Chapter 7
Nonparametric Distribution-Free Model Checks for Multivariate Dynamic Regressions

J. Carlos Escanciano and Miguel A. Delgado

7.1 Introduction

Parametric time series regression models continue being attractive among practitioners because they describe, in a concise way, the relation between the response or dependent variable and the explanatory variables. Much of the existing statistical literature is concerned with the parametric modelling in terms of the conditional mean function of a response variable $Y_t \in \mathbb{R}$, given some conditioning variable at time $t-1$, $I_{t-1} \in \mathbb{R}^d$, $d \in \mathbb{N}$, say. More precisely, let $Z_t \in \mathbb{R}^m$, $m \in \mathbb{N}$, be a m-dimensional observable random variable (r.v) and $W_{t-1} = (Y_{t-1}, \ldots, Y_{t-s}) \in \mathbb{R}^s$. The conditioning set we consider at time $t-1$ is given by $I_{t-1} = (W'_{t-1}, Z'_t)'$, so $d = s+m$. We assume throughout the article that the time series process $\{(Y_t, Z'_t)' : t = 0, \pm 1, \pm 2, \ldots\}$ is strictly stationary and ergodic. Henceforth, A' denotes the matrix transpose of A.

It is well-known that under integrability of Y_t, we can write the tautological expression

$$Y_t = f(I_{t-1}) + \varepsilon_t,$$

where $f(z) = E[Y_t \mid I_{t-1} = z]$, $z \in \mathbb{R}^d$, is the conditional mean function almost surely (a.s.) of Y_t, given $I_{t-1} = z$, and $\varepsilon_t = Y_t - E[Y_t \mid I_{t-1}]$ satisfies, by construction, that $E[\varepsilon_t \mid I_{t-1}] = 0$ a.s.

Then, in parametric modelling one assumes the existence of a parametric family of functions $\mathcal{M} = \{f(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ and considers the following regression model

$$Y_t = f(I_{t-1}, \theta) + e_t(\theta), \tag{7.1}$$

AMS 2000 subject classification. 62M07, 62G09, 62G10.

M. A. Delgado (✉)
Universidad Carlos III de Madrid, Madrid, Spain
e-mail: delgado@est-econ.uc3m.es

J. C. Escanciano
Indiana University, Bloomington, USA
e-mail: jescanci@indiana.edu

S. Lahiri et al. (eds.), *Contemporary Developments in Statistical Theory,*
Springer Proceedings in Mathematics & Statistics 68, DOI 10.1007/978-3-319-02651-0_7,

© Springer International Publishing Switzerland 2014

with $f(I_{t-1}, \theta)$ a parametric specification for the conditional mean $f(I_{t-1})$, and $\{e_t(\theta) : t = 0, \pm 1, \pm 2, \ldots\}$ a sequence of r.v.'s, deviations of the model. Model (7.1) includes classes of linear and nonlinear regression models and linear and nonlinear autoregression models, such as Markov-switching, exponential or threshold autoregressive models, among many others (see Fan and Yao 2003).

The condition $f(\cdot) \in \mathcal{M}$ is tantamount to

$$H_0 : E[e_t(\theta_0) \mid I_{t-1}] = 0 \quad \text{a.s. for some } \theta_0 \in \Theta \subset \mathbb{R}^p.$$

We aim to test H_0 against the alternative hypothesis

$$H_A : P(E[e_t(\theta) \mid I_{t-1}] \neq 0) > 0, \text{for all } \theta \in \Theta \subset \mathbb{R}^p,$$

where $(\Omega, \mathcal{F}, P)$ is the probability space in which all the r.v.'s of this article are defined.

There is a vast literature on testing the correct specification of regression models. In an independent and identically distributed (i.i.d) framework, some examples of those tests have been proposed by Bierens (1982, 1990), Eubank and Spiegelman (1990), Eubank and Hart (1992), Härdle and Mammen (1993), Horowitz and Härdle (1994), Hong and White (1995), Fan and Li (1996), Zheng (1996), Stute (1997), Stute et al. (1998), Li and Wang (1998), Fan and Huang (2001), Horowitz and Spokoiny (2001), Li (2003), Khamaladze and Koul (2004), Guerre and Lavergne (2005) and Escanciano (2006a), to mention a few. Whereas, in a time series context some examples are Bierens (1984), Li (1999), de Jong (1996), Bierens and Ploberger (1997), Koul and Stute (1999), Chen et al. (2003), Stute et al. (2006) and Escanciano (2006b, 2007). This extensive literature can be divided into two approaches. In the first approach test statistics are based on nonparametric estimators of the local measure of dependence $E[e_t(\theta_0) \mid I_{t-1}]$. This local approach requires smoothing of the data in addition to the estimation of the finite-dimensional parameter vector θ_0, and leads to less precise fits, see Hart (1997) for some review of the local approach when $d = 1$. Tests within the local approach are in general asymptotic distribution-free (ADF).

The second class of tests avoids smoothing estimation by means of an infinite number of unconditional moment restrictions over a parametric family of functions, i.e., it is based on the equivalence

$$E[e_t(\theta_0) \mid I_{t-1}] = 0 \; a.s. \Longleftrightarrow E[e_t(\theta_0) w(I_{t-1}, x)] = 0, \quad \text{almost everywhere (a.e.) in } \Pi \subset \mathbb{R}^q, \tag{7.2}$$

where $\Pi \subset \mathbb{R}^q$, $q \in \mathbb{N}$, is a properly chosen space, and the parametric family of functions $\{w(\cdot, x) : x \in \Pi\}$ is such that the equivalence (7.2) holds, see Stinchcombe and White (1998) and Escanciano (2006b) for primitive conditions on the family $\{w(\cdot, x) : x \in \Pi\}$ to satisfy this equivalence. We call the approach based on (7.2) the "integrated approach". In the integrated approach, test statistics are based on a distance from the sample analogue of $E[e_t(\theta_0) w(I_{t-1}, x)]$ to zero. This integrated approach is well known in the literature and was first proposed by Bierens (1982), who used the exponential function $w(I_{t-1}, x) = \exp(ix'I_{t-1})$, where $i = \sqrt{-1}$ denotes the imaginary unit, see also Bierens (1990) and Bierens and Ploberger (1999). Stute (1997) using empirical process theory, proposed to use the indicator function $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$ in an i.i.d context. Stinchcombe and White (1998) emphasized that there are many other possibilities in the choice of w. Recently, Escanciano (2006a) has considered in an i.i.d setup the family $w(I_{t-1}, x) = 1(\beta' I_{t-1} \leq u)$, $x = (\beta', u)' \in \Pi_{pro}$, where $\Pi_{pro} = \mathbb{S}^d \times [-\infty, \infty]$ is the auxiliary space with $\mathbb{S}^d$ the unit ball in $\mathbb{R}^d$, i.e., $\mathbb{S}^d = \{\beta \in \mathbb{R}^d : |\beta| = 1\}$. This new family combines the good properties of exponential and indicator families and delivers a Cramér-von Mises (CvM) test simple to compute and with excellent power properties in finite samples, see Escanciano (2006a) for further details. Escanciano (2007) provides a unified theory for specification tests based on the integrated approach for a general weighting function w, including but not restricting to indicators and exponential families.

A tenet in the integrated approach is that the asymptotic null distribution of resulting tests depends on the data generating process (DGP), the specified model and generally on the true parameter θ_0. Consequently, critical values for integrated tests have to be approximated with the assistance of resampling methods. In particular, Escanciano (2007) justified theoretically a wild bootstrap method to approximate the asymptotic critical values for general integrated-based tests. In contrast, Koul and Stute (1999) avoided resampling procedures by means of a martingale transformation in the spirit of that initially proposed by Khamaladze (1981). However, Koul and Stute's setup was restricted to homocedastic autoregressive models of order 1. Recently, Khamaladze and Koul (2004) have applied the martingale transform to residual marked processes in multivariate regressions with i.i.d data, but the resulting test is not ADF since it depends on the joint distribution of regressors. The main contribution of this article is to complement these approaches and extend them to heteroskedastic multivariate time series processes. We apply the martingale transform coupled with the Rossenblatt's transform on the multivariate regressors to get ADF test free of the joint design distribution. We formally justify the effect of these transformations on our test statistics using new asymptotic theory of function-parametric empirical processes under martingale conditions. Finally, we compare via a Monte Carlo experiment, our new model checks with existing bootstrap approximations.

The layout of the article is as follows. In Sect. 2 we present the ADF tests based on continuous functionals of a martingale transform of the function-parametric residual marked empirical process. We begin by establishing some heuristics for the martingale transform. In Sect. 3 we establish the asymptotic distribution of our test under the null. In Sect. 4 we compare the bootstrap approach with the martingale approach via a Monte Carlo experiment. Proofs are deferred to an appendix.

A word on notation. In the sequel C is a generic constant that may change from one expression to another. Throughout, $|A|$ denotes the Euclidean norm of A. $\overline{\mathbb{R}}^d$ denotes the extended d-dimensional Euclidean space, i.e., $\overline{\mathbb{R}}^d = [-\infty, \infty]^d$. Let $\|X\|_p$ be the L_p-norm of a r.v X, i.e., $\|X\|_p = (E|X|^p)^{1/p}$, $p \geq 1$. Let $N_{[]}(\varepsilon, \mathcal{H}, \|\cdot\|_p)$ be the ε-bracketing number of a class of functions $\mathcal{H}$ with respect to the norm $\|\cdot\|_p$, i.e., the minimal number N for which there exist ε-brackets $\{[l_j, u_j] : \|l_j - u_j\|_p \leq \varepsilon,$

$\|l_j\|_p < \infty$, $\|u_j\|_p < \infty$, $j =, 1 \ldots, N\}$ covering $\mathcal{H}$, see Definition 2.1.6 in van der Vaart and Wellner (1996). Let $\ell^\infty(\mathcal{H})$ be the metric space of all real-valued functions that are uniformly bounded on $\mathcal{H}$. As usual, $\ell^\infty(\mathcal{H})$ is endowed with the *sup* norm, i.e., $\|z\|_{\mathcal{H}} = \sup_{h\in\mathcal{H}} |z(h)|$. Let $\Longrightarrow$ denote weak convergence on $\ell^\infty(\mathcal{H})$, see Definition 1.3.3 in van der Vaart and Wellner (1996). Throughout the article, weak convergence on compacta in $\ell^\infty(\mathcal{H})$ means weak convergence on $\ell^\infty(\mathcal{C})$ for all compact subsets $\mathcal{C} \subset \mathcal{H}$. Also $\xrightarrow{P^*}$ and $\xrightarrow{as^*}$ denote convergence in outer probability and outer almost surely, respectively, see Definition 1.9.1 in Vaart and Wellner (1996). The symbol $\to_d$ denotes convergence in distribution of Euclidean random variables. All limits are taken as the sample size $n \to \infty$.

7.2 The Function-Parametric Residual Process and the Martingale Transform

In view of a sample $\{(Y_t, I'_{t-1})' : 1 \le t \le n\}$, and motivated from (7.2), we define the function-parametric empirical process,

$$R_n(b,\theta) = n^{-1/2}\sum_{t=1}^{n} e_t(\theta)b(I_{t-1}),$$

indexed by $(b,\theta) \in \mathcal{B}\times\Theta$, for a class of "check" functions $\mathcal{B}$ and a parameter space Θ. Examples of $\mathcal{B}$ will be specified later. Two important processes associated to $R_n(b,\theta)$ are the error-marked process $R_n(b) = R_n(b,\theta_0)$ and the residual-marked process

$$R_n^1(b) \equiv R_n(b,\theta_n) = n^{-1/2}\sum_{t=1}^{n} e_t(\theta_n)b(I_{t-1}),$$

where θ_n is a $\sqrt{n}$-consistent estimator for θ_0 (see Assumption A4 below). For convenience, we shall assume that $\mathcal{B} \subset L_2(\overline{\mathbb{R}}^d, G)$, the Hilbert space of all G -square integrable measurable functions, where $G(dx) = \sigma^2(x)F(dx)$, $F(\cdot)$ is the joint cumulative distribution function (cdf) of I_{t-1}, and $\sigma^2(\cdot)$ is the conditional error variance, i.e., $\sigma^2(y) = E[\varepsilon_t^2 \mid I_{t-1} = y]$. As usual, $L_2(\overline{\mathbb{R}}^d, G)$ is furnished with the inner-product

$$\langle f, g\rangle = \int_{\mathbb{R}^d} f(x)g(x)G(dx).$$

and the induced norm $\|h\| = \langle h, h\rangle^{1/2}$.

The aim of this section is to construct a suitable check space $\mathcal{B}$ such that the process $R_n^1(b)$, with $b \in \mathcal{B}$, delivers tests based on test statistics, $\Gamma(R_n^1)$ say, which are consistent and ADF. In this article we shall focus in a particular check space that makes use of the martingale transformation proposed by Khmaladze (1981, 1993) for the problem of goodness-of-fit tests of distributions.

Let $g(I_{0,}\theta_0) = (\partial/\partial\theta')f(I_0,\theta_0)$ and $s(I_0,\theta_0) = \sigma^{-2}(I_0)g(I_0,\theta_0)$ be the non-standardized and standardized scores, respectively. From Theorem 1 in Sect. 3, under the null hypothesis and some mild regularity conditions, we have the following relation between $R_n(b)$ and $R_n^1(b)$, uniformly in $b \in \mathcal{B}$,

$$R_n^1(b) = R_n(b) - \langle b, s' \rangle \sqrt{n}(\theta_n - \theta_0) + o_P(1). \tag{7.3}$$

This relation gives us a clue about how to choose b for the test based on $R_n^1(b)$ being ADF. Namely, if b is orthogonal to the score, i.e., $\langle b, s' \rangle = 0$, we have the uniform representation

$$R_n^1(b) = R_n(b) + o_P(1),$$

and the estimation of θ_0 does not have any effect in the asymptotic null distribution of $R_n^1(b)$. Furthermore, it can be shown that the limit process of $R_n(b)$ is a standard function-parametric Brownian motion in $L_2(\overline{\mathbb{R}}^d, G)$, that is, a Gaussian process with zero mean and covariance function $\langle b_1, b_2 \rangle$. Following ideas from Khmaladze (1993), a simple way to make b orthogonal to the score is to use a transformation $\mathcal{T}$ from $L_2(\overline{\mathbb{R}}^d, G)$ to $L_2(\overline{\mathbb{R}}^d, G)$ with values in the orthogonal complement of the space generated by the score s, and consider the transformed process $R_n^1(\mathcal{T}b)$. The covariance function of the limit process of $R_n^1(\mathcal{T}b)$ is then $\langle \mathcal{T}b_1, \mathcal{T}b_2 \rangle$, so unless $\mathcal{T}$ is an isometry (i.e., $\langle \mathcal{T}b_1, \mathcal{T}b_2 \rangle = \langle b_1, b_2 \rangle$), the Brownian motion structure is lost. Therefore, we observe that a way to make the asymptotic null distribution "immune" to the estimation effect and, at the same time, preserve the original covariance structure is to consider $R_n^1(\mathcal{T}b)$, where $\mathcal{T}$ is an isometry with image orthogonal to the score. In other words, a suitable check space to obtain consistent and ADF tests is $\mathcal{B} = \{\mathcal{T}h : h \in \mathcal{H}\}$, for an isometry $\mathcal{T}$ with image orthogonal to the score (to obtain the ADF property) and with suitable large class of functions $\mathcal{H}$ (to obtain consistency in the test procedure).

A large class of isometries with the previous properties is the class of shift isometries. Let $bas = \{s, f_1, f_2, \ldots\}$ be an orthogonal basis of $L_2(\overline{\mathbb{R}}^d, G)$. Let us define the isometry $\mathcal{T}_{bas}$ in the following way

$$\mathcal{T}_{bas}s = f_1 \qquad \mathcal{T}_{bas}f_j = f_{j+1}, j > 1.$$

Then, it is easy to show that $\mathcal{T}$ is an isometry from $L_2(\overline{\mathbb{R}}^d, G)$ to $L_2(\overline{\mathbb{R}}^d, G)$ with values in the orthogonal complement of the score s. A remarkable example of a shift isometry is the Khmaladze's martingale transform (cf. Khmaladze 1981, 1993), that posseses the added property of having an explicit formula. We use the same notation as in Khmaladze and Koul (2004). Introduce the so called scanning family of measurable subsets $\mathcal{A} = \{A_\lambda : \lambda \in \mathbb{R}\}$ of $\overline{\mathbb{R}}^d$, such that

1: $A_z \subseteq A_u, \forall z \leq u$.

2: $G(A_{-\infty}) = 0$, $G(A_{\infty}) = 1$
3: $G(A_z)$ is a strictly increasing and absolutely continuous function of $z \in \mathbb{R}$.

An example of scanning family is the following. Assuming that $G(\beta' y)$ is absolutely continuous for some $\beta \in \overline{\mathbb{R}}^d$, then the family $\mathcal{A} = \{A_z : z \in \mathbb{R}\}$ with $A_z = \{y \in \overline{\mathbb{R}}^d : \beta' y \leq z\}$ is a scanning family. Now define $z(y) = \inf\{z : y \in A_z\}$ and

$$C_z = \int_{A_z^c} s(x, \theta_0) s'(x, \theta_0) G(dx),$$

where A_z^c is the complement of A_z. The linear operator T is given by

$$T f(u) = f(u) - K f(u), \tag{7.4}$$

where

$$K f(u) = \int_{A_{z(u)}} f(x) s'(x, \theta_0) C_{z(x)}^{-1} G(dx) s(u, \theta_0) \tag{7.5}$$

and $f(\cdot) \in L_2(\overline{\mathbb{R}}^d, G)$. Such transformation was first proposed in the goodness-of-fit literature by Khmaladze (1981, 1993). In the statistical literature this transformation has been considered and extended to other problems in e.g. Stute et al. (1998), Koul and Stute (1999), Stute and Zhu (2002) or Koul and Khmaladze (2004). This transformation is becoming well-known in other areas and has been already applied to a variety of problems in Bai and Ng (2001), Koenker and Xiao (2002), Bai (2003), Delgado et al. (2008), Delgado and Stute (2008), Bai and Chen (2008), Song (2009, 2010) and Angrist and Kuersteiner (2011). It is not difficult to show that T defined by (7.4) is an isometry from $L_2(\overline{\mathbb{R}}^d, G)$ to $L_2(\overline{\mathbb{R}}^d, G)$ with values in the orthogonal complement of the score s, see Khmaladze and Koul (2004) for the proof.

The martingale transform[1] T depends on unknown quantities which can be estimated from a sample. The natural estimator of the transformation is

$$T_n f(u) = f(u) - \int_{A_{z(u)}} f(x) s_n'(x, \theta_n) C_{n,z(x)}^{-1} G_n(dx) s_n(u, \theta_n),$$

where

$$C_{n,z} = \int_{A_z^c} s_n(x, \theta_n) s_n'(x, \theta_n) G_n(dx),$$

with $G_n(dy) = \sigma_n^2(y) F_n(dy)$, F_n is the empirical cdf of $\{I_{t-1}\}_{t=1}^n$, $s_n(I_0, \theta) = \sigma_n^{-2}(I_0) g(I_0, \theta)$, θ_n is a $\sqrt{n}$-consistent estimator of θ_0, and $\sigma_n^2(y)$ is a consistent nonparametric estimator of $\sigma^2(y)$ (for instance, a Nadaraya-Watson estimator).

[1] The martingale trasform has also been variously referred to as: an innovation approach (Khmaladze, 1988), and an innovation process approach (Stute, Thies, and Zhu, 1998).

From the integrated approach we know that in the construction of consistent tests, it is not necessary to consider the whole space of functions $L_2(\overline{\mathbb{R}}^d, G)$. A parametric family that delivers well-known limit processes is the indicator class $\mathcal{B}_{ind} = \{1(\cdot \leq x) \equiv 1_x(\cdot) : x \in \overline{\mathbb{R}}^d\} \subset L_2(\overline{\mathbb{R}}^d, G)$. For the univariate case, i.e., $d = 1$, continuous functionals of standardizations of $R_n^1(T_n 1_x)$ deliver ADF tests for H_0, see Koul and Stute (1999). However, in the multivariate case, $d \geq 2$, the asymptotic null distribution of $R_n^1(T_n 1_x)$ still depends on the conditional variance and the design distribution. To overcome this problem we consider the so-called Rossenblatt's (1952) transformation. This transformation produces a multivariate distribution that is i.i.d on the d-dimensional unit cube, thereby, leading to tests that can be based on standardized tables. Let $I_t = (I_{t1}, I_{t2}, \ldots, I_{td})'$ and define the transformation $u = (u_1, \ldots, u_d)' = T_R(x)$ component-wise by $u_1 = F_1(x_1) = P(I_{t1} \leq x_1)$, $u_2 = F_2(x_2 \mid x_1) = P(I_{t2} \leq x_2 \mid I_{t1} = x_1), \ldots, u_d = F_d(x_d \mid x_1, \ldots, x_{d-1}) = P(I_{td} \leq x_d \mid I_{t1} = x_1, \ldots, I_{td-1} = x_{d-1})$. The inverse $x = T_R^{-1}(u)$ can be obtained recursively. Rossenblatt (1952) showed that $U_{t-1} = T_R(I_{t-1})$ has a joint distribution which marginals are uniform and independently distributed on $[0, 1]^d$.

In the next section, we shall show that under the null hypothesis and some mild regularity conditions the transformed process $J_n(u) = R_n^1(T_n(\sigma_n^{-1}(\cdot)1_u \circ T_R(\cdot)))$ converges weakly to a zero mean Gaussian process in $\ell^\infty(B_{x_0})$, for a suitable chosen set $B_{x_0} \subset [0, 1]^d$, with covariance function $u_1 \wedge u_2$, where for $a = (a_1, \ldots, a_d)'$ and $b = (b_1, \ldots, b_d)'$, $a \wedge b = \min\{a_1, b_1\} \times \cdots \times \min\{a_d, b_d\}$, that is, a standard Brownian sheet.

In practice the conditional distributions $F_1, \ldots, F_d$, are unknown and have to be estimated. Following Angrist and Kuersteiner (2004), we consider kernel estimators

$$\widehat{F}_1(x_1) = n^{-1}\sum_{t=1}^{n} 1(I_{t-11} \leq x_1)$$

$$\vdots$$

$$\widehat{F}_d(x_d \mid x_1, \ldots, x_{d-1}) = \frac{n^{-1}\sum_{t=1}^{n} 1(I_{t-1d} \leq x_d) K_{d-1}((x_d^- - I_{t-1d}^-)/h_n)}{n^{-1}\sum_{t=1}^{n} K_{d-1}((x_d^- - I_{t-1d}^-)/h_n)},$$

where $x_d^- = (x_1, \ldots, x_{d-1})'$, $I_{t-1d}^- = (I_{t-11}, \ldots, I_{t-1d-1})'$, $K_j(x) = (2\pi)^{-j/2} \sum_{h=1}^{w} \gamma_h \, |\sigma_h|^{-j} \exp(-0.5x'x/\sigma_h^2)$, $\sum_{h=1}^{w} \gamma_h = 1, \sum_{h=1}^{w} \gamma_h \, |\sigma_h|^{2\,l} = 0$, for $l = 1, 2, \ldots, w-1$, and $h_n = O(n^{-1/(2+d)})$ is a bandwidth sequence. Other higher order kernels or other nonparametric estimators are possible, as long as A6(ii) in the next section is satisfied.

Our final process is $\widehat{J}_n(u) = R_n^1(T_n(\sigma_n^{-1}(\cdot)1_u \circ \widehat{T}_R(\cdot)))$, where $\widehat{T}_R$ uses the previously described kernel estimation. $\widehat{J}_n(u)$ is called here the Khmaladze-Rossenblatt's transformed residual marked process. As a test statistic we consider in this article a

CvM functional

$$CvM_n = \int_{B_{x_0}} \left|\widehat{J}_n(u)\right|^2 F_{n,U}(du),$$

where $F_{n,U}(\cdot)$ is the empirical distribution function of the transformed sample $\{U_{t-1}\}_{t=1}^n$, $B_{x_0} = \{u \in [0,1]^d : \beta_1' T_R^{-1}(u) \leq x_0\}$, $\beta_1 \in \mathbb{R}^d$, and $x_0 < \infty$ is a user-chosen parameter necessary to avoid non-invertibility problems of the matrix $C_{n,z(x)}$, see Koul and Stute (1999) for a related situation. In the simulations we choose x_0 as the $(100-d)\%$ empirical quantile of the sample $\{\beta_1' I_{t-1}\}_{t=1}^n$. Other spaces B_{x_0}, threshold values x_0 and functionals different from the CvM are, of course, possible. Our test will reject the null hypothesis H_0 for "large" values of CvM_n. Next section establishes the asymptotic theory for CvM_n and Sect. 4 shows, via a Monte Carlo experiment, that it leads to a valuable diagnostic test.

7.3 Asymptotic Null Distribution

In this section we establish the limit distribution of $\widehat{J}_n$ under the null hypothesis H_0. First, we state a uniform representation for the function-parametric process $R_n^1(b)$, $b \in \mathcal{B}$, for a generic $\mathcal{B}$. This result is of independent interest. To derive these asymptotic results we consider the following notation and definitions. Let $\mathcal{F}_t = \sigma(I_t', I_{t-1}', \ldots, I_0')$ be the σ-field generated by the information set obtained up to time t. Let us endow $\mathcal{B}$ with the pseudo-metric $\|\cdot\|_{\mathcal{B}}$. Let us define $\mathcal{A} = \mathcal{B} \times \Theta$. For a given class of function $\mathcal{D}$ we define for $(r_1, r_2) \in \mathcal{D} \times \mathcal{D}$

$$d_{n,\mathcal{D}}^2(r_1, r_2) = n^{-1} \sum_{t=1}^n E\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right] |r_1(I_{t-1}) - r_2(I_{t-1})|^2$$

and

$$d_{\mathcal{D}}(r_1, r_2) = \|\varepsilon_t r_1(I_{t-1}) - \varepsilon_t r_2(I_{t-1})\|_2 .$$

Define the set $\Lambda_q = \{(r_1, r_2) \in \mathcal{D} \times \mathcal{D} : r_1 \leq r_2, d_{\mathcal{D}}^2(r_1, r_2) = 2^{-2q}\}$. If the family $\mathcal{D}$ satisfies that

$$\sup_{(r_1, r_2) \in \Lambda_q, q \in \mathbb{N}} \frac{d_{n,\mathcal{D}}^2(r_1, r_2)}{d_{\mathcal{D}}^2(r_1, r_2)} = O_P(1),$$

we say that $\mathcal{D}$ has bounded conditional quadratic variation with respect to $d_{\mathcal{D}}$. Also, we say that the class $\mathcal{D}$ satisfies a bracketing condition of order $p \geq 2$ and $s > 0$, in short $\mathcal{D}$ is $BEC(p, s)$, if

$$\int_0^\infty \left(\log\left(N_{[]}(\varepsilon^{1/s}, \mathcal{D}, \|\cdot\|_p)\right)\right)^{1/2} d\varepsilon < \infty.$$

The following assumptions are sufficient conditions for the weak convergence of $R_n^1(b)$ in $\ell^\infty(\mathcal{B})$ for a general $\mathcal{B}$.

Assumption A1: (*on the DGP*)

A1(a): $\{(Y_t, Z_t')' : t = 0, \pm1, \pm2, \ldots\}$ *is a strictly stationary and ergodic process.*

A1(b): $E[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ *a.s. for all* $t \geq 1$, *and* $E\,|\varepsilon_1|^2 < C$.

Assumption A2: (*on the set of functions* $\mathcal{B}$)

A2(a): (*Locally Uniform* L_p*-Smoothness*) *Suppose that for some* $s > 0, C_1 > 0$, *and for* $p \geq 2$, *the following holds: for each* $b_1 \in \mathcal{B}$,

$$\left\| \sup_{b_2 \in \mathcal{B}: \|b_1 - b_2\|_{\mathcal{B}} < \delta} |\varepsilon_t b_1(I_{t-1}) - \varepsilon_t b_2(I_{t-1})| \right\|_p \leq C_1 \delta^s.$$

A2(b): (*control the size of* $\mathcal{B}$) *The class of functions* $\mathcal{B}$ is $BEC(p, s)$ *for* p *and* s *as in A2(a).*

A2(c): *The class* $\mathcal{B}$ *has bounded conditional quadratic variation with respect to* $d_{\mathcal{B}}$ *and the parametric space* Θ *is compact in* $\mathbb{R}^p$.

Assumption A3: (*on the model*) $f(\cdot, \theta)$ *is twice continuously differentiable in a neighborhood of* $\theta_0 \in \Theta$. *There exists a function* $M(I_{t-1})$ *with* $\sup_{\theta \in \Theta} \left| g(I_{t-1,}\theta) \right| \leq M(I_{t-1})$, *such that* $M(I_{t-1})$ *is* $F(\cdot)$*-square integrable.*

Assumption A4: (*on the parameter*)

A4(a): *The true parameter* θ_0 *belongs to the interior of* Θ. *There exists a unique* θ_1 *such that* $|\theta_n - \theta_1| = o_P(1)$.

A4(b): *The estimator* θ_n *satisfies* $\sqrt{n}(\theta_n - \theta_0) = O_P(1)$.

Assumption A1(a) is standard in the model checks literature under time series, see, e.g., Koul and Stute (1999). A1(b) is weaker than other related moment conditions in the literature and allows for most empirically relevant conditional heteroskedastic models. A2 is needed for the asymptotic tightness of the process $R_n^1(b)$. The bracketing entropy condition has been frequently used in the literature. Combined with locally uniform L_p-continuity, the bracketing entropy condition can be used to establish the stochastic equicontinuity of a process that involves non-smooth functions containing infinite dimensional parameters. Assumption A3 is classical in the model checks literature, see, e.g., Stute and Zhu (2002). Assumption A4 is satisfied for most estimators in the literature, such as the conditional nonlinear least squares estimator (NLSE), or its robust modifications (under further regularity assumptions), see Koul's (1992, 2002) monographs. Under H_0, a more efficient estimator than the NLSE (see Wefelmeyer 1996) is given by the M-estimator satisfying the equation

$$\sum_{t=1}^{n} \sigma^{-2}(I_{t-1}) g(I_{t-1,}\theta_n)(Y_t - f(I_{t-1}, \theta_n) = 0. \tag{7.6}$$

A4(a) and A4(b) imply that under the null $\theta_0 = \theta_1$, but they might be different under the alternative. A2(c) is a standard assumption to obtain weak convergence theorems under martingale assumptions, see Bae and Levental (1995) and Nishiyama (2000). Because this assumption is crucial in most of our asymptotic results, we now give primitive and simple-to-check conditions for a class of functions $\mathcal{D}$ being of bounded conditional quadratic variation with respect to $d_{\mathcal{D}}$. See Escanciano and Mayoral (2010) for a related result. Let us define the quantity

$$G_t^{\mathcal{D}}(r) = E\left[E\left[\varepsilon_t^2 \mid I_{t-1}\right] r(I_{t-1}) \mid \mathcal{F}_{t-2}\right] \qquad r \in \mathcal{D},$$

Lemma 1: *Assume A1, A2(a-b) and that* $\left|G_t^{\mathcal{D}}(r_1) - G_t^{\mathcal{D}}(r_2)\right| \leq M_t d_{\mathcal{D}}^2(r_1, r_2)$, *where* M_t *is a stationary process with* $E[\,|M_1|\,] < \infty$. *Then,* $\mathcal{D}$ *has bounded conditional quadratic variation with respect to* $d_{\mathcal{D}}$.

Let V be a normal random vector with zero mean and variance–covariance matrix given by $L(\theta_0)$ (cf. A4(c)). Now, we are in position to state the asymptotic uniform representation of the process $R_n^1(b)$ and its weak convergence.

Theorem 1: *(i) Under Assumptions A1, A2 and A4(a) uniformly in* $b \in \mathcal{B}$,

$$\begin{aligned} R_n^1(\mathrm{b}) = {} & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{e_t(\theta_1) - E[e_t(\theta_1) \mid \mathcal{F}_{t-1}]\}\, b(I_{t-1}) \\ & + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ E[e_t(\theta) \mid \mathcal{F}_{t-1}]|_{\theta=\theta_n} - E[e_t(\theta_1) \mid \mathcal{F}_{t-1}]\right\} b(I_{t-1}) \\ & + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E[e_t(\theta_1) \mid \mathcal{F}_{t-1}] b(I_{t-1}) - E\left[E[e_t(\theta_1) \mid \mathcal{F}_{t-1}] b(I_{t-1})\right] \\ & + \sqrt{n} E\left[E[e_t(\theta_1) \mid \mathcal{F}_{t-1}] b(I_{t-1})\right] + o_P(1) \end{aligned}$$

(ii) If in addition, H_0, *A3 and A4(a) hold, then uniformly in* $b \in \mathcal{B}$,

$$R_n^1(b) = R_n(b) - \langle b, s' \rangle \sqrt{n}(\theta_n - \theta_0) + o_P(1).$$

The decomposition in Theorem 1(ii) paves the way for the discovery of appropriate martingale transforms of the residual marked process, see previous section. The analysis of function-parametric processes such as those considered in Theorem 1 provides simple methods of proof for the study of the asymptotic null distribution of $\widehat{J}_n$. To proceed further we need some regularity conditions.

Assumption A5: (*on the conditional variance and related quantities*)

A5(i): *The estimator* $\sigma_n^2(\cdot)$ *is a uniform consistent nonparametric estimator of* $\sigma^2(\cdot)$ *and* $0 < a \leq \sigma^2(y)$ for all $y \in \overline{\mathbb{R}}^d$ *and some positive* a.

A5(ii): $\sigma^{-j}(\cdot) \in \mathcal{W}$, $P(\sigma_n^{-j}(\cdot) \in \mathcal{W}) \to 1$ as $n \to \infty$ for $j = 1, 2$. *The class* $\mathcal{W}$ *satisfies A2(c), A2(a) for* $p > 2$ *and* $s = s_w > 0$ *and is* $BEC(p, r)$ *with* $r \leq \min(1, \mathrm{s_w})$. *Moreover,* $\mathcal{W}$ *has an envelope* $\overline{b}$, *such that* $\overline{b}(\cdot) < C < \infty$, *and the norm in* $\mathcal{W}$, $\|\cdot\|_{\mathcal{W}}$ *say, dominates the* L_2*-norm, i.e., there exists a* $C > 0$ *such that* $\|b\|_2 \leq C\,\|b\|_{\mathcal{W}}$, *for all* $b \in L_2(\overline{\mathbb{R}}^d, F)$.

A5(iii): $\mathcal{B}_{ind} = \{1_x(\cdot) : x \in \overline{\mathbb{R}}^d\}$ *satisfies A2(c) and* F *is absolutely continuous with respect to Lebesgue measure with density* $f(x) < \infty$ *for all* $x \in \overline{\mathbb{R}}^d$.

Assumption A6: A6(i): *The trimming constant* x_0 *is such that*

$$\inf_{x \in A_{x_0}} \left| C_{z(x)} \right| > \varepsilon > 0,$$

for some $\varepsilon > 0$ *and where* $A_{x_0} = \{x \in \overline{\mathbb{R}}^d : \beta_1' x \leq x_0\}$.

A6(ii): *The nonparametric estimators for the conditional distributions satisfy*

$$\sup_{x \in \mathbb{R}^d} \left| \widehat{F}_l(x_l \mid x_1, \ldots, x_{l-1}) - F_l(x_l \mid x_1, \ldots, x_{l-1}) \right| = o_P(1), l = 2, .., d,$$

A5(i) is standard in model checks under conditional heteroskedasticity, see Stute, Thies and Zhu (1998). Condition A5(ii) is necessary to obtain a uniform representation and tightness of the process $R_n^1(b)$ in $b \in \mathcal{B} = \{h1_x : h \in \mathcal{W} \text{ and } x \in \overline{\mathbb{R}}^d\}$. A5(ii) can be relaxed using results for degenerate U-processes, but it simplifies the theory and it gives us a clue about what are the properties necessary in $\mathcal{W}$ to obtain the asymptotic tightness of $R_n^1(b)$ in $b \in \mathcal{B}$. If we assume that $\sigma^{-2}(\cdot)$ is smooth, usual examples of $\mathcal{W}$ are spaces of smooth functions such as Sobolev, Hölder, or Besov classes. Therefore, the covering number condition of Assumptions A2 or A5(ii) can be found in many books and articles on approximation theory. To give an example, define for any vector $(a_1, \ldots, a_d)$ of d integers the differential operator $D^a = \partial^{|a|} / \partial x_1^{a_1} \ldots \partial x_q^{a_d}$, where $|a| = \sum_{i=1}^d a_i$. Let R be a bounded, convex subset of $\mathbb{R}^d$, with nonempty interior. For any smooth function $h : R \subset \mathbb{R}^d \to \mathbb{R}$ and some $\eta > 0$, let η be the largest integer smaller than η, and

$$\|h\|_{\infty,\eta} = \max_{|a| \leq \underline{\eta}} \sup_x \left| D^a h(x) \right| + \max_{|a| = \underline{\eta}} \sup_{x_1 \neq x_2} \frac{|D^a h(x_1) - D^a h(x_2)|}{\|x_1 - x_2\|^{\eta - \underline{\eta}}}.$$

Further, let $C_c^\eta(R)$ be the set of all continuous functions $h : R \subset \mathbb{R}^d \to \mathbb{R}$ with $\|h\|_{\infty,\eta} \leq c$. If $\mathcal{W} = C_c^\eta(R)$, then $\mathcal{W}$ satisfies Assumption A5(ii) provided that $\eta > d$, see van der Vaart and Wellner (1996, Theorem 2.7.1). A5(i) implies the invertibility of the matrix $C_{z(x)}$, and it is assumed only for simplicity in the exposition, see Nikabadze (1997). Conditions for A6(ii) to hold are in abundance in the literature, see, for instance, Andrews (1995). A6(ii) implies that

$$\sup_{x \in \mathbb{R}^d} \left| \widehat{T}_R(x) - T_R(x) \right| = o_P(1)$$

holds.

Theorem 2: *Under the null hypothesis* H_0, *and Assumptions A1 to A6*

$$\widehat{J}_n \Longrightarrow J_\infty, in \ell^\infty(B_{x_0}),$$

where J_∞ *is a standard Brownian Sheet, i.e, a continuous Gaussian process with zero mean and covariance function given by* $(u_{11} \wedge u_{21}) \times \cdots \times (u_{1d} \wedge u_{2d})$, *for* $u_1 = (u_{11}, \ldots u_{1d})'$ *and* $u_2 = (u_{21}, \ldots u_{2d})'$ *in* $[0, 1]^d$.

Next, using the last theorem and the Continuous Mapping Theorem (CMT), see, e.g., Theorem 1.3.6 in van der Vaart and Wellner (1996), we obtain the asymptotic null distribution of continuous functionals such as CvM_n.

Corollary 1: *Under the assumptions of Theorem 2, for any continuous (with respect to the sup norm) functional* $\Gamma(\cdot)$

$$\Gamma(\widehat{J}_n) \xrightarrow{d} \Gamma(J_\infty).$$

The integrating measure in CvM_n is a random measure, therefore, Corollary 1 is not readily applicable to the present case. However, an application of Lemma 3.1 in Chang (1990) shows that the estimation $F_{n,U}$ of the cdf of U_0, F_U say, does not affect the asymptotic theory for CvM_n as long as

$$\sup_{u \in B_{x_0}} \left| F_{n,U}(u) - F_U(u) \right| \longrightarrow 0 \text{ a.s.}$$

By the Glivenko-Cantelli's Theorem for ergodic and stationary time series, see e.g. Dehling and Philipp (2002, p. 4), jointly with A6(ii), the previous uniform convergence holds.

The power properties of CvM_n can be studied similarly to those established in Escanciano (2009). We do not discuss this issue here for the sake of space. A more important and difficult problem is the asymptotic power comparison between transformed and non-transformed tests from a theoretical point of view. This problem will be investigated elsewhere. Here, we focus on the finite-sample comparison between our ADF test and the bootstrap based tests via a Monte Carlo experiment in the next section.

7.4 Simulation Results

In this section we compare some bootstrap integrated CvM tests with our new ADF test via a Monte Carlo experiment. For the bootstrap CvM tests we consider the weighting functions $w(I_{t-1}, x) = \exp(ix' I_{t-1})$, $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$ and $w(I_{t-1}, x) = 1(\beta' I_{t-1} \leq u)$, $x = (\beta', u)' \in \Pi_{pro} = \mathbb{S}^d \times [-\infty, \infty]$. Our Monte Carlo experiment complements that of Koul and Sakhanenko (2005) in the context of goodness of fit for error distributions.

We briefly describe our simulation setup. Let $I_{t-1} = (Y_{t-1}, Y_{t-2})$ be the information set at time $t-1$. For our ADF test we consider $A_z = \{y \in \overline{\mathbb{R}}^2 : \beta_1' y \leq z\}$, with $\beta_1 = (1, 1)'$. Let $F_{n,\beta}(u)$ be the empirical distribution function of the projected information set $\{\beta' I_{t-1} : 1 \leq t \leq n\}$. Escanciano (2006a) proposed the CvM test

$$CVM_{n,pro} = \int_{\Pi_{pro}} (R^1_{n,pro}(\beta, u))^2 F_{n,\beta}(du) d\beta,$$

where

$$R^1_{n,pro}(\beta, u) = \frac{1}{\widehat{\sigma}_e \sqrt{n}} \sum_{t=1}^{n} e_t(\theta_n) 1(\beta' I_{t-1} \leq u)$$

and

$$\widehat{\sigma}_e^2 = \frac{1}{n} \sum_{t=1}^{n} e_t^2(\theta_n).$$

For a simple algorithm to compute $CVM_{n,pro}$ see Appendix B in Escanciano (2006a).

Bierens (1982) proposed to use $w(I_{t-1}, x) = \exp(iI'_{t-1}x)$ as the weighting function in (7.2) and considered the CvM test statistic

$$CvM_{n,\exp} = \int_{\Pi} \left| R^1_{n,\exp}(x) \right|^2 \Psi(dx),$$

where

$$R^1_{n,\exp}(x) = \frac{1}{\widehat{\sigma}_e \sqrt{n}} \sum_{t=1}^{n} e_t(\theta_n) \exp(ix' I_{t-1}),$$

and with $\Psi(dx)$ a suitable chosen integrating function. In order that $CvM_{n,\exp}$ has a closed expression, we consider the weighting function $\Psi(dx) = \phi(x)$, where $\phi(x)$ is the probability density function of the standard normal bivariate r.v. In that case, $CvM_{n,\exp}$ simplifies to

$$CvM_{n,\exp} = \frac{1}{\widehat{\sigma}_e^2 n} \sum_{t=1}^{n} \sum_{s=1}^{n} e_t(\theta_n) e_s(\theta_n) \exp\left(-\frac{1}{2} |I_{t-1} - I_{s-1}|^2\right).$$

Escanciano (2007) considered the CvM test based on the indicator function, which is given by

$$CvM_{n,ind} = \frac{1}{\widehat{\sigma}_e^2 n^2} \sum_{j=1}^{n} \left[\sum_{t=1}^{n} e_t(\theta_n) 1(I_{t-1} \le I_{j-1}) \right]^2.$$

We consider the wild bootstrap approximation for all these test statistics as described in Sect. 3 of Escanciano (2007).

Our null model is an AR(2) model:

$$Y_t = a + bY_{t-1} + cY_{t-2} + \varepsilon_t.$$

We examine the adequacy of this model under the following DGP:

1. AR(2) model: $Y_t = 0.6Y_{t-1} - 0.5Y_{t-2} + \varepsilon_t$.
2. AR(2) model with heteroskedasticity (ARHET): $Y_t = 0.6Y_{t-1} - 0.5Y_{t-2} + h_t\varepsilon_t$, where $h_t^2 = 0.1 + 0.1Y_{t-1}^2 + 0.3Y_{t-1}^2$.
3. Bilinear model (BIL): $Y_t = 0.6Y_{t-1} + 0.7\varepsilon_{t-1}Y_{t-2} + \varepsilon_t$.
4. Nonlinear Moving Average model (NLMA): $Y_t = 0.6Y_{t-1} + 0.7\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t$.
5. TAR(2) model: $Y_t = \begin{cases} 0.6Y_{t-1} + \varepsilon_t, & if\, Y_{t-2} < 1, \\ -0.5Y_{t-1} + \varepsilon_t, & if\, Y_{t-2} \ge 1. \end{cases}$

We consider for the experiments the sample sizes $n = 50, 100$, and 300. The number of Monte Carlo experiments is 1000 and the number of bootstrap replications is $B = 500$. In all the replications 200 pre-sample data values of the processes were

Table 7.1 *Empirical critical values for* CvM_n

$n\backslash\alpha$	10 %	5 %	1 %
50	0.55557	0.74353	1.18788
100	0.56371	0.75706	1.21756
300	0.61113	0.81060	1.35720

generated and discarded. For a fair comparison, the critical values for the new tests are approximated using 10000 replications of model 1. These critical values are given in Table 7.1.

In Table 7.2 we show the empirical rejection probabilities (RP) associated with the nominal levels 10, 5 and 1 %. The empirical levels of the test statistics are closed to the nominal level. Only in the heteroskedastic case the tests presents some small size distortion (underrejection).

In Table 7.3 we report the empirical power against the BIL, NLMA and TAR(2) alternatives. The RP increase with the sample size n for all test statistics, as expected.

Table 7.2 Empirical size of tests

		AR(2)			ARHET		
		10 %	5 %	1 %	10 %	5 %	1 %
$n = 50$	CvM_n	9.4	4.8	0.8	14.1	7.4	1.7
	$CvM_{n,\exp}$	10.5	5.5	1.1	13.6	7.8	0.8
	$CvM_{n,ind}$	10.3	4.3	1.3	12.4	6.5	1.0
	$CvM_{n,pro}$	11.6	5.7	0.8	13.1	5.9	1.0
$n = 100$	CvM_n	9.0	4.3	1.2	12.4	7.1	2.1
	$CvM_{n,\exp}$	13.4	7.0	1.0	11.7	6.9	2.7
	$CvM_{n,ind}$	11.3	6.5	1.4	12.7	5.8	1.4
	$CvM_{n,pro}$	11.2	6.4	1.6	13.4	7.1	2.0
$n = 300$	CvM_n	10.5	4.8	0.6	11.9	6.4	1.2
	$CvM_{n,\exp}$	10.3	6.0	1.9	12.3	6.1	1.5
	$CvM_{n,ind}$	9.6	4.7	0.5	11.8	6.2	2.0
	$CvM_{n,pro}$	12.5	5.7	1.8	13.2	7.1	1.6

Table 7.3 Empirical power of tests.

	BIL				NLMA			TAR(2)		
		10 %	*5 %*	*1 %*	*10 %*	*5 %*	*1 %*	*10 %*	*5 %*	*1 %*
$n = 50$	CvM_n	29.8	21.7	7.2	19.8	13.4	*4.7*	*53.3*	*40.8*	*19.6*
	$CvM_{n,\exp}$	29.4	18.0	4.4	16.0	8.6	1.5	23.0	13.4	2.0
	$CvM_{n,ind}$	32.2	22.8	8.1	*24.6*	*15.3*	4.6	39.8	30.0	10.5
	$CvM_{n,pro}$	*39.6*	*25.2*	*9.0*	22.9	11.6	2.3	38.5	27.2	9.7
$n = 100$	CvM_n	*56.1*	*43.0*	*24.6*	36.7	27.0	12.9	*76.3*	*69.1*	*49.7*
	$CvM_{n,\exp}$	43.8	30.0	10.7	28.6	16.2	3.8	43.2	27.5	8.2
	$CvM_{n,ind}$	50.0	39.4	19.1	*45.1*	*33.5*	*13.3*	65.4	54.8	34.9
	$CvM_{n,pro}$	55.7	42.3	20.1	41.0	26.8	9.0	62.0	51.3	28.2
$n = 300$	CvM_n	*96.6*	*93.1*	*81.5*	76.3	64.3	41.6	*99.5*	*99.0*	*95.9*
	$CvM_{n,\exp}$	77.2	66.0	36.9	75.6	61.0	28.4	92.5	86.4	61.1
	$CvM_{n,ind}$	76.2	68.4	50.8	88.8	*82.7*	*59.2*	98.5	96.9	88.1
	$CvM_{n,pro}$	75.2	65.8	44.8	*89.4*	80.8	51.9	98.7	96.6	86.5

The highest RP are presented in italics. It is shown that no test is better than the others uniformly for all alternatives, levels and sample sizes. The new ADF Cramér-von Mises test CvM_n performs quite well, being the best in many cases. In particular, it has the highest empirical power for BIL and TAR(2) alternatives uniformly in the level for $n = 300$. The empirical power for $CvM_{n,\exp}$ is low for these alternatives and, in general, less than $CvM_{n,ind}$. The test statistic $CvM_{n,ind}$ has good power against the BIL alternative for $n = 50$ and for the NLMA alternative for $n = 100$, and moderate power against the TAR(2). $CvM_{n,pro}$ performs similarly to $CvM_{n,ind}$, but with a little less empirical power in general.

Summarizing, we conclude from this limited Monte Carlo experiment that our new CvM test compares very well to bootstrap-based integrated tests, with power against all alternatives considered, and in many cases presenting the highest power performance. To conclude, we summarize the properties of our CvM test as follows: (i) it is asymptotically distribution-free; (ii) it is valid under fairly general regularity conditions on the underlying DGP, in particular, under conditional heteroskedasticity of unknown form and multivariate regressors; and (iii) it is simple to compute and has an excellent finite sample performance as has been shown in the Monte Carlo experiment. All these properties make of our test a valuable tool for time series modelling.

Acknowledgements Hira Koul aroused our interest in specification testing based on empirical processes many years ago. We are most thankful for his support to our research and the many discussions we had during these years. We are pleased to contribute to this volume with warm wishes for many more birthdays. Research funded by Spanish "Plan Nacional de I+D+i" reference number ECO2012–33053.

Appendix: Proofs

First, we shall state a weak convergence theorem which is a trivial extension of Theorem A1 in Delgado and Escanciano (2007). Let for each $n \geq 1$, $I'_{n,0}, \ldots, I'_{n,n-1}$, be an array of random vectors in $\mathbb{R}^p$, $p \in \mathbb{N}$, and $\varepsilon_{n,1}, \ldots, \varepsilon_{n,n}$, 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 σ-fields of $\mathcal{A}_n$ such that $\mathcal{F}_{n,t} \subset \mathcal{F}_{n,t+1}$, $t = 0, \ldots, n-1$ and such that for each $n \geq 1$ and each $\gamma \in \mathcal{H}$,

$$E[w(\varepsilon_{n,t}, I_{n,t-1}, \gamma) \mid \mathcal{F}_{n,t-1}] = 0 \qquad a.s., 1 \leq t \leq n, \forall n \geq 1. \tag{7.7}$$

Moreover, we shall assume that $\{w(\varepsilon_{n,t}, I_{n,t-1}, \gamma), \mathcal{F}_{n,t}, 0 \leq t \leq n\}$ is a square-integrable martingale difference sequence for each $\gamma \in \mathcal{H}$, that is, (7.7) holds, $Ew^2(\varepsilon_{n,t}, I_{n,t-1}, \gamma) < \infty$ and $w(\varepsilon_{n,t}, I_{n,t-1}, \gamma)$ is $\mathcal{F}_{n,t}$-measurable for each $\gamma \in \mathcal{H}$ and $\forall t, 1 \leq t \leq n, \forall n \in \mathbb{N}$. The following result gives sufficient conditions for the weak convergence of the empirical process

$$\alpha_{n,w}(\gamma) = n^{-1/2}\sum_{t=1}^{n} w(\varepsilon_{n,t}, I_{n,t-1}, \gamma) \qquad \gamma \in \mathcal{H}.$$

Under mild conditions the empirical process $\alpha_{n,w}$ can be viewed as a mapping from Ω_n to $\ell^\infty(\mathcal{H})$, the space of all real-valued functions that are uniformly bounded on $\mathcal{H}$. The weak convergence theorem that we present here is funded on results by Levental (1989), Bae and Levental (1995) and Nishiyama (2000). In Theorem A1 in Delgado and Escanciano (2007) $\mathcal{H}$ was finite-dimensional, but here we allow for an infinite-dimensional $\mathcal{H}$. The proof of theorem does not change by this possibility, however.

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 $\mathcal{H}$, which is defined as

$$\alpha_{n,w}(\mathcal{B}) = \max_{1\leq k\leq N} n^{-1}\sum_{t=1}^{n} E\left[\sup_{\gamma_1,\gamma_2\in H_k} \left|w(\varepsilon_{n,t}, I_{n,t-1}, \gamma_1) - w(\varepsilon_{n,t}, I_{n,t-1}, \gamma_2)\right|^2 \mid \mathcal{F}_{n,t-1}\right].$$

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

W1: For each $n \geq 1$, $\{(\varepsilon_{n,t}, I_{n,t-1})' : 1 \leq t \leq n\}$ is a strictly stationary and ergodic process. The sequence $\{w(\varepsilon_{n,t}, I_{n,t-1}, \gamma), \mathcal{F}_{n,t}, 0 \leq t \leq n\}$ is a square-integrable martingale difference sequence for each $\gamma \in \mathcal{H}$. Also, there exists a function $C_w(\gamma_1, \gamma_2)$ on $\mathcal{H} \times \mathcal{H}$ to $\mathbb{R}$ such that uniformly in $(\gamma_1, \gamma_2) \in \mathcal{H} \times \mathcal{H}$

$$n^{-1}\sum_{t=1}^{n} w(\varepsilon_{n,t}, I_{n,t-1}, \gamma_1) w(\varepsilon_{n,t}, I_{n,t-1}, \gamma_2) = C_w(\gamma_1, \gamma_2) + o_{P_n}(1).$$

W2: The family $w(\varepsilon_{n,t}, I_{n,t-1}, \gamma)$ is such that $\alpha_{n,w}$ is a mapping from Ω_n to $\ell^\infty(\mathcal{H})$ and for every $\varepsilon > 0$ there exists a finite partition $\mathcal{B}_\varepsilon = \{H_k; 1 \leq k \leq N_\varepsilon\}$ of $\mathcal{H}$, with N_ε being the elements of such partition, such that

$$\int_0^\infty \sqrt{\log(N_\varepsilon)}\, d\varepsilon < \infty \tag{7.8}$$

and

$$\sup_{\varepsilon\in(0,1)\cap\mathbb{Q}} \frac{\alpha_{n,w}(\mathcal{B}_\varepsilon)}{\varepsilon^2} = O_{P_n}(1). \tag{7.9}$$

Let $\alpha_{\infty,w}(\cdot)$ be a Gaussian process with zero mean and covariance function given by $C_w(\gamma_1, \gamma_2)$. We are now in position to state the following

Theorem A1: *If Assumptions W1 and W2 hold, then it follows that*

$$\alpha_{n,w} \Longrightarrow \alpha_{\infty,w} \text{ in } \ell^\infty(\mathcal{H}).$$

Proof of Theorem A1: Theorem A1 in Delgado and Escanciano (2007).

Proof of Lemma 1: By A2(a-b) we can form for any $\varepsilon > 0$ a finite partition $\mathcal{B}_\varepsilon = \{B_k; 1 \le k \le N_{[]}(\varepsilon, \mathcal{B}, \|\cdot\|_p)\}$ of $\mathcal{B}$ in ε-brackets $B_k = [\underline{b}_k, \overline{b}_k]$. Denote $v = 1/s$, with s as in A2(a), and define for every $q \in \mathbb{N}$, $q \ge 1$, $\varepsilon = 2^{-qv}$. We denote the previous partition associated to $\varepsilon = 2^{-qv}$ by $\mathcal{B}_q = \{B_{qk}; 1 \le k \le N_q \equiv N_{[]}(2^{-qv}, \mathcal{B}, \|\cdot\|_p)\}$. Without loss of generality we can assume that the finite partitions in the sequence $\{\mathcal{B}_q\}$ are nested. By A2(b), we have

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

Furthermore, by definition of the brackets

$$\begin{aligned} R_n(\mathcal{B}_q) &= \max_{1\le k\le N_q} \left| n^{-1} \sum_{t=1}^{n} E\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right] \sup_{r_1,r_2 \in B_{qk}} |r_1(I_{t-1}) - r_2(I_{t-1})|^2 \right| \\ &= \max_{1\le k\le N_q} \left| n^{-1} \sum_{t=1}^{n} E\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right] \left|\underline{b}_k(I_{t-1}) - \overline{b}_k(I_{t-1})\right|^2 \right| \\ &= \max_{1\le k\le N_q} d_n^2(\underline{b}_k, \overline{b}_k). \end{aligned} \tag{7.10}$$

Define the event

$$V_n = \left\{ \sup_{q\in\mathbb{N}} \max_{1\le k\le N_q} \frac{d_n^2(\underline{b}_k, \overline{b}_k)}{2^{-2q}} \ge \gamma \right\}.$$

We shall show that for all $\eta > 0$, there exists some $\gamma > 0$ such that $\limsup_{n\to\infty} P_n(V_n) \le \eta$. Note that

$$P_n(V_n) \le \sum_{q=1}^{\infty} P_n\left(\max_{1\le k\le N_q} \frac{d_n^2(\underline{b}_k, \overline{b}_k)}{2^{-2q}} \ge \gamma \right) \equiv \sum_{q=1}^{\infty} V_{nq} \tag{7.11}$$

Now, define the process

$$\widetilde{\alpha}_{n,w}(r) = n^{-1} \sum_{t=1}^{n} E\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right] r(I_{t-1}),$$

and the quantities for $1 \le t \le n$, $\widetilde{\beta}_t(r) = E\left[\varepsilon_t^2 \mid \mathcal{F}_{t-1}\right] r(I_{t-1}) - G_t^{\mathcal{B}}(r)$. Hence,

$$\widetilde{\alpha}_{n,w}(r) = n^{-1} \sum_{t=1}^{n} \widetilde{\beta}_t(r) + n^{-1} \sum_{t=1}^{n} G_t^{\mathcal{B}}(r).$$

By triangle's inequality

$$\begin{aligned} V_{nq} &\leq P_n\left(\max_{1\leq k\leq N_q}\left|n^{-1}\sum_{t=1}^{n}\left|\widetilde{\beta}_t(\underline{b}_k)-\widetilde{\beta}_t(\overline{b}_k)\right|\right|\geq 2^{-2q}\gamma\right) \\ &\quad +P_n\left(\max_{1\leq k\leq N_q}\left|n^{-1}\sum_{t=1}^{n}\left|G_t^{\mathcal{B}}(\underline{b}_k)-G_t^{\mathcal{B}}(\overline{b}_k)\right|\right|\geq 2^{-2q}\gamma\right) \\ &\equiv A_{1nq}+A_{2nq}. \end{aligned}$$

Notice that $\{\widetilde{\beta}_{n,w}(r),\mathcal{F}_{n,t-2}\}$ is a martingale difference sequence for each $r\in\mathcal{B}$, by construction. By a truncation argument, it can be assumed without loss of generality that $\max_{1\leq k\leq N_q}|\varepsilon_t|\left|\underline{b}_k(I_{t-1})-\overline{b}_k(I_{t-1})\right|^2\leq\sqrt{n}a_{q-1}$, where henceforth $a_q=2^{-q\rho}/\sqrt{\log(N_{q+1})}$ with $1<\rho<2$. See Theorem A1 in Delgado and Escanciano (2006). Define the set

$$B_n=\left\{\left(n^{-1}\sum_{t=1}^{n}M_t\right)\leq K\right\}.$$

Now, by Freedman's (1975) inequality in Lemma A2 and Lemma 2.2.10 in van der Vaart and Wellner (1996),

$$\begin{aligned} &E\max_{1\leq k\leq N_q}\left|n^{-1}\sum_{t=1}^{n}\left|\widetilde{\beta}_t(\underline{b}_k)-\widetilde{\beta}_t(\overline{b}_k)\right|\right|1(B_n) \\ &\qquad\leq C\left(a_{q-1}^2\log(1+N_q)+a_{q-1}2^{-q\nu/2}\sqrt{\log(1+N_q)}\right). \end{aligned}$$

Hence, by Markov's inequality and the definition of a_q, on the set B_n,

$$\begin{aligned} A_{1nq} &\leq C\frac{a_{q-1}^2\log(1+N_q)+a_{q-1}2^{-q\nu/2}\sqrt{\log(1+N_q)}}{2^{-2q}\gamma} \\ &= C\gamma^{-1}2^{-2q(\rho-1)}+C\gamma^{-1}2^{-q(\rho+\frac{\nu}{2}-1)}. \end{aligned}$$

On the other hand, by (D) and by Markov's inequality

$$\begin{aligned} A_{2nq} &\leq \gamma^{-1}s_n^{-2}\sum_{t=1}^{n}E\max_{1\leq k\leq N_q}2^{2q}\left|n^{-1}\sum_{t=1}^{n}\left|G_t^{\mathcal{B}}(\underline{b}_k)-G_t^{\mathcal{B}}(\overline{b}_k)\right|\right| \\ &\leq \gamma^{-1}2^{-q(\nu-2)}\left(n^{-1}\sum_{t=1}^{n}M_t\right)\leq K\gamma^{-1}2^{-q(\nu-2)}, \end{aligned}$$

on the set B_n. Therefore, by our previous arguments and the last three bounds,

$$P_n(V_n)\leq C\gamma^{-1}\sum_{q=1}^{\infty}\left(2^{-2q(\rho-1)}+2^{-q(\rho+\frac{\nu}{2}-1)}+2^{-q(\nu-2)}\right)+P_n(B_n^c),$$

which can be made arbitrarily small by choosing a sufficiently large γ and K. Hence, $\mathcal{B}$ has bounded quadratic variation. □

Lemma A0: (*Uniform Law of Large Numbers*) *If the class* $\mathcal{B}$ *is such that* $\log\left(N_{[]}(\varepsilon, \mathcal{B}, \|\cdot\|_1)\right) < \infty$ *for each* $\varepsilon > 0$, *with envelope* $\overline{b}$, $g(I_{t-1}, \theta)$ *satisfies A3 and* $E\left|M(I_{t-1})\overline{b}(I_{t-1})\right| < \infty$, *then uniformly in* $(\theta, b) \in \Theta \times \mathcal{B}$,

$$\left|\frac{1}{n}\sum_{t=1}^{n} g(I_{t-1},\theta)b(I_{t-1}) - E\left[g(I_{t-1},\theta)b(I_{t-1})\right]\right| = o_P(1).$$

Proof of Lemma A0: Under the assumptions of the lemma, the class $\{g(I_{t-1},\theta)b(I_{t-1}) : \theta \in \Theta, b \in \mathcal{B}\}$ is Glivenko-Cantelli. □

Proof of Theorem 1: First we shall show that the process

$$S_n(b,\theta) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{e_t(\theta) - E\left[e_t(\theta) \mid \mathcal{F}_{t-1}\right]\right\} b(I_{t-1}) \tag{7.12}$$

is asymptotically tight with respect to $(b,\theta) \in \mathcal{A}$.

Let us define the class $\mathcal{K} = \{\left\{e_t(\theta) - E\left[e_t(\theta) \mid \mathcal{F}_{t-1}\right]\right\} b(I_{t-1}) : (b,\theta) \in \mathcal{A}\}$. Denote $X_{t-1} = (I_{t-1}, I_{t-2}, \ldots)'$. Let $\mathcal{B}_\varepsilon = \{B_k; 1 \le k \le N_\varepsilon \equiv N_{[]}(\varepsilon, \mathcal{K}, \|\cdot\|_p\}$, with $B_k = [\underline{w}_k(Y_t, X_{t-1}), \overline{w}_k(Y_t, X_{t-1})]$, be a partition of $\mathcal{K}$ in ε-brackets with respect to $\|\cdot\|_p$. Notice that A2 implies

$$\begin{aligned}\Bigg\| \sup_{\substack{((b_2,\theta_2)\in\mathcal{A}:|\theta_1-\theta_2|<\delta \\ \|b_1-b_2\|_{\mathcal{B}}<\delta}} &\left|\left\{e_t(\theta_1) - E\left[e_t(\theta_1) \mid \mathcal{F}_{t-1}\right]\right\} b_1(I_{t-1})\right. \\ &\left. - \left\{e_t(\theta_2) - E\left[e_t(\theta_2) \mid \mathcal{F}_{t-1}\right]\right\} b_2(I_{t-1})\right| \Bigg\|_p \\ &\le C_1\delta^s.\end{aligned}$$

Theorem 3 in Chen et al. (2003) and A2 imply that (7.8) holds for such partition. On the other hand

$$\begin{aligned}&\max_{1\le k\le N_\varepsilon} n^{-1}\sum_{t=1}^{n} E\left[\left|\sup_{w_1,w_2\in B_k} |w_1(Y_t,X_{t-1}) - w_2(Y_t,X_{t-1})|\right|^2 \mid \mathcal{F}_{t-1}\right] \\ &\le \max_{1\le k\le N_\varepsilon} n^{-1}\sum_{t=1}^{n} E\left[\left|\underline{w}_k(Y_t,X_{t-1}) - \overline{w}_k(Y_t,X_{t-1})\right|^2 \mid \mathcal{F}_{t-1}\right].\end{aligned} \tag{7.13}$$

Therefore, A2(c) yields that (7.9) follows, and condition W2 of Theorem A1 holds. The asymptotically tightness of $S_n(b,\theta)$ is then proved.

Then, the last statement and A4(a)

$$R_n^1(\cdot) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{e_t(\theta_1) - E[e_t(\theta_1) \mid \mathcal{F}_{t-1}]\} b(I_{t-1})$$

$$+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{E[e_t(\theta)\mid\mathcal{F}_{t-1}]|_{\theta=\theta_n}-E[e_t(\theta_1)\mid\mathcal{F}_{t-1}]\right\}b(I_{t-1})$$
$$+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}E[e_t(\theta_1)\mid\mathcal{F}_{t-1}]b(I_{t-1})-E\left[E[e_t(\theta_1)\mid\mathcal{F}_{t-1}]b(I_{t-1})\right]$$
$$+\sqrt{n}E\left[E[e_t(\theta_1)\mid\mathcal{F}_{t-1}]b(I_{t-1})\right]+o_P(1),$$

uniformly in $b\in\mathcal{B}$. Part (i) is proved.

As for (ii), A3 and A4(a) imply by the Mean Value Theorem

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{E[e_t(\theta)\mid\mathcal{F}_{t-1}]|_{\theta=\theta_n}-E[e_t(\theta_0)\mid\mathcal{F}_{t-1}]\right\}b(I_{t-1})$$
$$=-n^{1/2}(\theta_n-\theta_0)'\frac{1}{n}\sum_{t=1}^{n}g(I_{t-1},\theta_{ni})b(I_{t-1}),$$

and where θ_{ni} satisfies $|\theta_{ni}-\theta_0|\leq|\theta_n-\theta_0|$. Now, A3, A2(b) and Lemma A0 imply that, uniformly in $b\in\mathcal{B}$,

$$\left|\frac{1}{n}\sum_{t=1}^{n}g(I_{t-1},\theta_{ni})b(I_{t-1})-E\left[g(I_{t-1},\theta_0)b(I_{t-1})\right]\right|=o_P(1).$$

From (i) and the last display, (ii) is proved. □

Before proving Theorem 2 we need several useful Lemmas. Let us define $A_{x_0}=\{x\in\overline{\mathbb{R}}^d:\beta_1'x\leq x_0\}$.

Lemma A1: *Under the assumptions of Theorem 2, uniformly in $x\in A_{x_0}$,*

$$R_n^1(T\sigma_n^{-1}(\cdot)1_x)=R_n(T\sigma^{-1}(\cdot)1_x)+o_P(1).$$

Lemma A2: *Under the assumptions of Theorem 2, uniformly in $x\in A_{x_0}$,*

$$R_n^1(T_n\sigma_n^{-1}(\cdot)1_x)=R_n^1(T\sigma_n^{-1}(\cdot)1_x)+o_P(1).$$

Lemma A3: *Under the assumptions of Theorem 2, uniformly in $u\in B_{x_0}$*

$$R_n^1(T_n(\sigma_n^{-1}(\cdot)1_u\circ\widehat{T}_R(\cdot)))=R_n^1(T_n(\sigma_n^{-1}(\cdot)1_u\circ T_R(\cdot)))+o_P(1).$$

Before proving Lemmas A1 to A3 we shall prove two more Lemmas. We need to define first the classes of functions $\mathcal{S}=\{Th1_x(\cdot):h\in\mathcal{W}\text{ and }x\in A_{x_0}\}$ and $\mathcal{B}=\{h1_x:h\in\mathcal{W}\text{ and }x\in A_{x_0}\}$. Define the semimetric

$$d_{ind}(x_1,x_2)=\left\|\varepsilon_t1_{x_1}(I_{t-1})-\varepsilon_t1_{x_2}(I_{t-1})\right\|_2,$$

and recall that $\mathcal{B}_{ind}=\{1(\cdot\leq x)\equiv1_x(\cdot):x\in\mathbb{R}^d\}$.

Lemma B1: *Assume that $\mathcal{B}_{ind}$ satisfies A2(c). Then, if $\mathcal{W}$ satisfies A5(ii) then $\mathcal{B}$ satisfies A2 with $p = 2$.*

Lemma B2: *Assume A3, A5 and A6(i). Then, if $\mathcal{B}$ satisfies A2 with $p = 2$ then $\mathcal{S}$ satisfies A2 with $p = 2$.*

Proof of Lemma B1: We shall start with A2(a). Assume $0 < \delta < 1$. By the triangle inequality, for each $h_1 \in \mathcal{W}$ and each $x_1 \in \overline{\mathbb{R}}^d$

$$\begin{aligned}
&\left\| \sup_{x_2 \in \overline{\mathbb{R}}^d, h_2 \in \mathcal{W}: \|h_1 - h_2\|_{\mathcal{W}} < \delta, d_{ind}(x_1, x_2) < \delta} \left| \varepsilon_t h_1 1_{x_1}(I_{t-1}) - \varepsilon_t h_2 1_{x_2}(I_{t-1}) \right| \right\|_2 \\
\leq\ & C \left\| \sup_{x_2 \in \overline{\mathbb{R}}^d, h_2 \in \mathcal{W}: \|h_1 - h_2\|_{\mathcal{W}} < \delta, d_{ind}(x_1, x_2) < \delta} |\varepsilon_t h_1(I_{t-1})| \left| 1_{x_1}(I_{t-1}) - 1_{x_2}(I_{t-1}) \right| \right\|_2 \\
&+ C \left\| \sup_{x_2 \in \overline{\mathbb{R}}^d, h_2 \in \mathcal{W}: \|h_1 - h_2\|_{\mathcal{W}} < \delta, d_{ind}(x_1, x_2) < \delta} 1_{x_2}(I_{t-1}) \left| \varepsilon_t h_1(I_{t-1}) - \varepsilon_t h_2(I_{t-1}) \right| \right\|_2 \\
\leq\ & C\delta^1 + C\delta^{s_w} \\
\leq\ & C\delta^s,
\end{aligned}$$

with $s = \min(1, \mathrm{s_w})$, where the second inequality is by A5(ii). A2(b) follows from Theorem 6 in Andrews (1994) and A5(ii), because $\mathcal{B}_{ind}$ is $BEC(p, 1/2)$ for all $p \geq 2$. A2(c) follows from the previous arguments, using A5(ii) and that $\mathcal{B}_{ind}$ and $\mathcal{W}$ satisfy A2(c). □

Proof of Lemma B2: We shall start with A2(a). Assume $0 < \delta < 1$. By the triangle inequality, for each $h_1 \in \mathcal{W}$ and each $x_1 \in \overline{\mathbb{R}}^d$

$$\begin{aligned}
&\left\| \sup_{x_2 \in \overline{\mathbb{R}}^d, h_2 \in \mathcal{W}: \|h_1 - h_2\|_{\mathcal{W}} < \delta, d_{ind}(x_1, x_2) < \delta,} \left| \varepsilon_t T h_1 1_{x_1}(I_{t-1}) - \varepsilon_t T h_2 1_{x_2}(I_{t-1}) \right| \right\|_2 \\
\leq\ & C \left\| \sup_{x_2 \in \overline{\mathbb{R}}^d, h_2 \in \mathcal{W}: \|h_1 - h_2\|_{\mathcal{W}} < \delta, d_{ind}(x_1, x_2) < \delta,} \left| \varepsilon_t h_1 1_{x_1}(I_{t-1}) - \varepsilon_t h_2 1_{x_2}(I_{t-1}) \right| \right\|_2 \\
& C \left\| \sup_{x_2 \in \overline{\mathbb{R}}^d, h_2 \in \mathcal{W}: \|h_1 - h_2\|_{\mathcal{W}} < \delta, d_{ind}(x_1, x_2) < \delta,} \left| \varepsilon_t K h_1 1_{x_1}(I_{t-1}) - \varepsilon_t K h_2 1_{x_2}(I_{t-1}) \right| \right\|_2,
\end{aligned}$$

where K is defined in (7.5). Then, it is only necessary to consider the second term in the last inequality. Now, by the linearity of K and the triangle inequality this term is bounded by

$$\leq\ C \left\| \sup_{x_2 \in \overline{\mathbb{R}}^d, h_2 \in \mathcal{W}: \|h_1 - h_2\|_{\mathcal{W}} < \delta, d_{ind}(x_1, x_2) < \delta,} \varepsilon_t K\{h_1(\cdot)(1_{x_1}(\cdot) - 1_{x_2}(\cdot))\}(I_{t-1}) \right\|_2$$

$$+C\left\| \sup_{x_2\in\overline{\mathbb{R}}^d, h_2\in\mathcal{W}:\|h_1-h_2\|_{\mathcal{W}}<\delta, d_{ind}(x_1,x_2)<\delta,} \varepsilon_t K 1_{x_2}(\cdot)(h_1(\cdot)-h_2(\cdot))(I_{t-1})\right\|_2$$
$$\equiv A_1 + A_2.$$

A_1^2 is equal to

$$E\left[\sup \varepsilon_t{}^2\left(\int 1(y\in A_{z(I_{t-1})})h_1(\cdot)(1_{x_1}(\cdot)-1_{x_2}(\cdot))s'(x,\theta_0)C_{z(x)}^{-1}G(dx)s(I_{t-1},\theta_0)\right)^2\right],$$

where the sup is computed over $d_{ind}(x_1,x_2)<\delta$. By Cauchy-Schwartz's inequality (C-S), A3, A5 and A6(i) the integral is bounded by

$$C\left|\int h_1^2(\cdot)(1_{x_1}(\cdot)-1_{x_2}(\cdot))^2 G(dx)\right| \leq C d_{ind}^2(x_1,x_2),$$

and hence $|A_1|\leq C\delta$. The proof for A_2 follows from the same steps that for A_1, and hence, it is omitted.

The proof of A2(b) is straightforward. A2(c) can be proved following the arguments in the proof of A2(a). These proofs are omitted for the sake of space. □

Proof of Lemma A1: By Lemmas B1 and B2, $\mathcal{B}$ and $\mathcal{S}$ satisfies A2 with $p=2$. Hence, by Theorem 1,

$$R_n^1(Tb(\cdot)1_x) = R_n(Tb(\cdot)1_x) + o_P(1),$$

uniformly in $x\in A_{x_0}$ and $b\in\mathcal{W}$. Now, by the convergence of σ_n^{-1},

$$R_n^1(T\sigma_n^{-1}(\cdot)1_x) = R_n^1(T\sigma^{-1}(\cdot)1_x) + o_P(1),$$

uniformly in $x\in A_{x_0}$. □

Proof of Lemma A2: Write $R_n^1((T-T_n)\sigma_n^{-1}(\cdot)1_x)$ as

$$\begin{aligned}
&\int \sigma_n^{-1}(y)1_x(y)R_n^1\left(s'(\cdot,\theta_0)1(\cdot\in A_{z(y)}^c)\right)C_{z(y)}^{-1}g(y,\theta_0)F(dy)\\
&-\int \sigma_n^{-1}(y)1_x(y)R_n^1\left(s_n'(\cdot,\theta_n)1(\cdot\in A_{z(y)}^c)\right)C_{n,z(y)}^{-1}g(y,\theta_n)F_n(dy)\\
=\;&\int \sigma_n^{-1}(y)1_x(y)\beta_n(\cdot,\sigma^{-2}(\cdot),\theta_0)\left[F(dy)-F_n(dy)\right]\\
&-\int \sigma_n^{-1}(y)1_x(y)\left[\beta_n(\cdot,\sigma_n^{-2}(\cdot),\theta_n)-\beta_n(\cdot,\sigma^{-2}(\cdot),\theta_0)\right]F_n(dy)\\
\equiv\;& A_{1n}(x)-A_{2n}(x),
\end{aligned}$$

where

$$\beta_n(y,b,\theta) = R_n^1\left(g'(\cdot,\theta)b(\cdot)1(\cdot\in A_{z(y)}^c)\right)C_{z(y)}^{-1}g(y,\theta). \tag{7.14}$$

Putting

$$\alpha_n(y) = \sigma_n^{-1}(y)1_x(y)\beta_n(\cdot, \sigma^{-2}(\cdot), \theta_0),$$

and using our Theorem 1 it is not difficult to show that the sequence $\{\alpha_n(\cdot)\}$ is asymptotically tight. Hence, by Lemma 3.4 in Stute, Thies and Zhu (1998)

$$\sup_{x \in A_{x_0}} |A_{1n}(x)| = o_P(1).$$

Similarly, it can be proved that $\beta_n(y, b, \theta)$ is uniformly tight in $(y, b, \theta) \in B_{x_0} \times \mathcal{W} \times \Theta$ (see Lemmas B1 and B2) and continuous in θ, but θ_n converges in probability to θ_0, and hence, again by Lemma 3.4 in Stute, Thies and Zhu (1998)

$$\sup_{x \in A_{x_0}} |A_{2n}(x)| = o_P(1).$$

□

Proof of Lemma A3: Define

$$\widehat{\gamma}_u(I_{t-1}) = 1_u \circ \widehat{T}_R(I_{t-1}),$$

$$\widetilde{\gamma}_u(I_{t-1}) = 1_u \circ T_R(I_{t-1})$$

and

$$d_u(\cdot) = \widehat{\gamma}_u(\cdot) - \widetilde{\gamma}_u(\cdot).$$

Then, write $R_n^1(T_n\sigma_n^{-1}(d_u(\cdot)))$ as

$$R_n^1(\sigma_n^{-1}(d_u(\cdot))) - \int d_u(\cdot)\sigma_n^{-1}(y)R_n^1\left(s_n'(\cdot, \theta_n)1(\cdot \in A_{z(y)}^c)\right) C_{n,z(y)}^{-1} g_n(y, \theta_n)F_n(dy)$$
$$\equiv A_{n1} - A_{n2}.$$

$|A_{n1}|$ is bounded by

$$\left| n^{-1/2}\sum_{t=1}^{n} e_t(\theta_0)\sigma_n^{-1}(I_{t-1})d_u(I_{t-1}) \right| + \left| n^{-1/2}\sum_{t=1}^{n}\{e_t(\theta_n) - e_t(\theta_0)\}\sigma_n^{-1}(I_{t-1})d_u(I_{t-1}) \right|$$
$$= \left| R_n\left(\sigma_n^{-1}d_u(\cdot)\right)\right| + \left|\sqrt{n}(\theta_n - \theta_0)'n^{-1}\sum_{t=1}^{n} g(I_{t-1}, \theta_{ni})\sigma_n^{-1}(I_{t-1})d_u(I_{t-1})\right|$$
$$\equiv |B_{n1}(u)| + |B_{n2}(u)|.$$

Now, the stochastic equicontinuity of $R_n b 1_x$ in $b \in \mathcal{W}$ and $1_x \in \mathcal{B}_{ind}$, and A6(ii) yield

$$\sup_{u \in [0,1]^d} |B_{1n}(u)| = o_P(1).$$

On the other hand, by Lemma A0, uniformly in $b \in \mathcal{B}$,

$$\left| \frac{1}{n}\sum_{t=1}^{n} g(I_{t-1},\theta_{ni})b(I_{t-1}) - E\left[g(I_{t-1},\theta_0)b(I_{t-1})\right] \right| = o_P(1).$$

Therefore, A4(b) and the last display yield

$$\sup_{u\in[0,1]^d} |B_{2n}(u)| = o_P(1).$$

As for A_{n2}, by C-S,

$$\left[\int \left[\widehat{\gamma}_u(y) - \widetilde{\gamma}_u(y)\right]^2 F_n(dy)\right]^{1/2} \left[\int \sigma_n^{-2}(y)\beta_n^2(y,\sigma_n^{-1},\theta_n)F_n(dy)\right]^{1/2},$$

where β_n is defined in (7.14). Both integrants are asymptotically tight (see the arguments of Lemma A2). Hence, Lemma 3.1 in Chang (1990) yields

$$\int \left[\widehat{\gamma}_u(y) - \widetilde{\gamma}_u(y)\right]^2 F_n(dy) = \int \left[\widehat{\gamma}_u(y) - \widetilde{\gamma}_u(y)\right]^2 F(dy) + o_P(1)$$

and

$$\int \sigma_n^{-2}(y)\beta_n^2(y,\sigma_n^{-1},\theta_n)F_n(dy) = O_P(1).$$

Now, we shall show that A6(ii) and A6(iii) imply

$$\sup_{u\in B_{x_0}} \left| \int \left[\widehat{\gamma}_u(y) - \widetilde{\gamma}_u(y)\right]^2 F(dy) \right| = o_P(1). \tag{7.15}$$

To that end, from A6(ii) we have that

$$\sup_{x\in\mathbb{R}^d} \left|\widehat{T}_R(x) - T_R(x)\right| = o_P(1),$$

Hence, for a given $\varepsilon > 0$, there exists and n_0 such that for all $n \geq n_0$

$$\sup_{x\in\mathbb{R}^d} \left|\widehat{T}_R(x) - T_R(x)\right| < \varepsilon$$

with probability tending to one. Therefore, on that set

$$\sup_{u\in B_{x_0}} \left| \int \left[\widehat{\gamma}_u(y) - \widetilde{\gamma}_u(y)\right]^2 F(dy) \right| \leq \sup_{u\in B_{x_0}} \left| E\left[1(u-\varepsilon \leq U_{t-1} \leq u+\varepsilon\right]\right| \leq 2\varepsilon.$$

Hence, as ε was arbitrary (7.15) holds, and Lemma A3 is proved. □

References

Andrews DWK (1994) Empirical process method in econometrics. In: Engle RF, McFadden DL (ed) The Handbook of Econometrics, vol. IV. North-Holland, Amsterdam

Andrews DWK (1995) Nonparametric kernel estimation for semiparametric models. Econometric Theor 11:560–596

Angrist JD, Kuersteiner GM (2011) Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. Rev Econ Stat 93:725–747

Bai J., 2003, Testing parametric conditional distributions of dynamic models. Review of Economics and Statistics 85:531–549

Bai J, Ng S (2001) A consistent test for conditional symmetry in time series models. J Economet 103:225–258

Bai J, Chen Z (2008) Testing multivariate distributions in GARCH models. J Economet 143:19–36

Bae J, Levental S (1995) Uniform CLT for markov chains and its invariance principle: A martingale approach. J Theor Probab 8:549–570

Bierens HJ (1982) Consistent model specification tests. J Economet 20:105–134

Bierens HJ (1984) Model specification testing of time series regressions. J Economet 26:323–353

Bierens HJ (1990) A consistent conditional moment test of functional form. Econometrica 58: 1443–1458

Bierens HJ, Ploberger W (1997) Asymptotic theory of integrated conditional moment tests. Econometrica 65:1129–1151

Chang NM (1990) Weak convergence of a self-consistent estimator of a survival function with doubly censored data. Ann Stat 18:391–404

Chen SX, Härdle W, Li M (2003) An empirical likelihood goodness-of-fit test for time series. J Roy Statist Soc Ser B 65:663–678

de Jong RM (1996) The Bierens' tests under data dependence. J Economet 72:1–32

Dehling H, Philipp W., 2002, Empirical process techniques for dependent data. In: Dehling H, Mikosch T, Sø rensen M (ed) Empirical process techniques for dependent data (Birkhaüser), 3–113

Delgado MA, Escanciano JC (2007) Nonparametric tests for conditional symmetry in dynamic models. J Economet 141:652–682

Delgado MA, Hidalgo J, Velasco C (2008) Distribution free goodness-of-fit tests for linear models. Ann Stat 33:2568–2609

Delgado MA, Stute W (2008) Distribution free specification tests of conditional models. J Economet 143:37–55

Escanciano JC (2006a) A consistent diagnostic test for regression models using projections. Economet Theor 22:1030–1051

Escanciano JC (2006b) Goodness-of-fit tests for linear and non-linear time series models. J Amer Stat Assoc 101:531–541

Escanciano JC (2007) Model checks using residual marked empirical processes. Stat Sin 17: 115–138

Escanciano JC (2009) On The Lack of Power of Omnibus Specification Tests. Economet Theor 25:162–194

Escanciano JC, Mayoral S (2010) Data-driven smooth tests for the martingale difference hypothesis. Comput Stat Data An 54:1983–1998

Eubank R, Hart J (1992) Testing goodness-of-fit in regression via order selection criteria. Ann Stat 20:1412–1425

Eubank R, Spiegelman S (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. J Amer Stat Assoc 85:387–392

Fan J, Huang L (2001) Goodness-of-fit tests for parametric regression models. J Amer Stat Assoc 96:640–652

Fan J, Yao Q, 2003, Nonlinear time series: Nonparametric and parametric methods. Springer-Verlag, New York

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

Guerre E, Lavergne P (2005) Rate-optimal data-driven specification testing for regression models. Ann Stat 33:840–870

Härdle W, Mammen E (1993) Comparing nonparametric versus parametric regression fits. Ann Stat 21:1926–1974

Hart JD, 1997, Nonparametric smoothing and lack-of-fit tests. Springer Verlag, New-York

Hong Y, White H (1995) Consistent specification testing via nonparametric series regression. Econometrica 63:1133–1159

Horowitz JL, Härdle W (1994) Testing a parametric model against a semiparametric alternative. Economet Theor 10:821–848

Horowitz JL, Spokoiny VG (2001) An adaptive rate-optimal test of a parametric mean-regression model against a nonparametric alternative. Econometrica 69:599–632

Khmaladze EV., 1981, Martingale approach to the goodness of fit tests. Theory Probab Appl 26:246–265

Khmaladze EV (1993) Goodness of fit problem and scanning innovation martingales. Ann Stat 21:798–829

Khmaladze EV, Koul H (2004) Martingale transforms goodness-of-fit tests in regression models. Ann Stat 32:995–1034

Koenker R, Xiao Z (2002) Inference on the quantile regression process. Econometrica 70:1583–1612

Koul HL 1992. Weighted empiricals and linear models. IMS Lecture Notes-Monograph Series, vol. 21. Hayward, California

Koul HL., 2002, Weighted Empirical Processes in Dynamic Nonlinear Models, 2nd ed, Lect Notes Stat, Vol. 166, Springer

Koul HL, Sakhanenko L (2005) Goodness-of-fit testing in regression: A finite sample comparison of bootstrap methodology and Khamaladze transformation. Stat Probab Lett 74:290–302

Koul HL, Stute W (1999) Nonparametric model checks for time series. Ann Stat 27:204–236

Levental S (1989) A uniform CLT for uniformly bounded families of martingale differences. J Theor Probab 2:271–287

Li Q (1999) Consistent model specification test for time series econometric models. J Economet 92:101–147

Li Q, Hsiao C, Zinn J (2003) Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. J Economet 112:295–325

Li Q, Wang S (1998) A simple consistent bootstrap test for a parametric regression functional form. J Economet 87:145–165

Nikabadze A., 1997, Scanning innovations and goodness of fit tests for vector random variables against the general alternative. A. Razmadze Mathematical Institute, Tbilisi, Preprint

Nishiyama Y (2000) Weak convergence of some classes of martingales with jumps. Ann Probab 28:685–712

Rossenblatt M (1952) Remark on multivariate transformation. Ann Math Stat 23:470–472

Song KK (2009) Testing conditional independence via rosenblatt transforms. Ann Stat 37: 4011–4045

Song KK (2010) Testing semiparametric conditional moment restrictions using conditional martingale transforms. J Economet 154:74–84

Stinchcombe M, White H (1998) Consistent specification testing with nuisance parameters present only under the alternative. Economet Theor 14:295–325

Stute W (1997) Nonparametric model checks for regression. Ann Stat 25:613–641

Stute W, Gonzalez-Manteiga W, Presedo-Quindimil M (1998) Bootstrap approximations in model checks for regression. J Amer Stat Assoc 93:141–149

Stute W, Thies S, Zhu LX (1998) Model checks for regression: An innovation process approach. Ann Stat 26:1916–1934

Stute W, Presedo-Quindimil M, González-Manteiga W, Koul HL (2006) Model checks of higher order time series. Stat Probab Lett 76:1385–1396

Stute W, Zhu LX (2002) Model checks for generalized linear models. Scand J Stat 29:535–545 van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer, New York

Wefelmeyer W (1996) Quasi-likelihood models and optimal inference. Ann Stat 24:405–422

Zheng X (1996) A consistent test of functional form via nonparametric estimation technique. J Economet 75:263–289