Delgado, M. A., & Song, X. (2018). Nonparametric tests for conditional symmetry. *Journal of Econometrics*, 206 (2), pp. 447-471.

DOI: 10.1016/j.jeconom.2018.06.010

Nonparametric Tests for Conditional Symmetry[1]

Miguel A. Delgado[2] Xiaojun Song [3]
Universidad Carlos III de Madrid Peking University

This version: October 23th 2017

Abstract: We propose omnibus tests for symmetry of the conditional distribution of a time series process about a nonparametric regression function. The test statistic is a weighted version of the integrated squared difference between the restricted and unrestricted estimators of the joint characteristic function of nonparametric residuals and explanatory variables, whose critical values are estimated with the assistance of a bootstrap technique. The test is sensitive to local alternatives converging to the null at the parametric rate $T^{-1/2}$, with T the sample size. We investigate the finite sample performance of the test by means of Monte Carlo experiments and two empirical applications to test whether losses are more likely than gains in financial markets, and whether expansions and contractions are equally likely in business cycles, given the relevant information.

Keywords: Conditional symmetry; Nonparametric testing; Permutation; Smoothing; Time series data.

JEL classification: C12; C14; C15.

[1]We are most thankful to the co-editors and two referees for their many insightful comments and suggestions, which have led to an improved version of the article. Support from Ministerio Economía y Competitividad (Spain), grant numbers ECO2014-55858-P and MDM 2014-0431, Comunidad de Madrid (Spain), MadEco-CM grant number S2015/HUM-3444, and the National Natural Science Foundation of China, grant number 71532001.

[2]Departamento de Economía, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain. E-mail: delgado@est-econ.uc3m.es.

[3]Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. E-mail: sxj@gsm.pku.edu.cn.

1 Introduction

This article proposes testing that the conditional distribution of an ergodic and stationary time series process given the available information is symmetric around a nonparametric regression function. The test is valid under fairly general conditions, which include situations where conditional location-scale models are too restrictive.

Testing for conditional inference is well motivated. Statistical inferences can be improved under the conditional symmetry assumption. Location and dispersion can be unambiguously defined under symmetry, and the center of symmetry can be robustly, even adaptively estimated. Also, testing for conditional symmetry is a useful specification tool, since most popular specifications are ruled out when the hypothesis of symmetry is rejected.

Conditional symmetry is also a relevant feature worth testing in econometrics and nonparametric modelling. 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. Checking asymmetries in business cycles is also well motivated. 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 behavior 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), and many alternative proposals have been proposed since then. Testing symmetry of the unknown marginal distribution after centering is an effective way of testing conditional symmetry when innovations are independent of the 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), Bhattacharya et al. (1982), Hušková (1984), Koziol (1985), Csörgo and Heathcote (1987), Schuster and Barker (1987), Hollander (1988), Ahmad and Li (1997), Henze et al. (2003) , Psaradakis (2003) and Fang et al. (2015). These tests are unable to detect infinitely many departures from the conditional symmetry hypothesis where innovations are not independent of the explanatory variables, e.g. under conditional heteroskedasticity. The hypothesis of independence between innovations and explanatory variables has been relaxed to allow conditional location-scale 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. The resulting test is in fact a specification test of the conditional symmetry center when innovations are independent of covariates. Dette et al. (2002), Neumeyer and Dette (2007) and Hušková and Meintanis (2012) considered conditional symmetry tests in the context of a nonparametric location-scale model. However, these tests are still inconsistent in directions where the scaled innovations are not independent of the covariates, which is likely in a serially dependent data context, e.g. conditional heteroskewness and heterokurtosis are expected when dealing with financial data (see, Harvey and Siddique, 1999, 2000 and Brooks et al., 2005).

Delgado and Escanciano (2007) proposed a test of conditional symmetry around a parametric location function in a serial dependence context, which is valid when the conditional location/scale assumption fails. The null hypothesis is rejected when the parametric center of symmetry is misspecified, even when the conditional distribution is symmetric with respect to some other center. Hydman and Yao (2002) introduced conditional density estimators under absolutely regular (ARE) serial dependence, which were applied to testing symmetry of the conditional density evaluated at a given conditioning point. We consider a test of the hypothesis that the conditional distribution is symmetric a.s. in a pure nonparametric context, which does not need to estimate the underlying conditional distribution, just the underlying nonparametric regression, and is also valid in a general nonparametric set up.

The test is based on the joint empirical characteristic function of nonparametric residuals and the explanatory variables and only needs to estimate the nonparametric regression using kernels. The empirical characteristic function has been used for testing symmetry of the innovations marginal distribution by Csörgo and Heathcote (1987), Henze et al. (2003) and Hušková and Meintanis (2012), amongst others. These type of tests are easier to justify under fairly weak regularity conditions compared with those based on comparing empirical distribution functions. Our test statistic is the integrated squared difference between the restricted and unrestricted estimators of the joint characteristic function of nonparametric residuals and explanatory variables with respect to a weighting function. Unlike tests based on smooth estimators of the conditional distribution, like the pointwise test of Hydman and Yao (2002), our test is sensitive in the direction of local alternatives converging to the null at the parametric rate $T^{-1/2}$, with T the sample size. The critical values can be estimated with the assistance of a bootstrap technique.

The rest of the article is organized as follows. In next section we present the testing procedure and the basic asymptotic results assuming that the regression function is known. These results are used in section 3 for deriving the asymptotic distribution of the test statistic when the nonparametric regression is estimated. Critical values are estimated using a bootstrap procedure. The finite sample performance of the test is studied in section 4 by means of Monte Carlo experiments. In section 5 we report the results of two applications of the test using real data to study whether losses are more likely than gains in stock markets and whether expansions and contractions are equally likely in business cycles. Section 6 is devoted to conclusions and final remarks. Mathematical proofs of the main results and discussion of regularity conditions are deferred to a mathematical appendix at the end of the article.

2 The testing procedure

Consider a $\mathbb{R}^{1+d}-$valued strictly stationary and ergodic time series process $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which satisfies the Markov's property

$$\mathbb{P}\left(Y_t \leq y\right| \mathcal{A}_{t-1}) = \mathbb{P}\left(Y_t \leq y\right| X_t) \text{ for all } y \in \mathbb{R} \ a.s., \tag{1}$$

where $\mathcal{A}_t := \sigma\left(\{Y_s, Z_{s+1}\}_{s=-\infty}^{t}\right)$, $\sigma(\cdot)$ means smallest sigma algebra and

$$X_t = (Y_{t-1}, Z_t, Y_{t-2}, Z_{t-1}, ..., Y_{t-k}, Z_{t-k+1})^{\tau} = (X_{1t}, ..., X_{pt})^{\tau}, \text{ with } p = k(d+1), \tag{2}$$

and "τ" denotes transpose. That is, the only relevant information for explaining Y_t are the first k lags of (Y_t, Z_{t+1}).

We propose a nonparametric test for the hypothesis that the conditional distribution of Y_t is symmetric about the (nonparametric) regression function, *i.e.*

$$H_0 : \mathbb{P}\left(U_t \leq u\right| X_t) = \mathbb{P}\left(-U_t \leq u\right| X_t) \text{ for all } u \in \mathbb{R} \ a.s., \tag{3}$$

where $U_t := Y_t - r(X_t)$ are innovations and $r(X_t) := \mathbb{E}\left(Y_t\right| X_t)$ is the regression function. The alternative hypothesis, H_1, consists of all non-possible events under H_0.

Remark 1 *A necessary, but non-sufficient, condition for* H_0, *is that the marginal distrib-*

ution of U_t is symmetric about zero, i.e. that $\mathbb{P}(U_t \le u) = \mathbb{P}(-U_t \le u)$. The conditional distribution of U_t is conditionally symmetric about zero iff $\mathbb{E}(\sin(U_t u)|\, X_t) = 0$ for all $u \in \mathbb{R}$ a.s., i.e. iff the conditional characteristic function is real, while the marginal distribution function of U_t is symmetric about zero iff $\mathbb{E}(\sin(U_t u)) = 0$ for all $u \in \mathbb{R}$, i.e. iff the marginal characteristic function is real. The conditional expectation can be different than zero, while the marginal expectation is zero. There are many consistent tests for marginal symmetry that are inconsistent for H_0 in infinitely many directions where U_t and X_t are not independent, as it was already pointed out in the introduction. Tests designed for testing symmetry of the marginal distribution of $S_t := U_t/\sqrt{\mathbb{E}(\, U_t^2|\, X_t)}$, in the context of non-parametric location/scale models, are also inconsistent for H_0 in infinitely many directions where the assumption of a conditional location/scale model is not satisfied, e.g. when higher conditional moments of S_t given X_t are not constant.

Since (3) is satisfied *iff* the conditional characteristic function of U_t given X_t is real, H_0 can be equivalently expressed as,

$$H_0 : \ \mathbb{E}(\sin(U_t u)|\, X_t) = 0 \text{ for all } u \in \mathbb{R} \ a.s.$$

Therefore, applying the integrated conditional moment (ICM) approach (see e.g. Bierens 1982, Bierens and Ploberger 1997 and references therein), H_0 and H_1 can be alternatively expressed in terms of the function

$$\begin{aligned} J(u, x) : &= 2i \cdot \mathbb{E}\left[\mathbb{E}(\sin(U_t u)|\, X_t) \exp(i X_t^\tau x)\right] \\ &= 2i \cdot \mathbb{E}\left[\sin(U_t u) \exp(i X_t^\tau x)\right] \\ &= \mathbb{E}\left[\exp(i(X_t^\tau x + U_t u)) - \exp(i(X_t^\tau x - U_t u))\right], \end{aligned}$$

with $i^2 = -1$, which is the difference between the joint characteristic functions of (U_t, X_t) and $(-U_t, X_t)$. That is, H_0 is satisfied *iff* $|J(v)| = 0$ for all $v \in \mathbb{R}^{1+p}$, where $|a + ib| = \sqrt{a^2 + b^2}$. Then, an ICM test is based on an estimator of J using nonparametric residuals. However, as it happens in most semiparametric problems involving integrals of nonparametric regression functions, or other conditional expectations, random denominators involve serious technical difficulties that are circumvented using different strategies. In testing problems like ours, a reasonable alternative consists of ridding off the random denominator problem by weighting by the density estimates evaluated at the data points (e.g. Delgado and González-Manteiga 2000.) We follow this approach. Assuming that X_t admits a bounded Lebesgue density f_X,

H_0 can be alternatively expressed in terms of the weighted version of $J(u,x)$,

$$\bar{J}(u,x):\ = 2i\cdot\mathbb{E}\left[\sin(U_t u)\exp\left(iX_t^\tau x\right)f_t^2\right],$$

with $f_t := f_X\left(X_t\right)$, i.e., we can express H_0 and H_1 as,

$$H_0:\left|\bar{J}(v)\right| = 0 \text{ for all } v\in\mathbb{R}^{1+p} \text{ vs } H_1:\left|\bar{J}(v)\right|\neq 0 \text{ for some } v\in\mathbb{R}^{1+p}. \tag{4}$$

Consider a Lebesgue integrable even function $W:\mathbb{R}^{1+p}\to\mathbb{R}^+$ such that

$$\left\{v\in\mathbb{R}^{1+p}:W(v)=0\right\} \text{ has Lebesgue measure } 0. \tag{5}$$

Then, H_0 and H_1 can be expressed as a significance test on the parameter,

$$\begin{aligned}
\eta(W):\ &= \int\left|\bar{J}(v)\right|^2 W(v)dv\\
&= 4\int\int\left|\mathbb{E}\left[\sin(U_t u)\left[\cos\left(X_t^\tau x\right)+i\sin\left(X_t^\tau x\right)\right]f_t^2\right]\right|^2 W(u,x)dudx\\
&= 4\int\int\left(\mathbb{E}^2\left[\sin(U_t u)\cos\left(X_t^\tau x\right)f_t^2\right]+\mathbb{E}^2\left[\sin(U_t u)\sin\left(X_t^\tau x\right)f_t^2\right]\right)W(u,x)dudx\\
&= \int\alpha^\tau(v)\alpha(v)W(v)dv,
\end{aligned}$$

where $\alpha(v) := \mathbb{E}\left[V_t(v)\right]$, $v^\tau = (u,x^\tau)$ and $V_t(v) := 2\sin(U_t u)\cdot f_t^2\cdot\theta_t(x)$, with $\theta_t(x)=\theta(x,X_t)$ and $\theta(x,\bar{x}) := \left(\sin(\bar{x}^\tau x),\cos(\bar{x}^\tau x)\right)^\tau$. Henceforth, an unspecified integral denotes integration over the whole space. That is, (4) can be alternatively expressed as

$$H_0:\eta(W)=0 \text{ vs } H_1:\eta(W)\neq 0.$$

Assume by the moment that $r(X_t)$ and the corresponding errors U_t are observable. Given a time series of length T, $\left\{(Y_t,Z_t)\right\}_{t=1}^T$, the sample analogue of $\eta(W)$ is

$$\eta_T(W) := \int\alpha_T^\tau(v)\alpha_T(v)W(v)dv,$$

with

$$\alpha_T(v) := \frac{1}{\tilde{T}}\sum_{t=k+1}^T V_t(v),$$

and $\tilde{T}=T-k$. If the regression function were known, we would consider the test $\Psi_{W,T}(c) = 1_{\{T\eta_T(W)>c\}}$, where c is the critical value.

Note $\mathbb{E}\left(V_t(v)|\,\mathcal{A}_{t-1}\right)=0$ $a.s.$ for each $v\in\mathbb{R}^{1+p}$, i.e. $\{V_t\left(v\right)\}_{t\in\mathbb{Z}}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{A}_t\}$. A convenient setting for studying the asymptotic distribution of $\sqrt{T}\alpha_T$ under H_0 and, hence, the distribution of $T\eta_T(W)$, is the infinite dimensional Hilbert space $\mathcal{L}^2_W$ of measurable $\mathbb{R}^2$-valued functions on $\mathbb{R}^{1+p}$ that are square integrable with respect to the measure $W(v)dv$. The norm in $\mathcal{L}^2_W$ will be denoted by

$$\|g\|_{\mathcal{L}^2_W}:=\left(\int g^{\tau}(v)g(v)W(v)dv\right)^{1/2}. \tag{6}$$

Since $\sqrt{T}\alpha_T$ is a random element of $\mathcal{L}^2_W$, the asymptotic distribution of $\sqrt{T}\alpha_T$ under H_0 is obtained by applying a central limit theorem (CLT) for real valued vectors of martingale differences taking values in the Hilbert space $\mathcal{L}^2_W$ (e.g. Walk 1977, Jakubowski 1980, Métivier and Nakao 1987, Xie 1995 or, more recently, Kundu, Majumdar and Mukherjee 2000.) Let α_∞ be a centered Gaussian process in $\mathcal{L}^2_W$ with covariance kernel

$$\Omega\left(v_1,v_2\right):=\mathbb{E}\left[\alpha_\infty(v_1)\alpha^{\tau}_\infty(v_2)\right]=\mathbb{E}\left[V_t(v_1)V^{\tau}_t(v_2)\right],\ v_1,v_2\in\mathbb{R}^{1+p}.$$

The law of large numbers (LLN) and CLT needed to justify the test $\Psi_{W,T}(c)$ require less restrictive assumptions than those assumed by Delgado and Escanciano (2007) to justify tests based on generic functionals of $\bar{J}$ estimators suitably scaled, e.g. Kolmogorov-Smirnov type test. Next proposition establishes the properties of the asymptotic power function $\beta_W(c)=\lim_{T\to\infty}\mathbb{E}\left[\Psi_{W,T}(c)\right]$ under H_0, H_1 and local alternatives of the form

$$H_{1T}:\mathbb{E}\left(\sin\left(U_tu\right)|\,X_t\right)=\frac{\gamma_t(u)}{\sqrt{T}}\text{ a.s.,}$$

where γ_t is a random element of $\mathcal{L}^2_W$ and $\delta(u,x)^{\tau}:=2\cdot\mathbb{E}\left[\gamma_t(u)f^2_t\theta_t(x)\right]\neq0$ on a set of positive Lebesgue measure. Notice that, if the conditional distribution of U_t given X_t admits a Lebesgue density, we can equivalently express H_{1T} as,

$$H_{1T}:f(u|\,X_t)=f_0(u|\,X_t)\left[1+\frac{h_t(u)}{\sqrt{T}}\right]\text{ a.s. for all }u\in\mathbb{R},$$

where $f_0(\cdot|\,X_t)$ is positive, symmetric, and integrates to one a.s., and h_t is a random element of $\mathcal{L}^2_W$ with $h_t(u)/\sqrt{T}\geq-1$ and $\int f_0(u|X_t)h_t(u)du=0$ a.s. Then $\gamma_t(u)=\int\sin(\bar{u}u)f_0(\bar{u}|\,X_t)h_t(\bar{u})d\bar{u}$. Define, $\Phi_W\left(\bullet\right)=\mathbb{P}\left(\|\alpha_\infty\|^2_{\mathcal{L}^2_W}>\bullet\right)$.

Proposition 1 *Assume that $\{(Y_t,Z_t)\}_{t\in\mathbb{Z}}$ is a strictly stationary ergodic process satisfying (1), such that X_t in (2) admits a bounded Lebesgue density f_X, and positive weights W satisfy*

(5). Then, for any $c>0$, under H_1

$$\beta_W(c) = 1, \tag{7}$$

under H_0,

$$\beta_W(c) = \Phi_W(c), \tag{8}$$

and under H_{1T},

$$\beta_W(c) = \mathbb{P}\left(\|\alpha_\infty + \delta\|^2_{\mathcal{L}^2_W} > c\right) \geq \Phi_W(c). \tag{9}$$

Therefore, the test $\Psi_{W,T}(c_\alpha)$ with critical value c_α, such that $\Phi_W(c_\alpha) = \alpha$ is valid for testing H_0 in the direction of H_1, and is sensitive to local alternatives H_{1T}. The asymptotic critical values are not pivotal and they must be estimated, e.g. using the bootstrap method introduced in next section.

A convenient weight function, introduced by Fan (1998) in the context of goodness-of-fit testing of a probability density function (pdf), is $\bar{W}_h(v) := (2\pi)^{-1-p}\bar{K}(hv)^2$, where $\bar{K}(v) = \int \exp(i\bar{v}'v)K(\bar{v})d\bar{v}$ is the Fourier's transform of $K(v) := \prod_{j=1}^{1+p} k(v_j)$ with $v = (v_1, ..., v_{p+1})^\tau$ and k a univariate pdf. With this choice of weights, applying Plancherel's theorem after a change of variable (see Fan 1998, Lemma 2.1),

$$\begin{aligned}
\eta\left(\bar{W}_h\right) &= \int \bar{J}^2(v)\frac{1}{(2\pi)^{1+p}}\left[\int \exp(ihv^\tau\bar{v})K(\bar{v})d\bar{v}\right]^2 dv \\
&= \int\left[\int f_X^2(\bar{x})K_h(u-\bar{u}, x-\bar{x})(F-G)(d\bar{u}, d\bar{x})\right]^2 dudx \\
&= \int\left(\bar{f}_h - \bar{g}_h\right)^2(v)dv,
\end{aligned}$$

where $F(u,x) := \mathbb{P}(U_t \leq u, X_t \leq x)$ and $G(u,x) := \mathbb{P}(-U_t \leq u, X_t \leq x)$ are the joint c.d.f.'s of (U_t, X_t) and $(-U_t, X_t)$, respectively, $\bar{f}_h$ and $\bar{g}_h$ are the densities of the convolutions $K_h * \bar{F}$ and $K_h * \bar{G}$, respectively, where $\bar{F}(u,x) := \int_{\{\bar{u}\leq u, \bar{x}\leq x\}} f_X^2(\bar{x})F(d\bar{u}, d\bar{x})$ and $\bar{G}(u,x) := \int_{\{\bar{u}\leq u, \bar{x}\leq x\}} f_X^2(\bar{x})G(d\bar{u}, d\bar{x})$. That is, the parameter $\eta\left(\bar{W}_h\right)$ is the squared Lebesgue integral of the difference between the densities of convolutions $K_h * \bar{F}$ and $K_h * \bar{G}$, with $K_h(v) := h^{-1-p}K(v/h)$. When F (resp. G) admits a Lebesgue density f (resp. g), under suitable conditions, $\lim_{h\to 0}\bar{f}_h(u,x) = f_X^2(x)f(u,x)$ and $\lim_{h\to 0}\bar{g}_h(u,x) = f_X^2(x)g(u,x)$ by Bochner's theorem (see e.g. Prakasa Rao 1983, Theorem 2.1.1).

The estimator of $\eta(\bar{W}_h)$ can be expressed as

$$\eta_T(\bar{W}_h) = \int\left(\bar{f}_{hT} - \bar{g}_{hT}\right)^2(v)dv,$$

where $\bar{f}_{hT}$ and $\bar{g}_{hT}$ are the densities of the convolutions $K_h * \bar{F}_T$ and $K_h * \bar{G}_T$, respectively, with $\bar{F}_T(u,x) = \tilde{T}^{-1}\sum_t f_t^2 1_{\{U_t \le u,\ X_t \le x\}}$ and $\bar{G}_T(u,x) = \tilde{T}^{-1}\sum_t f_t^2 1_{\{-U_t \le u,\ X_t \le x\}}$ the sample versions of $\bar{F}$ and $\bar{G}$, respectively. Henceforth, summations run from $k+1$ to T. That is, $\bar{f}_{hT}$ and $\bar{g}_{hT}$ are Rosenblatt-Parzen's kernel estimators of the densities $\bar{f}$ and $\bar{g}$, respectively, using a kernel K and a bandwidth h. It can be also justified that $\bar{f}_{hT}$ is a consistent estimator of $\bar{f}$ under suitable regularity conditions, with h converging to zero as T diverges to infinity at a convenient rate (e.g. Prakasa Rao 1983, Theorem 2.1.2). Reasoning as Bickel and Rosenblatt (1973) in the context of specification testing of a pdf, we can also justify that the distribution of $\eta_T(\bar{W}_h)$, suitably centered and scaled, can be approximated by a standard normal assuming that h converges to zero at a suitable rate and smoothness conditions for the density of F, f, and the regression function r. See Fan (1994) for discussion. In our proposal the weights $\bar{W}_h$ take h fixed.

This test is not feasible, because the functions r and f_X are unknown in practice under H_0 and the critical values are not pivotal. A feasible test is introduced in the next section.

3 Implementation and asymptotic justification of the test

A natural feasible version of $\eta_T(W)$ consists of substituting the innovations U_t by nonparametric residuals, once $r(X_t)$ is estimated. Assume that $F_X(x) = F(\infty, x)$ admits a Lebesgue density function f_X, and consider the kernel estimator of $r(x)$,

$$\hat{r}_T(x) := \frac{1}{\hat{f}_{XT}(x)\tilde{T}a_T^p}\sum\nolimits_t Y_t K_X\left(\frac{X_t - x}{a_T}\right),$$

where

$$\hat{f}_{XT}(x) := \frac{1}{\tilde{T}a_T^p}\sum\nolimits_t K_X\left(\frac{X_t - x}{a_T}\right),$$

estimates $f_X(x)$, $K_X(x) := \int_{\mathbb{R}} K(u,x)du = \prod_{j=1}^p k(x_j)$ with $x = (x_1, \ldots, x_p)^\tau$, and $\{a_T\}_{T\ge 1}$ is a sequence of positive bandwidth numbers. The sample analog of $\alpha(u,x)$ is

$$\hat{\alpha}_T(u,x) := \frac{2}{\tilde{T}}\sum_t \sin\left(u\hat{U}_t\right)\cdot \hat{f}_t^2 \cdot \theta_t(x).$$

where $\hat{U}_t := Y_t - \hat{r}_t$ with $\hat{r}_t := \hat{r}_T(X_t)$, and $\hat{f}_t := \hat{f}_{XT}(X_t)$. The feasible version of $\eta(W)$ is

$$\hat{\eta}_T(W) := \|\hat{\alpha}_T\|^2_{\mathcal{L}^2_W},$$

and the test is $\hat{\Psi}_{W,T}(c) = 1_{\{T\hat{\eta}_T(W) > c\}}$.

For each $v \in \mathbb{R}^{1+p}$ fixed, $\hat{\alpha}_T(v)$ is a type of $V-$ *statistics* that typically appears when making inferences on semiparametric models. The following proposition provides a first order asymptotic representation of $\hat{\eta}_T(W)$, which is crucial for deriving the asymptotic distribution of $T\hat{\eta}_T(W)$ and, hence, to justify the test. The proof of this proposition in the appendix uses a Hoeffding decomposition argument for $\hat{\alpha}_T$, as in Robinson (1988, 1989) for different statistics in other semiparametric problems. The asymptotic equivalence is proved assuming ARE serial dependence, that the functions r and f_X are smooth enough, and that Y_t admits enough moments. Our approach is based on results in Robinson (1989) in the context of testing restrictions on index models using ARE time series. The assumptions discussed in the appendix involve a relation between the rate of convergence of the bandwidth a_T in terms of T, the parameter governing the degree of dependence of the ARE time series, the number of moments assumed on Y_t, the number of derivatives imposed on r and f_X, the dimension of X_t, and the order of the higher order kernels used. These technical regularity conditions are presented in the mathematical appendix. Define

$$\tilde{\eta}_T(W) := \|\tilde{\alpha}_T\|^2_{\mathcal{L}^2_W},$$

where

$$\tilde{\alpha}_T(v) := \frac{1}{\tilde{T}} \sum\nolimits_t \tilde{V}_t(v),$$

with

$$\tilde{V}_t(v) := 2\left[\sin(uU_t) - uU_t\phi_t(u)\right] f_t^2\theta_t(x),$$

$\phi_t(u) := \phi(u, X_t)$ and $\phi(u,x) := \mathbb{E}\left[\cos(uU_t)|\, X_t = x\right]$. Henceforth, the notation "$\xrightarrow{d}$" means convergence in distribution of random elements taking values in $\mathcal{L}^2_W$, and also of random variables or vectors, $O_{\mathbb{P}}(1)$ stands for a sequence of random variables bounded in probability, $o_{\mathbb{P}}(1)$ for a sequence of random variables converging to zero in probability, and "$\stackrel{d}{=}$" equality in distribution.

Proposition 2 *Under assumptions in Proposition 1 and assumptions A.1-A.3 in the appendix,*

$$\hat{\eta}_T(W) = \tilde{\eta}_T(W) + o_{\mathbb{P}}\left(\frac{1}{T}\right).$$

Let $\tilde{\alpha}_\infty$ be a Gaussian element of $\mathcal{L}^2_W$ with zero mean and covariance function

$$\Sigma(v_1, v_2) := \mathbb{E}\left[\tilde{\alpha}_\infty(v_1)\tilde{\alpha}^\tau_\infty(v_2)\right] = \mathbb{E}\left[\tilde{V}_t(v_1) \cdot \tilde{V}^\tau_t(v_2)\right], \ v_1, v_2 \in \mathbb{R}^{1+p},$$

and $\tilde{\Phi}_W(c) = \mathbb{P}\left\{\|\tilde{\alpha}_\infty\|^2_{\mathcal{L}^2_W} > c\right\}$. Consider the asymptotic power function $\hat{\beta}_W(c) =$

$\lim_{T\to\infty}\mathbb{E}\left[\hat{\Psi}_{W,T}(c)\right]$.

Corollary 1 *Under the assumptions in Proposition 2, for any $c > 0$, under H_1*

$$\hat{\beta}_W(c) = 1, \tag{10}$$

under H_0,

$$\hat{\beta}_W(c) = \tilde{\Phi}_W(c), \tag{11}$$

and under H_{1T},

$$\hat{\beta}_W(c) = \mathbb{P}\left(\|\tilde{\alpha}_\infty + \delta\|^2_{\mathcal{L}^2_W} > c\right) \geq \tilde{\Phi}_W(c). \tag{12}$$

Since $\tilde{\Phi}_W$ depends on nuisance parameters, the test is implemented using a bootstrap technique.

Consider, the Rademacher's sequence $\{\zeta_t\}_{t=1+k}^T$ of *i.i.d.* r.v.'s with $\mathbb{P}_\zeta(\zeta_t = -1) = \mathbb{P}_\zeta(\zeta_t = 1) = 1/2$, where $\mathbb{P}_\zeta$ is the probability measure of the binary random variables $\{\zeta_t\}_{t=1+k}^T$. Define the $\alpha-level$ critical value $c_\alpha(W) : \mathbb{P}\left(\|\tilde{\alpha}_\infty\|^2_{\mathcal{L}^2_W} > c_\alpha(W)\right) = \alpha$. The bootstrap estimate of $c_\alpha(W)$ is based on the estimated asymptotic expansion in Proposition 2 using permutted residuals $\zeta_t\hat{U}_t$, i.e.,

$$\tilde{\alpha}^*_T(u,x) := \frac{2}{\tilde{T}}\sum_t \left[\sin(u\zeta_t\hat{U}_t) - u\zeta_t\hat{U}_t\hat{\phi}_t(u)\right]\hat{f}_t^2\theta_t(x),$$

where

$$\hat{\phi}_t(u) := \frac{1}{a_T^p\hat{f}_t\tilde{T}}\sum_\ell \cos\left(u\hat{U}_\ell\right)K_X\left(\frac{X_t - X_\ell}{a_T}\right)1_{\{|\hat{f}_t|>b_T\}},$$

and $\{b_T\}_{T\geq 1}$ is a sequence of positive trimming numbers converging to zero at a rate related to a_T and T, as indicated in $A.4$ in the appendix. The trimming is a technical device, which is introduced to prove the consistency of the bootstrap test, which does not have a practical effect, as shown in the simulations. The bootstrap critical values

$$\tilde{c}^*_{\alpha T}(W) := \inf\left\{c : \mathbb{P}_\zeta\left(T\,\|\tilde{\alpha}^*_T\|^2_{\mathcal{L}^2_W} \leq c\right) \geq 1-\alpha\right\},$$

forms a basis for the $\alpha-level$ test $\hat{\Psi}_{W,T}(\tilde{c}^*_{\alpha T}(W))$. The test can also be based on the $p-value$ estimate,

$$p-value^* := \mathbb{P}_\zeta\left(\|\tilde{\alpha}^*_T\|^2_{\mathcal{L}^2_W} > \hat{\eta}_T(W)\right).$$

Next proposition establishes the validity of the bootstrap test.

Proposition 3 *Under regularity conditions in Proposition 2 and assumption A.4 in the* *appendix, under* H_1,

$$\tilde{c}^{*}_{\alpha T}(W) = O_{\mathbb{P}}(1), \tag{13}$$

and

$$p - value^{*} \xrightarrow{p} 0, \tag{14}$$

under H_0,

$$\tilde{c}^{*}_{\alpha T}(W) = c_{\alpha}(W) + o_{\mathbb{P}}(1), \tag{15}$$

and

$$p - value^{*} \xrightarrow{d} U(0,1). \tag{16}$$

The critical values are approximated by Monte Carlo as accurately as desired. Then, the bootstrap critical values and $p - value$ are approximated by $\tilde{c}^{*}_{\alpha T,B}(W)$ and $p - value^{*}_{T,B}$, respectively, computed as follows,

1. Generate B independent Rademacher's sequences $\left\{\zeta_t^{(b)}\right\}_{t=1}^{T}$, $b = 1, ..., B$ with B large.
2. Compute
$$\tilde{\alpha}_T^{*(b)}(u,x) := \frac{2}{\tilde{T}} \sum_t \left[\sin(u\zeta_t^{(b)}\hat{U}_t) - u\zeta_t^{(b)}\hat{U}_t\hat{\phi}_t(u)\right] \hat{f}_t^2\theta_t(x), \ b = 1, ..., B.$$
3. Compute
$$\tilde{c}^{*}_{\alpha T,B}(W) = \inf\left\{c : \left(\frac{1}{B}\sum_{b=1}^{B} T\left\|\tilde{\alpha}_T^{*(b)}\right\|^2_{\mathcal{L}^2_W} \leq c\right) \geq 1-\alpha\right\},$$
and
$$p - value^{*}_{T,B} = \frac{1}{B}\sum_{b=1}^{B} 1_{\left\{\left\|\tilde{\alpha}_T^{*(b)}\right\|^2_{\mathcal{L}^2_W} > \hat{\eta}_T(W)\right\}}.$$
4. Reject H_0 at the $\alpha - level$ of significance when $\hat{\eta}_T(W) > \tilde{c}^{*}_{\alpha T,B}(W)$ or $p - value^{*}_{T,B} < \alpha$.

4 Finite sample performance

In this section we investigate the finite sample performance of the proposed test. The set up of our Monte Carlo studies is as follows. All simulations in this section are based on 2,000 replications of each experiment using 500 bootstrap resamples. We deliver simulation results for a 5% significance level and sample sizes of $T = 50$, 100 and 200 are considered. Nonparametric regression estimates are computed using the fourth order kernel $K_X(x) = 0.5(3 - x^2)\phi(x)$ with $\phi(x)$ the standard normal pdf, and bandwidth $a_T = c * s_X * T^{-1/3}$, where s_X is the sample standard deviation of X_t. We report results for $c = 0.5$, 1.0, and 1.5. In each bootstrap replication, the trimming parameter is chosen to be $b_T = 0.05 * T^{-1/12.5}$, which satisfies our assumptions. Results are fairly insensitive to the choice of b_T.

First, consider a weight function $\bar{W}_h^N(u,x) = \exp(-h(u^2+x^2))$, which corresponds to the standard normal characteristic function, with h fixed. Henceforth, $\hat{\eta}_T(h) := \hat{\eta}_T(\bar{W}_h^N)$, which can be expressed as

$$\hat{\eta}_T(h) = \frac{\pi^{1/2}}{2\tilde{T}^2 h^{1/2}} \sum\nolimits_t \sum\nolimits_s \left[e^{-\frac{(\hat{U}_t - \hat{U}_s)^2}{4h}} - e^{-\frac{(\hat{U}_t + \hat{U}_s)^2}{4h}} \right] e^{-\frac{(X_t - X_s)^2}{4h}} \hat{f}_t^2 \hat{f}_s^2, \tag{17}$$

where $\hat{U}_t$'s are nonparametric residuals, and $\hat{f}_t$'s are kernel density estimates. Note that alternative h provides alternative weight functions. We report $h = 0.5$ and $h = 1$, though we have run simulations with other h values as well as $\bar{W}_h$ corresponding to the Cauchy characteristic function. Results are similar with alternative weight functions.

Applying Proposition 4.1 in Henze, Klar and Meintanis (2003) adapted to our context, we obtain the computationally convenient expression for $\hat{\eta}_T^*(h) = \hat{\eta}_T^*(\bar{W}_h^N)$,

$$\begin{aligned}\hat{\eta}_T^*(h) = \frac{\pi^{1/2}}{2\tilde{T}^2 h^{1/2}} \sum\nolimits_t \sum\nolimits_s \Bigg\{ \Bigg[2 + \frac{\bar{U}_t^* \bar{U}_s^*}{2h} - \left(1 + \frac{(U_t^* - U_s^*)\, \bar{U}_s^*}{h} + \frac{(U_t^* - U_s^*)^2\, \bar{U}_t^* \bar{U}_s^*}{4h^2} \right) \Bigg] e^{-\frac{(U_t^* - U_s^*)^2}{4h}} \\ + \left[\frac{\bar{U}_t^* \bar{U}_s^*}{2h} - \left(1 + \frac{(U_t^* + U_s^*)\, \bar{U}_s^*}{h} + \frac{(U_t^* + U_s^*)^2\, \bar{U}_t^* \bar{U}_s^*}{4h^2} \right) \right] e^{-\frac{(U_t^* + U_s^*)^2}{4h}} \Bigg\} e^{-\frac{(X_t - X_s)^2}{4h}} \hat{f}_t^2 \hat{f}_s^2,\end{aligned}$$

where $U_t^* = \xi_t \hat{U}_t$ are permuted nonparametric residuals, and

$$\bar{U}_t^* = \frac{1}{\tilde{T} a_T^p} \sum\nolimits_s K_X \left(\frac{X_t - X_s}{a_T} \right) \frac{U_s^*}{\hat{f}_s} 1_{\{|\hat{f}_s| > b_T\}}.$$

The purpose of the first set of simulations is to study the finite sample performance of our test under different time series designs. We consider the following nonlinear autoregressive model of order 1, NLAR(1),

$$Y_t = 0.5 Y_{t-1} \exp(-0.5 Y_{t-1}^2) + \varepsilon_t,$$

where the errors ε_t are generated as follows:

(AU1) $\varepsilon_t \sim i.i.d.\ N(0,1)$.

(AU2) $\varepsilon_t \sim i.i.d.\ t_5$.

(AU3) $\varepsilon_t \sim i.i.d.\ e_1 1_{\{Z \leq 0.5\}} + e_2 1_{\{Z > 0.5\}}$, 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.

(AU4) $\varepsilon_t = \sigma_t e_t$, $\sigma_t^2 = \phi_0 + \phi_1 \sigma_{t-1}^2 + \phi_2 \varepsilon_{t-1}^2$, $e_t \sim i.i.d.\ N(0,1)$ with $\phi_0 = 2$, $\phi_1 = 0.5$, and $\phi_2 = 0.3$.

(AU5) Same as (AU4) except that $\phi_1 = 0.9$ and $\phi_2 = 0.05$.

(AU6) $\varepsilon_t = \lambda_t e_t$ and $e_t \sim t_{v_t}$, where

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

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

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2,$$

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

with $\alpha_0 = 2$, $\alpha_1 = 0.9$, $\alpha_2 = 0.05$, $\beta_0 = 3.5$, $\beta_1 = 0.2$, $\beta_2 = 0.5$.

(AU7) $\varepsilon_t \sim i.i.d.\ \exp(N(0,1))$.

(AU8) $\varepsilon_t \sim i.i.d.\ \chi_2^2$.

(AU9) $\varepsilon_t \sim i.i.d.\ -\ln(U(0,1))$.

(AU10) $\varepsilon_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.1$, $\lambda_4 = -0.18$.

(AU11) $\varepsilon_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.001$, $\lambda_4 = -0.13$.

(AU12) $\varepsilon_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0001$, $\lambda_4 = -0.17$.

Here, F_λ, the generalized lambda distribution, is defined in terms of the inverse of the cumulative distribution

$$F_\lambda^{-1}(u) = \lambda_1 + \frac{\left[u^{\lambda_3} - (1-u)^{\lambda_4}\right]}{\lambda_2}, \quad 0 < u < 1,$$

and the λ values here and below are taken from Table 1 of Randles et al. (1980), indicating different degrees of asymmetry and kurtosis. For instance, the λ values in (AU10) define a distribution with the associated skewness and kurtosis coefficients 2.0 and 21.2, respectively.

The symmetric distributions (AU1)-(AU5) and asymmetric distributions (AU7)-(AU12) are considered by Delgado and Escanciano (2007). The error terms in (AU4) and (AU5) follow a generalized autoregressive conditional heteroskedastic [GARCH (1,1)] model of Bollerslev (1986). 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. In (AU6), we consider a model for generalized autoregressive conditional heteroskedasticity and heterokurtosis proposed by Brooks et al. (2005), with time-varying degrees of freedom, and the conditional variance and conditional kurtosis are permitted to evolve separately. See also the autoregressive conditional skewness model developed in Harvey and Siddique (1999, 2000) where, instead of conditional kurtosis, the conditional skewness is allowed to vary over time. Notice that (AU1)-(AU6) fall under H_0, whereas (AU7)-(AU12) fall under H_1.

For the first experiment, we compute $\hat{\eta}_T(h)$ with $X_t = Y_{t-1}$. In designs (AU1)-(AU12), we generate $T + 300$ observations and then discard the first 300 to minimize the effect of initial values. We investigate the sensitivity of the test to the choice of weight functions $\bar{W}_h^N$ by considering two values of h (i.e. 0.5 and 1). Henceforth, $\hat{\eta}_T^{(j)}(h)$ for $j = 1$, 2, and 3 in the subsequent tables, corresponds to the $\hat{\eta}_T(h)$ test with c taking values 0.5, 1.0, and 1.5, respectively.

TABLE 1 & 2 ABOUT HERE

Tables 1 and 2 report the percentage of rejections for designs (AU1)-(AU12) for $\hat{\eta}_T(0.5)$ and $\hat{\eta}_T(1)$, respectively. The tests show a very good size accuracy for moderate sample sizes and reasonable powers. For (AU5) and (AU6), $\hat{\eta}_T(0.5)$ is slightly undersized for small sample sizes, which is corrected as sample size increases. On the other hand, both tests tend to be oversized for (AU2) for $T = 50$. Among the alternatives considered, (AU10) is harder to detect as it is fairly close to symmetry.

We also consider the NLAR(1) design in Hyndman and Yao (2002): $Y_t = 0.23Y_{t-1}(1.6 - Y_{t-1}) + 0.4\varepsilon_t$ with $\varepsilon_t \sim i.i.d.\ N(0,1)$ truncated in the interval $[-12, 12]$, which is in fact a quadratic AR(1) model. Note that Hyndman and Yao (2002)'s test can only be used to test symmetry of the conditional density evaluated at a given conditioning point, while our conditional test is designed to check symmetry uniformly in the conditioning variable. Hyndman and Yao (2002) report the p-values evaluated at a few X_t points using 500 time series observations. For this quadratic AR(1) model, the conditional distribution of Y_t given $X_t = Y_{t-m}$ is symmetric for $m = 1$ but not necessarily so for $m > 1$. This model is non-stationary. When $c = 1$, for sample sizes $T = 50$, 100 and 200, empirical sizes of $\hat{\eta}_T(0.5)$ for $m = 1$ (which falls under H_0) are respectively 0.04, 0.05 and 0.05. The empirical powers for $m = 3$ (which falls under H_1) are respectively 0.05, 0.06 and 0.07. Note that since the design $m = 3$ is very close to symmetry under our set up, the low power is expected, and even when sample size reaches 1000, power only increases to 0.16. Results for $c = 0.5$ and 1.5 are quantitatively similar and are omitted.

The second set of experiments compares the performance of our test with others designed to test marginal symmetry in two situations and in two directions. First, when the marginal and conditional distributions are both symmetric with $i.i.d.$ data. Second, when the marginal distribution is symmetric, but the conditional distribution is asymmetric under serial dependence. To this end, we introduce a marginal version of the test statistic in (17) intended to detect departures from the hypothesis of symmetry of the marginal distribution of U_t around zero, i.e

$$\hat{\eta}_{T,m}(h) = \frac{\pi^{1/2}}{2\tilde{T}^2 h^{1/2}} \sum_t \sum_s \left[e^{-\frac{(\hat{U}_t - \hat{U}_s)^2}{4h}} - e^{-\frac{(\hat{U}_t + \hat{U}_s)^2}{4h}} \right],$$

with critical values also estimated by the permutation procedure.

Specifically, we compare our test with Fang et al. (2015), T_S henceforth. Also, we consider for comparison the marginal test $\hat{\eta}_{T,m}(h)$ using $\hat{U}_t = (Y_t - \overline{Y}_T)/s_Y$ with $\overline{Y}_T$ and s_Y respectively the sample mean and sample standard deviation of Y_t. Tests T_S and $\hat{\eta}_{T,m}(h)$ are designed to detect departures of the null hypothesis of symmetry of the marginal distribution using *i.i.d.* data. Tests based on $\hat{\eta}_{T,m}(h)$ and T_S are inconsistent to test conditional symmetry when the marginal distribution is symmetric.

Table 3 reports the percentage of rejections under the null and alternative hypotheses under designs considered by Delgado and Escanciano (2007) and Fang et al. (2015). We use the test statistics $\hat{\eta}_T(0.5)$ and $\hat{\eta}_{T,m}(0.5)$, where the conditional test $\hat{\eta}_T(0.5)$ uses non-parametric residuals $\hat{U}_t = Y_t - \hat{r}(X_t)$ and $X_t = Y_{t-1}$, and the marginal test $\hat{\eta}_{T,m}(0.5)$ uses standardized data $\hat{U}_t = (Y_t - \overline{Y}_T)/s_Y$. Results for $\hat{\eta}_T(1)$ are similar and, hence, are omitted.

TABLE 3 ABOUT HERE

The designs considered are listed below.

- Symmetric distributions
  (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) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = 0.19754$, $\lambda_3 = \lambda_4 = 0.134915$.
  (S5) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = \lambda_4 = -0.08$.
  (S6) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -0.397912$, $\lambda_3 = \lambda_4 = -0.16$.
  (S7) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = \lambda_4 = -0.24$.

- Asymmetric distributions
  (A1) $Y_t \sim i.i.d.\ \exp(N(0,1))$.

(A2) $Y_t \sim i.i.d.\ \chi_2^2$.
(A3) $Y_t \sim i.i.d.\ -\ln(U(0,1))$.
(A4) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = 1.4$, $\lambda_4 = 0.25$.
(A5) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0075$, $\lambda_4 = -0.03$.
(A6) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.1$, $\lambda_4 = -0.18$.
(A7) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.001$, $\lambda_4 = -0.13$.
(A8) $Y_t \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0001$, $\lambda_4 = -0.17$.
(A9) $Y_t = e_t - e_{t-1}$ with $e_t \sim i.i.d.\ \exp(N(0,1))$.
(A10) $Y_t = e_t - e_{t-1}$ with $e_t \sim i.i.d.\ \chi_2^2$.
(A11) $Y_t = e_t - e_{t-1}$ with $e_t \sim i.i.d.\ -\ln(U(0,1))$.

In designs (S1)-(S7), Y_t's are *i.i.d.* symmetric around its mean with different kurtosis coefficients. In designs (A1)-(A8), Y_t's are *i.i.d.* and asymmetric around the mean with a wide range of kurtosis. In (A9)-(A11), Y_t's are identically distributed and not serially independent, but they are symmetric about its mean. However, in these cases, the conditional distribution of Y_t given Y_{t-1} is not symmetric. We observe in Table 3, that in designs (S1)-(S5), under the null hypothesis of marginal symmetry and *i.i.d.* data, all the tests have good size accuracy, even for the smallest sample size. In designs (S6) and (S7), $\hat{\eta}_{T,m}(0.5)$ and $\hat{\eta}_T(0.5)$ are slightly oversized. The three tests have reasonable power in most designs (A1)-(A8) under the alternative hypothesis of marginal asymmetry. Within this group, the alternative (A6) seems the most difficult to detect for all tests. However, our test $\hat{\eta}_T(0.5)$ behaves similarly than tests designed to detect marginal symmetry under these circumstances. It is worth noticing that our test is not designed to test marginal symmetry in the direction of marginal asymmetry, though it is valid to do so under designs (A1)-(A8). In the designs (A9)-(A11), under marginal symmetry, tests based on T_S and $\hat{\eta}_{T,m}(0.5)$ have very poor size, which is explained because they neglect the serial dependence structure. As expected, these tests have trivial power for detecting conditional asymmetry. However, since the conditional distribution of Y_t given Y_{t-1} is asymmetric in these cases, the test based on $\hat{\eta}_T(0.5)$ has reasonable power, which increases with the sample sizes.

The last set of simulations compares our tests with others designed to test symmetry of the marginal distribution of the scaled error term of a conditional location/scale model in the direction of nonparametric alternatives. Our test is compared with Hušková and Meintanis (2012)'s test, henceforth HM, under the model

$$Y_t = r(X_t) + U_t, \text{ with } U_t = \sigma(X_t) \cdot \varepsilon_t, \ t = 1, 2, ...,$$

with $\{X_t\}_{t\geq 1}$ *i.i.d.* $U(0,1)$ independent of $\{\varepsilon_t\}_{t\geq 1}$ *i.i.d.* according to (LS1)-(LS6) listed below.

(LS1) $\varepsilon \sim i.i.d.\ N(0,1)$.

(LS2) $\varepsilon \sim i.i.d.\ t_5$.

(LS3) $\varepsilon \sim i.i.d.\ e_1 1_{\{Z \leq 0.5\}} + e_2 1_{\{Z > 0.5\}}$, 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.

(LS4) $\varepsilon \sim i.i.d.\ \chi^2_2$.

(LS5) $\varepsilon \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.001$, $\lambda_4 = -0.13$.

(LS6) $\varepsilon \sim i.i.d.\ F_\lambda$ with $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -0.0001$, $\lambda_4 = -0.17$.

We consider $r(x) = \sin(2\pi x)$ and $\sigma(x) = |x|$. Notice that the scaled error term ε_t is symmetric around zero in the designs (LS1)-(LS3) and asymmetric in (LS4)-(LS6).

TABLE 4 ABOUT HERE

The HM test statistic is identical to $\hat{\eta}_{T,m}(0.5)$, but the critical values are estimated using a wild bootstrap procedure, rather than the permutation procedure that we suggest. The resulting test is denoted by $T^{(2)}_{0.5}$. The conditional test statistic $\hat{\eta}_T(h)$ is computed with the explanatory variable X_t, and nonparametric residuals $\hat{U}_t = Y_t - \hat{r}(X_t)$. Both statistics $\hat{\eta}_{T,m}(0.5)$ and $T^{(2)}_{0.5}$ are computed using the scaled nonparametric residuals $\hat{\varepsilon}_t = \hat{U}_t/\hat{\sigma}(X_t)$ with $\hat{\sigma}^2(X_t)$ the kernel estimator of $\sigma^2(X_t)$. The percentage of rejections for the three tests under designs (LS1)-(LS6) is reported in Table 4, in which Panels A, B and C use $c = 0.5$, 1.0, and 1.5, respectively, and Panel D $a_T = 0.075$ as used by HM. All three tests have good size accuracy. However, $\hat{\eta}_T(0.5)$ has much larger power than $\hat{\eta}_{T,m}(0.5)$ and $T^{(2)}_{0.5}$ under the designs (LS4)-(LS6) considered, while $T^{(2)}_{0.5}$ has the least power. In addition, each test behaves similarly under different choice of a_T's and is fairly insensitive to it.

Finally, we examine the power performance of our test under certain local alternatives. Let $Y_t = \sin(2\pi X_t) + U_t$, where $U_t = |X_t|\varepsilon_t$ with $\varepsilon_t = e_t - \mathbb{E}(e_t)$ and $X_t \sim i.i.d.\ U(0,1)$ independent of e_t. We consider e_t following a normal mixture distribution with the density given by

$$f_e(x) = \left(1 - \frac{2}{\sqrt{T}}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} + \frac{2}{\sqrt{T}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-1)^2}{2}}.$$

FIGURE 1 ABOUT HERE

Notice that when sample size T increases, the distribution of ε_t converges to a standard normal. We consider sample sizes running from 50 to 800 with an increment of 50. We observe from Figure 1 that our proposed test $\hat{\eta}_T(0.5)$ has non-trivial power against the local alternatives converging to the null at a parametric rate for $c = 0.5$, 1.0 and 1.5, with $c = 1.0$ appearing to give the most powerful test under this specific local alternative.

5 Applications in Stock Markets and Business Cycles

In this section, we revisit two problems of practical interests, i.e. the asymmetric behaviour for stock returns and business cycles. We apply our test to investigate whether losses are more likely than gains in stock markets, and whether expansions and contractions are equally likely in business cycles, given the relevant information.

We first consider four stock indices, namely, S&P 500, FTSE 100, Nikkei 225 and Shanghai A-Share (SSE-A). They represent various degrees of maturity and regulating conditions of the underlying stock markets. The four index series are collected using daily data from 1 January 2001 to 31 December 2015, with 3773, 3787, 3694, and 3786 observations, respectively. Returns series are obtained through $X_t = \log(P_t/P_{t-1})$, with P_t denoting the time series sequence for each stock index. Routine augmented Dickey-Fuller tests for the four indices indicate that there exist unit roots in all four index series but not in their returns series. Moreover, all returns series exhibit the well-recorded stylized facts of volatility clustering and high kurtosis indicating the existence of fat tails.

In our empirical application, we focus on the scenario where only the first lagged value will predict the stock returns. This feature is consistent with the stylized fact that today's financial markets are often more influenced by the most recent events. Specifically, we consider the following nonlinear autoregressive process of order 1, NLAR (1),

$$X_t = r(X_{t-1}) + U_t.$$

The hypothesis of interest is that X_t is symmetric around the (unknown) nonparametric regression function $r(X_{t-1})$ given the most recent information X_{t-1}.

To study the effect of 2008 financial crisis on the behavior of the financial markets, we split the sample period into two parts, (1) before the crisis: 1 January 2001 to 31 December 2006, and (2) after the crisis: 1 January 2007 to 31 December 2015. The (marginal) skewness coefficients for the four returns series before the crisis (resp. after the crisis) are respectively 0.16 (-0.32), -0.16 (-0.16), -0.08 (-0.52) and 0.64 (-0.51). Though these marginal coefficients seem to indicate a potential change of behavior for the returns series after the crisis, especially for Nikkei 225 and SSE-A, they are not justified theoretically to test the (unconditional) symmetry hypothesis. In addition, they cannot be used to study the conditional symmetry of the returns before and after the crisis.

Next, we consider testing the conditional expansion and contraction behaviour of business cycles in four major economies corresponding to USA, UK, Japan, and China. The data we used are in terms of the real measures of gross domestic products (GDPs) for each country, which are taken from the OECD's quarterly national accounts (QNA) data set. Specifically, the GDPs for USA, UK, Japan are seasonally adjusted quarterly series starting from the first quarter of 1960 until the second quarter of 2016, with a total number of observations 226, while China's quarterly GDP is only available from the first quarter of 1992 until the second quarter of 2016 with 98 observations. The four series of quarterly growth rates of real GDP (i.e. change over the previous quarter) are calculated. Like the stock returns, a NLAR(1) model is examined. The hypothesis of interest in this case is whether the errors U_t are symmetrically distributed given the previous period information of GDP growth rate, i.e. whether expansions and contractions are equally likely.

TABLE 5 ABOUT HERE

We report results with $\hat{\eta}_T(0.5)$ with $c = 0.5$, 1.0 and 1.5, results for $h = 1$ are similar. Table 5 reports the bootstrapped p-values for the four stock returns series S&P 500, FTSE 100, Nikkei 225 and (SSE-A) before the crisis (resp. after the crisis) as well as those for the four GDP quarterly growth rates in USA, UK, Japan and China. All p-values are based on 500 bootstrap draws. For the four returns series we do not reject the hypothesis of conditional symmetry at 5% significance level before the crisis. On the other hand, results in parentheses of columns 2-5 support the conditional asymmetry for the four stock markets after the crisis. Our findings for the conditional symmetry hypothesis thus confirm that there exists a distinct behavior for the four stock markets before and after the crisis. As to the business cycles, for all three c's, the bootstrapped p-values indicate asymmetric behavior for UK's macro-economy in contrast to Japan and China, for which we do not reject the hypothesis of symmetry at 5% significance level. Lastly, USA's growth appears to be symmetric, though the p-values are not as big as for Japan or China.

6 Conclusions

This article has proposed a test for symmetry of the conditional distribution about a non-parametric regression that can be implemented using stationary time series data under fairly weak regularity conditions. The methodological approach exploits the fact that the conditional distribution is symmetric *iff* the corresponding characteristic function is real, which suggests a Cramér-von Mises type test statistic based on the integrated joint characteristic function of regression errors and explanatory variables with respect to suitable weights. The test is consistent in the direction of alternatives where existing tests have trivial power, which includes situations under the alternative where the marginal distribution of the regression errors, or their conditionally scaled versions, is symmetric about zero.

The test is proven to enjoy good size accuracy and power properties using small samples, and is fairly insensitive to the smoothing parameter choice needed for estimating the nonparametric regression function. We have applied the test to study conditional symmetry of several stock indices returns, given the first lag, before and after 2008 crisis, and countries' GDP.

The testing methodology presented in this article consists of characterizing restrictions on the conditional distribution by means of the joint characteristic function, rather than the joint distribution function itself, can be applied to test other restrictions. For instance, the hypothesis of conditional independence between Y_t and $X_t^{(2)}$ given $X_t^{(1)}$, with $X_t = \left(X_t^{(1)\tau}, X_t^{(2)\tau}\right)^\tau$, can be formally stated as

$$H_0 : \mathbb{P}\left(Y_t \le y | \, X_t^{(1)}, X_t^{(2)}\right) = \mathbb{P}\left(Y_t \le y | \, X_t^{(1)}\right) \; a.s. \text{ for all } y \in \mathbb{R}.$$

This hypothesis can be equivalently expressed as

$$H_0 : |M(v)| = 0 \text{ for all } v \in \mathbb{R}^{1+p},$$

where $M(v) = \mathbb{E}\left\{U_t^{(1)}(y) \exp\left(iX_t^\tau x\right) f_{X^{(1)}t}\right\}$, $f_{X^{(1)}t} = f_{X^{(1)}}(X_t^{(1)})$ with $f_{X^{(1)}}$ the density of $X_t^{(1)}$ and $U_t^{(1)}(y) = \exp\left(iY_t y\right) - \mathbb{E}\left(\exp\left(iY_t y\right) | \, X_t^{(1)}\right)$. Then, given a suitable estimators of $U_t^{(1)}(y)$ and $f_{X^{(1)}t}$, $\hat{U}_t^{(1)}(y)$ and $\hat{f}_{X^{(1)}t}$ say, the test statistic for the omnibus test of H_0 would be

$$\hat{\eta}_T^{(1)}(W) = \int_{\mathbb{R}^{1+p}} \left|\hat{M}_T(v)\right|^2 W(v) dv,$$

where

$$\hat{M}_T(v) = \frac{1}{T} \sum_t \hat{U}_t^{(1)}(y) \exp\left(iX_t^\tau x\right) \hat{f}_{X^{(1)}t}.$$

This test is an alternative to Delgado and González-Manteiga's (2000) proposal.

7 Appendix

The proof of Proposition 1 applies Fatou's lemma and the ergodic theorem for consistency. Convergence in distribution of the test statistics under the null is proved applying the CLT for martingale differences taking values in separable Hilbert spaces. To this end we apply Theorem 1.4 of Kundu, Majumdar and Mukherjee (2000), which is reproduced below as a Lemma. Henceforth, for any $m-$dimensional random element $G(0,\Upsilon)$, of a separable infinite dimensional Hilbert space $\mathcal{H}$ with mean zero and covariance operator Υ, given $a,b\in\mathcal{H}$, we use the notation $\langle \Upsilon a,b\rangle = \mathbb{E}\left[\langle G(0,\Upsilon),a\rangle \langle G(0,\Upsilon),b\rangle\right].$

Lemma 1: *Let $\mathcal{H}$ stand for a real separable infinite dimensional Hilbert space endowed with an inner product $\langle\cdot,\cdot\rangle$, and corresponding norm $\|a\|_{\mathcal{H}}=\sqrt{\langle a,a\rangle}$. Let $\{\xi_{Tt}\}$ be a $\mathcal{H}-$ valued martingale difference array with respect to the filtration $\{\mathcal{F}_{Tt}\}$ such that $\mathbb{E}\|\xi_{Tt}\|^2_{\mathcal{H}}<\infty$ for each $1\le t\le T$, $T\ge 1$. Let $\{e_\ell:\ell\ge 1\}$ be an orthonormal basis of $\mathcal{H}$. Assume that the following conditions hold.*

- **(i)** *For every $b\in\mathcal{H}$, $\sum_{t=1}^{T}\mathbb{E}\left(\langle\xi_{Tt},b\rangle^2\middle|\mathcal{F}_{Tt-1}\right)=\sigma_b^2+o_{\mathbb{P}}(1)$ for some σ_b.*
- **(ii)** $\lim_{T\to\infty}\sum_{l=1}^{\infty}\sum_{t=1}^{T}\mathbb{E}\left(\langle\xi_{Tt},e_\ell\rangle^2\right)=\sum_{\ell=1}^{\infty}\sigma^2_{e_\ell}<\infty.$
- **(iii)** *$L_T(\varepsilon,e_\ell)=o_{\mathbb{P}}(1)$ for every $\varepsilon>0$ and every $\ell\ge 1$, where for $b\in\mathcal{H}$,*

$$L_T(\varepsilon,b)=\sum_{t=1}^{T}\mathbb{E}\left(\langle\xi_{Tt},b\rangle^2 1_{\{|\langle\xi_{Tt},b\rangle|>\varepsilon\}}\middle|\mathcal{F}_{T,t-1}\right).$$

Then, $\left\{\sum_{t=1}^{T}\xi_{Tt}\right\}$ converges in distribution to $G(0,\Upsilon)$, a centered Gaussian process in $\mathcal{H}$ with covariance operator Υ, which is characterized by $\langle\Upsilon b,b\rangle=\sigma_b^2$, for all $b\in\mathcal{H}$.

Proof of Proposition 1. By Fatou's lemma and the ergodic theorem, under H_1,

$$\liminf_{T\to\infty}\eta_T(W)\ge\int\left(\liminf_{T\to\infty}\alpha_T(v)\right)^{\tau}\left(\liminf_{T\to\infty}\alpha_T(v)\right)W(v)dv=\|\mathbb{E}V_t\|^2_{\mathcal{L}^2_W}>0\ a.s.,$$

which proves (7). Taking into account that, under H_0, $\{V_t(u)\}_{t\in\mathbb{Z}}$ is a bounded sequence of stationary ergodic martingale differences taking values on the real separable infinite dimensional Hilbert space of squared integrable functions with respect to the measure $W(v)dv$, $v\in\mathbb{R}^{1+p}$. The inner product in $\mathcal{L}^2_W$ is denoted by $\langle a,b\rangle=\int a(v)^{\tau}b(v)W(v)dv$, with corresponding norm $\|\cdot\|_{\mathcal{L}^2_W}$ in (6). Then, in order to prove (8), take $\xi_{Tt}=V_t/\sqrt{T}$ and $\mathcal{F}_{Tt}=\mathcal{A}_t$

and check conditions in Lemma 1 with $\sigma_b^2 = \mathbb{E}\left\langle \alpha_\infty, b\right\rangle^2 = \left\langle \Omega b, b\right\rangle$. (i) For every $b \in \mathcal{L}_W^2$,

$$\sum\nolimits_t \mathbb{E}\left(\left\langle V_t/\sqrt{T}, b\right\rangle^2 \middle| \mathcal{A}_{t-1}\right) = \frac{1}{\tilde{T}}\sum\nolimits_t \mathbb{E}\left(\left\langle V_t, b\right\rangle^2 \middle| X_t\right) = \sigma_b^2 + o(1)\ a.s.,$$

by the ergodic theorem, where $\sigma_b^2 = \mathbb{E}\left\langle V_t, b\right\rangle^2 = \mathbb{E}\left\langle \alpha_\infty, b\right\rangle^2$. (ii) Let $\{e_\ell\}$ be an orthonormal basis of $\mathcal{L}_W^2$, then,

$$\lim_{T\to\infty}\sum_{\ell=1}^{\infty}\sum\nolimits_t \mathbb{E}\left\langle V_t/\sqrt{T}, e_\ell\right\rangle^2 = \sum_{\ell=1}^{\infty}\mathbb{E}\left\langle V_t, e_\ell\right\rangle^2 = \sum_{\ell=1}^{\infty}\sigma_{e_\ell}^2 \leq \left\|\sum_{\ell=1}^{\infty} e_\ell\right\|_{\mathcal{L}_W^2}^2 < \infty.$$

(iii) Since $\left\langle V_t, b\right\rangle^2 < \infty\ \ a.s.$ uniformly in $t \in \mathbb{Z}$ for $b \in \mathcal{L}_W^2$, for each $\varepsilon > 0$ and all $\mu \geq 0$,

$$L_T\left(\varepsilon, e_\ell\right) \leq \frac{1}{\varepsilon^\mu}\sum\nolimits_t \mathbb{E}\left[\left\langle V_t/\sqrt{T}, e_\ell\right\rangle^{2+\mu} 1_{\left\{\left\langle V_t/\sqrt{T}, e_\ell\right\rangle > \varepsilon\right\}}\middle| X_t\right] \leq \frac{C}{T^{\frac{\mu}{2}}}\ a.s.,$$

for any $C < \infty$, where, henceforth, C denotes a generic bounded positive constant. Therefore, $\alpha_T \to_d \alpha_\infty$ as a random element of $\mathcal{L}_W^2$, and (8) follows applying the continuous mapping theorem (CMT). Finally, (9) is proved by writing, under H_{1T},

$$\left(\sqrt{T}\alpha_T - \delta_T\right)(u, x) = \frac{2}{\sqrt{T}}\sum\nolimits_t \left[\sin\left(U_t u\right) - \mathbb{E}\left(\sin\left(U_t u\right)\middle| X_t\right)\right] f_t^2 \theta_t(x),$$

where $\delta_T(u, x) = 2\tilde{T}^{-1}\sum_t \gamma_t(u) f_t^2 \theta_t(x)$. Then, using the fact that $\left\|\delta_T - \delta\right\|_{\mathcal{L}_W^2}^2 = o(1)\ a.s.$, by the ergodic theorem. Then, applying (ii), under H_{1T}, $\sqrt{T}\alpha_T \overset{d}{\to} \alpha_\infty + \delta$. Therefore, the proof of (9) is completed applying the CMT, i.e.

$$\lim_{T\to\infty}\mathbb{P}\left[T\left\|\alpha_T\right\|_{\mathcal{L}_W^2}^2 \geq c\right] = \mathbb{P}\left[\left\|\alpha_\infty + \delta\right\|_{\mathcal{L}_W^2}^2 \geq c\right] \geq \mathbb{P}\left[\left\|\alpha_\infty\right\|_{\mathcal{L}_W^2}^2 \geq c\right].$$

■

In order to justify the test $\hat{\Psi}_{W,T}(c)$ we need to impose new restrictions on the underlying DGP. Notice that, with v fixed, $\hat{\alpha}_T(v)$ is a standard $V-$ *statistic* involving kernels under ARE serial dependence. This type of $V-$ *statistics* appears in many inference procedures on semiparametric models. In order to prove Propositions 2, we follow Robinson (1989)'s approach, who provided a set of flexible sufficient regularity conditions to justify asymptotics on statistics similar to $T\hat{\eta}_T(W)$ in the context of testing restrictions on semiparametric index models with ARE time series data. These conditions involve using higher order kernels to make compatible the rate of convergence of bias and variance of integrated kernel estimators in high dimensions. The order of the kernel is related to the rate of convergence of the bandwidth, the smoothness conditions on the underlying nonparametric components, moments of Y_t, and the parameter of the ARE time series governing the severity of the serial dependence.

First, we introduce some definitions. We use a multiplicative kernel $K_X(x_1,...,x_p) = \prod_{j=1}^{p} k(x_j)$, where k belongs to the class of higher order kernels, according to the definition below.

Definition 1 *$\mathcal{K}_\ell$, $\ell \geq 1$, is the class of even functions $k : \mathbb{R} \to \mathbb{R}$ satisfying*

$$\begin{aligned} \int_{\mathbb{R}} x^i k(x)dx &= 1_{\{i=0\}},\ i = 0,...,\ell-1 \\ k(u) &= O\left(\left(1+|u|^{\ell+\epsilon+1}\right)^{-1}\right),\ \textit{some } \epsilon > 0. \end{aligned}$$

Definition 2 *Let $\{\xi_t\}_{t\in\mathbb{Z}}$ be a stationary process defined on $(\Omega, \mathcal{G}, \mathbb{P})$; $\{\xi_t\}_{t\in\mathbb{Z}}$ is ARE if*

$$\beta(j) = \mathbb{E}\left\{\sup_{A\in\mathcal{G}_j^\infty}\left|\mathbb{P}\left(A|\,\mathcal{G}_{-\infty}^0\right) - \mathbb{P}\left(A\right)\right|\right\} \to 0 \textit{ as } j \to \infty,$$

with $\mathcal{G}_a^b = \sigma\left(\{\xi_t\}_{t=a}^b\right) \subseteq \mathcal{G}$.

We have adapted conditions in Robinson (1989) to our context as follows.

A.1. a. $\mathbb{E}\,|Y_t|^\mu < \infty$ for some $\mu > 2$.

b. X_t admits a Lebesgue density f_X uniformly bounded, which is at least $N/2$ times boundedly differentiable.

c. The regression function r is $N/2$ times differentiable with derivatives $r^{(j)}$ such that $\mathbb{E}\left|r^{(j)}(X_j)\right|^\mu < \infty$, $j = 1,...,N/2$. The remainder term in the Taylor expansion to order $N/2$ of $r\,(x+\vartheta)$ in a neighbourhood of $\vartheta = 0$ is bounded by $|\vartheta|^{\frac{N}{2}+1}$ times a function $r'(x)$ a.s. such that $\mathbb{E}\,|r'\,(X_t)|^\mu < \infty$.

d. $\phi(u, X_t)$ is $N/2$ times differentiable for each $u \in \mathbb{R}$ a.s., with derivatives $\phi^{(j)}(u,\cdot)$ such that $\mathbb{E}\left\|\phi_t^{(j)}\right\|_{\mathcal{L}_W^2}^\mu < \infty$, with $\phi_t^{(j)}(u) := \phi^{(j)}(u, X_t)$, $j = 1,...,N/2$. The remainder term in the Taylor expansion of $\phi\,(\cdot, x+\vartheta)$ in a neighbourhood of $\vartheta = 0$ is bounded by $|\vartheta|^{\frac{N}{2}+1}$ times a function $\phi'(\cdot, x)$ such that $\mathbb{E}\left\|\phi_t'\right\|_{\mathcal{L}_W^2}^\mu < \infty$, with $\phi_t'(u) := \phi'(u, X_t)$.

e. The kernel function k belongs to $\mathcal{K}_N$, $N = 2, 3, ...$

f. $\{(Y_t, Z_t)\}_{t\in\mathbb{Z}}$ is ARE with coefficients

$$\beta(j) = O(j^{-v}), \text{ for some } v > 1 + \frac{2}{(\mu-2)}. \tag{18}$$

A.2. **a.** $Ta_T^{p\left(1+\frac{2}{\mu}+\frac{2}{v}\right)+\epsilon} \to \infty$ for some $\epsilon > 0$.

b. $Ta_T^{2p\left(1+\frac{2}{\mu}\right)+\epsilon} \to \infty$ for some $\epsilon > 0$.

c. $Ta_T^N \to 0$.

Conditions in A.1. are identical to the smoothness and moment conditions assumed by Robinson (1989), which in turn can be found in Robinson (1988) in a different semiparametric problem. Smoothness conditions on $r(\cdot)$ and $\phi(u,\cdot)$ are related to the order of the higher order kernel, N, related to the rate of convergence of the bandwidth in A.2. Condition A.1.f, which is also assumed in Robinson (1989), establishes the rate of convergence of the ARE parameter, which is related to the moment restriction A.1.a, and the bandwidth convergence rate in A.2. The order of the higher order kernel in A.1.e. and the rate of convergence of $\beta(j)$ in A.1.f. are related to the rate of convergence of the bandwidth a_T in A.2. Restrictions on the rate of convergence of the bandwidth A.2. are similar in Robinson (1989). In fact, A.2.a and A.2.c. are identical to Robinson (1989) assumptions (6.8) and (6.10), but Robinson (1989) assumption (6.9), $Ta_T^{2p} \to \infty$, is weaker than A.2.b. The weakest version of a_T in A.2.a and A.2.b is

$$Ta_T^{2p+\epsilon} \to \infty \text{ for some } \epsilon > 0 \tag{19}$$

when $\mu = v = \infty$. In the *i.i.d.* case, it is only required (19) and A.2.c. Notice that $Ta_T^{2p(1+2/\min(\mu,v))+\epsilon} \to \infty$ for some $\epsilon > 0$ suffices for A.2.a and A.2.b. To reconcile all the restrictions in $A.2.$, we need that $N > p(1+2/\mu+2/v)$ and $N > 2p(1+2/\mu)$.

We need restriction (5) on W for consistency, and also a technical condition on its tail behaviour, which is stated below.

A.3. $W : \mathbb{R}^{1+p} \to \mathbb{R}^+$ is an integrable even function that satisfies (5) and $\int u^4 W(u,x)dudx < \infty$.

Henceforth, we use the following notation, $Z_t = (Y_t, X_t)$, $a = a_T$, $K_{ts} = K_X((X_t - X_s)/a)$, and $s_t(u) = \cos(uU_t)$. The proof of Proposition 2 consists of applying Robinson (1988, 1989) results to this problem.

Proof of Proposition 2. Introduce the empirical process

$$\bar{\alpha}_T(u,x) := 2\frac{u}{\tilde{T}}\sum\nolimits_t \hat{f}_t^2\left(\hat{U}_t - U_t\right)s_t(u)\theta_t(x).$$

Since $\hat{\alpha}_T = \tilde{\alpha}_T + (\hat{\alpha}_T - \bar{\alpha}_T - \alpha_T) + (\alpha_T + \bar{\alpha}_T - \tilde{\alpha}_T)$, it suffices to show that

$$\|\hat{\alpha}_T - \bar{\alpha}_T - \tilde{\alpha}_T\|_{\mathcal{L}^2_W} = o_{\mathbb{P}}\left(T^{-1/2}\right), \quad (20)$$

$$\|\alpha_T + \bar{\alpha}_T - \tilde{\alpha}_T\|_{\mathcal{L}^2_W} = o_{\mathbb{P}}\left(T^{-1/2}\right). \quad (21)$$

Applying a mean value theorem (MVT) argument, uniformly in $u \in \mathbb{R}$,

$$\sin\left(u\hat{U}_t\right) = \sin\left(uU_t\right) + u\left(\hat{U}_t - U_t\right)s_t(u) + \varepsilon_{Tt}(u) \ a.s.$$

where $|\varepsilon_{Tt}(u)| \leq u^2\left|\hat{U}_t - U_t\right|^2$ a.s., and

$$\|\hat{\alpha}_T - \bar{\alpha}_T - \tilde{\alpha}_T\|_{\mathcal{L}^2_W} \leq \left(\int u^4 W(u,x)dudx\right)^{1/2} \frac{2}{T}\sum\nolimits_t \left|\hat{U}_t - U_t\right|^2 \hat{f}_t^2.$$

Hence, after applying A.3., we prove (20) by showing that,

$$\frac{1}{T}\sum\nolimits_t \left|\hat{U}_t - U_t\right|^2 \hat{f}_t^2 = o_{\mathbb{P}}\left(T^{-1/2}\right), \quad (22)$$

which is proved as follows. Write $\hat{U}_t - U_t = \tilde{T}^{-1}n_{t\ell}$, where $n_{t\ell} := n\left(Z_t, Z_\ell\right)$, with

$$n(z_1, z_2) = \frac{1}{a^p}\left(r\left(x_1\right) - r\left(x_2\right) - u_2\right)K_X\left(\frac{x_1 - x_2}{a}\right),$$

$z_i = \left(x_i, u_i\right)$, $i = 1, 2$, and $\tilde{n}_t = \tilde{n}(Z_t)$ with $\tilde{n}(z) = \mathbb{E}\left[n(z, Z_t)\right]$. Since,

$$\frac{1}{T}\sum\nolimits_t \left|\hat{U}_t - U_t\right|^2 \hat{f}_t^2 \leq \frac{2}{T^3}\sum\nolimits_t \left|\sum\nolimits_\ell \left(n_{t\ell} - \tilde{n}_t\right)\right|^2 + \frac{2}{T}\sum\nolimits_t \tilde{n}_t^2,$$

(22) follows from

$$\sum\nolimits_t \left|\sum\nolimits_{\ell \neq t}\left(n_{t\ell} - \tilde{n}_t\right)\right|^2 = o_{\mathbb{P}}\left(T^{5/2}\right), \quad (23)$$

$$\sum\nolimits_t n_{tt}^2 = o_{\mathbb{P}}\left(T^{5/2}\right), \quad (24)$$

$$\sum\nolimits_t \tilde{n}_t^2 = o_{\mathbb{P}}\left(T^{1/2}\right). \quad (25)$$

Applying Robinson (1989)'s Lemma (p.529) to $\mathbb{E}\left[\left(n_{ts} - \tilde{n}_t\right)\left(n_{tr} - \tilde{n}_t\right)\right]$ for $s, r \neq t$, $s < r$, treating separately the cases $t < s < r$, $s < t < r$ and $s < r < t$, as in Robinson (1989)

(7.17), for some $\delta > 0$,

$$\max_t \mathbb{E}\left|\sum\nolimits_{\ell}(n_{t\ell}-\tilde{n}_t)\right|^2 \leq CTs_{\delta}^{2/(2+\delta)}\sum\nolimits_{j=1}^{\infty}\beta^{\delta/(2+\delta)}(j), \tag{26}$$

with $s_\delta := \max_{t\neq\ell}\mathbb{E}\left|n_{t\ell}\right|^{2+\delta}$. By A.1.a, we can choose δ to satisfying $\delta \leq \mu - 2$. Then, by Hölder's inequality,

$$s_\delta = O\left(\max_{t\neq\ell}\left\{\left(\mathbb{E}\left|(r_t-r_\ell-U_\ell)\right|^\mu\right)^{\frac{2+\delta}{\mu}}\cdot\left(\mathbb{E}\left|K_{t\ell}\right|^{\frac{(2+\delta)\mu}{\mu-2-\delta}}\right)^{1-\frac{2+\delta}{\mu}}\right\}\right) = O\left(a^{-p(1+\delta+(2+\delta)/\mu)}\right), \tag{27}$$

since,

$$\max_{t\neq\ell}\mathbb{E}\left|K_{t\ell}\right|^\alpha = O\left(a^{p(1-\alpha)}\right) \text{ for } \alpha>0, \tag{28}$$

by Robinson (1989) (7.10). So,

$$s_\delta^{2/(2+\delta)} = O\left(a^{-p(1+2/\mu)-\epsilon/2}\right),$$

with $\epsilon = 2p\delta/(2+\delta)$ arbitrarily close to zero as δ is. Therefore,

$$\max_t \mathbb{E}\left|\sum\nolimits_{\ell\neq t}(n_{t\ell}-\tilde{n}_t)\right|^2 = O\left(T^{3/2}\right)\cdot O\left(\left(T^{1/2}a^{p(1+2/\mu)+\epsilon/2}\right)^{-1}\right) = o\left(T^{3/2}\right) \tag{29}$$

by A.2.b, which proves (23) using (26). To prove (24), notice that by the ergodic theorem,

$$\sum\nolimits_t n_{tt}^2 = \frac{1}{a^{2p}}K_X^2(0)\sum\nolimits_t U_t^2 = O_{\mathbb{P}}\left(Ta^{-2p}\right) = o_{\mathbb{P}}\left(T^2\right).$$

Now, by the smoothing assumptions in A.1, applying Robinson (1988)'s Lemma 5, $\max_t \mathbb{E}\left|\tilde{n}_t\right|^2 = O\left(a^N\right) = o\left(T^{-1}\right)$, which proves (25).

In order to prove (21), first write,

$$\begin{aligned}
(\bar{\alpha}_T+\alpha_T-\tilde{\alpha}_T)(u,x) &= \frac{2u}{\tilde{T}}\sum\nolimits_t\left[\hat{f}_t^2\left(\hat{U}_t-U_t\right)s_t(u)+f_t^2U_t\phi_t(u)\right]\theta_t(x)\\
&= \frac{2u}{\tilde{T}}\sum\nolimits_t\left(\hat{U}_t-U_t\right)\hat{f}_t\left(\hat{f}_t-f_t\right)s_t(u)\,\theta_t(x)\\
&\quad+\frac{2u}{\tilde{T}}\sum\nolimits_t\left[\left(\hat{U}_t-U_t\right)\hat{f}_tf_ts_t(u)+f_t^2U_t\phi_t(u)\right]\theta_t(x).\\
&= E_{1T}(u,x)+E_{2T}(u,x)
\end{aligned}$$

Applying Cauchy-Swartz's inequality and A.3.,

$$\mathbb{E}\left\|E_{1T}\right\|_{\mathcal{L}_W^2} \leq C\cdot\left(\frac{1}{T}\sum\nolimits_t\left|\hat{U}_t-U_t\right|^2\hat{f}_t^2\right)^{1/2}\left(\frac{1}{T}\sum\nolimits_t\left|\hat{f}_t-f_t\right|^2\right)^{1/2}=o_{\mathbb{P}}\left(T^{-1/2}\right)$$

by (22) and, defining $\tilde{K}_t := \tilde{K}(X_t)$ with $\tilde{K}(x) = a^{-p}\mathbb{E}\left(K_X(X_t-x)/a\right),$

$$\begin{aligned}\frac{1}{\tilde{T}}\sum\nolimits_t\left|\hat{f}_t-f_t\right|^2 &\leq \frac{2}{T^3}\sum\nolimits_t\left|\sum\nolimits_{\ell\neq t}\left(K_{t\ell}-\tilde{K}_t\right)+\frac{K_X(0)}{a^p}\right|^2+\frac{2}{T}\sum\nolimits_t\left|\tilde{K}_t-f_t\right|^2 \\ &= o_{\mathbb{P}}\left(T^{-1/2}\right)+o_{\mathbb{P}}\left(T^{-1}\right),\end{aligned} \tag{30}$$

since, using same arguments as in the proof of (22), $\max_t\mathbb{E}\left|\sum_{\ell\neq t}\left(K_{t\ell}-\tilde{K}_t\right)\right|^2 = o\left(T^{3/2}\right)$, and applying Robinson (1988)'s Lemma 4, $\max_t\mathbb{E}\left|\tilde{K}_t-f_t\right|^2 = O\left(a^N\right) = o\left(T^{-1}\right)$. Finally, we write E_{T2} in terms of the symmetric kernel $c(v,v_1,v_2) = d(v,v_1,v_2)+d(v,v_2,v_1)$ with $v=(u,x)$, $v_i=(u_i,x_i)$, $i=1,2$, and

$$d(v,v_1,v_2) = uf(x_1)\left(r(x_1)-r(x_2)-u_2\right)\cos(u_1u)\frac{1}{a^p}K_X\left(\frac{x_1-x_2}{a}\right)\theta(x,x_1).$$

That is,

$$\frac{1}{2}E_{2T}(v) = \frac{1}{2\tilde{T}^2}\sum\nolimits_t\sum\nolimits_\ell\left(c_{t\ell}-\tilde{c}_t-\tilde{c}_\ell+\tilde{c}\right)(v)+\frac{1}{\tilde{T}}\sum_t\left(\tilde{c}_t+g_t-\tilde{c}\right)(v)+\frac{\tilde{c}(v)}{2},$$

where $c_{t\ell} := c(v,Z_t,Z_\ell,)$, $\tilde{c}_t(v) := \tilde{c}(v,Z_t)$, $\tilde{c}(v,v_1) := \mathbb{E}\left(c(v,Z_t,v_1)\right)$, $\tilde{c}(v) := \mathbb{E}\left(\tilde{c}_t(v)\right)$ and $g_t := g_t(u,x) = uf_t^2U_t\phi_t(u)\theta_t(x)$. Therefore, in order to show that $\left\|E_{T2}\right\|_{\mathcal{L}_W^2} = o_{\mathbb{P}}\left(T^{-1/2}\right)$, it suffices to show that

$$\left\|\sum\sum_{t\neq\ell}\left(c_{t\ell}-\tilde{c}_t-\tilde{c}_\ell+\tilde{c}\right)\right\|_{\mathcal{L}_W^2} = o_{\mathbb{P}}\left(T^{3/2}\right), \tag{31}$$

$$\left\|\sum\nolimits_t c_{tt}\right\|_{\mathcal{L}_W^2} = o_{\mathbb{P}}\left(T^{3/2}\right), \tag{32}$$

$$\left\|\sum\nolimits_t\left(\tilde{c}_t-\tilde{c}+g_t\right)\right\|_{\mathcal{L}_W^2} = o_{\mathbb{P}}\left(T^{1/2}\right) \tag{33}$$

$$\tilde{c} = o_{\mathbb{P}}\left(T^{-1/2}\right). \tag{34}$$

Note that (31) is similar to (7.1) in Robinson (1989). Therefore, applying Denker and Keller (1983)'s Proposition 2 as in Robinson (1989), the mean square of the left-hand side of (31) is bounded by

$$b_{\delta}^{\frac{2}{2+\delta}} O\left(T^{2+\varepsilon}\right), \tag{35}$$

with $b_{\delta} := \max_{t\neq\ell} \mathbb{E}\left\|c_{t\ell}\right\|_{\mathcal{L}_W^2}^{2+\delta}$, if for some $\delta, \varepsilon > 0$, $\beta(j) = O\left(j^{(\varepsilon-2)(2+\delta)/\delta}\right)$. Then, mimicking the proof of (7.1.) in Robinson (1989) and, in view of (18), choose $\varepsilon = 2 - \delta v/(2+\delta)$, so

$$\frac{\delta}{2+\delta} < \frac{2}{v}. \tag{36}$$

Take $\delta \leq \mu - 2$, which is possible because $\mu > 2$. Then, use the fact that $b_{\delta} \leq C \cdot s_{\delta} = O\left(a^{-p(1+\delta+(2+\delta)/\mu)}\right)$ by (27). Therefore, (35)$=O\left(T^3\right) O\left(\left(Ta^{\kappa}\right)^{1-\delta v/(2+\delta)}\right)$, where $\kappa = p\left[(2+\delta)(1+2/\mu)+\delta\right]/(\delta v - 2 - \delta)$. Because of (18) we can choose δ to satisfy not only (36) and $\delta \leq \mu - 2$, but also

$$\frac{\delta}{2+\delta} > \frac{1}{v}. \tag{37}$$

Under (37), the condition $Ta^{\kappa} \to \infty$ implies that (35)$=o\left(T^3\right)$, which implies (31). Pick $\epsilon = \left[\varepsilon/(1-\varepsilon)\right]\left[p\left(1+2/\mu+1/v\right)\right]$, noticing that (37) implies $\varepsilon < 1$, and ε positive and arbitrarily close to 0 implies the same for ϵ. Then $\kappa = p\left(1+2/\mu+2/v\right)+\epsilon$ and, hence, $Ta^{\kappa} \to \infty$ by A.2.a, which proves (31). In order to prove (32), notice that $\mathbb{E}\left\|\sum_t c_{tt}\right\|_{\mathcal{L}_W^2}^2 \leq C \cdot Ta^{-2p}$ and, hence, $\left\|\sum_t c_{tt}\right\|_{\mathcal{L}_W^2} = o_{\mathbb{P}}\left(T^{1/2}a^{-p}\right) = o_{\mathbb{P}}\left(T^{3/2}\right)$ by A.2.b. For the proof of (33) notice that

$$\left(\tilde{c}_t + g_t - \tilde{c}\right)(u,x) = \sum_{i=1}^{3}\left[e_{it}(u,x) - \bar{e}_i\right](u,x),$$

where $e_{it}(v) = e_i(v, Z_t)$ and $\bar{e}_i(v) = \mathbb{E}\left[e_i(v,Z_t)\right]$, $i = 1,2,3$, with

$$\begin{aligned}
e_1(u,x,x_1) &= u\mathbb{E}\left[\left(r_t f_t \phi_t(u)\theta_t(x) - r(x_1)f_X(x_1)\phi(u,x_1)\theta(x,x_1)\right)\frac{1}{a^p}K_X\left(\frac{X_t - x_1}{a}\right)\right],\\
e_2(u,x,x_1) &= u\left(r(x_1) + u_1\right)\mathbb{E}\left[\left(f_X(x_1)\phi(u,x_1)\theta(x,x_1) - f_t\phi_t(u)\theta_t(x)\right)\frac{1}{a^p}K_X\left(\frac{X_t - x_1}{a}\right)\right],\\
e_3(u,x,x_1) &= uf_X(x_1)\cos(u_1 u)\theta(x,x_1)\mathbb{E}\left[\left(r(x_1) - r_t\right)\frac{1}{a^p}K_X\left(\frac{X_t - x_1}{a}\right)\right].
\end{aligned}$$

Now, use the fact that f_X, θ and ϕ are uniformly bounded, and $r(X_t)$ and U_t have at least two moments, to show that, under the smoothing assumptions and applying Robinson (1988) Lemma 5, A.2.c and A.3. imply that $\max_t \mathbb{E}\left\|e_{it}\right\|_{\mathcal{L}_W^2}^2 = O\left(a^N\right) = o\left(T^{-1}\right)$, $i = 1,2,3$, and $\mathbb{E}\left\|\tilde{c}\right\|_{\mathcal{L}_W^2}^2 \leq C \cdot \sum_{i=1}^3 \mathbb{E}\left\|\bar{e}_i\right\|_{\mathcal{L}_W^2}^2 = o\left(T^{-1}\right)$

$$\mathbb{E}\left\|\sum_t\left(\tilde{c}_t + g_t - \bar{c}\right)\right\|_{\mathcal{L}_W^2}^2 \leq C \cdot \sum_{i=1}^3\left[\mathbb{E}\left\|\sum_t e_{it}\right\|_{\mathcal{L}_W^2}^2 + T^2\mathbb{E}\left\|\bar{e}_i\right\|_{\mathcal{L}_W^2}^2\right] \leq C \cdot T^2 O\left(a^N\right) = o(T),$$

which proves (33) and (34). ■

Proof of Corollary 1. By Proposition 2, for each $c \in \mathbb{R}^+$, $\hat{\beta}_W(c) = \tilde{\Phi}_W(c) + o_{\mathbb{P}}(1)$, which proves (11) and (12), and, also (10), taking into account that by Fatou's lemma,

$$\liminf_{T\to\infty} \tilde{\eta}_T(W) \geq \int \left(\liminf_{T\to\infty} \tilde{\alpha}_T(v)\right)^{\tau} \left(\liminf_{T\to\infty} \tilde{\alpha}_T(v)\right) W(v)dv = \mathbb{E}\left\|\tilde{V}_t\right\|^2_{\mathcal{L}^2_W} > 0.$$

■

In order to justify consistency of the bootstrap test, we need the following assumption on the trimming parameter, which is related to the bandwidth parameter a_T.

A.4.

$$\frac{1}{T a_T^{2p(1+2/\mu)} b_T^4} + b_T \to 0.$$

Henceforth, we denote $b = b_T$. Let $\mathbb{E}_\zeta$ denote expectation with respect to the binary random variables $\{\zeta_t\}_{t=1+k}^{T}$, and for any random element of $\mathcal{L}^2_W$, $\{\vartheta^*_T\}_{T\geq 1}$, involving $\{\zeta_t\}_{t=1+k}^{T}$, $\vartheta^*_T = o_{\mathbb{P}^*}(1)$ means that for any $\epsilon > 0$, $\mathbb{P}_\zeta\left\{\|\vartheta^*_T\|^2_{\mathcal{L}^2_W} > \epsilon\right\} = o_{\mathbb{P}}(1)$. Consider the infeasible version of $\tilde{\alpha}^*_T$

$$\tilde{\alpha}^\dagger_T(u,x) := \frac{2}{\tilde{T}} \sum\nolimits_t \left[\sin(u\zeta_t U_t) - u\zeta_t U_t \phi_t(u)\right] f_t^2 \theta_t(x).$$

The proof of proposition 3, consists of three parts. First we show that

$$\left\|\sqrt{T}\left(\tilde{\alpha}^*_T - \tilde{\alpha}^\dagger_T\right)\right\|_{\mathcal{L}^2_W} = o_{\mathbb{P}^*}(1). \tag{38}$$

Then, (13) and (14) are proved using (38) and showing that for almost all sample sequences $(Y_1, Z_1), (Y_2, Z_2), \ldots$, under H_0, H_1 or H_{1T},

$$\limsup_{T\to\infty} \mathbb{E}_\zeta \left\|\sqrt{T}\tilde{\alpha}^\dagger_T\right\|^2_{\mathcal{L}^2_W} < \infty, \tag{39}$$

and (15) and (16) are proved using (39) and showing that for almost all sample sequences $(Y_1, Z_1), (Y_2, Z_2), \ldots$, under H_0 and H_{1T}

$$\lim_{T\to\infty} \mathbb{P}_\zeta\left\{T\left\|\tilde{\alpha}^\dagger_T\right\|^2_{\mathcal{L}^2_W} > c\right\} = \tilde{\Phi}_W(c) \text{ for all } c \in \mathbb{R}^+. \tag{40}$$

The proof of (38) is based on results in Robinson (1988, 1989) and (39) is the result of applying the ergodic theorem. The proof of (40) consists of showing that for almost all sample sequences $(Y_1, Z_1), (Y_2, Z_2), \ldots$, $\sqrt{T}\tilde{\alpha}^\dagger_T \to_d \tilde{\alpha}_\infty$ as a random element of $\mathcal{L}^2_W$ applying Theorem 1.1. in Kundu, Majumdar and Mukherjee (2000) for conditional convergence in distribution of Hilbert space valued martingale difference arrays, which is reproduced below as a Lemma. See Henze, Klar and Meintanis (2003) for application in a related context.

Lemma 2: *Let* $\{e_l\}_{l\geq 1}$ *be an orthonormal basis of the infinite dimensional Hilbert space* $\mathcal{H}$*, with inner product* $\langle\cdot,\cdot\rangle$ *and corresponding norm* $\|a\|_{\mathcal{L}^2_W} = \sqrt{\langle a,a\rangle}$*. Let* $\{\xi_{Tt}\}_{t=1}^T$ *be a finite sequence of independent* $\mathcal{H}-$*valued random elements with zero means and finite second moments, and put* $S_T = \sum_{t=1}^T \xi_{Tt}$*. Let* Υ_T *be the covariance operator of* S_T*. Assume that the following conditions hold.*

(i) $\lim_{T\to\infty}\langle\Upsilon_T e_j, e_\ell\rangle = a_{j\ell}$*, for all* $j\geq 1$ *and* $\ell\geq 1$*.*

(ii) $\lim_{T\to\infty}\sum_{\ell=0}^{\infty}\langle\Upsilon_T e_\ell, e_\ell\rangle = \sum_{\ell=0}^{\infty} a_{\ell\ell} < \infty$*.*

(iii) $\lim_{T\to\infty}\sum_{\ell=0}^{\infty} L_T(\varepsilon, e_\ell) = 0$ *for every* $\varepsilon > 0$ *and every* $\ell\geq 1$*, where for* $b\in\mathcal{H}$*,*

$$L_T(\varepsilon, b) = \sum_{\ell=1}^{T}\mathbb{E}\left(\langle\xi_{Tt}, b\rangle^2 1_{\{|\langle\xi_{Tt},b\rangle|>\varepsilon\}}\right).$$

Then S_T *converges in distribution to a centered Gaussian process* $G(0,\Upsilon)$ *in* $\mathcal{H}$ *with covariance operator* Υ *characterized by* $\langle\Upsilon h, e_\ell\rangle = \sum_{j=1}^{\infty}\langle h, e_j\rangle a_{j\ell}$*, for every* $\ell\geq 1$*.*

Proof of Proposition 3:. In order to prove (38), define,

$$\begin{aligned}
\tilde{\phi}_t &:= \frac{1}{\hat{f}_t\tilde{T}}\sum\nolimits_{\ell}\cos(uU_\ell)\,K_{t\ell}\hat{I}_\ell,\\
\bar{\phi}_t &:= \frac{1}{\hat{f}_t\tilde{T}}\sum\nolimits_{\ell}\cos(uU_\ell)\,K_{t\ell}I_\ell,\\
\check{\phi}_t &:= \frac{1}{\hat{f}_t\tilde{T}}\sum\nolimits_{\ell}\phi_\ell K_{t\ell}I_\ell,
\end{aligned}$$

with $\hat{I}_t = 1_{\{|\hat{f}_t|>b\}}$, $I_t = 1_{\{|f_t|>b/2\}}$, and consider the decomposition, $\left(\tilde{\alpha}^*_T - \tilde{\alpha}^\dagger_T\right)(v) = \sum_{j=1}^8 q_{jT}(v)$,

$$
\begin{aligned}
q_{1T}(u,x) &:= \frac{2}{\tilde{T}}\sum\nolimits_t \zeta_t\left[\left(\sin\left(u\hat{U}_t\right)-\sin\left(uU_t\right)\right)+u\left(U_t-\hat{U}_t\right)\hat{\phi}_t\right]\hat{f}_t\left(\hat{f}_t-f_t\right)\theta_t(x),\\
q_{2T}(u,x) &:= \frac{2}{\tilde{T}}\sum\nolimits_t \zeta_t\left[\left(\sin\left(u\hat{U}_t\right)-\sin\left(uU_t\right)\right)+u\left(U_t-\hat{U}_t\right)\hat{\phi}_t\right]\hat{f}_t f_t\theta_t(x),\\
q_{3T}(u,x) &:= \frac{2}{\tilde{T}}\sum\nolimits_t \zeta_t\left(\sin\left(uU_t\right)-uU_t\hat{\phi}_t(u)\right)\left(\hat{f}_t-f_t\right)^2\theta_t(x),\\
q_{4T}(u,x) &:= \frac{2}{\tilde{T}}\sum\nolimits_t \zeta_t\left(2\sin\left(uU_t\right)-uU_t\left(\hat{\phi}_t+\phi_t\right)(u)\right)f_t\left(\hat{f}_t-f_t\right)\theta_t(x),\\
q_{5T}(u,x) &:= \frac{2u}{\tilde{T}}\sum\nolimits_t \zeta_t U_t\left(\tilde{\phi}_t-\hat{\phi}_t\right)(u)\hat{f}_t f_t\theta_t(x),\\
q_{6T}(u,x) &:= \frac{2u}{\tilde{T}}\sum\nolimits_t \zeta_t U_t\left(\bar{\phi}_t-\tilde{\phi}_t\right)(u)\hat{f}_t f_t\theta_t(x),\\
q_{7T}(u,x) &:= \frac{2u}{\tilde{T}}\sum\nolimits_t \zeta_t U_t\left(\check{\phi}_t-\bar{\phi}_t\right)(u)\hat{f}_t f_t\theta_t(x),\\
q_{8T}(u,x) &:= \frac{2u}{\tilde{T}}\sum\nolimits_t \zeta_t U_t\left(\phi_t-\check{\phi}_t\right)(u)\hat{f}_t f_t\theta_t(x).
\end{aligned}
$$

Then (38) follows by showing that $\left\|\sqrt{T}q_{jT}\right\|_{\mathcal{L}^2_W}=o_{\mathbb{P}^*}(1)$, $j=1,...,8$. In order to show that $\left\|\sqrt{T}q_{1T}\right\|_{\mathcal{L}^2_W}=o_{\mathbb{P}^*}(1)$, apply the mean value theorem (MVT), Cauchy-Swartz's inequality, (22) and (30), to obtain that,

$$
\begin{aligned}
\mathbb{E}_\zeta\left\|q_{1T}\right\|_{\mathcal{L}^2_W} &\le \frac{4}{T}\sum\nolimits_t\left|\hat{U}_t-U_t\right|\left|\hat{f}_t\right|\left|\hat{f}_t-f_t\right|\\
&\le C\cdot\left(\frac{1}{T}\sum\nolimits_t\left|\hat{U}_t-U_t\right|^2\hat{f}_t^2\right)^{1/2}\left(\frac{1}{T}\sum\nolimits_t\left|\hat{f}_t-f_t\right|^2\right)^{1/2}\\
&= o_{\mathbb{P}}\left(T^{-1/2}\right).
\end{aligned}
$$

Henceforth, $h_t(u,x):=\theta_t^{\tau}(x)\theta_t(x)W(u,x)$. We show that $\left\|\sqrt{T}q_{2T}\right\|_{\mathcal{L}^2_W}=o_{\mathbb{P}^*}(1)$, applying the MVT and (22), i.e.

$$
\begin{aligned}
\mathbb{E}_\zeta\left\|q_{2T}\right\|^2_{\mathcal{L}^2_W} &= \frac{4}{\tilde{T}^2}\sum\nolimits_t f_t^2\hat{f}_t^2\int\left[\left(\sin\left(uU_t\right)-\sin\left(u\hat{U}_t\right)\right)+\left(\hat{U}_t-U_t\right)\right]^2 h_t(u,x)dudx\\
&\le \frac{C}{T^2}\sum\nolimits_t\left|\hat{U}_t-U_t\right|^2\hat{f}_t^2\\
&= o_{\mathbb{P}}\left(T^{-3/2}\right)
\end{aligned}
$$

We show $\left\|\sqrt{T}q_{3T}\right\|_{\mathcal{L}^2_W} = o_{\mathbb{P}^*}(1)$ mimicking the arguments to prove (22) and (30), with $n_{\ell t} = (1+|U_t|)^{1/2} K_{\ell t}$ and $\tilde{n}_t = (1+|U_t|)^{1/2}\left(\tilde{K}_t - f_t\right)$; then,

$$\begin{aligned}
\mathbb{E}_\zeta \left\|q_{3T}\right\|_{\mathcal{L}^2_W} &\leq \frac{C}{T}\sum\nolimits_t (1+|U_t|)\left(\hat{f}_t - f_t\right)^2 \\
&\leq \frac{C}{T^3}\sum\nolimits_t (n_{t\ell}-\tilde{n}_t)^2 + \frac{C}{T}\sum \tilde{n}_t^2 \\
&= o_{\mathbb{P}}\left(T^{-1/2}\right)
\end{aligned}$$

by (29), and applying Cauchy-Schwartz's inequality,

$$\mathbb{E}\left|\tilde{n}_t\right|^2 \leq C\mathbb{E}\left|U_t\right|\mathbb{E}\left|\tilde{K}_t - f_t\right|^2 = O(a^N) = o(T^{-1}) = o(T^{-1/2}).$$

We apply the same arguments to show that $\left\|\sqrt{T}q_{4T}\right\|_{\mathcal{L}^2_W} = o_{\mathbb{P}^*}(1)$ using $n_{\ell t} = (1+|U_t|)K_{\ell t}$ with $\tilde{n}_t = (1+|U_t|)\left(\tilde{K}_t - f_t\right)$; then

$$\begin{aligned}
\mathbb{E}_\zeta \left\|q_{4T}\right\|^2_{\mathcal{L}^2_W} &= \frac{4}{\tilde{T}^2}\int\sum\nolimits_t \left[\left(2\sin(uU_t) + U_t\cdot\left(\hat{\phi}_t + \phi_t\right)(u)\right) f_t\left(\hat{f}_t - f_t\right)\right]^2 h_t(u,x)dudx \\
&\leq \frac{C}{T^2}\sum\nolimits_t (1+|U_t|)^2\left(\hat{f}_t - f_t\right)^2 \\
&\leq \frac{C}{T^4}\sum\nolimits_t (n_{t\ell}-\tilde{n}_t)^2 + \frac{C}{T^2}\sum \tilde{n}_t^2 \\
&= o_{\mathbb{P}}\left(T^{-1}\right).
\end{aligned}$$

We show that $\left\|\sqrt{T}q_{5T}\right\|_{\mathcal{L}_W^2} = o_{\mathbb{P}^*}(1)$, applying Cauchy-Schwartz's inequality, the ergodic theorem, A.2.b and A.4. as follows,

$$\begin{aligned}
\mathbb{E}_\zeta \left\|q_{5T}\right\|^2_{\mathcal{L}_W^2} &= \frac{1}{T^2}\sum\nolimits_t U_t^2 f_t \int u^2 \left[\frac{1}{T}\sum\nolimits_{\ell\neq t}\left[\cos\left(uU_\ell\right)-\cos\left(u\hat{U}_\ell\right)\right] I_\ell K_{t\ell}\right.\\
&\quad \left.+\frac{K_X(0)}{a^pT}\left[\cos\left(uU_t\right)-\cos\left(u\hat{U}_t\right)\right] I_t\right]^2 h_t(u,x)dudx\\
&\leq \frac{C}{b^2T^2}\sum\nolimits_t U_t^2\left(\frac{1}{T}\sum\nolimits_{\ell\neq t}\left|\hat{U}_\ell-U_\ell\right|\left|\hat{f}_\ell\right|\left|K_{t\ell}\right|\right)^2+\frac{C}{a^{2p}T^4}\sum\nolimits_t U_t^2\\
&\leq \frac{C}{Tb^2}\left[\frac{1}{T^2}\sum\nolimits_t\sum\nolimits_{\ell\neq t}U_t^2K_{t\ell}^2\right]\left(\frac{1}{T^2}\sum\nolimits_\ell\left|\hat{U}_\ell-U_\ell\right|^2\hat{f}_\ell^2\right)+O_{\mathbb{P}}\left(\frac{1}{a^{2p}T^3}\right)\\
&= \frac{C}{Tb^2}O_{\mathbb{P}}\left(\frac{1}{a^p}\right)\cdot O_{\mathbb{P}}\left(\frac{1}{T^{3/2}}\right)+o_{\mathbb{P}}\left(\frac{1}{T^2}\right)\\
&= O_{\mathbb{P}}\left(\frac{1}{T}\right)\left[O_{\mathbb{P}}\left(\frac{1}{T^{3/2}b^2a^p}\right)+o_{\mathbb{P}}\left(\frac{1}{T}\right)\right]\\
&= o_{\mathbb{P}}\left(\frac{1}{T}\right).
\end{aligned}$$

To prove that $\left\|\sqrt{T}q_{6T}\right\|_{\mathcal{L}_W^2} = o_{\mathbb{P}^*}(1)$, use that $1_{\{|\hat{f}_\ell|>b\}} \leq 1_{\{|\hat{f}_\ell-f_\ell|>b/2\}} + 1_{\{|f_\ell|>b/2\}}$ and Cauchy-Swartz's inequality;that is,

$$\begin{aligned}
\mathbb{E}_\zeta \left\|q_{6T}\right\|^2_{\mathcal{L}_W^2} &= \frac{1}{\tilde{T}^2}\sum\nolimits_t U_t^2 f_t^2\int\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell s_\ell(u)\left(1_{\{|f_\ell|>b/2\}}-1_{\{|\hat{f}_\ell|>b\}}\right)K_{t\ell}\right)^2 h_t(u,x)dudx\\
&\leq \frac{C}{T^2}\sum\nolimits_t\sum\nolimits_{\ell\neq t}U_t^2\left(\frac{1}{T}\sum\nolimits_{\ell\neq t}1_{\{|\hat{f}_\ell-f_\ell|>b/2\}}\left|K_{t\ell}\right|\right)^2+\frac{C}{a^{2p}T^4}\sum\nolimits_t U_t^2\\
&\leq \frac{C}{T}\left(\frac{1}{T^2}\sum\sum\nolimits_{\ell\neq t}U_t^2K_{t\ell}^2\right)\left(\frac{1}{T}\sum\nolimits_{\ell\neq t}1_{\{|\hat{f}_\ell-f_\ell|>b/2\}}\right)+O_{\mathbb{P}}\left(\frac{1}{T^3a^{2p}}\right)\\
&\leq \frac{C}{T}O_{\mathbb{P}}\left(\frac{1}{a^{p\left(1+\frac{2}{\mu}\right)}}\right)O_{\mathbb{P}}\left(\frac{1}{T^{1/2}b^2}\right)\\
&= o_{\mathbb{P}}\left(\frac{1}{T}\right),
\end{aligned}$$

by A.4, since by Hölder's inequality,

$$\begin{aligned}
\max_{t\neq\ell}\mathbb{E}\left(U_t^2K_{t\ell}^2\right) &\leq \max_{t\neq\ell}\left\{\left[\mathbb{E}\left|U_t\right|^\mu\right]^{\frac{2}{\mu}}\left[\mathbb{E}\left|K_{t\ell}\right|^{\frac{2\mu}{\mu-2}}\right]^{\frac{\mu-2}{\mu}}\right\}\\
&= O\left(\frac{1}{a^{p\left(1+\frac{2}{\mu}\right)}}\right)
\end{aligned}$$

by (28), and by Markov inequality and (30),

$$\max_t \mathbb{E}\left(1_{\{|\hat{f}_t - f_t|>b/2\}}\right) \leq \frac{4}{b^2}\max_t \mathbb{E}\left|\hat{f}_t - f_t\right|^2 = O\left(\frac{1}{T^{1/2}b^2}\right). \tag{41}$$

We show that $\left\|\sqrt{T}q_{7T}\right\|_{\mathcal{L}^2_W} = o_{\mathbb{P}^*}(1)$, first noticing that, since $\sup_u \mathbb{E}\left(\left(s_\ell - \phi_\ell\right)(u)\right|\mathcal{A}_\ell) = 0$ a.s.,

$$\begin{aligned}
\sup_u \mathbb{E}\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \left(\phi_\ell - s_\ell\right)(u) I_\ell K_{t\ell}\right)^2 &= \sup_u \frac{1}{T^2}\sum\nolimits_\ell \mathbb{E}\left(\left(\phi_\ell - s_\ell\right)^2(u) I_\ell K_{t\ell}^2\right) \\
&\leq C\cdot\frac{1}{T}\left[\max_{\ell\neq t}\mathbb{E}\left(K_{t\ell}^2\right) + \frac{1}{a^{2p}T}\right] \\
&= O\left(\frac{1}{a^pT}\right) + O\left(\frac{1}{a^{2p}T^2}\right) \\
&= O\left(\frac{1}{T^{1/2}}\right) + O\left(\frac{1}{T}\right),
\end{aligned}$$

by (28) and A.2.b. Then, applying Hölder's inequality and the ergodic theorem,

$$\begin{aligned}
\mathbb{E}_\zeta \left\|q_{7T}\right\|^2_{\mathcal{L}^2_W} \quad :&= \frac{1}{\tilde{T}^2}\sum\nolimits_t U_t^2 f_t^2 \int\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \left(\phi_\ell - s_\ell\right)(u) I_\ell K_{t\ell}\right)^2 h_t(u,x)dudx, \\
&\leq \frac{C}{T}\left(\frac{1}{T}\sum\nolimits_t |U_t|^\mu\right)^{\frac{2}{\mu}} \\
&\quad \times\left[\frac{1}{T}\sum\nolimits_t\left(\int\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \left(\phi_\ell - s_\ell\right)(u) I_\ell K_{t\ell}\right)^2 h_t(u,x)dudx\right)^{\frac{2\mu}{\mu-2}}\right]^{\frac{\mu-2}{2\mu}} \\
&= \frac{C}{T}\cdot O_{\mathbb{P}}(1)\cdot O_{\mathbb{P}}\left(\frac{1}{T^{1/2}}\right) \\
&= o_{\mathbb{P}}\left(\frac{1}{T}\right)
\end{aligned}$$

In order to show that $\left\|\sqrt{T}q_{8T}\right\|_{\mathcal{L}^2_W} = o_{\mathbb{P}^*}(1)$ first notice that,

$$\begin{aligned}
\mathbb{E}_\zeta \left\|q_{8T}\right\|^2_{\mathcal{L}^2_W} &= \frac{1}{\tilde{T}^2}\sum\nolimits_t U_t^2 f_t^2 \int\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \left(\phi_t(u) - \phi_\ell(u) I_\ell\right) K_{t\ell}\right)^2 h_t(u,x)dudx \\
&\leq 2\frac{1}{\tilde{T}^2}\sum\nolimits_t U_t^2 f_t^2 \int\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \phi_\ell(u)\left(1 - I_\ell\right) K_{t\ell}\right)^2 h_t(u,x)dudx \qquad (42) \\
&\quad +2\frac{1}{\tilde{T}^2}\sum\nolimits_t U_t^2 f_t^2 \int\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \left(\phi_t - \phi_\ell\right)(u) K_{t\ell}\right)^2 h_t(u,x)dudx. \qquad (43)
\end{aligned}$$

In order to prove (42)$= o_{\mathbb{P}}(T^{-1})$, we first show that

$$\sup_u \mathbb{E}\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \phi_\ell(u)\left(1-I_\ell\right)K_{t\ell}\right)^2 = o_{\mathbb{P}}(1). \tag{44}$$

To this end, we mimic the arguments in the proof of (22). Write $\tilde{T}^{-1}\sum_\ell \phi_\ell(u)\left(1-I_\ell\right)K_{t\ell} = \tilde{T}^{-1}\sum_\ell w_{t\ell}$ with $w_{t\ell}(u) = w(u, X_\ell, X_t)$,

$$w(u,x_1,x_2) = \frac{1}{a^p}\phi(u,x_1)\left(1-1_{\{|f(x_1)|>b/2\}}\right)K_X\left(\frac{x_1-x_2}{a}\right).$$

Define $\tilde{w}(u,x) = \mathbb{E}\left[w(u,X_t,x)\right]$ and $\tilde{w}_t(u) = \tilde{w}(u,X_t)$. The left-hand side of (44) is bounded by a constant times

$$\begin{aligned} & \frac{K_X^2(0)}{T^2 a^{2p}} + \sup_u \mathbb{E}\left(\frac{1}{T}\sum\nolimits_{\ell\neq t}\left(w_{t\ell}-\tilde{w}_t\right)(u)\right)^2 + \sup_u \mathbb{E}\left(\tilde{w}_t^2(u)\right) \qquad (45) \\ = \; & o_{\mathbb{P}}(T^{-1}) + o_{\mathbb{P}}(T^{-1/2}) + o_{\mathbb{P}}(1), \end{aligned}$$

by A.2.b, and, since the second term of (45) is $o_{\mathbb{P}}(T^{-1/2})$ as (29),

$$\begin{aligned} \sup_u \mathbb{E}\left(\tilde{w}_t^2(u)\right) &= \int\left[\int \frac{1}{a^p}\phi(u,x_1)\left(1_{\{|f(x_1)|>b/2\}}-1\right)K_X\left(\frac{x_1-x_2}{a}\right)f(x_1)dx_1\right]^2 f(x_2)dx_2 \\ &= \int\left[\int_{\{|f(v+ax_2)|\le b/2\}} K_X(v)\,f(v+ax_2)dv\right]^2 f(x_2)dx_2 \\ &= o(1), \end{aligned}$$

as $a, b \to 0$ by dominated convergence. Then, applying Hölder's inequality, (42) is bounded by

$$\begin{aligned} & \frac{C}{T}\left(\frac{1}{T}\sum\nolimits_t |U_t|^\mu\right)^{\frac{2}{\mu}}\left[\frac{1}{T}\sum\nolimits_t\left(\int\left(\frac{1}{\tilde{T}}\sum\nolimits_\ell \phi_\ell(u)\left(I_\ell-1\right)K_{t\ell}\right)^2 h_t(u,x)dudx\right)^{\frac{2\mu}{\mu-2}}\right]^{\frac{\mu-2}{2\mu}} \\ = \; & \frac{C}{T}\cdot O_{\mathbb{P}}(1)\cdot o_{\mathbb{P}}(1) \\ = \; & o_{\mathbb{P}}\left(\frac{1}{T}\right) \end{aligned}$$

by the ergodic theorem and (44). We show that (43)$= o_{\mathbb{P}}(T^{-1})$ reasoning as in the proof of (44) and using the fact that

$$\sup_u \mathbb{E}\left(\frac{1}{\tilde{T}}\sum\nolimits_{\ell}\left(\phi_t-\phi_\ell\right)(u)K_{t\ell}\right)^2 = o\left(\frac{1}{T^{1/2}}\right),$$

applying the same arguments as in the proof (30) with $n_{t\ell} = \left(\phi_\ell - \phi_t\right)(u)K_{t\ell}$ and noticing that $\phi(u,\cdot)$ is uniformly bounded and satisfies the same smoothing properties than $r(\cdot)$.

Now, under H_0, H_1 or H_{1T}, by dominated convergence and the ergodic theorem,

$$\limsup_{T\to\infty} T\mathbb{E}_\zeta \left\|\tilde{\alpha}_T^\dagger\right\|^2_{\mathcal{L}^2_W} = \limsup_{T\to\infty}\frac{1}{T}\sum\nolimits_t \left\|\tilde{V}_t\right\|^2_{\mathcal{L}^2_W} = \mathbb{E}\left\|\tilde{V}_t\right\|^2_{\mathcal{L}^2_W} \ a.s.,$$

where $\mathbb{E}\left\|\tilde{V}_t\right\|^2_{\mathcal{L}^2_W} < \infty$, which shows (39). We show (40) by applying Lemma 2 with $\xi_{Tt}(v) = \zeta_t\tilde{V}_t(v)\Big/\sqrt{T}$. For each fixed sample $\{(Y_t,Z_t)\}_{t=1}^T$, $\{\xi_{Tt}\}_{t=1+k}^T$ is a sequence of independent $\mathcal{L}^2_W - valued$ random elements with $\mathbb{E}_\zeta\left(\xi_{Tt}(v)\right) = 0$ and $\mathbb{E}_\zeta\left(\xi^2_{Tt}(v)\right) < \infty$ uniformly in $v \in \mathbb{R}^{1+p}$. Before checking (i)-(iii) in Lemma 2, we first obtain the (conditional) covariance function of $\sqrt{T}\tilde{\alpha}_T^\dagger$,

$$T\mathbb{E}_\zeta\left[\tilde{\alpha}_T^\dagger(v_1)\,\tilde{\alpha}_T^{\dagger\tau}(v_2)\right] = \frac{1}{\tilde{T}}\sum\nolimits_t \tilde{V}_t(v_1)\,\tilde{V}_t^\tau(v_2) =: \Sigma_T(v_1,v_2).$$

Then, to prove (i), by dominated convergence and the ergodic theorem,

$$\begin{aligned}\lim_{T\to\infty}\langle\Sigma_T e_j, e_\ell\rangle &= \int\!\!\int e_j^\tau(v_1)\lim_{T\to\infty}\Sigma_T(v_1,v_2)\,e_\ell(v_2)\,W(v_1)W(v_2)dv_1dv_2 \ a.s. \qquad (46)\\ &= \int\!\!\int e_j^\tau(v_1)\,\Sigma(v_1,v_2)\,e_\ell(v_2)\,W(v_1)W(v_2)dv_1dv_2 \ a.s.\\ &=: a_{j\ell} \text{ for all } j,\ell \geq 0.\end{aligned}$$

To prove (ii), use (46) to obtain

$$\lim_{T\to\infty}\sum_{\ell=0}^\infty\langle\Sigma_T e_\ell, e_\ell\rangle = \sum_{\ell=0}^\infty a_{\ell\ell} \leq C\left\|\sum_{\ell=0}^\infty a_{\ell\ell}\right\|^2_{\mathcal{L}^2_W} < \infty \ a.s.$$

Finally, to prove (iii), for every $\varepsilon > 0$ and every $\ell \geq 0$

$$\begin{aligned}
& \sum\nolimits_t \mathbb{E}_\zeta \left(\left\langle \zeta_t \tilde{V}_t(v) \middle/ \sqrt{T}, e_\ell \right\rangle^2 1_{\left\{ \left| \left\langle \zeta_t \tilde{V}_t(v)/\sqrt{T}, e_\ell \right\rangle \right| > \varepsilon \right\}} \right) \\
= & \sum\nolimits_t \left\langle \tilde{V}_t(v) \middle/ \sqrt{T}, e_\ell \right\rangle^2 1_{\left\{ \left| \left\langle \tilde{V}_t(v)/\sqrt{T}, e_\ell \right\rangle \right| > \varepsilon \right\}} \ a.s. \\
\leq & \frac{1}{\varepsilon^\mu T^{1+\mu/2}} \sum_t \left\langle \tilde{V}_t(v), e_\ell \right\rangle^{2+\mu} 1_{\left\{ \left| \left\langle \tilde{V}_t(v)/\sqrt{T}, e_\ell \right\rangle \right| > \varepsilon \right\}} \ a.s. \\
= & O(T^{-\mu/2}) \ a.s.
\end{aligned}$$

by the ergodic theorem. ■

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.

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

Bai, J., 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.

Bickel, P.J., Rosenblatt, M. (1973). "On some global measures of the deviations of density function estimates", *Annals of Statistics*, 1, 1071-1095.

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

Bierens, H.J., Ploberger, W. (1997). "Asymptotic theory of integrated conditional moment tests", *Econometrica,* 65, 1129-1151

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

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.

Csörgo, S., Heathcote, C.R. (1987). “Testing for symmetry”, *Biometrika*, 74, 177-184.

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. (2000). “Significance testing in nonparametric regression based on the bootstrap”, *Annals of Statistics*, 1469-1507.

Denker, M., Keller, G. (1983). “On U-statistics and v. Mises' statistics for weakly dependent processes”, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 64, 505-522.

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.

Fan, Y. (1994). “Testing the goodness of fit of a parametric density function by kernel method”, *Econometric Theory*, 10, 316-356.

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. (1998). “Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters”, *Econometric Theory*, 14, 604-621.

Fang, Y., Q. Li, X. Wu and D. Zhang (2015). “A data driven smooth test of symmetry”, *Journal of Econometrics*, 188, 490-501.

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

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

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.

Henze, N., Klar, B., Meintanis, S.G. (2003). "Invariant tests for symmetry about an unspecified point based on the empirical characteristic function", *Journal of Multivariate Analysis*, 87, 275-297.

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.

Hušková, M., Meintanis, S.G. (2012). "Tests for symmetric error distribution in linear and nonparametric regression models", *Communications in Statistics - Simulation and Computation*, 41, 833-851.

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.

Jakubowski, A. (1980). "On limit theorems for sums of dependent Hilbert space valued random variables", *In: Lecture Notes in Statistics, Vol. 2. Springer, New York, pp. 178-187.*

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

Kundu, S., Majumdar, S., Mukherjee, K. (2000). "Central limit theorems revisited", *Statistics and Probability Letters*, 47, 265-275.

Métivier, M., Nakao, S. (1987). "Equivalent conditions for the tightness of a sequence of continuous Hilbert valued martingales", *Nagoya Mathematical Journal*, 106, 113-119.

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

Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*, Academic Press.

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.

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

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.

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

Walk, H. (1977). "An invariance principle for the Robbins-Monro process in a Hilbert space", *Zeitschrift fur Warscheinlichkeitstheorie und Verwandte Gebiete*, 39, 135-150.

Xie, Y. (1995). "Limit theorems of Hilbert valued semimartingales and Hilbert valued martingale measures", *Stochastic Processes and Their Applications*, 59, 277-293.

Table 1: Empirical rejection frequency of $\hat{\eta}_T(0.5)$ for (AU1)-(AU12)

DGPs	$\hat{\eta}_T^{(1)}(0.5)$	$\hat{\eta}_T^{(2)}(0.5)$	$\hat{\eta}_T^{(3)}(0.5)$	$\hat{\eta}_T^{(1)}(0.5)$	$\hat{\eta}_T^{(2)}(0.5)$	$\hat{\eta}_T^{(3)}(0.5)$	$\hat{\eta}_T^{(1)}(0.5)$	$\hat{\eta}_T^{(2)}(0.5)$	$\hat{\eta}_T^{(3)}(0.5)$
		$T = 50$			$T = 100$			$T = 200$	
(AU1)	0.03	0.04	0.05	0.04	0.05	0.05	0.04	0.05	0.05
(AU2)	0.07	0.07	0.06	0.07	0.05	0.05	0.06	0.06	0.06
(AU3)	0.03	0.03	0.05	0.03	0.04	0.04	0.04	0.03	0.06
(AU4)	0.04	0.04	0.04	0.04	0.05	0.06	0.04	0.04	0.05
(AU5)	0.02	0.03	0.04	0.02	0.04	0.04	0.03	0.04	0.05
(AU6)	0.03	0.03	0.04	0.05	0.03	0.04	0.04	0.04	0.05
(AU7)	0.92	0.94	0.94	0.99	0.99	0.99	1.00	1.00	1.00
(AU8)	0.75	0.83	0.84	0.98	0.99	0.99	1.00	1.00	1.00
(AU9)	0.81	0.89	0.91	0.99	0.99	0.99	1.00	1.00	1.00
(AU10)	0.26	0.27	0.27	0.43	0.46	0.48	0.71	0.74	0.76
(AU11)	0.89	0.96	0.96	0.99	1.00	1.00	1.00	1.00	1.00
(AU12)	0.91	0.96	0.97	0.99	1.00	0.99	1.00	1.00	1.00

Table 2: Empirical rejection frequency of $\hat{\eta}_T(1)$ for (AU1)-(AU12)

DGPs	$\hat{\eta}_T^{(1)}(1)$	$\hat{\eta}_T^{(2)}(1)$	$\hat{\eta}_T^{(3)}(1)$	$\hat{\eta}_T^{(1)}(1)$	$\hat{\eta}_T^{(2)}(1)$	$\hat{\eta}_T^{(3)}(1)$	$\hat{\eta}_T^{(1)}(1)$	$\hat{\eta}_T^{(2)}(1)$	$\hat{\eta}_T^{(3)}(1)$
		$T = 50$			$T = 100$			$T = 200$	
(AU1)	0.04	0.05	0.06	0.05	0.06	0.06	0.05	0.06	0.06
(AU2)	0.08	0.07	0.06	0.08	0.06	0.05	0.06	0.06	0.06
(AU3)	0.04	0.04	0.06	0.03	0.05	0.06	0.04	0.05	0.05
(AU4)	0.05	0.04	0.04	0.05	0.05	0.05	0.06	0.05	0.06
(AU5)	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.04
(AU6)	0.05	0.04	0.04	0.05	0.04	0.04	0.06	0.05	0.04
(AU7)	0.89	0.91	0.93	0.99	0.99	0.99	1.00	1.00	1.00
(AU8)	0.78	0.84	0.85	0.99	0.99	0.99	1.00	1.00	1.00
(AU9)	0.82	0.89	0.91	0.99	0.99	0.99	1.00	1.00	1.00
(AU10)	0.30	0.30	0.29	0.49	0.46	0.45	0.74	0.75	0.75
(AU11)	0.89	0.94	0.94	0.99	1.00	0.99	1.00	1.00	1.00
(AU12)	0.92	0.95	0.95	0.99	0.99	0.99	1.00	1.00	1.00

Table 3: Empirical rejection frequency of T_S, $\hat{\eta}_{T,m}(0.5)$, and $\hat{\eta}_T(0.5)$ for (S1)-(A11)

DGPs	T_S	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T^{(1)}(0.5)$	$\hat{\eta}_T^{(2)}(0.5)$	$\hat{\eta}_T^{(3)}(0.5)$	T_S	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T^{(1)}(0.5)$	$\hat{\eta}_T^{(2)}(0.5)$	$\hat{\eta}_T^{(3)}(0.5)$	T_S	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T^{(1)}(0.5)$	$\hat{\eta}_T^{(2)}(0.5)$	$\hat{\eta}_T^{(3)}(0.5)$
			$T = 25$					$T = 50$					$T = 100$		
(S1)	0.03	0.04	0.02	0.03	0.03	0.03	0.04	0.03	0.03	0.05	0.05	0.05	0.03	0.04	0.04
(S2)	0.03	0.06	0.05	0.05	0.05	0.06	0.07	0.04	0.05	0.05	0.06	0.06	0.05	0.05	0.05
(S3)	0.03	0.05	0.02	0.03	0.02	0.03	0.05	0.02	0.03	0.03	0.07	0.05	0.03	0.03	0.04
(S4)	0.04	0.05	0.03	0.03	0.04	0.05	0.05	0.04	0.04	0.03	0.04	0.05	0.03	0.03	0.04
(S5)	0.02	0.05	0.05	0.06	0.05	0.04	0.05	0.04	0.07	0.06	0.05	0.05	0.06	0.06	0.06
(S6)	0.03	0.08	0.07	0.07	0.05	0.04	0.07	0.07	0.06	0.06	0.05	0.06	0.07	0.07	0.06
(S7)	0.05	0.09	0.09	0.08	0.08	0.05	0.08	0.09	0.08	0.07	0.06	0.07	0.09	0.07	0.06
(A1)	0.53	0.92	0.48	0.54	0.58	0.91	0.99	0.83	0.88	0.89	0.97	1.00	0.99	0.99	0.99
(A2)	0.48	0.81	0.29	0.37	0.40	0.94	0.99	0.64	0.75	0.80	1.00	1.00	0.94	0.98	0.99
(A3)	0.51	0.81	0.31	0.42	0.47	0.95	0.99	0.66	0.80	0.85	1.00	1.00	0.96	0.99	0.99
(A4)	0.22	0.25	0.06	0.11	0.11	0.53	0.49	0.13	0.21	0.27	0.89	0.81	0.34	0.48	0.55
(A5)	0.20	0.42	0.15	0.19	0.23	0.63	0.74	0.33	0.46	0.50	0.97	0.97	0.70	0.82	0.85
(A6)	0.12	0.22	0.10	0.14	0.14	0.24	0.35	0.19	0.24	0.24	0.58	0.59	0.38	0.42	0.44
(A7)	0.56	0.86	0.39	0.53	0.59	0.95	0.99	0.79	0.88	0.93	0.99	1.00	0.99	1.00	1.00
(A8)	0.54	0.89	0.42	0.58	0.63	0.93	1.00	0.81	0.90	0.93	0.99	1.00	0.99	1.00	1.00
(A9)	0.00	0.01	0.23	0.18	0.14	0.00	0.00	0.45	0.37	0.30	0.00	0.00	0.72	0.68	0.63
(A10)	0.02	0.01	0.10	0.10	0.08	0.01	0.00	0.20	0.21	0.17	0.01	0.00	0.41	0.41	0.40
(A11)	0.02	0.02	0.11	0.13	0.11	0.01	0.01	0.21	0.26	0.24	0.01	0.00	0.41	0.51	0.46

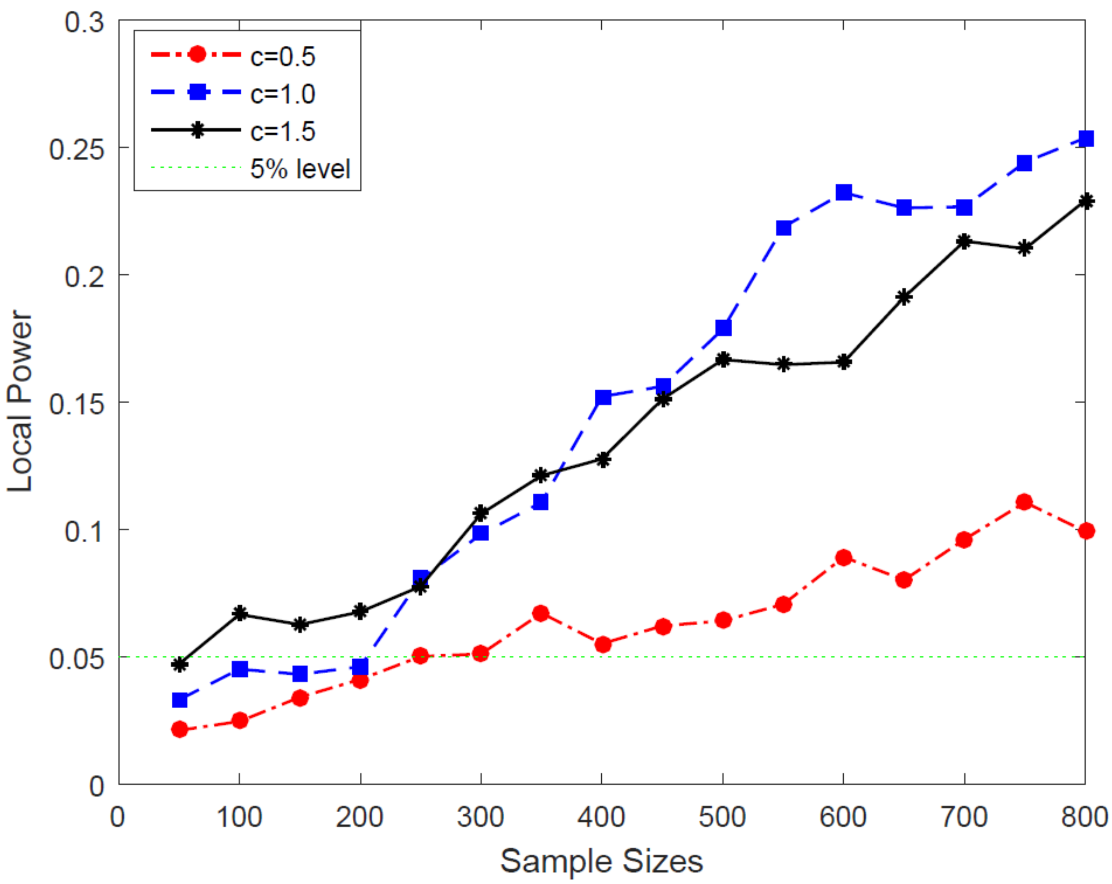

Table 4: Empirical rejection frequency of $T_{0.5}^{(2)}$, $\hat{\eta}_{T,m}(0.5)$ and $\hat{\eta}_T(0.5)$ for (LS1)-(LS6)

Panel A ($a_T = 0.5 * s_X * T^{-1/3}$)

DGPs	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$
		$T = 50$			$T = 100$			$T = 200$	
(LS1)	0.04	0.03	0.03	0.04	0.03	0.02	0.04	0.05	0.03
(LS2)	0.04	0.03	0.04	0.04	0.03	0.04	0.05	0.05	0.04
(LS3)	0.05	0.03	0.03	0.04	0.03	0.02	0.04	0.04	0.03
(LS4)	0.04	0.18	0.37	0.04	0.32	0.81	0.05	0.38	0.99
(LS5)	0.04	0.18	0.45	0.04	0.28	0.87	0.05	0.32	1.00
(LS6)	0.04	0.19	0.45	0.05	0.25	0.88	0.05	0.30	1.00

Panel B ($a_T = 1.0 * s_X * T^{-1/3}$)

DGPs	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$
		$T = 50$			$T = 100$			$T = 200$	
(LS1)	0.04	0.04	0.04	0.03	0.05	0.03	0.04	0.06	0.04
(LS2)	0.04	0.03	0.05	0.04	0.05	0.05	0.05	0.06	0.05
(LS3)	0.03	0.03	0.04	0.04	0.04	0.03	0.05	0.05	0.04
(LS4)	0.04	0.31	0.58	0.07	0.53	0.94	0.10	0.60	1.00
(LS5)	0.05	0.35	0.67	0.06	0.47	0.98	0.08	0.54	1.00
(LS6)	0.05	0.34	0.70	0.07	0.43	0.97	0.08	0.49	1.00

Panel C ($a_T = 1.5 * s_X * T^{-1/3}$)

DGPs	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$
		$T = 50$			$T = 100$			$T = 200$	
(LS1)	0.04	0.04	0.05	0.07	0.05	0.04	0.05	0.05	0.04
(LS2)	0.04	0.05	0.07	0.05	0.05	0.06	0.06	0.06	0.06
(LS3)	0.04	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.05
(LS4)	0.09	0.44	0.63	0.16	0.55	0.96	0.20	0.59	1.00
(LS5)	0.09	0.43	0.70	0.16	0.53	0.98	0.17	0.54	1.00
(LS6)	0.08	0.42	0.74	0.15	0.50	0.98	0.16	0.54	1.00

Panel D ($a_T = 0.075$)

DGPs	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$	$T_{0.5}^{(2)}$	$\hat{\eta}_{T,m}(0.5)$	$\hat{\eta}_T(0.5)$
		$T = 50$			$T = 100$			$T = 200$	
(LS1)	0.04	0.02	0.04	0.05	0.04	0.05	0.05	0.05	0.05
(LS2)	0.04	0.03	0.06	0.05	0.05	0.06	0.06	0.05	0.05
(LS3)	0.04	0.03	0.04	0.05	0.05	0.05	0.06	0.05	0.05
(LS4)	0.05	0.30	0.56	0.09	0.52	0.95	0.19	0.58	0.99
(LS5)	0.04	0.32	0.67	0.07	0.48	0.98	0.16	0.54	1.00
(LS6)	0.05	0.32	0.70	0.08	0.47	0.98	0.16	0.53	1.00

Table 5: Bootstrapped p-values of $\hat{\eta}_T(0.5)$ of stock markets before (resp. after) the crisis and business cycles

Bandwidths	S&P500	FTSE100	Nikkei225	SSE-A	USA	UK	Japan	China
$c = 0.5$	0.802 (0.000)	0.194 (0.054)	0.168 (0.004)	0.942 (0.008)	0.134	0.006	0.374	0.358
$c = 1.0$	0.754 (0.000)	0.112 (0.022)	0.236 (0.000)	0.778 (0.002)	0.074	0.010	0.438	0.656
$c = 1.5$	0.728 (0.000)	0.096 (0.020)	0.228 (0.000)	0.718 (0.008)	0.078	0.006	0.438	0.682