

## MODEL CHECKS USING RESIDUAL MARKED EMPIRICAL PROCESSES

J. Carlos Escanciano

*Indiana University*

*Abstract:* This article proposes omnibus goodness-of-fit tests of a parametric regression time series model. We use a general class of residual marked empirical processes as the buildingblocks for our testing problem. First, we establish a new weak convergence theorem under mild assumptions, one that extends previous existing asymptotic results and which may be of independent interest. This result allows us to study the asymptotic null distribution of the tests statistics and their asymptotic behavior against Pitman's local alternatives in a unified way. To approximate the asymptotic null distribution of test statistics we give a theoretical justification of a bootstrap procedure. Our bootstrap tests are robust to conditional higher moments of unknown form, in particular to conditional heteroskedasticity. Finally, a Monte Carlo study shows that the bootstrap and the asymptotic results provide good approximations for small sample sizes and an empirical application to the Canadian lynx data set is considered.

*Key words and phrases:* Canadian lynx data set, conditional mean, diagnostic tests, marked empirical processes, time series, weak convergence, wild bootstrap.

### 1. Introduction

The last decades have seen increased interest in time series modelling. Much of the existing statistical and econometrics literature has been concerned with the parametric modelling of the conditional mean function of a response variable  $Y_t \in \mathbb{R}$ , given some information set at time  $t - 1$ ,  $I_{t-1} \in \mathbb{R}^d$ , say. More precisely, let  $Z_t$  be an  $m$ -dimensional observable random variable and let  $W_{t-1} = (Y_{t-1}, \dots, Y_{t-s})'$ . The information set we consider at time  $t - 1$  is  $I_{t-1} = (W'_{t-1}, Z'_t)'$ , so  $d = s + m$ . We assume throughout that the time series process  $\{(Y_t, I'_{t-1})' : t = 0, \pm 1, \pm 2, \dots\}$  is strictly stationary and ergodic. 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  $f(z) = E[Y_t | I_{t-1} = z]$ , is the conditional mean function and  $\varepsilon_t = Y_t - E[Y_t | I_{t-1}]$  satisfies  $E[\varepsilon_t | I_{t-1}] = 0$  a.s..

In parametric regression modelling one assumes the existence of a parametric family of functions  $\mathcal{M} = \{f(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ , and proceeds to make inferences

on  $\theta$  or to test the hypothesis that  $f(\cdot) \in \mathcal{M}$ . Parametric time series regression models continue to be attractive among practitioners because the parameter  $\theta$  together with the functional form  $f(I_{t-1}, \theta)$  describe, in a concise way, the relation between the response  $Y_t$  and the information set  $I_{t-1}$ . Examples of models  $\mathcal{M}$  include 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 e.g., Tong (1990), or more recently, Fan and Yao (2003). Define  $e_t(\theta_0) = Y_t - f(I_{t-1}, \theta_0)$ . When  $f(I_{t-1}, \theta)$  is correctly specified for  $f(I_{t-1})$ ,

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

There is a huge 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), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Horowitz and Härdle (1994), Hong and White (1995), Fan and Li (1996), Zheng (1996), Stute (1997), Stute, Thies and Zhu (1998), Li and Wang (1998), Fan and Huang (2001), Horowitz and Spokoiny (2001), Li, Hsiao and Zinn (2003), Zhu (2003), Zhu and Ng (2003), Khmaladze and Koul (2004), Koul and Ni (2004), Escanciano (2006) and Guerre and Lavergne (2005), 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, Härdle and Li (2003), Escanciano (2006), Hong and Lee (2005) and Guerre and Guay (2005). This huge literature can be divided into two approaches. The first approach is called the “local approach”, because it is based on nonparametric estimators of the local measure of dependence  $E[e_t(\theta_0) | I_{t-1}]$ . The local approach requires smoothing of the data in addition to the estimation of the finite-dimensional parameter vector  $\theta_0$  and leads to less precise fits, see Hart (1997) for some review of the local approach when  $d = 1$ .

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

$$\begin{aligned} E[e_t(\theta_0) | I_{t-1}] = 0 \text{ a.s.} &\iff E[e_t(\theta_0)w(I_{t-1}, x)] = 0, \\ &\text{almost everywhere (a.e.) in } \Pi \subset \mathbb{R}^q, \end{aligned} \quad (2)$$

where  $\Pi \subset \mathbb{R}^q$  is a properly chosen space, and the parametric family of function  $\{w(\cdot, x) : x \in \Pi\}$  is such that (2) holds, see Stinchcombe and White (1998) and Escanciano (2006) for primitive conditions on the family  $\{w(\cdot, x) : x \in \Pi\}$  to satisfy this equivalence. We call the approach based on (2) the “integrated

approach”, because it uses integrated (or cumulative) measures of dependence. 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  $w(I_{t-1}, x) = \exp(ix' I_{t-1})$ , see also Bierens (1990) and Bierens and Ploberger (1997). 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, see also Stute, Thies and Zhu (1998). More recently, Koul and Stute (1999) have proposed asymptotic distribution-free tests for nonlinear autoregressive models of order 1, using again the indicator function. The exponential and indicator families are the most used in the literature. Stinchcombe and White (1998) emphasized that there are many other possibilities in the choice of  $w$ , such as  $w(I_{t-1}, x) = \sin(x' I_{t-1})$  or  $w(I_{t-1}, x) = 1/(1+\exp(c-x' I_{t-1}))$ , both of them with  $\Pi \subset \mathbb{R}^d$ . Recently, Escanciano (2006) 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  $\mathbb{S}^d = \{\beta \in \mathbb{R}^d : |\beta| = 1\}$ . This new family has the property that overcomes the curse of dimensionality because it is based on one-dimensional projections and, at the same time, avoids the choice of a subjective integrating measure in the Cramér-von Mises (CvM) test. In addition, the CvM test based on this new family has excellent power properties in finite samples and formalizes traditional inferences based on residual-fitted values plots for linear models, see Escanciano (2006) for details. Note that different families  $w$  have different power properties of the integrated-based test. The “optimal”  $w$  will depend on the true alternative at hand as well as the functional used to measure the orthogonality restrictions. It is worth stressing that the choice of  $w$  gives us flexibility to direct power in desired directions. So, it would be important to establish a general theory for integrated-based tests that covers a large class of weighting functions  $w$ .

The main aim of this article is to present a unified theory for the goodness-of-fit tests of regression time series models based on the integrated approach for a general weighting function  $w$ . The main contributions of the article are as follows. We establish a new weak convergence theorem for marked processes under martingale difference conditions, which improves previous results. As an application of the new weak convergence theorem, we extend the tests used in Stute (1997) to a multivariate time series framework under mild conditions. We also extend the test of Escanciano (2006) to a time series setup. We theoretically justify a new bootstrap method based on the wild bootstrap to approximate the asymptotic critical values for general integrated-based tests. Finally, we compare the proposed and existing tests in a Monte Carlo experiment to show that the proposed tests can play a valuable role in time series model checks.

The layout of the article is as follows. In Section 2 we consider the general theory for the integrated approach and we introduce the residual marked empirical processes that are the basis for the test statistics. We begin by establishing a weak convergence theorem for a general class of marked empirical processes. This allows us to study, in a unified way, the asymptotic distribution of the test statistics under the null and under local alternatives. In Section 3 a bootstrap procedure for approximating the asymptotic null distribution of the omnibus tests is considered and theoretically justified. In Section 4 a simulation exercise compares different test under the null and under the alternative, and we study the well-known Canadian lynx data set. Proofs are deferred to an appendix.

In the sequel,  $C$  is a generic constant that may change from one expression to another. Throughout,  $A'$ ,  $A^c$  and  $|A|$  denote the matrix transpose, the complex conjugate and the Euclidean norm of  $A$ , respectively. As usual,  $O_P(1)$  and  $o_P(1)$  denote bounded in probability and convergence in probability to zero under the probability  $P$ , respectively. All limits are taken as the sample size  $n \rightarrow \infty$ .

## 2. Asymptotic Theory: Residual Marked Empirical Processes

Denote by  $S$  the class of all strictly stationary ergodic processes with marginals in  $\mathbb{R}^{d+1}$ ,  $d \in \mathbb{N}$ , such that the first marginal component is integrable, and let  $\{Y_t, I'_{t-1}\}' : t = 0, \pm 1, \pm 2, \dots\}$  with  $0 < E|Y_t| < \infty$ , be one of these processes. Let  $(\Omega, \mathcal{A}, P)$  denote the probability space in which previous random variables (r.v's) are defined. The main goal in this article is to test the null hypothesis

$$H_0 : E[Y_t | I_{t-1}] = f(I_{t-1}, \theta_0) \text{ a.s., for some } \theta_0 \in \Theta \subset \mathbb{R}^p,$$

against the general nonparametric alternative

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

Note that we have restricted ourselves under both hypotheses to processes in  $S$ . One way to characterize  $H_0$  is by the infinite number of unconditional moment restrictions

$$E[e_t(\theta_0)w(I_{t-1}, x)] = 0 \text{ a.e., } x \in \Pi, \quad (3)$$

where  $\{w(\cdot, x) : x \in \Pi\}$  is such that the equivalence in (2) holds, see Section 1.

In view of a sample  $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$ , define the marked empirical process

$$R_{n,w}(x, \theta) = n^{-\frac{1}{2}} \sum_{t=1}^n e_t(\theta)w(I_{t-1}, x),$$

where  $e_t(\theta) = Y_t - f(I_{t-1}, \theta)$ ,  $\theta \in \Theta$ ,  $t \in \mathbb{Z}$ . The associated error-marked process is  $R_{n,w} = R_{n,w}(x, \theta_0)$  and the residual-marked process is  $R_{n,w}^1(x) = R_{n,w}(x, \theta_n)$ , for an estimator  $\theta_n$  for  $\theta_0$ .

Because of (2), test statistics are based on a distance from the standardized sample analogue of  $E[e_t(\theta_0)w(I_{t-1}, x)]$  to zero, i.e., on a norm of  $R_{n,w}^1$ ,  $\Gamma(R_{n,w}^1)$ , say. The most common of these are

$$\begin{aligned} CvM_{n,w} &= \int_{\Pi} |R_{n,w}^1(x)|^2 \Psi(dx), \\ KS_{n,w} &= \sup_{x \in \Pi_c} |R_{n,w}^1(x)|, \end{aligned}$$

where  $\Psi(x)$  is an integrating function satisfying some mild conditions, see A4(b) below. Other functionals are possible. Then, the tests we consider reject the null hypothesis  $H_0$  for “large” values of  $\Gamma(R_{n,w}^1)$ .

To study the asymptotic distribution of functionals of  $R_{n,w}^1$  for different families  $w$ , we need a sufficiently general weak convergence theorem that allows for continuous and discontinuous (with respect to  $x$ ) weighting functions. The next section gives the answer to this problem under very mild assumptions. This result is the main result of the article and may be of independent interest.

## 2.1. Weak convergence theorem

In this section we consider a weak convergence theorem for a large class of marked empirical processes of which  $R_{n,w}$  is a special case. Let, for each  $n \geq 1$ ,  $X'_{n,0}, \dots, X'_{n,n-1}$  be an array of random vectors in  $\mathbb{R}^d$ , and let  $\mathcal{F}_{n,t} = \sigma(X'_{n,t}, X'_{n,t-1}, \dots, X'_{n,0})$ ,  $0 \leq t \leq n$ , be the  $\sigma$ -field generated by the observations obtained up to time  $t$ . Furthermore let, for each  $n \geq 1$ ,  $\varepsilon_{n,1}, \dots, \varepsilon_{n,n}$  be an array of square integrable real r.v's such that for each  $t$ ,  $1 \leq t \leq n$ ,  $\varepsilon_{n,t}$  is  $\mathcal{F}_{n,t}$ -measurable and such that a.s.

$$E[\varepsilon_{n,t} \mid \mathcal{F}_{n,t-1}] = 0 \quad 1 \leq t \leq n, \quad \forall n \geq 1. \tag{4}$$

Denote by  $(\Omega_n, \mathcal{A}_n, P_n)$ ,  $n \geq 1$ , the probability space in which all the r.v's  $\{\varepsilon_{n,t}, X'_{n,t-1} : 1 \leq t \leq n\}$  are defined. The main goal of this section is to establish the weak convergence of the marked empirical process

$$\alpha_{n,w}(x) = n^{-\frac{1}{2}} \sum_{t=1}^n \varepsilon_{n,t} w(X_{n,t-1}, x), \quad x \in \Pi \subset \mathbb{R}^q.$$

Usually, different families  $w$  deliver different technical approaches for the asymptotic theory, essentially due to the continuity of the family  $w$  with respect to the auxiliary parameter  $x$ . Compare, for instance, the tightness conditions in

Bierens and Ploberger (1997) and Khmaladze and Koul (2004). One possibility for a unified theory is to embed the empirical process  $\alpha_{n,w}$  in a suitable large function space. Here, we formulate assumptions that guarantee the weak convergence of  $\alpha_{n,w}$  to a Gaussian limit in  $\ell^\infty(\Pi)$ , the space of all complex-valued functions that are uniformly bounded on any compact subset of  $\Pi \subset \mathbb{R}^q$ . Throughout the article  $\Pi_c$  denotes a general compact subset of  $\Pi$ . Of course, the sample paths of  $\alpha_{n,w}$  are usually contained in much a smaller space (such as the *cadlag* space  $D(\Pi)$ ), but as long as this space is equipped with the supremum metric, this is irrelevant for the weak convergence theorem. For the indicator family  $w(X_t, x) = 1(X_t \leq x)$  our assumptions are weaker than those considered in related weak convergence theorems, and similar to the mildest obtained in the i.i.d. case, see Stute (1997). More concretely, for non-smooth functions only finite second moments and a mild smooth condition on a conditional distribution is necessary. For the family  $w(I_{t-1}, x) = 1(\beta' I_{t-1} \leq u)$ , our results are new and are similar to those obtained in the i.i.d. setup by Escanciano (2006). For i.i.d. sequences, our assumptions reduce to those considered by Stute (1997) and Escanciano (2006). The weak convergence theorem that we give is founded on results by Levental (1989), Bae and Levental (1995) and Nishiyama (2000). From now on, the symbol  $\Rightarrow$  denotes weak convergence on compacta in  $\ell^\infty(\Pi)$ , see Definition 1.3.3 and Chapter 1.6 in van der Vaart and Wellner (1996, hereafter VW).

Define 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  $\Pi$  as

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

Then we need the following assumptions.

**Assumption W1.** For each  $n \geq 1$ ,  $\{(\varepsilon_{n,t}, X'_{n,t})' : 1 \leq t \leq n\}$  is a strictly stationary and ergodic process with  $E[\varepsilon_{n,t} | \mathcal{F}_{n,t-1}] = 0$  a.s.,  $\varepsilon_{n,t}$  is  $\mathcal{F}_{n,t}$ -measurable and  $0 < E\varepsilon_{n,t}^2 < \infty$ ,  $\forall t, 1 \leq t \leq n$ . Also, there exists a function  $K_w(x_1, x_2)$  on  $\Pi \times \Pi$  to  $\mathbb{R}$  such that, uniformly in  $(x_1, x_2) \in \Pi_c \times \Pi_c$ ,

$$n^{-1} \sum_{t=1}^n \varepsilon_{n,t}^2 w(X_{n,t-1}, x_1) w^c(X_{n,t-1}, x_2) = K_w(x_1, x_2) + o_{P_n}(1).$$

**Assumption W2.** For every compact subset  $\Pi_c$ , the family  $w(\cdot, x)$  is uniformly bounded (a.s) on  $\Pi_c$  and for every  $\varepsilon \in (0, 1)$  of the form  $\varepsilon = 2^{-q}$  there exists a finite partition  $\mathcal{P}_q = \{H_k; 1 \leq k \leq N_\varepsilon\}$  of  $\Pi_c$ , with  $N_\varepsilon$  the elements of such

partition, such that

$$\int_0^1 \sqrt{\log(N_\varepsilon)} d\varepsilon < \infty, \quad (5)$$

$$\sup_{q \in N} \frac{\alpha_{n,w}(\mathcal{P}_q)}{2^{-2q}} = O_{P_n}(1). \quad (6)$$

Let  $\alpha_{n,w}(\cdot)$  be a Gaussian process with zero mean and covariance function given by  $K_w(x_1, x_2)$ .

**Theorem 1.** *If Assumptions W1 and W2 hold, then  $\alpha_{n,w} \Rightarrow \alpha_{\infty,w}$ .*

Condition W1 allows us to establish the convergence of the finite-dimensional distributions of  $\alpha_{n,w}$ , whereas W2 is responsible for the asymptotic tightness. Loosely speaking, (5) controls the size of the family  $\{w(\cdot, x) : x \in \Pi\}$ , and (6) allows us to control the suprema over the links in a chaining argument by means of Freedman's (1975) inequality after applying a truncation at level  $\sqrt{n}$ , see (11).

Now we show that assumption W2 is satisfied (under W1 and some mild conditions) for all the families  $w$  considered in the literature. We start with the smooth case. Note that under W1 and for smooth functions  $w(X_{n,t-1}, x)$  satisfying

$$|w(X_{n,t-1}, x_1) - w(X_{n,t-1}, x_2)| \leq K_{n,t} \rho(x_1, x_2),$$

where  $\rho(\cdot, \cdot)$  such that  $(\Pi_c, \rho)$  is a totally bounded metric space and  $K_{n,t}$  is, for each  $n \geq 1$ , a stationary process with  $E[\varepsilon_{n,t}^2 K_{n,t}^2] < \infty$ ,  $\forall t, 1 \leq t \leq n$ , 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  $w(X_{t-1}, x) = \exp(ix'X_{t-1})$ ,  $w(X_{t-1}, x) = \sin(x'X_{t-1})$  or  $w(X_{t-1}, x) = 1/(1 + \exp(c - x'X_{t-1}))$ ,  $c \in \mathbb{R}$ .

For non-smooth functions, such as  $w(X_t, x) = 1(X_{t-1} \leq x)$  or  $w(X_{t-1}, x) = 1(\beta'X_{t-1} \leq u)$ ,  $x = (\beta', u)'$ , the situation is much more involved. For  $w(X_{t-1}, x) = 1(X_{t-1} \leq x)$ , Koul and Stute (1999) proved the weak convergence of the process  $\alpha_{n,w}$  for  $d = 1$  under slightly more than fourth moment, Markov and bounded densities assumptions, see Domínguez and Lobato (2003) for the case  $d > 1$ . To the best of our knowledge, these are the weakest assumptions in the literature for the stationary and ergodic case. The fourth moment assumption can be restrictive in applications, for instance it rules out most empirically relevant conditional heteroskedastic processes whose fourth moments are often infinite. For  $w(X_{t-1}, x) = 1(\beta'X_{t-1} \leq u)$ ,  $x = (\beta', u)'$ , Escanciano (2006) proved a weak

convergence theorem in an i.i.d. setup using the techniques of VW. These techniques cannot be applied directly in a time series context. The next result is an application of Theorem 1 to these particular weighting functions and provides an improvement of Lemma 3.1 in Koul and Stute (1999) and an extension to time series of Escanciano (2006). We need some further notation and an assumption. Define the semimetric

$$d_w(x_1, x_2) = \lim_{n \rightarrow \infty} \left( n^{-1} \sum_{t=1}^n E[\varepsilon_{n,t}^2 \{w(X_{n,t-1}, x_1) - w(X_{n,t-1}, x_2)\}^2] \right)^{\frac{1}{2}} \quad (7)$$

and let  $G_{n,t,w}(x) = E[E[\varepsilon_{n,t}^2 | X_{n,t-1}]w(X_{n,t-1}, x) | \mathcal{F}_{n,t-2}]$ .

**Assumption W3.**  $|G_{n,t,w}(x_1) - G_{n,t,w}(x_2)| \leq M_{n,t} d_w^2(x_1, x_2)$ , for each  $(x_1, x_2) \in \Pi \times \Pi$ , where  $M_{n,t}$  is, for each  $n \geq 1$ , a stationary process with  $E[|M_{n,t}|] < \infty$ ,  $\forall t, 1 \leq t \leq n$ .

Assumption W3 is sufficient for W2 (under W1) for the weighting functions  $w(\cdot, x) = 1(\cdot \leq x)$  or  $w(\cdot, x) = 1(\beta' \cdot \leq u)$ . Notice that W3 is a mild condition. For i.i.d. sequences it is trivially satisfied, whereas for time series sequences it follows if, for instance, the conditional density of  $X_{n,t-1}$  given  $\mathcal{F}_{n,t-2}$ ,  $f_{n,\mathcal{F}_{n,t-2}}(\cdot)$  say, satisfies (a.s)  $f_{n,\mathcal{F}_{n,t-2}}(\cdot) \leq C f_n(\cdot)$  for some  $C > 0$ , where  $f_n(\cdot)$  is the density of  $X_{n,t-1}$ .

**Corollary 1.** *Under W1 and W3 the weak convergence of Theorem 1 holds for  $w(\cdot, x) = 1(\cdot \leq x)$  and  $w(\cdot, x) = 1(\beta' \cdot \leq u)$ .*

## 2.2. Asymptotic distribution under the null

We establish the limit distribution of  $R_{n,w}^1$  under the null hypothesis  $H_0$ . To derive asymptotic results we consider the following assumptions. First, define the score  $g(I_{t-1}, \theta_0) = (\partial/\partial\theta')f(I_{t-1}, \theta_0)$ . Let  $\mathcal{F}_t = \sigma(I'_t, I'_{t-1}, \dots, I'_0)$ , be the  $\sigma$ -field generated by the information set obtained up to time  $t$ . Recall that  $\varepsilon_t = Y_t - E[Y_t | I_{t-1}]$  and  $e_t(\theta_0) = \varepsilon_t$  a.s.

**Assumption A1.**

A1(a):  $\{Y_t, I'_{t-1}\}' : t = 0, \pm 1, \pm 2, \dots\} \in S$  with joint cumulative distribution function (cdf)  $F(\cdot)$  and marginal distributions  $F_Y(\cdot)$  and  $F_I(\cdot)$ , respectively.

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

**Assumption A2.**  $f(\cdot, \theta)$  is twice continuously differentiable in a neighborhood of  $\theta_0 \in \Theta$ . There exists a function  $M(I_{t-1})$  with  $|g(I_{t-1}, \theta)| \leq M(I_{t-1})$ , such that  $M(I_{t-1})$  is  $F_I(\cdot)$ -integrable.

**Assumption A3.**

- A3(a): The parameter space  $\Theta$  is compact in  $\mathbb{R}^p$ . The true parameter  $\theta_0$  belongs to the interior of  $\Theta$ . There exists a unique  $\theta_1$  such that  $|\theta_n - \theta_1| = o_{P_n}(1)$ .
- A3(b): Under  $H_0$ ,  $\sqrt{n}(\theta_n - \theta_0) = n^{-1/2} \sum_{t=1}^n e_t(\theta_0)k(I_{t-1}, \theta_0) + o_P(1)$ , where  $k(\cdot)$  is such that  $L(\theta_0) = E[e_t^2(\theta_0)k(I_{t-1}, \theta_0)k'(I_{t-1}, \theta_0)]$  exists and is positive definite.

**Assumption A4.**

- A4(a): The weighting function  $w(\cdot)$  is such that the equivalence in (2) holds. For any compact set  $\Pi_c$  of  $\Pi$ ,  $w(I_{t-1}, x)$  is uniformly bounded (a.s) on  $\Pi_c$ , and satisfies, under the null, the assumption W2 above.
- A4(b):  $w(I_{t-1}, x)$  satisfies the uniform law of large numbers (ULLN)

$$\sup_{x \in \Pi_c} \left| n^{-1} \sum_{t=1}^n \varepsilon_t w(X_t, x) - E[\varepsilon_t w(X_t, x)] \right| \xrightarrow{as^*} 0,$$

whenever  $Z = \{(\varepsilon_t, X'_t)', t = 0, \pm 1, \dots\} \in S$ .

- A4(c): The integrating function  $\Psi(\cdot)$  is a probability distribution function which is absolutely continuous with respect to Lebesgue measure.

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 and allows for most empirically relevant conditional heteroskedastic models. Assumption A2 is classical in the model checks literature, see, e.g., Stute and Zhu (2002). Assumption A3 is satisfied, for instance, for the nonlinear least squares estimator (NLSE), or its robust modifications (under further regularity assumptions), see Koul (2002). The assumption that  $w$  satisfies (2) is needed only for the consistency of the tests against nonparametric alternatives. W2 usually holds under previous assumptions, see Section 2.1. The ULLN in A4(b) usually follows from the Ergodic Theorem and a Glivenko-Cantelli argument. Note that under A4,  $R_{n,w}^1$  can be viewed as a random element with values in  $\ell^\infty(\Pi)$ . The choice of  $\Psi(\cdot)$  depends on the space  $\Pi$  and  $w$ , and is crucial for the power properties of the CvM test. Some discussion about the choice of  $\Psi(\cdot)$  for a given  $w$  can be found in a similar context in Escanciano and Velasco (2003). A4(c) is only needed for the consistency of the CvM tests.

Under A1, using the Central Limit Theorem (CLT) for stationary ergodic martingale difference sequences, cf., Billingsley (1961), we have that the finite-dimensional distributions of  $R_{n,w}$  converge to those of a multivariate normal distribution with a zero mean vector and variance-covariance matrix given by the covariance function

$$K_w(x_1, x_2) = E[\varepsilon_t^2 w(I_{t-1}, x_1) w^c(I_{t-1}, x_2)]. \quad (8)$$

The next result is an extension of this convergence to weak convergence in the space  $\ell^\infty(\Pi)$ .

**Theorem 2.** *Under the null hypothesis  $H_0$ , A1 and A4(a),  $R_{n,w} \Rightarrow R_{\infty,w}$ , where  $R_{\infty,w}(\cdot)$  is a Gaussian process with zero mean and covariance function given by (8).*

In practice,  $\theta_0$  is unknown and has to be estimated from a sample  $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$  by an estimator  $\hat{\theta}_n$ . The next result shows the effect of the parameter uncertainty on the asymptotic null distribution of  $R_{n,w}^1$ . To this end, define the function  $G_w(x) \equiv G_w(x, \theta_0) = E[g(I_{t-1}, \theta_0)w(I_{t-1}, x)]$  and let  $V$  be a normal random vector with zero mean and variance-covariance matrix given by  $L(\theta_0)$ .

**Theorem 3.** *Under the null hypothesis  $H_0$  and Assumptions A1–A3 and A4(a-b),  $R_{n,w}^2(\cdot) \Rightarrow R_{\infty,w}(\cdot) - G'_w(\cdot)V \equiv R_{\infty,w}^2(\cdot)$ , where  $R_{\infty,w}$  is the same process as in Theorem 2 and  $\text{Cov}(R_{\infty,w}(x_1), V) = E[\varepsilon_t^2 k(I_{t-1}, \theta_0)w(I_{t-1}, x_1)]$ . Consequently, for any continuous functional  $\Gamma(\cdot)$ ,  $\Gamma(R_{n,w}^1) \xrightarrow{d} \Gamma(R_{\infty,w}^1)$ .*

Notice that the integrating measure  $\Psi(dx)$  in  $CvM_{n,w}$  can be chosen as a random measure  $\Psi_n(x)$ , say. An application of Lemma 3.1 in Chang (1990) shows that this choice does not change the asymptotic theory for  $CvM_{n,w}$  as long as  $\Psi_n(x)$  converges uniformly a.s. on compacta to a measure satisfying A4(c), i.e.,  $\sup_{x \in \Pi_c} |\Psi_n(x) - \Psi(x)| \rightarrow 0$  a.s., for every compact set  $\Pi_c \subset \Pi$ . This is the case for the family  $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$ , where the natural integrating measure is the empirical distribution function of the information sample  $\{I_{t-1} : 1 \leq t \leq n\}$ , or for  $w(I_{t-1}, x) = 1(\beta'I_{t-1} \leq u)$  where the integrating measure is the product of  $F_{n,\beta}(du)$  and  $d\beta$ , the empirical distribution function of the series  $\{\beta'I_{t-1} : 1 \leq t \leq n\}$  and the uniform density on the unit sphere, respectively, see Escanciano (2006). By the Glivenko-Cantelli Theorem for ergodic and stationary time series, see e.g., Dehling and Philipp (2002, p.4) and Wolfowitz (1954), the uniform convergence holds for  $1(I_{t-1} \leq x)$  and  $1(\beta'I_{t-1} \leq u)$ , respectively.

### 2.3. Consistency and local alternatives

In this section we study the consistency properties of the test based on continuous functionals  $\Gamma(R_{n,w}^1)$ . First, we show that these tests are consistent, that is, they are able to detect any alternative in  $S$ . Throughout the paper  $\xrightarrow{P^*}$  denotes convergence in outer probability, see Definition 1.9.1 in VW.

**Theorem 4.** *Under the alternative hypothesis  $H_A$  and A1–A4(a-b),  $n^{-1/2}R_{n,w}^1(\cdot) \xrightarrow{P^*} E[e_t(\theta_1)w(I_{t-1}, \cdot)]$ .*

Consequently, since  $H_A$  implies that  $E[e_t(\theta_1)w(I_{t-1}, \cdot)] \neq 0$  in a set with positive measure, the test statistic  $\Gamma(R_{n,w}^1)$  will diverge to  $+\infty$  under the alternative.

The next result shows the asymptotic distribution of  $R_{n,w}^1$  under a sequence of local alternatives converging to null at a parametric rate  $n^{-1/2}$ , say

$$H_{A,n} : Y_{t,n} = f(I_{t-1}, \theta_0) + \frac{a(I_{t-1})}{n^{1/2}} + \varepsilon_t, \quad a.s., \quad (9)$$

where the function  $a(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$  is  $F_I$ -measurable and  $F_I$ -integrable.

**Assumption A3'.** Under  $H_{A,n}$ ,  $\sqrt{n}(\theta_n - \theta_0) = \xi_a + n^{-1/2} \sum_{t=1}^n e_t(\theta_0)k(I_{t-1}, \theta_0) + o_P(1)$ , where the function  $k(\cdot)$  is as in A3 and  $\xi_a \in \mathbb{R}^p$ .

Take  $D_{w,a}(\cdot) = E[a(I_{t-1})w(I_{t-1}, \cdot)] - C'_w(\cdot)\xi_a$ .

**Theorem 5.** Under (9), Assumptions A1, A2, A3' and A4(a-b),  $R_{n,w}^1 \Rightarrow R_{\infty,w}^1 + D_{w,a}$ , where  $R_{\infty,w}^1$  is the process defined in Theorem 3.

Note that from (2),  $D_{w,a} = 0$  a.e.  $\Leftrightarrow a(I_{t-1}) = \xi'_a g(I_{t-1}, \theta_0)$  a.s. Therefore, for directions  $a(\cdot)$  not collinear to the score, the shift function  $D_{w,a}$  is non-trivial. As a matter of fact, it can be shown that for such alternatives the test based on any continuous even functional  $\Gamma(\cdot)$  is asymptotically unbiased.

### 3. Bootstrap Approximation of Residual Marked Empirical Processes

We have seen that the asymptotic null distribution of continuous functions of  $R_{n,w}^1$  depends in a complicated way on the data generating process (DGP) as well as the specification under the null hypothesis. Therefore, critical values for the tests statistics cannot be tabulated for general cases. A rather recently approach to solving this problem is that of Khmaladze and Koul (2004), who consider a martingale transformation of the process  $R_{n,w}^1$ , with  $w(I_{t-1}, B) = 1(I_{t-1} \in B)$ ,  $B$  a Borel set, that delivers asymptotically distributed-free tests. As they comment, their approach can be easily generalized to time series autoregressions. In fact, our theory can help to this end. Unfortunately, their approach is only useful for the indicator weighting family and, more importantly when  $d > 1$  the asymptotic null distribution of the transformed process still depends on the cdf  $F_I$ . Here, we propose a bootstrap method to solve the problem of approximating the asymptotic null distribution of an integrated-based test under time series and a general  $w$ . Resampling methods have been used extensively in the goodness-of-fit literature of regression models, see, e.g., Härdle and Mammen (1993), Stute, González-Manteiga and Presedo-Quindimil (1998) and Li, Hsiao and Zinn (2003) in the i.i.d. context, and Franke, Kreiss and Mammen (2002) for time series sequences. It is shown in these articles that the most relevant bootstrap method for regression problems is the wild bootstrap (WB) introduced in Wu (1986) and Liu (1988). Our approach is an

extension to nonlinear time series regressions of the WB approach. Other resampling schemes are of course possible in our context, e.g., the stationary bootstrap of Politis and Romano (1994). More concretely, we approximate the asymptotic null distribution of  $R_{n,w}^1$  by that of

$$R_{n,w}^{1*}(x) = n^{-\frac{1}{2}} \sum_{t=1}^n e_t^*(\theta_n^*) w(I_{t-1}, x) \quad x \in \Pi,$$

where the sequence  $\{e_t^*(\theta_n^*)\}_{t=1}^n$  are the fixed design wild bootstrap (FDWB) residuals obtained from the following algorithm.

- (1) Estimate the original model and obtain the residuals  $e_t(\theta_n)$  for  $t = 1, \dots, n$ .
- (2) Generate WB residuals according to  $e_t^*(\theta_n) = e_t(\theta_n)V_t$  for  $t = 1, \dots, n$ , where  $\{V_t : 1 \leq t \leq n\}$  is a sequence of independent random variables (r.v's) with zero mean, unit variance, bounded support, and independent of the sequence  $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$ .
- (3) Given  $\theta_n$  and  $e_t^*(\theta_n)$ , generate bootstrap data for the dependent variable  $Y_t^*$  as  $Y_t^* = f(I_{t-1}, \theta_n) + e_t^*(\theta_n)$  for  $t = 1, \dots, n$ .
- (4) Compute  $\theta_n^*$  from the data  $\{(Y_t^*, I'_{t-1})' : 1 \leq t \leq n\}$  and compute the residuals  $e_t^*(\theta_n^*) = Y_t^* - f(I_{t-1}, \theta_n^*)$  for  $t = 1, \dots, n$ .

Examples of  $\{V_t\}$  sequences are i.i.d. Bernoulli variates with

$$P(V_t = 0.5(1 - \sqrt{5})) = b, \quad P(V_t = 0.5(1 + \sqrt{5})) = 1 - b, \quad (10)$$

where  $b = (1 + \sqrt{5})/2\sqrt{5}$ , used in, e.g., Stute, González-Manteiga and Presedo-Quindimil (1998), or  $P(V_t = 1) = 0.5$  and  $P(V_t = -1) = 0.5$  as in Liu (1988). To justify this bootstrap approximation we need an additional assumption on the behaviour of the bootstrap estimator.

#### Assumption A5.

- A5(a): The estimator  $\theta_n^*$  satisfies  $\sqrt{n}(\theta_n^* - \theta_n) = n^{-1/2} \sum_{t=1}^n e_t^*(\theta_n) k(I_{t-1}, \theta_n) + o_P(1)$  a.s., where the function  $k(\cdot)$  is as in A3 with  
A5(b):  $|k(I_{t-1}, \theta)| \leq K(I_{t-1})$ , such that  $K(I_{t-1})$  is  $F_I$ -integrable.

This bootstrap procedure allows us to approximate the asymptotic critical values of the tests based on  $\Gamma(R_{n,w}^1)$ .

**Theorem 6.** *Assume A1–A5. Then, under the null Hypothesis  $H_0$ , under any fixed alternative hypothesis or under the local alternatives (9),  $R_{n,w}^{1*} \xrightarrow{*} \tilde{R}_{\infty,w}^1$ , a.s., where  $\tilde{R}_{\infty,w}^1$  is the Gaussian process of Theorem 3, but with  $\theta_1$  replacing  $\theta_0$ , and  $\xrightarrow{*}$  denoting weak convergence almost surely under the bootstrap law, see Giné and Zinn (1990).*

Note that under the null  $\theta_1$  coincides with  $\theta_0$ , and then  $\tilde{R}_{\infty,w}^1 = R_{\infty,w}^1$  in distribution. Therefore, we can approximate the asymptotic null distribution of the process  $R_{n,w}^1$  by that of  $R_{n,w}^{1*}$ . In particular, we can simulate the critical values for the tests statistics  $D_n = \Gamma(R_{n,w}^1)$  by the usual bootstrap algorithm. Section 4 below shows that this bootstrap procedure provides good approximations in finite samples.

#### 4. Finite Sample Performance and Empirical Application

In order to examine the finite sample performance of some integrated-based tests we carry out a simulation experiment. We compare the CvM tests based on 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  $w(I_{t-1}, x=1(\beta' I_{t-1} \leq u))$ ,  $x = (\beta', u)'$ . Throughout the simulations  $I_t \equiv I_{t-1,P} = (Y_{t-1}, \dots, Y_{t-P})$  will be the information set at time  $t - 1$ .

Let  $F_{n,\beta,P}(u)$  be the empirical distribution function of the projected information set  $\{\beta' I_{t-1,P} : 1 \leq t \leq n\}$ . Escanciano (2006) proposed in the i.i.d. setup the CvM test

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

where  $R_{n,pro,P}^1(\beta, u) = \hat{\sigma}_e^{-1} n^{-1/2} \sum_{t=1}^n e_t(\theta_n) 1(\beta' I_{t-1,P} \leq u)$  and  $\hat{\sigma}_e^2 = n^{-1} \sum_{t=1}^n e_t^2(\theta_n)$ . For a simple algorithm to compute  $CVM_{n,pro,P}$  see Appendix B in Escanciano (2006).

Bierens (1982) proposed to use  $w(I_{t-1}, x) = \exp(iI_{t-1}'x)$  as the weighting function in (2) and  $\Psi(dx) = \phi(x)$ , where  $\phi(x)$  is the probability density function of the standard normal  $P$ -variate r.v, delivering the CvM test statistic

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

We also consider the two test statistics based on the usual indicator function,

$$\begin{aligned} CvM_{n,ind,P} &= \frac{1}{\hat{\sigma}_e^2 n^2} \sum_{j=1}^n \left[ \sum_{t=1}^n e_t(\theta_n) 1(I_{t-1,P} \leq I_{j-1,P}) \right]^2, \\ KS_{n,ind,P} &= \max_{1 \leq i \leq n} \left| \frac{1}{\hat{\sigma}_e \sqrt{n}} \sum_{t=1}^n e_t(\theta_n) 1(I_{t-1,P} \leq I_{i-1,P}) \right|. \end{aligned}$$

Note that,  $CvM_1$  and  $PCvM_{n,1}$  are the same test statistic by definition.

We use the FDWB approximation for all the test statistics. Our theory applies to these test statistics and, in particular, our Theorem 6 validates these bootstrap approximations. In the sequel  $\varepsilon_t \sim i.i.d. N(0, 1)$ . Our models are

motivated by the well-studied Canadian lynx data set. Moran (1953) fitted an AR(2) model  $Y_t = a + bY_{t-1} + cY_{t-2} + \varepsilon_t$  to these data. We examine the adequacy of this model under the following DGP.

1. An AR(2) model:  $Y_t = 1.05 + 1.41Y_{t-1} - 0.77Y_{t-2} + \varepsilon_t$ .
2. An AR(2) with heteroskedasticity (ARHET):  $Y_t = 1.05 + 1.41Y_{t-1} - 0.77Y_{t-2} + h_t\varepsilon_t$ , where  $h_t^2 = 0.1 + 0.1Y_{t-1}^2$ .
3. An AR(3) model:  $Y_t = 1.05 + 1.41Y_{t-1} - 0.77Y_{t-2} + 0.33Y_{t-3} + \varepsilon_t$ .
4. ARMA(2, 2) model:  $Y_t = 1.05 + 1.41Y_{t-1} - 0.77Y_{t-2} + 0.33\varepsilon_{t-1} + 0.21\varepsilon_{t-2} + \varepsilon_t$ .
5. A TAR(2) model:  $Y_t = \begin{cases} 0.62 + 1.25Y_{t-1} - 0.43Y_{t-2} + \varepsilon_t, & \text{if } Y_{t-2} \leq 3.25, \\ 2.25 + 1.52Y_{t-2} - 1.24Y_{t-2} + \varepsilon_t, & \text{if } Y_{t-2} > 3.25. \end{cases}$
6. An EXPAR(2) model:  $Y_t = a_t Y_{t-1} - b_t Y_{t-2} + 0.2\varepsilon_t$ , where  $a_t = 0.138 + (0.316 + 0.982Y_{t-1}) \exp(-3.89Y_{t-1}^2)$  and  $b_t = 0.437 + (0.659 + 1.260Y_{t-1}) \exp(-3.89Y_{t-1}^2)$ .

All the models except the ARHET have been fitted to the Canadian lynx data set, see Tong (1990) for a survey. We consider for the experiments under the null a sample size of  $n = 100$ , and under the alternative,  $n = 100$ , and 200. The number of Monte Carlo experiments is 1,000 and the number of bootstrap replications is  $B = 500$ . In all the replications 200 pre-sample data values of the processes were generated and discarded. For the information set we consider the values  $P = 3, 5$  and  $7$ . We employ a sequence  $\{V_t\}$  of i.i.d. Bernoulli variates as in (10).

In Table 1 we show the empirical rejection probabilities (RP) associated with the nominal level 5%. The results with other nominal levels are similar. The empirical levels of the test statistics are closed to the nominal level. Only the heteroskedastic case  $CvM_{n,\text{exp},P}$  presents some small size distortion (under-in-rejection).

Table 1. Size of 5% tests.

<b><math>n = 100</math></b>	<b>AR(2)</b>			<b>ARHET</b>			
	<b><math>P</math></b>	<b>3</b>	<b>5</b>	<b>7</b>	<b>3</b>	<b>5</b>	<b>7</b>
$CVM_{n,\text{pro},P}$		4.6	5.2	5.1	5.3	4.3	4.2
$CvM_{n,\text{exp},P}$		4.4	5.4	5.4	3.2	2.1	2.0
$CvM_{n,\text{ind},P}$		5.2	5.0	4.0	5.5	5.1	5.1
$KS_{n,\text{ind},P}$		4.7	5.4	5.1	6.0	7.4	5.5

In Table 2 we report the empirical power against the AR(3) and ARMA(2, 2) processes. The empirical power increases with the sample size  $n$  for all test statistics, as expected. It is shown that the Cramér-von Mises test  $CVM_{n,\text{pro},P}$  has the best empirical power in all cases. The empirical power of  $CvM_{n,\text{exp},P}$  is low for these alternatives and, in general, less than those of  $CvM_{n,\text{ind},P}$  and  $KS_{n,\text{ind},P}$ , especially for large  $P$ . In Table 3 we show the RP for the TAR

and EXPAR models. Again, the test statistic  $CVM_{n,pro,P}$  has more empirical power in almost all cases.  $CvM_{n,exp,P}$  has good empirical power properties for these models, in particular, it overtakes  $CvM_{n,ind,P}$  and  $KS_{n,ind,P}$ . For the last two models the empirical power of the test statistics  $CvM_{n,ind,P}$  and  $KS_{n,ind,P}$  decrease as the lag parameter  $P$  increases.

Table 2. Power of 5% tests.

$P$	AR(3)			ARMA(2, 2)		
	3	5	7	3	5	7
$n = 100$	$CVM_{n,pro,P}$	<b>61.9</b>	<b>56.4</b>	<b>40.7</b>	<b>30.7</b>	<b>46.7</b>
	$CvM_{n,exp,P}$	25.0	18.2	14.2	11.3	10.6
	$CvM_{n,ind,P}$	22.1	33.2	25.9	15.2	25.4
	$KS_{n,ind,P}$	22.0	26.3	17.7	13.7	21.9
$n = 200$	$CVM_{n,pro,P}$	<b>91.2</b>	<b>88.4</b>	<b>76.7</b>	<b>61.2</b>	<b>80.7</b>
	$CvM_{n,exp,P}$	45.9	32.3	21.8	16.4	12.6
	$CvM_{n,ind,P}$	34.3	54.7	46.1	24.1	49.1
	$KS_{n,ind,P}$	36.5	48.4	36.5	21.6	44.9

Table 3. Power of 5% tests.

$P$	TAR			EXPAR		
	3	5	7	3	5	7
$n = 100$	$CVM_{n,pro,P}$	<b>76.2</b>	<b>40.0</b>	<b>27.9</b>	<b>81.2</b>	57.1
	$CvM_{n,exp,P}$	52.2	31.3	19.4	73.1	<b>65.7</b>
	$CvM_{n,ind,P}$	60.9	4.9	9.0	72.7	11.5
	$KS_{n,ind,P}$	40.4	7.4	11.5	53.3	4.7
$n = 200$	$CVM_{n,pro,P}$	<b>98.2</b>	<b>79.3</b>	<b>57.4</b>	98.9	94.4
	$CvM_{n,exp,P}$	89.0	69.4	44.1	97.8	<b>96.6</b>
	$CvM_{n,ind,P}$	90.9	15.2	16.1	97.3	34.4
	$KS_{n,ind,P}$	73.6	22.0	27.6	89.3	12.8

The well-known Canadian lynx data set consists of the annual record of the Canadian lynx trapped in the Mackenzie River district of northwest Canada for the period 1821-1834 inclusive, with a total of 114 observations. For an exhaustive description of these data see Tong (1990, pp.357-418) and references therein. The first time series model built on this particular data set was probably that of Moran (1953), a linear AR(2) model to the logarithm of the lynx data,  $Y_t$ , say. We consider this specification under the null and the same implementation as in the Monte Carlo simulations, except that now we take  $P = 2, 4, 6$  and 10. We report the empirical p-values for the test statistics  $CVM_{n,pro,P}$ ,  $CvM_{n,exp,P}$ ,  $CvM_{n,ind,P}$  and  $KS_{n,ind,P}$  in Table 4.

Table 4. p-values for the Canadian lynx data.

<i>P</i>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>
$CVM_{n,pro,P}$	0.000	0.000	0.042	0.000
$CvM_{n,exp,P}$	0.002	0.000	0.016	0.000
$CvM_{n,ind,P}$	0.000	0.004	0.066	0.090
$KS_{n,ind,P}$	0.000	0.008	0.196	0.090

The AR(2) specification is rejected at 5% by  $CVM_{n,pro,P}$  and  $CvM_{n,exp,P}$  for all values of  $P$ , whereas  $CvM_{n,ind,P}$  and  $KS_{n,ind,P}$  fail to reject it for large values of  $P$  ( $P = 6$  and 10). Therefore, this specification is not satisfactory; this fact was realized by many authors, including Moran (1953). Two further specifications for this data set were considered in Tong (1990). First we consider the TAR(2, 2, 2) model given in DGP 5, with  $P = 2, 4, 6$  and 10, and we report the empirical p-values for the test statistics in Table 5.

Table 5. p-values for the Canadian lynx data TAR(2, 2, 2) model.

<i>P</i>	<b>2</b>	<b>4</b>	<b>6</b>	<b>10</b>
$CVM_{n,pro,P}$	0.322	0.008	0.086	0.012
$CVM_{n,exp,P}$	0.180	0.004	0.094	0.016
$CvM_{n,ind,P}$	0.174	0.002	0.098	0.114
$KS_{n,ind,P}$	0.408	0.022	0.182	0.140

For  $P = 2$  all the test statistics fail to reject the TAR(2, 2, 2) specification, whereas for  $P = 4$  and 10,  $CVM_{n,pro,P}$  and  $CvM_{n,exp,P}$  reject it at 5% level. The test statistics  $CvM_{n,ind,P}$  and  $KS_{n,ind,P}$  support the TAR(2, 2, 2) model for  $P = 4$  and 10, but this may be due to the curse of dimensionality and not because of a correct specification. The model selected by the AIC among some TAR models, see (Tong, 1990, p.387), is the following TAR(2, 7, 2):

$$Y_t = \begin{cases} 0.54 + 1.032Y_{t-1} - 0.173Y_{t-2} + 0.171Y_{t-3} - 0.431Y_{t-4} \\ + 0.332Y_{t-5} - 0.284Y_{t-6} + 0.210Y_{t-7}, & \text{if } Y_{t-2} \leq 3.116 \\ 2.25 + 1.52Y_{t-1} - 1.24Y_{t-2} + \varepsilon_t, & \text{if } Y_{t-2} > 3.116. \end{cases}$$

For this specification we consider  $P = 7, 8, 9$  and 10. The empirical p-values are reported in Table 6.

Table 6. p-values for the Canadian lynx data TAR(2, 7, 2) model.

<i>P</i>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
$CVM_{n,pro,P}$	0.990	0.974	0.958	0.966
$CVM_{n,exp,P}$	0.966	0.874	0.834	0.824
$CvM_{n,ind,P}$	0.674	0.178	0.476	0.600
$KS_{n,ind,P}$	0.342	0.132	0.390	0.486

The results in Table 6 show that the TAR(2, 7, 2) is a well-specified model for this data set. For this model  $CVM_{n,pro,P}$  has the highest p-value for all values of the lag parameter  $P$ . The worth of the TAR(2, 7, 2) model for the Canadian lynx data was pointed out by Tong (1990).

## Appendix A. Proofs

The next lemma corresponds to Theorem 1.5.4 and Theorem 1.5.6 of VW.

**Lemma A1.** *Let  $\mathcal{T}$  be a non-empty set. For every  $n$ , 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  $\mathcal{B} = \{T_k; 1 \leq k \leq N\}$  to  $\mathcal{T}$  such that

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

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), there exists a version  $\tilde{X}$  of  $X$  such that  $X_n \Rightarrow \tilde{X}$  in  $\ell^\infty(\mathcal{T})$ .

**Proof of Theorem 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 the previous lemma holds. To this end, fix a compact subset  $\Pi_c \subset \Pi$  and, using W2, choose a nested sequence of finite partitions  $\mathcal{P}_q = \{B_{qk}; 1 \leq k \leq N_q\}$  of  $\Pi_c$ ,  $q \geq 1$ , such that  $\sum_{q=1}^{\infty} 2^{-q} \sqrt{\log N_q} < \infty$ . 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

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

To simplify notation define  $M_t^n(x) = n^{-1/2} \varepsilon_{n,t} w(X_{n,t-1}, x)$ . Then, by Lemma A1, see also the proof of Theorem 2.5.6 of VW, it is sufficient to prove that for every  $\varepsilon, \eta > 0$  there exists a  $q_0$  such that

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

where  $\|\cdot\|_{\Pi_c}$  denotes the uniform norm on  $\Pi_c$ . To this end, fix any  $q_0$  for a while, and let us define 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 indicator functions  $C_{t,q-1}^n = 1(\Delta_t^n(B_{q_0}x) \leq a_{q_0}, \dots, \Delta_t^n(B_{q-1}x) \leq a_{q-1})$ ,  $D_{t,q}^n = 1(\Delta_t^n(B_{q_0}x) \leq a_{q_0}, \dots, \Delta_t^n(B_{q-1}x) \leq a_{q-1}, \Delta_t^n(B_q x) > a_q)$  and  $D_{t,q_0}^n = 1(\Delta_t^n(B_{q_0}x) > a_{q_0})$ . Now, similarly to VW p.131, we apply a truncation argument at level  $\sqrt{n}$  and write

$$\begin{aligned} M_t^n(x) - M_t^n(\pi_{q_0}x) &= (M_t^n(x) - M_t^n(\pi_{q_0}x))D_{t,q_0}^n + \sum_{q=q_0+1}^{\infty} (M_t^n(x) - M_t^n(\pi_q x))D_{t,q}^n \\ &\quad + \sum_{q=q_0+1}^{\infty} (M_t^n(\pi_q x) - M_t^n(\pi_{q-1}x))C_{t,q-1}^n. \end{aligned} \quad (11)$$

On the other hand, by (4),

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

Now, by (11) and the last display

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

where

$$\begin{aligned} I_1 &= \left\| \sum_{t=1}^n \Delta_t^n(B_{q_0}x) D_{t,q_0}^n \right\|_{\Pi_c}, \\ I_2 &= \left\| \sum_{t=1}^n E[\Delta_t^n(B_{q_0}x) D_{t,q_0}^n \mid \mathcal{F}_{n,t-1}] \right\|_{\Pi_c}, \\ II_1 &= \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \Delta_t^n(B_q x) D_{t,q}^n \right\|_{\Pi_c}, \\ II_2 &= \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} E[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t-1}] \right\|_{\Pi_c}, \\ III &= \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} (M_t^n(\pi_q x) - M_t^n(\pi_{q-1}x)) C_{t,q-1}^n \right. \\ &\quad \left. - E[(M_t^n(\pi_q x) - M_t^n(\pi_{q-1}x)) C_{t,q-1}^n \mid \mathcal{F}_{n,t-1}] \right\|_{\Pi_c}. \end{aligned}$$

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

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

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  $\Delta_t^n(B_{q_0} x) D_{t,q_0}^n \leq Cn^{-1/2} |\varepsilon_{n,t}| 1(|\varepsilon_{n,t}| > C\sqrt{n}a_{q_0})$  eventually. Then from (4) and W1 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 A2 in Stute, González-Manteiga and Presedo-Quindimil (1998).

By assumption W2, for any  $\eta > 0$  there exists a constant  $K = K_\eta > 0$ , such that  $\limsup_{n \rightarrow \infty} P(\Omega_n \setminus \Omega_K^n) \leq \eta$ , where  $\Omega_K^n = \{\sup_{q \in N} \alpha_{n,w}(\mathcal{B}_\varepsilon)/2^{-2q} \leq K\}$ . Then, for the estimation of  $II_2$ , we see that

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

As for  $II_3$ , since  $|\Delta_t^n(B_q x) D_{t,q}^n - E[\Delta_t^n(B_q x) D_{t,q}^n | \mathcal{F}_{n,t-1}]| \leq 2a_{q-1}$  eventually, and  $\sum_{t=1}^n E[|\Delta_t^n(B_q x)|^2 D_{t,q}^n | \mathcal{F}_{n,t-1}] \leq K 2^{-2q}$  a.s. on the set  $\Omega_K^n$ , it follows from the Freedman's (1975) inequality, which plays here the same role as the Bernstein's inequality does in the i.i.d. setup, and Lemma 2.11.17 of VW, that for any measurable set  $A$

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

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

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

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

$$\begin{aligned}
& E|II_3|1(\Omega_K^n) \\
& \leq E \left\| \sum_{t=1}^n \sum_{q=q_0+1}^{\infty} \Delta_t^n(B_q x) D_{t,q}^n - E[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t-1}] \right\|_{\Pi_c} 1(\Omega_K^n) \\
& \leq \sum_{q=q_0+1}^{\infty} E \left\| \sum_{t=1}^n \Delta_t^n(B_q x) D_{t,q}^n - E[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t-1}] \right\|_{\Pi_c} 1(\Omega_K^n) \\
& \leq \sum_{q=q_0+1}^{\infty} \sum_{k=1}^{N_q} E \left| \sum_{t=1}^n \Delta_t^n(B_q x) D_{t,q}^n - E[\Delta_t^n(B_q x) D_{t,q}^n \mid \mathcal{F}_{n,t-1}] \right| 1(\Omega_{qk}^n \cap \Omega_K^n) \\
& \leq C(2 + \sqrt{K}) \sum_{q=q_0+1}^{\infty} \sum_{k=1}^{N_q} 2^{-q} \sqrt{\log(N_q)} \left( P(\Omega_{qk}^n) + \frac{1}{N_q} \right) \\
& \leq C(2 + \sqrt{K}) \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log(N_q)}.
\end{aligned}$$

Finally, the estimation of  $III$  follows from the same arguments as for  $II_3$ , and therefore we obtain  $E|III|$  as with  $II_3 \leq C(2 + \sqrt{K}) \sum_{q=q_0+1}^{\infty} 2^{-q} \sqrt{\log(N_q)}$ . The theorem follows from choosing a large  $K$ , a large  $q_0$ , and then letting  $n \rightarrow \infty$ .

**Proof of Corollary 1.** Throughout this proof  $w(\cdot, x) = 1(\cdot \leq x)$  or  $w(\cdot, x) = 1(\beta' \cdot \leq u)$ . Define the marked process  $\tilde{\alpha}_{n,w} = n^{-1} \sum_{t=1}^n E[\varepsilon_{n,t}^2 \mid X_{n,t-1}] w(X_{n,t-1}, x)$ , and the oscillation modulus of  $\tilde{\alpha}_{n,w}$  as  $\varpi_{n,d}(a) = \sup_{d^2(x,y) \leq a} |\tilde{\alpha}_{n,w}(x) - \tilde{\alpha}_{n,w}(y)|$ . Write

$$\begin{aligned}
\tilde{\alpha}_{n,w}(x) &= n^{-1} \sum_{t=1}^n \{E[\varepsilon_{n,t}^2 \mid X_{n,t-1}] w(X_{n,t-1}, x) - G_{n,t,w}(x)\} + n^{-1} \sum_{t=1}^n G_{n,t,w}(x) \\
&\equiv \tilde{\beta}_{n,w}(x) + F_{n,w}(x).
\end{aligned}$$

Hence, by triangle inequality,

$$|\tilde{\alpha}_{n,w}(x) - \tilde{\alpha}_{n,w}(y)| \leq |\tilde{\beta}_{n,w}(x) - \tilde{\beta}_{n,w}(y)| + |F_{n,w}(x) - F_{n,w}(y)|. \quad (12)$$

Notice that  $\{\tilde{\beta}_{n,w}(x), \mathcal{F}_{n,t-2}\}$  is a martingale for each  $x \in \Pi$ , by construction. By a truncation argument, it can be assumed without loss of generality and by W1 that  $|\varepsilon_{n,t}| \leq Cn^{1/2}$  on a set with arbitrarily large probability, see (11). Hence, on that set  $E[\varepsilon_{n,t}^2 \mid X_{n,t-1}] \leq Cn$ . Now, Freedman's (1975) inequality and Lemma 2.2.10 in VW, jointly with fact that  $\{\cdot w(\cdot, x) : x \in \Pi\}$  is a VC class, yield

$$E \sup_{d_w^2(x,y) \leq n^{-\frac{1}{2}}} |\tilde{\beta}_{n,w}(x) - \tilde{\beta}_{n,w}(y)| = O_{P_n}(n^{-\frac{1}{4}}).$$

On the other hand, W3 implies that  $\sup_{d_w^2(x,y) \leq n^{-1/2}} |F_{n,w}(x) - F_{n,w}(y)| = O_{P_n}(n^{-1/4})$ . Hence for  $a_n = n^{-1/2}$ ,  $\varpi_{n,d_w}(a_n) = O_{P_n}(a_n^2)$ . Take a partition of  $\Pi$  in  $\varepsilon$ -brackets  $H_k = [x_k, y_k]$  with respect to the semimetric  $d_w(x_1, x_2)$ . Then (5) follows because the class is VC, and (6) follows from the previous arguments.

**Proof of Theorem 2.** It follows from Theorem 1.

**Proof of Theorem 3.** Applying the classical mean value theorem argument we have

$$R_{n,w}^1(x) = R_{n,w}(x) - n^{1/2}(\theta_n - \theta_0)'(I - II - III),$$

where  $I = n^{-1} \sum_{t=1}^n \{g(I_{t-1}, \theta_{ni}) - g(I_{t-1}, \theta_0)\}w(I_{t-1}, x)$ ,  $II = n^{-1} \sum_{t=1}^n [g(I_{t-1}, \theta_0)w(I_{t-1}, x) - G_w(x, \theta_0)]$ , and  $III = G_w(x, \theta_0)$ , where  $\theta_{ni}$  satisfies  $|\theta_{ni} - \theta_0| \leq |\theta_n - \theta_0|$  a.s.. By A1–A3, A4(b) and the uniform law of large numbers (ULLN) of Jennrich (1969, Theorem 2), it is easy to show that  $I = o_P(1)$  and  $II = o_P(1)$  uniformly in  $x \in \Pi_c$ . So, the theorem follows from Theorem 2 and A3(b).

**Proof of Theorem 4.** From A4(b), under A1–A3, uniformly in  $x \in \Pi_c$ ,

$$\frac{1}{n} \sum_{t=1}^n [e_t(\theta_n)w(I_{t-1}, x) - E[e_t(\theta_1)w(I_{t-1}, x)]] = o_P(1).$$

**Proof of Theorem 5.** Under the local alternatives (9) write

$$R_{n,w}^1(x) = R_{n,w}(x) + A_1 + A_2, \quad (13)$$

with  $A_l = n^{-1/2} \sum_{t=1}^n \{f(I_{t_1}, \theta_0) - f(I_{t-1}, \theta_n)\}w(I_{t-1}, x)$  and  $A_2 = n^{-1} \sum_{t=1}^n a(I_{t-1})w(I_{t-1}, x)$ . Using A3' as in Theorem 3, we obtain  $|A_1 + n^{1/2}(\theta_n - \theta_0)'G_w(x, \theta_0)| = o_P(1)$ , uniformly in  $x \in \Pi_c$ . On the other hand, A4(b) yields that, uniformly in  $x \in \Pi_c$ ,  $A_2 - E[a(I_{t-1})w(I_{t-1}, x)] = o_P(1)$ . Using the preceding and (13), the theorem holds by A3' and Theorem 3.

**Proof of Theorem 6.** We need to show that the process  $R_{n,w}^{1*}$  (conditionally on the sample) has the same asymptotic finite dimensional distributions as the process  $R_{n,w}^1$ , with  $\theta_1$  replacing  $\theta_0$ , and that  $R_{n,w}^{1*}$  is asymptotically tight, both with probability one. Then, similarly to Theorem 3 we obtain, uniformly in  $x \in \Pi_c$ ,

$$R_{n,w}^{1*}(x) = R_{n,w}^*(x) - n^{\frac{1}{2}}(\theta_n^* - \theta_n)'G_w(x, \theta_1) + o_P(1)\text{a.s.}.$$

The convergence of the finite-dimensional distributions follows from the last expression, A5 and from the Cramér-Wold device. The tightness (a.s.) follows from the same steps as in Theorem 1 in Stute, González-Manteiga and Presedo-Quindimil (1998). The proof is finished.

### Acknowledgements

I would like to thank my thesis advisor, Carlos Velasco, for his guidance in this work. I would also like to thank Miguel A. Delgado and Efstathios Paparoditis for helpful comments and suggestions, and Winfried Stute for his hospitality during my visit to Giessen's University. Any errors are solely mine. I would like to thank two referees, an associate editor and Co-editor Jane-Ling Wang for very helpful comments and suggestions. Research funded by the Spanish Dirección General de Enseñanza Superior (DGES) reference numbers SEJ2004-04583/ECON and SEJ2005-07657/ECON.

### References

- Bae, J. and Levental, S. (1995). Uniform CLT for Markov chains and its invariance principle: a martingale approach. *J. Theoret. Probab.* **8**, 549-570.
- Bierens, H. J. (1982). Consistent model specification tests. *J. Econometrics* **20**, 105-134.
- Bierens, H. J. (1984). Model specification testing of time series regressions. *J. Econometrics* **26**, 323-353.
- Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443-1458.
- Bierens, H. J. and Ploberger, W. (1997). Asymptotic theory of integrated conditional moment test. *Econometrica* **65**, 1129-1151.
- Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proc. Amer. Math. Soc.* **12**, 788-792.
- Chang, N. M. (1990). Weak convergence of a self-consistent estimator of a survival function with doubly censored data. *Ann. Statist.* **18**, 391-404.
- Chen, S. X., Härdle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *J. Roy. Statist. Soc. Ser. B* **65**, 663-678.
- de Jong, R. M. (1996). The Bierens' tests under data dependence. *J. Econometrics* **72**, 1-32.
- Dehling, H. and Philipp, W. (2002). Empirical process techniques for dependent data. In *Empirical processes techniques for dependent data*. (Edited by H. Dehling, T. Mikosch and M. Sørensen), 1-115. Birkhauser Boston.
- Domínguez, M. and Lobato, I. N. (2003). A consistent test for the martingale difference hypothesis. *Econom. Rev.* **22**, 351-377.
- Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22**, 1030-1051.
- Escanciano, J. C. (2006). Goodness-of-fit tests for linear and non-linear time series models. *J. Amer. Statist. Assoc.* **101**, 531-541.
- Escanciano, J. C. and Velasco, C. (2006). Generalized spectral tests for the martingale difference hypothesis. *J. Econometrics* **134**, 151-185.
- Eubank, R. and Hart, J. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20**, 1412-1425.
- Eubank, R. and Spiegelman, S. (1990). Testing the goodness of fit of a linear model via non-parametric regression techniques. *J. Amer. Statist. Assoc.* **85**, 387-392.

- Fan, J. and Huang, L. (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96**, 640-652.
- Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables, parametric and semiparametric functional forms. *Econometrica* **64**, 865-890.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.
- Franke, J., Kreiss, J. P. and Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8**, 1-37.
- Freedman, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3**, 100-118.
- Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18**, 851-869.
- Guerre, E. and Guay, A. (2005). A data-driven nonparametric specification test for dynamic regression models. *Econom. Theory*. To appear.
- Guerre, E. and Lavergne, P. (2005). Rate-optimal data-driven specification testing for regression models. *Ann. Statist.* **33**, 840-870.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926-1974.
- Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer Verlag, New York.
- Hong, Y. and Lee, Y. J. (2005). Generalized spectral tests for conditional mean models in time series with conditional heteroskedasticity of unknown form. *Rev. Econom. Stud.* **72**, 499-541.
- Hong, Y. and White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica* **63**, 1133-1159.
- Horowitz, J. L. and Härdle, W. (1994). Testing a parametric model against a semiparametric alternative. *Econom. Theory* **10**, 821-848.
- Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean regression model against a nonparametric alternative. *Econometrica* **69**, 599-632.
- Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40**, 633-643.
- Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.* **32**, 995-1034.
- Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Model*. 2nd edition. Lecture Notes in Statistics, Vol. 166, Springer.
- Koul, H. L. and Ni, P. (2004). Minimum distance regression model checking. *J. Statist. Plann. Inference* **119**, 109-144.
- Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27**, 204-236.
- Levental, S. (1989). A uniform CLT for uniformly bounded families of martingale differences. *J. Theoret. Probab.* **2**, 271-287.
- Li, Q. (1999). Consistent model specification test for time series econometric models. *J. Econometrics* **92**, 101-147.
- Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression functional form. *J. Econometrics* **87**, 145-165.
- Li, Q., Hsiao, C. and Zinn, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *J. Econometrics* **112**, 295-325.

- Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16**, 1696-1708.
- Moran, P. A. P. (1953). The statistical analysis of the Canadian lynx cycle, I: structure and prediction. *Aust. J. Zool.* **1**, 163-173.
- Nishiyama, Y. (2000). Weak convergence of some classes of martingales with jumps. *Ann. Probab.* **28**, 685-712.
- Politis, D. and Romano, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statist. Sinica* **4**, 461-476.
- Stinchcombe, M. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econom. Theory* **14**, 295-325.
- Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-641.
- Stute, W., González-Manteiga, W. and Presedo-Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93**, 141-149.
- Stute, W., Thies, S. and Zhu, L. X. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.* **26**, 1916-1934.
- Stute, W. and Zhu L. X. (2002). Model checks for generalized linear models. *Scand. J. Statist.* **29**, 535-545.
- Tong, H. (1990). *Nonlinear Time Series: A Dynamic System Approach*. Clarendon Press, Oxford.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- Wolfowitz, J. (1954). Generalization of the theorem of Glivenko-Cantelli. *Ann. Math. Statist.* **25**, 131-138.
- Wooldridge, J. (1992). A test for functional form against nonparametric alternatives. *Econom. Theory* **8**, 452-475.
- Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1261-1350.
- Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econom. Theory* **8**, 435-451.
- Zheng, X. (1996). A consistent test of functional form via nonparametric estimation technique. *J. Econometrics* **75**, 263-289.
- Zhu, L. X. (2003). Model cheking of dimension reduction type for regression. *Statist. Sinica* **13**, 283-296.
- Zhu, L. X. and Ng, K. W. (2003). Checking the adequacy of a partial linear model. *Statist. Sinica* **13**, 763-781.

Department of Economics, Indiana University, Wylie Hall, 100 S. Woodlawn, Bloomington, IN, 47405, U.S.A.

E-mail: jescanci@indiana.edu

(Received March 2005; accepted December 2005)