

# Nonparametric tests for conditional symmetry in dynamic models<sup>☆</sup>

Miguel A. Delgado<sup>a,\*</sup>, J. Carlos Escanciano<sup>b</sup>

<sup>a</sup>*Department of Economics, Universidad Carlos III de Madrid, C./Madrid 126–128, 28903 Getafe, Spain*

<sup>b</sup>*Department of Economics, Indiana University, 100 S. Woodlawn, Wylie Hall, Bloomington, IN 47405-7104, USA*

---

## Abstract

This article proposes omnibus tests for conditional symmetry around a parametric function in a dynamic context. Conditional moments may not exist or may depend on the explanatory variables. Test statistics are suitable functionals of the empirical process of residuals and explanatory variables, whose limiting distribution under the null is nonpivotal. The tests are implemented with the assistance of a bootstrap method, which is justified assuming very mild regularity conditions on the specification of the center of symmetry and the underlying serial dependence structure. Finite sample properties are examined by means of a Monte Carlo experiment.

*JEL classification:* C12; C14; C15

*Keywords:* Omnibus tests; Symmetry tests; Conditional distributions; Time series; Empirical processes; Bootstrap

---

## 1. Introduction

Testing symmetry of a distribution is a useful model specification tool. Location and dispersion can be unambiguously defined under symmetry and, unlike other location parameters, the center of symmetry can be robustly, even adaptively, estimated. Smirnov (1947) first proposed an omnibus test for the simple hypothesis of symmetry around a

---

<sup>☆</sup>Research funded by the Spanish Ministerio de Educación y Ciencia, reference number SEJ2004-04583/ECON.

\*Corresponding author.

*E-mail addresses:* miguelangel.delgado@uc3m.es (M.A. Delgado), jescanci@indiana.edu (J. Carlos Escanciano).

known value based on the standard empirical process. See also the related works by [Butler \(1969\)](#), [Nadaraya \(1975\)](#), [Hill and Rao \(1977, 1981\)](#), [Aki \(1981\)](#), [Antille et al. \(1982\)](#), [Bhattacharya et al. \(1982\)](#) or more recently [Neumeyer and Dette \(2003\)](#).

In this article, we are concerned with omnibus testing of the composite hypothesis of conditional symmetry around a parametric function in a time series context. Such tests are well motivated in econometrics practice. For instance, it is interesting to test, given the available information at a given period of time, whether or not profits and losses are equally likely in financial markets, or whether or not positive and negative shocks are equally likely in macroeconomic models. [Bai and Ng \(2001\)](#) provide an excellent motivation of testing conditional symmetry in econometric applications. Unlike other symmetry tests, our testing procedure allows for higher conditional moments of unknown functional form, e.g., nonparametric conditional heteroskedasticity or heterokurtosis. In fact, it is not even assumed that any innovation moment exists, which is convenient when dealing with financial data, where conditional distributions frequently exhibit fat tails.

Consider a  $\mathbb{R}^{1+k}$ -valued strictly stationary multivariate time series process  $(Y, X) = (Y_t, X_t)_{t \in \mathbb{Z}}$ , with an information set  $I_t = \{(Y_s, X_s), t - m + 1 \leq s \leq t\}$  at time  $t$ , i.e.,  $I_t \in \mathbb{R}^p$  with  $p = m(1 + k)$ . Given a suitable parameter space  $\Theta \subset \mathbb{R}^q$  and a function  $g : \mathbb{R}^p \times \Theta \rightarrow \mathbb{R}$ , we are interested in testing that the conditional distribution function of  $Y_t$  given  $I_t = \cdot$  is symmetric around  $g(\cdot, \theta_0)$  for some  $\theta_0 \in \Theta$ . Consider the family of symmetric distributions around zero

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

Thus, the null hypothesis can be expressed as

$$H_0 : \text{There exists a } \theta_0 \in \Theta \text{ such that } F_{\theta_0}(\cdot|u) \in \mathcal{G} \text{ a.s.}, \quad (1)$$

where  $F_{\theta}(v|u)$  is the conditional distribution of the residuals  $\varepsilon_t(\theta) = Y_t - g(I_t, \theta)$  given  $I_t = u$  evaluated at  $\varepsilon_t(\theta) = v$ . We consider omnibus tests, i.e., the alternative hypothesis is the negation of the null.

The null hypothesis states the correct specification of the dynamic model

$$Y_t = g(I_t, \theta_0) + \varepsilon_t, \quad t \in \mathbb{Z},$$

where  $\varepsilon_t = \varepsilon_t(\theta_0)$  are innovations with a conditional symmetry center equal to zero. The parameter vector  $\theta_0$  can be identified under  $H_0$  as the solution of the moment equations

$$E[\psi(Y_t, I_t, \theta_0)] = 0 \quad (2)$$

for some vector of functions  $\psi(\cdot)$  or, alternatively as the optimizing value

$$\theta_0 = \arg \min_{\theta \in \Theta} E[\rho(\varepsilon_1(\theta))], \quad (3)$$

for some function  $\rho(\cdot)$ . For instance, if  $\rho(v) = v^2$ ,  $E(Y_t|I_t) = g(I_t, \theta_0)$ , and if  $\rho(v) = |v|$ ,  $\text{Median}(Y_t|I_t) = g(I_t, \theta_0)$ . The corresponding  $\psi$  functions for the conditional mean and median are  $\psi(v, u, \theta) = \nabla_{\theta} g(u, \theta) \cdot (v - g(u, \theta))$  and  $\psi(v, u, \theta) = \nabla_{\theta} g(u, \theta) \cdot \text{sign}(v - g(u, \theta))$ , respectively. Henceforth,  $\nabla_a$  means derivative w.r.t.  $a$ . Under  $H_0$ ,  $g(\cdot, \theta_0)$  is the conditional center of symmetry, which is equal to the conditional median and also to the conditional mean when these conditional location functions exist. Under  $H_1$ , there is also a  $\theta_0$  defined by (2) or (3) and  $g(\cdot, \theta_0)$  is no longer the center of symmetry, but any other conditional location function.

The many procedures for testing the symmetry of the marginal distribution of data around an unknown parameter can be also applied for testing the symmetry of the marginal distribution of the regression errors  $\varepsilon_t = Y_t - g(I_t, \theta_0)$  around zero. That is, for testing

$$\dot{H}_0 : E[F_{\theta_0}(v|I_1)] = 1 - E[F_{\theta_0}(-v|I_1)] \quad \text{a.s. for some } \theta_0 \in \Theta. \quad (4)$$

Fan and Gencay (1995) and Ahmad and Li (1997) proposed omnibus tests of  $\dot{H}_0$ , consistent in the direction of general nonparametric alternatives, based on smooth estimates of the marginal probability density of  $\varepsilon_t$  with independent and identically distributed (*iid*) observations. These tests are inconsistent for testing  $H_0$  in any direction where  $\dot{H}_0$  holds. Zheng (1998) proposed an omnibus test for  $H_0$  based on smooth nonparametric estimates of the conditional distribution function. In these tests, the testing decision often depends on the choice of a smoothing parameter, despite of the satisfaction of several smoothness assumptions on the underlying conditional probability density. In this paper, we propose omnibus tests for the composite hypothesis  $H_0$  without using smoothers and under fairly general regularity conditions on the underlying data generating process (DGP).

The null hypothesis  $H_0$  in (1) can be equivalently expressed as

$$H_0 : S_{\theta_0}(u, v) = 0 \quad \text{a.s. for some } \theta_0 \in \Theta, \quad (5)$$

where

$$S_{\theta}(u, v) = K_{\theta}(u, v) - K_{\theta}(u, \infty) + K_{\theta}(u, -v),$$

and

$$\begin{aligned} K_{\theta}(u, v) &= E[1(I_1 \leq u)F_{\theta}(v|I_1)] \\ &= E[1(I_1 \leq u)1(\varepsilon_1(\theta) \leq v)] \end{aligned}$$

is the joint cumulative distribution function (cdf) of the  $(p + 1)$ -valued random variable  $(I_1, \varepsilon_1(\theta))$ . Henceforth, inequalities are coordinatewise. Given some suitable square-root- $n$ -consistent estimator of  $\theta_0$ ,  $\theta_n$  say, this formulation of  $H_0$  suggests to use the empirical process

$$S_{n,\theta_n}(u, v) = \sqrt{n}[K_{n,\theta_n}(u, v) - K_{n,\theta_n}(u, \infty) + K_{n,\theta_n}(u, -v)],$$

where

$$K_{n,\theta}(u, v) = \frac{1}{n} \sum_{i=1}^n 1(I_i \leq u)1(\varepsilon_i(\theta) \leq v)$$

is the joint empirical distribution of  $\{\varepsilon_i(\theta), I_i\}_{i=1}^n$ , which estimates  $K_{\theta}(u, v)$ .

The empirical process  $S_{n,\theta_n}(\infty, \cdot)$  forms a basis for testing symmetry of the marginal distribution of the innovations. Tests based on  $S_{n,\theta_0}(\infty, \cdot)$  were first proposed by Butler (1969) for testing the simple symmetry hypothesis, when parameters are known, as a variation of the empirical process introduced by Smirnov (1947). The asymptotic distribution of  $S_{n,\theta_n}(\infty, \cdot)$  depends on unknown features of the underlying DGP, because of the effect of estimated parameters. Therefore, the asymptotic distribution of functionals of  $S_{n,\theta_n}(\infty, \cdot)$ , used as test statistics, cannot be tabulated. This is why Bai and Ng (2001) proposed to use a martingale transform of  $S_{n,\theta_n}(\infty, \cdot)$  resulting in an asymptotically distribution free empirical process, as suggested by Khmaladze (1981) in a different

context. See also [Bai \(2003\)](#) and [Delgado and Stute \(2005\)](#) for martingale transform applications to testing conditional distribution model specification. These transformations are computationally challenging, specially in the multiparameter case, like the one considered in this article. Bootstrap assisted tests are well motivated under these circumstances.

The test proposed by [Bai and Ng \(2001\)](#) is omnibus for testing symmetry around zero of the marginal distribution of the conditionally scaled innovations  $U_t = \varepsilon_t/\sigma(I_t, \gamma_0)$ , where  $\text{Var}(Y_t|I_t = \cdot) = \sigma^2(\cdot, \gamma_0)$ , with  $\sigma^2$  a known function and  $\gamma_0$  an unknown parameter vector. The resulting tests are sensitive to the parametric specification of the conditional variance, and they are inconsistent in any of the infinite directions where the marginal distribution of the standardized errors is symmetric around zero. In this article, we propose conditional symmetry tests where higher order conditional moments of  $\varepsilon_t(\theta_0)$  given  $I_t$  are nonparametric, i.e., unknown functions of the information set. That is, the serial dependence structure of the innovations is unknown. Furthermore, conditional moments may not exist.

By defining  $S_{n,\theta_n}(-\infty, \cdot) = S_{n,\theta_n}(\cdot, -\infty) = 0$ , the sample paths of  $S_{n,\theta_n}$  belong to the space  $\ell^\infty(\overline{\mathbb{R}}^d)$ , the space of all uniformly bounded real functions on  $\overline{\mathbb{R}}^d := [-\infty, \infty]^d$ , with  $d = p + 1$ , which is equipped with the sup-norm. Tests statistics are continuous functionals of  $S_{n,\theta_n}$ , say  $\varphi(S_{n,\theta_n})$ , for some suitable continuous functional  $\varphi : \ell^\infty(\overline{\mathbb{R}}^d) \mapsto \mathbb{R}^+ \equiv [0, \infty)$ . The most popular criteria are the Kolmogorov–Smirnov  $\varphi(g) = \sup_{(u,v) \in \mathbb{R}^d} |g(u, v)|$  and the Cramér–von Mises  $\varphi(g) = \int_{\mathbb{R}^2} g(u, v)^2 \Phi(du, dv)$  for some suitable measure function  $\Phi$ , which is usually a consistent estimate of the distribution  $K_{\theta_0}$ , e.g.,  $K_{n,\theta_n}$ . Once a Functional Central Limit Theorem (FCLT) for  $S_{n,\theta_n}$  is provided, the limiting distribution of  $\varphi(S_{n,\theta_n})$  under  $H_0$  is obtained by applying the Continuous Mapping Theorem (CMT).

In this article we consider convergence in distribution of empirical processes in the metric space  $\ell^\infty(\overline{\mathbb{R}}^d)$  with the sup-norm in the sense of Hoffmann–Jørgensen (see, e.g., [Dudley, 1999](#), p. 94). The convergence in distribution of the standard residual empirical process

$$T_{n,\theta_n}(u, v) := n^{1/2}[K_{n,\theta_n}(u, v) - K_{\theta_0}(u, v)]$$

has been obtained for a variety of models under fairly weak regularity conditions on  $g$  and the underlying serial dependence structure, but assuming that  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  are *iid*, see, e.g., [Koul \(2002\)](#) monograph. Once the limiting distribution of  $T_{n,\theta_n}$  is established, the limiting distribution of  $S_{n,\theta_n}$  follows straightforwardly. The *iid* innovations assumption rules out important situations where conditional moments are not expected to be constant, e.g., models for financial data with conditional heteroskedasticity or conditional heterokurtosis of unknown form. See, for instance, [Harvey and Siddique \(1999, 2000\)](#).

The weak convergence of  $S_{n,\theta_n}$  in  $\ell^\infty(\overline{\mathbb{R}}^d)$  with *iid* innovations seems difficult to be extended to the case where  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  exhibit an unknown serial dependence structure. Such extension is one of the main contributions of this paper. To this end, we need first, an asymptotic representation of  $S_{n,\theta_n}$  in terms of  $S_{n,\theta_0}$  when  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  are not independent and second, we need a FCLT for  $S_{n,\theta_0}$ . We take advantage of the fact that, under  $H_0$ ,  $S_{n,\theta_0}(u, v)$  is a martingale for each  $(u, v) \in \mathbb{R}^d$ , which allows to apply the weak convergence results of [Levental \(1989\)](#), [Bae and Levental \(1995\)](#) and [Nishiyama \(2000\)](#).

It is worth noticing that the conditional distribution is symmetric if and only if the conditional characteristic function is real-valued, i.e., it does not have imaginary part. This fact has been exploited by [Feuerverger and Mureika \(1977\)](#), [Csörgö and Heathcote \(1982,](#)

1987), Koutrovelis (1985), Ghosh and Ruymgaart (1992) and Heathcote et al. (1995) amongst others, to construct consistent tests for unconditional symmetry. More formally, the null hypothesis can equivalently expressed as

$$\int_{-\infty}^{\infty} \sin(\bar{v}v)F_{\theta_0}(d\bar{v}|u) = 0 \quad \text{a.s. for some } \theta_0 \in \Theta$$

or, equivalently,

$$\int_{-\infty}^{\infty} \sin(\bar{v}v)K_{\theta_0}(u, d\bar{v}) = 0 \quad \text{a.s. for some } \theta_0 \in \Theta.$$

This suggests to use as test statistics continuous functionals of the empirical process

$$\begin{aligned} R_{n,\theta_n}(u, v) &= \sqrt{n} \int_{-\infty}^{\infty} \sin(\bar{v}v)K_{n,\theta_n}(u, d\bar{v}) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \sin(\varepsilon_t(\theta_n)v)1(I_t \leq u). \end{aligned}$$

Interestingly, under  $H_0$ ,

$$R_{n,\theta_n}(u, v) = \frac{1}{2} \int_{-\infty}^{\infty} \sin(\bar{v}v)S_{n,\theta_n}(u, d\bar{v}).$$

Therefore, the limiting distribution of test statistics  $\varphi(R_{n,\theta_n})$ , based on a suitable continuous functional  $\varphi : \ell^\infty(\overline{\mathbb{R}}^d) \mapsto \mathbb{R}$ , is obtained as a straightforward consequence of the CMT, once it has been derived the limiting distribution of  $S_{n,\theta_n}$ .

The rest of the paper is organized as follows. In the next section we provide the asymptotic distribution of the test statistics. The asymptotic power of the tests are studied in Section 3. We suggest and justify, in Section 4, to implement the tests with the assistance of a bootstrap method. The practical performance of the tests is illustrated by means of a Monte Carlo experiment in Section 5. Mathematical proofs and some instrumental results are confined to an Appendix, at the end of the paper.

## 2. Limiting distribution of test statistics under the null hypothesis

The assumptions on the underlying serial dependence structure are summarized by the following regularity conditions,

(A1)  $\{Y_t, I_t\}_{t \in \mathbb{Z}}$  is a strictly stationary and ergodic process.

(A2) The joint cdf of  $(\varepsilon_1, I_1)$ ,  $K$ , is uniformly continuous on  $\overline{\mathbb{R}}^d$  and  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  is a Markov's process, in the sense that, under  $H_0$ ,

$$E(1_{\{\varepsilon_t \leq \cdot\}} | \mathcal{F}_t) = F_{\theta_0}(\cdot | I_t) \quad \text{a.s. for each } t \in \mathbb{N},$$

where  $\mathcal{F}_t = \sigma(I'_t, I'_{t-1}, \dots)$  is the  $\sigma$ -field generated by the information set obtained up to time  $t$ .

(A3) The family of distribution functions  $\{F_{\theta_0}(\cdot | u) : u \in \overline{\mathbb{R}}^p\}$  has Lebesgue densities  $\{f_{\theta_0}(\cdot | u) : u \in \overline{\mathbb{R}}^p\}$  that are uniformly bounded

$$\sup_{u \in \overline{\mathbb{R}}^p, v \in \mathbb{R}} |f_{\theta_0}(v | u)| < \infty.$$

Furthermore, there exists a constant  $C$  such that

$$\sup_{u \in \overline{\mathbb{R}}^p, v_1, v_2 \in \mathbb{R}} \frac{|F_{\theta_0}(v_1|u) - F_{\theta_0}(v_2|u)|}{|F_{\theta_0}(v_1) - F_{\theta_0}(v_2)|} \leq C,$$

where  $F_{\theta_0}$  is the marginal distribution of the errors  $\varepsilon_t(\theta_0)$ ,  $t \in \mathbb{Z}$ .

A FCLT for  $S_{n,\theta_0}$  can be obtained easily from a FCLT for the standard empirical process  $T_{n,\theta_0}$  and an application of the CMT. Though FCLTs for  $T_{n,\theta_0}$  are available in generous supply, assuming that  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  are *iid* or satisfy some short of mixing condition, it seems hard to prove it under general serial dependence assumptions like (A1) and (A2). This is why we prove directly the weak convergence of  $S_{n,\theta_0}$ , rather than  $T_{n,\theta_0}$ , taking advantage of the fact that, under  $H_0$ , (A1) and (A2),

$$S_{n,\theta_0}(u, v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \omega_i(v) \cdot 1(I_i \leq u),$$

with  $\omega_t(v) = 1\{\varepsilon_t \leq v\} - 1\{-\varepsilon_t \leq v\}$  being a martingale differenced sequence with respect to the filtration  $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$  for each  $v \in \mathbb{R}$ , i.e.,  $E(\omega_t(v)|\mathcal{F}_t) = 0 \forall v \in \mathbb{R}$ . Therefore, applying a standard CLT for martingales, see e.g., [Hall and Heyde \(1980\)](#), the finite-dimensional distributions of  $S_{n,\theta_0}$  converge to those of  $S_\infty$ , a Gaussian process with continuous sample paths and covariance function,

$$E[S_\infty(u_1, v_1)S_\infty(u_2, v_2)] = E[\omega_1(v_1)\omega_1(v_2)1(I_1 \leq u_1 \wedge u_2)].$$

Next theorem extends the finite-dimensional convergence of  $S_{n,\theta_0}$  to weak convergence in  $\ell^\infty(\overline{\mathbb{R}}^d)$ , which is a direct consequence of Theorem A.1 in the Appendix.

**Theorem 1.** *Under  $H_0$ , if (A1)–(A3) hold, then,*

$$S_{n,\theta_0} \text{ converges in distribution to } S_\infty \text{ in } \ell^\infty(\overline{\mathbb{R}}^d).$$

The limiting distribution of  $S_{n,\theta_n}$  is obtained from Theorem 1 and an asymptotic expansion of  $S_{n,\theta_n}$  in terms of  $S_{n,\theta_0}$ . Such expansion requires the following regularity conditions on  $g$ .

(A4) There exists a vector of functions  $\dot{g} : \mathbb{R}^p \times \Theta \rightarrow \mathbb{R}^q$  such that  $\dot{g}(I_t, \theta)$  is  $\mathcal{F}_t$ -measurable for each  $t \in \mathbb{Z}$ ,  $E[|\dot{g}(I_1, \theta_0)|^2] < \infty$ , and satisfies, for all  $\alpha, k$  and  $\theta_1 \in \Theta$ ,

$$\limsup_{n \rightarrow \infty} \Pr \left\{ \sup_{|\theta_1 - \theta_2| \leq kn^{-1/2}} \frac{|g(I_1, \theta_1) - g(I_1, \theta_2) - (\theta_1 - \theta_2)' \dot{g}(I_1, \theta_2)|}{|\theta_1 - \theta_2|} > \alpha \right\} = 0.$$

(A5)  $r_{\theta_0}(u, v) = E[\dot{g}(I_1, \theta_0)f_{\theta_0}(v|u)1(I_1 \leq u)]$  is an absolutely continuous function in all its arguments.

These assumptions are standard when dealing with empirical process of residuals in dynamic models, see, e.g., [Koul \(1996\)](#). Under these regularity conditions, we can obtain the following asymptotic expansion of  $S_{n,\theta_n}$ .

**Theorem 2.** Under  $H_0$ , if (A1)–(A5) hold, then, uniformly in  $(u, v) \in \overline{\mathbb{R}}^d$ ,

$$S_{n,\theta_n}(u, v) = S_{n,\theta_0}(u, v) - 2r_{\theta_0}(u, v)' \sqrt{n}(\theta_n - \theta_0) + o_p(1).$$

The asymptotic distribution of  $S_{n,\theta_n}$  is a straightforward consequence of Theorems 1 and 2, once it is assumed that  $\theta_n$  satisfies the following asymptotic representation.

$$(A6) \quad \sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n l_{\theta_0}(\varepsilon_t, I_t) + o_p(1),$$

where  $l_{\theta_0}(\cdot)$  is such that  $E[l_{\theta_0}(\varepsilon_1, I_1)] = 0$  and  $L(\theta_0) = E[l_{\theta_0}(\varepsilon_1, I_1)l_{\theta_0}'(\varepsilon_1, I_1)]$  exists and is positive definite.

The estimator  $\theta_n$  can be a  $Z$ -estimator, the sample analog of  $\theta_0$  defined in (2), i.e.,

$$\sum_{t=1}^n \psi(Y_t, I_t, \theta_n) = 0, \quad (6)$$

or a  $M$ -estimator, the sample analog of  $\theta_0$  defined in (3), i.e.,

$$\theta_n = \arg \min_{\theta \in \Theta} \sum_{t=1}^n \rho(\varepsilon_t(\theta)). \quad (7)$$

The expansion for  $\theta_n$  is satisfied, both under the null and the alternative hypothesis, for a variety of estimators. For instance, if  $\theta_n$  is the nonlinear least-squares (NLS) estimator, i.e.,  $\rho(v) = v^2$ ,  $l_{\theta_0}(v, u) = E[\nabla_{\theta} g(I_1, \theta_0) \nabla_{\theta} g(I_1, \theta_0)']^{-1} \nabla_{\theta} g(u, \theta_0) \cdot v$ . If  $\theta_n$  is the nonlinear least absolute deviation (NLAD) estimator, i.e.,  $\rho(v) = |v|$ ,  $l_{\theta_0}(v, u) = E[\nabla_{\theta} g(I_1, \theta_0) \nabla_{\theta} g(I_1, \theta_0)']^{-1} \nabla_{\theta} g(u, \theta_0) \cdot \text{sign}(v)$ . The parameter  $\theta_0$  defined in (2) or (3) is consistently estimated by  $\theta_n$  in (6) or (7), even when  $H_0$  is not satisfied.

**Theorem 3.** Under  $H_0$ , if (A1)–(A6) hold, then,

$$S_{n,\theta_n} \text{ converges in distribution to } \hat{S}_{\infty} \text{ in } \ell^{\infty}(\overline{\mathbb{R}}^d),$$

with

$$\hat{S}_{\infty}(u, v) \stackrel{d}{=} S_{\infty}(u, v) - r_{\theta_0}(u, v)' \int_{\mathbb{R}^d} l_{\theta_0}(\bar{v}, \bar{u}) S_{\infty}(d\bar{u}, d\bar{v}).$$

The limiting distribution of  $R_{n,\theta_n}$ , as well as of test statistics based on  $S_{n,\theta_n}$  or  $R_{n,\theta_n}$ , is an immediate consequence of Theorem 3 and the CMT, as stated in the following corollary.

**Corollary 1.** Under the conditions of Theorem 3,

$$R_{n,\theta_n} \rightarrow_d \hat{R}_{\infty} \text{ in } \ell^{\infty}(\overline{\mathbb{R}}^d),$$

with

$$\hat{R}_{\infty}(u, v) \stackrel{d}{=} \int_{-\infty}^{\infty} \sin(\bar{v}v) \hat{S}_{\infty}(u, d\bar{v}).$$

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

$$\varphi(S_{n,\theta_n}) \rightarrow_d \varphi(\hat{S}_{\infty}).$$

In the empirical processes literature, the most popular functionals are the sup-norm,  $\varphi(g) = \sup_{(u,v) \in \overline{\mathbb{R}}^d} |g(u,v)|$ , which provides the Kolmogorov–Smirnov-type statistics

$$KS_n = \sup_{(u,v) \in \overline{\mathbb{R}}^d} |S_{n,\theta_n}(u,v)| \quad (8)$$

and

$$\overline{KS}_n = \sup_{(u,v) \in \Pi_c} |R_{n,\theta_n}(u,v)|,$$

where  $\Pi_c \subset \mathbb{R}^d$  is a compact subset containing the origin. Under the conditions in Corollary 1,

$$KS_n \rightarrow_d \sup_{(u,v) \in \overline{\mathbb{R}}^d} |\hat{S}_\infty(u,v)|$$

and

$$\overline{KS}_n \rightarrow_d \sup_{(u,v) \in \Pi_c} |\hat{R}_\infty(u,v)|.$$

Other popular choice is the  $L_2$ -distance with respect to a suitable measure,  $\Phi$  say,  $\varphi(g) = \int_{\mathbb{R}^2} g(u,v)^2 \Phi(du, dv)$ . The choice of the integrating function  $\Phi$  has implications on the power performance of the Cramér–von Mises-type statistics. In the standard goodness-of-fit tests,  $\Phi$  is the distribution function under the null hypothesis. Since, it is not known, it is reasonable to use the empirical joint distribution. In our case, this choice yields the Cramér–von Mises-type statistic

$$\begin{aligned} CvM_n &= \int_{\mathbb{R}^d} S_{n,\theta_n}^2(u,v) K_{n,\theta_n}(du, dv) \\ &= \frac{1}{n} \sum_{t=1}^n S_{n,\theta_n}^2(\varepsilon_t(\theta_n), I_t). \end{aligned} \quad (9)$$

A Cramér–von Mises-type statistics based on  $R_{n,\theta_n}$  is

$$\overline{CvM}_n = \int_{\mathbb{R}^d} R_{n,\theta_n}^2(u,v) \Phi_n(du, dv).$$

Epps and Pulley (1983) have discussed the choice of  $\Phi_n$  in the context of goodness-of-fit testing based on the empirical characteristic function. Following the arguments of these authors, we consider a weighting function of the form  $\Phi_n(u,v) = F_{n,I}(u) \cdot \Phi_1(v)$ , where  $F_{n,I}$  is the empirical distribution function of  $\{I_t\}_{t=1}^n$  and  $\Phi_1$  is the standard normal distribution function. With such a choice,

$$\overline{CvM}_n = n^{-2} \sum_{r=1}^n \sum_{s=1}^n a_{rs} b_{rs}, \quad (10)$$

where

$$a_{rs} = \sum_{t=1}^n 1(I_r \leq I_t) 1(I_s \leq I_t)$$



and

$$b_{rs} = 0.5 \exp(-0.5(\varepsilon_r(\theta_n) - \varepsilon_s(\theta_n))^2) - 0.5 \exp(-0.5(\varepsilon_r(\theta_n) + \varepsilon_s(\theta_n))^2).$$

Corollary 1 and an application of Lemma 3.1 in [Chang \(1990\)](#) imply that as long as the Glivenko-Cantelli result

$$\sup_{x \in \mathbb{R}^d} |K_{n,\theta_n}(x) - K_{\theta_0}(x)| \rightarrow 0 \quad \text{a.s.}$$

is satisfied,

$$CvM_{n \rightarrow d} \int_{\mathbb{R}^d} \hat{S}_{\infty}^2(u, v) K_{\theta_0}(du, dv)$$

and

$$\overline{CvM}_{n \rightarrow d} \int_{\mathbb{R}^d} \hat{R}_{\infty}^2(u, v) \Phi(du, dv),$$

where  $\Phi(u, v) = F_I(u) \cdot \Phi_1(v)$  and  $F_I$  is the cdf of  $I_1$ .

### 3. Asymptotic power

Let us consider contiguous asymmetric alternatives of the form

$$H_{A_n} : f_{\theta_0}^{(n)}(v|u) = f_{\theta_0}(v|u) \left[ 1 + \frac{1}{\sqrt{n}} h_{n,\theta_0}(u, v) \right] \quad \text{a.s. for some } \theta_0 \in \Theta, \quad (11)$$

where  $f_{\theta_0}$  is a symmetric density, i.e.,  $f_{\theta_0}(v|u) = f_{\theta_0}(-v|u)$  for each  $(u, v) \in \mathbb{R}^d$ , and  $h_{n,\theta_0} : \mathbb{R}^d \rightarrow \mathbb{R}$  is a function such that for each  $n \geq 1$  and each  $(u, v) \in \mathbb{R}^d$ ,

$$\frac{1}{\sqrt{n}} h_{n,\theta_0}(u, v) \geq -1, \quad h_{n,\theta_0}(u, v) \neq h_{n,\theta_0}(u, -v), \quad \sup_{u \in \mathbb{R}^d} \left| \int_{-\infty}^{\infty} h_{n,\theta_0}(u, v) f_{\theta_0}(dv|u) \right| = 0,$$

and

$$h_{n,\theta_0} \rightarrow h_{\theta_0} \quad \text{in } L_2(K_{\theta_0}) \text{ for all } \theta_0 \in \Theta,$$

where  $L_2(K_{\theta_0})$  is the Hilbert space of all  $K_{\theta_0}$ -square integrable real-valued functions on  $\mathbb{R}^d$ . These contiguous alternatives have been considered in the classical goodness-of-fit testing problem of parametric distribution functions (see [Neuhaus, 1973, 1976](#)).

Notice that, under  $H_{A_n}$ ,

$$\begin{aligned} \mathbb{E}[K_{n,\theta_0}(u, v)] &= \int_{-\infty}^v \mathbb{E}[f_{\theta_0}^{(n)}(\bar{v}|I_1) \mathbf{1}(I_1 \leq u)] d\bar{v} \\ &= \int_{-\infty}^v \mathbb{E}[f_{\theta_0}(\bar{v}|I_1) \mathbf{1}(I_1 \leq u)] d\bar{v} + \frac{1}{\sqrt{n}} \Delta_{\theta_0}^1(u, v) \end{aligned}$$

with

$$\Delta_{\theta_0}^1(u, v) = \int_{-\infty}^v \mathbb{E}[f_{\theta_0}(\bar{v}|I_1) h_{\theta_0}(I_1, \bar{v}) \mathbf{1}(I_1 \leq u)] d\bar{v}.$$

Hence, the expansion of  $S_{n,\theta_0}$  under  $H_{A_n}$ , now becomes (uniformly in  $(u, v) \in \overline{\mathbb{R}^d}$ )

$$S_{n,\theta_n}(u, v) = S_{n,\theta_0}(u, v) - 2r_{\theta_0}(u, v)\sqrt{n}(\theta_n - \theta_0) \\ + \Delta_{\theta_0}^1(u, v) - \Delta_{\theta_0}^1(u, \infty) + \Delta_{\theta_0}^1(u, -v) + o_P(1).$$

Under contiguous alternatives  $H_{A_n}$ , the expansion (A6) for  $\theta_n$  still continues to hold, but  $I_{\theta_0}$  terms are not centered anymore. See [Behnen and Neuhaus \(1975\)](#). This results in the additional shift,

$$\Delta_{\theta_0}^2(u, v) = 2r_{\theta_0}(u, v)' \int_{-\infty}^{\infty} \mathbb{E}[h_{\theta}(I_1, \bar{v})l_{\theta}(I_1, \bar{v})f_{\theta}(\bar{v}|I_1)] d\bar{v}.$$

Put

$$\Delta_{\theta_0}(u, v) = \Delta_{\theta_0}^1(u, v) + \Delta_{\theta_0}^1(u, -v) - \Delta_{\theta_0}^1(\infty, v) - \Delta_{\theta_0}^2(u, v).$$

The following theorem provides the asymptotic distribution of  $S_{n,\theta_n}$  under  $H_{A_n}$ .

**Theorem 4.** *Under the alternative hypothesis  $H_A$  and (A1)–(A6),*

$$S_{n,\theta_n} - \Delta_{\theta_0} \rightarrow_d \hat{S}_{\infty}.$$

This theorem shows that the limiting distribution of  $S_{n,\theta_n}$  is shifted under  $H_{A_n}$ . This fact guarantees that the corresponding test statistics based on suitable continuous functionals are able to detect contiguous alternatives  $H_{A_n}$ . Though the additional term  $\Delta_{\theta_0}^2$ , it is possible that, though parameters may be known, their estimation increases the power of the test.

Notice that  $H_{A_n}$  nests the mixtures considered by [Bai and Ng \(2001\)](#),

$$\dot{H}_{A_n} : f_{\theta_0}^{(n)}(v|u) = \left(1 - \frac{\delta}{\sqrt{n}}\right) f_{\theta_0}(v|u) + \frac{\delta}{\sqrt{n}} a_{\theta_0}(v|u) \quad \text{a.s. for some } \theta_0 \in \Theta,$$

where  $a_{\theta_0}$  is an asymmetric conditional density and  $\delta \in \mathbb{R}$ , taking

$$h_{n,\theta_0}(u, v) = \delta \left[ \frac{a_{\theta_0}(v|u)}{f_{\theta_0}(v|u)} - 1 \right].$$

However, other local alternatives are possible. For instance, taking

$$h_{n,\theta_0}(u, v) = v \cdot m_{n,\theta_0}(u) \cdot 1(|v \cdot m_{n,\theta_0}(u)| \leq \sqrt{n}),$$

produces conditional heteroskewness, i.e., for any function  $m_{n,\theta_0}$  satisfying that  $m_{n,\theta_0} \rightarrow m_{\theta_0}$  in  $L_2(F_I)$ ,

$$\mathbb{E}(\varepsilon_t^3 | I_t = u) = \frac{1}{\sqrt{n}} m_{n,\theta_0}(u) \int_{\sqrt{n}/|m_{n,\theta_0}(u)|}^{\sqrt{n}/|m_{n,\theta_0}(u)|} v^4 f_{\theta_0}(v|u) dv.$$

#### 4. Bootstrap approximation

Bootstrap assisted tests have been extensively used in the specification testing literature when the limiting distribution of the test statistics is not pivotal, see, e.g., [Stute et al. \(1998\)](#), [Delgado and González-Manteiga \(2001\)](#) and [Li et al. \(2003\)](#). The wild bootstrap (WB) introduced in [Wu \(1986\)](#) and [Liu \(1988\)](#) appears to be relevant for respecting the underlying relation between innovations and explanatory variables. We adapt the WB

approach to test for conditional symmetry with time series data, extending the method proposed by [Neumeyer and Dette \(2003\)](#) for the *iid* linear regression case. Other proposal, only valid for linear processes, is that of [Psaradakis \(2003\)](#) who considered a sieve bootstrap procedure for testing unconditional symmetry based on residuals resampled from an autoregressive approximation of the given process. In a related but different problem, [Corradi and Swanson \(2006\)](#) use the block bootstrap for Kolmogorov-type conditional distribution tests under dynamic misspecification and parameter estimation error.

Here, we approximate the asymptotic null distribution of the test statistics  $S_{n,\theta_n}$  by the “bootstrap distribution” of

$$S_{n,\theta_n}^*(u, v) = n^{-1/2} \sum_{t=1}^n w_{t,\theta_n}^*(v) 1(I_t \leq u) \quad (u, v) \in \overline{\mathbb{R}}^d,$$

where  $w_{t,\theta_n}^*(v) = \{1(\varepsilon_t^*(\theta_n^*) \leq v) - 1(-\varepsilon_t^*(\theta_n^*) \leq v)\}$  and the sequence  $\{\varepsilon_t^*(\theta_n^*)\}_{t=1}^n$  are the WB residuals obtained from the following algorithm:

- (1) Estimate the original model and obtain the residuals  $\varepsilon_t(\theta_n)$  for  $t = 1, \dots, n$ .
- (2) Generate WB residuals according to  $\varepsilon_t^*(\theta_n) = \varepsilon_t(\theta_n)V_t$  for  $t = 1, \dots, n$ , where  $\{V_t\}_{t=1}^n$  is a sequence of *iid* Bernoulli variates with  $\Pr(V_1 = 1) = \Pr(V_1 = -1) = 0.5$  (Rademacher random variables) and independent of the sample  $\{Y_t, I_t\}_{t=1}^n$ .
- (3) Given  $\theta_n$  and  $\varepsilon_t^*(\theta_n)$ , generate bootstrap observations for the dependent variable  $Y_t^*$  according to

$$Y_t^* = g(I_t, \theta_n) + \varepsilon_t^*(\theta_n) \quad \text{for } t = 1, \dots, n.$$

- (4) Compute the bootstrap analog of  $\theta_n$ ,  $\theta_n^*$  say, using the bootstrap observations  $\{Y_t^*, I_t\}_{t=1}^n$  and compute the residuals  $\hat{\varepsilon}_t^*(\theta_n^*) = Y_t^* - g(I_t, \theta_n^*)$  for  $t = 1, \dots, n$ .

The unknown limiting null distribution of  $\varphi(S_{n,\theta_n})$ , i.e., the distribution of  $\varphi(\hat{S}_\infty)$ , is approximated by the bootstrap distribution of  $\varphi(S_{n,\theta_n}^*)$ . That is, the bootstrap distribution

$$F_{\varphi(S_{n,\theta_n}^*)}^*(x) = \Pr(\varphi(S_{n,\theta_n}^*) \leq x | \{Y_t, I_t\}_{t=1}^n)$$

estimates the asymptotic null distribution function

$$F_{\varphi(\hat{S}_\infty)}(x) = \Pr(\varphi(\hat{S}_\infty) \leq x).$$

Thus,  $H_0$  will be rejected at the  $100\alpha\%$  of significance when  $\varphi(S_{n,\theta_n}) \geq c_{n,\alpha}^*$ , where  $F_{\varphi(S_{n,\theta_n}^*)}^*(c_{n,\alpha}^*) = 1 - \alpha$ . Also, we can use the bootstrap  $p$ -values,  $p_n^*$  say, rejecting  $H_0$  when  $p_n^* < \alpha$ , where  $p_n^* = \Pr(\varphi(S_{n,\theta_n}^*) \geq \varphi(S_{n,\theta_n}) | \{Y_t, I_t\}_{t=1}^n)$ . The bootstrap assisted test is valid if  $F_{\varphi(S_{n,\theta_n}^*)}^*$  is a consistent estimator of  $F_{\varphi(\hat{S}_\infty)}$  at each continuity point of  $F_{\varphi(\hat{S}_\infty)}$ . When consistency is a.s., it is expressed as  $\varphi(S_{n,\theta_n}^*) \rightarrow_d \varphi(\hat{S}_\infty)$  a.s. See [Giné and Zinn \(1990\)](#) or [van der Vaart and Wellner \(1996\)](#) for discussion.

Likewise, we can construct tests based on

$$R_{n,\theta_n}^*(u, v) = \frac{1}{2} \int_{-\infty}^{\infty} \sin(\bar{v}v) S_{n,\theta_n}^*(u, d\bar{v}).$$

In order to show that the bootstrap assisted tests are valid, we need to assume that the bootstrap analogs of  $\theta_n$  satisfy an asymptotic expansion like (A6) in the bootstrap world. Remark that we say that the bootstrap statistic  $\eta_n^*$  converges in probability a.s. to  $\eta_n$  if for all  $\delta > 0$ ,  $\Pr(|\eta_n^* - \eta_n| \geq \delta | \{Y_t, I_t\}_{t=1}^n) \rightarrow 0$  a.s., which is expressed as  $\eta_n^* = \eta_n + o_P(1)$  a.s. Also, bootstrap expectations are denoted by  $E^*(\eta_n^*) = E(\eta_n^* | \{Y_t, I_t\}_{t=1}^n)$ .

(A7(a)) There exists a unique  $\theta_1$  such that under both, the null and the alternative hypotheses,  $|\theta_n - \theta_1| = o_P(1)$ . The estimator  $\theta_n^*$  satisfies the following asymptotic expansion:

$$\sqrt{n}(\theta_n^* - \theta_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n l_{\theta_n}^*(Y_t^*, I_t) + o_P(1) \quad \text{a.s.},$$

where the function  $l_{\theta_n}^*$  is such that

(A7(b))  $E^*[l_{\theta_n}^*(Y_t^*, I_t)] = 0$ , a.s.

(A7(c))  $L(\theta_n) = E^*[l_{\theta_n}^*(Y_t^*, I_t)l_{\theta_n}^{*'}(Y_t^*, I_t)]$  exists and is positive definite (a.s.) with  $L(\theta_n) \rightarrow L(\theta_1)$  a.s.

(A7(d)) For all  $(u_1, v_1), (u_2, v_2) \in \overline{\mathbb{R}}^d$ ,

$$\begin{aligned} n^{-1} \sum_{t=1}^n w_{t, \theta_n}(v_1) w_{t, \theta_n}(v_2) 1(I_t \leq u_1) 1(I_t \leq u_2) \\ \xrightarrow{\text{as}} E[w_{t, \theta_1}(v_1) w_{t, \theta_1}(v_2) 1(I_t \leq u_1) 1(I_t \leq u_2)]. \end{aligned}$$

(A7(e)) Uniformly in  $(u, v) \in \mathbb{R}^d$ ,

$$n^{-1} \sum_{t=1}^n \sum_{s=1}^n (E^*[w_{t, \theta_n}^*(v) 1(I_t \leq u) l_{\theta_n}^*(Y_s^*, I_s)] - w_{t, \theta_1}(v) 1(I_t \leq u) l_{\theta_1}(Y_s, I_s)) \xrightarrow{\text{as}} 0.$$

Henceforth, almost sure convergence of nonmeasurable maps is understood, as usual, as outer almost sure convergence, see van der Vaart and Wellner (1996) for definitions. It is not difficult to show that assumption (A7) is satisfied for  $Z$  and  $M$  estimators under suitable regularity conditions, see, e.g., Koul (2002, Chapter 7). Sufficient conditions for (A7) are easily obtained from results of Wooldridge (1994), White (1994) or Koul (2002). In many cases, the functions  $l_{\theta_0}$  and  $l_{\theta_n}^*$  required in (A6) and (A7) can be expressed as  $l_{\theta_0}(\varepsilon_t, I_t) = \varepsilon_t(\theta_0)k(I_t, \theta_0)$  and  $l_{\theta_n}^*(Y_t^*, I_t) = V_t \varepsilon_t(\theta_n)k(I_t, \theta_n)$ , respectively, for some function  $k(\cdot)$ , see, e.g., the NLS, or, more generally, estimators resulting from a martingale estimating equation (see Heyde, 1990). Then, in those cases, (A7(e)) reduces to the uniform convergence

$$\sup_{(u, v) \in \overline{\mathbb{R}}^d} \left| \frac{1}{n} \sum_{t=1}^n (w_{t, \theta_n}(v) 1(I_t \leq u) \varepsilon_t(\theta_n) k(I_t, \theta_n) - w_{t, \theta_1}(v) 1(I_t \leq u) \varepsilon_t(\theta_1) k(I_t, \theta_1)) \right| \xrightarrow{\text{as}} 0,$$

which is satisfied under some mild conditions on the function  $k(\cdot)$  and (A1)–(A5).

Next theorem justifies the validity of bootstrap assisted tests.

**Theorem 5.** Assume (A1)–(A7), and let  $\varphi : \ell^\infty(\overline{\mathbb{R}^d}) \mapsto \mathbb{R}$  be a continuous functional. Then,

$$\varphi(S_{n,\theta_n}^*) \rightarrow_d \varphi(\widetilde{S}_\infty) \quad a.s.$$

where  $\widetilde{S}_\infty$  is the same Gaussian process as in Theorem 3 but with  $\theta_1$  replacing  $\theta_0$ .

Since the theorem is satisfied under the null and the alternatives, it justifies the consistency of bootstrap assisted tests.

## 5. Monte Carlo

We investigate in this section, by means of a Monte Carlo experiment, the finite sample performance of Cramér–von Mises-type statistics (9) and (10), as well as their one parameter versions

$$CvM_{n,u} = \int_{\mathbb{R}^d} S_{n,\theta_n}^2(\infty, v) K_{n,\theta_n}(\infty, dv)$$

and

$$\overline{CvM}_{n,u} = \int_{\mathbb{R}^d} R_{n,\theta_n}^2(\infty, v) \Phi_1(dv),$$

respectively, with  $\Phi_1(v)$  the cdf of standard normal r.v. These tests are also compared with the asymptotically pivotal tests of [Bai and Ng \(2001\)](#), henceforth BN, based on the martingale transform of  $S_{n,\theta_n}(\infty, \cdot)$  using the conditionally scaled residuals, and using the same smooth estimates of nonparametric functions as BN recommends. The BN's test is denoted by CS.

Bootstrap critical values are approximated by Monte Carlo using 500 replications. We consider two sample sizes,  $n = 50$  and 200. The Monte Carlo experiments are based on 1000 replications. We only report results for the 5% significance level. We have also considered size-corrected critical values, which are not reported since they do not provide any additional information.

First, we consider the case where the conditional center of symmetry does not depend on explanatory variables. That is, we consider  $g(I_t, \theta_0) = \mu$ , for all  $t \in \mathbb{Z}$ , say, where  $\mu$  is a constant and  $I_t = Y_{t-1}$ . The unknown mean  $\mu$  is estimated by the sample mean and the residuals are  $\varepsilon_t(\overline{Y}_n) = Y_t - \overline{Y}_n$ . As in BN, the demeaned data are standardized by the sample standard deviation.

We investigate the size accuracy of the test in the context of the following designs:

- (S1)  $Y_t \sim iid \ N(0, 1)$ .
- (S2)  $Y_t \sim iid \ t_5$ .
- (S3)  $Y_t \sim iid \ e_1 1(Z \leq 0.5) + e_2 1(Z > 0.5)$  with  $e_1 \sim iid \ N(-1, 1)$ ,  $e_2 \sim iid \ N(1, 1)$  and  $Z \sim iid \ U(0, 1)$  mutually independent.
- (S4–S7)  $Y_t$  are *iid* according to a symmetric  $\lambda$  distributions with

$$F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}] / \lambda_2, \quad 0 < u < 1,$$

with the  $\lambda$  values taken from [Randles et al. \(1980\)](#),

$$(S4) \ \lambda_1 = 0, \ \lambda_2 = 0.19754, \ \lambda_3 = 0.134915 \text{ and } \lambda_4 = 0.134915.$$

$$(S5) \ \lambda_1 = 0, \ \lambda_2 = -1, \ \lambda_3 = -0.08 \text{ and } \lambda_4 = -0.08.$$

(S6)  $\lambda_1 = 0$ ,  $\lambda_2 = -0.397912$ ,  $\lambda_3 = -0.16$  and  $\lambda_4 = -0.16$ .

(S7)  $\lambda_1 = 0$ ,  $\lambda_2 = -1$ ,  $\lambda_3 = -0.24$  and  $\lambda_4 = -0.24$ .

Table 1 reports the percentage of rejections for models (S1)–(S7).

All the bootstrap tests exhibit good size accuracy. The asymptotic BN test also behaves fairly well under  $H_0$  but for the design (S5).

In order to study the power in finite samples, we consider the following designs:

(A1)  $Y_t \sim iid \exp(N(0, 1))$ .

(A2)  $Y_t \sim iid \chi_{(2)}^2$ .

(A3)  $Y_t \sim iid -\ln(U(0, 1))$ .

(A4)–(A8)  $Y_t$  are *iid* according to asymmetric  $\lambda$  distributions (see (S4)–(S7)) with

(A4)  $\lambda_1 = 0$ ,  $\lambda_2 = 1$ ,  $\lambda_3 = 1.4$  and  $\lambda_4 = 0.25$ .

(A5)  $\lambda_1 = 0$ ,  $\lambda_2 = -1$ ,  $\lambda_3 = -0.0075$  and  $\lambda_4 = -0.03$ .

(A6)  $\lambda_1 = 0$ ,  $\lambda_2 = -1$ ,  $\lambda_3 = -0.1$  and  $\lambda_4 = -0.18$ .

(A7)  $\lambda_1 = 0$ ,  $\lambda_2 = -1$ ,  $\lambda_3 = -0.001$  and  $\lambda_4 = -0.13$ .

(A8)  $\lambda_1 = 0$ ,  $\lambda_2 = -1$ ,  $\lambda_3 = -0.0001$  and  $\lambda_4 = -0.17$ .

(A9)  $Y_t = X_t - X_{t-1}$  with  $X_t \sim iid \chi_{(2)}^2$ .

(A10)  $Y_t = X_t - X_{t-1}$  with  $X_t \sim iid -\ln(U(0, 1))$ .

The alternatives (A1)–(A8) were considered by BN. Under the alternatives (A9) and (A10)  $\dot{H}_0$  in (4) (unconditional symmetry) is satisfied, though  $H_0$  does not hold (Table 2).

As expected, all the tests are able to detect alternatives (A1)–(A8), but the “marginal” tests  $CvM_{n,u}$ ,  $\overline{CvM}_{n,u}$  and  $CS$ , which are consistent for testing  $\dot{H}_0$ , have trivial power for testing  $H_0$  in the direction of alternatives (A9) and (A10). However, the bootstrap tests  $CvM_n$  and  $\overline{CvM}_n$  also exhibit good power in the direction of (A9) and (A10). It is worth mentioning that the test  $CvM_n$  performs better than  $\overline{CvM}_n$  for alternatives (A1), (A9) and (A10), whereas  $\overline{CvM}_n$  rejects more in the direction of alternatives (A2)–(A8). This behavior may be explained by the fact that alternatives (A1), (A9) and (A10) are “low-frequency” alternatives, which are well detected by standard empirical process based tests, whereas tests based on the empirical characteristic function are designed for detecting “high-frequency” alternatives.

Table 1  
Empirical size at 5% of significance level

	$n = 50$					$n = 200$				
	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	$CS$	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	$CS$
(S1)	5.2	5.6	3.8	5.6	2.9	6.1	6.4	3.7	4.8	4.4
(S2)	5.0	5.0	6.6	6.6	4.5	5.2	6.1	5.6	6.8	4.8
(S3)	5.6	5.6	3.8	4.2	4.6	5.8	5.2	6.4	5.2	4.8
(S4)	6.5	6.1	4.5	5.4	3.3	3.6	4.8	5.6	6.1	3.6
(S5)	5.0	5.7	5.4	21.8	13.4	4.1	5.6	6.6	7.0	3.2
(S6)	5.9	6.1	5.5	6.8	4.8	5.0	4.8	5.9	6.6	7.0
(S7)	7.8	7.2	6.6	6.1	8.6	5.6	5.0	5.9	5.4	10.4

Table 2  
Empirical power at 5% of significance level

	$n = 50$					$n = 200$				
	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	$CS$	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	$CS$
(A1)	91.6	87.9	99.7	99.9	92.1	100.0	100.0	100.0	100.0	100.0
(A2)	78.2	86.2	98.6	99.8	78.9	100.0	100.0	100.0	100.0	100.0
(A3)	77.6	87.0	98.6	99.8	78.2	100.0	100.0	100.0	100.0	100.0
(A4)	23.5	37.0	45.8	57.6	29.8	77.2	90.8	100.0	100.0	92.0
(A5)	40.5	99.9	44.4	100.0	70.4	95.3	100.0	100.0	100.0	97.3
(A6)	24.3	27.7	21.2	29.0	23.3	71.3	75.6	85.0	100.0	72.6
(A7)	84.1	97.8	99.0	100.0	82.9	100.0	100.0	100.0	100.0	100.0
(A8)	87.2	95.4	99.0	100.0	86.6	100.0	100.0	100.0	100.0	100.0
(A9)	85.9	77.0	0.60	0.20	3.30	99.4	99.6	0.00	0.00	3.20
(A10)	47.0	39.2	1.40	0.60	2.20	99.6	99.3	0.00	0.00	2.00

Table 3  
Proportion of rejections at 5% of significance level for the AR(1) model

$\rho = 0.5$	$n = 50$					$n = 200$				
	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	$CS$	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	$CS$
(AU1)	6.6	5.6	4.8	5.6	3.7	5.6	6.6	5.0	4.6	11.4
(AU2)	7.2	6.8	5.7	6.2	3.5	5.2	5.2	5.6	5.0	4.8
(AU3)	4.8	4.3	5.2	5.4	3.3	5.9	5.9	5.0	5.0	5.0
(AU4)	94.6	96.8	97.2	99.2	71.2	100.0	100.0	100.0	100.0	100.0
(AU5)	96.4	99.4	99.0	100.0	78.4	97.8	97.2	100.0	100.0	95.8
(AU6)	96.8	98.4	99.4	100.0	83.1	100.0	100.0	100.0	100.0	99.0

Also, as in BN, we consider the following autoregressive process of order 1, AR(1),  $Y_t = \rho Y_{t-1} + \varepsilon_t$ , where  $\varepsilon_t$  are:

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

(AU2)  $\varepsilon_t \sim iid t_5$ .

(AU3)  $\varepsilon_t \sim iid$  mixture of normals as (S3).

(AU4)  $\varepsilon_t \sim iid \chi_2^2$ .

(AU5)  $\varepsilon_t \sim iid \lambda$ -distribution with parameters  $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.001$  and  $\lambda_4 = -0.13$ .

(AU6)  $\varepsilon_t \sim iid \lambda$ -distribution with parameters  $\lambda_1 = 0, \lambda_2 = -1, \lambda_3 = -0.0001$  and  $\lambda_4 = -0.17$ .

The proportion of rejections for  $\rho = 0.5$  are reported in Table 3.

Table 3 confirms that bootstrap tests perform better than asymptotically pivotal test based on  $CS$ . Surprisingly enough, the conditional tests perform similarly to unconditional tests, even when the latter take into account the information that the errors are independent of the regressors, as is the case with the AR(1) alternatives considered here.

Now, we consider the following conditional heteroskedastic model, GARCH(1,1),  $Y_t = 1 + \varepsilon_t$ ,  $\varepsilon_t = \sigma_t e_t$ ,  $\sigma_t^2 = \phi_0 + \phi_1 \sigma_{t-1}^2 + \phi_2 \varepsilon_{t-1}^2$ ;  $\phi_0 = 2, \phi_1 = 0.5, \phi_2 = 0.3$  (GARCH1);  $\phi_0 = 2, \phi_1 = 0.9, \phi_2 = 0.05$  (GARCH2). The residuals for *CS* are standardized using the parametric estimated standard deviation under models GARCH1 and GARCH2. The distributions for  $\varepsilon_t$  are the same that in the AR(1) model (AU1)–(AU6). We report the results for the GARCH(1,1) models in Tables 4 and 5, respectively.

The empirical size of tests for the GARCH1 and GARCH2 models are quite accurate. Only BN's test statistic *CS* shows large overrejection for the (AU3) model. Our tests statistics  $CvM_n$  and  $\overline{CvM}_n$  have excellent empirical power against alternatives (AU4)–(AU6). The unconditional test  $CvM_{n,u}$  has no power against the alternative (AU4), specially for large sample sizes. This contradictory behavior may be due to the fact that  $CvM_{n,u}$  cannot discriminate between conditional symmetry and other conditional dependence structure.

In models GARCH1 and GARCH2, the BN's test *CS* uses the parametric estimator for the conditional variance. In order to study the sensitivity of this test under a misspecified conditional variance, we consider the same models as before but where the incorrectly estimated variance is that of a ARCH(1) model. The true DGP are the GARCH1 and GARCH2 models considered before. The proportion of rejections for the latter experiment are reported in Table 6.

Table 6 shows that *CS* has large size distortions for model (AU3). Furthermore, under the alternatives (AU5) and (AU6) and for  $n = 50$ , the proportion of rejections has

Table 4  
Proportion of rejections at 5% of significance level for the GARCH1 model

	$n = 50$					$n = 200$				
	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	<i>CS</i>	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	<i>CS</i>
(AU1)	4.9	5.6	4.6	5.6	2.3	8.2	9.7	6.6	5.3	3.9
(AU2)	6.4	5.7	5.6	6.4	7.9	6.0	5.2	5.3	4.3	15.4
(AU3)	3.7	4.8	3.9	3.8	8.5	6.4	6.1	5.0	4.0	27.0
(AU4)	98.5	100.0	14.0	96.2	99.0	100.0	100.0	0.00	99.6	100.0
(AU5)	47.8	56.1	63.0	76.6	79.9	93.2	100.0	100.0	100.0	100.0
(AU6)	67.0	71.6	82.2	91.2	83.9	100.0	100.0	100.0	100.0	100.0

Table 5  
Proportion of rejections at 5% of significance level for the GARCH2 model

	$n = 50$					$n = 200$				
	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	<i>CS</i>	$CvM_n$	$\overline{CvM}_n$	$CvM_{n,u}$	$\overline{CvM}_{n,u}$	<i>CS</i>
(AU1)	5.2	5.9	3.8	4.6	2.2	6.4	7.4	2.9	3.3	2.4
(AU2)	4.3	5.0	6.6	6.0	2.8	3.7	5.0	4.3	4.3	4.4
(AU3)	5.4	5.5	2.9	2.8	3.0	3.4	4.0	5.3	4.6	3.5
(AU4)	96.2	99.8	57.8	100.0	98.1	100.0	100.0	2.33	100.0	99.8
(AU5)	75.7	80.0	90.0	93.8	83.4	100.0	100.0	100.0	100.0	100.0
(AU6)	86.4	89.1	96.2	98.0	89.6	100.0	100.0	100.0	100.0	100.0



Table 6

Proportion of rejections at 5% of significance level for CS under misspecified GARCH models

DGP $n$	GARCH1 50	GARCH1 200	GARCH2 50	GARCH2 200
Estimated model: ARCH(1)				
(AU1)	2.2	2.6	2.2	1.4
(AU2)	7.4	11.8	4.1	4.6
(AU3)	5.6	24.0	2.4	2.9
(AU4)	97.6	99.0	82.6	99.8
(AU5)	43.8	99.4	39.0	98.4
(AU6)	46.4	99.2	45.6	98.0

decreased in the GARCH1 model from 79.9 and 83.9 to 43.8 and 46.4, respectively, and for the GARCH2 model from 83.4 and 89.6 to 39.0 and 45.6, respectively. This behavior has been due to the misspecification of the conditional variance.

This small simulation study suggests that even with relative small sample sizes the bootstrap test proposed in this paper exhibits fairly good size accuracy and power. Our tests are able to detect alternatives where the innovations' marginal distribution is symmetric, which go unnoticed by alternative procedures designed for testing the symmetry of the errors' marginal distribution. Also, unlike alternative procedures, our tests are insensitive to misspecification of higher conditional moments and, in particular, there is no need of assuming any conditional variance model (e.g., GARCH) in the presence of conditional heteroskedasticity.

## Appendix A. Mathematical proofs

First, we shall consider in this section a FCLT for a large class of empirical processes under martingale difference conditions which is essential for providing the different results in the paper. Let for each  $n \geq 1$ ,  $I'_{n,1}, \dots, I'_{n,n}$ , be an array of random vectors in  $\mathbb{R}^p$ ,  $p \in \mathbb{N}$ , and  $\varepsilon_{n,1}, \dots, \varepsilon_{n,n}$ , be an array of real random variables (r.v.'s). Denote by  $(\Omega_n, \mathcal{A}_n, P_n)$ ,  $n \geq 1$ , the probability space in which all the r.v.'s  $\{\varepsilon_{n,t}, I'_{n,t}\}_{t=1}^n$  are defined. Let  $\mathcal{F}_{n,t}$ ,  $0 \leq t \leq n$ , be a double array of sub- $\sigma$ -fields of  $\mathcal{A}_n$  such that  $\mathcal{F}_{n,t} \subset \mathcal{F}_{n,t+1}$ ,  $t = 0, \dots, n-1$  and such that for each  $n \geq 1$  and each  $x \in \Pi \subseteq \overline{\mathbb{R}^d}$ ,  $d \in \mathbb{N}$ ,

$$E[w(\varepsilon_{n,t}, I_{n,t}, x) \mid \mathcal{F}_{n,t}] = 0 \quad \text{a.s., } 1 \leq t \leq n, \quad \forall n \geq 1. \quad (12)$$

Moreover, we shall assume that  $\{w(\varepsilon_{n,t}, I_{n,t}, x), \mathcal{F}_{n,t}, 0 \leq t \leq n\}$  is a square-integrable martingale difference sequence for each  $x \in \Pi \subseteq \overline{\mathbb{R}^d}$ , that is, (12) holds,  $Ew^2(\varepsilon_{n,t}, I_{n,t}, x) < \infty$  and  $w(\varepsilon_{n,t}, I_{n,t}, x)$  is  $\mathcal{F}_{n,t+1}$ -measurable for each  $x \in \Pi \subseteq \overline{\mathbb{R}^d}$  and  $\forall t, 1 \leq t \leq n, \forall n \in \mathbb{N}$ . The main goal of this section is to establish the weak convergence of the empirical process

$$\alpha_{n,w}(x) = n^{-1/2} \sum_{t=1}^n w(\varepsilon_{n,t}, I_{n,t}, x), \quad x \in \Pi.$$

Under mild conditions the empirical process  $\alpha_{n,w}$  can be viewed as a mapping from  $\Omega_n$  to  $\ell^\infty(\Pi)$ , the space of all real-valued functions that are uniformly bounded on any compact subset of  $\Pi \subseteq \overline{\mathbb{R}^d}$ . Let  $\rightarrow_d$  denote weak convergence on compacta in  $\ell^\infty(\Pi)$ , see van der

Vaart and Wellner (1996, Definition 1.3.3, Chapter 1.6). Note that if  $\Pi$  is compact, then  $\rightarrow_d$  reduces to the classical weak convergence concept of Hoffmann–Jørgensen (Dudley, 1999, p. 94). Of course, the sample paths of  $\alpha_{n,w}$  are usually contained in a much smaller space (such as the *cadlag* space  $D(\Pi)$ , the space of real-valued functions on  $\Pi$  with jump discontinuities), but as long as this space is equipped with the *sup*-metric, this is irrelevant for the weak convergence theorem. The weak convergence theorem that we present here is funded on results by Levental (1989), Bae and Levental (1995) and Nishiyama (2000).

An important role in the weak convergence theorem is played by the conditional quadratic variation of the empirical process  $\alpha_{n,w}$  on a finite partition  $\mathcal{B} = \{H_k; 1 \leq k \leq N\}$  of  $\Pi_c$ , where hereafter  $\Pi_c$  is any compact subset of  $\Pi$  which is defined as

$$\alpha_{n,w}(\mathcal{B}) = \max_{1 \leq k \leq N} n^{-1} \sum_{t=1}^n \mathbb{E} \left[ \sup_{x_1, x_2 \in H_k} |w(\varepsilon_{n,t}, I_{n,t}, x_1) - w(\varepsilon_{n,t}, I_{n,t}, x_2)|^2 \middle| \mathcal{F}_{n,t} \right].$$

Then, for the weak convergence theorem we need the following assumptions:

(W1) For each  $n \geq 1$ ,  $\{(\varepsilon_{n,t}, I'_{n,t})' : 1 \leq t \leq n\}$  is a strictly stationary and ergodic process. The sequence  $\{w(\varepsilon_{n,t}, I_{n,t}, x), \mathcal{F}_{n,t}, 0 \leq t \leq n\}$  is a square-integrable martingale difference sequence for each  $x \in \Pi \subseteq \overline{\mathbb{R}^d}$ . Also, there exists a function  $C_w(x_1, x_2)$  on  $\Pi_c \times \Pi_c$  to  $\mathbb{R}$  such that uniformly in  $(x_1, x_2) \in \Pi_c \times \Pi_c$

$$n^{-1} \sum_{t=1}^n w(\varepsilon_{n,t}, I_{n,t}, x_1) w(\varepsilon_{n,t}, I_{n,t}, x_2) = C_w(x_1, x_2) + o_{P_n}(1).$$

(W2) For every compact subset  $\Pi_c$ , the family  $w(\varepsilon_{n,t}, I_{n,t}, x)$  is such that  $\alpha_{n,w}$  is a mapping from  $\Omega_n$  to  $\ell^\infty(\Pi_c)$  and for every  $\varepsilon > 0$  there exists a finite partition  $\mathcal{B}_\varepsilon = \{H_k; 1 \leq k \leq N_\varepsilon\}$  of  $\Pi_c$ , with  $N_\varepsilon$  being the elements of such partition, such that

$$\int_0^\infty \sqrt{\log(N_\varepsilon)} d\varepsilon < \infty \quad (13)$$

and

$$\sup_{\varepsilon \in (0,1) \cap \mathbb{Q}} \frac{\alpha_{n,w}(\mathcal{B}_\varepsilon)}{\varepsilon^2} = O_{P_n}(1). \quad (14)$$

Let  $\alpha_{\infty,w}(\cdot)$  be a Gaussian process with zero mean and covariance function given by  $C_w(x_1, x_2)$ . We are now in position to state the following

**Theorem A.1.** *If Assumptions (W1) and (W2) hold, then it follows that*

$$\alpha_{n,w} \rightarrow_d \alpha_{\infty,w} \quad \text{in } \ell^\infty(\Pi).$$

Now, we shall show that assumption (W2) is satisfied (under (W1) and some mild conditions) for most families  $w$  considered in the literature. First, we start with smooth functions  $w$ , which may arise, for instance, when conditional specifications are made in terms of the conditional characteristic function. Note that under (W1) and for smooth functions  $w$  satisfying

$$|w(\varepsilon_{n,t}, I_{n,t}, x_1) - w(\varepsilon_{n,t}, I_{n,t}, x_2)| \leq K_{n,t} \rho(x_1, x_2),$$

with  $\rho(\cdot, \cdot)$  such that  $(\Pi_c, \rho)$  is a totally bounded metric space and  $K_{n,t}$  is, for each  $n \geq 1$ , a strictly stationary process with  $E[K_{n,t}^2] < \infty$ ,  $\forall t, 1 \leq t \leq n$ , a sufficient condition for (W2) is that

$$\int_0^\infty \sqrt{\log(N(\Pi_c, \rho, \varepsilon))} d\varepsilon < \infty,$$

where  $N(\Pi_c, \rho, \varepsilon)$  is the  $\varepsilon$ -covering number of  $\Pi_c$  with respect to  $\rho$ , i.e., the minimum number of  $\rho$ -balls needed to cover  $\Pi_c$ . This assumption is satisfied, for instance, for  $w(\varepsilon_{n,t}, I_{n,t}, x) = [\sin(v\varepsilon_{n,t}) \sin(u'I_{n,t})]$ ,  $x = (u', v)' \in \Pi_c$ , with  $\Pi_c$  a compact subset of  $\mathbb{R}^d$ . For nonsmooth functions, such as  $w(\varepsilon_{n,t}, I_{n,t}, x) = \{1(\varepsilon_{n,t} \leq v) - 1(-\varepsilon_{n,t} \leq v)\} 1(I_{n,t} \leq u)$ ,  $x = (u', v)' \in \mathbb{R}^d$ , the situation is more involved, see the proof of Theorem 1.

To prove Theorem A.1 we consider two lemmas. The first lemma corresponds to Theorems 1.5.4 and 1.5.6 of van der Vaart and Wellner (1996).

**Lemma A.1.** *Let  $\mathcal{T}$  be a nonempty set. For every  $n \in \mathbb{N}$  let  $(\Omega_n, \mathcal{F}_n, P_n)$  be a probability space, and  $X_n$  be a mapping from  $\Omega_n$  to  $\ell^\infty(\mathcal{T})$ . Consider the following statements:*

- (i)  $X_n$  converges weakly to a tight Borel law;
- (ii) every finite-dimensional marginal of  $X_n$  converges weakly to a (tight) Borel law;
- (iii) for every  $\varepsilon, \eta > 0$  there exists a finite partition  $\mathcal{B} = \{T_k; 1 \leq k \leq N\}$  of  $\mathcal{T}$  such that

$$\limsup_{n \rightarrow \infty} \mathbf{P}^* \left[ \max_{1 \leq k \leq N} \sup_{t, s \in T_k} |X_n(t) - X_n(s)| > \varepsilon \right] \leq \eta.$$

*Then, there is the equivalence (i)  $\iff$  (ii) + (iii). Furthermore, if the marginals of a stochastic process  $X$  have the same laws as the limits in (ii), then there exists a version  $\tilde{X}$  of  $X$  such that  $X_n \rightarrow_d \tilde{X}$  in  $\ell^\infty(\mathcal{T})$ .*

Next, lemma is the so-called Bernstein–Freedman inequality for martingale difference arrays. See Freedman (1975) for the proof.

**Lemma A.2.** *Let  $\{M_{n,t} : 1 \leq t \leq n\}$  be an  $\mathbb{R}$ -valued martingale difference array with respect to the filtration  $\mathcal{F}_{n,t}$ , such that  $|M_{n,t}| < a$ ,  $\forall n, 1 \leq t \leq n$ . Let  $\sigma$  be a bounded stopping time. Then for any  $b > 0$*

$$\mathbf{P} \left( \max_{1 \leq s \leq \sigma} \left| \sum_{t=1}^s M_{n,t} \right| > \varepsilon, \sum_{t=1}^{\sigma} E[M_{n,t}^2 \mid \mathcal{F}_{n,t}] \leq b \right) \leq 2 \exp \left( - \frac{\varepsilon^2}{2(a\varepsilon + b)} \right) \quad \forall \varepsilon > 0.$$

**Proof of Theorem A.1.** Apply the Central Limit Theorem (CLT) for stationary and ergodic martingale difference sequences, cf. Billingsley (1961), to show that the finite-dimensional distributions of  $\alpha_{n,w}$  converge to those of the Gaussian process  $\alpha_{\infty,w}$ . To complete the proof we need to show that (iii) in Lemma A.1 holds. To this end, fix a compact subset  $\Pi_c \subset \Pi$ , and using (W2) we can choose a nested sequence of finite partitions  $\mathcal{P}_q = \{B_{qk}; 1 \leq k \leq N_q\}$  of  $\Pi_c$ , for every  $q \in \mathbb{N}$ ,  $q \geq 1$ , such that

$$\sum_{q=1}^{\infty} 2^{-q} \sqrt{\log N_q} < \infty.$$

Let define  $a_q = 2^{-q}/\sqrt{\log(N_{q+1})}$ . Now, choose an element  $x_{qk}$  for each  $B_{qk}$  and define for every  $x \in \Pi_c$  the events

$$\begin{aligned} \pi_q x &= x_{qk} \\ B_q x &= B_{qk} \end{aligned} \quad \text{if } x \in B_{qk}.$$

To simplify notation define  $M_t^n(x) = n^{-1/2}w(\varepsilon_{n,t}, I_{n,t}, x)$ . Then, by Lemma A.1, see also the proof of Theorem 2.5.6 of [van der Vaart and Wellner \(1996\)](#), it is sufficient to prove that for every  $\varepsilon, \eta > 0$  there exists a  $q_0 \in \mathbb{N}$  such that

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left[ \left\| \sum_{t=1}^n M_t^n(x) - \sum_{t=1}^n M_t^n(\pi_{q_0} x) \right\|_{\Pi_c} > \varepsilon \right] \leq \eta,$$

where  $\|\cdot\|_{\Pi_c}$  denotes the uniform norm on  $\Pi_c$  and where from now on probabilities of nonmeasurable maps are understood as outer probabilities. To this end, fix any  $q_0$  for a while, and let define the quantities for each fixed  $n$  and large  $q \geq q_0$

$$\Delta_t^n(B) = \sup_{x_1, x_2 \in B} |M_t^n(x_1) - M_t^n(x_2)|,$$

and the events

$$C_{t,q-1}^n = 1(\Delta_t^n(B_{q_0} x) \leq a_{q_0}, \dots, \Delta_t^n(B_{q-1} x) \leq a_{q-1}),$$

$$D_{t,q}^n = 1(\Delta_t^n(B_{q_0} x) \leq a_{q_0}, \dots, \Delta_t^n(B_{q-1} x) \leq a_{q-1}, \Delta_t^n(B_q x) > a_q)$$

and

$$D_{t,q_0}^n = 1(\Delta_t^n(B_{q_0} x) > a_{q_0}).$$

Now, similarly to [van der Vaart and Wellner \(1996, p. 131\)](#), we decompose

$$\begin{aligned} M_t^n(x) - M_t^n(\pi_{q_0} x) &= (M_t^n(x) - M_t^n(\pi_{q_0} x))D_{t,q_0}^n + \sum_{q=q_0+1}^{\infty} (M_t^n(x) - M_t^n(\pi_q x))D_{t,q}^n \\ &\quad + \sum_{q=q_0+1}^{\infty} (M_t^n(\pi_q x) - M_t^n(\pi_{q-1} x))C_{t,q}^n. \end{aligned} \quad (15)$$

On the other hand, by (12)

$$\begin{aligned} 0 &= \mathbb{E}[(M_t^n(x) - M_t^n(\pi_{q_0} x))D_{t,q_0}^n \mid \mathcal{F}_{n,t}] + \sum_{q=q_0+1}^{\infty} \mathbb{E}[(M_t^n(x) - M_t^n(\pi_q x))D_{t,q}^n \mid \mathcal{F}_{n,t}] \\ &\quad + \sum_{q=q_0+1}^{\infty} \mathbb{E}[(M_t^n(\pi_q x) - M_t^n(\pi_{q-1} x))C_{t,q}^n \mid \mathcal{F}_{n,t}]. \end{aligned}$$

Now, by (15) and the last display

$$\left\| \sum_{t=1}^n M_t^n(x) - \sum_{t=1}^n M_t^n(\pi_q x) \right\|_{\Pi_c} \leq I_1 + I_2 + II_1 + II_2 + III,$$

where

$$I_1 = \left\| \sum_{t=1}^n \Delta_t^n(B_{q_0} x) D_{t,q_0}^n \right\|_{\Pi_c},$$

$$I_2 = \left\| \sum_{t=1}^n \mathbb{E}[\Delta_t^n(B_{q_0} x) D_{t,q_0}^n | \mathcal{F}_{n,t}] \right\|_{\Pi_c},$$

$$II_1 = \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \Delta_t^n(B_q x) D_{t,q}^n \right\|_{\Pi_c},$$

$$II_2 = \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n | \mathcal{F}_{n,t}] \right\|_{\Pi_c}$$

and

$$III = \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} (M_t^n(\pi_q x) - M_t^n(\pi_{q-1} x)) C_{t,q}^n - \mathbb{E}[(M_t^n(\pi_q x) - M_t^n(\pi_{q-1} x)) C_{t,q}^n | \mathcal{F}_{n,t}] \right\|_{\Pi_c}.$$

Further, it holds by the triangle inequality that  $II_1 \leq II_3 + II_2$ , where

$$II_3 = \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \Delta_t^n(B_q x) D_{t,q}^n - \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n | \mathcal{F}_{n,t}] \right\|_{\Pi_c}.$$

Hereafter, we perform estimations for terms  $I_1$ ,  $I_2$ ,  $II_3$ ,  $II_2$  and  $III$ . First, from  $\Delta_t^n(B_q x) \leq 2\|M_t^n(x)\|_{\Pi_c}$ , we have that under our assumptions it can be easily proved that  $I_1$  and  $I_2$  converge in probability to zero for any fixed  $q_0$ , see for instance Lemma A.2 in [Stute et al. \(1998\)](#).

By assumption (W2), for any  $\eta > 0$  there exists a constant  $K = K_\eta > 0$ , such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\Omega_n \setminus \Omega_K^n) \leq \eta,$$

where

$$\Omega_K^n \left\{ \sup_{q \in \mathbb{N}} \frac{\alpha_{n,w}(\mathcal{B}_\varepsilon)}{2^{-2q}} \leq K \right\}.$$

Then, for the estimation of  $II_2$ , we see that

$$\begin{aligned} II_2 &\leq \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \frac{1}{a_q} \mathbb{E}[|\Delta_t^n(B_q x)|^2 D_{t,q}^n \mid \mathcal{F}_{n,t}] \right\|_{\Pi_c} \\ &\leq \sup_{q \geq q_0+1} \left\| \sum_{t=1}^n \frac{\mathbb{E}[|\Delta_t^n(B_q x)|^2 D_{t,q}^n \mid \mathcal{F}_{n,t}]}{2^{-2q}} \right\|_{\Pi_c} \sum_{q=q_0+1}^{\infty} \frac{2^{-2q}}{a_q} \\ &\leq K \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log N_{q+1}} \quad \text{a.s. on the set } \Omega_K^n. \end{aligned}$$

As for  $II_3$ , since

$$|\Delta_t^n(B_q x) D_{t,q}^n - \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t}]| \leq 2a_{q-1} \quad \text{identically,}$$

and

$$\sum_{t=1}^n \mathbb{E}[|\Delta_t^n(B_q x)|^2 D_{t,q}^n \mid \mathcal{F}_{n,t}] \leq K 2^{-2q} \quad \text{a.s. on the set } \Omega_K^n,$$

it follows from the [Freedman's \(1975\)](#) inequality in Lemma A.2, which plays here the same role as the Bernstein's inequality does in the *iid* setup, and Lemma 2.11.17 of [van der Vaart and Wellner \(1996\)](#) that for any measurable set  $A$

$$\begin{aligned} \mathbb{E} \left[ \sum_{t=1}^n \Delta_t^n(B_q x) D_{t,q}^n - \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t}] \right] 1(A \cap \Omega_K^n) \\ \leq C(2a_{q-1} \log(N_q) + \sqrt{K} 2^{-q} \sqrt{\log(N_q)}) \left( P(A) + \frac{1}{N_q} \right) \\ \leq C((2 + \sqrt{K}) 2^{-q} \sqrt{\log(N_q)}) \left( P(A) + \frac{1}{N_q} \right). \end{aligned}$$

Thus, using the last inequality and defining for every  $q \in \mathbb{N}$ ,  $q \geq 1$ , a partition  $\{\Omega_{qk}^n : 1 \leq k \leq N_q\}$  of  $\Omega_n$  such that the maximum

$$\left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \Delta_t^n(B_q x) D_{t,q}^n - \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t}] \right\|_{\Pi_c}$$

is achieved at  $B_{qk}$  on the set  $\Omega_{qk}^n$ . Then, we have

$$\begin{aligned} \mathbb{E}[II_3 | 1(\Omega_K^n)] &\leq \mathbb{E} \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \Delta_t^n(B_q x) D_{t,q}^n - \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t}] \right\|_{\Pi_c} 1(\Omega_K^n) \\ &\leq \sum_{q=q_0+1}^{\infty} \mathbb{E} \left\| \sum_{t=1}^n \Delta_t^n(B_q x) D_{t,q}^n - \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t}] \right\|_{\Pi_c} 1(\Omega_K^n) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_q \sum_{q_0+1}^{\infty} \sum_k^{N_q} \left| \mathbb{E} \sum_{t=1}^n \Delta_t^n(B_q x) D_{t,q}^n - \mathbb{E}[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t}] \right| \mathbf{1}(\Omega_{qk}^n \cap \Omega_K^n) \\
&\leq C(2 + \sqrt{K}) \sum_q \sum_{q_0+1}^{\infty} \sum_k^{N_q} 2^{-q} \sqrt{\log(N_q)} \left( P(\Omega_{qk}^n) + \frac{1}{N_q} \right) \\
&\leq C(2 + \sqrt{K}) \sum_q \sum_{q_0+1}^{\infty} 2^{-q} \sqrt{\log(N_q)}.
\end{aligned}$$

Finally, the estimation of  $III$  follows from the same arguments as for  $II_3$ , and therefore, we obtain

$$\mathbb{E}|III| \mathbf{1}(\Omega_K^n) \leq C(2 + \sqrt{K}) \sum_q \sum_{q_0+1}^{\infty} 2^{-q} \sqrt{\log(N_q)}.$$

The theorem follows from choosing a large  $K$ , a large  $q_0$  and then, letting  $n \rightarrow \infty$ .  $\square$

**Proof of Theorem 1.** To prove Theorem 1 we need that the conditions of Theorem A.1 hold. To that end, let us define the semimetric

$$d^2(x, y) := |F_{\theta_0}(v_y) - F_{\theta_0}(v_x)| + |F_I(u_y) - F_I(u_x)|,$$

where  $y := (u'_y, v_y)'$ ,  $x := (u'_x, v_x)' \in \overline{\mathbb{R}}^d$  and  $F_I$  is the stationary marginal distribution of  $I_1$ . Then, the equicontinuity of the joint cdf guarantees that for any  $\varepsilon > 0$  we can form a partition  $\mathcal{B}_\varepsilon = \{B_k; 1 \leq k \leq N_\varepsilon\}$  of  $\overline{\mathbb{R}}^d$  in  $\varepsilon$ -brackets  $B_k = [x_k, y_k]$ , i.e.,  $\{B_k\}_{k=1}^{N_\varepsilon}$  covers  $\overline{\mathbb{R}}^d$ ,  $x_k \leq y_k$ , and  $d^2(x_k, y_k) \leq \varepsilon^2$ . Let us denote for each  $k$ ,  $1 \leq k \leq N_\varepsilon$ ,  $y_k := (u'_{y_k}, v_{y_k})'$ ,  $x := (u'_{x_k}, v_{x_k})'$ . For every  $q \in \mathbb{N}$ ,  $q \geq 1$ , when  $\varepsilon = 2^{-q}$  we denote the previous partition by  $\mathcal{B}_q = \{B_{qk}; 1 \leq k \leq N_q \equiv N_{2^{-q}}\}$ . From standard results on VC-classes, see [van der Vaart and Wellner \(1996\)](#), we have that (13) of Theorem A.1 holds for these partitions. Furthermore, simple algebra yields

$$\begin{aligned}
\alpha_n(\mathcal{B}_q) &:= \max_{1 \leq k \leq N_q} n^{-1} \sum_{t=1}^n \mathbb{E} \left[ \left| \sup_{x, y \in B_{qk}} |w(\varepsilon_t, I_t, x) - w(\varepsilon_t, I_t, y)| \right|^2 \middle| \mathcal{F}_t \right] \\
&\leq \max_{1 \leq k \leq N_q} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[ \sup_{x, y \in B_{qk}} |1(\varepsilon_t \leq v_x) 1(I_t \leq u_x) - 1(\varepsilon_t \leq v_y) 1(I_t \leq u_y)|^2 \middle| \mathcal{F}_t \right] \right| \\
&\quad + \max_{1 \leq k \leq N_q} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{E} \left[ \sup_{x, y \in B_{qk}} |1(-\varepsilon_t \leq v_x) 1(I_t \leq u_x) - 1(-\varepsilon_t \leq v_y) 1(I_t \leq u_y)|^2 \middle| \mathcal{F}_t \right] \right| \\
&= 2 \max_{1 \leq k \leq N_q} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{E}[1(v_{xk} \leq \varepsilon_t \leq v_{yk}) \mid \mathcal{F}_t] 1(u_{xk} \leq I_t \leq u_{yk}) \right|. \tag{16}
\end{aligned}$$

Therefore, the last display and (A3) imply (14). Therefore, (W2) of Theorem A.1 holds and Theorem 1 is proved.  $\square$

Before proving Theorems 2 and 3, we need some additional lemmas which generalize Theorem 2.2.3 in [Koul \(2002\)](#) and Lemma 1.1 in [Koul \(1996\)](#) under only martingale difference assumptions. For a sequence of r.v.'s  $\delta_{n,t}$ ,  $1 \leq t \leq n$ , and  $x = (u', v)' \in \overline{\mathbb{R}}^d$ , let

define the processes

$$V_n(x) = \frac{1}{n} \sum_{t=1}^n 1(\varepsilon_{n,t} \leq v + \delta_{n,t}) 1(I_{n,t} \leq u),$$

$$J_n(x) = \frac{1}{n} \sum_{t=1}^n F_{\theta_0, I_{n,t}}(v + \delta_{n,t}) 1(I_{n,t} \leq u),$$

where  $F_{\theta_0, I_{n,t}}(v) = F_{\theta_0}(v | I_{n,t})$ . Let us also define

$$V_n^*(x) = \frac{1}{n} \sum_{t=1}^n 1(\varepsilon_{n,t} \leq v) 1(I_{n,t} \leq u), \quad x = (u', v)' \in \overline{\mathbb{R}}^d,$$

$$J_n^*(x) = \frac{1}{n} \sum_{t=1}^n F_{\theta_0, I_{n,t}}(v) 1(I_{n,t} \leq u),$$

$$U_n(x) = \sqrt{n}(V_n(x) - J_n(x))$$

and

$$U_n^*(x) = \sqrt{n}(V_n^*(x) - J_n^*(x)).$$

**Lemma A.3.** *Under the assumptions of Theorem 3 with  $\{\varepsilon_{n,t}, I'_{n,t}\}_{t=1}^n$  replacing  $\{\varepsilon_t, I'_t\}_{t=1}^n$  there and assuming that  $\delta_{n,t}$  is  $\mathcal{F}_{n,t}$ -measurable, where  $\mathcal{F}_{n,t} = \sigma(I'_{n,t}, I'_{n,t-1}, \dots)$ ,  $0 \leq t \leq n$ ,  $\max_{1 \leq t \leq n} |\delta_{n,t}| = o_P(1)$  and that  $\varepsilon_{n,t}$  is  $\mathcal{F}_{n,t+1}$ -measurable, then*

$$\|U_n - U_n^*\|_{\overline{\mathbb{R}}^d} = o_P(1).$$

**Proof.** The fact that  $U_n - U_n^*$  is a sum of conditionally centered bounded r.v.'s yields that

$$\text{Var}(U_n(x) - U_n^*(x)) \leq \text{E} n^{-1} \sum_{t=1}^n |F_{\theta_0, I_{n,t}}(v + \delta_{n,t}) - F_{\theta_0, I_{n,t}}(v)| 1(I_{n,t} \leq u) = o(1)$$

by (A3) and the Ergodic Theorem. The lemma follows from showing the asymptotic uniform equicontinuity of  $U_n$  and  $U_n^*$ . For the latter condition apply our Theorem A.1 as in Theorem 1 in a routine fashion. See Theorem 2.2.3 in [Koul \(2002\)](#) for details.  $\square$

For  $s \in \Theta$ , let define the processes

$$V_n(x, s) = \frac{1}{\sqrt{n}} \sum_{t=1}^n 1(Y_{n,t} - g_{n,t}(s) \leq v) 1(I_{n,t} \leq u),$$

$$d_{n,t}(s) = g_{n,t}(\theta_0 + n^{-1/2}s) - g_{n,t}(\theta_0),$$

$$v_n(x, s) = n^{-1/2} \sum_{t=1}^n F_{\theta_0, I_{n,t}}(v + d_{n,t}(s)) 1(I_{n,t} \leq u),$$

$$W_n(x, t) = V_n(x, \theta_0 + n^{-1/2}s) - v_n(x, s),$$

where  $g_{n,t}(\theta) \equiv g(I_{n,t}, \theta)$  and  $N_b = \{s \in \Theta : |s| \leq b\}$ .



**Lemma A.4.** *Under the assumptions of Lemma A.3, assume that,*

$$n^{1/2} \max_{1 \leq t \leq n} |\dot{g}(I_{n,t}, \theta_0)| = o_{\mathbb{P}}(1) \quad (17)$$

and

$$n^{-1} \sum_{t=1}^n |\dot{g}(I_{t,n}, \theta_0)| = O_{\mathbb{P}}(1). \quad (18)$$

Then, for any  $b \in \mathbb{R}$ ,

$$\sup |W_n(x, s) - W_n(x, 0)| = o_{\mathbb{P}}(1),$$

where the supremum is over  $(x, s) \in \overline{\mathbb{R}}^d \times N_b$ .

**Proof.** From (17) we readily obtain

$$\sup_{t,s} |d_{n,t}(s)| = o_{\mathbb{P}}(1).$$

Hence from Lemma A.3 we conclude that

$$\sup_x |W_n(x, s) - W_n(x, 0)| = o_{\mathbb{P}}(1), \quad s \in N_b.$$

To complete the proof, because the compactness of  $N_b$ , it suffices to show that  $\forall \alpha > 0$ ,  $\exists \delta > 0$  and  $n_0 < \infty$ , such that  $\forall s \in N_b$

$$\Pr \left( \sup_{x \in \overline{\mathbb{R}}^d, |r-s| \leq \delta} |D_n(x, r) - D_n(x, s)| > \alpha \right) \leq \alpha, \quad n > n_0, \quad (19)$$

where  $D_n(x, r) := W_n(x, r) - W_n(x, 0)$ . Write  $D_{n,2}(x, s, r) := D_n(x, r) - D_n(x, s)$ .

Now fix an  $\eta > 0$ ,  $s \in N_b$  and a  $\delta > 0$ . Let  $\Delta_{n,t} := n^{-1/2}(\delta |\dot{g}(I_{t,n}, \theta_0)| + 2b\eta)$  and

$$A_n := \left\{ \sup_{r \in N_b, |r-s| \leq \delta} |d_{n,t}(r) - d_{n,t}(s)| \leq \Delta_{n,t}, 1 \leq t \leq n \right\}.$$

From (A4), it follows that

$$\Pr(A_n) \geq 1 - \eta, \quad n > n_1. \quad (20)$$

Next, define, for  $x \in \overline{\mathbb{R}}^d$ ,  $a \in \mathbb{R}$

$$D_{n,3}(x, s, a) := n^{-1/2} \sum_{t=1}^n \{1(\varepsilon_{n,t} \leq v + d_{n,t}(s) + a\Delta_{n,t}) - F_{\theta_0, I_{n,t}}(v + d_{n,t}(s) + a\Delta_{n,t})\} 1(I_{n,t} \leq u).$$

By definition and from (17),  $d_{n,t}(s) + a\Delta_{n,t}$  is  $\mathcal{F}_{n,t}$ -measurable with

$$\max_{1 \leq t \leq n} |d_{n,t}(s) + a\Delta_{n,t}| = o_{\mathbb{P}}(1).$$

Therefore an application of Lemma A.3 yields that

$$\sup_x |D_{n,3}(x, s, a) - D_{n,3}(x, s, 0)| = o_{\mathbb{P}}(1), \quad a \in \mathbb{R}. \quad (21)$$

Now, using the monotonicity of the indicator function and the cdf  $F_{\theta_0, I_{n,t}}$ , we obtain, that on  $A_n$ ,  $\forall r \in N_b, |r - s| \leq \delta$ ,

$$\begin{aligned} |D_{n,2}(x, s, r)| &\leq |D_{n,3}(x, s, 1) - D_{n,3}(x, s, 0)| + |D_{n,3}(x, s, -1) - D_{n,3}(x, s, 0)| \\ &\quad + \left| n^{-1/2} \sum_{t=1}^n \{F_{\theta_0, I_{n,t}}(v + d_{n,t}(s) + \Delta_{n,t}) \right. \\ &\quad \left. - F_{\theta_0, I_{n,t}}(v + d_{n,t}(s) - \Delta_{n,t})\} 1(I_{n,t} \leq u) \right|. \end{aligned}$$

By (A3) the last term in this upper bound is no larger than

$$Cn^{-1} \sum_{t=1}^n (\delta |\dot{g}(I_{t,n}, \theta_0)| + 2b\eta) 1(I_{n,t} \leq u),$$

which, in view of (18), can be made smaller than  $\alpha$  with an arbitrary large probability for sufficiently large  $n$  by the choice of  $\delta$  and  $\eta$ . This together with (20) and (21) completes the proof of (19) and hence that of the theorem.  $\square$

**Proof of Theorems 2 and 3.** Define  $\hat{\varepsilon}_t := \varepsilon_t(\theta_n)$  and  $\varepsilon_t := \varepsilon_t(\theta_0)$  and write

$$\begin{aligned} S_{n, \theta_n}(u, v) &= n^{-1/2} \sum_{t=1}^n w_{\theta_0}(v) 1(I_t \leq u) + n^{-1/2} \sum_{t=1}^n \{w_{\theta_n}(v) - w_{\theta_0}(v)\} 1(I_t \leq u) \\ &:= S_{n, \theta_0}(u, v) + S_{2n, \theta_n}(u, v), \end{aligned}$$

where

$$S_{2n, \theta_n}(u, v) = n^{-1/2} \sum_{t=1}^n \{w_{\theta_n}(v) - w_{\theta_0}(v)\} 1(I_t \leq u) := D_{n1}(x) - D_{n2}(x) + D_{n3}(x),$$

with

$$D_{n1}(x) = n^{-1/2} \sum_{t=1}^n \{1(\hat{\varepsilon}_t \leq v) - 1(\varepsilon_t \leq v) - F_{\theta_0, I_{n,t}}(v + \hat{\varepsilon}_t - \varepsilon_t) + F_{\theta_0, I_{n,t}}(v)\} 1(I_t \leq u),$$

$$D_{n2}(x) = n^{-1/2} \sum_{t=1}^n \{1(-\hat{\varepsilon}_t \leq v) - 1(-\varepsilon_t \leq v) - F_{\theta_0, I_{n,t}}(v - \hat{\varepsilon}_t + \varepsilon_t) + F_{\theta_0, I_{n,t}}(v)\} 1(I_t \leq u)$$

and

$$D_{n3}(x) = n^{-1/2} \sum_{t=1}^n \{F_{\theta_0, I_{n,t}}(v - \hat{\varepsilon}_t + \varepsilon_t) - F_{\theta_0, I_{n,t}}(v + \hat{\varepsilon}_t - \varepsilon_t)\} 1(I_t \leq u).$$

From Lemma A.4 we have that

$$\sup_{x \in \overline{\mathbb{R}}^d} |D_{nj}(x)| = o_{\mathbb{P}}(1), \quad j = 1, 2.$$

Whereas, from (A3) and the Glivenko Cantelli's Theorem

$$\sup_{x \in \overline{\mathbb{R}}^d} |D_{n3}(x) + 2\sqrt{n}(\theta_n - \theta_0)r_{\theta_0}(x)| = o_{\mathbb{P}}(1).$$

As for  $S_{n,\theta_0}(u, v)$ , the weak convergence follows from our Theorem A.1. This proves Theorems 2 and 3.  $\square$

**Proof of Theorem 4.** First, note that under the local alternatives (11)

$$\mathbb{E}(\omega_t(v)|I_t) = \Delta_{\theta_0}^1(I_t, v) - \Delta_{\theta_0}^1(I_t, \infty) + \Delta_{\theta_0}^1(I_t, -v) := A_t(v),$$

where

$$\Delta_{\theta_0}^1(I_t, v) = \int_{-\infty}^v f_{\theta_0}(\bar{v}|I_t)h_{\theta_0}(I_t, \bar{v}) d\bar{v}.$$

Then, we write

$$\begin{aligned} S_{n,\theta_n}(u, v) &= n^{-1/2} \sum_{t=1}^n \left\{ w_t(v) - \frac{A_t(v)}{n^{1/2}} + w_{t,\theta_n}(v) - w_t(v) + \frac{A_t(v)}{n^{1/2}} \right\} 1(I_t \leq u) \\ &:= \tilde{S}_{n,\theta_0}(u, v) + A_1 + A_2, \end{aligned} \quad (22)$$

with  $w_{t,\theta_n}(v) = 1\{\varepsilon_t(\theta_n) \leq v\} - 1\{-\varepsilon_t(\theta_n) \leq v\}$ ,

$$\tilde{S}_{n,\theta_0}(u, v) = n^{-1/2} \sum_{t=1}^n \left\{ w_t(v) - \frac{A_t(v)}{n^{1/2}} \right\} 1(I_t \leq u),$$

$$A_1 = n^{-1/2} \sum_{t=1}^n \{w_{t,\theta_n}(v) - w_t(v)\} 1(I_t \leq u)$$

and

$$A_2 = n^{-1} \sum_{t=1}^n A_t(v) 1(I_t \leq u).$$

Arguing as in Theorem 3, we obtain

$$|A_1 + 2\sqrt{n}(\theta_n - \theta_0)r_{\theta_0}(x)| = o_p(1)$$

uniformly in  $x \in \overline{\mathbb{R}}^d$ . On the other hand, by the Ergodic Theorem, uniformly in  $x \in \overline{\mathbb{R}}^d$

$$|A_2 - \mathbb{E}[A_1(v)1(I_1 \leq u)]| = o_p(1).$$

As for  $\tilde{S}_{n,\theta_0}$ , because  $\{w_t(v) - n^{-1/2}A_t(v)\}1(I_t \leq u)$  is a zero mean square-integrable martingale difference sequence with respect to  $\mathcal{F}_t$ , for each  $x \in \overline{\mathbb{R}}^d$ , we can use our Theorem A.1 to conclude that

$$\tilde{S}_{n,\theta_0} \implies S_\infty.$$

Using the preceding equations and (22), the theorem holds by (A6).  $\square$

**Proof of Theorem 5.** We need to show that the process  $S_{n,\theta_n}^*$  (conditionally on the sample) has the same asymptotic finite-dimensional distributions that the process  $S_{n,\theta_n}$ , with  $\theta_1$  replacing  $\theta_0$ , and that  $S_{n,\theta_n}^*$  is asymptotically tight, both with probability one. Let us denote  $w_{t,\theta_n}^*(v) = \{1(\varepsilon_t^*(\theta_n) \leq v) - 1(-\varepsilon_t^*(\theta_n) \leq v)\}$  and  $w_{t,\theta_n}^*(v) = \{1(\varepsilon_t^*(\theta_n) \leq v) - 1(-\varepsilon_t^*(\theta_n) \leq v)\}$ ,

where  $e_t^*(\theta_n) = e_t(\theta_n)V_t$ . Then, write similarly to Theorem 3

$$\begin{aligned} S_{n,\theta_n}^* &= n^{-1/2} \sum_{t=1}^n w_{t,\theta_n}^*(v)1(I_t \leq u) + n^{-1/2} \sum_{t=1}^n \{w_{t,\theta_n}^*(v) - w_{t,\theta_n}^*(v)\}1(I_{t-1} \leq x) \\ &:= S_{1n,\theta_n}^*(u, v) + S_{2n,\theta_n}^*(u, v). \end{aligned}$$

First, we prove

$$S_{1n,\theta_n}^* \rightarrow_d \tilde{S}_\infty \quad \text{a.s.}$$

To this end, we need to show that the finite-dimensional distributions of  $S_{1n,\theta_n}^*$  converge (conditional on the original sample) to those of  $\tilde{S}_\infty$  a.s. for all samples. Let us consider a finite set of points of  $\mathbb{R}^d$ ,  $x_1 = (u'_1, v_1)', \dots, x_r = (u'_1, v_1)'$  and a real vector  $\lambda = (\lambda_1, \dots, \lambda_r)'$  with  $|\lambda| = 1$ . Define

$$Z_{n,r}^* = n^{-1/2} \sum_{t=1}^n \sum_{j=1}^r \lambda_j w_{t,\theta_n}^*(v_j)1(I_t \leq u_j) := \sum_{t=1}^n \zeta_{nt}^{r*},$$

where  $\zeta_{nt}^{r*}$  is implicitly defined. Then, noting that conditional on the original data,  $\zeta_{nt}^{r*}$  is an independent (not identically distributed) array of random variables,

$$\mathbb{E}^* \left( \sum_{t=1}^n \zeta_{nt}^{r*} \right) = \sum_{t=1}^n n^{-1/2} \sum_{j=1}^r \lambda_j \mathbb{E}^*(w_{t,\theta_n}^*(v_j)1(I_t \leq u_j)) = 0,$$

while

$$\begin{aligned} V^* \left( \sum_{t=1}^n \zeta_{nt}^{r*} \right) &= \sum_{t=1}^n V^*(\zeta_{nt}^{r*}) \\ &= \sum_{j=1}^r \sum_{h=1}^r \lambda_j \lambda_h \left( n^{-1} \sum_{t=1}^n w_{t,\theta_n}(v_j)w_{t,\theta_n}(v_h)1(I_t \leq u_j)1(I_t \leq u_h) \right) := \tilde{\sigma}_{n,r}^2. \end{aligned}$$

By (A7(d))  $\tilde{\sigma}_{h,r}^2 \rightarrow \sigma_{h,r}^2$  a.s., where  $\sigma_{h,r}^2$  is the variance associated to  $\tilde{S}_\infty$ . Then, it is easy to show that for some positive constants  $\delta$

$$\sum_{t=1}^n \mathbb{E}^* [|\zeta_{nt}^{r*}|^2 1(|\zeta_{nt}^{r*}| > \delta)]$$

converges almost surely to zero, see [Stute et al. \(1998, p. 149\)](#). Then the triangular array  $\{\zeta_{nt}^{r*}\}$  satisfies the conditions of the Lindeberg–Feller’s CLT, conditionally on almost all samples, so that  $\sum_{t=1}^n \zeta_{nt}^{r*} \Rightarrow^* \mathcal{N}(0, \sigma_{h,r}^2)$  a.s. Next, the asymptotic uniform equicontinuity almost sure in all samples follows from Theorem 2.11.9 in [van der Vaart and Wellner \(1996\)](#) or from our Theorem A.1.

As for  $S_{2n,\theta_n}^*$ ,

$$S_{2n,\theta_n}^*(u, v) = D_{n1}^*(x) - D_{n2}^*(x) + D_{n3}^*(x),$$

with

$$D_{n1}^*(x) = n^{-1/2} \sum_{t=1}^n \{1(\varepsilon_t^*(\theta_n^*) \leq v) - 1(\varepsilon_t^*(\theta_n) \leq v) - F_{\theta_n, J_{n,t}}(v + \varepsilon_t^*(\theta_n^*) - \varepsilon_t^*(\theta_n)) + F_{\theta_n, J_{n,t}}(v)\} 1(I_t \leq u),$$

$$D_{n2}^*(x) = n^{-1/2} \sum_{t=1}^n \{1(-\varepsilon_t^*(\theta_n^*) \leq v) - 1(-\varepsilon_t^*(\theta_n) \leq v) - F_{\theta_n, J_{n,t}}(v - \varepsilon_t^*(\theta_n^*) + \varepsilon_t^*(\theta_n)) + F_{\theta_n, J_{n,t}}(v)\} 1(I_t \leq u)$$

and

$$D_{n3}^*(x) = n^{-1/2} \sum_{t=1}^n \{F_{\theta_n, J_{n,t}}(v - \varepsilon_t^*(\theta_n^*) + \varepsilon_t^*(\theta_n)) - F_{\theta_n, J_{n,t}}(v + \varepsilon_t^*(\theta_n^*) - \varepsilon_t^*(\theta_n))\} 1(I_t \leq u).$$

A similar argument to Lemma A.4 yields that

$$\sup_{x \in \overline{\mathbb{R}}^d} |D_{nj}^*(x)| = o_p(1) \quad \text{a.s., } j = 1, 2.$$

Whereas, from the (A3) and the Glivenko Cantelli's Theorem

$$\sup_{x \in \overline{\mathbb{R}}^d} |D_{n3}^*(x) + 2\sqrt{n}(\theta_n^* - \theta_n)r_{\theta_n^*}(x)| = o_p(1) \quad \text{a.s.}$$

The theorem follows from (A7).  $\square$

## References

- Ahmad, I.A., Li, Q., 1997. Testing symmetry of an unknown density function by kernel method. *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 American Statistical Association* 77, 639–646.
- Bae, J., Levental, S., 1995. Uniform CLT for Markov chains and its invariance principle: a martingale approach. *Journal of Theoretical Probability* 8, 549–570.
- Bai, J., 2003. Testing parametric conditional distributions of dynamic models. *Review of Economic and Statistics* 85, 531–549.
- Bai, J., Ng, S., 2001. A test for conditional symmetry in time series models. *Journal of Econometrics* 103, 225–258.
- Behnen, K., Neuhaus, G., 1975. A central limit theorem under contiguous alternatives. *Annals of Statistics* 3, 1349–1353.
- Bhattacharya, P.K., Gastwirth, J.L., Wright, A.L., 1982. Two modified Wilcoxon tests for symmetry about an unknown location parameter. *Biometrika* 69, 377–382.
- Billingsley, P., 1961. The Lindeberg Levy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.
- Butler, C.C., 1969. A test for symmetry using the sample distribution function. *Annals of Mathematical Statistics* 40, 2209–2210.
- Chang, N.M., 1990. Weak convergence of a self consistent estimator of a survival function with doubly censored data. *Annals of Statistics* 18, 391–404.
- Corradi, V., Swanson, N.R., 2006. Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779–806.

- Csörgő, S., Heathcote, C.R., 1982. Some results concerning symmetric distributions. *Bulletin of the Australian Mathematical Society* 25, 327–335.
- Csörgő, S., Heathcote, C.R., 1987. Testing for symmetry. *Biometrika* 74, 177–184.
- Delgado, M.A., González Manteiga, W., 2001. Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics* 29, 1469–1507.
- Delgado, M.A., Stute, W., 2005. Distribution free specification tests of conditional models. Mimeo.
- Dudley, R.M., 1999. *Uniform Central Limit Theorems*. Cambridge University Press, Cambridge, MA.
- Epps, T.W., Pulley, L.B., 1983. A test for normality based on the empirical characteristic function. *Biometrika* 70, 723–726.
- Fan, Y., Gencay, R., 1995. A consistent nonparametric test of symmetry in linear regression models. *Journal of American Statistical Association* 90, 551–557.
- Feuerverger, A., Mureika, R.A., 1977. The empirical characteristic function and its applications. *Annals of Statistics* 5, 88–97.
- Freedman, D.A., 1975. On tail probabilities for martingales. *Annals of Probability* 3, 100–118.
- Ghosh, S., Ruymgaart, F., 1992. Applications of empirical characteristic functions in some multivariate problems. *Canadian Journal of Statistics* 20, 429–440.
- Giné, E., Zinn, J., 1990. Bootstrapping empirical measures. *Annals of Statistics* 18, 851–869.
- Hall, P., Heyde, C.C., 1980. *Martingale Limit Theory and its Applications*. Wiley, New York.
- Harvey, C., Siddique, A., 1999. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34, 465–487.
- Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55, 1263–1296.
- Heathcote, C., Rachev, S., Cheng, B., 1995. Testing multivariate symmetry. *Journal of Multivariate Analysis* 54, 91–112.
- Heyde, C.C., 1990. *Quasi likelihood and its applications: a general approach to optimal parameter estimation*. Springer, New York.
- Hill, D.L., Rao, P.V., 1977. Tests of symmetry based on Cramér von Mises statistics. *Biometrika* 64, 489–494.
- Hill, D.L., Rao, P.V., 1981. Tests of symmetry based on Watson statistics. *Communications in Statistics A* 10, 1111–1125.
- Khmaladze, E.V., 1981. Martingale approach to the goodness of fit tests. *Theory of Probability and its Applications* 26, 246–265.
- Koul, H.L., 2002. *Weighted Empirical Processes in Dynamic Nonlinear Models*, second ed. *Lecture Notes in Statistics*, vol. 166, Springer, Berlin.
- Koutrovelis, I., 1985. Distribution free procedures for location and symmetry inference problems based on the empirical characteristic function. *Scandinavian Journal of Statistics* 12, 257–269.
- Levental, S., 1989. A uniform CLT for uniformly bounded families of martingale differences. *Journal of Theoretical Probability* 2, 271–287.
- Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics* 112, 295–325.
- Liu, R.Y., 1988. Bootstrap procedures under some non i.i.d models. *Annals of Statistics* 16, 1696–1708.
- Nadaraya, E.A., 1975. Testing the symmetry of a distribution law (in Russian). *Soobsheniya Akademii Nauk Gruzinskovi SSR* 78, 277–280.
- Neuhaus, G., 1973. Asymptotic properties of the Cramér von Mises statistic when parameters are estimated. *Proceedings of the Prague Symposium on Asymptotic Statistics*, vol. 2, pp. 257–297.
- Neuhaus, G., 1976. Weak convergence under Contiguous Alternatives of the Empirical Process when Parameters are Estimated: The  $D_k$  Approach. *Lecture Notes in Mathematics*, vol. 566. Springer, Berlin, pp. 68–82.
- Neumeyer, N., Dette, H., 2003. Testing for symmetric error distribution in nonparametric regression models. Preprint.
- Nishiyama, Y., 2000. Weak convergence of some classes of martingales with jumps. *Annals of Probability* 28, 685–712.
- Psaradakis, Z., 2003. A bootstrap test for symmetry of dependent data based on a Kolmogorov Smirnov type statistic. *Communications in Statistics, Series B, Simulations 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 American Statistical Association* 75, 168–172.
- Smirnov, N.V., 1947. On criteria for the symmetry of distribution laws of random variables. *Doklady Akademii Nauk SSSR*.

- Stute, W., González Manteiga, W., Presedo Quindimil, M., 1998. Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association* 93, 141–149.
- van der Vaart, A.W., Wellner, J.A., 1996. *Weak Convergence and Empirical Processes*. Springer, New York.
- White, H., 1994. *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.
- Wooldridge, J., 1994. Estimation and Inference for Dependent Processes. In: Engle, R., McFadden, D. (Eds.), *Handbook of Econometrics*, vol. IV. North Holland, Amsterdam, pp. 2641–2739 (Chapter 45).
- Wu, C.F.J., 1986. Jackknife, Bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics* 14, 1261–1350.
- Zheng, X., 1998. Consistent specification testing for conditional symmetry. *Econometric Theory* 14, 139–149.