Y}_{t-1})-\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}}\int\mathrm{e}^{\mathrm{i}y\bar{y}}f_k(\underline{Y}_{t-1},\bar{y})\,d\bar{y}$. Let $\hat{\psi}_{t-p-k}(\underline{x},y)=\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}\hat{f}(\underline{Y}_{t-1})-\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}}\int\mathrm{e}^{\mathrm{i}y\bar{y}}\hat{f}_k(\underline{Y}_{t-1},\bar{y})\,d\bar{y}$. Then, by simple algebra

$$
\begin{aligned}
\hat{\gamma}_{nk}(\underline{x}, y) =& \frac{1}{n-k} \sum_{t=k+1}^{n} \left(Y_t - m(\underline{Y}_{t-1}) - \frac{g_{tk}}{\sqrt{n}} \right) \hat{f}(\underline{Y}_{t-1}) \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})} \\
& - \frac{1}{n-k} \sum_{t=k+1}^{n} \left(\hat{m}(\underline{Y}_{t-1}) - m(\underline{Y}_{t-1}) \right) \hat{f}(\underline{Y}_{t-1}) \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})} \\
& + \frac{1}{\sqrt{n}} \frac{1}{n-k} \sum_{t=k+1}^{n} g_{tk} \hat{f}(\underline{Y}_{t-1}) \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})} \\
=& \frac{1}{n-k} \sum_{t=k+1}^{n} \varepsilon_{nt} \hat{\psi}_{t-p-k}(\underline{x}, y) \\
& + \left(\frac{1}{n-k} \sum_{t=k+1}^{n} \left(Y_t - \hat{m}(\underline{Y}_{t-1}) \right) \hat{f}(\underline{Y}_{t-1}) \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})} \right. \\
& \left. - \frac{1}{n-k} \sum_{t=k+1}^{n} \left(Y_t - m(\underline{Y}_{t-1}) \right) \hat{\psi}_{t-p-k}(\underline{x}, y) \right) \\
& + \frac{1}{\sqrt{n}} \frac{1}{n-k} \sum_{t=k+1}^{n} g_{tk} \hat{\psi}_{t-p-k}(\underline{x}, y) \\
:=& \hat{\gamma}_{1nk}(\underline{x}, y) + \hat{\gamma}_{2nk}(\underline{x}, y) + \frac{1}{\sqrt{n}} \hat{\gamma}_{3nk}(\underline{x}, y),
\end{aligned}
$$

where $\hat{\gamma}_{2nk}(\underline{x}, y) = \hat{\gamma}_{21nk}(\underline{x}, y) - \hat{\gamma}_{22nk}(\underline{x}, y)$ with $\hat{\gamma}_{21nk}(\underline{x}, y)$ and $\hat{\gamma}_{22nk}(\underline{x}, y)$ defined accordingly. Substituting the expression into $\hat{S}_n(\eta)$, we have

$$
\begin{aligned}
\hat{S}_n(\eta) =& \sum_{k=1}^{n-1} \sqrt{n-k} \hat{\gamma}_{nk}(\underline{x}, y) \frac{\sqrt{2} \sin k\pi\lambda}{k\pi} \\
:=& \hat{S}_{1n}(\eta) + \hat{S}_{2n}(\eta) + \hat{G}_n(\eta),
\end{aligned}
$$

where

$$
\hat{S}_{jn}(\eta) = \sum_{k=1}^{n-1} \sqrt{n-k} \hat{\gamma}_{jnk}(\underline{x}, y) \frac{\sqrt{2} \sin k\pi\lambda}{k\pi}
$$

for $j = 1, 2$, and

$$
\hat{G}_n(\eta) = \sum_{k=1}^{n-1} \frac{1}{\sqrt{n}} \sqrt{n-k} \hat{\gamma}_{3nk}(\underline{x}, y) \frac{\sqrt{2} \sin k\pi\lambda}{k\pi}
$$

First of all, the density estimators $\hat{f}(\underline{Y}_{t-1})$ and $\hat{f}_k(\underline{Y}_{t-1}, Y_{t-p-k})$ are consistent estimators for $f(\underline{Y}_{t-1})$ and $f_k(\underline{Y}_{t-1}, Y_{t-p-k})$, respectively, under Assumptions A.1, A.3 and A.4, see e.g. Györfi et al. (1989), Härdle (1992) or more recently Theorem 6 in Hansen (2008). Thus, applying Lemma B.2 and following the same arguments as in the proof of Lemma A.1, one can show that $\hat{S}_{1n}(\eta)$ can be approximated by $\tilde{S}_{1n}(\eta)$ in Hilbert space $L_2(\Pi, \nu)$, where

$$\tilde{S}_{1n}(\eta) = \sum_{k=1}^{n-1} \sqrt{n-k}\tilde{\gamma}_{1nk}(\underline{x}, y)\frac{\sqrt{2}\sin k\pi\lambda}{k\pi}$$

with

$$\tilde{\gamma}_{1nk}(\underline{x}, y) = \frac{1}{n-k}\sum_{t=k+1}^{n} \varepsilon_{nt}\psi_{t-p-k}(\underline{x}, y).$$

Now, noticing that under the local alternatives, the sequence of innovations ε_{nt} is a MDS with respect to the σ-field $\mathscr{F}_t$. Following the steps in Theorem 1, it is straightforward to show that the process $\tilde{S}_{1n}$ converges weakly to S_∞. The weak convergence of $\hat{S}_{1n}$ to S_∞ follows immediately.

Applying a similar argument as in Lemma B.3, we can show

$$\begin{aligned}\sqrt{n-k}\hat{\gamma}_{2nk}(\underline{x}, y) = \Bigg\{ & \frac{1}{n}\sum_{t=1}^{n}\int e_{(\underline{x},y)}(\bar{y})g(\bar{y})\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1}-\underline{\bar{x}}}{h}\right)d\underline{\bar{x}}d\bar{y} \\ & -\mathrm{E}\left[e_{(\underline{x},y)}(Y_{t-p-k})g(Y_{t-p-k})\right]\Bigg\} + o_p(1), \end{aligned} \tag{1.12}$$

where $g(\bar{y})$ is the marginal density of Y_t and $e_{(\underline{x},y)}(\bar{y}) = \mathrm{E}[g_{tk}\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}|Y_{t-p-k} = \bar{y}]$. Under suitable smoothness assumptions on $e_{(\underline{x},y)}$, then, the first term in the right hand side of (1.12) is $o_p(1)$. We have used the fact

$$\int \frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1}-\underline{\bar{x}}}{h}\right)d\underline{\bar{x}} = 1.$$

So $\sqrt{n-k}\hat{\gamma}_{2nk}(\underline{x}, y) = o_p(1)$ uniformly in $(\underline{x}, y)$. Thus, applying Lemma B.2, $\hat{S}_{2n}(\eta)$ converges in probability to zero in $L_2(\Pi, \nu)$, that is $\|\hat{S}_{2n}\|^2 = o_p(1)$.

Finally, by the uniform ergodic theorem (UET) for stationary and ergodic time series, see e.g. Dehling and Philipp (2002, p. 4), we get

$$\sup_{(\underline{x},y)\in\mathbb{R}^{p+1}} |\hat{\gamma}_{3nk}(\underline{x},y) - L_k(\underline{x},y)| = o_p(1), \quad \forall k \geq 1,$$

where $L_k(\underline{x},y) = \mathrm{E}[g_{tk}\psi_{t-p-k}(\underline{x},y)]$. Using Lemma B.2, we can easily show that $\hat{G}_n$ converges in probability to G_∞ in $L_2(\Pi,\nu)$, where

$$G_\infty(\eta) = \sum_{k=1}^{\infty} L_k(\underline{x},y)\frac{\sqrt{2}\sin k\pi\lambda}{k\pi}$$

is a deterministic shift function. So, under the local alternatives,

$$\hat{S}_n \Rightarrow S_\infty + G_\infty$$

in Hilbert space $L_2(\Pi,\nu)$ by Slutsky's Theorem. This proves Theorem 3. □

Proof of Theorem 4: Our bootstrap is a new type called multiplier bootstrap. It is easy to prove its asymptotic validity. Define

$$S_n^{0*}(\eta) = \sum_{k=1}^{n-1} \sqrt{n-k}\,\gamma_{nk}^{0*}(\underline{x},y)\,\frac{\sqrt{2}\sin k\pi\lambda}{k\pi},$$

where

$$\gamma_{nk}^{0*}(\underline{x},y) = \frac{1}{n-k}\sum_{t=k+1}^{n} \varepsilon_t \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}}\left[\mathrm{e}^{\mathrm{i}yY_{t-p-k}}f(\underline{Y}_{t-1}) - \int \mathrm{e}^{\mathrm{i}y\bar{y}}f(\underline{Y}_{t-1},\bar{y})\,d\bar{y}\right]v_t,$$

with $\varepsilon_t = Y_t - m(\underline{Y}_{t-1})$ and $\{v_t\}_{t=1}^n$ a sequence of random variables with zero mean, unit variance and independent of the sample $\{Y_t\}_{t=1}^n$.

Following the same argument as Delgado and Gonzalez-Manteiga (2001) and Escanciano and Velasco (2006), it suffices to first establish that the norm of $\hat{S}_n^*$ and the norm of S_n^{0*} are asymptotically equivalent in the Hilbert space $L_2(\Pi,\nu)$, i.e.

$$\left|\|\hat{S}_n^*\|^2 - \|S_n^{0*}\|^2\right| \to_p 0,$$

where $\|\hat{S}_n^*\|^2 := \langle \hat{S}_n^*, \hat{S}_n^* \rangle = \int_\Pi \hat{S}_n^*(\eta)\hat{S}_n^{*c}(\eta)\, d\nu(\eta)$ is the squared norm of $\hat{S}_n^*(\eta)$ and $\|S_n^{0*}\|^2$ is defined in a similar way. Then, we show that the finite dimensional projections of $S_n^{0*}(\eta)$ converge (conditional on the original sample) to those of $\hat{S}_n(\eta)$ in probability for all samples and that the sequence $\{S_n^{0*}(\eta)\}$ is tight in probability.

Notice that we can decompose $\hat{S}_n^*(\eta)$ into the following,

$$\hat{S}_n^*(\eta) = S_n^{0*}(\eta) + R_n^*(\eta)$$

with

$$R_n^*(\eta) = \sum_{k=1}^{n-1} \sqrt{n-k}\left[\hat{\gamma}_{nk}^*(\underline{x}, y) - \gamma_{nk}^{0*}(\underline{x}, y)\right] \frac{\sqrt{2}\sin k\pi\lambda}{k\pi}.$$

As in Lemma A.1, to show that $\|\hat{S}_n^*\|^2 \to_p \|S_n^{0*}\|^2$, it suffices to show that $\|R_n^*\|^2 = o_p(1)$ and $\|S_n^{0*}\|^2 = O_p(1)$. We now write

$$\begin{aligned}
&\sqrt{n-k}\left[\hat{\gamma}_{nk}^*(\underline{x}, y) - \gamma_{nk}^{0*}(\underline{x}, y)\right] \\
=&\frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \hat{\varepsilon}_t \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left[\mathrm{e}^{\mathrm{i}yY_{t-p-k}} \hat{f}(\underline{Y}_{t-1}) - \int \mathrm{e}^{\mathrm{i}y\bar{y}} \hat{f}_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y}\right] v_t \\
&- \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \varepsilon_t \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left[\mathrm{e}^{\mathrm{i}yY_{t-p-k}} f(\underline{Y}_{t-1}) - \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y}\right] v_t \\
=&\left\{\frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})} \left[\hat{\varepsilon}_t \hat{f}(\underline{Y}_{t-1}) - \varepsilon_t f(\underline{Y}_{t-1})\right] v_t\right\} \\
&- \left\{\frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left[\hat{\varepsilon}_t \int \mathrm{e}^{\mathrm{i}y\bar{y}} \hat{f}_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y} - \varepsilon_t \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y}\right] v_t\right\} \\
=&B_{1n}(\underline{x}, y) - B_{2n}(\underline{x}, y).
\end{aligned}$$

By triangle inequality,

$$\|R_n^*\|^2 \leq 2\left(\|R_{1n}^*\|^2 + \|R_{2n}^*\|^2\right),$$

where, for $j = 1, 2$,

$$R^*_{jn}(\eta) = \sum_{k=1}^{n-1} B_{jn}(\underline{x}, y) \frac{\sqrt{2} \sin k\pi\lambda}{k\pi},$$

it suffices to show that $\|R^*_{jn}\|^2 = o_p(1)$ for $j = 1, 2$, where

$$\|R^*_{jn}\|^2 = \int_0^1 \int_{\mathbb{R}^{p+1}} \left| R^*_{jn}(\eta) \right|^2 W(d\underline{x}, dy) d\lambda.$$

To do so, we need to verify the two conditions in Lemma B.2. Denote by E^* and V^* the expectation and the variance, respectively, given the sample $\{Y_t\}_{t=1}^n$. The first condition is straightforward. We only have to check the second condition. Write

$$\begin{aligned} B_{1n}(\underline{x}, y) =& \frac{1}{n^{3/2}} \sum_{t=1}^{n} \sum_{s \neq t}^{n} \mathrm{e}^{\mathrm{i}(\underline{x}' \underline{Y}_{t-1} + y Y_{t-p-k})} \\ & \times \left[\frac{1}{h^p} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h} \right) (Y_t - Y_s) - \varepsilon_t f(\underline{Y}_{t-1}) \right] v_t \\ & \times [1 + o_p(1)]. \end{aligned}$$

Using the similar proof of Lemma B.3 and taking into account that v_t's are i.i.d. and independent of the sample $\{Y_t\}_{t=1}^n$,

$$\begin{aligned} &\sup_{(\underline{x},y)} |B_{1n}(\underline{x}, y)| \\ =& \sup_{(\underline{x},y)} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} D_h(Y_t, \underline{Y}_{t-1}) \mathrm{e}^{\mathrm{i}(\underline{x}' \underline{Y}_{t-1} + y Y_{t-p-k})} v_t \right| + o_p(1), \end{aligned} \tag{1.13}$$

where

$$D_h(y, \underline{x}) = \int (y - m(\bar{\underline{x}})) f(\bar{\underline{x}}) \frac{1}{h^p} K\left(\frac{\underline{x} - \bar{\underline{x}}}{h} \right) - (y - m(\underline{x})) f(\underline{x}).$$

Reasoning as in the proof of Lemma B.3 again, we can show that the first term in (1.13) is $o_p(1)$ under suitable smoothness assumptions. Thus, $\sup_{(\underline{x},y)} |B_{1n}(\underline{x}, y)| = o_p(1)$. We conclude

$\|R_{1n}^{*}\|^2 = o_p(1)$.

We further write

$$
\begin{aligned}
B_{2n}(\underline{x}, y) =& \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \varepsilon_t \left(\int \mathrm{e}^{\mathrm{i}y\bar{y}} \left[\hat{f}_k(\underline{Y}_{t-1}, \bar{y}) - f_k(\underline{Y}_{t-1}, \bar{y}) \right] d\bar{y} \right) v_t \\
& - \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} (\varepsilon_t - \hat{\varepsilon}_t) \left(\int \mathrm{e}^{\mathrm{i}y\bar{y}} \hat{f}_k(\underline{Y}_{t-1}, \bar{y}) \, d\bar{y} \right) v_t \\
=& B_{2n}^{1}(\underline{x}, y) - B_{2n}^{2}(\underline{x}, y).
\end{aligned}
$$

Again, by triangle inequality, we have

$$
\|R_{2n}^{*}\|^2 \leq 2 \left(\|R_{2n}^{1*}\|^2 + \|R_{2n}^{2*}\|^2 \right),
$$

where, for $j = 1, 2$,

$$
R_{2n}^{j*}(\eta) = \sum_{k=1}^{n-1} B_{2n}^{j}(\underline{x}, y) \frac{\sqrt{2} \sin k\pi\lambda}{k\pi}.
$$

So, it suffices to show that $\|R_{2n}^{j*}\|^2 = o_p(1)$ for $j = 1, 2$ with

$$
\|R_{2n}^{j*}\|^2 = \int_0^1 \int_{\mathbb{R}^{p+1}} \left| R_{2n}^{j*}(\eta) \right|^2 W(d\underline{x}, dy) d\lambda.
$$

First of all, since

$$
\begin{aligned}
B_{2n}^{1}(\underline{x}, y) =& \frac{1}{n^{3/2}} \sum_{t=1}^{n} \sum_{s \neq t}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \varepsilon_t \\
& \times \left[\frac{1}{h^p} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h} \right) \mathrm{e}^{\mathrm{i}yY_{s-p-k}} - \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{t-1}, \bar{y}) \, d\bar{y} \right] v_t \\
& \times [1 + o_p(1)],
\end{aligned}
$$

Reasoning as before and taking into account that v_t's are i.i.d. and independent of the sample, we have

$$\sup_{(\underline{x},y)} \left| B_{2n}^{1}(\underline{x}, y) \right| = o_p(1)$$

by applying the fact

$$\frac{1}{\sqrt{n}} \sum_t \int \left[\frac{1}{h^p} K\left(\frac{\underline{Y}_{t-1} - \bar{\underline{x}}}{h} \right) \mathrm{e}^{\mathrm{i}y Y_{t-p-k}} - \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y} \right] d\bar{\underline{x}} = o_p(1).$$

Thus, $\|R_{2n}^{1*}\|^2 = o_p(1)$.

Now, for term $B_{2n}^2(\underline{x}, y)$, we write

$$\begin{aligned}
B_{2n}^2(\underline{x}, y) =& \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left(\hat{m}(\underline{Y}_{t-1}) - m(\underline{Y}_{t-1}) \right) \int \mathrm{e}^{\mathrm{i}y\bar{y}} \hat{f}_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y} v_t \\
=& \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \left(\hat{m}(\underline{Y}_{t-1}) - m(\underline{Y}_{t-1}) \right) \int \mathrm{e}^{\mathrm{i}y\bar{y}} \frac{1}{f_k(\underline{Y}_{t-1}, \bar{y})} \\
& \left(f_k(\underline{Y}_{t-1}, \bar{y}) - \hat{f}_k(\underline{Y}_{t-1}, \bar{y}) \right) \hat{f}_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y} v_t \\
& + \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} (\hat{m}(\underline{Y}_{t-1}) - m(\underline{Y}_{t-1})) \int \mathrm{e}^{\mathrm{i}y\bar{y}} \frac{\hat{f}_k^2(\underline{Y}_{t-1}, \bar{y})}{f_k(\underline{Y}_{t-1}, \bar{y})}\, d\bar{y} v_t \\
=& B_{2n}^{21}(\underline{x}, y) + B_{2n}^{22}(\underline{x}, y).
\end{aligned}$$

Since

$$\mathrm{E}\left[\left(\hat{m}(\underline{Y}_{t-1}) - m(\underline{Y}_{t-1}) \right)^2 \hat{f}_k^2(\underline{Y}_{t-1}, Y_{t-p-k}) \right] = O_p\left(\frac{1}{nh^p} \right)$$

and

$$\mathrm{E}\left[\left(f_k(\underline{Y}_{t-1}, Y_{t-p-k}) - \hat{f}_k(\underline{Y}_{t-1}, Y_{t-p-k})) \right)^2 \right] = O_p\left(\frac{1}{nh^p} \right),$$

by applying the Cauchy-Schwarz and the Markov inequalities, we have

$$\begin{aligned}
&\sup_{(\underline{x},y)} \left|B_{2n}^{21}(\underline{x},y)\right| \\
=&O_p\left(n^{1/2}\left\{\mathrm{E}\left[\left(\hat{m}(\underline{Y}_{t-1})-m(\underline{Y}_{t-1})\right)^2 \hat{f}_k^2(\underline{Y}_{t-1},Y_{t-p-k})\right]\right.\right. \\
&\left.\left.\times\ \mathrm{E}\left[\left(f_k(\underline{Y}_{t-1},Y_{t-p-k})-\hat{f}_k(\underline{Y}_{t-1},Y_{t-p-k}))\right)^2\right]\right\}^{1/2}\right) \\
=&O_p\left(n^{1/2}\left(\frac{1}{n^{1/2}h^{p/2}}\right)\left(\frac{1}{n^{1/2}h^{p/2}}\right)\right) \\
=&o_p(1),
\end{aligned}$$

where we have used $nh^{2p}\to\infty$ in Assumption A.4'. So, by Lemma B.2, $\|R_{2n}^{21*}\|^2 = o_p(1)$.

Following the same arguments as above, we can show that $\sup_{(\underline{x},y)}|B_{2n}^{22}(\underline{x},y)| = o_p(1)$, so that $\|R_{2n}^{22*}\|^2 = o_p(1)$ too. Thus, $\|R_{2n}^{2*}\|^2 = o_p(1)$.

Combining all the results, we have shown that $\|R_n^*\|^2 = o_p(1)$ holds. It is straightforward to see that

$$\begin{aligned}
&\mathrm{E}^*\left[\|S_n^{0*}\|^2\right] = \mathrm{E}^*\left\{\int_0^1\int_{\mathbb{R}^{p+1}}\left|S_n^{0*}(\eta)\right|^2 W(d\underline{x},dy)d\lambda\right\} \\
=&\sum_{k=1}^{n-1}\frac{n-k}{(k\pi)^2}\mathrm{E}^*\left\{\int_{\mathbb{R}^{p+1}}\left|\gamma_{nk}^{0*}(\underline{x},y)\right|^2 W(d\underline{x},dy)\right\} \\
\leq&C\sum_{k=1}^{\infty}\frac{1}{(k\pi)^2} \\
=&O(1).
\end{aligned}$$

Thus we have proved $\|\hat{S}_n^*\|^2 \to_p \|S_n^{0*}\|^2$.

The proofs that (1) finite dimensional projections of $\langle S_n^{0*}(\eta), g\rangle$ (conditional on the original sample) are asymptotically normal $\forall g\in L_2(\Pi,\nu)$ with the same asymptotic variance of $\hat{S}(\eta)$ in probability for all samples, and that (2) the sequence $\{S_n^{0*}(\eta)\}$ is tight in probability are based on the similar arguments to those used in the poof of Theorem 1. We complete the proof of Theorem 4. □

1.9 Appendix B: Lemmas

We first state a lemma by Yoshihara (1976), which is useful to prove Lemma B.3. Its proof is omitted.

Lemma B.1: *Let $\{U_t\}$, $t = 1, \dots, T$, be a d-dimensional strictly stationary absolutely regular stochastic process with mixing coefficient $\beta(k)$. Let $t_1 < \cdots < t_k$ be integers. Let $F(i, j)$, $i \leq j$, be the distribution function of $U_{t_i}, \dots, U_{t_j}$. Let $h(\phi) := h(\phi_1, \dots, \phi_k)$ be a Borel-measurable function on $\mathbb{R}^{kd}$ such that for some $\delta > 0$ and given j, $M \equiv \int |h(\phi)|^{1+\delta} \, dF(1, j)dF(j + 1, k)$ exists. Then,*

$$\left| \int h(\phi) \, dF(1, k) - \int h(\phi) \, dF(1, j)dF(j + 1, k) \right| \leq 4M^{1/(1+\delta)} \beta(l)^{\delta/(1+\delta)},$$

where $l = t_{j+1} - t_j$.

We then adapt Lemma 1 of Escanciano and Velasco (2006) to our context, which is needed to prove the main asymptotic results. Its proof is similar to that in Escanciano and Velasco (2006) and hence it is omitted.

Lemma B.2: *For $\eta = (\lambda, \underline{x}, y)$, suppose we have a random element in $L_2(\Pi, \nu)$ of the form*

$$h_n(\eta) = \sum_{k=1}^{n-1} h_{k,n}(\underline{x}, y) \frac{\sqrt{2} \sin k\pi\lambda}{k\pi}$$

If Assumption A.5 and the following two conditions hold,

(i) $\int_{\mathbb{R}^{p+1}} E|h_{k,n}(\underline{x}, y)|^2 W(d\underline{x}, dy) < C$ *uniformly in* $1 \leq k < n$,

(ii) $\sup_{(\underline{x},y)\in[-a,a]^{p+1}} |h_{k,n}(\underline{x}, y)| = o_p(1)$, $\forall 1 \leq k < n$, $\forall a > 0$,

then, $h_n(\eta)$ converges in probability to zero in $L_2(\Pi, \nu)$, i.e. $\|h_n\|^2 = o_p(1)$.

The next lemma establishes an asymptotic equivalence between $\hat{\gamma}_{nk}(\underline{x}, y)$ and $\gamma_{nk}(\underline{x}, y)$ uniformly in $(\underline{x}, y) \in \mathbb{R}^{p+1}$ under the null hypothesis.

Lemma B.3: *Under Assumptions A.1-A.4 and the null hypothesis, we have for all k, $1 \leq k <$*

n,

$$\sup_{(\underline{x},y)\in\mathbb{R}^{p+1}}\left|\sqrt{n-k}\left[\hat{\gamma}_{nk}(\underline{x},y)-\gamma_{nk}(\underline{x},y)\right]\right|=o_p(1).$$

Proof of Lemma B.3: First, $\hat{\gamma}_{nk}(\underline{x},y)$ can be written as

$$\begin{aligned}\hat{\gamma}_{nk}(\underline{x},y) &= \frac{1}{n-k}\sum_{t=k+1}^{n}\hat{\varepsilon}_t\hat{f}(\underline{Y}_{t-1})\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}\\ &= \frac{1}{(n-k)(n-1)}\sum_{t=k+1}^{n}\sum_{s=1,s\neq t}^{n}\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1}-\underline{Y}_{s-1}}{h}\right)(Y_t-Y_s)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}.\end{aligned}$$

We decompose

$$\begin{aligned}\hat{\gamma}_{nk}(\underline{x},y) =& \frac{1}{(n-k)(n-1)}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1}-\underline{Y}_{s-1}}{h}\right)(Y_t-Y_s)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}\\ &-\frac{1}{(n-k)(n-1)}\sum_{t=1}^{k}\sum_{s=1,s\neq t}^{n}\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1}-\underline{Y}_{s-1}}{h}\right)(Y_t-Y_s)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}\end{aligned}$$

Let

$$\hat{\gamma}^1_{nk}(\underline{x},y)=\frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1}-\underline{Y}_{s-1}}{h}\right)(Y_t-Y_s)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}$$

and

$$\hat{\gamma}^2_{nk}(\underline{x},y)=\frac{1}{n(n-1)}\sum_{t=1}^{k}\sum_{s=1,s\neq t}^{n}\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1}-\underline{Y}_{s-1}}{h}\right)(Y_t-Y_s)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}.$$

Since $\hat{\gamma}_{nk}(\underline{x},y)=n/(n-k)\left(\hat{\gamma}^1_{nk}(\underline{x},y)+\hat{\gamma}^2_{nk}(\underline{x},y)\right)$ for all k, $1\leq k<n$, it suffices to study the asymptotic behaviour for $\hat{\gamma}^1_{nk}(\underline{x},y)$ and $\hat{\gamma}^2_{nk}(\underline{x},y)$, respectively. We shall resort to the theory of U-statistics. We start with $\hat{\gamma}^1_{nk}(\underline{x},y)$, and then we show that the second term $\hat{\gamma}^2_{nk}(\underline{x},y)$ is asymptotically of order $o_p(n^{-1/2})$ under the assumptions stated in section 3.

We have to show that for any fixed $(\underline{x}, y) \in \mathbb{R}^{p+1}$,

$$\mathrm{E}\left[\left|\sqrt{n-k}\left[\hat{\gamma}_{nk}(\underline{x}, y) - \gamma_{nk}(\underline{x}, y)\right]\right|^2\right] = o(1). \tag{1.14}$$

Now, for all k, $1 \leq k < n$, put $\mathcal{W}_{kt} = (Y_t, \underline{Y}_{t-1}, Y_{t-p-k})$. We introduce

$$\begin{aligned} U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}) =& \frac{1}{2}\left[\frac{1}{h^p} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h}\right)(Y_t - Y_s)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})}\right. \\ & \left. + \frac{1}{h^p} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h}\right)(Y_s - Y_t)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{s-1} + yY_{s-p-k})}\right] \end{aligned}$$

For notational simplicity, we have suppressed the dependence on $(\underline{x}, y)$ of $U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})$. So, $\hat{\gamma}^1_{nk}(\underline{x}, y)$ can be written as a U-statistic of the following form,

$$\sqrt{n}\hat{\gamma}^1_{nk}(\underline{x}, y) = \sqrt{n}\frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}),$$

However, it is important to emphasize that this U-statistic is not a standard one since the kernel $U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})$ depends on the sample size n implicitly through the bandwidth h. The most powerful tool used to study the asymptotic behaviour of a U-statistic is the Hoeffding's decomposition, see Hoeffding (1948). It has been shown, both in the i.i.d. and the weak dependent contexts, that a U-statistic can be decomposed into several terms having different orders of magnitudes, and that in general only the one term with the leading order will determine the asymptotic behaviour of the U-statistic, see Serfling (1980) and Borovkova et al. (2001) for further details.

So, according to U-statistic theory, we calculate the projection term of $U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})$. Denote $\mathcal{W}_k = (Y, \underline{Y}_{-1}, Y_{-p-k})$ and $\varepsilon = Y - m(\underline{Y}_{-1})$. We obtain

$$
\begin{aligned}
u_{k1}(\mathcal{W}_k) =& E[U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})|\mathcal{W}_{ks} = \mathcal{W}_k] \\
=& \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right) m(\underline{Y}_{t-1})\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}\right] \\
& - \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right) \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1}+yY_{t-p-k})}\right] Y \\
& + \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right)\right] \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{-1}+yY_{-p-k})}Y \\
& - \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right) m(\underline{Y}_{t-1})\right] \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{-1}+yY_{-p-k})} \\
:=& A_{11} + A_{12} + A_{13} + A_{14},
\end{aligned}
$$

where we have used the fact the under the null hypothesis, $E(Y_t|\underline{Y}_{t-1}, Y_{t-p-k}) = m(\underline{Y}_{t-1})$ a.s. for all $1 \leq k < n$.

Let $f_k(\underline{Y}_{t-1}|Y_{t-p-k})$ denote the conditional density function of $\underline{Y}_{t-1}$ given Y_{t-p-k} and $f_k(\underline{Y}_{t-1}, Y_{t-p-k})$ the joint density function of $(\underline{Y}_{t-1}, Y_{t-p-k})$. And let $g(Y_t)$ denote the marginal density of Y_t. Now let's calculate A_{1j} for $j = 1, \ldots, 4$. First of all, by using the change of variable technique and Taylor expansion of order l around Y_{-1}, we can show that

$$
\begin{aligned}
A_{11} =& \frac{1}{2}\int \mathrm{e}^{\mathrm{i}y\bar{y}} \left[\int \frac{1}{h^p}K\left(\frac{z - \underline{Y}_{-1}}{h}\right) m(z)\mathrm{e}^{\mathrm{i}\underline{x}'z} f_k(z|\bar{y})\, dz\right] g(\bar{y})d\bar{y} \\
=& \frac{1}{2}\int \mathrm{e}^{\mathrm{i}yY_{t-p-k}} \left[\int K(u)m(\underline{Y}_{-1} + hu)\mathrm{e}^{\mathrm{i}\underline{x}'(\underline{Y}_{-1}+hu)} f_k(\underline{Y}_{-1} + hu|\bar{y})\, du\right] g(\bar{y})d\bar{y} \\
=& \frac{1}{2}\int \mathrm{e}^{\mathrm{i}y\bar{y}} \left\{m(\underline{Y}_{-1})f_k(\underline{Y}_{-1}|\bar{y}) + \left[\frac{1}{2}m(\underline{Y}_{-1})f_k^{''}(\underline{Y}_{-1}|\bar{y}) + m'(\underline{Y}_{-1})f_k'(\underline{Y}_{-1}|\bar{y})\right.\right. \\
& \left.\left. + \frac{1}{2}m^{''}(\underline{Y}_{-1})f_k(\underline{Y}_{-1}|\bar{y})\right] h^2 \int u^2K(u)\, du + O(h^l)\right\} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{-1}}\, g(\bar{y})d\bar{y} + o(h^l) \\
=& \frac{1}{2}m(\underline{Y}_{-1})\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{-1}} \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{-1}|\bar{y})\, g(\bar{y})d\bar{y} + O(h^l) \\
=& \frac{1}{2}m(\underline{Y}_{-1})\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{-1}} \int \mathrm{e}^{\mathrm{i}y\bar{y}} f_k(\underline{Y}_{-1}, \bar{y})\, d\bar{y} + O(h^l),
\end{aligned}
$$

where in the calculation we have exploited the higher order kernel properties for K in Assumption A.3. Similarly, we can calculate

$$A_{12} = -\frac{1}{2} Y e^{i\underline{x}'\underline{Y}_{-1}} \int e^{iy\bar{y}} f_k(\underline{Y}_{-1}, \bar{y})\, d\bar{y} + O(h^l),$$

$$A_{13} = \frac{1}{2} Y f(\underline{Y}_{-1}) e^{i(\underline{x}'\underline{Y}_{-1} + yY_{-p-k})} + O(h^l),$$

and

$$A_{14} = -\frac{1}{2} m(\underline{Y}_{-1}) f(\underline{Y}_{-1}) e^{i(\underline{x}'\underline{Y}_{-1} + yY_{-p-k})} + O(h^l).$$

Thus, we obtain

$$\begin{aligned} & u_{k1}(\mathcal{W}_k) \\ =& -\frac{1}{2}\varepsilon e^{i\underline{x}'\underline{Y}_{-1}} \int e^{iy\bar{y}} f_k(\underline{Y}_{-1}, \bar{y})\, d\bar{y} + \frac{1}{2}\varepsilon e^{i\underline{x}'\underline{Y}_{-1}} f(\underline{Y}_{-1}) e^{iyY_{-p-k}} + O(h^l) \\ =& \frac{1}{2}\varepsilon e^{i\underline{x}'\underline{Y}_{-1}} \left\{ e^{iyY_{-p-k}} f(\underline{Y}_{-1}) - \int e^{iy\bar{y}} f_k(\underline{Y}_{-1}, \bar{y})\, d\bar{y} \right\} + O(h^l). \end{aligned}$$

Apparently, once we get the projection term like the one in the above equation, we obtain that $u_{k1}(\mathcal{W}_{kt}) \approx \varepsilon_t e^{i\underline{x}'\underline{Y}_{t-1}} \{ e^{iyY_{t-p-k}} f(\underline{Y}_{t-1}) - \int e^{iy\bar{y}} f_k(\underline{Y}_{t-1}, \bar{y})\, d\bar{y} \}/2 := \xi_t/2$. Let's denote $\phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}) = U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}) - u_{k1}(\mathcal{W}_{kt}) - u_{k1}(\mathcal{W}_{ks})$. Therefore, by Hoeffding's decomposition, $\sqrt{n}\hat{\gamma}^1_{nk}(x, y)$ could be expressed as

$$\begin{aligned} & \sqrt{n}\hat{\gamma}^1_{nk}(x, y) \\ =& \frac{2}{\sqrt{n}} \sum_{t=1}^{n} u_{k1}(\mathcal{W}_{kt}) + \sqrt{n}\frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} \phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}) \\ =& \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \xi_t + \sqrt{n}\frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} \phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}) + O_p((nh^{2l})^{1/2}). \end{aligned}$$

We have seen that $n^{-1/2}\sum_{t=1}^{n} \xi_t = \sqrt{n}\gamma_{nk}(x, y) + o_p(1)$. By Assumption A.4, we have $O_p((nh^{2l})^{1/2}) = o_p(1)$. The remaining proof about the expansion of $\sqrt{n}\hat{\gamma}_{n1k}(x, y)$ consists of showing the following result, that is, for any fixed $(\underline{x}, y) \in \mathbb{R}^{p+1}$,

$$\sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}) = o_p(1).$$

To this end, we shall prove that for any fixed $(\underline{x}, y) \in \mathbb{R}^{p+1}$,

$$\mathrm{E}\left[\left|\sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})\right|^2\right] = o(1). \tag{1.15}$$

Combining the results (1.14) and (1.15) and applying Chebyshev's inequality, we conclude that

$$\left|\sqrt{n-k}\left[\hat{\gamma}_{nk}(\underline{x}, y) - \gamma_{nk}(\underline{x}, y)\right]\right| = o_p(1), \quad \forall(\underline{x}, y) \in \mathbb{R}^{p+1}.$$

Now, let's show (1.15). The basic idea to prove it is to find a suitable upper bound for the expectation as demonstrated in Robinson (1989). One of the most useful tools used to derive this bound is one fundamental lemma for absolutely regular (ARE) processes established in Yoshihara (1976). To this end, we first need to decompose the expectation into the following three terms of different natures and show that they are all asymptotically negligible. Formally, we shall show that

$(i1)$ the double summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})^2\right\} = o(1), \tag{1.16}$$

$(i2)$ the triple summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-2}\sum_{s=t+1}^{n-1}\sum_{u=s+1}^{n}\phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})\phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ku})\right\} = o(1), \tag{1.17}$$

and $(i3)$ the quadruple summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-3}\sum_{s=t+1}^{n-2}\sum_{u=s+1}^{n-1}\sum_{v=u+1}^{n}\phi_k(\mathcal{W}_{kt},\mathcal{W}_{ks})\phi_k(\mathcal{W}_{ku},\mathcal{W}_{kv})\right\}=o(1), \tag{1.18}$$

respectively. Moreover, to prove (1.18), as in Yoshihara (1976), we need to consider the following two different cases separately, case $(i3.1)$ with $t < s < u < v$ and $s - t > v - u$, and case $(i3.2)$ $t < s < u < v$ and $s - t \le v - u$. Since both cases are of the same nature and can be handled by similar techniques, we only focus on the first case $(i3.1)$, that is, we shall show

$$\mathrm{E}\left\{\frac{1}{n^3}\sum\sum\sum\sum_{t<s<u<v,s-t>v-u}\phi_k(\mathcal{W}_{kt},\mathcal{W}_{ks})\phi(\mathcal{W}_{ku},\mathcal{W}_{kv})\right\}=o(1). \tag{1.19}$$

It suffices to prove that the three terms in (1.16), (1.17), and (1.19) are all asymptotically negligible. Let $F_k(w_{kt})$ be the marginal distribution function of $\mathcal{W}_{kt}$ and $F_{kt_1,\ldots,kt_L}(w_{kt_1},\ldots,w_{kt_L})$ be the joint distribution function of $\mathcal{W}_{kt_1},\ldots,\mathcal{W}_{kt_L}$. Henceforth, for notational economy, the dependence of $U_k(w_{kt_1},w_{kt_1})$, $\phi_k(w_{kt_1},w_{kt_2})$, $F_k(w_{kt})$ and $F_{kt_1,\ldots,kt_L}(w_{kt_1},\ldots,w_{kt_L})$ on k will be suppressed.

For term (1.16), by Assumptions A.1 and A.2, we have

$$\begin{aligned}
&\int|\phi(w_{t_1},w_{t_2})|^{2+\delta}\,dF(w_{t_1})dF(w_{t_2})\\
\le&C\int|U(w_{t_1},w_{t_2})|^{2+\delta}\,dF(w_{t_1})dF(w_{t_2})+C\\
\le&C\int\frac{1}{h^{p(2+\delta)}}\left|K\left(\frac{\underline{Y}_{t_1-1}-\underline{Y}_{t_2-1}}{h}\right)\right|^{2+\delta}\left|m(\underline{Y}_{t_1-1})-Y_{t_2}\right|^{2+\delta}\left|\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t_1-1}+yY_{t_1-p-k})}\right|^{2+\delta}\\
&dF(w_{t_1})dF(w_{t_2})+C\\
\le&C\{h^{-p(1+\delta)}+1\}:=C_{1h}.
\end{aligned} \tag{1.20}$$

The first inequality follows since $\mathrm{E}|u_{k1}(\mathcal{W}_{kt})|^{2+\delta}$ is of smaller order than $\mathrm{E}|U_k(\mathcal{W}_{kt},\mathcal{W}_{ks})|^{2+\delta}$ and the last equality is because of change of variables and Assumption A.1. Noticing that when

$\delta = 0$, the inequality in (1.20) holds too, so that

$$\int \phi(w_{t_1}, w_{t_2})^2 \, dF(w_{t_1}) dF(w_{t_2}) \leq C(h^{-p} + 1) = O(h^{-p}). \tag{1.21}$$

Further, by Lemma B.1, we have

$$\begin{aligned}
&\left| \int \phi(w_{t_1}, w_{t_2})^2 \, dF_{t_1,t_2}(w_{t_1}, w_{t_2}) - \int \phi(w_{t_1}, w_{t_2})^2 \, dF(w_{t_1}) dF(w_{t_2}) \right| \\
\leq & 4C_{1h}^{2/(2+\delta)} \beta(s-t)^{\delta/(2+\delta)} \\
\leq & Ch^{-2p(1+\delta)/(2+\delta)} (s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha},
\end{aligned} \tag{1.22}$$

where we have used the β-mixing condition in Assumption A.1. Combing inequalities (1.21) and (1.22), we have

$$\begin{aligned}
\mathrm{E}\{\phi(\mathcal{W}_t, \mathcal{W}_s)^2\} = & \int \phi(w_{t_1}, w_{t_2})^2 \, dF_{t_1,t_2}(w_{t_1}, w_{t_2}) \\
& \leq C \left\{ h^{-2p(1+\delta)/(2+\delta)} (s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha} + h^{-p} \right\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathrm{E} \left\{ \frac{1}{n^3} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} \phi(\mathcal{W}_t, \mathcal{W}_s)^2 \right\} \\
= & \frac{1}{n^3} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} \mathrm{E}\{\phi(\mathcal{W}_t, \mathcal{W}_s)^2\} \\
\leq & C \left(\frac{1}{n^3} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} \left\{ h^{-2p(1+\delta)/(2+\delta)} (s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha} + h^{-p} \right\} \right) \\
\leq & C \left(\frac{1}{n^2 h^{2p(1+\delta)/(2+\delta)}} + \frac{1}{nh^p} \right) \\
= & C \left(\frac{h^{2p/(2+\delta)}}{(nh^p)^2} + \frac{1}{nh^p} \right) = o(1),
\end{aligned}$$

by applying $h \to 0$ and $nh^p \to \infty$ in Assumption A.4. The last inequality follows due to the fact that $\sum_{t=1}^{n-1} \sum_{s=t+1}^{n} (s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha} = O(n)$ by noticing that $\delta > \alpha$ and $(2+\alpha)\delta/(2+$

$\delta)\alpha > 1$.

For term (1.17), an upper bound of $\int |\phi(w_{t_1}, w_{t_2})\phi(w_{t_1}, w_{t_3})|^{1+\delta/2}\, dF_{t_1,t_2}(w_{t_1}, w_{t_2}) dF_{t_3}(w_{t_3})$ is computed in a similar way,

$$\begin{aligned}
&\int |\phi(w_{t_1}, w_{t_2})\phi(w_{t_1}, w_{t_3})|^{1+\delta/2}\, dF_{t_1,t_2}(w_{t_1}, w_{t_2}) dF_{t_3}(w_{t_3}) \\
=&\int |\phi(w_{t_1}, w_{t_2})|^{1+\delta/2} \left[\int |\phi(w_{t_1}, w_{t_3})|^{1+\delta/2}\, dF_{t_3}(w_{t_3})\right] dF_{t_1,t_2}(w_{t_1}, w_{t_2}) \\
\leq& C\{h^{-p\delta} + 1\} := C_{2h}.
\end{aligned}$$

Applying Lemma B.1, we have

$$\begin{aligned}
&\bigg|\int \phi(w_{t_1}, w_{t_2})\phi(w_{t_1}, w_{t_3})\, dF_{t_1,t_2,t_3}(w_{t_1}, w_{t_2}, w_{t_3}) \\
&- \int \phi(w_{t_1}, w_{t_2})\phi(w_{t_1}, w_{t_3})\, dF_{t_1,t_2}(w_{t_1}, w_{t_2}) dF_{t_3}(w_{t_3})\bigg| \\
\leq& 4C_{2h}^{2/(2+\delta)} \beta(s-t)^{\delta/(2+\delta)} \\
\leq& Ch^{-2p\delta/(2+\delta)}(s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha}. \qquad (1.23)
\end{aligned}$$

Noticing that $\int \phi(w_{t_1}, w_{t_3})\, dF_{t_3}(w_{t_3}) = 0$ by construction, based on (1.23), we immediately have

$$\begin{aligned}
|\mathrm{E}\{\phi(\mathcal{W}_t, \mathcal{W}_s)\phi(\mathcal{W}_t, \mathcal{W}_u)\}| &= \left|\int \phi(w_{t_1}, w_{t_2})\phi(w_{t_1}, w_{t_3})\, dF_{t_1,t_2,t_3}(w_{t_1}, w_{t_2}, w_{t_3})\right| \\
&\leq Ch^{-2p\delta/(2+\delta)}(s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha}.
\end{aligned}$$

We then get

$$
\begin{aligned}
&\left|\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-2}\sum_{s=t+1}^{n-1}\sum_{u=s+1}^{n}\phi(\mathcal{W}_t,\mathcal{W}_s)\phi(\mathcal{W}_t,\mathcal{W}_u)\right\}\right| \\
\leq&\frac{1}{n^3}\sum_{t=1}^{n-2}\sum_{s=t+1}^{n-1}\sum_{u=s+1}^{n}\left|\mathrm{E}\left\{\phi(\mathcal{W}_t,\mathcal{W}_s)\phi(\mathcal{W}_t,\mathcal{W}_u)\right\}\right| \\
\leq&C\left(\frac{1}{n^3}\sum_{t=1}^{n-2}\sum_{s=t+1}^{n-1}\sum_{u=s+1}^{n}\left\{h^{-2p\delta/(2+\delta)}(s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha}\right\}\right) \\
\leq&C\frac{1}{nh^{2p\delta/(2+\delta)}} \\
=&C\frac{h^{p(2-\delta)/(2+\delta)}}{nh^p}=o(1)
\end{aligned}
$$

by $h\to 0$, $nh^p\to\infty$ and the fact that $\delta<2$.

Finally, for the case $(i3.1)$ in (1.19), we obtain an upper bound,

$$
\begin{aligned}
&\int|\phi(w_{t_1},w_{t_2})\phi(w_{t_3},w_{t_4})|^{1+\delta/2}\,dF_{t_1,t_2,t_3}(w_{t_1},w_{t_2},w_{t_3})dF_{t_4}(w_{t_4}) \\
\leq&\int|\phi(w_{t_1},w_{t_2})|^{1+\delta/2}\left[\int|\phi(w_{t_3},w_{t_4})|^{1+\delta/2}\,dF_{t_4}(w_{t_4})\right]\,dF_{t_1,t_2,t_3}(w_{t_1},w_{t_2},w_{t_3}) \\
\leq&C\{h^{-p\delta}+1\}:=C_{3h}.
\end{aligned}
$$

Applying Lemma B.1 again,

$$
\begin{aligned}
&\left|\int\phi(w_{t_1},w_{t_2})\phi(w_{t_3},w_{t_4})\,dF_{t_1,t_2,t_3,t_4}(w_{t_1},w_{t_2},w_{t_3},w_{t_4})\right. \\
&\left.-\int\phi(w_{t_1},w_{t_2})\phi(w_{t_3},w_{t_4})\,dF_{t_1,t_2,t_3}(w_{t_1},w_{t_2},w_{t_3})dF_{t_4}(w_{t_4})\right| \\
\leq&4C_{3h}^{2/(2+\delta)}\beta(s-t)^{\delta/(2+\delta)} \\
\leq&Ch^{-2p\delta/(2+\delta)}(s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha}.
\end{aligned}
$$

Noticing that we still have $\int \phi(w_{t_3}, w_{t_4})\, dF_{t_4}(w_{t_4}) = 0$ by construction. We obtain

$$
\begin{aligned}
|\mathrm{E}\{\phi(\mathcal{W}_t, \mathcal{W}_s)\phi(\mathcal{W}_u, \mathcal{W}_v)\}| &= \left| \int \phi(w_{t_1}, w_{t_2})\phi(w_{t_3}, w_{t_4})\, dF_{t_1,t_2,t_3,t_4}(w_{t_1}, w_{t_2}, w_{t_3}, w_{t_4}) \right| \\
&\leq Ch^{-2p\delta/(2+\delta)}(s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha}.
\end{aligned}
$$

So, we get

$$
\begin{aligned}
&\left| \mathrm{E}\left\{ \frac{1}{n^3} \sum\sum\sum\sum_{t<s<u<v, s-t>v-u} \phi(\mathcal{W}_t, \mathcal{W}_s)\phi(\mathcal{W}_t, \mathcal{W}_u) \right\} \right| \\
\leq & \frac{1}{n^3} \sum\sum\sum\sum_{t<s<u<v, s-t>v-u} |\mathrm{E}\left\{\phi(\mathcal{W}_t, \mathcal{W}_s)\phi(\mathcal{W}_u, \mathcal{W}_v)\right\}| \\
\leq & C\left(\frac{1}{n^3} \sum\sum\sum\sum_{t<s<u<v, s-t>v-u} \left\{ h^{-2p\delta/(2+\delta)}(s-t)^{-(2+\alpha)\delta/(2+\delta)\alpha} \right\} \right) \\
\leq & C \frac{1}{n^{2(\delta-\alpha)/\alpha(2+\delta)} h^{2p\delta/(2+\delta)}} \\
= & C \frac{h^{2p(\delta-\alpha-\delta\alpha)/\alpha(2+\delta)}}{(nh^p)^{2(\delta-\alpha)/\alpha(2+\delta)}} = o(1),
\end{aligned}
$$

where the last equality follows by assumptions $nh^p \to \infty$ and $\delta > \alpha/(1-\alpha)$ so that $\delta > \alpha$ and $\delta - \alpha - \delta\alpha > 0$. We have proved (1.15).

We prove $\sqrt{n}\hat{\gamma}^2_{nk}(\underline{x}, y)$ is negligible. Observe that

$$
\begin{aligned}
\hat{\gamma}^2_{nk}(\underline{x}, y) = & \left[\frac{1}{n(n-1)} \sum_{t=1}^{k} \sum_{s=1, s\neq t}^{k} \frac{1}{h^p} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h} \right) (Y_t - Y_s) \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})} \right] \\
& + \left[\frac{1}{n(n-1)} \sum_{t=1}^{k} \sum_{s=k+1}^{n} \frac{1}{h^p} K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{s-1}}{h} \right) (Y_t - Y_s) \mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})} \right] \\
:= & \hat{\gamma}^{21}_{nk}(\underline{x}, y) + \hat{\gamma}^{22}_{nk}(\underline{x}, y).
\end{aligned}
$$

Since $\mathrm{E}\{\sqrt{n}\hat{\gamma}^2_{nk}(\underline{x}, y)\}^2 \leq 2(\mathrm{E}\{\sqrt{n}\hat{\gamma}^{21}_{nk}(\underline{x}, y)\}^2 + E\{\sqrt{n}\hat{\gamma}^{22}_{nk}(\underline{x}, y)\}^2)$, we shall now prove that $E\{\sqrt{n}\hat{\gamma}^{21}_{nk}(\underline{x}, y)\}^2 \to 0$ and $E\{\sqrt{n}\hat{\gamma}^{22}_{nk}(\underline{x}, y)\}^2 \to 0$, respectively. Then, by Chebyshev's inequality, we finish showing $\sqrt{n}\hat{\gamma}^{21}_{nk}(\underline{x}, y) = o_p(1)$ and $\sqrt{n}\hat{\gamma}^{22}_{nk}(\underline{x}, y) = o_p(1)$. First noting that

$\sqrt{n}\hat{\gamma}_{nk}^{21}(\underline{x}, y)$ can be written as

$$\sqrt{n}\hat{\gamma}_{nk}^{21}(\underline{x}, y) = \frac{2}{\sqrt{n}(n-1)} \sum_{t=1}^{k-1} \sum_{s=t+1}^{k} U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks}),$$

As in the proof for $\hat{\gamma}_{nk}^{1}(\underline{x}, y)$, the expectation $\mathrm{E}\{\sqrt{n}\hat{\gamma}_{nk}^{21}(x, y)\}^2$ consists of three terms of different nature. For example, the double summation term,

$$\mathrm{E}\left\{ \frac{1}{n^3} \sum_{t=1}^{k-1} \sum_{s=t+1}^{k} U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})^2 \right\} \tag{1.24}$$

converges to zero since it is always bounded by $C\left(\sum_{t=1}^{n-1}\sum_{s=t+1}^{n} E\{\phi_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})\}^2/n^3\right) = o(1)$ as shown before. Likewise, we can show that the triple summation terms and the quadruple summation terms are all asymptotically negligible. Thus we have shown that $\sqrt{n}\hat{\gamma}_{nk}^{21}(\underline{x}, y) = o_p(1)$ uniformly in $(\underline{x}, y)$. Similarly, we can prove that $\sqrt{n}\hat{\gamma}_{nk}^{22}(\underline{x}, y) = o_p(1)$ uniformly in $(\underline{x}, y)$. Combined with the previous expansion of $\sqrt{n}\hat{\gamma}_{nk}^{1}(\underline{x}, y)$, we then finish the proof of Lemma B.3. $\Box$

The final lemma provides an asymptotically expansion for $\hat{\gamma}_k(\underline{x}, y)$ in $(\underline{x}, y) \in \mathbb{R}^{p+1}$ under the alternative hypothesis.

Lemma B.4: *Under Assumptions A.1-A.4 and the alternative hypothesis, there exists at least one $k \geq 1$ such that*

$$\sup_{(\underline{x},y)\in\mathbb{R}^{p+1}} \left| \sqrt{n-k}\left[\hat{\gamma}_{nk}(\underline{x}, y) - \gamma_{nk}(\underline{x}, y) - \gamma_{nk}^{0}(\underline{x}, y)\right] \right| = o_p(1),$$

where

$$\gamma_{nk}^{0}(\underline{x}, y) = \frac{1}{n-k} \sum_{t=k+1}^{n} \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}} \int (Y_t - m(\underline{Y}_{t-1}))\mathrm{e}^{\mathrm{i}yY_{t-p-k}} f(Y_t, \underline{Y}_{t-1}, Y_{t-p-k})\, dY_t\, dY_{t-p-k}$$

and $\gamma_{nk}(\underline{x}, y)$ is defined the same as in Lemma B.3.

Proof of Lemma B.4: This lemma is proved using a similar argument as in the proof of Lemma B.3. Hence only a sketch of its proof is provided here. Under the alternative hypothesis, the projection term is obtained by

$$
\begin{aligned}
u_{k1}(\mathcal{W}_k) =& E[U_k(\mathcal{W}_{kt}, \mathcal{W}_{ks})|\mathcal{W}_{ks} = \mathcal{W}_k] \\
=& \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right)(Y_t - m(\underline{Y}_{t-1}))\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})}\right] \\
&+ \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right)m(\underline{Y}_{t-1})\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})}\right] \\
&- \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right)\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{t-1} + yY_{t-p-k})}\right]Y \\
&+ \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right)\right]\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{-1} + yY_{-p-k})}Y \\
&- \frac{1}{2}E\left[\frac{1}{h^p}K\left(\frac{\underline{Y}_{t-1} - \underline{Y}_{-1}}{h}\right)m(\underline{Y}_{t-1})\right]\mathrm{e}^{\mathrm{i}(\underline{x}'\underline{Y}_{-1} + yY_{-p-k})} \\
:=& A_{10} + A_{11} + A_{12} + A_{13} + A_{14},
\end{aligned}
$$

where A_{1j} for $j = 1, \ldots, 4$ are defined the same as in Lemma B.3. We have an extra term comparing to Lemma B.3. Notice that under the alternative hypothesis, there exists at least one $k \geq 1$ such that $E(Y_t|\underline{Y}_{t-1}, Y_{t-p-k}) \neq E(Y_t|\underline{Y}_{t-1}) = m(\underline{Y}_{t-1})$. Therefore, the term A_{10} is not zero identically. In fact, we have

$$
\begin{aligned}
A_{10} =& \frac{1}{2}\int \frac{1}{h^p}K\left(\frac{z - \underline{Y}_{-1}}{h}\right)(Y_t - m(z))\mathrm{e}^{\mathrm{i}(\underline{x}'z + yY_{t-p-k})}f(Y_t, \underline{Y}_{-1}, Y_{t-p-k})\,dY_t\,d\underline{Y}_{-1}\,dY_{t-p-k} \\
=& \frac{1}{2}\int K(u)\left(Y_t - m(\underline{Y}_{-1} + hu)\right)\mathrm{e}^{\mathrm{i}\underline{x}'(\underline{Y}_{-1} + hu) + \mathrm{i}yY_{t-p-k}} \\
&\times f(Y_t, \underline{Y}_{-1} + hu, Y_{t-p-k})\,dY_t\,du\,dY_{t-p-k} \\
=& \frac{1}{2}\mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{-1}}\int (Y_t - m(\underline{Y}_{-1}))\mathrm{e}^{\mathrm{i}yY_{t-p-k}}f(Y_t, \underline{Y}_{-1}, Y_{t-p-k})\,dY_t\,dY_{t-p-k} + O(h^l).
\end{aligned}
$$

So, $u_{k1}(\mathcal{W}_{kt}) \approx \{\xi_t + \mathrm{e}^{\mathrm{i}\underline{x}'\underline{Y}_{t-1}}\int(Y_t - m(\underline{Y}_{t-1}))\mathrm{e}^{\mathrm{i}yY_{t-p-k}}f(Y_t, \underline{Y}_{t-1}, Y_{t-p-k})\,dY_t\,dY_{t-p-k}\}/2$, where ξ_t is defined in Lemma B.3.

It is important to notice that $\gamma^0_{nk}(\underline{x}, y) \to_p \gamma_k(\underline{x}, y)$. Therefore, $\gamma^0_{nk}(\underline{x}, y) \to_p 0$ almost everywhere in $(\underline{x}, y) \in \mathbb{R}^{p+1}$ if and only if under H_0, so that Lemma B.4 reduces to Lemma B.3 under the null hypothesis. However, under the alternative, $\gamma^0_{nk}(\underline{x}, y)$ converges to a non-zero dependence measure $\gamma_k(\underline{x}, y)$ in probability for some $k \geq 1$, causing an additional non-trivial shift term asymptotically. As a matter of fact, this distinctive behaviour under the null and under the alternative is the key to guaranteeing the consistency of our test. See also the proof of Theorem 2 for further explanation.

Then, following the same steps as in Lemma B.3, we finish the proof of Lemma B.4. □

1.10 References

Arcones, M.A., Yu, B (1994). “Central limit theorems for empirical and U-processes of stationary mixing sequences”, *Journal of Theoretical Probability*, 7, 47-71.

Bierens, H.J. (1982). “Consistent model specification tests”, *Journal of Econometrics*, 20, 105-134.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley Series in Probability and Statistics. Wiley, New York.

Bollerslev, T. (1986). “Generalized autoregressive conditional heteroskedasticity”, *Journal of Econometrics*, 31, 307-327.

Borovkova, S., Burton, R., Dehling, H. (2001). “Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation”, *Transactions of the American Mathematical Society*, 353, 4261-4318.

Chen, B., Hong, Y.M. (2012). “Testing for the Markov property in time series”, *Econometric Theory*, 28, 130-178.

Chen, X.H., White, H. (1996). “Laws of large numbers for Hilbert space-valued mixingales with applications”, *Econometric Theory*, 12, 284-304.

Chen, X.H., White, H. (1998). “Central limit and functional central limit theorem for Hilbert-valued dependent heterogeneous arrays with applications”, *Econometric Theory*, 14, 260-284.

Chen, X.H., Fan, Y.Q. (1999). “Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series”, *Journal of Econometrics*, 91, 373-401.

Dawid, A.P. (1979). “Conditional independence in statistical theory”, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 41, 1-31.

de Jong, R.M. (1996). “The Bierens test under data dependence”, *Journal of Econometrics*, 72, 1-32.

Dehling, H., Philipp, W. (2002). “Empirical Process Techniques for Dependent Data”, *In: Dehling, H., Mikosch, T., Sørensen, M. (Eds.), Empirical Process Techniques for Dependent*

Data, Birkhaüser, Basel, pp. 3-113.

Delgado, M.A., Gonzalez-Manteiga, W. (2001). "Significance testing in nonparametric regression based on the bootstrap", *Annals of Statistics*, 29, 1469-1507.

Engle, R.F. (1982). "Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation", *Econometrica*, 50, 987-1008.

Escanciano, J.C., Lobato., I.N. (2009) "Testing the martingale hypothesis", *In: Palgrave Handbook of Econometrics*, Patterson, K. and Mills, T.C. (Eds.), Palgrave, MacMilan.

Escanciano, J.C., Velasco, C. (2006). "Generalized spectral tests for the martingale difference hypothesis", *Journal of Econometrics*, 134, 151-185.

Fan, Y.Q., Li, Q. (1996). "Consistent model specification tests: Omitted variables and semiparametric functional forms", *Econometrica*, 64, 865-890.

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

Finkenstädt, B.F., Yao, Q.W., Tong, H. (2001). "A conditional density approach to the order determination of time series", *Statistics and Computing*, 11, 229-240.

Györfi, L., Härdle, W., Sarda, P., Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*, Springer Verlag, Heidelberg.

Härdle, W. (1992). *Applied Nonparametric Regression (Econometric Society Monographs)*, Cambridge University Press, United Kingdom.

Hoeffding, W. (1948). "A class of statistics with asymptotically normal distribution", *Annals of Mathematical Statistics*, 19, 293-325.

Hong, Y.M. (1998). "Testing for pairwise serial independence via the empirical distribution function", *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 60, 429-453.

Hong, Y.M. (1999). "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach", *Journal of the American Statistical Association*, 84, 1201-1220.

Hong, Y.M. (2000). “Generalized spectral tests for serial dependence”, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62, 557-574.

Horowitz, J.L., Spokoiny, V. (2001). “An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative”, *Econometrica*, 69, 599-631.

Huang, J.Z., Yang, L. (2004).“Identification of nonlinear additive autoregressive models”, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 66, 463-477.

Kuan, C.-M., Lee, W.-M. (2004). “A New Test of the Martingale Difference Hypothesis”, *Studies in Nonlinear Dynamics and Econometrics*, 8, 1-24.

Liu, R.Y. (1988). “Bootstrap procedures under some non-i.i.d. models”, *Annals of Statistics*, 16, 1696-1708.

Lobato, I. N. (2003). “Testing for nonlinear autoregression”, *Journal of Business and Economic Statistics*, 21, 164-173.

Neumeyer, N., Dette, H. (2007). “Testing for symmetric error distribution in nonparametric regression models”, *Statistica Sinica*, 17, 775-795.

Nishiyama, Y., Hitomi, K., Kawasaki, Y., Jeong, K. (2011). “A consistent nonparametric test for nonlinear causality - Specification in time series regression”, *Journal of Econometrics*, 165, 112-127.

Parthasarathy, K.R. (1967). *Probability Measures on Metric Spaces*, Academic Press, New York.

Politis, D., Romano, J. (1994). “Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap”, *Statistica Sinica*, 4, 461-476.

Robinson, P.M. (1988). “Root-N-consistent semiparametric regression”, *Econometrica*, 56, 931-954.

Robinson, P.M. (1989). “Hypothesis testing in semiparametric and nonparametric models for econometric time series”, *Review of Economic Studies* , 56, 511-534.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

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

Stute, W. (1994). "U-statistic processes: a martingale approach", *Annals of Probability*, 22, 1725-1744.

Stute, W. (1997). "Nonparametric model checks for regression", *Annals of Statistics*, 25, 613-641.

Stute, W., Gonzalez-Manteiga, W., Quindimil, M.P. (1998). "Bootstrap Approximations in Model Checks for Regression", *Journal of the American Statistical Association*, 93, 141-149.

Tong, W. (1990). *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press, United Kingdom.

Tschernig, R., Yang, L. (2000). "Nonparametric lag selection for time series", *Journal of Time Series Analysis*, 21, 457-487.

Wang, X., Hong, Y.M. (2012). "Characteristic function based testing for conditional independence: A nonparametric regression approach", *Working Paper*, Xiamen University.

Wu, C.F.J. (1986). "Jacknife, bootstrap, and other resampling methods in regression analysis", *Annals of Statistics*, 14, 1261-1295.

Yoshihara, K. (1976). "Limiting behavior of U-statistics for stationary, absolutely regular processes", *Zeitschrift fur Wahrscheinlichkeitstheorie und verwande Gebiete*, 35, 237-252.

Chapter 2

Testing Symmetry of a Nonparametric Conditional Distribution

2.1 Introduction

Testing symmetry of a conditional distribution is useful in model checking, but also on its own. Location and dispersion can be unambiguously defined under symmetry, and the center of symmetry can be robustly, even adaptively, estimated. That is, statistical inferences can be improved under the symmetry assumption. Most popular specifications are ruled out when the symmetry assumption is rejected. Despite specification related issues, it is interesting to test symmetry in many circumstances. For instance, we may be interested in testing whether losses are more likely than gains in stock markets controlling for the available information, or whether negative and positive shocks are equally likely in macroeconomic models. For example, checking asymmetries in business cycles. The rich body of empirical studies suggests that business cycle expansions appear to be more persistent and less volatile than contractions. For example, DeLong and Summers (1986), Hussey (1992), Verbrugge (1997) and Belaire-Franch and Contreras (2002) all showed that economic time series tend to behave asymmetrically over the business cycle. Brunner (1992) argued that the assumption of Gaussian shocks places strong restrictions on the time series behaviour of economic fluctuations. Models built upon the Gaussian assumption would be too restrictive and even produce unreliable conclusions. The assumption of symmetry would also affect our forecasts. Symmetry implies that positive shocks to the conditional mean are as likely as negative shocks. If this is not the case, forecasts should adjust to the possibility that positive and negative forecast errors are not equally likely. For example, Campbell and Hentschel (1992) proposed the 'No news is good news' model in which the residuals in a model of log returns conditional on volatility are asymmetrically distributed. Therefore, both theoretically and empirically speaking, whether or not to impose symmetric Gaussian shocks to the conditional mean is a crucial problem to be addressed in macro-model-building exercises before exploring more complicated business cycle structures.

The first symmetry test is due to Smirnov (1947) as an extension of the classical goodness-of-fit tests. Testing symmetry of the unknown marginal distribution of residuals coming from some parametric specification of the regression curve is an effective way of testing symmetry of the conditional distribution when innovations are independent of explanatory variables. See, for instance, Gupta (1967), Butler (1969), Gastwirth (1971), Doksum et al. (1977), Randles et al. (1980), Aki (1981), Antille et al. (1982), Battacharya et al. (1982), Hušková (1984), Koziol (1985), Schuster and Barker (1987), Hollander (1988), Ahmad and Li (1997), Hyndman and Yao (2002) or Psaradakis (2003). These tests are unable to detect infinitely many departures from the conditional symmetry hypothesis where innovations are not independent of the explanatory variables. The hypothesis of independence between innovations and explanatory variables has been relaxed to allow conditional scale-location models, where rather than the innovations, only suitably scaled innovations are assumed to be independent of covariates. For instance, Fan and Gencay (1995) and Bai and Ng (2001) considered fully parametric location and scale functions, motivated by testing conditional symmetry in GARCH-type models. Dette et al. (2002) and Neumeyer and Dette (2007) considered tests with nonparametric location and scale functions. However, these tests are still inconsistent in directions where the scaled innovations are not independent of the covariates, which is likely in a serial dependent data context, e.g. heterokurtosis is likely when dealing with financial data (Harvey and Siddique 1999, 2000). Delgado and Escanciano (2007) proposed a test of conditional symmetry around a parametric location function in a serial dependence context, which is consistent in general situations where higher order conditional moments may be dependent of functions different to those characterizing location and scale. Misspecification of the regression function may lead to misleading conclusion, but this nuisance is avoided in this article by estimating the location using smoothing techniques.

The rest of the paper is organized as follows. In the next section, we describe in details our testing problem and how to construct the test statistics. In Section 3, we provide the asymptotic distribution of the test statistics. In Section 4, the asymptotic power of the tests is studied. In Section 5, we suggest and validate a wild bootstrap method in order to implement our tests in practice. Extensive Monte Carlo simulation results are shown in Section 6. We include an empirical application of the proposed tests for stock market returns in Section 7. Mathematical proofs of the main results and some auxiliary results are deferred to the mathematical Appendices A and B, respectively.

2.2 The testing procedure

Consider a $\mathbb{R}^{1+d}-$valued strictly stationary time series process $(Y, X) = (Y_t, X_t)_{t\in\mathbb{Z}}$ and an information set $I_t = \{(Y_{s-1}, X_s)\,,\ t-k+1 \leq s \leq t\}$ at time t, i.e. $I_t \in \mathbb{R}^p$ with $p = k(1+d)$. We are interested in testing whether the conditional distribution of Y_t given $I_t = u$ is symmetric around $m(u)$ for each $u \in \mathbb{R}^p$, where $m : \mathbb{R}^p \to \mathbb{R}$ is an unknown but smooth function. Consider the family of symmetric distributions around zero

$$\mathcal{G} = \{G : G(v) = 1 - G(-v)\}\,,$$

and define the conditional distribution of innovations $\varepsilon_t = Y_t - m(I_t)$ given $I_t = u$ evaluated at $\varepsilon_t = v$ as $F(v|u) = \mathbb{P}(\varepsilon_t \leq v|\, I_t = u)$. The null hypothesis can be expressed as

$$H_0 : F(\cdot|u) \in \mathcal{G} \text{ for each } u \in \mathbb{R}^p \text{ a.s.}$$

This hypothesis can be equivalently characterized in terms of the joint distribution of the innovations and the corresponding information set, i.e. $K(v,u) = \mathbb{P}(\varepsilon_t \leq v,\ I_t \leq u) = \mathbb{E}(1(I_t \leq u)F(v|I_t))$, where $1(A)$ is the indicator function of the event A and event $(I_t \leq u)$ indicates that each component of the vector I_t is less than or equal to the corresponding element in u.

First, we notice that H_0 is satisfied if and only if the conditional characteristic function of ε_t given I_t is real valued, i.e. it does not have an imaginary part, see Ghosh and Ruymgaart (1992) and Heathcote et al. (1995) and references therein. That is, H_0 holds if and only if (iff)

$$H_0 : \mathbb{E}\left(\sin\left(w\varepsilon_t\right)\right| I_t = u) = \int_{\mathbb{R}} \sin(ws)F(ds|u) = 0 \text{ for each } (w,u) \in \mathbb{R}^{p+1}.$$

Applying the fundamental theorem of calculus, we can express H_0 as a moment restriction with respect to the joint distribution, i.e.

$$H_0 : \mathbb{E}\left(\sin\left(w\varepsilon_t\right) 1(I_t \leq u)\right) = 0 \text{ for each } (w,u) \in \mathbb{R}^{p+1}. \tag{2.1}$$

If the innovations $\{\varepsilon_t\}_{t=1}^n$ were observed, a natural estimator of the expectation in (2.1) is

$$\begin{aligned} R_n^0(w,u) &= \int_{\mathbb{R}} \sin(wv)K_n^0(dv,u) \\ &= \frac{1}{n}\sum_{t=1}^{n} \sin(w\varepsilon_t)1(I_t \leq u) \end{aligned}$$

where $K_n^0(v,u) = n^{-1}\sum_{t=1}^n 1(\varepsilon_t \leq v)1(I_t \leq u)$ is the empirical distribution function of $\{\varepsilon_t, I_t\}_{t=1}^n$. There exist functional central limit theorems (FCLT) for $\alpha_n^0 = \sqrt{n}\left(K_n^0 - K\right)$ under $i.i.d.$ observations and specific serial dependence structures, and the limiting distribution of $\sqrt{n}R_n^0$ under H_0 is obtained from these FCLT's. However, it is hard to get the limiting distribution of α_n^0 under a general serial dependence structure. However, we can take advantage of the fact that, H_0 holds iff $S^0 = 0$ a.s. almost everywhere, where,

$$\begin{aligned} S^0(v,u) &= \mathbb{E}\left(1\left(I_t \le u\right)\left(F(v|I_t) - 1 - F(-v|I_t)\right)\right) \\ &= K^0\left(v,u\right) - K^0(\infty,u) - K^0(-v,u). \end{aligned}$$

Therefore, under H_0,

$$R_n^0(w,u) = \frac{1}{2}\int_{\mathbb{R}} \sin(wv) S_n^0(dv,u)$$

with

$$S_n^0(v,u) = K_n^0\left(v,u\right) - K_n^0(\infty,u) - K_n^0(-v,u).$$

It is easy to handle the limiting distribution of $\sqrt{n}S_n^0$ under general serial dependence assumptions by exploiting the fact that, for $(v,u) \in \mathbb{R}^{p+1}$ fixed, $\omega_t\left(v\right) = \left(1\left(\varepsilon_t \le v\right) - 1\left(-\varepsilon_t \le v\right)\right)$ is a martingale difference with respect to the $\sigma-$field generated by the information set obtained up to time t, $\mathcal{F}_t = \sigma\left(I_t, I_{t-1}, ...\right)$, i.e. $\mathbb{E}\left(\omega_t\left(v\right)\middle|\, \mathcal{F}_t\right) = 0$ a.s. for each $(v,u) \in \mathbb{R}^{p+1}$. Hence, the asymptotic distribution of $\sqrt{n}S_n^0$ under H_0 can be obtained by applying weak convergence results for martingales, e.g. Levental (1989), Bae and Levental (1995) and Nishiyama (2000).

The next two regularity conditions summarize the restrictions on the serial dependence structure of the underlying time series process.

A.1 $\{Y_t, I_t\}_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic process.

A.2 The joint distribution of (ε_1, I_1), K, is uniformly continuous on $\overline{\mathbb{R}}^{p+1}$ and $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a Markov's process, in the sense that, under H_0,

$$\mathbb{P}\left(\varepsilon_t \le \cdot \middle|\, \mathcal{F}_t\right) = F\left(\cdot \middle|\, I_t\right) \text{ a.s. for each } t \in \mathbb{Z}.$$

By defining $S_n^0(-\infty,\cdot) = S_n^0(\cdot,-\infty) = 0$, the sample paths of $\sqrt{n}S_n^0$ belong to the space $\ell^\infty(\overline{\mathbb{R}}^{p+1})$, the space of all uniformly bounded real functions on $\overline{\mathbb{R}}^{p+1} := [-\infty,\infty]^{p+1}$, which is equipped with the sup-norm. Assuming A.1 and A.2, Delgado and Escanciano (2007) show that under H_0, $\sqrt{n}S_n^0$ converges weakly on the topology of $\ell^\infty(\overline{\mathbb{R}}^{p+1})$ endowed with the sup-norm to S^0_∞, a Gaussian process centered at zero with continuous sample paths and covariance function

$$\mathbb{E}\left(S^0_\infty\left(v_1,u_1\right)S^0_\infty\left(v_2,u_2\right)\right)=\mathbb{E}\left(\omega_1\left(v_1\right)\omega_1\left(v_2\right)1\left(u_1\wedge u_2\right)\right),$$

where $a\wedge b\equiv\min\{a,b\}$ element-wisely for any vectors a and b. Therefore, applying the continuous mapping theorem (CMT), $\sqrt{n}R^0_n$ converges in distribution to R^0_∞ under H_0, with

$$R^0_\infty(w,u)=\frac{1}{2}\int_{\mathbb{R}}\sin(wv)S^0_\infty(dv,u).$$

Furthermore, for any continuous functional $\varphi:\ell^\infty(\overline{\mathbb{R}}^{p+1})\mapsto\mathbb{R}$,

$$\varphi(S^0_n)\to_d\varphi(S^0_\infty).$$

If $\{\varepsilon_t\}^n_{t=1}$ were observed, tests statistics could be based on $\|\sqrt{n}R^0_n\|$, where $\|\cdot\|$ is any norm satisfying the Riesz property, i.e. if $g\le h$, then $\|g\|\le\|h\|$. The test would be given by the statistic $\phi^0_n(c)=1\left(\|\sqrt{n}R^0_n\|\ge c\right)$, where $1(A)$ is the indicator function of the event A and c is the critical values which determines the size of the test. In particular, a test at a significance level α uses the critical value c_α, such that $\mathbb{P}\left(\|R^0_\infty\|\ge c_\alpha\right)=\alpha$.

A feasible test is based on residuals $\hat{\varepsilon}_t=Y_t-\hat{m}(I_t)$, $t=1,\ldots,n$, for some suitable nonparametric estimator $\hat{m}$ of m. Let

$$K_n\left(v,u\right)=\frac{1}{n}\sum_{t=1}^{n}1(\hat{\varepsilon}_t\le v)1(I_t\le u),$$

and define the feasible estimator of S^0 by

$$S_n(v,u)=K_n\left(v,u\right)-K_n(\infty,u)-K_n(-v,u),$$

where $K_n(\infty,u)\equiv K^0_n(\infty,u)=F_{nI}(u)$ is the empirical distribution function of $\{I_t\}^n_{t=1}$. The estimator R_n based on S_n is therefore given by

$$R_n(w,u) = \frac{1}{2}\int_{\mathbb{R}} \sin(wv) S_n(dv,u).$$

To test H_0, we propose test statistics based on S_n, i.e. the Cramér-von Mises-type statistic

$$\begin{aligned} CvM_n &= \int_{\mathbb{R}^{p+1}} \left(\sqrt{n} S_n(v,u)\right)^2 \, K_n(dv,du) \\ &= \sum_{t=1}^{n} S_n(\hat{\varepsilon}_t, I_t)^2, \end{aligned}$$

and the Kolmogorov-Smirnov-type statistic

$$KS_n = \sup_{(v,u)\in\overline{\mathbb{R}}^{p+1}} |\sqrt{n} S_n(v,u)|.$$

Furthermore, a test statistic of the Cramér-von Mises-type based on R_n is given by

$$\overline{CvM}_n = \int_{\mathbb{R}^{p+1}} \left(\sqrt{n} R_n(w,u)\right)^2 \, H_n(\,dw,\,du),$$

where following the arguments of Epps and Pulley (1983), we choose H_n to be a weighting function of the form $H_n(w,u) = \Phi(w)F_{nI}(u)$ with $\Phi(w)$ the standard normal cumulative distribution function and $F_{nI}(u)$ the empirical distribution function of $\{I_t\}_{t=1}^n$ defined before. Likewise, a Kolmogorov-Smirnov-type statistic based on R_n is

$$\overline{KS}_n = \sup_{(w,u)\in\Pi_c} |\sqrt{n} R_n(w,u)|, \tag{2.2}$$

where $\Pi_c \subset \mathbb{R}^{p+1}$ is a compact subset containing the origin.

2.3 Asymptotic results

The aim of this section is to study the asymptotic distribution of the test statistics under the null hypothesis.

We start by estimating the location function $m(u)$ for an arbitrary point $u = (u_1, \ldots, u_p)$. There are many available nonparametric estimators in the literature. We propose to estimate $m(u)$ by the Nadaraya-Watson estimator, see Nadaraya (1964) and Watson (1964), i.e.

$$\hat{m}(u) = \frac{n^{-1} \sum_{t=1}^{n} Y_t K_h (u - I_t)}{\hat{f}_I(u)}, \tag{2.3}$$

where $\hat{f}_I(u)$ is the Nadaraya-Watson estimator of the density function $f_I(u)$,

$$\hat{f}_I(u) = n^{-1} \sum_{t=1}^{n} K_h (u - I_t).$$

Here, for $u = (u_1, \ldots, u_p) \in \mathbb{R}^p$, $K(u) = \prod_{j=1}^{p} k(u_j)$ is a p-dimensional product kernel, $k : \mathbb{R} \to \mathbb{R}$ is a symmetric univariate kernel function, $h = (h_1, \ldots, h_p) \in \mathbb{R}_+^p$ is a p-dimensional bandwidth vector converging to zero when n tends to infinity, and $K_h(u) = \prod_{j=1}^{p} k(u_j/h_j)/h_j$. Notice that, to simplify our analysis, we will let $h_1 = \ldots = h_p = h$. However, different bandwidths could be easily allowed and, with slightly more complex notation, the asymptotic theory developed in this paper will still be valid when h_j $(j = 1, \ldots, p)$ satisfies $h_j/h \to c_j$ for some $0 < c_j < \infty$ and some baseline bandwidth h satisfying Assumption A.5 below.

One remark is in order. Since we have to estimate ε_t nonparametrically, there is a technical issue of random denominators ($\hat{f}_I(u)$ in the expression of (2.3)). To avoid such a problem, we can introduce a trimming parameter $b > 0$ to trim out those close-to-zero values of the density estimates $\hat{f}_I(I_t)$ such that $1_t = 1(|\hat{f}_I(I_t)| > b)$, see e.g. Robinson (1988). However, the choice of the trimming parameter b is introduces some difficulty. Another way to circumvent the random denominator problem consists of restricting the justification of our tests to the case where the density function $f_I(u)$ is bounded from below by some positive constant δ, i.e. $\inf_{u \in \mathbb{R}^p} f_I(u) \geq \delta > 0$. This assumption may look restrictive since it would rule out any regressors whose distribution has unbounded support such as the commonly used normal regressors. But in the present paper, this assumption appears not causing severe problems to the size and power performance of the proposed tests when it is not fulfilled, as illustrated in the Monte Carlo part of Section 6.

To derive our asymptotic results, we need A.2 and modify A.1 to A.1' below.

A.1' (i) $\{Y_t, I_t\}_{t\in\mathbb{Z}}$ is a strictly stationary, ergodic and absolutely regular process with β-mixing coefficients $\beta(j) = O(j^{-(2+\eta)/\eta})$ for some constant $0 < \eta < 1$; (ii) the marginal density $f_I(u)$ of I_t is bounded from below and from above, its partial derivatives exist up to order L for some integer $L \geq 2$, and they are uniformly continuous; (iii) $\mathbb{E}\left(|Y_t|^{2+\delta}\right) < \infty$ for some $\delta > \eta/(1-\eta)$.

We also impose the following regularity conditions.

A.3 All partial derivatives of location m up to order L exist a.s., and they are uniformly continuous and bounded.

A.4 The p-th product function K satisfies $\int u^i K(u)\,du = \delta_{0i}$ for $i = 0, 1, \cdots, L-1$ and $\int u^l K(u)\,du \neq 0$ with δ_{ij} the delta function equal to one if $i = j$ and zero otherwise. The univariate function k is a bounded, symmetric probability density function on $\mathbb{R}$.

A.5 The bandwidth h is such that: (i) $h \to 0$; (ii) $nh^{3p+\gamma} \to \infty$ for some small $\gamma > 0$; and (iii) $nh^{2L} \to 0$, as $n \to \infty$.

A.6 The distribution function of ε_t, F_ε, is twice continuously differentiable a.s. Its density function $f_\varepsilon(v)$ is bounded from above a.s. and satisfies $\sup_v f'_\varepsilon(v) < \infty$.

All the assumptions are quite standard in nonparametric time series analysis. Assumption A.1' is a regularity condition on the underlying data generating process (DGP) of $\{Y_t, I_t\}_{t\in\mathbb{Z}}$. Assumption A.1'(i) restricts the degree of temporal dependence in $\{Y_t, I_t\}$, which is generally adopted in the nonparametric time series literature, see, e.g. Hjellvik et al. (1998), Su and White (2007, 2008), Chen and Hong (2010, 2012) and Wang and Hong (2012) amongst others.

Let $\{V_t\}_{t\in\mathbb{Z}}$ be any strictly stationary stochastic process. We say that $\{V_t\}$ is β-mixing (or absolutely regular) with mixing coefficient $\beta(j)$ if

$$\beta(j) = \sup_{s\in\mathbb{N}} \mathrm{E}\left[\sup_{A\in\mathscr{F}_{s+j}^{\infty}} |\Pr(A|\mathscr{F}_{-\infty}^{s}) - \Pr(A)|\right] \to 0$$

as $j \to \infty$, where $\mathscr{F}_s^t$ is the σ-algebra generated by $\{V_s, \cdots, V_t\}$ for $s \leq t$. For the notion of β-mixing and other mixing conditions such as φ-mixing and strong mixing, see e.g. Doukhan (1994) and Fan and Yao (2003) amongst others. Many well-known processes, such as stationary autoregressive moving average (ARMA) processes and a large class of nonlinear processes, including bilinear, nonlinear autoregressive (NLAR), and autoregressive conditional heteroskedasticity (ARCH) models, satisfy the β-mixing condition, see e.g. Fan and Li (1999). We believe that the β-mixing condition in Assumption A.1'(i) can be relaxed to a strong mixing condition. Assumption A.1'(ii) and Assumption A.1'(iii) impose some smoothness and moment conditions. Assumption A.3 is a smoothness condition on the location function m. Assumptions A.4 and A.5 are standard in deriving asymptotic theory of nonparametric regression. A.6 is needed to prove that the first order Taylor expansion around the distribution function of true innovations is allowed and the reminder term in the expansion is asymptotically negligible. Neumeyer and Van Keilegom (2010) also assume a similar assumption to A.6.

We are now ready to state the main results of this section.

Theorem 1: *Under the null, and under Assumptions **A.1-A.2**,*

$$\sqrt{n}S_n^0 \Rightarrow S_\infty^0$$

in $\ell^\infty(\overline{\mathbb{R}}^{p+1})$, where S_∞^0 is a zero mean Gaussian process with covariance function

$$\mathbb{E}\left(S_\infty^0(v_1,u_1), S_\infty^0(v_2,u_2)\right) = \mathbb{E}[\omega_1(v_1)\omega_1(v_2)1(I_1 \leq u_1 \wedge u_2)].$$

As discussed already in the Introduction, the proof of Theorem 1 above is Theorem 1 in

Delgado and Escanciano (2007). Notice that, if we let $u = \infty$ in S_n^0, we will obtain a test of unconditional symmetry of ε_t. In fact, we can calculate that in the unconditional case, the covariance expression above reduces to $2F_\varepsilon\left(-|v_1| \vee |v_2|\right)$ with $a \vee b \equiv \max\{a, b\}$, which coincides with the covariance function of the limit of the classical empirical symmetry process $S_n^0(v, \infty)$ based on an $i.i.d.$ sample, see e.g. Smirnov (1947).

The following theorem establishes that S_n can be represented in terms of S_n^0 and some shift term asymptotically.

Theorem 2: *Under the null, if Assumptions* ***A.1'-A.6*** *hold, then, uniformly in* $(v, u) \in \overline{\mathbb{R}}^{p+1}$,

$$S_n(v, u) = S_n^0(v, u) + W_n^0(v, u) + o_p(n^{-1/2}),$$

where

$$W_n^0(v, u) = \left(f_\varepsilon(v) + f_\varepsilon(-v)\right) \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t 1(I_t \leq u),$$

with $\varepsilon_t = Y_t - m(I_t)$ *the symmetric errors conditional on* I_t.

Now notice that

$$W_n^0(v, u) = \frac{1}{2}\left(f_\varepsilon(v) + f_\varepsilon(-v)\right) \int_{\mathbb{R}} \bar{v}\, S_n^0(d\bar{v},\, u).$$

The asymptotic distribution of $\sqrt{n}S_n$ is then a straightforward consequence of Theorems 1 and 2. We immediately have the following result.

Corollary 1: *Under the null and Assumptions* ***A.1'-A.6***,

$$\sqrt{n}S_n \Rightarrow S_\infty^1$$

in $\ell^\infty(\overline{\mathbb{R}}^{p+1})$*, where*

$$S_\infty^1(v, u) \stackrel{d}{=} S_\infty^0(v, u) + \frac{1}{2}\left(f_\varepsilon(v) + f_\varepsilon(-v)\right) \int_{\mathbb{R}} \bar{v}\, S_\infty^0(d\bar{v},\, u).$$

It is interesting to compare our asymptotic results with those obtained by Neumeyer and Dette (2007) in an $i.i.d.$ context for a univariate case, where they assume explicitly independence of I and ε. Again, under the null hypothesis of a symmetric error distribution, by letting $u=\infty$ in $S_n(v,u)$, we can obtain a covariance structure given by

$$2F_{\varepsilon}\left(-|v_1|\vee|v_2|\right)+4f_{\varepsilon}(v_1)f_{\varepsilon}(v_2)+4f_{\varepsilon}(v_1)\int_{-\infty}^{v_2}\bar{v}f_{\varepsilon}(\bar{v})\,d\bar{v}+4f_{\varepsilon}(v_2)\int_{-\infty}^{v_1}\bar{v}f_{\varepsilon}(\bar{v})\,d\bar{v}.$$

The above covariance function coincides with that obtained by Neumeyer and Dette (2007). Comparing to the classical covariance function obtained by Smirnov (1947), there are three additional terms depending on the density of the error distribution. As in Neumeyer and Dette (2007), this complication is caused by the estimation of the location function in our procedure.

We state in Corollaries 2 and 3 the asymptotic null distributions of $\sqrt{n}R_n$ and the test statistics based on S_n or R_n as a straightforward application of Corollary 1 and the continuous mapping theorem, see e.g. Billingsley (1968) Theorem 5.1.

Corollary 2: *Under H_0, if Assumptions **A.1'-A.6** hold, then*

$$\sqrt{n}R_n\Rightarrow R_{\infty}^{1}$$

in $\ell^{\infty}(\overline{\mathbb{R}}^{p+1})$ with

$$R_{\infty}(w,u)\overset{d}{=}\frac{1}{2}\int_{\mathbb{R}}\sin(wv)S_{\infty}^{1}(dv,\,u).$$

Corollary 3: *Suppose Assumptions **A.1'-A.6** hold. Then, under the null,*

$$CvM_n\rightarrow_d\int_{\mathbb{R}^{p+1}}S_{\infty}^{1}(v,u)^2K(dv,\,du),$$

,

$$\overline{CvM}_n\rightarrow_d\int_{\mathbb{R}^{p+1}}R_{\infty}^{1}(w,u)^2H(dw,\,du),$$

where $H(w,u)=\Phi(w)F_I(u)$, $F_I(u)$ is the CDF of I_1 and $\Phi(v)$ is the CDF of standard normal *distribution, and*

$$KS_n \to_d \sup_{(v,u)\in\overline{\mathbb{R}}^{p+1}} |S^1_\infty(v,u)|,$$

$$\overline{KS}_n \to_d \sup_{(w,u)\in\Pi_c} |R^1_\infty(w,u)|,$$

where $\Pi_c \subset \mathbb{R}^{p+1}$ *is a compact subset containing the origin.*

2.4 Behaviour under local alternatives

We consider contiguous asymmetric nonparametric alternatives of the following form:

$$H_{An} : f^{(n)}(v|u) = f(v|u)\left[1 + \frac{1}{\sqrt{n}}\delta_n(v,u)\right] \quad \text{a.s.,} \tag{2.4}$$

where $f(\cdot|\cdot)$ is a symmetric conditional density, i.e. $f(v|u) = f(-v|u)$ for each $(v,u) \in \mathbb{R}^{p+1}$, and $\delta_n : \mathbb{R}^{p+1} \to \mathbb{R}$ is a function such that for each $n \geq 1$ and each $(v,u) \in \mathbb{R}^{p+1}$,

$$\frac{1}{\sqrt{n}}\delta_n(v,u) \geq -1, \quad \delta_n(v,u) \neq \delta_n(-v,u), \quad \sup_{u\in\mathbb{R}^p}\left|\int_{\mathbb{R}} \delta_n(v,u) f(dv|u)\right| = 0,$$

and

$$\delta_n \to \delta \quad \text{in} \quad L_2(H),$$

where $L_2(H)$ is the Hilbert space of all H-square integrable real-valued functions on $\mathbb{R}^{p+1}$.

We will show in the proof of Theorem 3 in Appendix A that, the asymptotic expansion of S_n under the local alternatives (2.4) can be expressed as

$$\begin{aligned}\sqrt{n}S_n(v,u) =& \sqrt{n}\tilde{S}_n(v,u) + (f_\varepsilon(v) + f_\varepsilon(-v))\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(Y_t - m(I_t))1(I_t \leq u) \\ &+ \Xi(v,u) - \Xi(\infty,u) + \Xi(-v,u) + o_p(1),\end{aligned} \tag{2.5}$$

uniformly in $(v,u) \in \mathbb{R}^{p+1}$, where $\tilde{S}_n$ is a new stochastic process. One can show that $\sqrt{n}\tilde{S}_n$ converges weakly to the same Gaussian process S^0_∞ with zero mean and covariance structure

$\mathbb{E}\left(\omega_1(v_1)\omega_2(v_2)1(I_1 \leq u_1 \wedge u_2)\right)$ in Theorem 1. It is important to stress that the functions $\Xi(v,u)$, $\Xi(\infty,u)$, and $\Xi(-v,u)$ are three nontrivial shift terms caused by the asymmetric local alternatives $f^{(n)}(v|u)$. Formally, $\Xi(v,u)$ is given by

$$\Xi(v,u) = \int_{-\infty}^{v} \mathbb{E}\left[f(\bar{v}|I_1)\delta_n(\bar{v},I_1)1(I_1 \leq u)\right] d\bar{v}.$$

On the other hand, in the context of parametrically specified location $m(u) = m(u,\theta_0)$, Delgado and Escanciano (2007) assume that the consistent estimator θ_n of θ_0 has to satisfy the following asymptotic representation,

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} l_{\theta_0}(\varepsilon_t, I_t) + o_p(1),$$

where $l_{\theta_0}(\cdot,\cdot)$ is such that $\mathbb{E}\left[l_{\theta_0}(\varepsilon_1, I_1)\right] = 0$ and $\mathbb{E}\left[l_{\theta_0}(\varepsilon_1, I_1)l'_{\theta_0}(\varepsilon_1, I_1)\right]$ exists and is positive definite. However, unlike Delgado and Escanciano (2007), here we don't have an additional shift term caused by the parametric model. Since in our nonparametric context, the crucial assumption that the $\sqrt{n}$-consistent estimators θ_n of the true parameters θ_0 must satisfy an asymptotic representation of Bahadur form is no longer needed. In fact, the asymptotic behaviour of θ_n under the contiguous alternatives H_{An} is no longer a concern to us at all. Nevertheless, we notice that comparing to the expansion of S_{n,θ_n} under their H_{An} in Delgado and Escanciano (2007), its nonparametric counterpart of asymptotic representation of S_n under the local alternatives (2.4), which takes the form in (2.5), is very similar to that of S_{n,θ_n} except without an additional shift term resulting from the parametric assumption (i.e. term $\Delta^2_\theta(u,v)$ in their expansion). This result is not unexpected given that we do not need to model the location using a parametric specification.

Denote by $\Xi^1(v,u) = \Xi(v,u) - \Xi(\infty,u) + \Xi(-v,u)$. We obtain the following theorem for the asymptotic distribution of S_n under the class of local alternative H_{An}.

Theorem 3: *Under the alternative hypothesis H_{An} in* (2.4)*, if Assumptions **A.1'-A.6** hold, then*

$$\sqrt{n}S_n - \Xi^1 \Rightarrow S^1_\infty$$

in $\ell^\infty(\overline{\mathbb{R}}^{p+1})$.

This theorem shows that the asymptotic distribution of S_n is nontrivially shifted under the local alternatives H_{An} and hence guarantees that the corresponding test statistics based on suitable continuous functionals has power against the class of local alternatives H_{An}. That is, our tests are able to detect non-parametric alternatives converging to the null at the parametric rate $\sqrt{n}$ with n the sample size. Generally speaking, the shift term Ξ^1 is not identically zero which would guarantee the good power performance of the tests. Only under very rare and uninteresting circumstances would Ξ^1 be zero. In those cases, the tests do not have power against them since it cannot detect those types of alternatives which happen to make $\Xi^1 = 0$. A detailed exploration of such cases is beyond the scope of the present paper.

2.5 Bootstrap

Since the asymptotic distribution of the process S_n under the null hypothesis depends on the unknown data generating process in a highly complicated way, the test statistics are not asymptotically distribution-free. We cannot compute the critical values without estimating the unknown features of the true data generating processes. Furthermore, we cannot tabulate the critical values because different critical values will be needed for different DGPs. To overcome all these difficulties, we propose to implement our tests through the wild bootstrap procedure to test for conditional symmetry in a dynamic nonparametric regression framework. See, e.g. Wu (1986), Liu (1988), Härdle and Mammen (1993), Li and Wang (1998), Stute et al. (1998), Whang (2000) and Neumeyer and Dette (2007) amongst many others, for detailed discussions about the advantages of wild bootstrap approach in various contexts.

The essential problem here is to find a bootstrap distribution that mimics the null distribution of the test statistics even though the data may fail to satisfy the null hypothesis. In this paper, to successfully impose the null restriction and to allows heteroskedastic errors, we consider the following wild bootstrap procedure for the test statistic CvM_n (or KS_n):

Step 1: Estimate the nonparametric model $Y_t = m(I_t) + \varepsilon_t$ using the original sample

$\{(Y_t, I_t)\}_{t=1}^n$ and obtain the nonparametric residuals $\hat{\varepsilon}_t = Y_t - \hat{m}(I_t)$ for $t = 1, \ldots, n$.

Step 2: Obtain the wild bootstrap residuals ε_t^* for $t = 1, \ldots, n$ using a two point distribution, i.e., $\varepsilon_t^* = v_t \hat{\varepsilon}_t$, where $\{v_t\}_{t=1}^n$ is a sequence of i.i.d. Rademacher random variables, i.e. $P(v_t = 1) = P(v_t = -1) = 0.5$ and is independent of the sample $\{(Y_t, I_t)\}_{t=1}^n$.

Step 3: Obtain $Y_t^* = \hat{m}(I_t) + \varepsilon_t^*$ for $t = 1, \ldots, n$. The resulting sample $\{(Y_t^*, I_t)\}_{t=1}^n$ is the bootstrap sample.

Step 4: Generate new estimated nonparametric residuals according to $\hat{\varepsilon}_t^* = Y_t^* - \hat{m}^*(I_t)$, where the regression function $\hat{m}^*(\cdot)$ is defined analogously to $\hat{m}(\cdot)$ in (2.3) but is based on the bootstrap sample $\{(Y_t^*, I_t)\}_{t=1}^n$.

Step 5: Compute the bootstrapped value of CvM_n by applying the definition of CvM_n to the bootstrap sample $\{(Y_t^*, I_t)\}_{t=1}^n$ in place of the original sample $\{(Y_t, I_t)\}_{t=1}^n$. We denote the bootstrapped test statistics by CvM_n^*.

Step 6: Repeat Steps 2-5 above B times to give a sample $\{CvM_{n,b}^*\}_{b=1}^B$ of the bootstrapped value of CvM_n. The distribution of this sample, which is commonly called the "bootstrap distribution" in the literature, mimics the distribution of CvM_n under the null hypothesis.

Step 7: Let $c_{\alpha,B}^{CvM*}$ be the $(1-\alpha)$-th sample quantile of the "bootstrap distribution" of CvM_n^*. It is the bootstrap estimate of the α-level critical value. More formally, let $CvM_{n,(1)}^* \leq CvM_{n,(2)}^* \leq \cdots \leq CvM_{n,(B)}^*$ denote the ordered values of the B realizations of CvM_n^*, we choose $c_{\alpha,B}^{CvM*} = CvM_{n,([B(1-\alpha)+1])}^*$. For instance, in the case of $\alpha = 0.05$ and $B = 100$, we would take $c_{\alpha,B}^{CvM*} = CvM_{n,(96)}^*$. We reject the null hypothesis at the significance level α if $CvM_n > c_{\alpha,B}^{CvM*}$.

The asymptotic behaviour of the bootstrapped process S_n^* conditional on the original sample $\{Y_t, I_t\}_{t=1}^n$ will be established. Formally, the bootstrapped version of the empirical process S_n is defined as the following,

$$S_n^*(v, u) = \frac{1}{n}\sum_{t=1}^n \hat{\omega}_t^*(v) 1(I_t \leq u), \quad (v, u) \in \mathbb{R}^{p+1},$$

where $\hat{\omega}_t^*(v) = 1(\hat{\varepsilon}_t^* \leq v) - 1(-\hat{\varepsilon}_t^* \leq v)$ is the bootstrap counterpart of $\hat{\omega}_t(v)$ and the sequence

$\{\hat{\varepsilon}_t^*\}_{t=1}^n$ are the wild bootstrap nonparametric residuals obtained from the above proposed bootstrap procedure. The "bootstrap distribution" of S_n^* is expected to mimic the asymptotic distribution of S_n under the null hypothesis.

We now establish the first order asymptotic validity of the wild bootstrap procedure. Therefore, we can approximate the asymptotic distribution of the process S_n by that of the bootstrapped process S_n^*.

Theorem 4: *Suppose Assumptions **A.1'-A.6** hold,*

$$\sqrt{n}S_n^* \Rightarrow S_\infty^1 \quad \textit{in probability}$$

in $\ell^\infty(\overline{\mathbb{R}}^{p+1})$, where S_∞^1 is the Gaussian process defined in Corollary 1 and $\Rightarrow$ in probability denotes the weak convergence in probability under the bootstrap law, i.e., conditional on the original sample $\{(Y_t, I_t)\}_{t=1}^n$.

Let $K_n^*(v,u) = n^{-1}\sum_{t=1}^n 1(\hat{\varepsilon}_t^* \le v)1(I_t \le u)$. Define by

$$\begin{aligned} CvM_n^* &= \int_{\mathbb{R}^{p+1}} \left(\sqrt{n}S_n^*(v,u)\right)^2 \, K_n^*(dv,du) \\ &= \sum_{t=1}^n S_n^*(\hat{\varepsilon}_t^*, I_t)^2 \end{aligned}$$

and

$$KS_n^* = \sup_{(u,v)\in\overline{\mathbb{R}}^{p+1}} |\sqrt{n}S_n^*(v,u)|$$

the bootstrapped version of our test statistics. Likewise, we can get

$$R_n^*(w,u) = \frac{1}{2}\int_{\mathbb{R}} \sin(wv) S_n^*(dv,\, u),$$

and the bootstrapped test statistics based on it:

$$\overline{CvM}_n^* = \int_{\mathbb{R}^{p+1}} \left(\sqrt{n}R_n^*(w,u)\right)^2 \, H_n(dw,\, du),$$

and

$$\overline{KS}_n^* = \sup_{(w,u)\in\Pi_c} |\sqrt{n}R_n^*(w,u)|.$$

The following corollary establishes the limiting distributions of our bootstrapped test statistics CvM_n^* and KS_n^* based on S_n^*, and $\overline{CvM}_n^*$ and $\overline{KS}_n^*$ based on R_n^*.

Corollary 4: *If Assumptions **A.1'-A.7** hold, then,*

$$CvM_n^* \to_d \int_{\mathbb{R}^{p+1}} S_\infty^1(v,u)^2 K(dv,\, du),$$

$$KS_n^* \to_d \sup_{(v,u)\in\overline{\mathbb{R}}^{p+1}} |S_\infty^1(v,u)|,$$

$$\overline{CvM}_n^* \to_d \int_{\mathbb{R}^{p+1}} R_\infty^1(w,u)^2 H(dw,\, du),$$

and

$$\overline{KS}_n^* \to_d \sup_{(w,u)\in\Pi_c} |R_\infty^1(w,u)|,$$

where S_∞^1 and R_∞^1 are the same Gaussian processes stated in Corollary 1 and Corollary 2, respectively.

As before, the proof of Corollary 4 can be obtained straightforwardly by exploiting Theorem 4 and the continuous mapping theorem.

2.6 Monte Carlo Simulations

In this section, the finite sample performance of the proposed tests is studied by means of Monte Carlo experiments.

Let us consider a general nonparametric dynamic model $Y_t = m(I_t) + \varepsilon_t$ without exogenous explanatory variables. Recall that our primary interest is whether the conditional distribution of Y_t given I_t is symmetric around some function $m(I_t)$, or equivalently, whether the conditional distribution of ε_t given I_t is symmetric around zero. We consider the cases where $I_t = Y_{t-1}$. Specifically, the conditional location function $m(u)$ is specified according to the following four basic types of data generating processes:

DGP$_1$: $m(Y_{t-1}) = \mu$ and ε_t is $i.i.d.$;

DGP$_2$: $m(Y_{t-1}) = \alpha + \beta Y_{t-1}$ and ε_t is $i.i.d.$;

DGP$_3$: $m(Y_{t-1})$ is a smooth nonlinear function of Y_{t-1} and ε_t is $i.i.d.$;

or

DGP$_4$: $m(Y_{t-1}) = \mu$ and $\varepsilon_t = \sigma_t e_t$ is conditionally heteroskedastic, where e_t is a sequence of $i.i.d.$ random variables with zero mean and unit variance, and σ_t^2 is generated by a GARCH (p, q) process.

In the sequel, let ε_t be $i.i.d.$ random variables according to symmetric or asymmetric distributions. To examine the size accuracy of the test under the null hypothesis, we consider the following designs:

(S1) $Y_t \sim i.i.d.\ N(0, 1)$.

(S2) $Y_t \sim i.i.d.\ t_5$.

(S3) $Y_t \sim i.i.d.\ e_1 1(Z \leq 0.5) + e_2 1(Z > 0.5)$ with $e_1 \sim i.i.d.\ N(-1, 1)$, $e_2 \sim i.i.d.\ N(1, 1)$ and $Z \sim i.i.d.\ U(0, 1)$ mutually independent.

(S4) $Y_t = 0.5Y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ N(0, 1)$.

(S5) $Y_t = 0.5Y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ t_5$.

(S6) $Y_t = 0.5Y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ e_1 1(Z \leq 0.5) + e_2 1(Z > 0.5)$ with $e_1 \sim i.i.d.\ N(-1, 1)$, $e_2 \sim i.i.d.\ N(1, 1)$ and $Z \sim i.i.d.\ U(0, 1)$ mutually independent.

(S7) $Y_t = 0.23Y_{t-1}(1.6 - Y_{t-1}) + 0.4\varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ N(0, 1)$ truncated in the interval $[-12, 12]$.

(S8) $Y_t = 0.5Y_{t-1} + \varepsilon_t$, where $\varepsilon_t = \sigma_t e_t$, $\sigma_t^2 = \phi_0 + \phi_1 \sigma_{t-1}^2 + \phi_2 \varepsilon_{t-1}^2$, $e_t \sim i.i.d.\ N(0, 1)$; $\phi_0 = 2$, $\phi_1 = 0.5$ and $\phi_2 = 0.3$ (S8.1); $\phi_0 = 2$, $\phi_1 = 0.9$ and $\phi_2 = 0.05$ (S8.2).

(S9) $Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \gamma_2 \varepsilon_{t-1} + \varepsilon_t$ with

$$\varepsilon_t = \lambda_t e_t, \quad e_t \sim t_{v_t},$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2 + \alpha_3 1(\varepsilon_{t-1} < 0)\varepsilon_{t-1}^2,$$

$$k_t = \beta_0 + \beta_1 k_{t-1} + \beta_2 \frac{\varepsilon_{t-1}^4}{\sigma_{t-1}^4} + \beta_3 1(\varepsilon_{t-1} < 0) \frac{\varepsilon_{t-1}^4}{\sigma_{t-1}^4},$$

$$v_t = \frac{2(2k_t - 3)}{k_t - 3},$$

$$\lambda_t = \sigma_t \sqrt{\frac{v_t - 2}{v_t}}.$$

We set $\gamma_0 = 0.1$, $\gamma_1 = 0.5$, $\gamma_2 = 0.2$, $\alpha_0 = 0.01$, $\alpha_1 = 0.8$, $\alpha_2 = 0.05$, $\alpha_3 = 0.2$, $\beta_0 = 3.5$, $\beta_1 = 0.2$, $\beta_2 = 0.5$, $\beta_3 = 0.3$.

In order to investigate the empirical power of the test in finite samples, we consider the following eight designs:

(P1) $Y_t = 0.5Y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ \chi_2^2$.

(P2) $Y_t = 0.5Y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ \exp(N(0,1))$.

(P3) $Y_t = 0.5Y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ -\ln(U(0,1))$.

(P4) $Y_t = 0.5Y_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.$ asymmetric λ-distribution with parameters $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.001$ and $\lambda_4 = -0.13$.

(P5) $Y_t = 0.5Y_{t-1}\exp(-0.5Y_{t-1}^2) + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ \chi_{(2)}^2$.

(P6) $Y_t = |Y_{t-1}|^{0.8} + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ \exp(N(0,1))$.

(P7) $Y_t = -Y_{t-1}/(1 + Y_{t-1}^2) + \varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ -\ln(U(0,1))$.

(P8) $Y_t = X_t - X_{t-1}$ with $X_t \sim i.i.d.\ \chi_{(2)}^2$.

(P9) $Y_t = X_t - X_{t-1}$ with $X_t \sim i.i.d.\ -\ln(U(0,1))$.

(P10) $Y_t = 1 + \varepsilon_t$, where $\varepsilon_t = \sigma_t e_t$, $\sigma_t^2 = \phi_0 + \phi_1\sigma_{t-1}^2 + \phi_2\varepsilon_{t-1}^2$, $e_t \sim i.i.d.$ asymmetric λ-distribution with parameters $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0001$ and $\lambda_4 = -0.17$.; $\phi_0 = 2$, $\phi_1 = 0.5$ and $\phi_2 = 0.3$ (P10.1); $\phi_0 = 2$, $\phi_1 = 0.9$ and $\phi_2 = 0.05$ (P10.2).

All designs except (S7), (S9), (P5)-(P7) are based on or slightly modified from those in Delgado and Escanciano (2007). The λ values in designs (P4) and (P10) are taken from Randles et al. (1980). Design (S7) is a quadratic AR(1) time series model considered in Hyndman and Yao (2002), which by construction is a nonlinear autoregressive model of order 1 too. It is interesting to notice that in this design the conditional distribution of Y_t given $I_t = Y_{t-m}$ is symmetric for $m = 1$ but not necessarily so for $m > 1$. In this case, the conditional center of symmetry is the quadratic function $m(x) = 0.23x(16 - x)$. Designs (P5)-(P7) are included to study the effects of nonlinear location functions under the alternatives. We also study the generalized autoregressive conditional heteroskedastic model of Bollerslev (1986), especially the GARCH(1,1) model, in designs (S8) and (P10). Two different sets of parameters are considered, among which the choice $(\phi_0, \phi_1, \phi_2) = (2, 0.9, 0.05)$ is close to being an IGARCH(1,1) model. It is interesting to remark that, under the alternatives in designs (P8) and (P9), unconditional symmetry is satisfied, but conditional symmetry does not hold. All parameter combinations considered were selected to make the results of our study comparable with those obtained by Delgado and Escanciano (2007), whenever this is possible.

In design (S9), we consider a new model for autoregressive conditional heteroskedasticity and kurtosis proposed by Brooks et al. (2005), where the conditional variance and conditional kurtosis are permitted to evolve separately through a time-varying degrees of freedom parameter. Notice that the last terms in the conditional variance and conditional kurtosis equations permit the next period values of these quantities to have asymmetric responses to the signs of the realized innovations from the previous period, in the style of Glosten et al (1993). The design (S9) is a design that specifically allows for the time-varying conditional higher order moments, e.g. conditional kurtosis, through an autoregressive fashion. Unlike all the other designs, where $E(\varepsilon_t^4|I_t)$ equals to a constant, for design (S9), we have that $E(\varepsilon_t^4|I_t) = g(I_t)$ is not constant, i.e. $g(I_t)$ a non-degenerate function. See also the closely related autoregressive conditional skewness model developed in Harvey and Siddique (1999, 2000) where $E(\varepsilon_t^3|I_t)$ is allowed to be time-varying.

Three sample sizes, n=50, 100, 200 are considered in the simulation study. For each design, we first generate $n + 200$ observations and then discard the first 200 observations to minimize the initial value effect. The Monte Carlo experiments are based on 500 replications and the bootstrap critical values are approximated by $B = 500$ bootstrap replications. We only report results for the nominal size of 5%. Results for the other nominal sizes are available from the authors upon request. Standard normal density is used as our kernel function K. In all the simulations, we have adopted bandwidth of the form $h = c \times n^{-0.2}$ as in the univariate nonparametric estimation and we choose optimal bandwidth h^* for estimation (however, not for testing) according to the rule-of-thumb suggestion of Bowman and Azzalini (1997). That is, optimal bandwidth $h^* = c^* \times n^{-0.2}$ is calculated by $h_1 = \text{median}(|x - \text{median}(x)|)/0.6745 \times (4/3/n)^{0.2}$, $h_2 = \text{median}(|y - \text{median}(y)|)/0.6745 \times (4/3/n)^{0.2}$ and $h^* = \sqrt{h_1 \times h_2}$, so that optimal c^* is implicitly decided.

Tables 1 and 2 report the percentage of rejections for designs (S1)-(S9). They show fairly accurate size properties of our proposed tests CvM_n and KS_n with the exception of desgin (S9), where both tests appear to have an over-sized problem. It will require relatively large sample size to guarantee good size performance in the case of time varying conditional kurtosis. As a matter of fact, when sample size reaches 300, the empirical sizes for both tests are already very close to the nominal size 5%. In Tables 3 and 4 we report the empirical power performance against the alternatives (P1)-(P10). In summary, the proposed tests CvM_n and KS_n, both implemented with the wild bootstrap procedure suggested in Section 5, perform quite well in terms of empirical size and empirical power and are preserving the intrinsic advantage of robust to various kinds of parametric specification for the location.

2.7 Application to Stock Indices

In this section, we revisit the real problem of asymmetric behaviour for stock returns. That is, we shall apply the proposed test to investigate whether losses are more likely than gains given the available information in stock markets.

Four important stock indices across the world including both developed and emerging markets, namely S&P 500 index, FTSE 100 index, Nikkei 225 index and Shanghai A-Share index, which represent very distinct maturity and regulation conditions of the corresponding stock market, are considered. The four indices series are collected using daily data from 1 January 2001 to 31 December 2004 with a total of 1045 observations after deleting all public holidays and non-trading days. Returns series are calculated by $r_{it} = \log(P_{it}/P_{i,t-1}) \times 100\%$, where P_{it} denotes the time series sequence for any of the four stock indices with $i = 1$ denoting S&P

500, $i = 2$ FTSE 100, $i = 3$ Nikkei 225, and $i = 4$ Shanghai A-Share. Figure 1 provides plots for both time series of the indices and returns. Table 5 reports some descriptive statistics for the returns series. The augmented Dickey-Fuller test for the indices indicates that there exists a unit root in all three index series but not in their returns series. All four returns series exhibit the well-recorded stylized facts of volatility clustering and high kurtosis indicating the existence of fat tails, and hence all of these returns are highly leptokurtic. For example, the returns series from the Shanghai A-Share index has an excess kurtosis more than 7. However, conclusions drawn from information about the skewness coefficients are somewhat mixed. S&P 500 returns and Shanghai returns are positively skewed and the other two returns series are negatively skewed. Surprisingly, the skewness coefficient of Shanghai A-Share is more than 0.8 in sharp comparison to the magnitudes of the other series, while Nikkei 225 only has a slightly negative skewness coefficient. We therefore suspect our tests can detect this abnormal behaviour in Shanghai A-Share. Moreover, Jarque-Bera test statistics indicate that they are highly non-normal. We also plot the kernel density estimates for the four returns series (but not reported here). Judging from the (unconditional) density plots, we find that, although they are highly unlikely to be normally distributed, they all appear to be unconditionally symmetric except returns from Shanghai index. But as we have demonstrated, unconditional symmetry does not necessarily imply conditional symmetry.

We shall focus on the case where only the first lagged value will predict the stock returns, i.e. the returns series follow a Markov structure of order one. Specifically, we consider the following nonlinear autoregressive process of order 1, NLAR(1),

$$r_{it} = m(r_{i,t-1}) + \varepsilon_t, \quad , i = 1, 2, 3, 4.$$

We are of interest to test whether r_{it}, for $i = 1, \ldots, 4$, is symmetric around the unknown center $m(r_{i,t-1})$ given the information $r_{i,t-1}$, at the significance level 5%. Apart from the proposed two-sided tests of CvM_n and KS_n, we also implement two one-sided tests of the Kolmogorov-Smirnov-type to determine whether the conditional distribution is skewed to the right or to the left, which are given by

$$KS_n^+ = \sup_{(v,u)\in\overline{\mathbb{R}}^{p+1}} \left(\sqrt{n}S_n(v,u)\right)$$

in order to test the null hypothesis of conditional symmetry against the one-sided alternatives skewing to the right, and

$$KS_n^- = \sup_{(v,u)\in\overline{\mathbb{R}}^{p+1}} \left(-\sqrt{n}S_n(v,u)\right)$$

in order to test conditional symmetry against the one-sided alternatives skewing to the left.

In Table 6, we summarize the results from the tests CvM_n, KS_n, KS_n^+ and KS_n^- assisted with the help of the wild bootstrap procedure proposed in Section 5. The implementation is as in the Monte Carlo simulations part. To facilitate interpretations we present both the bootstrapped critical values and bootstrapped p-values for the four returns series. We observe that, for returns from S&P 500, FTSE 100 and Nikkei 225, both two-sided tests CvM_n and KS_n fail to reject the hypothesis of conditional symmetry. Results support the conclusion that losses are equally likely than gains given the information of previous period returns for these three stock markets. These findings are further confirmed by the one-sided tests of KS_n^+ and KS_n^-. On the other hand, there is a strong indication of conditional asymmetric behaviour for returns from Shanghai A-Share stock market. Moreover, one-sided test KS_n^- suggests that the conditional distribution of r_{4t} given $r_{4,t-1}$ is skewed to the left, indicating that losses are more likely than gains in Chinese's stock market.

2.8 Conclusions

In this paper we have investigated a useful model specification tool for nonparametric innovations in a dynamic context, i.e. the symmetry of conditional distributions around a nonparametric location function. Therefore, the (conditional) center is not parametrically specified and robust to the misspecifcation of the regression function. Test statistics of Cramér-von Mises-type and Kolmogorov-Smirnov-type are proposed. They are able to detect nonparametric alternatives converging to the null hypothesis at the parametric rate $\sqrt{n}$, where n is the sample size. A wild bootstrap procedure is suggested to obtain the critical values and the validity of the resulting bootstrap assisted test is formally justified. Extensive Monte Carlo simulation indicates that the proposed tests work very well in fairly small sample sizes. An empirical application is conducted to examine whether losses are more likely that gains given the available information in four major stock markets.

2.9 References

Ahmad, I.A., Li, Q. (1997). "Testing symmetry of an unknown density function by kernel method", *Journal of Nonparametric Statistics*, 7, 279-293.

Aki, S. (1981). "Asymptotic distribution of a Cramér-von Mises type statistic for testing symmetry when the center is estimated", *Annals of Institute of Statistical Mathematics*, 33, 1-14.

Akritas, M. G., Van Keilegom, I. (2001). "Nonparametric estimation of the residual distribution", *Scandinavian Journal of Statistics*, 28, 549-567.

Antille, A., Kersting, G., Zucchini, W. (1982). "Testing symmetry", *Journal of the American Statistical Association*, 77, 639-646.

Bae, J., Levental, S. (1995). "Uniform CLT for Markov chains and its invariance principle: a martingale approach", *Journal of Theoretical Probability*, 8, 549-570.

Bai, J.S., Ng, S. (2001). "A test for conditional symmetry in time series models", *Journal of Econometrics*, 103, 225-258.

Belaire-Franch, J., Contreras, D. (2002). "A Pearson's test for symmetry with an application to the Spanish business cycle", *Spanish Economic Review*, 4, 221-238.

Bhattacharya, P.K., Gastwirth, J.L., Wright, A.L. (1982). "Two modified Wilkoxon tests for symmetry about an unknown location parameter", *Biometrika*, 69, 377-382.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley Series in Probability and Statistics. Wiley, New York.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

Bowman, A.W., Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, United Kingdom.

Brooks, C., Burke, S.P., Heravi, S., Persand G. (2005). "Autoregressive conditional kurtosis", *Journal of Financial Econometrics*, 3, 399-421.

Brunner, A.D. (1992). "Conditional asymmetries in real GNP: a seminonparametric approach", *Journal of Business and Economics Statistics*, 10, 65-72.

Butler, C.C. (1969). "A test for symmetry using the sample distribution function", *Annals of Mathematical Statistics*, 40, 2209-2210.

Campbell, J.Y., Hentschel, L. (1992). "No news is good news: An asymmetric model of changing volatility in stock returns", *Journal of Financial Economics*, 31, 281-318.

Chen, B., Hong, Y. (2010). "Characteristic function-based testing for multifactor continuous time Markov models via nonparametric regression", *Econometric Theory*, 26, 1115-1179.

Chen, B., Hong, Y. (2012). "Testing for the Markov property in time series", *Econometric Theory*, 28, 130-178.

Dehling, H., Philipp, W. (2002). "Empirical Process Techniques for Dependent Data", *In: Dehling, H., Mikosch, T., Sørensen, M. (Eds.), Empirical Process Techniques for Dependent Data*, Birkhaüser, Basel, pp. 3-113.

DeLong, J.B., Summers, L.H. (1986). "Are business cycles symmetrical", *In: Gordon, R. (Eds.), American Business Cycles: Continuity and Change*, National Bureau of Economic Research and University of Chicago Press, Chicago, pp. 166-179.

Delgado, M.A., Escanciano, J.C. (2007). "Nonparametric tests for conditional symmetry in dynamic models", *Journal of Econometrics*, 141, 652-682.

Delgado, M.A., González-Manteiga, W. (2001). "Significance testing in nonparametric regression based on the bootstrap", *Annals of Statistics*, 29, 1469-1507.

Dette, H., Kusi-Appiah, S., Neumeyer, N. (2002). "Testing symmetry in nonparametric regression models", *Journal of Nonparametric Statistics*, 14, 477-494.

Doksum, K., Fenstad, G., Aaberge, R. (1977). "Plots and tests for symmetry", *Biometrika*, 64, 473-487.

Doukhan, P. (1994). *Mixing: Properties and Examples*, Springer-Verlag, New York.

Epps, T.W., Pulley, L.B. (1983). "A test for normality based on the empirical characteristic function", *Biometrika*, 70, 723-726.

Fan, J., Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York.

Fan, Y., Gencay, R. (1995). "A consistent nonparametric test of symmetry in linear regression models", *Journal of the American Statistical Association*, 90, 551-557.

Fan, Y., Li, Q. (1999). "Root-n-consistent estimation of partially linear time series models", *Journal of Nonparametric Statistics*, 10, 245-271.

Gastwirth, J.L. (1971) "On the sign test for symmetry", *Journal of the American Statistical Association*, 66, 821-828.

Gao, J., Hong, Y. (2008). "Central limit theorems for generalized U-statistics with applications in nonparametric specification", *Journal of Nonparametric Statistics*, 20, 61-76.

Ghosh, S., Ruymgaart, F. (1992). "Applications of empirical characteristic functions in some multivariate problems", *Canadian Journal of Statistics*, 20, 429-440.

Glosten, L.R., Jagannathan, R., Runkle, D.E. (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks", *Journal of Finance*, 48, 1779-1801.

Gupta, M.K. (1967). "An asymptotically nonparametric test of symmetry", *Annals of Mathematical Statistics*, 38, 849-866.

Hansen, B.E. (2008). "Uniform convergence rates for kernel estimation with dependent data", *Econometric Theory*, 24, 726-748.

Härdle, W., Mammen, E. (1993). "Comparing nonparametric versus parametric regression fits", *Annals of Statistics*, 21, 1926-1947.

Harvey, C.R., Siddique, A. (1999). "Autoregressive conditional skewness", *Journal of Financial and Quantitative Analysis*, 34, 465-487.

Harvey, C.R., Siddique, A. (2000). "Conditional skewness in asset pricing tests", *Journal of Finance*, 55, 1263-1295.

Heathcote, C., Rachev, S., Cheng, B. (1995). "Testing multivariate symmetry", Journal of Multivariate Analysis, 54, 91-112.

Hjellvik, V., Yao, Q., Tjøstheim, D. (1998) "Linearity testing using local polynomial approximation", *Journal of Statistical Planning and Inference*, 68, 295-321.

Hollander, M. (1988). "Testing for symmetry", *In: Encyclopedia of Statistical Sciences,* *Volume 9*, Wiley, New York, pp. 211-216.

Hušková, M. (1984). "Hypothesis of symmetry", *In: Handbook of Statistics, Nonparametric Methods, Volume 4*, North Holland, Amsterdam, pp. 63-78.

Hussey, R. (1992). "Nonparametric evidence on asymmetry in business cycles using aggregate employment time series", *Journal of Econometrics*, 51, 217-231.

Hyndman, R., Yao, Q. (2002). "Nonparametric estimation and symmetry tests for conditional density functions", *Journal of Nonparametric Statistics*, 14, 1-16.

Jarque, C., Bera, A. (1980). "Efficient tests for normality homoscedasticity and serial independence of regression residuals", *Economic Letters*, 6, 255-259.

Koziol, J.A. (1985). "A note on testing symmetry with estimated parameters", *Statistics and Probability Letters*, 3, 227-230.

Kristensen, D. (2009). "Uniform convergence rates of kernel estimators with heterogeneous dependent data", *Econometric Theory*, 25, 1433-1445.

Levental, S. (1989). "A uniform CLT for uniformly bounded families of martingale differences", *Journal of Theoretical Probability*, 2, 271-287.

Li, Q., Wang, S. (1998). "A simple consistent bootstrap test for a parametric regression function", *Journal of Econometrics*, 87, 145-165.

Liu, R.Y. (1988). "Bootstrap procedures under some non-i.i.d models", *Annals of Statistics*, 16, 1696-1708.

Masry, E. (1996). "Multivariate local polynomial regression for time series: Uniform strong consistency and rates", *Journal of Time Series Analysis*, 17, 571-599.

Nadaraya, E.A. (1964). "On estimating regression", *Theory of Probability and its Applications*, 9, 141-142.

Neumeyer, N., Dette, H. (2007). "Testing for symmetric error distribution in nonparametric regression models", *Statistica Sinica*, 17, 775-795.

Neumeyer, N., Van Keilegom, I. (2010). "Estimating the error distribution in nonparametric multiple regression with applications to model testing", *Journal of Multivariate Analysis*, 101, 1067-1078.

Nishiyama, Y. (2000). “Weak convergence of some classes of martingales with jumps”, *Annals of Probability*, 28, 685-712.

Ojeda, J.L. (2008). “Hölder continuity properties of the local polynomial estimator”, *Working paper*, Universidad de Zaragoza.

Psaradakis, Z. (2003). “A bootstrap test for symmetry of dependent data based on a Kolmogorov Smirnov type statistic”, *Communications in Statistics - Simulation and Computation*, 32, 113-126.

Randles, R.H., Fligner, M.A., Policello, G.E., Wolfe, D.A. (1980). “An asymptotically distribution-free test for symmetry versus asymmetry”, *Journal of the American Statistical Association*, 75, 168-172.

Robinson, P.M. (1988). “Root-N-consistent semiparametric regression”, *Econometrica*, 56, 931-954.

Said, S.E., Dickey, D.A. (1984). “Testing for unit roots in autoregressive-moving average models of unknown order”, *Biometrika*, 71, 599-607.

Schuster, E.F., Barker, R.C. (1987). “Using the bootstrap in testing symmetry versus asymmetry”, *Communications in Statistics - Simulation and Computation*, 16, 69-84.

Smirnov, N.V. (1947). “On criteria for the symmetry of distribution laws of random variables (in Russian)”, *Doklady Akademii Nauk SSSR*, 56, 11-14.

Stute, W. (1994). “U-statistic processes: A martingale approach”, *Annals of Probability*, 22, 1725-1744.

Stute, W., González-Manteiga, W., Presedo-Quindimil, M. (1998). “Bootstrap approximations in model checks for regression”, *Journal of the American Statistical Association*, 93, 141-149.

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

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

van der Vaart, A.W., Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.

Verbrugge, R.J. (1997). “Investigating cyclical asymmetries”, *Studies in Nonlinear Dynamics and Econometrics*, 2, 15-22.

Wang, X., Hong, Y. (2012). “Characteristic function based testing for conditional independence: A nonparametric regression approach”, *Working Paper*, Xiamen University.

Watson, G.S. (1964). “Smooth regression analysis”, *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 359-372.

Whang, Y.-J. (2000). “Consistent bootstrap tests of parametric regression functions”, *Journal of Econometrics*, 98, 27-46.

Wu, C.F.J. (1986). “Jacknife, bootstrap, and other resampling methods in regression analysis”, *Annals of Statistics*, 14, 1261-1295.

Yoshihara, K. (1976). “Limiting behavior of U-statistics for stationary, absolutely regular processes”, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwande Gebiete*, 35, 237-252.

Table 2.1: Empirical size of CvM_n at 5%

	$h = c^* \times n^{-0.2}$									
	(S1)	(S2)	(S3)	(S4)	(S5)	(S6)	(S7)	(S8.1)	(S8.2)	(S9)
$n = 50$	5.2	5.7	4.5	5.3	7.7	5.5	5.7	4.2	6.0	9.4
$n = 100$	5.1	5.1	4.0	6.4	7.0	6.8	6.9	5.8	6.4	10.8
$n = 200$	5.1	5.7	4.8	5.9	6.5	4.9	6.0	6.2	6.6	8.8

Table 2.2: Empirical size of KS_n at 5%

	$h = c^* \times n^{-0.2}$									
	(S1)	(S2)	(S3)	(S4)	(S5)	(S6)	(S7)	(S8.1)	(S8.2)	(S9)
$n = 50$	4.0	4.8	2.8	3.6	6.4	5.0	5.6	3.8	4.8	7.0
$n = 100$	5.4	4.0	3.4	5.4	5.8	5.4	6.8	4.8	4.8	8.8
$n = 200$	4.6	6.0	3.6	3.4	6.0	7.4	5.2	4.0	5.4	8.8

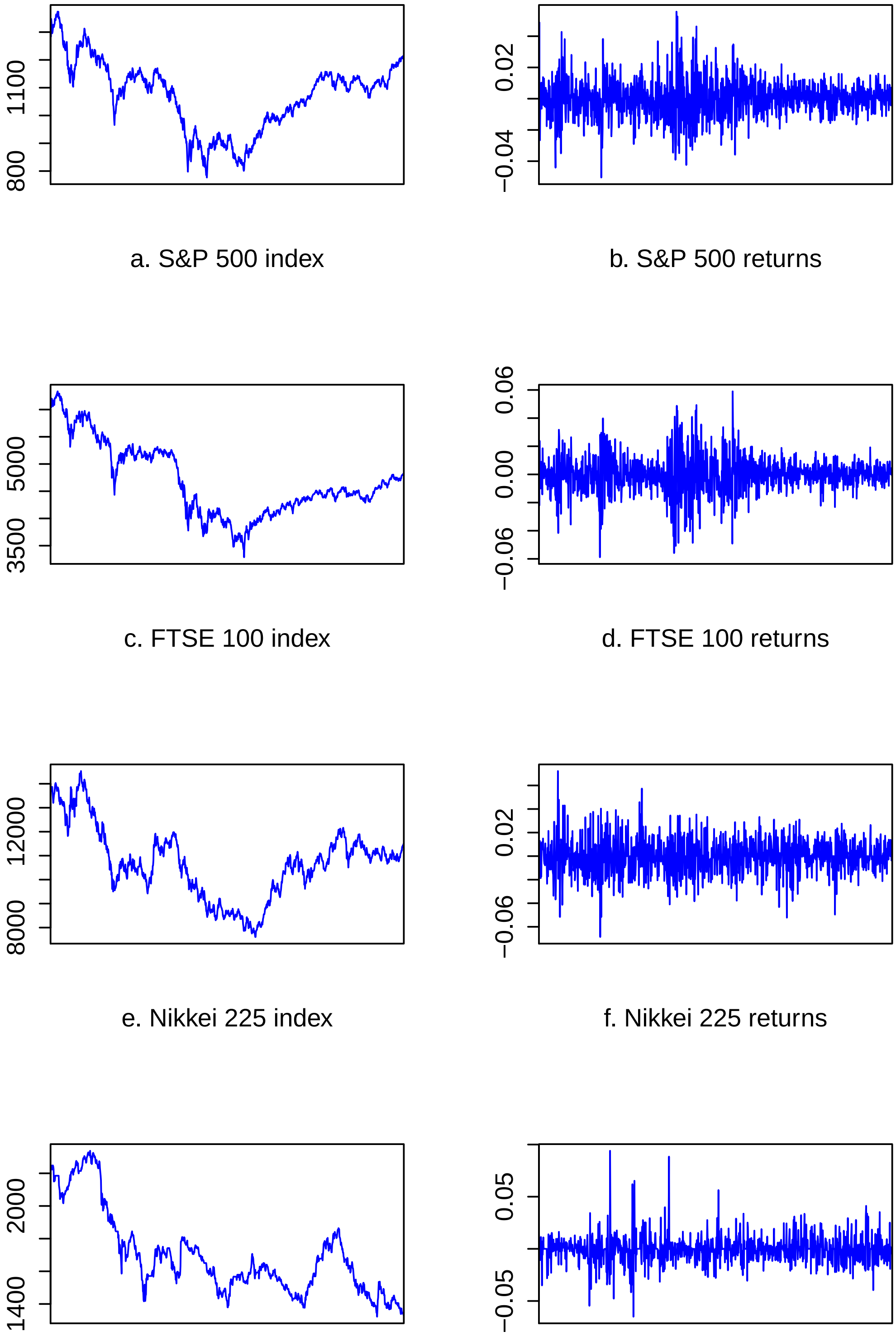
Table 2.3: Empirical power of CvM_n at 5%

	$h = c^* \times n^{-0.2}$										
	(P1)	(P2)	(P3)	(P4)	(P5)	(P6)	(P7)	(P8)	(P9)	(P10.1)	(P10.2)
$n = 50$	93.1	99.0	92.2	96.0	91.8	98.3	86.3	41.4	39.2	96.8	97.4
$n = 100$	99.9	100.0	100.0	100.0	99.9	100.0	99.6	68.6	71.2	100.0	100.0
$n = 200$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	94.6	96.6	100.0	100.0

Table 2.4: Empirical power of KS_n at 5%

	$h = c^* \times n^{-0.2}$										
	(P1)	(P2)	(P3)	(P4)	(P5)	(P6)	(P7)	(P8)	(P9)	(P10.1)	(P10.2)
$n = 50$	89.0	98.2	88.4	93.6	88.0	97.4	86.6	47.4	46.6	93.8	95.4
$n = 100$	100.0	100.0	99.4	99.8	100.0	100.0	99.2	77.6	80.6	100.0	100.0
$n = 200$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.8	99.0	100.0	100.0

Figure 2.1: Time series plots for stock indices and returns



a. S&P 500 index

b. S&P 500 returns

c. FTSE 100 index

d. FTSE 100 returns

e. Nikkei 225 index

f. Nikkei 225 returns

g. Shanghai index

h. Shanghai returns

Table 2.5: Descriptive statistics for stock returns

	01/01/2001-31/12/2004			
	S&P 500	FTSE 100	Nikkei 225	Shanghai
Mean	-0.0001	-0.0002	-0.0002	-0.0007
Std	0.0122	0.0128	0.0149	0.0129
Skewness	0.1740	-0.1536	-0.0151	0.8052
Kurtosis	5.0627	5.9779	4.4240	10.1822
JB	190.3467	389.8607	88.2508	2356.7000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ADF	-33.5576	-34.3813	-33.3503	-31.8028
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Note: ADF denotes the augmented Dickey-Fuller test, e.g. Said and Dickey (1984). JB denotes the Jarque-Bera test for normality proposed by Jarque and Bera (1980). p-values are reported in parentheses.

Table 2.6: Conditional symmetry tests for stock returns

	S&P 500		FTSE 100		Nikkei 225		Shanghai	
	Statistics	p-values	Statistics	p-values	Statistics	p-values	Statistics	p-values
CvM_n	0.1864	0.1140	0.1936	0.1470	0.0727	0.6480	0.5327	0.0490
	(0.2601)		(0.3073)		(0.2787)		(0.5251)	
KS_n	1.2386	0.2870	1.7959	0.0840	1.1457	0.5450	2.8797	0.0160
	(1.7340)		(2.0127)		(2.1830)		(2.5081)	
KS_n^+	0.8051	0.4880	0.6193	0.6600	1.1457	0.2240	0.9908	0.5930
	(1.6721)		(1.8578)		(1.8424)		(2.3223)	
KS_n^-	1.2386	0.1290	1.7959	0.0450	0.5574	0.8520	2.8797	0.0020
	(1.5482)		(1.7650)		(2.0746)		(2.1520)	

Note: Bootstrapped critical values at level 5% are reported in parentheses.

2.10 Appendix A

We provide the proofs of the main theoretical results in this appendix. Denote $\hat{F}_\varepsilon(v) = n^{-1}\sum_{t=1}^{n} 1(\varepsilon_t \leq v)$ and $\hat{F}_{\hat{\varepsilon}}(v) = n^{-1}\sum_{t=1}^{n} 1(\hat{\varepsilon}_t \leq v)$. The following lemmas are needed to establish the asymptotic theory for $S_n(v,u)$.

Lemma A.1: *Under Assumptions **A.1'-A.6**, then,*

$$\|\hat{m} - m\|_{p+\alpha} = o_p(1),$$

*where $0 < \alpha < \delta/2$, δ is defined as in Assumption **A.5**, and where for any function f defined on $\mathbb{R}^p$,*

$$\|f\|_{p+\alpha} = \max_{k.\leq p}\sup_u |D^k f(u)| + \max_{k.=p}\sup_{u,u'} \frac{|D^k f(u) - D^k f(u')|}{\|u-u'\|^\alpha}$$

with $k = (k_1, \ldots, k_p)$,

$$D^k = \frac{\partial^{k.}}{\partial u_1^{k_1}\cdots\partial u_p^{k_p}},$$

$k. = \sum_{j=1}^{p} k_j$, and $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^p$.

Proof of Lemma A.1: It follows from Theorem 8 in Hansen (2008) or Theorem 1 in Kristensen (2009) that

$$\sup_u |\hat{m}(u) - m(u)| = O_p((nh^p)^{-1/2}(\log n)^{1/2}) + O(h^L),$$

which is $o_p(1)$ by Assumption A.5.

On the basis of this result it can be shown that for $k. \leq p$,

$$\sup_u \left|D^k\hat{m}(u) - D^k m(u)\right| = O_p((nh^{p+2k.})^{-1/2}(\log n)^{1/2}) + O(h^{L-k.}) = o_p(1),$$

and for $k. = d$,

$$\sup_{u,u'} \frac{|D^k\hat{m}(u) - D^k m(u) - D^k\hat{m}(u') + D^k m(u')|}{\|u-u'\|^\alpha}$$
$$= O_p((nh^{3p+2\alpha})^{-1/2}(\log n)^{1/2}) + O(h^{L-p-\alpha}) = o_p(1),$$

See Proposition 3.2 and Theorem 3.2 in Ojeda (2008) for a detailed proof for the case $p = 1$. For the $p \geq 2$ case, the proof is similar but more technical, see also Masry (1996) and Hansen (2008). The rest of the proof is therefore omitted. □

Lemma A.2: *Under Assumptions **A.1'-A.6**,*

$$\sup_v \left| \hat{F}_{\hat{\varepsilon}}(v) - \hat{F}_{\varepsilon}(v) - F_{\hat{\varepsilon}}(v) + F_{\varepsilon}(v) \right| = o_p(n^{-1/2}),$$

and

$$\sup_{v,u} \left| \frac{1}{n} \sum_{t=1}^{n} \left[1(\hat{\varepsilon}_t \leq v) - 1(\varepsilon_t \leq v) - F_{\hat{\varepsilon}}(v) + F_{\varepsilon}(v)\right] \left[1(I_t \leq u) - F_I(u)\right] \right| = o_p(n^{-1/2}),$$

where $F_{\hat{\varepsilon}}(v)$ is the distribution of residuals $\hat{\varepsilon} = Y - \hat{m}(I)$ conditional on the data $\{Y_t, I_t\}_{t=1}^n$, i.e., considering $\hat{m}$ as a fixed function, and where $F_{\varepsilon}(v)$ is the distribution of errors $\varepsilon = Y - m(I)$.

Proof of Lemma A.2: The proof is similar to that of Lemma 1 in Akritas and Van Keilegom (2001). We will show the first statement. The second one can be proved in a similar way. The first statement above can be expressed as the following:

$$\begin{aligned}
&\hat{F}_{\hat{\varepsilon}}(v) - \hat{F}_{\varepsilon}(v) - F_{\hat{\varepsilon}}(v) + F_{\varepsilon}(v) \\
&= \frac{1}{n} \sum_{t=1}^{n} \{1(\hat{\varepsilon}_t \leq v) - 1(\varepsilon_t \leq v) - P(\hat{\varepsilon} \leq v | (Y, I)) + P(\varepsilon \leq v)\} \\
&= \frac{1}{n} \sum_{t=1}^{n} \{1(\varepsilon_t \leq v + d_n(I_t)) - 1(\varepsilon_t \leq v) - P(\varepsilon \leq v + d_n(I)) + P(\varepsilon \leq v)\}
\end{aligned}$$

where $d_n(I_t) = \hat{m}(I_t) - m(I_t)$. We will focus our attention on $1(\varepsilon_t \leq v + d_n(I_t)) - 1(\varepsilon_t \leq v) - P(\varepsilon \leq v + d_n(I)) + P(\varepsilon \leq v)$. The following proof is mostly based on results from van der Vaart and Wellner (1996). Let

$$\begin{aligned}
\mathscr{F} = \{(u, e) \to 1(e \leq v + d(u)) - 1(e \leq v) - P(\varepsilon \leq v + d(u)) + P(\varepsilon \leq v); \\
-\infty < v < \infty, d \in C^{p+\alpha}(R_I)\},
\end{aligned}$$

where $C^{p+\alpha}(R_I)$ is the class of p-times-differentiable functions d defined on support R_I of I such that $\|d\|_{p+\alpha} \leq 1$ with $\|d\|_{p+\alpha}$ defined in Lemma A.1. First notice that Lemma A.1 implies that $P(d_n \in C^{p+\alpha}(R_I)) \to 1$ as $n \to \infty$ with $d_n(u) = \hat{m}(u) - m(u)$. It follows that it suffices to show that the class $\mathscr{F}$ is Donsker, i.e. we will establish the weak convergence of $n^{-1/2} \sum_{t=1}^{n} f(\varepsilon_t, I_t)$, $f \in \mathscr{F}$, and that the variance of $1(\varepsilon \leq v + d_n(I)) - 1(\varepsilon \leq v) - P(\varepsilon \leq v + d_n(I)) + P(\varepsilon \leq v)$ tends to zero, uniformly in v.

In a first step, we will show that the class $\mathscr{F}$ is Donsker. According to Theorem 2.10.6 in van der Vaart and Wellner (1996), we can deal with the four terms in the definition of $\mathscr{F}$ separately. It suffices to prove that the class

$$\mathscr{F}_1 = \{(u, e) \to 1(e \leq v + d(u)); -\infty < v < \infty, d \in C^{p+\delta}(R_I)\}$$

is Donsker, since the other terms are similar, but much easier. This is done by verifying the condition of Theorem 2.5.6 in van der Vaart and Wellner (1996):

$$\int_0^\infty \sqrt{\log N_{[\,]}(\bar{\varepsilon}, \mathscr{F}_1, L_2(P))}\, d\bar{\varepsilon} < \infty,$$

where $N_{[\,]}(\bar{\varepsilon}, \mathscr{F}_1, L_2(P))$ is the $\bar{\varepsilon}^2$-bracketing number of the class $\mathscr{F}_1$, i.e., the smallest number of balls of $L_2(P)$-radius $\bar{\varepsilon}$ needed to cover $\mathscr{F}_1$, P is the probability measure corresponding to the joint distribution of (ε, I), and where $L_2(P)$ is the L_2-norm. To this end, by Theorem 2.7.1 of van der Vaart and Wellner (1996), for any $\bar{\varepsilon} > 0$, the $\bar{\varepsilon}^2$-bracketing numbers of the class $C^{p+\alpha}(R_I)$ are bounded by

$$m = N_{[\,]}(\bar{\varepsilon}^2, C^{p+\alpha}(R_I), L_2(P)) \leq \exp(K\bar{\varepsilon}^{-2p/(p+\alpha)}),$$

where $K > 0$. Let $d_1^L \leq d_1^U, \ldots, d_m^L \leq d_m^U$ be the functions defining the m brackets for $C^{p+\alpha}(R_I)$. For each d and each fixed v, we have:

$$1(\varepsilon \leq v + d_i^L(I)) \leq 1(\varepsilon \leq v + d(I)) \leq 1(\varepsilon \leq v + d_i^U(I)).$$

Define $F_i^L(v) = P(\varepsilon \le v + d_i^L(I))$ and let $-\infty = v_{i1}^L < v_{i2}^L < \ldots < v_{i,m_L}^L = \infty$ ($m_L = O(\bar{\varepsilon}^{-2})$) partition the line in segments having F_i^L-probability less than or equal to a fraction of $\bar{\varepsilon}^2$.

Similarly, define $F_i^U(v) = P(\varepsilon \le v + d_i^U(I))$ and let $-\infty = v_{i1}^U < v_{i2}^U < \ldots < v_{i,m_U}^U = \infty$ ($m_U = O(\bar{\varepsilon}^{-2})$) partition the line in segments having F_i^U-probability less than or equal to a fraction of $\bar{\varepsilon}^2$. Now define the following bracket for v:

$$v_{ik_1}^L \le v \le v_{ik_2}^U,$$

where $v_{ik_1}^L$ is the largest of the v_{ik}^L with the property of being less than or equal to v and $v_{ik_2}^U$ is the smallest of the v_{ik}^U with the property of being greater than or equal to v. We will show that

$$1(\varepsilon \le v_{ik_1}^L + d_i^L(I)) \le 1(\varepsilon \le v + d(I)) \le 1(\varepsilon \le v_{ik_2}^U + d_i^U(I)).$$

Since we have

$$\begin{aligned}
&\|1(\varepsilon \le v_{ik_2}^U + d_i^U(I)) - 1(\varepsilon \le v_{ik_1}^L + d_i^L(I))\|_2^2 \\
&= F_i^U(v_{ik_2}^U) - F_i^L(v_{ik_1}^L) \\
&= F_i^U(v) - F_i^L(v) + K\bar{\varepsilon}^2,
\end{aligned}$$

and also by applying a Taylor expansion to the function F_ε we get

$$\begin{aligned}
&F_i^U(v) - F_i^L(v) \\
&= \int [F_\varepsilon(v + d_i^U(u)) - F_\varepsilon(v + d_i^L(u))]\, dF_I(u) \\
&= \int f_\varepsilon(v + \bar{d}(u))[d_i^U(u) - d_i^L(u)]\, dF_I(u),
\end{aligned}$$

where $\bar{d}(u)$ is between $d_i^U(u)$ and $d_i^L(u)$, the above expression is bounded in absolute value by $K_1\|d_i^U - d_i^L\|_{P,1} \le K_1\bar{\varepsilon}^2$. Hence, for the class $\mathscr{F}_1$ and for each $\bar{\varepsilon} > 0$, we have at most $O(\bar{\varepsilon}^{-2}\exp(K\bar{\varepsilon}^{-2p/(p+\alpha)}))$ brackets in total. However, for $\bar{\varepsilon} > 1$, one bracket is enough. So we have,

$$\int_0^\infty \sqrt{\log N_{[\,]}(\bar{\varepsilon}, \mathscr{F}_1, L_2(P))}\, d\bar{\varepsilon} < \infty.$$

This shows that the class $\mathscr{F}_1$ (and hence $\mathscr{F}$) is Donsker.

Next, we show that the variance of $1(\varepsilon \le v + d_n(I)) - 1(\varepsilon \le v) - P(\varepsilon \le v + d_n(I)) + P(\varepsilon \le v)$ tends to zero, uniformly in v. We calculate

$$\begin{aligned}
&\text{Var}(1(\varepsilon \le v + d_n(I)) - 1(\varepsilon \le v) - P(\varepsilon \le v + d_n(I)) + P(\varepsilon \le v)) \\
&= \text{Var}(1(\varepsilon \le v + d_n(I)) - 1(\varepsilon \le v)) \\
&\le E[E(\{1(\varepsilon \le v + d_n(I)) - 1(\varepsilon \le v)\}^2 | (Y, I))] \\
&= E[F_\varepsilon(v + d_n(I)) - F_\varepsilon(\min(v, v + d_n(I)))] \\
&+ E[F_\varepsilon(v) - F_\varepsilon(\min(v, v + d_n(I)))] \\
&= E\left[f_\varepsilon(v + \theta d_n(I)) |d_n(I)|\right],
\end{aligned}$$

for some $0 < \theta < 1$ by Taylor expansion of F_ε. By Assumption **A.7**, $\sup_{-\infty < v < \infty} f_\varepsilon(v) \le C < \infty$, the above expression is bounded by

$$CE|d_n(I)| \le C\|\hat{m}(u) - m(u)\|_{p+\alpha} \to_p 0$$

by Lemma A.1. Since the class $\mathscr{F}$ is Donsker, it follows from Corollary 2.3.12 in van der Vaart and Wellner (1996) that

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P\left(\sup_{f \in \mathscr{F}, \text{Var}(f) < \delta} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n f(\varepsilon_t, I_t) \right| > \bar{\varepsilon} \right) = 0$$

for every $\bar{\varepsilon} > 0$ By restricting the supremum inside this probability to the elements in $\mathscr{F}$ corresponding to $d(I) = d_n(I)$ as defined above, the result follows. $\square$

Lemma A.3: *Under Assumptions **A.1'-A.6**,*

$$\sup_v \left| \hat{F}_{-\hat{\varepsilon}}(v) - \hat{F}_{-\varepsilon}(v) - F_{-\hat{\varepsilon}}(v) + F_{-\varepsilon}(v) \right| = o_p(n^{-1/2}),$$

and

$$\sup_{v,u}\left|\frac{1}{n}\sum_{t=1}^{n}\left[1(-\hat{\varepsilon}_t\le v)-1(-\varepsilon_t\le v)-F_{-\hat{\varepsilon}}(v)+F_{-\varepsilon}(v)\right]\left[1(I_t\le u)-F_I(u)\right]\right|$$
$$=o_p(n^{-1/2}),$$

where $\hat{F}_{-\hat{\varepsilon}}(v)$ and $F_{-\hat{\varepsilon}}(v)$ are defined in a similar way to $\hat{F}_{\hat{\varepsilon}}(v)$ and $F_{\hat{\varepsilon}}(v)$ in Lemma A.2.

Proof of Lemma A.3: The proof of Lemma A.3 is identical to that of Lemma A.2 simply with ε replaced by $-\varepsilon$ and $\hat{\varepsilon}$ replaced by $-\hat{\varepsilon}$, and is therefore omitted. □

In the sequel, denote $x=(v,u)\in\mathbb{R}^{p+1}$. Notice that $S_n(x)=S_n^0(x)+W_n(x)$ with

$$W_n(x)=\frac{1}{n}\sum_{t=1}^{n}\left[\hat{\omega}_t(v)-\omega_t(v)\right]1(I_t\le u).$$

We further can decompose $W_n(x)$ into the following six components,

$$W_n(x):=(W_{n1}(x)+W_{n2}(x)+W_{n3}(x))-(W_{n4}(x)+W_{n5}(x)+W_{n6}(x))$$

with

$$W_{n1}(x)=F_I(u)\frac{1}{n}\sum_{t=1}^{n}\left[1(\hat{\varepsilon}_t\le v)-1(\varepsilon_t\le v)-F_{\hat{\varepsilon}}(v)+F_{\varepsilon}(v)\right],$$

$$W_{n2}(x)=\frac{1}{n}\sum_{t=1}^{n}\left[1(\hat{\varepsilon}_t\le v)-1(\varepsilon_t\le v)-F_{\hat{\varepsilon}}(v)+F_{\varepsilon}(v)\right]\left[1(I_t\le u)-F_I(u)\right],$$

$$W_{n3}(x)=\frac{1}{n}\sum_{t=1}^{n}\left[F_{\hat{\varepsilon}}(v)-F_{\varepsilon}(v)\right]1(I_t\le u),$$

$$W_{n4}(x) = F_I(u)\frac{1}{n}\sum_{t=1}^{n}\left[1(-\hat{\varepsilon}_t \leq v) - 1(-\varepsilon_t \leq v) - F_{-\hat{\varepsilon}}(v) + F_{-\varepsilon}(v)\right],$$

$$W_{n5}(x) = \frac{1}{n}\sum_{t=1}^{n}\left[1(-\hat{\varepsilon}_t \leq v) - 1(-\varepsilon_t \leq v) - F_{-\hat{\varepsilon}}(v) + F_{-\varepsilon}(v)\right]\left[1(I_t \leq u) - F_I(u)\right],$$

and

$$W_{n6}(x) = \frac{1}{n}\sum_{t=1}^{n}\left[F_{-\hat{\varepsilon}}(v) - F_{-\varepsilon}(v)\right]1(I_t \leq u).$$

Lemma A.4 below establishes the asymptotic behaviour of $W_{nj}(x)$ for $j = 1, \ldots, 6$. It turns out that $\sqrt{n}W_{nj}(x)$ for $j = 1, 2, 4, 5$ is asymptotically negligible uniformly in x, while both $\sqrt{n}W_{n3}(x)$ and $\sqrt{n}W_{n6}(x)$ admit an asymptotic representation composed of true errors ε_t.

Lemma A.4: *Assume **A.1'-A.6**, uniformly in* $x \in \mathbb{R}^{p+1}$,

$$W_{nj}(x) = o_p(n^{-1/2}), \quad j = 1, 2, 4, 5,$$

$$W_{n3}(x) = f_\varepsilon(v)\frac{1}{n}\sum_{t=1}^{n}(Y_t - m(I_t))1(I_t \leq u) + o_p(n^{-1/2}),$$

and

$$W_{n6}(x) = -f_\varepsilon(-v)\frac{1}{n}\sum_{t=1}^{n}(Y_t - m(I_t))1(I_t \leq u) + o_p(n^{-1/2}).$$

Proof of Lemma A.4: This first statement follows directly from Lemma A.2 and Lemma A.3. We now prove the second statement. Using $\hat{\varepsilon}_t = \varepsilon_t - (\hat{m}(I_t) - m(I_t))$, $W_{n3}(x)$ can be rewritten as

$$W_{n3}(x) = \frac{1}{n}\sum_{t=1}^{n}\left[F_\varepsilon\left(v+\hat{m}(I_t)-m(I_t)\right)-F_\varepsilon\left(v\right)\right]1(I_t\le u),$$

by Taylor expansion of F_ε, we get

$$\begin{aligned}W_{n3}(x) =& f_\varepsilon(v)\frac{1}{n}\sum_{t=1}^{n}\left(\hat{m}(I_t)-m(I_t)\right)1(I_t\le u)\\ &+\frac{1}{n}\sum_{t=1}^{n}\frac{1}{2}f'_\varepsilon(\bar{v})\left(\hat{m}(I_t)-m(I_t)\right)^2 1(I_t\le u)\end{aligned}$$

where $\bar{v}$ is between $v+\hat{m}(I_t)-m(I_t)$ and v, and $\bar{v}$ may depend on I_t. By Assumption **A.6**, $f'_\varepsilon(v)$ is uniformly bounded, the second term above is bounded by

$$C\cdot\frac{1}{n}\sum_{t=1}^{n}\left(\hat{m}(I_t)-m(I_t)\right)^2,$$

which is the the mean square errors (MSE) of the Nadaraya-Watson estimator $\hat{m}(I_t)$ for $m(I_t)$. It is standard to establish that MSE of $\hat{m}$ is $O_p((1/\sqrt{nh^p}+h^L)^2)=o_p(n^{-1/2})$ from Assumptions **A.3-A.5**, e.g. Masry (1996) or more recently Hansen (2008). Therefore,

$$\begin{aligned}&W_{n3}(x)\\ =&f_\varepsilon(v)\frac{1}{n}\sum_{t=1}^{n}(Y_t-m(I_t))1(I_t\le u)-f_\varepsilon(v)\frac{1}{n}\sum_{t=1}^{n}(Y_t-\hat{m}(I_t))1(I_t\le u)+o_p(n^{-1/2}),\end{aligned}$$

where the second term above is $o_p(n^{-1/2})$ uniformly in x according to Lemma B.2 and Assumption **A.6**.

The third statement can be proved in a similar way. Since

$$W_{n6}(x) = -\frac{1}{n}\sum_{t=1}^{n}\left[F_\varepsilon\left(-v+\hat{m}(I_t)-m(I_t)\right)-F_\varepsilon\left(-v\right)\right]1(I_t\le u),$$

the same reasoning as for $W_{n3}(x)$ applies. Hence, we finish the proof. □

Proof of Theorem 1: It is omitted. See the proof of Theorem A.1 and Theorem 1 in Delgado and Escanciano (2007). □

Proof of Theorem 2: Its proof consists of applying Lemma A.4 and the decomposition that $S_n(x) = S_n^0(x) + W_n(x)$. □

Proof of Theorem 3: Notice that under the local alternatives in (2.4),

$$\begin{aligned}
&E(\omega_t(v)|I_t)\\
=&E(1(\varepsilon_t \le v)|I_t) - 1 + E(1(\varepsilon_t \le -v)|I_t)\\
=&\int_{-\infty}^{\infty} 1(\bar{v} \le v) f^{(n)}(\bar{v}|I_t)\, d\bar{v} - 1 + \int_{-\infty}^{\infty} 1(\bar{v} \le -v) f^{(n)}(\bar{v}|I_t)\, d\bar{v}\\
=&\frac{1}{\sqrt{n}}[\Delta(v, I_t) - \Delta(\infty, I_t) + \Delta(-v, I_t)]\\
=&\frac{A_t(v)}{\sqrt{n}}
\end{aligned}$$

where $A_t(v) = \Delta(v, I_t) - \Delta(\infty, I_t) + \Delta(-v, I_t)$, and

$$\Delta(v, I_t) = \int_{-\infty}^{v} f(\bar{v}|I_t)\delta_n(\bar{v}, I_t)\, d\bar{v},$$

and the second equality follows by the symmetry of the density function $f(v|u)$ and the specification of local alternatives in (2.4).

Thus, we can rewrite

$$\begin{aligned}
&\sqrt{n}S_n(x)\\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\omega_t(v) - \frac{A_t(v)}{\sqrt{n}} + \hat{\omega}_t(v) - \omega_t(v) + \frac{A_t(v)}{\sqrt{n}}\right\} 1(I_t \le u)\\
:=&\sqrt{n}\tilde{S}_n(x) + A_{n1}(x) + A_{n2}(x),
\end{aligned} \tag{2.6}$$

where again $\hat{\omega}_t(v) = 1(\hat{\varepsilon}_t \le v) - 1(-\hat{\varepsilon}_t \le v)$,

$$\sqrt{n}\tilde{S}_n(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\omega_t(v) - \frac{A_t(v)}{\sqrt{n}}\right\} 1(I_t \le u),$$

$$A_{n1}(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\hat{\omega}_t(v) - \omega_t(v)\right] 1(I_t \leq u),$$

and

$$A_{n2}(x) = \frac{1}{n} \sum_{t=1}^{n} A_t(v) 1(I_t \leq u).$$

Let's first establish the asymptotic behaviour of $\sqrt{n}\tilde{S}_n$. It is easy to see that the term $\left\{\omega_t(v) - n^{-1/2}A_t(v)\right\} 1(I_t \leq u)$ is a zero mean square-integrable martingale difference sequence with respect to the filtration $\mathscr{F}_t$ for each (v, u). Thus, by Theorem A.1 in Delgado and Escanciano (2007), we can obtain

$$\sqrt{n}\tilde{S}_n \Rightarrow S_\infty$$

in $l^\infty(\overline{\mathbb{R}}^{p+1})$.

Similar to the arguments in Lemma A.4, we can show that, uniformly in x,

$$A_{n1}(x) = (f_\varepsilon(v) + f_\varepsilon(-v)) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} [Y_t - m(I_t)] 1(I_t \leq u) + o_p(1).$$

On the other hand, uniformly in x,

$$\begin{aligned} A_{n2}(x) &= \frac{1}{n} \sum_{t=1}^{n} A_t(v) 1(I_t \leq u) \\ &= E\left(A_1(v) 1(I_1 \leq u)\right) + o_p(1) \\ &= \Xi(v, u) - \Xi(\infty, u) + \Xi(-v, u) + o_p(1) \\ &:= \Xi^1(v, u) + o_p(1), \end{aligned}$$

where the second equality follows by the uniform ergodic theorem (UET) for stationary and ergodic sequences, see e.g. Dehling and Philipp (2002, p. 4), and where $\Xi_\varepsilon(u, v)$ is given by

$$\Xi(u, v) = \int_{-\infty}^{v} E[f(\bar{v}|I_1)\delta_n(\bar{v}, I_1) 1(I_1 \leq u)]\, d\bar{v}.$$

Thus, uniformly in (v, u), we have

$$\sqrt{n}S_n(v,u) = \sqrt{n}\tilde{S}_n(v,u) + (f_\varepsilon(v) + f_\varepsilon(-v))\,\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t 1(I_t \le u) + \Xi^1(v,u) + o_p(1),$$

where $\Xi^1(v,u) = \Xi(v,u) - \Xi(\infty,u) + \Xi(-v,u)$ is a deterministic shift function. So, by Slutsky's theorem, $\sqrt{n}S_n$ converges weakly to $S^1_\infty + \Xi^1$ in $l^\infty(\overline{\mathbb{R}}^{p+1})$ under the local alternatives H_{An} in (2.4). Hence, we finish the proof of Theorem 3. $\square$

Proof of Theorem 4: Denote $\omega^*_t(v) = 1(\varepsilon^*_t \le v) - 1(-\varepsilon^*_t \le v)$ with $\varepsilon^*_t = v_t\hat{\varepsilon}_t$, where $\{v_t\}_{t=1}^n$ is a sequence of random variables with zero mean and unit variance and also independent of the original sample. As before, we first decompose $\sqrt{n}S^*_n$ in the following way,

$$\begin{aligned}
&\sqrt{n}S^*_n(x)\\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\omega^*_t(v)1(I_t \le u) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\hat{\omega}^*_t(v) - \omega^*_t(v)\right]1(I_t \le u)\\
:=&\sqrt{n}S^*_{n1}(x) + \sqrt{n}S^*_{n2}(x).
\end{aligned}$$

First of all, for term $\sqrt{n}S^*_{n2}(x)$, we decompose

$$\sqrt{n}S^*_{n2}(x) = (B^*_{n1}(x) + B^*_{n2}(x) + B^*_{n3}(x)) - (B^*_{n4}(x) + B^*_{n5}(x) + B^*_{n6}(x)),$$

where

$$B^*_{n1}(x) = F_I(u)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[1(\hat{\varepsilon}^*_t \le v) - 1(\varepsilon^*_t \le v) - F_{\hat{\varepsilon}^*}(v) + F_{\varepsilon^*}(v)\right],$$

$$B^*_{n2}(x) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[1(\hat{\varepsilon}^*_t \le v) - 1(\varepsilon^*_t \le v) - F_{\hat{\varepsilon}^*}(v) + F_{\varepsilon^*}(v)\right]\left[1(I_t \le u) - F_I(u)\right],$$

$$B_{n3}^*(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[F_{\hat{\varepsilon}^*}(v) - F_{\varepsilon^*}(v) \right] 1(I_t \leq u),$$

$$B_{n4}^*(x) = F_I(u) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[1(-\hat{\varepsilon}_t^* \leq v) - 1(-\varepsilon_t^* \leq v) - F_{-\hat{\varepsilon}^*}(v) + F_{-\varepsilon^*}(v) \right],$$

$$B_{n5}^*(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[1(-\hat{\varepsilon}_t^* \leq v) - 1(-\varepsilon_t^* \leq v) - F_{-\hat{\varepsilon}^*}(v) + F_{-\varepsilon^*}(v) \right] \left[1(I_t \leq u) - F_I(u) \right],$$

and

$$B_{n6}^*(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[F_{-\hat{\varepsilon}^*}(v) - F_{-\varepsilon^*}(v) \right] 1(I_t \leq u).$$

Using the same arguments as in Lemma A.4, we can show

$$\sup_x |B_{nj}^*(x)| = o_p(1),$$

for $j = 1, 2, 4, 5$, and conditional on the original sample $\{Y_t, I_t\}_{t=1}^n$,

$$B_{n3}^*(v, u) = f_{\varepsilon^*}(v) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t^* 1(I_t \leq u)$$

and

$$B_{n6}^*(v, u) = -f_{\varepsilon^*}(-v) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t^* 1(I_t \leq u).$$

On the other hand, for the first term S_{n1}^*, we only have to show that $\sqrt{n} S_{n1}^* \Rightarrow S_\infty$ in probability conditional on the original sample. That is, we show that the finite-dimensional distributions of $\sqrt{n} S_{n1}^*$ converge (conditional on the original sample) to those of S_∞ in probability for all samples and it is asymptotically tight in probability. To this end, let's consider a finite set of points $(u_1, v_1), \ldots, (u_r, v_r)$ and a real vector $\lambda = (\lambda_1, \ldots, \lambda_r)'$ with $|\lambda| = 1$. We define

$$Z_{n,r}^{*}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j=1}^{r}\lambda_j\omega_t^{*}(v_j)1(I_t\leq u_j)\equiv\sum_{t=1}^{n}\xi_{n,t}^{r*}.$$

We notice that, conditional on the original sample $\{Y_t, I_t\}_{t=1}^{n}$, the array of random variables $\xi_{n,t}^{r*}$ is an independent (however, not necessarily identically distributed) array of random variables.

Denote by $\mathbb{E}^*$ and $\mathbb{V}^*$ the expectation and the variance, respectively, given the sample $\{Y_t, I_t\}_{t=1}^{n}$. We have

$$\mathbb{E}^*\left(Z_{n,r}^{*}\right)=\mathbb{E}^*\left(\sum_{t=1}^{n}\xi_{n,t}^{r*}\right)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{j=1}^{r}\lambda_j\mathbb{E}^*(\omega_t^{*}(v_j))1(I_t\leq u_j)=0,$$

and

$$\begin{aligned}
&\mathbb{V}^*\left(Z_{n,r}^{*}\right)=\mathbb{V}^*\left(\sum_{t=1}^{n}\xi_{n,t}^{r*}\right)\\
=&\sum_{t=1}^{n}V^*(\xi_{n,t}^{r*})\\
=&\sum_{j=1}^{r}\sum_{h=1}^{r}\lambda_j\lambda_h\left(\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}^*(\omega_t^{*}(v_j)\omega_t^{*}(v_h))1(I_t\leq u_j)1(I_t\leq u_h)\right)\\
=&\sum_{j=1}^{r}\sum_{h=1}^{r}\lambda_j\lambda_h\left(\frac{1}{n}\sum_{t=1}^{n}\hat{\omega}_t(v_j)\hat{\omega}_t(v_h)1(I_t\leq u_j)1(I_t\leq u_h)\right)\\
:=&\hat{\sigma}_{h,r}^2\\
\rightarrow_p&\sigma_{h,r}^2:=\sum_{j=1}^{r}\sum_{h=1}^{r}\lambda_j\lambda_h\mathbb{E}\left(\omega_t(v_j)\omega_t(v_h)1(I_t\leq u_j)1(I_t\leq u_h)\right),
\end{aligned}$$

where $\sigma_{h,r}^2$ is the covariance function of S_∞. Notice that the last result follows immediately by Lemma B.3. Besides $\hat{\sigma}_{h,r}^2\rightarrow_p\sigma_{h,r}^2$, we can also show that $\sum_{t=1}^{n}\mathbb{E}^*\left(|\xi_{n,t}^{r*}|^2 1(|\xi_{n,t}^{r*}|>\delta)\right)\rightarrow_p 0$ for some positive constant δ, see Stute et al. (1998, p. 149). By Lindeberg-Feller's central limit theorem, conditional on almost all samples, $\sum_{t=1}^{n}\xi_{n,t}^{r*}\Longrightarrow_* N(0,\sigma_{h,r}^2)$ in probability. The asymptotic uniform equicontinuity in probability in all samples follows directly from Theorem 2.11.9 in van der Vaart and Wellner (1996) or from Theorem A.1 in Delgado and Escanciano (2007). □

2.11 Appendix B

The first lemma from Yoshihara (1976) is needed to prove Lemma B.2. Its proof is omitted.

Lemma B.1: *Let $\{U_t\}$, $t = 1, \ldots, T$, be a d-dimensional strictly stationary absolutely regular stochastic process with mixing coefficient $\beta(k)$. Let $t_1 < \cdots < t_k$ be integers. Let $F(i, j)$, $i \leq j$, be the distribution function of $U_{t_i}, \ldots, U_{t_j}$. Let $h(\phi) := h(\phi_1, \ldots, \phi_k)$ be a Borel-measurable function on $\mathbb{R}^{kd}$ such that for some $\delta > 0$ and given j, there exists*

$$M \equiv \int |h(\phi)|^{1+\delta}\, dF(1, j) dF(j+1, k) < \infty$$

Then,

$$\left| \int h(\phi)\, dF(1, k) - \int h(\phi)\, dF(1, j) dF(j+1, k) \right| \leq 4M^{1/(1+\delta)} \beta(l)^{\delta/(1+\delta)},$$

where $l = t_{j+1} - t_j$.

The next lemma establishes a result useful to prove the main asymptotic theory.

Lemma B.2: *Under Assumptions **A.1'-A.6**,*

$$\sup_u \left| \frac{1}{n} \sum_{t=1}^{n} (Y_t - \hat{m}(I_t))\, 1(I_t \leq u) \right| = o_p(n^{-1/2}).$$

Proof of Lemma B.2: First noticing that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (Y_t - \hat{m}(I_t))\, 1(I_t \leq u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big((Y_t - \hat{m}(I_t))\, \hat{f}_I(I_t) \Big) \frac{1(I_t \leq u)}{\hat{f}_I(I_t)}.$$

Now substituting Nadaraya-Watson kernel estimators $\hat{m}(I_t)$ and $\hat{f}_I(I_t)$ into the above expression, we have

$$
\begin{aligned}
&\frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{s\neq t}^{n}\frac{1}{h^p}K\left(\frac{I_t-I_s}{h}\right)(Y_t-Y_s)\frac{1(I_t\leq u)}{\hat{f}_I(I_t)}\\
=&\frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{s\neq t}^{n}\frac{1}{h^p}K\left(\frac{I_t-I_s}{h}\right)(Y_t-Y_s)\frac{1(I_t\leq u)}{f_I(I_t)}\\
&+\frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{s\neq t}^{n}\frac{1}{h^p}K\left(\frac{I_t-I_s}{h}\right)(Y_t-Y_s)\frac{1(I_t\leq u)}{f_I(I_t)}\left(\frac{f_I(I_t)}{\hat{f}_I(I_t)}-1\right)\\
:=&D_{n1}(u)+D_{n2}(u).
\end{aligned}
$$

Following Stute (1994) or Delgado and González-Manteiga (2001), D_{n1} and D_{n2} are two U-processes indexed by u. Applying standard techniques from the theory of U-statistics, we will show that $D_{n1}=o_p(1)$ and $D_{n2}=o_p(1)$ uniformly in u. Here we only prove $D_{n1}=o_p(1)$, the proof of D_{n2} is similar but easier by noticing that $\hat{f}_I$ is a strongly consistent estimator of f_I.

We can rewrite D_{n1} as a U-statistic. To this end, let $\mathcal{W}_t=(Y_t,I_t)$ and introduce kernel

$$
\begin{aligned}
U(\mathcal{W}_t,\mathcal{W}_s)=&\frac{1}{2}\left[\frac{1}{h^p}K\left(\frac{I_t-I_s}{h}\right)(Y_t-Y_s)\frac{1(I_t\leq u)}{f_I(I_t)}\right.\\
&\left.+\frac{1}{h^p}K\left(\frac{I_t-I_s}{h}\right)(Y_s-Y_t)\frac{1(I_s\leq u)}{f_I(I_s)}\right],
\end{aligned}
$$

where we have suppressed the dependence on u of $U(\mathcal{W}_t,\mathcal{W}_s)$. Thus,

$$
D_{n1}=\sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}U(\mathcal{W}_t,\mathcal{W}_s)
$$

has a U-statistic form, see e.g. Gao and Hong (2008). However, it is easy to notice that the above (generalized) U-statistic is not a standard one since the kernel $U(\mathcal{W}_t,\mathcal{W}_s)$ depends on the sample size n implicitly through bandwidth h, which is a function of n.

Put $\mathcal{W} = (Y, I)$ and $\varepsilon = Y - m(I)$. We first calculate the projection term as follows,

$$
\begin{aligned}
u_1(\mathcal{W}) =& E[U(\mathcal{W}_t, \mathcal{W}_s)|\mathcal{W}_s = \mathcal{W}] \\
=& \frac{1}{2} E\left[\frac{1}{h^p} K\left(\frac{I_t - I}{h}\right) m(I_t) \frac{1(I_t \le u)}{f_I(I_t)}\right] \\
& - \frac{1}{2} E\left[\frac{1}{h^p} K\left(\frac{I_t - I}{h}\right) \frac{1(I_t \le u)}{f_I(I_t)}\right] Y \\
& + \frac{1}{2} E\left[\frac{1}{h^p} K\left(\frac{I_t - I}{h}\right)\right] Y \frac{1(I \le u)}{f_I(I)} \\
& - \frac{1}{2} E\left[\frac{1}{h^p} K\left(\frac{I_t - I}{h}\right) m(I_t)\right] \frac{1(I \le u)}{f_I(I)} \\
:=& D_{11} + D_{12} + D_{13} + D_{14}.
\end{aligned}
$$

We can also calculate that $D_{11} = m(I)1(I \le u) + O(h^L)$, $D_{12} = -Y1(I \le u) + O(h^L)$, $D_{13} = Y1(I \le u) + O(h^L)$ and $D_{14} = -m(I)1(I \le u) + O(h^L)$, so that $u_1(\mathcal{W}) = O(h^L)$.

Put $\phi(\mathcal{W}_t, \mathcal{W}_s) = U(\mathcal{W}_t, \mathcal{W}_s) - u_1(\mathcal{W}_t) - u_1(\mathcal{W}_s)$. By the theory of Hoeffding's decomposition, we write

$$
\begin{aligned}
& \sqrt{n} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} U(\mathcal{W}_t, \mathcal{W}_s) \\
=& \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_1(\mathcal{W}_t) + \sqrt{n} \frac{2}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} \phi(\mathcal{W}_t, \mathcal{W}_s).
\end{aligned}
$$

Then, we want to show that

$$
Var\left\{\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_1(\mathcal{W}_t)\right\} \to 0, \tag{2.7}
$$

and

$$
E\left\{\sqrt{n} \frac{1}{n(n-1)} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} \phi(\mathcal{W}_t, \mathcal{W}_s)\right\}^2 \to 0 \tag{2.8}
$$

uniformly in u. It is straightforward to prove (2.7). We immediately have this result because of the fact $Var\left(u_1(\mathcal{W}_t)\right) = O(h^{2L})$ due to Assumption A.4 and A.5.

We can follow the same idea as Yoshihara (1976) to prove (2.8). To this end, it suffices to show that the following three terms are asymptotically negligible, that is,

(a) the double summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\phi(\mathcal{W}_t,\mathcal{W}_s)^2\right\}\to 0, \tag{2.9}$$

(b) the triple summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-2}\sum_{s=t+1}^{n-1}\sum_{u=s+1}^{n}\phi(\mathcal{W}_t,\mathcal{W}_s)\phi(\mathcal{W}_t,\mathcal{W}_u)\right\}\to 0, \tag{2.10}$$

and (c) the quadruple summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-3}\sum_{s=t+1}^{n-2}\sum_{u=s+1}^{n-1}\sum_{v=u+1}^{n}\phi(\mathcal{W}_t,\mathcal{W}_s)\phi(\mathcal{W}_u,\mathcal{W}_v)\right\}\to 0.$$

Notice that, in order to show the last result, two distinct scenarios stand out, i.e. case $(c.1)$ with $t<s<u<v$ and $s-t>v-u$, and case $(c.2)$ $t<s<u<v$ and $s-t\leq v-u$. Given that both cases can be handled in a similar way, it suffices to show

$$\mathrm{E}\left\{\frac{1}{n^3}\sum\sum\sum\sum_{t<s<u<v,s-t>v-u}\phi(\mathcal{W}_t,\mathcal{W}_s)\phi(\mathcal{W}_u,\mathcal{W}_v)\right\}\to 0. \tag{2.11}$$

Lemma B.1 will be frequently used to prove the above claims. For example, for term (2.9), we have

$$\begin{aligned}
&\int|\phi(w_{t_1},w_{t_2})|^{2+\delta}\,dF(w_{t_1})dF(w_{t_2})\\
\leq& C\int|U(w_{t_1},w_{t_2})|^{2+\delta}\,dF(w_{t_1})dF(w_{t_2})+C\\
\leq& C\int\frac{1}{h^{p(2+\delta)}}\left|K\left(\frac{I_{t_1}-I_{t_2}}{h}\right)\right|^{2+\delta}|m(I_{t_1})-Y_{t_2}|^{2+\delta}\\
&dF(w_{t_1})dF(w_{t_2})+C\\
\leq& C\{h^{-p(1+\delta)}+1\}:=C_{1h}
\end{aligned}\tag{2.12}$$

by Assumptions A.1', A.4 and A.5, where the first inequality follows as $\mathrm{E}|u_1(\mathcal{W}_t)|^{2+\delta}$ is of smaller order than $\mathrm{E}|U(\mathcal{W}_t,\mathcal{W}_s)|^{2+\delta}$. By change of variables and Assumption A.1', we can get the last equality. When $\delta=0$, the inequality in (2.12) holds too, thus

$$\int \phi(w_{t_1},w_{t_2})^2\,dF(w_{t_1})dF(w_{t_2}) \le C(h^{-p}+1)=O(h^{-p}). \tag{2.13}$$

Now, by the β-mixing condition in Assumption A.1' and Lemma B.1,

$$\begin{aligned}&\left|\int \phi(w_{t_1},w_{t_2})^2\,dF_{t_1,t_2}(w_{t_1},w_{t_2}) - \int \phi(w_{t_1},w_{t_2})^2\,dF(w_{t_1})dF(w_{t_2})\right|\\ \le& 4C_{1h}^{2/(2+\delta)}\beta(s-t)^{\delta/(2+\delta)}\\ \le& Ch^{-2p(1+\delta)/(2+\delta)}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta},\end{aligned} \tag{2.14}$$

Combing inequalities (2.13) and (2.14), we get

$$\begin{aligned}\mathrm{E}\{\phi(\mathcal{W}_t,\mathcal{W}_s)^2\} &= \int \phi(w_{t_1},w_{t_2})^2\,dF_{t_1,t_2}(w_{t_1},w_{t_2})\\ &\le C\left\{h^{-2p(1+\delta)/(2+\delta)}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta}+h^{-p}\right\}.\end{aligned}$$

Therefore,

$$\begin{aligned}&\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\phi(\mathcal{W}_t,\mathcal{W}_s)^2\right\}\\ =&\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\mathrm{E}\{\phi(\mathcal{W}_t,\mathcal{W}_s)^2\}\\ \le& C\left(\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\left\{h^{-2p(1+\delta)/(2+\delta)}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta}+h^{-p}\right\}\right)\\ \le& C\left(\frac{1}{n^2h^{2p(1+\delta)/(2+\delta)}}+\frac{1}{nh^p}\right)\\ =&C\left(\frac{h^{2p/(2+\delta)}}{(nh^p)^2}+\frac{1}{nh^p}\right)\end{aligned}$$

converges to zero by Assumption A.5, where we have exploited the fact $\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}(s-$

$t)^{-(2+\eta)\delta/(2+\delta)\eta} = O(n)$ by noticing that $\delta > \eta$ and $(2+\eta)\delta/(2+\delta)\eta > 1$ under Assumption A.1'. We therefore finish proving (2.9). The proofs of the terms (2.10) and (2.11) are similar but lengthy, and is available upon request. □

The last lemma is needed to show the validity of bootstrap assisted test in Theorem 5.

Lemma B.3: *For all* $(v_1, u_1), (v_2, u_2) \in \overline{\mathbb{R}}^{p+1}$,

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\hat{\omega}_t(v_1)\hat{\omega}_t(v_2)1(I_t \le u_1)1(I_t \le u_2)\\ \to_p&\mathbb{E}\left(\omega_t(v_1)\omega_t(v_2)1(I_t \le u_1)1(I_t \le u_2)\right).\end{aligned}$$

Proof of Lemma B.3: Let's rewrite

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\hat{\omega}_t(v_1)\hat{\omega}_t(v_2)1(I_t \le u_1)1(I_t \le u_2)\\ =&\frac{1}{n}\sum_{t=1}^{n}\omega_t(v_1)\omega_t(v_2)1(I_t \le u_1)1(I_t \le u_2)\\ &+\frac{1}{n}\sum_{t=1}^{n}\omega_t(v_1)[\hat{\omega}_t(v_2)-\omega_t(v_2)]1(I_t \le u_1)1(I_t \le u_2)\\ &+\frac{1}{n}\sum_{t=1}^{n}\omega_t(v_2)[\hat{\omega}_t(v_1)-\omega_t(v_1)]1(I_t \le u_1)1(I_t \le u_2)\\ &+\frac{1}{n}\sum_{t=1}^{n}[\hat{\omega}_t(v_1)-\omega_t(v_1)][\hat{\omega}_t(v_2)-\omega_t(v_2)]1(I_t \le u_1)1(I_t \le u_2)\\ &=B_1+B_2+B_3+B_4.\end{aligned}$$

It is immediate to have that $B_1 \to_p \mathbb{E}\left(\omega_t(v_1)\omega_t(v_2)1(I_t \le u_1)1(I_t \le u_2)\right)$. Now we simply need to show that $B_j = o_p(1)$ for $j = 2, 4$ since B_3 can be proven in the same way as B_2. And they are proven by using the similar but easier arguments as in Lemmas A.1-A.4. □

Chapter 3

Nonparametric Tests of Conditional Independence for Weakly Dependent Data

3.1 Introduction

A variable Y is said to be conditionally independent of Z given X if and only if the conditional density of Y given Z and X equals to the conditional density of Y given X, that is, the realization of Z does not carry any information about Y when the realization of X is already known. Following David (1979), we write

$$Y \perp Z | X$$

to denote the hypothesis that Y is independent of Z given X. The hypothesis $Y \perp Z | X$ is related directly to but stronger than the hypothesis that Y is independent of Z, i.e. $Y \perp Z$ (unconditional independence) or the mean independence conditions $E(Y|Z, X) = E(Y|X)$.

The assumption of conditional independence plays an important role and is a widely imposed one in both statistical and econometric literature. For example, Markov property of a time series process, Granger non-causality, the assumption of missing at random (MAR) and exogeneity all can be formulated as a conditional independence problem, see Wang and Hong (2013) for detailed explanation of the relevance and importance of testing the conditional independence hypothesis in economics and econometrics. However, in stark contrast to the numerous tests of (unconditional) independence proposed either for independent and identically distributed (i.i.d.) data or for weakly dependent data, by far, there are not many tests designed especially for checking conditional independence assumption. Nonparametric tests for (unconditional) independence between random variables and/or vectors, and nonparametric tests for serial independence are abundant, e.g. a nonparametric test of Cramér-von Mises type first introduced by Hoeffding (1948), the empirical distribution function-based tests of Blum et al. (1961), Skaug and Tjøstheim (1993) for testing independence of raw data and Delgado and Mora (2000) or Ghoudi et al. (2001) for testing serial independence of time series or regression errors, the empirical characteristic function-based test of Csörgö (1985), kernel smoothing-based tests like Rosenblatt (1975), Robinson (1991), and Hong and White (2005), and tests based on measures of association and dependence between random variables and/or vectors such as Bakirov et al. (2006), Székely et al. (2007) or Diks and Panchenko (2007). See also Diks (2009) for a brief summary on various nonparametric tests for independence in the literature. On the other hand, among the recently available tests for conditional independence, the majority are designed only for i.i.d. data, e.g. Linton and Gozalo (1996) develop a non-pivotal nonparametric test based on the empirical distribution function, Song (2009) employs Rosenblatt transforms to obtain a distribution-free test, Huang (2010) propose a test based on maximal nonlinear conditional correlation, and Huang et al. (2013) develop an integrated conditional moment (ICM)-type test. Even fewer tests are tailored for dependence processes, exceptions are Su and White (2007, 2008, 2011), Bouezmarni et al. (2012) and Wang and Hong (2013) and some references therein.

In the present paper we aim to further fill the gap among the literature of testing conditional independence and propose new consistent nonparametric tests for conditional independence which are especially applicable to dependent data with weak dependence of unknown form. Our approach exploits a proper conditional moment restriction and is in the same spirit with Delgado and González-Manteiga (2001), which partially circumvents the "curse of dimensionality" problem. In comparison to the existing tests based on smoothers, our new test is able to detect local alternatives converging to the null at a parametric $\sqrt{n}$ rate, where n is the sample size.

We now lay out some preliminaries for testing the conditional independence hypothesis in a time series framework. Formally, let X_t, Y_t, and Z_t be three random vectors defined on some probability space $(\Omega, \mathscr{F}, P)$, with dimensions d_x, d_y, and d_z, respectively. Denote $d = d_x + d_y + d_z$. The hypothesis we are interested in throughout this paper is that Y_t is independent of Z_t given X_t, i.e. $Y_t \perp Z_t | X_t$ using the previous notation. Intuitively speaking, the conditional independence of Y_t and Z_t given X_t implies that, given the information realized already in X_t, Z_t can not provide any additional information in predicting Y_t in any way. We say that conditional on random vector X_t, the random vectors Y_t and Z_t are independent, if and only if

$$F_{Y_t,Z_t|X_t}(y,z|x) = F_{Y_t|X_t}(y|x)F_{Z_t|X_t}(z|x),$$

almost everywhere for $(x, y, z) \in \mathbb{R}^d$, where $F_{Y_t,Z_t|X_t}$, $F_{Y_t|X_t}$ and $F_{Z_t|X_t}$ denote the corresponding conditional CDFs.

The plan of the rest of the paper is as follows. In Section 2 we examine the testing problem and provide the test statistic. Section 3 establishes the asymptotic null distribution and Section 4 studies the consistency property of the test. The asymptotic power of the test under certain nonparametric alternatives is investigated in Section 5 and a bootstrap assisted procedure to implement the test is proposed and formally justified in Section 6. In Section 7, we study the finite sample performance of our test by means of an extensive set of Monte Carlo simulations. Section 8 presents an empirical example of using variance risk premium to predict equity risk premium. Finally, Section 9 concludes the paper. All mathematical proofs are deferred to the Appendices A and B.

3.2 The testing problem

We assume that we observe a time series of random vectors $\{X_t, Y_t, Z_t\}_{t=1}^n$. We aim to test the null hypothesis of $Y_t \perp Z_t | X_t$. It is important to stress that our formulation specifically allows for weak dependence in the data. Hence, our test is a new test for conditional independence designed towards time series data with weak dependence of unknown form. In principle, the nonstationary time series data are also permitted by applying our methodology. This is evident from the usage of time index t throughout the Introduction part. However, in the present paper, to simplify the asymptotic analysis, we shall focus on the strictly stationary case and develop our theoretical results based on this assumption, so that all the conditional distribution functions do not depend on time index t any more. Henceforth, we drop the subscript t in the conditional CDFs.

The null hypothesis of interest is that conditional on the information on random vector X_t, the random vectors Y_t and Z_t are independent, i.e.

$$\mathrm{H}_0 : F_{Y,Z|X} = F_{Y|X} F_{Z|X} \quad \text{a.s.} \tag{3.1}$$

The alternative hypothesis consists of the cases where H_0 doesn't hold. It is important to emphasize that this new formulation of testing conditional independence in (3.1) is highly attractive and useful, since it partly circumvents the notorious problem of "curse of dimensionality" by directly conditioning on only the variable X_t in all three conditional CDFs, whereas, in the usual formulation, one tend to condition on (Z_t, X_t), e.g. Su and White (2007, 2008), Bouezmarni and Taamouti (2012) and many others. The alternative hypothesis consists of all possible dependent relationships between Y_t and Z_t conditional on X_t, that is, given X_t, Y_t can be dependent of Z_t through mean, variance, and skewness and kurtosis, or even higher moments. For instance, it is possible to have situations where the dependence between Y_t and Z_t in low moments (mean, variance) does not exist, but it does exist in higher moments (e.g. skewness, kurtosis). We shall abbreviate here the rejection of the null hypothesis as $Y \not\perp Z|X$. See the Monte Carlo section for more concrete examples of possible data generating processes under the alternative hypothesis.

A more direct and commonly treated approach in order to check the null hypothesis (3.1) would consist of the following two steps. First of all, we obtain appropriate (consistent) nonparametric estimators for the three unknown conditional CDFs: $F_{Y,Z|X}(y, z|X)$, $F_{Y|X}(y|X)$

and $F_{Z|X}(z|X)$. For example, we can use the traditional Nadaraya-Watson (NW) kernel estimator proposed by Nadaraya (1964) and Watson (1964), or the local logistic distribution estimator and the weighted Nadaraya-Watson (WNW) estimator investigated by Hall et al. (1999) to estimate the conditional distribution functions. These methods have the advantage of producing distribution function estimators that are always constrained to lie between 0 and 1 and to be monotone increasing. Other choices are local linear methods suggested by Yu and Jones (1998). However, for the sake of exposition, in the present paper, we simply use NW estimator without worrying the efficiency issues of using different methods to estimate conditional distribution functions. To be concrete, a NW kernel estimator for $F_{Y,Z|X}(y,z|x)$ is defined as follows

$$\hat{F}_{Y,Z|X}(y,z|x) = \frac{\frac{1}{nh^{d_x}}\sum_{t=1}^{n} 1(Y_t \leq y)1(Z_t \leq z)K\left(\frac{X_t - x}{h}\right)}{\hat{f}_X(x)},$$

where $\hat{f}_X(x) = (nh^{d_x})^{-1}\sum_{t=1}^{n} K((X_t - x)/h)$ is the NW kernel estimator of marginal density function of X_t, $f_X(x)$, K is a d_x-th product kernel function (usually a density) and $h = h_n \in \mathbb{R}^+$ is a sequence of smoothing parameters (i.e. bandwidths). Similarly, we can obtain NW kernel estimators $\hat{F}_{Y|X}$ and $\hat{F}_{Z|X}$ for $F_{Y|X}$ and $F_{Z|X}$, respectively. Comparing to previous approaches in the literature, one important advantage here is that we only have to choose one bandwidth parameter h for the estimation of all three conditional CDFs.

Once we have obtained the three estimators for the conditional distribution functions, the next step is to measure how close the distance is to zero between $\hat{F}_{Y,Z|X}$ and the product of $\hat{F}_{Y|X}$ and $\hat{F}_{Z|X}$. To measure this closeness, popular choices are the well known Kolmogorov-Smirnov-type (KS) statistics of sup-norm or Cramér-von Mises-type (CvM) statistics of a L_2-norm with respect to a suitably chosen probability measure. For example, a CvM type test statistic based on a L_2-distance can be constructed in the following way,

$$\Gamma_n(h) = \frac{1}{n}\sum_{t=1}^{n}\left\{\hat{F}_{Y,Z|X}(Y_t, Z_t|X_t) - \hat{F}_{Y|X}(Y_t|X_t)\hat{F}_{Z|X}(Z_t|X_t)\right\}^2 W(X_t), \tag{3.2}$$

where $W(X_t)$ is a user chosen nonnegative weighting function, which may also depend on h. One practical choice of W is the nonparametric estimator $\hat{f}_X^4(X_t)$ in order to circumvent the problem of random denominators arising from the nonparametric estimation of the conditional CDFs. It is worthy to mention that the above formulation reduces effectively the dimensions of the testing problem, since we only face three d_x-dimensional nonparametric estimation problems. In this case, only one bandwidth h is necessary.

Several alternative forms similar to (3.2) are proposed and extensively studied in the literature of testing conditional independence. They are designed to test the hypotheses of the form $\Pr\{F_{Y|Z,X}(y|Z,X) = F_{Y|X}(y|X)\} = 1$ almost everywhere for $y \in \mathbb{R}^{d_y}$. For example, Bouezmarni and Taamouti (2012) exploit exactly the above formulation to construct tests for the conditional independence hypothesis using conditional CDFs. Formally, they propose a test statistic based on

$$\Gamma_n(h_1, h_2) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \hat{F}_{Y|Z,X}(Y_t|Z_t, X_t; h_1) - \hat{F}_{Y|X}(Y_t|X_t; h_2) \right\}^2 W(X_t, Z_t), \tag{3.3}$$

with $W(X_t, Z_t)$ again a non-negative weighting function, where

$$\hat{F}_{Y|Z,X}(y|z,x;h_1) = \frac{\left(nh_1^{d_z+d_x}\right)^{-1} \sum_{t=1}^{n} 1(Y_t \le y) K\left(\frac{Z_t - z}{h_1}, \frac{X_t - x}{h_1}\right)}{\left(nh_1^{d_z+d_x}\right)^{-1} \sum_{t=1}^{n} K\left(\frac{Z_t - z}{h_1}, \frac{X_t - x}{h_1}\right)}$$

and

$$\hat{F}_{Y|X}(y|x;h_2) = \frac{\left(nh_2^{d_x}\right)^{-1} \sum_{t=1}^{n} 1(Y_t \le y) K^*\left(\frac{X_t - x}{h_2}\right)}{\left(nh_2^{d_x}\right)^{-1} \sum_{t=1}^{n} K^*\left(\frac{X_t - x}{h_2}\right)},$$

are NW estimator of $F_{Y|Z,X}$ and $F_{Y|X}$, respectively. In the above expressions, h_1 and h_2 are two different bandwidths and K and K^* are two different kernel functions. It is immediate to notice that their test statistic is subject to the problem of "curse of dimensionality" more seriously than (3.2) since they have to face a $(d_x + d_z)$-dimensional problem, that is, the nonparametric estimation of $F_{Y|Z,X}(y|z,x)$. Furthermore, they have to chose two different bandwidths h_1 and h_2 while only one bandwidth h is needed in (3.2).

In this paper, we propose new tests based on the attractive formulation in (3.1) with the inherent advantage of dimension reduction. According to the null hypothesis in (3.1), testing independence of Y_t and Z_t conditional on X_t is equivalent to testing that the conditional CDF of $(Y_t, Z_t)'$ given X_t is the product of two marginal conditional distributions of Y_t given X_t and Z_t given X_t. Notice that $F_{Y,Z|X}(y, z|x) = E[1(Y_t \leq y)1(Z_t \leq z)|X_t = x]$. This useful fact has been exploited by many authors, e.g. Delgado and González-Manteiga (2001) in their test for conditional independence in an i.i.d framework. Clearly, the null hypothesis of conditional independence can be re-expressed as

$$\mathrm{H}_0 : E[1(Y_t \leq y)1(Z_t \leq z)|X_t] = E[1(Y_t \leq y)|X_t]E[1(Z_t \leq z)|X_t] \quad \text{a.s.}$$

a.e. for any $(y, z) \in \mathbb{R}^{d_y+d_z}$. It is immediate to see that we can further transform the above null hypothesis to

$$\mathrm{H}_0 : E\left\{1(Y_t \leq y)(1(Z_t \leq z) - E[1(Z_t \leq z)|X_t])|X_t\right\} = 0 \quad \text{a.s.} \tag{3.4}$$

a.e. for any $(y, z) \in \mathbb{R}^{d_y+d_z}$. The alternative hypothesis is the negation of the null. Therefore, the problem of testing hypothesis (3.1) is essentially equivalent to testing the properly modified conditional moment restriction in (3.4) which by construction, comparing to the previous methods, already exploits the advantage of a certain degree of dimension reduction in formulating the testing conditional independence problem at hand to the form of (3.4). It is also important to mention that the above formulation of conditional moment restriction is indexed by parameters (y, z) so that, for a given fixed pair of (y, z), we have one conditional moment restriction. Since this restriction has to hold for all possible (y, z), as a matter of fact, we have an infinite number (i.e. continuum) of conditional moment restrictions.

Now, put $\varepsilon_t(y, z) = 1(Y_t \leq y)\epsilon_t(z)$ for any $(y, z) \in \mathbb{R}^{d_y+d_z}$ with $\epsilon_t(z) = 1(Z_t \leq z) - E[1(Z_t \leq z)|X_t]$. Obviously, the errors $\epsilon_t(z)$ form a martingale process. We call $\varepsilon_t(y, z)$ by the name of generalized errors. We express the null hypothesis in terms of the generalized errors $\varepsilon_t(y, z)$, that is, (3.4) holds if and only if

$$E[\varepsilon_t(y, z)|X_t] = 0 \quad \text{a.s.} \quad \forall (y, z) \in \mathbb{R}^{d_y+d_z}.$$

To test the above conditional moment restriction, popular kernel based approach could be applied, that is, we could follow methods from Fan and Li (1996) and Zheng (1998)[1] and construct tests using the fact

$$\begin{aligned} &E\{\varepsilon_t(y,z)E[\varepsilon_t(y,z)|X_t]\} \\ =&E\{(E[\varepsilon_t(y,z)|X_t])^2\} \\ \geq&0, \end{aligned}$$

where the equality holds if and only if $E[\varepsilon_t(y,z)|X_t] = 0$ a.s. Therefore, one possible consistent test could be proposed basing on an estimator of the above expression. We leave this possibility for future research.

Another widely adopted methodology to test a conditional moment restriction consists of first transforming such conditional moment restriction to an infinite number of unconditional moment restrictions. By a measure-theoretic argument, the conditional restriction can be characterized by the following transformation, that is, it is equivalent to testing

$$E[\varepsilon_t(y,z)1(X_t \leq x)] = 0 \quad \text{a.s.} \quad \forall (x,y,z) \in \mathbb{R}^{d_x+d_y+d_z}. \tag{3.5}$$

This type of transformation is commonly seen and well studied in the literature of model checks for regression, significance testing in regression analysis or other contexts for testing, e.g. Stute (1997), Delgado and González-Manteiga (2001), and Escanciano and Velasco (2006). This integrated regression function approach is the approach that we are going to concentrate on. Our test statistics are based on the stochastic process derived from a proper estimator of the previous expression. Formally, based on (3.5), to avoid certain technical issues, we propose to test that the following slightly modified (nonlinear) dependence measures $\gamma(x,y,z)$ are identically zero

[1] See also Fan and Li (1999), Li (1999), Gu et al (2007) or more recently Jeong et al. (2012) and Dette et al. (2014) and references therein for different applications of this specific methodology.

for all (x, y, z) in their supports, i.e. we can characterize H_0 as

$$\mathrm{H}_0 : \gamma(x, y, z) = 0 \quad \text{a.s.}, \quad \forall (x, y, z) \in \mathbb{R}^d, \tag{3.6}$$

with $d = d_x + d_y + d_z$, where

$$\gamma(x, y, z) = \mathrm{E}[1(X_t \leq x)\varepsilon_t(y, z) f_X(X_t)]. \tag{3.7}$$

The inclusion of marginal density estimator $f_X(X_t)$ in the modified formulation in (3.7) is mainly because of technical reason. This so-called “density-weighted” null hypothesis helps to avoid the random denominators problem arising from the kernel estimation of $F_{Z|X}$, see e.g. Powell et al. (1989). Theoretically speaking, there is an optimal weight function in the sense of maximizing local power against the class of local alternatives (1.7) in Section 5. But using the density function $f_X(X_t)$ as the weight function appears not causing severe problems to the size and power performance of the proposed test, as shown in the Monte Carlo simulation of Section 7. Another solution to avoid this random denominators problem consists of introducing a trimming parameter $b > 0$ to trim out those close-to-zero values of $\hat{f}_X(X_t)$, i.e. $1\left(\hat{f}_X(X_t) > b\right)$, used in e.g. Robinson (1988). We shall not pursue this possibility here.

If we could observe $\varepsilon_t(y, z)$ (hence, there is no need to include $f_X(X_t)$), then simple test statistics could be formulated based on the following (infeasible) stochastic process

$$\hat{S}_n^0(x, y, z) := \sqrt{n}\hat{\gamma}_{0n}(x, y, z) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1(X_t \leq x)\varepsilon_t(y, z).$$

By defining $\hat{S}_n^0(-\infty, \cdot, \cdot) = \hat{S}_n^0(\cdot, -\infty, \cdot) = \hat{S}_n^0(\cdot, \cdot, -\infty) = 0$, the sample paths of $\hat{S}_n^0$ belong to the space $\ell^\infty(\overline{\mathbb{R}}^d)$, the space of all uniformly bounded real functions on $\overline{\mathbb{R}}^d := [-\infty, \infty]^d$, which is equipped with the sup-norm. Assuming some regularity conditions, one can show that under the null hypothesis, $\hat{S}_n^0$ converges weakly on the topology of $\ell^\infty(\overline{\mathbb{R}}^d)$ endowed with the sup-norm to S_∞^0, a Gaussian process with zero mean, continuous sample paths, and covariance structure

$$\begin{aligned}
&E[S^0_\infty(x_1,y_1,z_1)S^0_\infty(x_2,y_2,z_2)]\\
=&E\left(1(X_1\le x_1\wedge x_2))F_{Y|X}(y_1\wedge y_2|X_1)[F_{Z|X}(z_1\wedge z_2|X_1)-F_{Z|X}(z_1|X_1)F_{Z|X}(z_2|X_1)]\right),
\end{aligned}$$

where $a\wedge b\equiv\min\{a,b\}$ element-wisely for any vectors a and b.

Since we can not observe $\varepsilon_t(y,z)$ in practice, we need to estimate it by its counterpart $\hat{\varepsilon}_t(y,z)=1(Y_t\le y)\hat{\epsilon}_t(z)$ with $\hat{\epsilon}_t(z)=1(Z_t\le z)-\hat{F}_{Z|X}(z|X_t)$ the nonparametric residuals from regression $1(Z_t\le z)=F_{Z|X}(z|X_t)+\epsilon_t(z)$, where

$$\hat{F}_{Z|X}(z|X_t)=\frac{\frac{1}{(n-1)h^{d_x}}\sum_{s=1,s\neq t}^n 1(Z_s\le z)K\left(\frac{X_t-X_s}{h}\right)}{\hat{f}_X(X_t)}$$

is the leave-one-out NW kernel estimator for the conditional distribution function $F_{Z|X}(z|X_t)$ and

$$\hat{f}_X(X_t)=\frac{1}{(n-1)h^{d_x}}\sum_{s=1,s\neq t}^n K\left(\frac{X_t-X_s}{h}\right)$$

is the leave-one-out NW kernel estimator for the probability density function $f_X(X_t)$ evaluated at the data point X_t. Therefore, a feasible estimator of $\gamma(x,y,z)$ based on a sample of observations $\{(X_t,Y_t,Z_t)\}_{t=1}^n$ is given by

$$\hat{\gamma}_n(x,y,z)=\frac{1}{n}\sum_{t=1}^n 1(X_t\le x)\hat{\varepsilon}_t(y,z)\hat{f}_X(X_t),$$

which is algebraically equivalent to

$$\begin{aligned}
\hat{\gamma}_n(x,y,z)=&\frac{1}{n(n-1)}\sum_{t=1}^n\sum_{s=1,s\neq t}^n\frac{1}{h^{d_x}}K\left(\frac{X_t-X_s}{h}\right)\\
&\times 1(X_t\le x)1(Y_t\le y)[1(Z_t\le z)-1(Z_s\le z)]. \qquad (3.8)
\end{aligned}$$

Notice that the expression (3.8) is a variant of standard U-process, e.g. Stute (1994) or Delgado and González-Manteiga (2001). In fact, we rely on U-statistic theory to establish our asymptotic results in the Appendix.

Finally, let's put $\hat{S}_n(x, y, z) = \sqrt{n}\hat{\gamma}_n(x, y, z)$. Test statistics can be constructed based on any continuous functionals. We propose a test statistic of the Cramér-von Mises-type based on the L_2 distance with respect to a suitable measure, in our case, the empirical joint distribution. This yields

$$\begin{aligned} CvM_n &= \int_{\mathbb{R}^d} \left(\hat{S}_n(x, y, z)\right)^2 d\hat{F}_n(x, y, z) \\ &= \sum_{t=1}^{n} \hat{\gamma}_n^2(X_t, Y_t, Z_t), \end{aligned} \tag{3.9}$$

where

$$\hat{F}_n(x, y, z) = \frac{1}{n}\sum_{t=1}^{n} 1(X_t \leq x)1(Y_t \leq y)1(Z_t \leq z)$$

is the empirical distribution function of $\{X_t, Y_t, Z_t\}$. A test statistic of the Kolmogorov-Smirnov-type based on the sup-norm is given by

$$KS_n = \sup_{(x,y,z)\in\overline{\mathbb{R}}^d} |\hat{S}_n(x, y, z)|.$$

Under the null hypothesis, CvM_n (or KS_n) is expected to be very close to zero and converges to a finite distribution, while CvM_n will diverge to infinity under the alternatives. Hence, the CvM_n test is a consistent test. We reject the null hypothesis of conditional independence whenever CvM_n exceeds a certain "big" value. However, since the asymptotic distribution is not pivotal and depends on the underlying data generating process (DGP) in a complicated way, this critical value may depend on the underlying DGP as well. Therefore, in practice, some kind of bootstrap or re-sampling procedures are necessary in order to implement our tests. This issue will be addressed in section 6.

3.3 Asymptotic null distribution

In this section, we will establish the null limiting distribution of our test. Since in the present paper we are mainly focusing on data with weak dependence, we need to specify the amount of dependence allowed in the processes of interest. In what follows, we introduce the notion of β-mixing dependence. Let us recall the definition of a β-mixing process (see e.g. Doukhan (1994) and Fan and Yao (2003) amongst others). For $\left\{\mathcal{W}_t = (X_t, Y_t, Z_t)'\,;\ t \geq 1\right\}$ a strictly stationary stochastic process and $\mathcal{F}_t^s$ a sigma algebra generated by $(\mathcal{W}_s, \ldots, \mathcal{W}_t)$ for $s \leq t$, the process $\{\mathcal{W}\}$ is called β-mixing (or absolutely regular) with mixing coefficient $\beta(j)$, if

$$\beta\left(j\right) = \sup_{s\in\mathbb{N}} \mathbb{E}\left[\sup_{\mathcal{A}\in\mathcal{F}_{s+j}^{+\infty}} \left|P\left(\mathcal{A}\right|\mathcal{F}_{-\infty}^{s}\right) - P\left(\mathcal{A}\right)\right|\right] \to 0,\ a.s.\ j \to \infty.$$

We impose the following regularity conditions.

Assumption A.1 (Data Generating Process): (a) $\{X_t, Y_t, Z_t\}_{t\in\mathbb{Z}}$ is a strictly stationary, ergodic and absolutely regular process with β-mixing coefficients $\beta(j) = O(j^{-(2+\eta)/\eta})$ for some constant $0 < \eta < 1$; (b) the conditional distribution functions $F_{Y|X}(y|x)$ and $F_{Z|X}(z|x)$ are continuously differentiable with respect to x up to order l for some integer $l \geq 2$.

Assumption A.2 (Kernel Function): $K(u)$ is a bounded, continuous, and symmetric function on $\mathbb{R}^{d_x}$ such that $K(u) = \prod_{j=1}^{d_x} k(u_j)$, $\int u^i K(u)\, du = \delta_{0i}$ for $i = 0, 1, \cdots, l-1$ and $\int u^l K(u)\, du \neq 0$, where $k : \mathbb{R} \to \mathbb{R}$ is a bounded, symmetric univariate function and δ_{ij} is the delta function taking value one when $i = j$ and zero otherwise.

Assumption A.3 (Bandwidth): The sequence of bandwidths h satisfies (a) $h \to 0$; and (b) $nh^{d_x} \to \infty$ and $nh^{2l} \to 0$ as $n \to \infty$.

Assumption A.1(a) restricts the amount of temporal dependence in $\{X_t, Y_t, Z_t\}$, which is a standard assumption in nonparametric time series analysis. Assumption A.1(b) is a smoothness condition and Assumption A.1(c). Assumption A.2 imposes a higher order kernel condition and is standard in asymptotic theory of nonparametric regression. Bandwidth condition in Assumption A.3 is minimal and guarantees that the projection terms of U-statistics in the proof are asymptotically negligible.

We now state the asymptotic distribution of $\hat{S}_n$ under H_0.

Theorem 1: *Let Assumptions A.1-A.3 hold. Then, under the null hypothesis,*

$$\hat{S}_n \Rightarrow S_\infty$$

in $\ell^\infty(\overline{\mathbb{R}}^d)$*, where* S_∞ *is a zero mean Gaussian process with covariance function*

$$\mathbb{E}\left(S_\infty(x_1, y_1, z_1), S_\infty(x_2, y_2, z_2)\right) = \mathbb{E}\left(f_X^2(X_1)1(X_1 \leq x_1 \wedge x_2)\phi_1(y_1)\phi_1(y_2)\epsilon_1(z_1)\epsilon_1(z_2)\right)$$

with

$$\phi_t(y) = 1(Y_t \leq y) - F_{Y|X}(y|X_t)$$

and

$$\epsilon_t(z) = 1(Z_t \leq z) - F_{Z|X}(z|X_t).$$

We state the null limiting distributions of our test statistics CvM_n and KS_n in the corollary below. Its proof is given in the appendix.

Corollary 1: *Let Assumptions A.1-A.3 hold. Then, under the null hypothesis,*

$$CvM_n \rightarrow_d CvM_\infty = \int_{\mathbb{R}^d} \left(S_\infty(x, y, z)\right)^2 \, dF(x, y, z)$$

and

$$KS_n \rightarrow_d KS_\infty = \sup_{(x,y,z)\in\overline{\mathbb{R}}^d} |S_\infty(x, y, z)|,$$

where S_∞ *is the Gaussian process defined in Theorem 1.*

3.4 Consistency

The consistency property of the second test based on rejecting H_0 for large values of CvM_n is stated in the following theorems.

Theorem 2: *Under Assumptions A.1-A.3 and under the alternative hypothesis,*

$$\frac{1}{n}CvM_n \to_p \int_{\mathbb{R}^d} (\gamma(x,y,z))^2 \; dF(x,y,z)$$

and

$$\frac{1}{\sqrt{n}}KS_n \to_p \sup_{(x,y,z)\in\overline{\mathbb{R}}^d} |\gamma(x,y,z)|.$$

Clearly, $\gamma(x,y,z)$, under the alternative hypothesis, is not identically zero for a positive Lebesgue measure of $(x,y,z) \in \mathbb{R}^d$, cf. Theorem 1 in Bierens (1982). This fact guarantees that the tests will be consistent because

$$\int_{\mathbb{R}^d} (\gamma(x,y,z))^2 \; dF(x,y,z) > 0$$

or

$$\sup_{(x,y,z)\in\overline{\mathbb{R}}^d} |\gamma(x,y,z)| > 0.$$

That is, the test is consistent against all alternatives of the null hypothesis in (3.6).

3.5 Asymptotic power

To further investigate the consistency properties of the proposed test, we first introduce the following class of nonparametric local alternatives:

$$\mathrm{H}_{1n} : F_{Y,Z|X}(y,z|x) = F_{Y|X}(y|x)F_{Z|X}(z|x) + \frac{1}{\sqrt{n}}\Delta(x,y,z), \tag{3.10}$$

where $\Delta(x,y,z)$ satisfies $\Delta(x,\infty,z) = \Delta(x,y,\infty) = \Delta(x,-\infty,z) = \Delta(x,y,-\infty) = 0$, $\Delta(x,y,z) \neq 0$, and is a non-random, twice continuously differentiable function with uniformly bounded first order derivatives with respect to x, y and z. The additional term $n^{-1/2}\Delta(x,y,z)$ characterizes the departure of the conditional joint distribution function from the product of conditional marginal distribution functions, whereas the rate $n^{-1/2}$ is the exact speed at which the deviation between $F_{Y,Z|X}(y,z|x)$ and $F_{Y|X}(y|x)F_{Z|X}(z|x)$ vanishes to zero as the sample size n tending to infinity. In fact, when we characterize the local alternatives in terms of conditional density functions, that is,

$$\mathrm{H}_{1n} : f_{Y,Z|X}(y,z|x) = f_{Y|X}(y|x)f_{Z|X}(z|x) + \frac{1}{\sqrt{n}}\delta(x,y,z),$$

where the deviation term $\delta(x,y,z)$ satisfies $\delta(x,y,z) \neq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(x,y,z)\,dydz = 0$. It is immediate to infer that in our formulation $\Delta(x,y,z)$ needs to satisfy $\Delta(x,y,z) = \int_{-\infty}^{y}\int_{-\infty}^{z}\delta(x,\bar{y},\bar{z})\,d\bar{y}d\bar{z}$.

Furthermore, we assume the function $\Delta(x,y,z)$ to satisfy the following regularity condition.

Assumption A.4 (Local Alternatives) There exists some $(x,y,z) \in \mathbb{R}^d$ with a positive Lebesgue measure such that the function $\Delta(x,y,z)$ satisfies that

$$\int_{-\infty}^{x} f_X^2(\bar{x})\Delta(\bar{x},y,z)\,d\bar{x} \neq 0.$$

Next theorem shows the asymptotic behaviour of $\hat{S}_n$ under the sequences of alternative hypotheses tending to the null at the parametric rate $n^{1/2}$.

Theorem 3: *Under the local alternatives in* (3.10)*, suppose Assumptions A.1-A.4 hold, then*

$$\hat{S}_n \Rightarrow S_\infty^1$$

with

$$S_\infty^1 \stackrel{d}{=} S_\infty + G_\infty$$

in $\ell^{\infty}(\overline{\mathbb{R}}^d)$, where S_∞ is the zero mean Gaussian process defined in Theorem 1 and G_∞ is a deterministic shift function with

$$G_\infty(\eta) \equiv G_\infty(x, y, z) = E\left[f_X(X_1)1(X_1 \leq x)\Delta(X_1, y, z)\right].$$

Several remarks are in order. Since in general the deterministic shift function G_∞ is not identically zero under the regularity assumption, our test has nontrivial asymptotic local power and is able to detect the class of local alternatives H_{1n} converging to the null at a parametric rate $n^{1/2}$. The convergence rate $n^{1/2}$ is much faster than the usual convergence rate obtained from many other smoothed nonparametric tests. For example, Su and White's (2008) test only has power against local alternatives at distance $n^{-1/2}h^{-(d_x+d_y+d_z)/4}$, while tests proposed by Su and White (2007) and Bouezmarni and Taamouti (2012) have power against local alternatives at distance $n^{-1/2}h^{-(d_x+d_z)/4}$. Among the smoothed nonparametric tests, Wang and Hong's (2013) test is the only test that has a relatively faster convergence rate and it can detect a class of local alternatives that converges to the null at the rate $n^{-1/2}h^{-d_x/4}$. Nonetheless, tests of the Wang and Hong's (2013) type are able to detect other hight frequency local alternatives considered by Rosenblatt (1975) and Horowitz and Spokoiny (2001) amongst others, while our tests cannot detect such type of alternatives.

We also notice that there could exist some peculiar non-conditional independence DGPs (i.e. $Y \not\perp Z|X$) such that the function $\Delta(x, y, z)$ happens to make $G_\infty \equiv 0$, or in other words, this class of $\Delta(x, y, z)$ violates the Assumption A.4. It is clear that our test will not be able to detect such types of local alternatives included in H_{1n} but failing our assumption. Fortunately, these DGPs are not common in economics and finance. However, it would be interesting to formally characterize these classes of non-conditional independence process that do not satisfy Assumption A.4, but this is beyond the scope of this paper.

The next corollary is a direct application of Continuous Mapping Theorem and Theorem 3. Its proof is similar to that of Corollary 1 and is therefore omitted.

Corollary 2: *Under the local alternatives* (3.10)*, if Assumptions A.1-A.4 hold,*

$$CvM_n \to_d CvM_\infty^1 = \int_{\mathbb{R}^d} \left(S_\infty(x,y,z) + G_\infty(x,y,z)\right)^2 \, dF(x,y,z)$$

and

$$KS_n \to_d KS_\infty^1 = \sup_{(x,y,z)\in\overline{\mathbb{R}}^d} |S_\infty(x,y,z) + G_\infty(x,y,z)|.$$

We conclude that under the class of local alternatives specified in (3.10) and Assumption A.4, the limiting distributions of the test statistics CvM_n and KS_n shift in a nontrivial way asymptotically. The shifting term G_∞ will guarantee that our tests have nontrivial power against the class of local alternatives H_{1n}. Therefore, the test is able to detect nonparametric alternatives converging to the null at the parametric rate $\sqrt{n}$ with sample size n.

3.6 Bootstrap

Since the null limiting distribution of our test statistic CvM_n depends heavily on the underlying DGP, it is difficult to obtain critical values by using the asymptotic distribution. So we use a bootstrap procedure to estimate the critical values for our tests. Our bootstrap is of a multiplier bootstrap type proposed by Delgado and González Manteiga (2001).

Taking advantage of the asymptotic theory developed in Theorem 1, we define the bootstrapped process as

$$\hat{S}_n^*(x,y,z) = \sqrt{n}\hat{\gamma}_n^*(x,y,z)$$

with

$$\begin{aligned}\hat{\gamma}_n^*(x,y,z) =& \frac{1}{n}\sum_{t=1}^{n} \hat{f}_X(X_t) 1(X_t \le x) \left[1(Y_t \le y) - \hat{F}_{Y|X}(y|X_t)\right] \\ &\times \left[1(Z_t \le z) - \hat{F}_{Z|X}(z|X_t)\right] v_t, \end{aligned} \tag{3.11}$$

where $F_{Y|X}(y|X_t)$ and $F_{Z|X}(z|X_t)$ are the NW estimators of conditional CDFs $F_{Y|X}(y|X_t)$ and

$F_{Z|X}(z|X_t)$, and $\{v_t\}_{t=1}^n$ is a sequence of independent random variables with zero mean, unit variance, bounded support and is also independent of $\{Y_t\}_{t=1}^n$. One example of $\{v_t\}$ sequences is i.i.d. Bernoulli variates with

$$\Pr\left(v_t = \frac{1-\sqrt{5}}{2}\right) = \frac{1+\sqrt{5}}{2\sqrt{5}}$$

and

$$\Pr\left(v_t = \frac{1+\sqrt{5}}{2}\right) = \frac{-1+\sqrt{5}}{2\sqrt{5}},$$

see Mammen (1993) for motivation on this popular choice, see the application of this choice in Delgado and González-Manteiga (2001) or Escanciano and Velasco (2006). The other choice of $\{v_t\}$ sequences is i.i.d. Bernoulli variates such that $\Pr(v_t = 1) = 0.5$ and $\Pr(v_t = -1) = 0.5$ (Rademacher variates) used by Liu (1988) and de Jong (1996). It is important to notice that the third moment of v_t in those cases is equal to 1, and hence the first three moments of the bootstrap series coincide with the three moments of the original series. These properties have implications on the second order asymptotic properties of the bootstrap approximation, see Liu (1988).

In the next theorem we justify the asymptotic validity of our multiplier bootstrap method and it allows us to approximate the critical values of the test.

Theorem 4: *Suppose Assumptions A.1-A.3 hold, then under the null hypothesis* (3.1)*, under any fixed alternative hypothesis or under the local alternatives* (3.10)*,*

$$\hat{S}_n^* \Rightarrow S_\infty \quad \textit{in probability}$$

in $\ell^\infty(\overline{\mathbb{R}}^d)$*, where* S_∞ *is the Gaussian process defined in Theorem 1 and* $\Rightarrow$ *in probability denotes the weak convergence in probability under the bootstrap law, i.e., conditional on the original sample* $\{(X_t', Y_t', Z_t')'\}_{t=1}^n$.

Therefore, we can approximate the asymptotic distribution of the stochastic process $\hat{S}_n$ by that of $\hat{S}_n^*$. In particular, we can simulate the critical values for the test statistics CvM_n (or

KS_n) by the following algorithm:

Step 1: Estimate the nonparametric regression $1(Z_t \leq z) = F_{Z|X}(z|X_t) + \epsilon_t(z)$ using the original sample $\{(X_s', Z_s')\}_{s=1}^n$, i.e. getting a $n \times n$ matrix of the NW estimates of $F_{Z|X}(Z_t|X_s)$ for (t, s), and obtain the corresponding matrix of nonparametric residuals $\hat{\epsilon}_t(Z_s)$ for (t, s). Compute $\hat{\gamma}_n(X_t, Y_t, Z_t)$ for each t and compute the test statistic CvM_n using (3.9).

Step 2: Generate a sequence of i.i.d. Bernoulli variates $\{v_t\}_{t=1}^n$ independent of the original sample. Compute $\hat{\gamma}_n^*(X_t, Y_t, Z_t)$ based on (3.11) for $t = 1, \cdots, n$ and get the bootstrapped test statistic CvM_n^*.

Step 3: Repeat Step 2 B times to give a sample of bootstrapped test statistic $\{CvM_{n,b}^*\}_{b=1}^B$. Compute the empirical $(1-\alpha)$-th sample quantile of $\{CvM_{n,b}^*\}_{b=1}^B$, $CvM_n^{*\alpha}$ say. The proposed test rejects the null hypothesis at the significance level α if $CvM_n > CvM_n^{*\alpha}$.

Notice that given the result obtained in Theorem 4, the proposed bootstrap assisted tests have a correct asymptotic level, are consistent and are able to detect local alternatives converging to the null at the parametric rate $\sqrt{n}$, where n is the sample size.

3.7 Monte Carlo simulation

In this section, we carry out an extensive set of experiments to examine the finite sample performance of the proposed test CvM_n under the null hypothesis and under the alternative hypothesis. Results from the test KS_n are available upon request. As noted in Section 1, testing conditional independence can be seen as a tool of testing Granger causality in certain ways. We shall focus on these Granger causality (or non-causality) types of data generating processes (DGPs). Formally, we consider the following eleven DGPs taken or slightly modified from

Wang and Hong (2013):

$$
\begin{aligned}
&\text{(S1):} \quad Y_t = \varepsilon_{1,t}, Z_t = \varepsilon_{2,t}, X_t = \varepsilon_{3,t} \\
&\text{(S2):} \quad Y_t = 0.5Y_{t-1} + \varepsilon_{1,t} \\
&\text{(S3):} \quad Y_t = 0.5Y_{t-1} \exp\left(-0.5Y_{t-1}^2\right) + \varepsilon_{1,t} \\
&\text{(S4):} \quad Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}, h_{1,t} = 0.01 + 0.9h_{1,t-1} + 0.05Y_{t-1}^2 \\
&\qquad\quad\;\; Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}, h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2 \\
&\text{(P1):} \quad Y_t = 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{1,t} \\
&\text{(P2):} \quad Y_t = 0.5Y_{t-1}Z_{t-1} + \varepsilon_{1,t} \\
&\text{(P3):} \quad Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}^2 + \varepsilon_{1,t} \\
&\text{(P4):} \quad Y_t = 0.3 + 0.2\log(h_t) + \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2 + 0.3Z_{t-1}^2 \\
&\text{(P5):} \quad Y_t = 0.5Y_{t-1} + 0.5Z_{t-1}\varepsilon_{1,t} \\
&\text{(P6):} \quad Y_t = \sqrt{h_t}\varepsilon_{1,t}, h_t = 0.01 + 0.5Y_{t-1}^2 + 0.25Z_{t-1}^2 \\
&\text{(P7):} \quad Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}, h_{1,t} = 0.01 + 0.1h_{1,t-1} + 0.4Y_{t-1}^2 + 0.5Z_{t-1}^2 \\
&\qquad\quad\;\; Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}, h_{2,t} = 0.01 + 0.9h_{2,t-1} + 0.05Z_{t-1}^2
\end{aligned}
$$

where $\varepsilon_{1,t}$, $\varepsilon_{2,t}$ and $\varepsilon_{3,t}$ are three i.i.d. $N(0,1)$ sequences independent of each other and Z_t in DGPs (S2) and (S3) and DGPs (P1)-(P6) is generated by the following autoregressive process of order 1, AR(1),

$$Z_t = 0.5Z_{t-1} + \varepsilon_{2,t}.$$

These DGPs cover a wide range of linear and nonlinear time series processes. DGPs (S1)-(S4), serve to investigate the size performance of our test CvM_n, while DGPs (P1)-(P7) allow us to examine the powers. Specifically, for DGP (S1), the pure innovations case, we simply test $Y_t \perp Z_t | X_t$, that is, $\varepsilon_{1,t} \perp \varepsilon_{2,t} | \varepsilon_{3,t}$. But from DGP (S2) to DGP (P7), we will test whether Y_t

is independent of Z_{t-1} conditional on Y_{t-1}, i.e. $Y_t \perp Z_{t-1} | Y_{t-1}$, which is equivalent to testing whether Z_t Granger causes Y_t by setting the lag order to 1.

For each DGP, we first generate $n + 200$ observations and then discard the first 200 observations to minimize the possible effects caused by initial values. The number of Monte Carlo simulations is 500 and the bootstrap critical values are calculated from $B = 500$ bootstrap replications. Three sample sizes, $n = 100, 200500$ are considered in all the simulation study. We only report results for the nominal size of 5%. We choose the standard normal density $K(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$ as our kernel function. Bandwidth of the form $h = c \times n^{-1/3}$ is chosen and results with $c = 0.5, 1.0, 1.5$ are reported. However, how to optimal the bandwidth h^* in order to maximize the performance of our test is beyond the scope of this paper.

Table 1 reports the empirical sizes of test under DGPs (S1)-(S3) at the 5% significance level using bootstrapped critical values. Our test has reasonable sizes in the case of small and moderate sample sizes. When sample size increases, the test shows fairly accurate sizes especially for the choice of $c = 1.0$. Power performance is reported in Table 2. For small sample size $n = 100$, the test is not powerful. However, it gains power rapidly for mildly large sample sizes. A general pattern we observe is that larger c delivers higher power but produces larger size distortion too.

It is important to mention that our test CvM_n is also applicable for i.i.d. data, as demonstrated already from the previous simulation result for DGP (S1). However, to further illustrate the level and power performance of the test in an i.i.d. context, we conduct the following simulation. We generate an i.i.d. sample $\{X_i, Y_i, Z_i\}_{i=1}^n$ using the DGP taken from Huang et al. (2013):

$$Y = \theta Z + X + \varepsilon_Y$$
$$Z = X + X^2 + \varepsilon_Z,$$

Table 3.1: Empirical size of CvM_n

Bandwidth	DGPs			
$h = cn^{-1/3}$	(S1)	(S2)	(S3)	(S4)
		$n = 100$		
$c = 0.5$	0.036	0.036	0.064	0.088
$c = 1.0$	0.066	0.068	0.052	0.072
$c = 1.5$	0.082	0.080	0.078	0.098
		$n = 200$		
$c = 0.5$	0.050	0.056	0.052	0.070
$c = 1.0$	0.042	0.064	0.052	0.064
$c = 1.5$	0.056	0.064	0.060	0.080
		$n = 500$		
$c = 0.5$	0.038	0.048	0.032	0.062
$c = 1.0$	0.050	0.051	0.054	0.050
$c = 1.5$	0.060	0.054	0.058	0.064

Table 3.2: Empirical power of CvM_n

Bandwidth	DGPs						
$h = cn^{-1/3}$	(P1)	(P2)	(P3)	(P4)	(P5)	(P6)	(P7)
				$n = 100$			
$c = 0.5$	0.972	0.112	0.280	0.220	0.140	0.126	0.104
$c = 1.0$	0.980	0.160	0.354	0.230	0.266	0.170	0.146
$c = 1.5$	0.990	0.178	0.408	0.284	0.270	0.164	0.162
				$n = 200$			
$c = 0.5$	1.000	0.370	0.862	0.608	0.360	0.216	0.124
$c = 1.0$	1.000	0.380	0.886	0.634	0.524	0.282	0.142
$c = 1.5$	1.000	0.388	0.910	0.622	0.432	0.300	0.186
				$n = 500$			
$c = 0.5$	1.000	0.912	1.000	0.998	0.988	0.808	0.408
$c = 1.0$	1.000	0.938	1.000	1.000	0.996	0.860	0.416
$c = 1.5$	1.000	0.942	1.000	1.000	1.000	0.886	0.456

where

$$\begin{pmatrix} \varepsilon_Y \\ \varepsilon_Z \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_Y^2 & 0 \\ 0 & \sigma_Z^2 \end{pmatrix}\right) = N\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}\right)$$

and

$$X \sim N(0, \sigma_X^2) = N(0, 3).$$

When $\theta = 0$, the null is true; otherwise the alternative holds.

As in Huang et al. (2013), we use the above DGP to study the finite sample size and power of the test against conditional mean dependence. We use $n = 50, 100, 200$ and bandwidth $h = 0.5n^{-1/3}$. Let

$$\rho_{Y,Z|X} = \frac{Cov(Y, Z|X)}{\sigma_{Y|X}\sigma_{Z|X}} = \frac{\theta\sigma_Z^2}{\sigma_Z\sqrt{\sigma_Y^2 + \theta^2\sigma_Z^2}} = \frac{2\theta}{\sqrt{4\theta^2 + 1}},$$

which indicates the strength of the dependence between Y and Z, conditional on X. Since both $Y|X$ and $Z|X$ are normal, $\rho_{Y,Z|X}$ fully captures the dependence between Y and Z,conditional on X. For given $\rho_{Y,Z|X}$, θ is determined by

$$\theta = \frac{\rho_{Y,Z|X}}{2\sqrt{1 - \rho_{Y,Z|X}^2}}.$$

Figure 1 plots the rejection frequency of test CvM_n for $\rho_{Y,Z|X} = -0.9, -0.8, \ldots, 0.8, 0.9$. The size and power look fairly satisfactory for small sample sizes. They look very good when the sample size reaches 200. When the sample size is small, the levels of Huang et al.'s (2013) test approaches 5% nominal value from below, delivering conservative test. In comparison to Huang et al.'s (2013) test, the levels of our test remain fairly accurate. However, for the specific bandwidth choice of $h = 0.5 \times n^{-1/3}$, when the sample size increases to 200, our test is slightly oversized.

3.8 An empirical study

In this section, we aim to examine and test whether there exists (nonlinear) predictability of equity risk premium using variance risk premium. The variance risk premium is defined as the difference between the risk-neutral and objective expectations of realized variance, where the risk-neutral expectation of variance is measured as the end-of-month Volatility Index-squared de-annualized and the realized variance is the sum of squared 5-minute log returns of the S&P 500 index over the month.

By now, there is a vast growing literature which focuses on its study on the predictive power of variance risk premium for the aggregate stock market returns, bond returns or exchange rate returns. For example, Bollerslev et al. (2009) first discover that variance risk premium is able to explain a nontrivial fraction of the time series variation in post 1990 aggregate stock market returns, with high (low) premia predicting high (low) future returns; Wang et al. (2013) find the empirical evidence suggesting that the firm-level variance risk premium has a prominent explanatory power for credit spreads in the presence of market- and firm-level control variables; by defining a "global" variance risk premium, Bollerslev et al. (2013) uncover stronger predictability of aggregate stock market returns using variance risk premium across countries; while Della Corte et al. (2013) investigate the predictive information content in foreign exchange volatility risk premia for exchange rate returns and find that a portfolio that sells currencies with high insurance costs and buys currencies with low insurance costs generates sizeable out-of-sample returns and Sharpe ratios.

We use monthly aggregate S&P 500 composite index over the period January 1996 to September 2008. Our empirical analysis is based on the logarithmic return on the S&P 500 in excess of the 3-month T-bill rate. The monthly variance risk premium is downloaded from Hao Zhou's website. Now, let $RP_{t+\tau}$ be the risk premium τ months ahead and VRP_t be the variance risk premium at time t. In this empirical study, we take $\tau =$1, 3, 6, and 9 months. We shall examine if the variance risk premium VRP_t explains (either in a linear or nonlinear fashion) the risk premium $RP_{t+\tau}$ given the information RP_t, which is equivalent to stating whether

VRP_t Granger causes RP_t by setting the lag order to τ. Formally, to test for the presence of (nonlinear) predictability of VRP_t, we consider the following null hypotheses

$$\mathrm{H}_0 : \Pr\left\{F(RP_{t+\tau}, VRP_t|RP_t) = F(RP_{t+\tau}|RP_t)F(VRP_t|RP_t)\right\} = 1$$

against the alternative hypothesis

$$\mathrm{H}_1 : \Pr\left\{F(RP_{t+\tau}, VRP_t|RP_t) = F(RP_{t+\tau}|RP_t)F(VRP_t|RP_t)\right\} < 1.$$

That is, for a given horizon τ, we test the conditional independence of $RP_{t+\tau}$ and VRP_t given RP_t, i.e. $RP_{t+\tau} \perp VRP_t | RP_t$.

For the purpose of comparison, we also perform the popular linear causality analysis in the literature. To this end, we consider the following linear regression model:

$$RP_{t+\tau} = \mu_\tau + \beta_\tau RP_t + \alpha_\tau VRP_t + \varepsilon_{t+\tau}.$$

The hypothesis of interest is that VRP does not Granger cause RP for τ months ahead in a linear way, i.e. testing the null hypothesis $\mathrm{H}_0 : \alpha_\tau = 0$ against the alternative hypothesis $\mathrm{H}_1 : \alpha_\tau \neq 0$. To test H_0, standard t-statistic given by $t_{\hat{\alpha}_\tau} = \hat{\alpha}_\tau / \hat{\sigma}_{\hat{\alpha}_\tau}$ will be calculated, where $\hat{\alpha}_\tau$ is the least squares estimator of α_τ and $\hat{\sigma}_{\hat{\alpha}_\tau}$ is the estimator of its standard error $\sigma_{\hat{\alpha}_\tau}$. Moreover, to avoid the impact of possible dependence in the residual terms $\hat{\varepsilon}_{t+\tau}$ on our inference, $\hat{\sigma}_{\hat{\alpha}_\tau}$ is calculated using the commonly used heteroscedasticity autocorrelation consistent (HAC) robust variance estimator suggested by Newey and West (1987).

Table 3 reports the testing results for Granger causality (nonlinear predictability) from variance risk premium to risk premium, at four different horizons, using our proposed test CvM_n and the linear test. The implementation is as in the Monte Carlo simulations part. Results from linear test fail to reject the null hypothesis of no causality, it only indicates some predictability until at the long 9 month horizon. On the other hand, by using our test, at the 5% significance level, we have found convincing evidence that risk premium can be predicted using variance risk premium at both mid-run and long-run horizons. We also find that there is a very high degree of predictability at horizons more than one-month which can be attributed to a nonlinear causal effect.

3.9 Conclusion

This paper proposes a new consistent test of conditional independence for data with weak dependence using the empirical process method. The asymptotic properties of the proposed test is developed. To implement the test in practice, a multiplier bootstrap procedure is suggested and its asymptotic validity is formally justified. The test can be applied to testing for possible conditional dependence in a wide variety of nonparametric models. Using the proposed test, we also study whether there exists some nonlinear predictability of equity risk premium using variance risk premium.

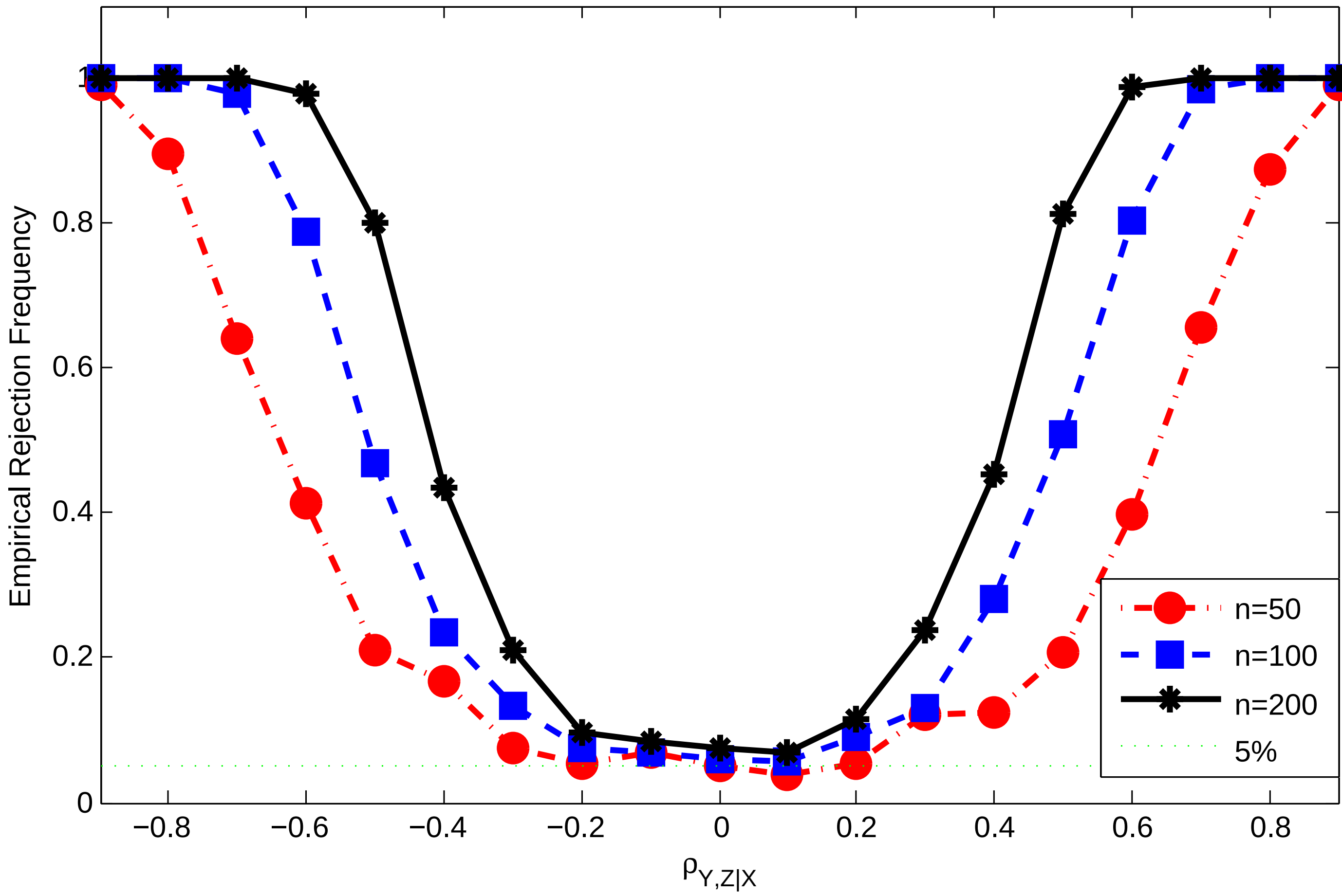
Further research is needed to examine the testing of conditional independence with data dependent bandwidths in order to maximize the performance of the tests. Allowing for the bandwidths to be data dependent instead of fixed ones in the testing of econometric restrictions, as studied by Li and Li (2010), may well be a useful approach to investigate.

Table 3.3: Testing (nonlinear) Granger causality from VRP to RP

Direction of Causality	$h = c \times n^{-1/3}$	CvM_n	p-value	LIN
		Horizon: One Month		
VRP→RP				
	$c = 0.5$	0.0057	0.2250	0.0017
		(0.0094)		(0.4172)
	$c = 1.0$	0.0050	0.1950	
		(0.0081)		
	$c = 1.5$	0.0050	0.1340	
		(0.0067)		
		Horizon: Three Months		
VRP→RP				
	$c = 0.5$	0.0319	0.0950	0.0016
		(0.0357)		(1.5009)
	$c = 1.0$	0.0312	0.0250	
		(0.0263)		
	$c = 1.5$	0.0277	0.0120	
		(0.0206)		
		Horizon: Six Months		
VRP→RP				
	$c = 0.5$	0.0976	0.0000	0.0002
		(0.0426)		(0.3329)
	$c = 1.0$	0.1160	0.0000	
		(0.0306)		
	$c = 1.5$	0.0953	0.0000	
		(0.0257)		
		Horizon: Nine Months		
VRP→RP				
	$c = 0.5$	0.4254	0.0000	0.0005
		(0.0597)		(2.0807)
	$c = 1.0$	0.3734	0.0000	
		(0.0406)		
	$c = 1.5$	0.2550	0.0000	
		(0.0303)		

Note: For test CvM_n, both bootstrapped critical values (in parentheses) at the 5% level and bootstrapped p-values are reported. LIN corresponds to the linear test, where the least squares estimate $\hat{\alpha}_\tau$ and its t-statistic $t_{\hat{\alpha}_\tau}$ (in parentheses) based on HAC robust variance estimator, are reported.

Figure 3.1: Power functions of CvM_n for DGP i.i.d. with nominal size 5%



Appendix A

In this appendix, we provide proofs of the main asymptotic results.

First, recall that for fixed (y, z), $\varepsilon_t(y, z) := 1(Y_t \leq y)[1(Z_t \leq z) - F_{Z|X}(z|X_t)]$. Denote

$$\gamma_n^0(x, y, z) = \frac{1}{n}\sum_{t=1}^{n} e_t(x, y, z),$$

where

$$e_t(x, y, z) = f_X(X_t)1(X_t \leq x)\phi_t(y)\epsilon_t(z)$$

with $\phi_t(y) = 1(Y_t \leq y) - F_{Y|X}(y|X_t)$ and $\epsilon_t(z) = 1(Z_t \leq z) - F_{Z|X}(z|X_t)$.

Proof of Theorem 1: By Lemma B.1, uniformly in (x, y, z),

$$\begin{aligned}\hat{S}_n(x, y, z) &= \sqrt{n}\gamma_n^0(x, y, z) + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(x, y, z) + o_p(1).\end{aligned}$$

Noticing that, under the null hypothesis , $e_t(x, y, z)$ forms a martingale differenced sequence with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$ for each $(x, y, z) \in \mathbb{R}^d$, that is, $E\left(e_t(x, y, z)|\mathcal{F}_t\right) = 0$ $\forall(x, y, z) \in \mathbb{R}^d$. Under Assumptions A.1–A.3, by applying a standard central limit theorem (CLT) for martingales, see e.g. Hall and Heyde (1980), it can be shown that the finite-dimensional distributions of the stochastic process $\hat{S}_n$ converge to those of S_∞, a Gaussian process with continuous sample paths and covariance function

$$\mathbb{E}\left(S_\infty(x_1, y_1, z_1), S_\infty(x_2, y_2, z_2)\right) = \mathbb{E}\left(e_1(x_1, y_1, z_1)e_1(x_2, y_2, z_2)\right).$$

We need to extend the convergence of the finite dimensional distributions of $\hat{S}_n$ to the weak convergence in $\ell^\infty(\overline{\mathbb{R}}^d)$, which is a direct consequence of Theorem A.1 in Delgado and Escanciano (2007). $\square$

Proof of Corollary 1: By continuous mapping theorem, see e.g. Billingsley (1968) Theorem 5.1. and the weak convergence of the process $\hat{S}_n$, we immediately have the convergence of KS_n in distribution.

For the null limit distribution of CvM_n, we write

$$\begin{aligned}
&|CvM_n - CvM_\infty| \\
=&\left|\int_{\mathbb{R}^d} \left(\hat{S}_n(x,y,z)\right)^2 d\hat{F}_n(x,y,z) - \int_{\mathbb{R}^d} \left(S_\infty(x,y,z)\right)^2 dF(x,y,z)\right| \\
\leq&\left|\int_{\mathbb{R}^d} \left(\hat{S}_n^2(x,y,z) - S_\infty^2(x,y,z)\right) d\hat{F}_n(x,y,z)\right| \\
&+\left|\int_{\mathbb{R}^d} S_\infty^2(x,y,z)\, d(\hat{F}_n(x,y,z) - F(x,y,z))\right|.
\end{aligned}$$

The first term of the right-hand side of the above inequality is $o_p(1)$ by Theorem 1. Glivenko-Cantelli's Theorem yields that $\sup_{(x,y,z)} |\hat{F}_n(x,y,z) - F(x,y,z)| = o(1)$ a.s. Then, taking into account that the trajectories of the limit process $S_\infty(x,y,z)$ are bounded and continuous almost surely and applying Helly-Bray Theorem (see p. 97 in Rao, 1965) to each of these trajectories, we obtain

$$\left|\int_{\mathbb{R}^d} S_\infty^2(x,y,z)\, d(\hat{F}_n(x,y,z) - F(x,y,z))\right| \to_{a.s.} 0$$

This concludes the proof of the Corollary. □

Proof of Theorem 2: We only prove the consistence result for CvM_n since the proof for KS_n is similar but easier. From Lemma B.2, we get

$$\begin{aligned}
\frac{1}{n}CvM_n =&\frac{1}{n}\int_{\mathbb{R}^d} \left(\sqrt{n}\hat{\gamma}_n(x,y,z)\right)^2 d\hat{F}_n(x,y,z) \\
=&\frac{1}{n}\int_{\mathbb{R}^d} \left(\sqrt{n}\left(\gamma_n^0(x,y,z) + \gamma_n^1(x,y,z)\right)\right)^2 d\hat{F}_n(x,y,z)[1 + o_p(1)] \\
=&(D_{1n} + D_{2n} + D_{3n})[1 + o_p(1)],
\end{aligned}$$

where

$$D_{1n} = \frac{1}{n} \int_{\mathbb{R}^d} \left(\sqrt{n} \gamma_n^0(x,y,z) \right)^2 \, d\hat{F}_n(x,y,z),$$

$$D_{2n} = \int_{\mathbb{R}^d} \left(\gamma_n^1(x,y,z) \right)^2 \, d\hat{F}_n(x,y,z),$$

and

$$D_{3n} = 2 \int_{\mathbb{R}^d} \gamma_n^0(x,y,z) \gamma_n^1(x,y,z) \, d\hat{F}_n(x,y,z).$$

We shall prove that $D_{1n} = o_p(1)$ and $D_{2n} = O_p(1)$. From Lemma B.1 and Theorem 1, we have that nD_{1n} converges in distribution. Thus, $D_{1n} = o_p(1)$. Noting that, for stationary and ergodic sequences,

$$\begin{aligned}
\gamma_n^1(x,y,z) &= \frac{1}{n} \sum_{t=1}^{n} f_X(X_t) 1(X_t \leq x) \Delta_t(y,z) \\
&\to_p E\left(f_X(X_1) 1(X_1 \leq x) \left[F_{Y,Z|X}(y,z|X_1) - F_{Y|X}(y|X_1) F_{Z|X}(z|X_1) \right] \right) \\
&:= \gamma(x,y,z),
\end{aligned}$$

we get

$$\begin{aligned}
D_{2n} &= \int_{\mathbb{R}^d} \left(\gamma_n^1(x,y,z) \right)^2 \, dF(x,y,z) + \int_{\mathbb{R}^d} \left(\gamma_n^1(x,y,z) \right)^2 \, d\left(\hat{F}_n(x,y,z) - F(x,y,z) \right) \\
&\to_p \int_{\mathbb{R}^d} \left(\gamma(x,y,z) \right)^2 \, dF(x,y,z) > 0,
\end{aligned}$$

where we have used the Glivenko-Cantelli's Theorem for stationary and ergodic sequences, i.e. $\sup_{(x,y,z)} |\hat{F}_n(x,y,z) - F(x,y,z)| = o(1)$ a.s. From the Cauchy-Schwartz's inequality, we also conclude that $D_{3n} = o_p(1)$. Hence, we finish the proof. $\square$

Proof of Theorem 3: Recall that $\hat{\varepsilon}_t(y,z) = 1(Y_t \leq y)[1(Z_t \leq z) - \hat{F}_{Z|X}(z|X_t)]$. Let's denote $\varepsilon_{nt}(y,z) = F_{Y,Z|X}(y,z|X_t) - F_{Y|X}(y|X_t)F_{Z|X}(z|X_t) - n^{-1/2}\Delta(X_t,y,z)$ under the local alternatives H_{1n} in (1.7). Then, by simple rearranging, we get

$$
\begin{aligned}
&\hat{\gamma}_n(x,y,z)\\
=&\frac{1}{n}\sum_{t=1}^{n}\hat{f}_X(X_t)1(X_t\le x)\varepsilon_{nt}(y,z)\\
&+\frac{1}{n}\sum_{t=1}^{n}\hat{f}_X(X_t)1(X_t\le x)\left[\hat{\varepsilon}_t(y,z)-F_{Y,Z|X}(y,z|X_t)+F_{Y|X}(y|X_t)F_{Z|X}(z|X_t)\right]\\
&+\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{t=1}^{n}\hat{f}_X(X_t)1(X_t\le x)\Delta(X_t,y,z)\\
:=&\hat{\gamma}_{1n}(x,y,z)+\hat{\gamma}_{2n}(x,y,z)+\frac{1}{\sqrt{n}}\hat{\gamma}_{3n}(x,y,z).
\end{aligned}
$$

First of all, by utilizing standard kernel estimation theory, we can show that

$$
\hat{\gamma}_{1n}(x,y,z)=\frac{1}{n}\sum_{t=1}^{n}f_X(X_t)1(X_t\le x)\varepsilon_{nt}(y,z)+o_p\left(\frac{1}{\sqrt{n}}\right)
$$

Now it is clear to see that under the local alternatives H_{1n}, the sequence of innovations ε_{nt} is a martingale difference sequence with respect to the σ-field $\mathscr{F}_t$. Then, following similar arguments as in the proof of Theorem 5, we can immediately prove that

$$
\hat{S}_{1n} := \sqrt{n}\hat{\gamma}_{1n} \Rightarrow S_\infty,
$$

where S_∞ is the same zero mean Gaussian process as in Theorem 5.

For the term $\hat{\gamma}_{2n}(x, y, z)$, we notice that it can be decompose into three parts,

$$
\begin{aligned}
&\hat{\gamma}_{2n}(x, y, z) \\
=&\frac{1}{n}\sum_{t=1}^{n}\hat{f}_X(X_t)1(X_t \leq x)\left[1(Y_t \leq y)1(Z_t \leq z) - F_{Y,Z|X}(y, z|X_t)\right] \\
&-\frac{1}{n}\sum_{t=1}^{n}\hat{f}_X(X_t)1(X_t \leq x)\left[1(Y_t \leq y) - F_{Y|X}(y|X_t)\right]F_{Z|X}(z|X_t) \\
&-\frac{1}{n}\sum_{t=1}^{n}\hat{f}_X(X_t)1(X_t \leq x)1(Y_t \leq y)\left[\hat{F}_{Z|X}(z|X_t) - F_{Z|X}(z|X_t)\right] \\
:=&\hat{\gamma}_{2n}^{1}(x, y, z) - \hat{\gamma}_{2n}^{2}(x, y, z) - \hat{\gamma}_{2n}^{3}(x, y, z).
\end{aligned}
$$

It suffices to establish that, uniformly in (x, y, z), all three terms $\hat{\gamma}_{2n}^{j}(x, y, z) = o_p(n^{-1/2})$ for $j = 1, 2, 3$. This is achieved by using the uniform ergodic theorem and the Glivenko-Cantelli's Theorem for ergodic and stationary time series.

Finally, for the third term, by standard law of large numbers for stationary sequences, it is straightforward to show that $\hat{\gamma}_{3n}(x, y, z)$ converges in probability to a deterministic shift function $G_\infty(x, y, z)$, where

$$G_\infty(x, y, z) = E\left[f_X(X_1)1(X_1 \leq x)\Delta(X_1, y, z)\right].$$

Thus, under the local alternatives H_{1n} stated in (1.7),

$$\hat{S}_n := \sqrt{n}\hat{\gamma}_n \Rightarrow S_\infty + G_\infty$$

by Slutsky's Theorem. we conclude the proof of the theorem. □

Proof of Theorem 4: Define

$$\hat{S}_n^{0*}(x, y, z) = \sqrt{n}\hat{\gamma}_n^{0*}(x, y, z),$$

with

$$\hat{\gamma}_n^{0*}(x,y,z) = \frac{1}{n}\sum_{t=1}^{n} f_X(X_t) 1(X_t \le x) \left[1(Y_t \le y) - F_{Y|X}(y|X_t)\right] \\ \times \left[1(Z_t \le z) - F_{Z|X}(z|X_t)\right] v_t,$$

where $\{v_t\}_{t=1}^n$ is a sequence of independent random variables with zero mean, unit variance and is independent of the original sample.

It suffices to prove that, uniformly in (x,y,z), the process $\hat{S}_n^*(x,y,z)$ and the process $\hat{S}_n^{0*}(x,y,z)$ are asymptotically equivalent, i.e.

$$\sup_{(x,y,z)\in\mathbb{R}^d} \left|\sqrt{n}\left[\hat{\gamma}_n^*(x,y,z) - \hat{\gamma}_n^{0*}(x,y,z)\right]\right| = o_p(1).$$

To achieve this, first note that

$$\begin{aligned}
&\sqrt{n}\left[\hat{\gamma}_n^*(x,y,z) - \hat{\gamma}_n^{0*}(x,y,z)\right] \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\hat{f}_X(X_t) - f_X(X_t)\right] 1(X_t \le x)\left[1(Y_t \le y) - \hat{F}_{Y|X}(y|X_t)\right] \\
&\times\left[1(Z_t \le z) - \hat{F}_{Z|X}(z|X_t)\right] \\
&-\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_X(X_t) 1(X_t \le x)\left[\hat{F}_{Y|X}(y|X_t) - F_{Y|X}(y|X_t)\right]\left[1(Z_t \le z) - F_{Z|X}(z|X_t)\right] \\
&-\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_X(X_t) 1(X_t \le x)\left[\hat{F}_{Z|X}(z|X_t) - F_{Z|X}(z|X_t)\right]\left[1(Y_t \le y) - F_{Y|X}(y|X_t)\right] \\
&+\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_X(X_t) 1(X_t \le x)\left[\hat{F}_{Y|X}(y|X_t) - F_{Y|X}(y|X_t)\right]\left[\hat{F}_{Z|X}(z|X_t) - F_{Z|X}(z|X_t)\right] \\
:=&B_{1n} - B_{2n} - B_{3n} + B_{4n}.
\end{aligned}$$

We shall show that $B_{jn} = o_p(1)$, for $j = 1, \ldots, 4$, uniformly in (x,y,z). Let C denote a generic positive and bounded constant. For term B_{1n}, the expectation of its square is bounded by

$$C \cdot E\left(\hat{f}_X(X_t) - f_X(X_t)\right)^2 \to 0$$

since the mean squared errors of kernel estimator $\hat{f}_X(X_t)$ vanishes asymptotically. Thus, by Markov's inequality, $B_{1n} = o_p(1)$.

The proof that B_{2n} and B_{3n} are asymptotically negligible is similar, so we only show $B_{2n} = o_p(1)$. Again, the expectation of B_{2n}^2 is bounded by

$$C \cdot E\left(\hat{F}_{Y|X}(y|X_t) - F_{Y|X}(y|X_t)\right)^2 \to 0.$$

The proof of $B_{4n} = o_p(1)$ follows the same steps. □

Appendix B

In the first lemma we obtain the asymptotic behaviour for $\hat{\gamma}_n(x, y, z)$ under the null hypothesis.

Lemma B.1: Under Assumptions A.1-A.3 and the null hypothesis,

$$\sup_{(x,y,z)\in\mathbb{R}^d} \left|\sqrt{n}\left[\hat{\gamma}_n(x, y, z) - \gamma_n^0(x, y, z)\right]\right| = o_p(1),$$

with

$$\gamma_n^0(x, y, z) = \frac{1}{n}\sum_{t=1}^{n} f_X(X_t)1(X_t \leq x)\phi_t(y)\epsilon_t(z),$$

where

$$\phi_t(y) = 1(Y_t \leq y) - F_{Y|X}(y|X_t),$$

and

$$\epsilon_t(z) = 1(Z_t \leq z) - F_{Z|X}(z|X_t).$$

Proof of Lemma B.1: In the sequel, for notational simplicity, we will suppress the dependence on (x, y, z) of quantities like $\hat{\gamma}_n(x, y, z)$. We first express $\hat{\gamma}_n$ as a form of U-statistic. To this end, define $\mathcal{W}_t = (X_t, Y_t, Z_t)$ and introduce

$$\begin{aligned} U(\mathcal{W}_t, \mathcal{W}_s) =& \frac{1}{2}\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X_s}{h}\right)1(X_t \leq x)1(Y_t \leq y)[1(Z_t \leq z) - 1(Z_s \leq z)]\right. \\ &\left. + \frac{1}{h^{d_x}}K\left(\frac{X_t - X_s}{h}\right)1(X_s \leq x)1(Y_s \leq y)[1(Z_s \leq z) - 1(Z_t \leq z)]\right]. \end{aligned}$$

Therefore, we rewrite

$$\sqrt{n}\hat{\gamma}_n = \sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n} U(\mathcal{W}_t, \mathcal{W}_s),$$

We shall use the technique of Hoeffding's decomposition to study the asymptotic behaviour of a generalized U-statistic like the above one, .

By standard U-statistic theory, see Serfling (1980) for further details, we first need to calculate the projection term of $U(\mathcal{W}_t, \mathcal{W}_s)$ given $\mathcal{W}_s$. Denote $\mathcal{W} = (X, Y, Z)$. We get

$$\begin{aligned} U_1(\mathcal{W}) =& E[U(\mathcal{W}_t, \mathcal{W}_s)|\mathcal{W}_s = \mathcal{W}] \\ =& \frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)1(X_t \leq x)F_{Y|X}(y|X_t)F_{Z|X}(z|X_t)\right] \\ & - \frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)1(X_t \leq x)1(Y_t \leq y)\right]1(Z \leq z) \\ & + \frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)\right]1(X \leq x)1(Y \leq y)1(Z \leq z) \\ & - \frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)F_{Z|X}(z|X_t)\right]1(X \leq x)1(Y \leq y) \\ :=& \sum_{j=1}^{4} A_{1j}, \end{aligned}$$

where we have applied the law of iterated expectation and the fact that $E(1(Z_t \leq z)|X_t) = F_{Z|X}(z|X_t)$, and under the null hypothesis, $F_{Y,Z|X}(y, z|X_t) = F_{Y|X}(y|X_t)F_{Z|X}(z|X_t)$.

Now denote $f_{Y,X}(y, x)$ to be the joint density of (Y, X). We further could calculate that

$$2A_{11} = 1(X \leq x)F_{Y|X}(y|X)F_{Z|X}(z|X)f_X(X) + O(h^l),$$

$$2A_{12} = -1(X \leq x)F_{Y|X}(y|X)1(Z \leq z)f_X(X) + O(h^l),$$

$$2A_{13} = 1(X \leq x)1(Y \leq y)1(Z \leq z)f_X(X) + O(h^l),$$

and

$$2A_{14} = -1(X \leq x)1(Y \leq y)F_{Z|X}(z|X)f_X(X) + O(h^l),$$

so that

$$2U_1(\mathcal{W}) = 1(X \leq x)[1(Y \leq y) - F_{Y|X}(y|X)][1(Z \leq z) - F_{Z|X}(z|X)]f_X(X) + O(h^l).$$

Thus, $U_1(\mathcal{W}_t) := \xi_t/2 + O_p(h^l) := 1(X_t \leq x)\phi_t(y)\epsilon_t(z)f_X(X_t)/2 + O_p(h^l)$. Now, according to U-statistic theory, let's denote $\psi(\mathcal{W}_t, \mathcal{W}_s) = U(\mathcal{W}_t, \mathcal{W}_s) - U_1(\mathcal{W}_t) - U_1(\mathcal{W}_s)$. By Hoeffding's decomposition, $\sqrt{n}\hat{\gamma}_n(x, y, z)$ can be written as

$$\begin{aligned}
&\sqrt{n}\hat{\gamma}_n(x, y, z) \\
=&\frac{2}{\sqrt{n}}\sum_{t=1}^{n} U_1(\mathcal{W}_t) + \sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi(\mathcal{W}_t, \mathcal{W}_s) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_t + \sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi(\mathcal{W}_t, \mathcal{W}_s) + O_p((nh^{2L})^{1/2}).
\end{aligned}$$

It is clear that $O_p((nh^{2l})^{1/2}) = o_p(1)$ by Assumption A.3. We finish the proof about the uniform expansion of $\sqrt{n}\hat{\gamma}_n(x, y, z)$ by showing that

$$\sqrt{n}\frac{1}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi(\mathcal{W}_t, \mathcal{W}_s) = o_p(1),$$

uniformly in (x, y, z). By Markov's inequality, this can be achieved by proving

$$\mathrm{E}\left[\left|\sqrt{n}\frac{1}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi(\mathcal{W}_t, \mathcal{W}_s)\right|^2\right] \to 0. \tag{3.12}$$

The following proof is similar to that of Theorem 1 in Nishiyama et al. (2011). Observing first the components from left-hand side of (3.12), we only need to show that, separately, the following three terms are all asymptotically negligible, that is,

(a) the double summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi(\mathcal{W}_t, \mathcal{W}_s)^2\right\} \to 0, \tag{3.13}$$

(b) the triple summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-2}\sum_{s=t+1}^{n-1}\sum_{u=s+1}^{n}\psi(\mathcal{W}_t, \mathcal{W}_s)\psi(\mathcal{W}_t, \mathcal{W}_u)\right\} \to 0, \tag{3.14}$$

and (c) the quadruple summation term

$$\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-3}\sum_{s=t+1}^{n-2}\sum_{u=s+1}^{n-1}\sum_{v=u+1}^{n}\psi(\mathcal{W}_t, \mathcal{W}_s)\psi(\mathcal{W}_u, \mathcal{W}_v)\right\} \to 0. \tag{3.15}$$

Furthermore, in order to show (3.15), we have to consider two different scenarios. Specifically speaking, case $(c.1)$ with $t < s < u < v$ and $s - t > v - u$, and case $(c.2)$ $t < s < u < v$ and $s - t \leq v - u$. Since both cases are of the same spirits and can be handled similarly, we simply have to focus on case $(c.1)$, i.e., we prove

$$\mathrm{E}\left\{\frac{1}{n^3}\sum\sum\sum\sum_{t<s<u<v, s-t>v-u}\psi(\mathcal{W}_t, \mathcal{W}_s)\psi(\mathcal{W}_u, \mathcal{W}_v)\right\} \to 0. \tag{3.16}$$

One useful lemma from Yoshihara (1976) for strictly stationary mixing processes will be exploited. For illustration, we only prove (3.13). First of all, we calculate

$$\begin{aligned}
&\int |\psi(w_1,w_2)|^{2+\delta}\, dF(w_1)dF(w_2) \\
\leq& C\int |U(w_1,w_2)|^{2+\delta}\, dF(w_1)dF(w_2) + C \\
\leq& C\int \frac{1}{h^{d_x(2+\delta)}}\left|K\left(\frac{X_1-X_2}{h}\right)\right|^{2+\delta}\, dF(w_1)dF(w_2) + C \\
\leq& C\{h^{-d_x(1+\delta)}+1\} := C_{1h} \qquad (3.17)
\end{aligned}$$

The first inequality holds because $\mathrm{E}|U_1(\mathcal{W}_t)|^{2+\delta}$ is of smaller order than $\mathrm{E}|U(\mathcal{W}_t,\mathcal{W}_s)|^{2+\delta}$. By change of variables, we get the last equality. Lemma B.1 yields

$$\begin{aligned}
&\left|\int \psi(w_1,w_2)^2\, dF_{1,2}(w_1,w_2) - \int \psi(w_1,w_2)^2\, dF(w_1)dF(w_2)\right| \\
\leq& 4C_{1h}^{2/(2+\delta)}\beta(s-t)^{\delta/(2+\delta)} \\
\leq& Ch^{-2d_x(1+\delta)/(2+\delta)}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta}. \qquad (3.18)
\end{aligned}$$

Since when $\delta=0$, the inequality in (3.17) holds too, so that

$$\int \psi(w_1,w_2)^2\, dF(w_1)dF(w_2) \leq C(h^{-d_x}+1) = O(h^{-d_x}). \qquad (3.19)$$

Combing inequalities (3.19) and (3.18), we have

$$\begin{aligned}
\mathrm{E}\{\psi(\mathcal{W}_t,\mathcal{W}_s)^2\} &= \int \psi(w_1,w_2)^2\, dF_{1,2}(w_1,w_2) \\
&\leq C\left\{h^{-2d_x(1+\delta)/(2+\delta)}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta} + h^{-d_x}\right\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathrm{E}\left\{\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi(\mathcal{W}_t,\mathcal{W}_s)^2\right\}\\
=&\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\mathrm{E}\{\psi(\mathcal{W}_t,\mathcal{W}_s)^2\}\\
\leq&C\left(\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\left\{h^{-2d_x(1+\delta)/(2+\delta)}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta}+h^{-d_x}\right\}\right)\\
=&C\left(h^{-2d_x(1+\delta)/(2+\delta)}\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta}+h^{-d_x}\frac{1}{n^3}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}1\right)\\
\leq&C\left(\frac{1}{n^2h^{2d_x(1+\delta)/(2+\delta)}}+\frac{1}{nh^{d_x}}\right)\\
=&C\left(\frac{h^{2d_x/(2+\delta)}}{(nh^{d_x})^2}+\frac{1}{nh^{d_x}}\right),
\end{aligned}$$

which goes to zero because of Assumption A.3. The last inequality holds because

$$\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}(s-t)^{-(2+\eta)\delta/(2+\delta)\eta}=\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}(s-t)^{-1-\gamma}\leq Cn,$$

because $\gamma=2(\delta-\eta)\delta/((2+\delta)\eta)>0$ by picking some $\delta>\eta$. The proofs that terms (3.14) and (3.16) are all asymptotically negligible, are similar but lengthy. We hence omit the detailed steps. □

The next lemma establishes the asymptotic representation of $\hat{\gamma}_n(x,y,z)$ under the alternative hypothesis.

Lemma B.2: Under Assumptions A.1-A.3 and the alternative hypothesis,

$$\sup_{(x,y,z)\in\mathbb{R}^d}\left|\sqrt{n}\left[\hat{\gamma}_n(x,y,z)-\gamma_n^0(x,y,z)-\gamma_n^1(x,y,z)\right]\right|=o_p(1),$$

with

$$\gamma_n^1(x,y,z) = \frac{1}{n}\sum_{t=1}^{n} f_X(X_t)1(X_t \le x)\Delta_t(y,z),$$

where

$$\Delta_t(y,z) = F_{Y,Z|X}(y,z|X_t) - F_{Y|X}(y|X_t)F_{Z|X}(z|X_t),$$

and $\gamma_n^0(x,y,z)$ is defined the same as in Lemma B.1.

Proof of Lemma B.2: The proof of Lemma B.2 follows by using the similar steps as in Lemma B.1. Hence we omit the details. Instead, we only provide the calculation of the projection term of $U(\mathcal{W}_t,\mathcal{W}_s)$ given $\mathcal{W}_s = \mathcal{W}$ under the alternative hypothesis. Formally,

$$\begin{aligned}
&U_1(\mathcal{W}) = E[U(\mathcal{W}_t,\mathcal{W}_s)|\mathcal{W}_s = \mathcal{W}]\\
=&\frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)1(X_t \le x)1(Y_t \le y)(1(Z_t \le z) - F_{Z|X}(z|X_t))\right]\\
&+\frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)1(X_t \le x)1(Y_t \le y)F_{Z|X}(z|X_t)\right]\\
&-\frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)1(X_t \le x)1(Y_t \le y)\right]1(Z \le z)\\
&+\frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)\right]1(X \le x)1(Y \le y)1(Z \le z)\\
&-\frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)F_{Z|X}(z|X_t)\right]1(X \le x)1(Y \le y)\\
:=&\sum_{j=0}^{4} A_{1j},
\end{aligned}$$

where A_{1j} for $j = 1,\ldots,4$ is defined to be the same as in Lemma B.1 and A_{10} is a non-trivially shifted term because of the alternative hypothesis. To see this, notice that by law of iterated expectation, A_{10} can be rewritten as

$$
\begin{aligned}
&A_{10} \\
=&\frac{1}{2}E\left[\frac{1}{h^{d_x}}K\left(\frac{X_t - X}{h}\right)1(X_t \le x)(F_{Y,Z|X}(y,z|X_t) - F_{Y|X}(y|X_t)F_{Z|X}(z|X_t))\right].
\end{aligned}
$$

Clearly, A_{10} will be identically zero when under the null hypothesis of conditional independence of Y and Z given X, hence we go back to Lemma B.1 when under the null. Furthermore, it is not too difficult to calculate that

$$
A_{10} = \frac{1}{2}f_X(X)1(X \le x)[F_{Y,Z|X}(y,z|X) - F_{Y|X}(y|X)F_{Z|X}(z|X)] + O(h^L).
$$

Now, we have

$$
\begin{aligned}
&U_1(\mathcal{W}_t) \\
=&\frac{1}{2}f_X(X_t)1(X_t \le x)\phi_t(y)e_t(z) + \frac{1}{2}f_X(X_t)1(X_t \le x)\Delta_t(y,z) + O_p(h^L)
\end{aligned}
$$

Therefore, again by U-statistic theory and Hoeffding's decomposition, if let $\psi^1(\mathcal{W}_t, \mathcal{W}_s) = U(\mathcal{W}_t, \mathcal{W}_s) - U_1(\mathcal{W}_t) - U_1(\mathcal{W}_s)$, now $\sqrt{n}\hat{\gamma}_n(x,y,z)$ can be rewritten as

$$
\begin{aligned}
&\sqrt{n}\hat{\gamma}_n(x,y,z) \\
=&\frac{2}{\sqrt{n}}\sum_{t=1}^{n} U_1(\mathcal{W}_t) + \sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi^1(\mathcal{W}_t, \mathcal{W}_s) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_X(X_t)1(X_t \le x)\phi_t(y)e_t(z) \\
&+ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} f_X(X_t)1(X_t \le x)\Delta_t(y,z) \\
&+ \sqrt{n}\frac{2}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi^1(\mathcal{W}_t, \mathcal{W}_s) + O_p((nh^{2L})^{1/2}).
\end{aligned}
$$

It is important to remark that the additional shift term $\gamma_n^1(x,y,z)$ converges in probability to $\gamma(x,y,z)$ by law of large numbers, where $\gamma(x,y,z)$, under the alternative hypothesis, is not identically zero for a positive Lebesgue measure of $(x, y, z) \in \mathbb{R}^d$. This fact will guarantee the consistency of our test. See the proof of the theorem below.

To finish the proof of Lemma B.1, it suffices to prove that, uniformly in (x, y, z)

$$\sqrt{n}\frac{1}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi^1(\mathcal{W}_t, \mathcal{W}_s) = o_p(1),$$

which, by Markov's inequality, is equivalent to showing

$$\mathrm{E}\left[\left(\sqrt{n}\frac{1}{n(n-1)}\sum_{t=1}^{n-1}\sum_{s=t+1}^{n}\psi^1(\mathcal{W}_t, \mathcal{W}_s)\right)^2\right] \to 0. \tag{3.20}$$

Now we use again the arguments for (3.12) in the proof of (3.20) (here with $\psi^1(\mathcal{W}_t, \mathcal{W}_s)$ in the place of $\psi(\mathcal{W}_t, \mathcal{W}_s)$). We conclude the proof. $\square$

3.10 References

Bakirov N.K., Rizzo M.L. and Székely G.J. (2006). "A multivariate nonparametric test of independence", *Journal of Multivariate Analysis*, 97, 1742-1756.

Bierens H.J. (1982). "Consistent model specification tests", *Journal of Econometrics*, 20, 105-134.

Billingsley P. (1968). *Convergence of Probability Measures*, Wiley Series in Probability and Statistics, Wiley, New York.

Bollerslev T., Marrone J., Xu L. and Zhou H. (2013). "Stock return predictability and variance risk premia: statistical inference and international evidence", *Journal of Financial and Quantitative Analysis*, forthcoming.

Bollerslev T., Tauchen G. and Zhou H. (2009). "Expected stock returns and variance risk premia", *Review of Financial Studies*, 22, 4463-4492.

Bouezmarni T. and Taamouti A. (2012). "Nonparametric tests for conditional independence using conditional distributions", *Working Paper*, Universidad Carlos III de Madrid.

Blum J.R., Kiefer J. and Rosenblatt M. (1961). "Distribution free tests of independence based on the sample distribution function", *Annals of Mathematical Statistics*, 32, 485-498.

Bouezmarni T., Rombouts J.V.K. and Taamouti A. (2012). "Nonparametric copula-based test for conditional independence with applications to Granger causality", *Journal of Business and Economic Statistics*, 30, 275-287.

Csörgö S. (1985). "Testing for independence by the empirical characteristic function", 16, 290-299.

Dawid A.P. (1979). "Conditional independence in statistical theory", *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 41, 1-31.

de Jong R. (1996). "The Bierens test under data dependence", *Journal of Econometrics*, 72, 1-32.

Delgado M.A. and Escanciano J.C. (2007). "Nonparametric tests for conditional symmetry in dynamic models", *Journal of Econometrics*, 141, 652-682.

Delgado M.A. and González-Manteiga W. (2001). "Significance testing in nonparametric regression based on the bootstrap", *Annals of Statistics*, 29, 1469-1507.

Delgado M.A. and Mora J. (2000). "A nonparametric test for serial independence of regression errors", *Biometrika*, 87, 228-234.

Della Corte P., Ramadorai T. and Sarno L. (2013). "Volatility risk premia and exchange rate predictability", *Working paper*, Imperial College London.

Dette H., Guhlich M. and Neumeyer N. (2014). "Testing for additivity in nonparametric quantile regression", *Annals of the Institute of Statistical Mathematics*, forthcoming.

Diks C. and Panchenko V. (2007). "Nonparametric tests for serial independence based on quadratic forms", *Statistica Sinica*, 17, 81-98.

Diks C. (2009). "Nonparametric tests for independence", 6252-6271, *In: Meyers R.A. (Eds.), Encyclopedia of Complexity and Systems Science*, Springer, New York.

Doukhan P. (1994). *Mixing: Properties and Examples*, New York: Springer-Verlag.

Escanciano J.C. and Velasco C. (2006). "Testing the martingale difference hypothesis using integrated regression functions", *Computational Statistics and Data Analysis*, 51, 2278-2294.

Fan J. and Yao Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer.

Fan Y. and Li Q. (1996). "Consistent model specification tests: Omitted variables and semiparametric functional forms", *Econometrica*, 64, 865-890.

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

Ghoudi K., Kulperger R.J. and Rémillard B. (2001). "A nonparametric test of serial independence for time series and residuals", *Journal of Multivariate Analysis*, 79, 191-218.

Gu J., Li D. and Liu D. (2007). "Bootstrap non-parametric significance test", *Journal of Nonparametric Statistics*, 19, 215-230.

Hall P. and Heyde C.C. (1980). *Martingale Limit Theory and its Applications*, Wiley Series in Probability and Statistics, Wiley, New York.

Hall P., Wolff C.L. and Yao Q. (1999). "Methods for estimating a conditional distribution function", *Journal of the American Statistical Association*, 94, 154-163.

Hoeffding W. (1948). "A non-parametric test of independence", *Annals of Mathematical Statistics*, 19, 546-557.

Hong Y. and White H. (2005). "Asymptotic distribution theory for nonparametric entropy measures of serial dependence", *Econometrica*, 73, 837-901.

Horowitz J.L. and Spokoiny V. (2001). "An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative", *Econometrica*, 69, 599-631.

Huang M, Sun Y. and White H. (2013). "A flexible nonparametric test for conditional independence", *Working paper*, University of California, San Diego.

Huang T.-M. (2010). "Testing conditional independence using maximal nonlinear conditional correlation", *Annals of Statistics*, 38, 2047-2091.

Jeong K., Härdle W.K. and Song S. (2012). "A consistent nonparametric test for causality in quantile", *Econometric Theory*, 28, 861-887.

Li D. and Li Q.(2010). "Nonparametric/semiparametric estimation and testing of econometric models with data dependent smoothing parameters", *Journal of Econometrics*, 157, 179-190.

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

Linton O. and Gozalo P. (1996). "Conditional independence restrictions: testing and estimation", *Cowles Foundation Discussion Paper No. 1140*, Yale University.

Liu R.Y. (1988). "Bootstrap procedures under some non-i.i.d models", *Annals of Statistics*, 16, 1696-1708.

Mammen E. (1993). "Bootstrap and wild bootstrap for high dimensional linear models", *Annals of Statistics*, 21, 255-285.

Nadaraya E.A. (1964). "On estimating regression", *Theory of Probability and its Applications*, 9, 141-142.

Newey W. and West D. (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix”, *Econometrica*, 55, 703-708.

Nishiyama Y., Hitomi K., Kawasaki Y. and Jeong K. (2011). “A consistent nonparametric test for nonlinear causality - Specification in time series regression”, *Journal of Econometrics*, 165, 112-127.

Powell J., Stock J. and Stoker T. (1989). “Semiparametric estimation of index coefficients”, *Econometrica*, 57, 1403-1430.

Rao C.R. (1965). *Linear Statistical Inference and its Applications*, Wiley Series in Probability and Statistics, Wiley, New York.

Robinson P. (1988). “Root-N-consistent semiparametric regression”, *Econometrica*, 56, 931-954.

Robinson P. (1991). “Consistent nonparametric entropy-based testing”, *Review of Economic Studies*, 58, 437-453.

Rosenblatt M. (1975). “A quadratic measure of deviation of two-dimensional density estimates and a test of independence”, *Annals of Statistics*, 3, 1-14.

Serfling R.J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Statistics, Wiley, New York.

Skaug H.J. and Tjøstheim D. (1993). “A nonparametric test of serial independence based on the empirical distribution function”, *Biometrika*, 80, 591-602.

Song K. (2009). “Testing conditional independence via Rosenblatt transforms”, *Annals of Statistics*, 37, 4011-4045.

Stute W. (1994). “U-statistic processes: A martingale approach”, *Annals of Probability*, 22, 1725-1744.

Stute W. (1997). “Nonparametric model checks for regression”, *Annals of Statistics*, 25, 613-641.

Su L. and White H. (2007). “A consistent characteristic function-based test for conditional independence”, *Journal of Econometrics*, 141, 807-834.

Su L. and White H. (2008). “Nonparametric Hellinger metric test for conditional independence”, *Econometric Theory*, 24, 829-864.

Su L. and White H. (2011). “Testing conditional independence via empirical likelihood”, *Working paper*, Singapore Management University.

Székely G.J., Rizzo M.L. and Bakirov N.K. (2007). “Measuring and testing dependence by correlation of distances”, *Annals of Statistics*, 35, 2769-2794.

Wang H., Zhou H. and Zhou Y. (2013). “Credit default swap spreads and variance risk premia”, *Journal of Banking and Finance*, 37, 3733-3746.

Wang X. and Hong Y. (2012). “Characteristic function based testing for conditional independence: A nonparametric regression approach”, *Working paper*, Xiamen University.

Watson G.S. (1964). “Smooth regression analysis”, *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 359-372.

Yoshihara K. (1976). “Limiting behavior of U-statistics for stationary, absolutely regular processes”, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwande Gebiete*, 35, 237-252.

Yu K. and Jones M.C. (1998). “”Local linear quantile regression”, *Journal of the American Statistical Association*, 93, 228-237.

Zheng J.X. (1998). “Consistent specification testing for conditional symmetry”, *Econometric Theory*, 14, 139-149.