



Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Working Paper nº 13/04

Model Checks Using Residual Marked Empirical
Processes

J. Carlos Escanciano

Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

MODEL CHECKS USING RESIDUAL MARKED EMPIRICAL PROCESSES

J. Carlos Escanciano
Working Paper No.13/04
September 2004
JEL No. C12.

ABSTRACT

This paper proposes omnibus and directional tests for testing the goodness-of-fit of a parametric regression time series model. We use a general class of residual marked empirical processes as the building-blocks for estimation and testing of such models. First, we establish a weak convergence theorem under mild assumptions, which allows us to study in a unified way the asymptotic null distribution of the test statistics and their asymptotic behavior against Pitman's local alternatives. To approximate the asymptotic null distribution of test statistics we justify theoretically a bootstrap procedure. Also, some asymptotic theory for the estimation of the principal components of the residual marked processes is considered. This asymptotic theory is used to derive optimal directional tests and efficient estimation of regression parameters. 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.

Juan Carlos Escanciano Reyero
Universidad de Navarra
Departamento de Métodos Cuantitativos
Campus Universitario
31080 Pamplona
jescanci@unav.es

Acknowledgments

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. Any errors are solely mine. This work is part of my PhD. Thesis at Universidad Carlos III de Madrid. Research funded by the Spanish Dirección General de Enseñanza Superior (DGES) reference number BEC2001-1270.

1. INTRODUCTION

The purposes of the present paper are threefold; first, we elaborate a new asymptotic theory (weak convergence theorem) for a general class of marked processes that arise in model checks, second, we use this new asymptotic theory to construct consistent and asymptotically normal estimation of parameters in time series regression models, and third, we study a general method for testing the goodness-of-fit of such models. The building-blocks for both inference problems, estimation and testing, are a general class of residual marked processes. All tests considered in this paper, omnibus and directional, are functionals of these residual marked processes. The new asymptotic theory allows us to study, in a unified way, the asymptotic distribution of the test statistics under the null and under Pitman's local alternatives converging to the null at the parametric rate $n^{-1/2}$, with n the sample size. In the omnibus case, the asymptotic null distributions of the tests depend on certain features of the data generating process and the specification under the null, making the implementation of the asymptotic results difficult in practice. In order to estimate the asymptotic critical values, we extend to time series regressions a resampling procedure based on the wild-bootstrap. Also, we shall show that the principal components of the residual marked processes play an important role in the design of optimal directional tests and efficient estimation of regression parameters. To establish the asymptotic theory for such optimal inferences, we need to provide estimations and consistency results for the principal components. The methodology of this paper can be also extended to other conditional moment restrictions tests, in particular, to model checks for joint conditional mean and variance specifications, which are of great interest in, e.g., financial applications.

Parametric time series modelling continues to be a major interest of social and natural scientists. These models permit that behavioral patterns potentially can be explained or predicted simply by studying the past history of a variable and/or employing the relationships between several variables. An important task of the time series analyst consists in examining the intrinsic nature of the type and form of the dependence between values at different time points, and to try to construct statistical models that can reproduce this dependency. To this end, much of the existing literature is concerned with the parametric modelling of such dependencies in terms of the conditional mean function of the response variable $Y_t \in \mathbb{R}$, given some information set at time $t - 1$, $I_{t-1} \in \mathbb{R}^d$, $d \in \mathbb{N}$, say. More specifically, given the strictly stationary time series process $\{(Y_t, I_{t-1}')' : t = 0, \pm 1, \pm 2, \dots\}$, under integrability of Y_t , we can write the tautological expression

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

where $f(z) := E[Y_t | I_{t-1} = z]$, $z \in \mathbb{R}^d$, is the conditional mean almost surely (a.s.) of Y_t given

the information set I_{t-1} , and $\varepsilon_t := Y_t - E[Y_t | I_{t-1}]$ is, by construction, independent in conditional mean with respect to \mathcal{F}_{t-1} , the σ -field generated by I_{t-1} .

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

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

with $f(I_{t-1}, \theta)$ a parametric specification for the conditional mean $f(I_{t-1})$, and $\{e_t(\theta) : t = 0, \pm 1, \pm 2, \dots\}$ a sequence of random variables (r.v), deviations of the model. Parametric time series regression models continue to be attractive among practitioners because the parameter θ 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 specifications (1) include linear and nonlinear autoregressive models, such as Markov-switching, exponential or threshold autoregressive models among many others, see Tong (1990), or more recently, Fan and Yao (2003). When $f(I_{t-1}, \theta)$ is correctly specified for $f(I_{t-1})$, that is, when (and only when) there exists some θ_0 in $\Theta \subset \mathbb{R}^p$ such that $f(I_{t-1}, \theta_0) = f(I_{t-1})$ a.s., the r.v $e_t(\theta_0)$ coincides (a.s.) with $\{\varepsilon_t\}$, and hence, $e_t(\theta_0)$ will be independent in conditional mean with respect to \mathcal{F}_{t-1} . Thus, the correct specification is tantamount to

$$E[e_t(\theta_0) | I_{t-1}] = 0 \text{ a.s.}, \text{ for some } \theta_0 \in \Theta \subset \mathbb{R}^p. \quad (2)$$

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), Li, Hsiao and Zinn (2003), Khamaladze and Koul (2004) or Koul and Ni (2004), to mention a few. Whereas in a time series context some examples are Bierens (1984), Li (1999), de Jong (1996), Bierens and Ploberger (1997), Kreiss, Neumann and Yao (1998) or Koul and Stute (1999). Although the idea of these tests is the same in all cases, namely, comparing a parametric and a (semi-) non-parametric estimation of a functional of the conditional mean in (2), they can be divided in two classes of tests, both based on the equivalence

$$E[e_t(\theta_0) | I_{t-1}] = 0, \text{ a.s.} \iff E[e_t(\theta_0)w(I_{t-1})] = 0, \forall w \in BM(\mathbb{R}^d), \quad (3)$$

where $BM(\mathbb{R}^d)$ denotes the space of bounded, \mathcal{F}_{t-1} -measurable real-valued functions on \mathbb{R}^d . The first class of consistent tests uses the fact that (3) holds with $w(\cdot)$ equal to the regression function $E[e_t(\theta_0) | I_{t-1}]$, and hence, one can estimate this regression function by nonparametric smoothing

techniques and to check a unique orthogonality condition. We call this approach the “local approach”, because of the use of the local measure of dependence $E[e_t(\theta_0) | I_{t-1}]$. 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 $d = 1$. Recently, Chen, Härdle and Li (2003) have proposed a model diagnostic test for time series regressions using empirical likelihood procedures, see also Tripathi and Kitamura (2003). Chen, Härdle and Li (2003) show that their test is asymptotically equivalent to the Härdle and Mammen’s (1993) test, with the added property that the empirical likelihood test Studentizes internally, avoiding asymptotic variance estimation.

The second class of tests avoids smoothing estimation by means of reducing the conditional mean independence in (2) to an infinite number of unconditional moment restrictions over a parametric family of functions, i.e.,

$$E[e_t(\theta_0) | I_{t-1}] = 0 \text{ a.s.} \iff E[e_t(\theta_0)w(I_{t-1}, x)] = 0, \text{ almost everywhere (a.e.) in } \Pi \subset \mathbb{R}^q, \quad (4)$$

where $\Pi \subset \mathbb{R}^q$, $q \in \mathbb{N}$, is a properly chosen space, and the parametric family $\{w(\cdot, x) : x \in \Pi\}$ is such that the equivalence (4) holds, see Stinchcombe and White (1998) and Escanciano (2004b) for primitive conditions on the family $\{w(\cdot, x) : x \in \Pi\}$ to satisfy this equivalence. We call the approach based on (4) 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. The most frequently used weighting functions have been exponential functions, e.g. $w(I_{t-1}, x) = \exp(ix'I_{t-1})$ in Bierens (1990), where $i = \sqrt{-1}$ denotes the imaginary unit, and the indicator function $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$, see, for instance, Koul and Stute (1999). The former has the advantage of being analytic but in order to achieve consistency against all alternatives, the user has to choose an integrating measure on Π . In the indicator case, x lives on the information variable’s space, and hence, a natural measure is the empirical distribution function of the information set, although the family of indicator functions is not analytic.

On the other hand, when the dimension of the information set I_{t-1} is high or even moderate, the sparseness of the data in high-dimensional spaces leads to most of the above test statistics to suffer a considerable bias. This is a common problem in both approaches, the local and the integrated. In particular, tests based on the family $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$ tend usually to underrejection when the dimension d of the information set is large or moderate and the alternative at hand is nonlinear, even for large sample sizes, see Section 5. On the other hand, tests with the exponential family are more robust to this problem because they are based on one-dimensional projections $x'I_{t-1}$.

Recently, Escanciano (2004a) has considered in a 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 has the property that overcomes the problem of the

curse of dimensionality because is based on projections, and at the same time, avoids the choice of a subjective integrating measure in the Cramér-von Mises test. In addition, the Cramér-von Mises test based on this new family presents excellent power properties in finite samples, see Section 5 below.

Note that different families w deliver different power properties of the integrated based tests. The “optimal” family will depend on the true alternative at hand as well as the functional used to measure the orthogonality restrictions, see (6) below. So, it would be important to establish a general theory for the integrated based test in order to cover a large class of weighting functions w . Note that the choice of w gives us flexibility in directing power toward desired directions.

The main aim of this paper is to present a unified theory for the goodness-of-fit tests (omnibus and directional) based on the integrated approach for a general weighting function w , including but not restricting to indicators and exponential families. The second important goal of this paper is to provide new estimators for parameters in the regression function, using the same underlying processes as in the testing procedure.

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, which are the basis for the test statistics and for a new class of minimum distance estimators. We begin by establishing a weak convergence theorem for a general class of marked empirical processes, which allows us to study, in a unified way, the asymptotic distribution of the test statistics under the null and under local alternatives, and also to prove the consistency and asymptotic normality of the new proposed estimators. In Section 3, a bootstrap procedure for approximating the asymptotic null distribution of the omnibus tests is considered and theoretically justified. We study the estimation and consistency of the principal components of the residual marked processes to derive, in Section 4, optimal directional tests against a particular local alternative and efficient estimation under a minimum distance principle. In Section 5, we make a simulation exercise comparing different tests under the null and under the alternative, and we apply previous methodology to study the conditional mean specification of the well-studied Canadian lynx data set. Proofs are deferred to Section 6.

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. In what follows, Π_c will denote a compact subset of $\Pi \subset \mathbb{R}^q$, and let $\ell^\infty(\Pi_c)$ be the space of all complex-valued functions that are uniformly bounded on Π_c . Let \implies denote weak convergence on compacta in $\ell^\infty(\Pi)$, i.e., weak convergence on $\ell^\infty(\Pi_c)$ for any compact subset Π_c of Π , see Definition 1.3.3 and Chapter 1.6 in van der Vaart and Wellner (1996, hereafter VW). Also $\xrightarrow{P^*}$ and $\xrightarrow{as^*}$ denote convergence in outer probability and outer almost surely, respectively, see Definition 1.9.1 in VW. All convergence 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 $Z \in S$, $Z = \{(Y_t, I'_{t-1})' : t = 0, \pm 1, \pm 2, \dots\}$ with $0 < E|Y_t| < \infty$, be one of these processes. The main goal in this paper 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 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 . As arguing above, one way to characterize H_0 is by the infinite number of unconditional moment restrictions

$$E[e_t(\theta_0)w(I_{t-1}, x)] = 0, \text{ a.e., } x \in \Pi, \quad (5)$$

where the parametric family $\{w(\cdot, x) : x \in \Pi\}$ is such that the equivalence in (4) holds. Examples of such families are $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$, $w(I_{t-1}, x) = \exp(ix'I_{t-1})$, $w(I_{t-1}, x) = \sin(x'I_{t-1})$ or $w(I_{t-1}, x) = 1/(1 + \exp(c - x'I_{t-1}))$ with $c \neq 0$, all of them with $\Pi \subset \mathbb{R}^d$, or $w(I_{t-1}, x) = 1(\beta'I_{t-1} \leq u)$, with $x = (\beta', u)' \in \Pi_{pro}$, see Stinchcombe and White (1998) or Escanciano (2004b).

In view of a sample $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$, we have that the standardized sample version of (5) is given by the marked empirical process

$$R_{n,w}(x, \theta_0) \equiv R_{n,w}(x) = n^{-1/2} \sum_{t=1}^n e_t(\theta_0)w(I_{t-1}, x),$$

if θ_0 is known, or

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

if θ_0 is unknown and has to be estimated by θ_n , say. We elaborate theory for both $R_{n,w}$ and $R_{n,w}^1$, although being the latter the most interesting case, we consider throughout the paper the composite hypothesis and we assume that θ_0 is unknown. The marks in $R_{n,w}^1$ are given by the classical residuals, so we call $R_{n,w}^1$ a residual marked empirical process.

Because of (4), 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., a norm of $R_{n,w}^1$, $\Gamma(R_{n,w}^1)$, say. The most used norms are the Cramér-von Mises (CvM) and Kolmogorov-Smirnov (KS) functionals

$$CvM_{n,w} := \int_{\Pi} |R_{n,w}^1(x)|^2 \Psi(dx)$$

and

$$KS_{n,w} := \sup_{x \in \Pi_c} |R_{n,w}^1(x)|,$$

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

The power properties of the integrated test based on $\Gamma(R_{n,w}^1)$ depend on the family w and the functional Γ chosen. Now, we shall see this for the particular case of the CvM functional using different families w and different integrating measures $\Psi(\cdot)$. We use some arguments that nicely reflect the relationship between different CvM tests and that also serve to compare the local and the integrated approaches. Consider the Fourier transform of $R_{n,w}^1$

$$\begin{aligned} & \frac{1}{(2\pi)^{1/2}} \int \exp(iz'x) (R_{n,w}^1(x))^c \Psi(dx) \\ = & \frac{1}{n^{1/2}(2\pi)^{1/2}} \sum_{t=1}^n e_t(\theta_n) \int \exp(iz'x) w^c(I_{t-1}, x) \Psi(dx) \\ = & \frac{1}{n^{1/2}(2\pi)^{1/2}} \sum_{t=1}^n e_t(\theta_n) \exp(iz'I_{t-1}) \int \exp(iz'(x - I_{t-1})) w^c(I_{t-1}, x) \Psi(dx). \end{aligned}$$

Then, for many weighting families w , for instance $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$, the integral in the last equality does not depend on I_{t-1} , i.e., $(2\pi)^{-1/2} \int \exp(iz'(x - I_{t-1})) w^c(I_{t-1}, x) \Psi(dx) := \overline{K}_{w,\Psi}(z)$, and therefore, by Parseval's identity, see Taylor and Lay's (1979) Theorem 6.10,

$$\int |R_{n,w}^1(x)|^2 \Psi(dx) = \int |R_{n,\exp}^1(z)|^2 |\overline{K}_{w,\Psi}(z)|^2 dz, \quad (6)$$

with

$$R_{n,\exp}^1(x) := n^{-1/2} \sum_{t=1}^n e_t(\theta_n) \exp(ix'I_{t-1}).$$

Expression (6) is useful, because in particular, it shows that if we consider the local approach, as in Härdle and Mammen (1993), with a Nadaraya-Watson estimator for the regression function, i.e., $w_K(I_{t-1}, x) = K((I_{t-1} - x)/h)$, but with a fixed bandwidth h , then (6) implies that the test based on $w_K(I_{t-1}, x)$ would be consistent even when the regression function is not estimated consistently (h is fixed), see Fan and Li (2000) for a comparison between the kernel based tests and the Bierens' (1990) test. Also, for $w_K(I_{t-1}, x)$

$$\frac{1}{(2\pi)^{1/2} h^d} \int \exp(iz'(x - I_{t-1})) K\left(\frac{I_{t-1} - x}{h}\right) \Psi(dx) = \overline{K}(hz),$$

where \overline{K} is the Fourier transform of the kernel K . Hence, we see that the effect of tending h to zero is to put more weight on high frequency alternatives. One advantage of fixing h is that there is a

one-to-one correspondence between values of h and the asymptotic distribution of the CvM based on R_{n,w_K}^1 . Thus, one would expect to obtain a better approximation of the finite sample distribution of the CvM by its asymptotic distribution with a fixed h than with a vanishing h , see Fan (1998, note 4). Also, by fixing h , the CvM test based on R_{n,w_K}^1 is able to detect local alternatives converging to the null at the parametric rate $n^{-1/2}$, whereas with vanishing h it is able to detect only local alternatives converging at the rate $O((nh^{d/2})^{-1/2})$, slower than $n^{-1/2}$. The price of fixing h , is that the asymptotic null distribution is no longer standard and has to be approximated, for example, by resampling methods. Because in this paper we consider a general weighting function w , our theory also covers the case of R_{n,w_K}^1 with fixed h , whenever $w_K(I_{t-1}, x)$ satisfies A4(a) below. This is the case for most common kernel functions used in the literature.

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 mild assumptions. This result is of interest in its own.

2.1 Weak convergence.

In this section we consider a weak convergence theorem for a large class of marked empirical processes for which the process $R_{n,w}$ is a special case. Let $Z = \{(\varepsilon_t, X_t)'\}, t = 0, \pm 1, \dots\} \in S$ satisfying

$$E[\varepsilon_t | X_t] = 0 \text{ a.s.} \quad t = 0, \pm 1, \dots$$

Motivated from equivalence (4), our goal in this section is to establish the weak convergence of the empirical process based on a family $\{w(\cdot, x) : x \in \Pi\}$, i.e.,

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

Usually, different families w deliver different technical approaches for the asymptotic theory, essentially this is due to the continuity of the family with respect to the auxiliary parameter x , compare, for instance, the tightness condition in Bierens and Ploberger (1997) and Koul and Stute (1999). One possibility to propose 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 $\ell^\infty(\Pi_c)$, the space of all complex-valued functions that are uniformly bounded on Π_c . Of course, the sample paths of $\alpha_{n,w}$ are usually contained in much a smaller space (such as $D([\infty, \infty]^d)$), but as long as this space is equipped with the supremum metric, this is irrelevant for the weak convergence theorem. For some families w , such as the indicator family $w(X_t, x) = 1(X_t \leq x)$, our assumptions are weaker than those considered in other related weak

convergence theorems, and are similar to the mildest obtained in the i.i.d. case, see Stute (1997). The weak convergence theorem that we present here is founded on a remarkable result by Nishiyama (2000, Corollary 4.3), which generalizes Theorem 3.1 in Ossiander (1987) and Theorem 2.11.9 in VW to empirical processes under possibly non-stationary martingale difference sequences. Earlier results in this direction can be found in Levental (1989) and Bae and Levental (1995).

Let 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 Π_c as

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

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

Assumption W1: $\{(\varepsilon_t, X_t)' : t = 0, \pm 1, \pm 2, \dots\} \in S$ with $E[\varepsilon_1 | X_1] = 0$ a.s. and $0 < E\varepsilon_1^2 < \infty$.

Assumption W2: For every compact subset Π_c , the family $w(I_{t-1}, x)$ is uniformly bounded (a.s.) on Π_c and for every $\varepsilon > 0$ there exists a finite partition $\mathcal{B}_\varepsilon = \{H_k; 1 \leq k \leq N_\varepsilon\}$ of Π_c such that

$$\int_0^\infty \sqrt{\log(N_\varepsilon)} d\varepsilon < \infty \quad (7)$$

and

$$\sup_{\varepsilon \in (0,1) \cap \mathbb{Q}} \frac{\alpha_{n,w}(\mathcal{B}_\varepsilon)}{\varepsilon^2} = O_P(1). \quad (8)$$

Let $\alpha_{\infty,w}(\cdot)$ be a Gaussian process with zero mean and covariance function given by

$$K_w(x_1, x_2) := E[\varepsilon_t^2 w(X_t, x_1) w^c(X_t, x_2)].$$

We are now in position to state our first main result.

THEOREM 1: *If Assumptions W1 and W2 hold, then it follows that*

$$\alpha_{n,w} \Longrightarrow \alpha_{\infty,w}.$$

Now, we shall show that assumption W2 is satisfied (under W1 and some mild conditions) for all the families w considered in the literature. First, we start with the smooth case. Note that under W1 and for smooth functions $w(X_t, x)$ satisfying

$$|w(X_t, x_1) - w(X_t, x_2)| \leq K_t \rho(x_1, x_2),$$

with $\rho(\cdot, \cdot)$ such that (Π_c, ρ) is a totally bounded metric space and K_t is a stationary process with $E[\varepsilon_t^2 K_t^2] < \infty$, 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 ε -covering number of Π_c with respect to ρ , i.e., the minimum number of ρ -balls needed to cover Π_c . This assumption is satisfied, for instance, for $w(X_t, x) = \exp(ix'X_t)$, $w(I_t, x) = \sin(x'I_t)$ or $w(I_t, x) = 1/(1 + \exp(c - x'I_t))$, $c \in \mathbb{R}$.

For non-smooth functions, such as $w(X_t, x) = 1(X_t \leq x)$ or $w(I_{t-1}, x) = 1(\beta' I_{t-1} \leq u)$, $x = (\beta', u)'$, the situation is more involved. For $w(X_t, x) = 1(X_t \leq x)$, 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. 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 found to be infinite. On the other hand, for $w(I_{t-1}, x) = 1(\beta' I_{t-1} \leq u)$, $x = (\beta', u)'$, Escanciano (2004a) proves a weak convergence theorem in a i.i.d setup using the techniques of VW. These techniques can not be applied directly to a time series context. Next result is an application of Theorem 1 to these particular weighting functions and provides an extension of the weak convergence theorems of Koul and Stute (1999) and Escanciano (2004a) to the multivariate case $d > 1$ and a time series framework, respectively. First, let define the canonical semimetrics

$$d_{ind}(x_1, x_2) := \left(E[\varepsilon_t^2 \{1(X_t \leq x_1) - 1(X_t \leq x_2)\}^2] \right)^{1/2} \quad x_1, x_2 \in \Pi_{ind} := [-\infty, \infty]^d$$

and

$$d_{pro}(x_1, x_2) := \left(E[\varepsilon_t^2 \{1(\beta'_1 X_t \leq u_1) - 1(\beta'_2 X_t \leq u_2)\}^2] \right)^{1/2},$$

where $x_1 = (\beta'_1, u_1)'$ and $x_2 = (\beta'_2, u_2)'$, $x_1, x_2 \in \Pi_{pro} = \mathbb{S}^d \times [-\infty, \infty]$.

COROLLARY 1: *Under W1 and the uniform continuity of $d_{ind}(\cdot, \cdot)$ ($d_{pro}(\cdot, \cdot)$) on $\Pi_{ind} \times \Pi_{ind}$ ($\Pi_{pro} \times \Pi_{pro}$) the weak convergence of Theorem 1 holds.*

2.2 Asymptotic distribution under the null.

Now, we establish the limit distribution of the marked empirical process $R_{n,w}^1$ under the null hypothesis H_0 . The null limit distributions of the tests are the limit distributions of some functionals of $R_{n,w}^1$. To derive asymptotic results we consider the following assumptions. First, let define the semimetric $d_w(x_1, x_2) := \left(E[\varepsilon_1^2 \{w(I_0, x_1) - w(I_0, x_2)\}^2] \right)^{1/2}$ and the score $g(I_{t-1}, \theta_0) := (\partial/\partial\theta')f(I_{t-1}, \theta_0)$.

Assumption A1:

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

A1(b): $E|\varepsilon_1|^2 < C$.

A1(c): $d_w(x_1, x_2)$ is continuous on $\Pi_c \times \Pi_c$, for any compact subset $\Pi_c \subset \Pi$.

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 parametric space Θ is compact in \mathbb{R}^p . The true parameter θ_0 belongs to the interior of Θ . There exists a unique θ_* such that $|\theta_n - \theta_*| = o_P(1)$. Obviously, under the null hypothesis H_0 , $\theta_* = \theta_0$.

A3(b): The estimator θ_n satisfies the following asymptotic expansion under H_0

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n l(Y_t, I_{t-1}, \theta_0) + o_P(1),$$

where $l(\cdot)$ is such that $E[l(Y_t, I_{t-1}, \theta_0)] = 0$ and $L(\theta_0) = E[l(Y_t, I_{t-1}, \theta_0)l'(Y_t, I_{t-1}, \theta_0)]$ exists and is positive definite.

Assumption A4:

A4(a): The weighting function $w(\cdot)$ is such that the equivalence in (4) holds. For any compact set Π_c of Π , $w(I_{t-1}, x)$ is uniformly bounded (a.s.) on Π_c , satisfies under the null the assumption W2 above and under both, the null and the alternative, the uniform law of large numbers (ULLN)

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

when $Z = \{(\varepsilon_t, X_t)', t = 0, \pm 1, \dots\} \in S$.

A4(b): The integrating function $\Psi(\cdot)$ is a probability distribution function which is chosen 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. A1(c) is necessary for the asymptotic tightness of the process $R_{n,w}$ in the non-smooth case. It guarantees the continuity of the limit process. 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), for its robust modifications (under further regularity assumptions), see Chapter 5 in Koul (2002), and as we shall show below for a large class of minimum distance estimators constructed through the residual marked empirical process $R_{n,w}^1$. The assumption that w satisfies (4) is needed only for the consistency of the tests. W2 usually holds under previous assumptions, see Section 2.1. The ULLN in A4(a) usually follows from the Ergodic Theorem and a Glivenko-Cantelli's argument. Under the null hypothesis is consequence of previous assumptions, cf. Theorem 1. Note that under A4, $R_{n,w}^1$ can be viewed as a random

element with values in $\ell^\infty(\Pi_c)$. The choice of $\Psi(\cdot)$ depends on the space Π and w , and is crucial for the power properties of the CvM test, cf. (6). Some discussions about the choice of $\Psi(\cdot)$ for a given w can be found in Escanciano and Velasco (2003). A4(b) is only needed for the consistency of the CVM tests.

Under A1 and (2), 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

$$K_w(x_1, x_2) = E[\varepsilon_t^2 w(I_{t-1}, x_1) w^c(I_{t-1}, x_2)]. \quad (9)$$

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

THEOREM 2: *Under the null hypothesis H_0 , A1 and A4(a)*

$$R_{n,w} \Longrightarrow R_{\infty,w},$$

where $R_{\infty,w}(\cdot)$ is a continuous Gaussian process with zero mean and covariance function given by (9).

In practice, θ_0 is unknown and has to be estimated from a sample $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$ by an estimator θ_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)*

$$R_{n,w}^1(\cdot) \Longrightarrow R_{\infty,w}(\cdot) - G'_w(\cdot)V \equiv R_{\infty,w}^1(\cdot),$$

where $R_{\infty,w}$ is the same process as in Theorem 2 and

$$\text{Cov}(R_{\infty,w}(x_1), V) = E[\varepsilon_t l(Y_t, I_{t-1}, \theta_0) w(I_{t-1}, x_1)].$$

Next, using the last theorem and the Continuous Mapping Theorem (CMT), see, e.g., Theorem 1.3.6 in VW, we obtain the asymptotic null distribution of continuous functionals $CvM_{n,w}$ and $KS_{n,w}$.

COROLLARY 2: *Under the assumptions of Theorem 3, for any continuous functional $\Gamma(\cdot)$*

$$\Gamma(R_{n,w}^1) \xrightarrow{d} \Gamma(R_{\infty,w}^1).$$

REMARK 1: Note that the asymptotic null distributions of $CvM_{n,w}$ and $KS_{n,w}$ depend in a complex way of the data generating process (DGP) as well as the hypothesized model under the null, so critical values have to be tabulated for each model and each DGP, making the application of these asymptotic results difficult in practice.

REMARK 2: 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(b), i.e.,

$$\sup_{x \in \Pi_c} |\Psi_n(x) - \Psi(x)| \longrightarrow 0 \text{ a.s.},$$

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 (2004a). By the Glivenko-Cantelli's 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.

In Assumption A3 we require that the estimator of θ_0 admits an asymptotic linear representation. For completeness of the presentation, we give some mild sufficient conditions under which a class of minimum distance estimators, see Chapter 5 in Koul (2002) and references therein, are asymptotically linear. Motivated from equivalence (4) and under suitable conditions on Ψ , see A4(b), we have that under the null

$$\theta_0 = \arg \min_{\theta \in \Theta} \int_{\Pi} |E[e_t(\theta)w(I_{t-1}, x)]|^2 \Psi(dx), \quad (10)$$

and θ_0 is the unique value that satisfies (10). Then, we propose estimating θ_0 by the sample analogue of (10), that is,

$$\theta_n := \arg \min_{\theta \in \Theta} \int_{\Pi} n^{-1} |R_{n,w}^1(x, \theta)|^2 \Psi(dx). \quad (11)$$

This estimator is a minimum distance estimator and extends in some sense the Generalized Method of Moments (GMM) estimator, frequently used in econometric and statistical applications. Similar generalizations of GMM have been considered first in Carrasco and Florens (2000) for i.i.d data. Recently, and for $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$, Domínguez and Lobato (2004) have proposed a particular case of the estimator (11) for a conditional moment restriction under time series. Also recently, Koul and Ni (2004) have proposed a minimum distance estimation for θ_0 under i.i.d series using a L_2 -distance similar to that used in Härdle and Mammen (1993) in the “local approach”. Our procedure

complements and, in some cases, extends these approaches to other frameworks, including time series, other functions w , weaker assumptions or the possibility of considering a fixed parameter bandwidth in the local approach. In addition, we study how to obtain efficient estimation under this procedure and the relationship between this efficient estimation and a maximum likelihood estimator (MLE), see Section 4. The following matrices are involved in the asymptotic variance-covariance matrix of the estimator,

$$C := \int_{\Pi} G_w(x) G_w'(x) \Psi(dx),$$

$$D := \int_{\Pi} G_w(x) G_w'(x) K_w(x, y) \Psi(dx) \Psi(dy).$$

For the consistency and asymptotic normality of the estimator we need an additional assumption.

Assumption A1': The regression function $f(\cdot, \theta)$ satisfies that there exists a function $K(I_{t-1})$ with $|f(I_{t-1}, \theta)| \leq K(I_{t-1})$, such that $K(I_{t-1})$ is $F_I(\cdot)$ -integrable.

THEOREM 4: *Under the null hypothesis H_0 , Assumptions A1-A2, A4(a) and A1'*

- (i) The estimator given in (11) is consistent, i.e., $\theta_n \rightarrow \theta_0$ a.s..
- (ii) If in addition, the matrix C is nonsingular, then

$$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} N(0, C^{-1}DC^{-1}).$$

From the proof of Theorem 4 we have the asymptotic linear expansion required in A3(b)

$$\sqrt{n}(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n l(Y_t, I_{t-1}, \theta_0) + o_P(1),$$

with

$$l(Y_t, I_{t-1}, \theta_0) = -C^{-1}h(I_{t-1}, \theta_0)\{Y_t - f(I_{t-1}, \theta_0)\},$$

where

$$h(I_{t-1}, \theta_0) := \int_{\Pi} G_w(x) w^e(I_{t-1}, x) \Psi(dx).$$

Note that in general the estimator given in (11) is not asymptotically efficient. An asymptotically efficient estimator based on the same minimum distance principle is constructed in Section 4.

2.3 Consistency and local alternatives.

In this section we study the consistency properties of the test based on 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 .

THEOREM 5: *Under the alternative hypothesis H_A and A1-A4(a)*

$$n^{-1/2}R_{n,w}^1(\cdot) \xrightarrow{P^*} E[e_t(\theta_*)w(I_{t-1}, \cdot)].$$

Hence, using the CMT we conclude that under the assumptions of Theorem 5 and A4(b)

$$\int_{\Pi} \left| n^{-1/2}R_{n,w}^1(x) \right|^2 \Psi(dx) \xrightarrow{P} \int_{\Pi} |E[e_t(\theta_*)w(I_{t-1}, x)]|^2 \Psi(dx) > 0$$

and

$$\sup_{x \in \Pi_c} \left| n^{-1/2}R_{n,w}^1(x) \right| \xrightarrow{P} \sup_{x \in \Pi_c} |E[e_t(\theta_*)w(I_{t-1}, x)]| > 0.$$

Therefore, the test statistic $\Gamma(R_{n,w}^1)$ will go to $+\infty$ under the alternative and the test will gain power.

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

$$H_{A,n} : Y_{t,n} = f(I_{t-1}, \theta_0) + \frac{a(I_{t-1})}{n^{1/2}} + \varepsilon_t, \text{ a.s.}, \quad (12)$$

where the function $a(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ is F_I -measurable, F_I -integrable, with zero mean and satisfies

$$Pr(a(I_{t-1}) = 0) < 1. \quad (13)$$

To derive the next result we need the following assumption.

Assumption A3': The estimator θ_n satisfies the following asymptotic expansion under $H_{A,n}$

$$\sqrt{n}(\theta_n - \theta_0) = \xi_a + \frac{1}{\sqrt{n}} \sum_{t=1}^n l(Y_t, I_{t-1}, \theta_0) + o_P(1),$$

where the function $l(\cdot)$ is as in A3 and $\xi_a \in \mathbb{R}^p$.

REMARK 3: It is not difficult to show that θ_n in (11) satisfies A3' under A1-A2 and A1' with

$$\xi_a = C^{-1} \int_{\Pi} E[a(I_{t-1})w(I_{t-1}, x)]G_w(x)\Psi(dx).$$

The deterministic function $D_{w,a}(\cdot) := E[a(I_{t-1})w(I_{t-1}, \cdot)] - G'_w(\cdot)\xi_a$ plays an important roll in the following result.

THEOREM 6: Under the local alternatives (12), Assumptions A1, A2, A3' and A4(a)

$$R_{n,w}^1 \implies R_{\infty,w}^1 + D_{w,a},$$

where $R_{\infty,w}^1$ is the process defined in Theorem 3.

Note that from the equivalence (4), we have that

$$D_{w,a} = 0 \text{ a.e.} \iff a(I_{t-1}) = \xi'_a g(I_{t-1}, \theta_0) \text{ a.s..}$$

Therefore, for directions $a(\cdot)$ not collinear to the score the shift function $D_{w,a}$ is non-trivial. For some estimators, $D_{w,a}$ has an intuitive geometric interpretation. For instance, for the new minimum distance estimators (11) the shift function is given by

$$D_{w,a}(\cdot) = E[a(I_{t-1})w(I_{t-1}, \cdot)] - G'_w(\cdot)C^{-1} \int_{\Pi} E[a(I_{t-1})w(I_{t-1}, x)]G_w(x)\Psi(dx),$$

and represents the orthogonal projection in $L_2(\Pi, \Psi)$ of $E[a(I_{t-1})w(I_{t-1}, \cdot)]$ parallel to G_w , where $L_2(\Pi, \Psi)$ is the Hilbert space of all complex-valued and Ψ -square integrable functions on Π , see a similar result in Khamaladze and Koul (2004) for a general class of M-estimators. The next corollary is consequence of the CMT and the last theorem.

COROLLARY 3: Under the local alternatives (12), and Assumptions A1, A2, A3' and A4(a), for any continuous functional $\Gamma(\cdot)$

$$\Gamma(R_{n,w}^1) \xrightarrow{d} \Gamma(R_{\infty,w}^1 + D_{w,a}).$$

To gain some insights in the asymptotic power properties of the integrated based tests, we give conditions guarantying that tests based on $\Gamma(R_{n,w}^1)$, for a continuous even functional $\Gamma(\cdot)$, are asymptotically strictly unbiased. Let define the asymptotic local power function of the test based on rejecting if $\Gamma(R_n) > c_\alpha$ as

$$\Pi_\Gamma(\alpha, a) := \lim_{n \rightarrow \infty} P(\Gamma(R_{n,w}^1) > c_\alpha \mid H_{A,n}),$$

where c_α is such that $P(\Gamma(R_{\infty,w}^1) > c_\alpha \mid H_0) = \alpha$. We find conditions on $a(\cdot)$ in order to

$$\Pi_\Gamma(\alpha, a) > \alpha$$

holds. From Corollary 2, we know that

$$\Pi_\Gamma(\alpha, a) = \mathbb{P}_0(\Gamma(R_{\infty,w}^1 + D_{w,a}) > c_\alpha),$$

where \mathbb{P}_0 is the distribution of $R_{\infty,w}^1$ under the null. If the condition $\xi_{-a} = -\xi_a$ holds and $\Gamma(\cdot)$ is even, by symmetry of R_∞ we have that $\Pi_\Gamma(\alpha, -a) = \Pi_\Gamma(\alpha, a)$. Anderson's Lemma, see Anderson

(1955), yields that $\Pi_\Gamma(\alpha, ca)$ is a non-decreasing function of $|c|$. It is not difficult to show that first derivative of $\Pi_\Gamma(\alpha, ca)$ at $c = 0$ is equal to zero and that the second derivative is strictly positive if $D_{w,a} \neq 0$ in a set with positive Lebesgue measure, see Milbrodt and Strasser's (1990) Theorem 2.8. In the latter case, $\Pi_\Gamma(\alpha, a) > \alpha$. Obviously, if $\Pi_\Gamma(\alpha, a) > \alpha$, then $D_{w,a} \neq 0$ with positive Lebesgue measure. Therefore, we see that for any direction not parallel to $g(I_{t-1}, \theta_0)$, tests based on $\Gamma(R_{n,w}^1)$, for a continuous even $\Gamma(\cdot)$, are able to detect it asymptotically. This property is not attainable for those tests based on the local approach.

3. BOOTSTRAP APPROXIMATION OF RESIDUAL MARKED EMPIRICAL PROCESSES

We have seen that the asymptotic null distribution of continuous functionals of $R_{n,w}^1$ depends in a complicated way of the DGP and the specification under the null. Therefore, critical values for the test statistics can not be tabulated for general cases. A rather recently approach to solve this problem is that of Khamaladze and Koul (2004), who consider a martingale transformation of the process $R_{n,w}^1$, with $w(I_{t-1}, B) = 1(I_{t-1} \in B)$, and where B is a Borel set, that delivers asymptotically-free distributed tests. As they comment, their approach can be easily generalized to time series autoregressions and the theory of the present paper can help to this end. Unfortunately, this approach is only useful for the indicator weighting family. Here, we propose a bootstrap method to solve the problem of approximating the asymptotic null distribution of an integrated based test under time series. 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) or Li, Hsiao and Zinn (2003), in and i.i.d context, and Kreiss, Neumann and Yao (1998) or Franke, Kreiss and Mammen (2002) for time series sequences. It is shown in these papers 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 as considered for instance in Stute, González-Manteiga and Presedo-Quindimil (1998). Our bootstrap approximation can be useful for a large class of test statistics in regression model checks, in particular, those tests based on the integrated approach, e.g. Koul and Stute (1999), or for the local approach, for instance, the generalization of Härdle and Mammen's (1993) test to time series by Kreiss, Neumann and Yao (1998) (with or without h fixed). Other resampling schemes are 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

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

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 with zero mean, unit variance, bounded support and also independent of the sequence $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$.
- 3) Given θ_n and $e_t^*(\theta_n)$ generate bootstrap data for the dependent variable Y_t^* according to

$$Y_t^* = f(I_{t-1}, \theta_n) + e_t^*(\theta_n) \text{ for } t = 1, \dots, n.$$

- 4) Compute θ_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 $t = 1, \dots, n$.

In the remaining of this section and using standard bootstrap notation, denote by E^* the expectation operator given the sample $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$. Examples of $\{V_t\}$ sequences are i.i.d. Bernoulli variates with

$$P(V_t = 0.5(1 - \sqrt{5})) = (1 + \sqrt{5})/2\sqrt{5} \quad (14)$$

and

$$P(V_t = 0.5(1 + \sqrt{5})) = 1 - (1 + \sqrt{5})/2\sqrt{5}, \quad (15)$$

used in, e.g., Mammen (1993), 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) or de Jong (1996), for other sequences see Mammen (1993). To justify theoretically this bootstrap approximation we need an additional assumption on the behaviour of the bootstrap estimator.

Assumption A5:

A5(a): The estimator θ_n^* satisfies the following asymptotic expansion

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

where the function $l(\cdot)$ is as in A3 with

A5(b): $E^*[l(Y_t^*, I_{t-1}, \theta_n)] = 0$, a.s..

A5(c): $L(\theta_n) = E^*[l(Y_t^*, I_{t-1}, \theta_n)l'(Y_t^*, I_{t-1}, \theta_n)]$ exists and is positive definite (a.s.) with $L(\theta_n) \longrightarrow L(\theta_*)$ a.s..

A5(d): $n^{-1} \sum_{t=1}^n E^*[e_t(\theta_n)w(I_{t-1}, x)V_t l(Y_t^*, I_{t-1}, \theta_n)] \longrightarrow E[e_t(\theta_*)w(I_{t-1}, x)l(Y_t, I_{t-1}, \theta_*)]$ a.s..

REMARK 4: Again, it is not difficult to show that θ_n^* using the principle (11), satisfies A5 for sufficiently large n , under A1-A2, A1' and the invertibility of the matrix C .

This bootstrap procedure allows us to approximate the asymptotic critical values of the tests based on $\Gamma(R_{n,w}^1)$. We use the concept of convergence in distribution in probability one, a less restrictive concept is convergence in distribution in probability, see Giné and Zinn (1990) for more detailed discussions on these concepts.

THEOREM 7: *Assume A1-A5. Then, under the null Hypothesis H_0 , under any fixed alternative hypothesis or under the local alternatives (12),*

$$R_{n,w}^{1*} \xrightarrow[*]{\implies} \tilde{R}_{\infty,w}^1, \text{ a.s.},$$

where $\tilde{R}_{\infty,w}^1$ is the same Gaussian process as in Theorem 3 but with θ_* replacing θ_0 , and $\xrightarrow[*]{\implies}$ denote weak convergence almost surely under the bootstrap law, see Giné and Zinn (1990).

Note that, under the null θ_* coincides with θ_0 , and then, $\tilde{R}_{\infty,w}^1 \equiv R_{\infty,w}^1$ 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:

1. Calculate the test statistic D_n with the original sample.
2. Generate V_t , a sequence of i.i.d random variables with zero mean, unit variance and bounded support, independent of the original sample.
3. Compute $R_{n,w}^{1*}$ and $D_n^* = \Gamma(R_{n,w}^{1*})$.
4. Repeat steps 2 and 3 B times and compute the empirical $(1 - \alpha)th$ sample quantile of D_n^* with the B values, $D_{n,[\alpha]}^*$ say. The proposed test rejects the null hypothesis at the significance level α if $D_n > D_{n,[\alpha]}^*$.

Note that given the result obtained in Theorem 7 the proposed bootstrap tests have a correct asymptotic level, are consistent and are able to detect alternatives tending to the null at the parametric rate $n^{-1/2}$. Section 5 below shows that this bootstrap procedure provides good approximations in finite samples.

4. OPTIMAL DIRECTIONAL TESTS AND EFFICIENT ESTIMATION

In this section we shall employ the principal components of the residual marked process $R_{n,w}^1$ in order to construct asymptotic optimal directional tests and asymptotic efficient estimators based on

(11). To establish the asymptotic theory of these optimal inference procedures we need to consider estimation and consistency results of such principal components. To begin with, we need some notation. Throughout this section we view $R_{n,w}^1$ as a random element with values in H_1 instead of $\ell^\infty(\Pi_c)$, where $H_1 \equiv L_2(\Pi, \Psi)$ is the Hilbert space of all complex-valued and Ψ -square integrable functions on Π with the inner-product

$$\langle f, g \rangle_{H_1} = \int_{\Pi} f(x)g^c(x)\Psi(dx),$$

and the induced norm $\|\cdot\|_{H_1} = \langle \cdot, \cdot \rangle_{H_1}^{1/2}$. H_1 is endowed with the natural Borel σ -field, see e.g. Chapter VI in Parthasarathy (1967) for a study of random elements with values on Hilbert spaces. As a mapping in H_1 , $R_{\infty,w}^1$ is a Gaussian random element and has characteristic functional $\chi(h) = \exp(-\frac{1}{2}\langle C_w h, h \rangle_{H_1})$, $h \in H_1$, where C_w is its covariance operator, which is given by

$$C_w h(\cdot) = E[\langle R_{\infty,w}^1, h \rangle_{H_1} R_{\infty,w}^1(\cdot)] \quad h \in H_1. \quad (16)$$

Let H_1^0 be the nullspace of C_w , and H_1^1 its orthogonal complement in H_1 . Because C_w is a compact linear operator, we have that $\{(\lambda_{i,w}, \varphi_{i,w}) : i = 1, 2, \dots\}$ is a complete sequence of eigenelements of it, i.e., $\{\lambda_{i,w} : i = 1, 2, \dots\}$ are real-valued and positive, and the corresponding eigenfunctions $\{\varphi_{i,w} : i = 1, 2, \dots\}$ form a complete orthonormal basis for H_1^1 . Hence any H_1^1 -valued random element has a Fourier expansion in terms of $\{\varphi_{i,w} : i = 1, 2, \dots\}$. In particular, we have the Fourier representations (in distribution)

$$R_{n,w}^1 = \sum_{i=1}^{\infty} \lambda_{i,w}^{1/2} \epsilon_{n,i,w} \varphi_{i,w},$$

$$R_{\infty,w}^1 = \sum_{i=1}^{\infty} \lambda_{i,w}^{1/2} \epsilon_{i,w} \varphi_{i,w},$$

where

$$\epsilon_{i,w} = \lambda_{i,w}^{-1/2} \langle R_{\infty,w}^1, \varphi_{i,w} \rangle_{H_1},$$

and

$$\epsilon_{n,i,w} = \lambda_{i,w}^{-1/2} \langle R_{n,w}^1, \varphi_{i,w} \rangle_{H_1}.$$

Note that by Theorem 3, $\{\epsilon_{i,w} : i = 1, 2, \dots\}$ are i.i.d. $N(0, 1)$ r.v. and $\{\epsilon_{n,i,w} : i = 1, 2, \dots\}$ are, at least, uncorrelated with unit variance. Then, by Parseval's identity

$$CvM_{\infty,w} = \sum_{i=1}^{\infty} \lambda_{i,w} \epsilon_{i,w}^2, \quad \text{in distribution.} \quad (17)$$

Therefore, the asymptotic null distribution of $CvM_{n,w}$ can be expressed as a weighted sum of independent χ_1^2 r.v. with weights depending on the DGP. The principal components $\{\epsilon_{i,w} : i = 1, 2, \dots\}$ play a central role in the power properties of the CvM tests, see, e.g., Neuhaus (1976). Although the CvM test is consistent against all alternatives in H_A , in practice is not able to detect specific alternatives one might have in mind. In particular, it is possible to show that there exist directions $a(\cdot)$ for which the asymptotic local power function is as near to α as desired. This can be immediately seen from (17), since possible high-frequency deviations from H_0 are downweighted by $\lambda_{i,w}$ and $\lambda_{i,w} \downarrow 0$, given the compactness of C_w .

Now, we shall start with the estimation of the eigenelements $\{(\lambda_{i,w}, \varphi_{i,w}) : i = 1, 2, \dots\}$ of C_w . Note that the empirical counterpart of (16) under the null hypothesis is given by

$$C_{n,w}h(\cdot) = \frac{1}{n} \sum_{t=1}^n e_t^2(\theta_n) w(I_{t-1}, \cdot) \int_{\Pi} w(I_{t-1}, x) h^c(x) \Psi(dx),$$

where θ_n is any \sqrt{n} -consistent estimator of θ_0 . Note that, contrary to C_w , the operator $C_{n,w}$ has a finite dimensional closed range (that is spanned by the functions $w(I_{t-1}, \cdot)$, $t = 1, \dots, n$). Therefore, the number of eigenvalues and eigenfunctions of $C_{n,w}$ is finite and bounded by n , and they can be computed by solving a linear system. Let $\lambda_{n,i,w}$ and $\varphi_{n,i,w}$, $1 \leq i \leq n$, denote an eigenvalue and eigenfunction of $C_{n,w}$, respectively. The eigenfunction $\varphi_{n,i,w}$ necessarily has the form $n^{-1} \sum_{t=1}^n \beta_{i,t} w(I_{t-1}, \cdot)$, for some coefficients $\beta_{i,t}$, $t = 1, \dots, n$, and the equation to solve becomes

$$\frac{1}{n} \sum_{t=1}^n e_t^2(\theta_n) w(I_{t-1}, \cdot) \left[\frac{1}{n} \sum_{s=1}^n \beta_{i,s} \int_{\Pi} w(I_{t-1}, x) w^c(I_{s-1}, x) \Psi(dx) \right] = \lambda_{n,i,w} \frac{1}{n} \sum_{t=1}^n \beta_{i,t} w(I_{t-1}, \cdot).$$

Here $\beta_{i,t}$, $t = 1, \dots, n$, and $\lambda_{n,i,w}$ are the solutions of the system of n equations

$$\frac{1}{n} \sum_{s=1}^n \beta_{i,s} a_{ts} = \lambda_{n,i,w} \beta_{i,t} \quad 1 \leq i, t \leq n,$$

with $a_{ts} = \int_{\Pi} e_t^2(\theta_n) w(I_{t-1}, x) w^c(I_{s-1}, x) \Psi(dx)$. The solutions $\beta_i = (\beta_{i,1}, \dots, \beta_{i,n})'$ and $\lambda_{n,i,w}$ are the eigenelements of the $n \times n$ matrix A of elements $(1/n)a_{ts}$. From now on, $\varphi_{n,i,w}$ will be an orthonormalized eigenfunction associated to $\lambda_{n,i,w}$, with $\{\lambda_{n,i,w} : 1 \leq i \leq n\}$ ranked in decreasing order. Next result shows the consistency of these estimators. First, let denote by $\|\cdot\|$ the usual norm for linear bounded operators on H_1 , i.e.,

$$\|l\| = \sup_{\|h\|_{H_1} \leq 1} \|lh\|_{H_1}.$$

THEOREM 8: *Assume A1-A4. Then, under H_0*

$$\|C_{n,w} - C_w\| \longrightarrow 0 \text{ a.s..}$$

Note that the following inequalities hold

$$\sup_{i \geq 1} |\lambda_{n,i,w} - \lambda_{i,w}| \leq \|C_{n,w} - C_w\|$$

and

$$\|\varphi_{n,i,w} - \tilde{\varphi}_{i,w}\|_{H_1} \leq c_i \|C_{n,w} - C_w\|, \quad i \geq 1,$$

where c_i is a real number that depends only on $\lambda_{i,w}$ and $\tilde{\varphi}_{i,w} = \text{sgn} \left(\left\langle \varphi_{i,w}^{(n)}, \varphi_{i,w} \right\rangle_{H_1} \right) \varphi_{i,w}$ (sgn is the sign function, i.e., $\text{sgn}(x) = 1(x > 0) - 1(x < 0)$). Last inequalities and Theorem 8 imply the consistency of the estimated eigenlements.

Now, we shall discuss how to use the above asymptotic results to perform an approximated optimal directional test for testing H_0 against the family of alternatives parameterized by c , $c \in \mathbb{R} \setminus \{0\}$,

$$H_{A,n}(c) : Y_{t,n} = f(I_{t-1}, \theta_0) + \frac{ca(I_{t-1})}{n^{1/2}} + \varepsilon_t, \quad a.s., \quad (18)$$

where $a(\cdot)$ is as in Section 2.3 but with $E[a^2(I_{t-1})] = 1$.

We have seen that for directions a such that $D_{w,a} \neq 0$ in a subset of positive Lebesgue measure, the change from H_0 to $H_{A,n}(c : c \neq 0)$ delivers in a non-random shift in the mean function of the Gaussian process $R_{\infty,w}^1$. Therefore, tests for H_0 against $H_{A,n}(c : c \neq 0)$ can be viewed as tests for $\tilde{H}_0 : E[R_{\infty,w}^1] = 0$ against $\tilde{H}_{A,n} : E[R_{\infty,w}^1(\cdot)] = D_{w,a}(\cdot)$. In a fundamental work, Grenander (1952) generalized the optimal Neyman-Pearson theory to this framework, see also Neuhaus (1976) and Stute (1997). In particular, we can deduce optimal directional tests for testing H_0 against $H_{A,n}$ by means of the Neyman-Pearson Lemma in its functional form. Let assume that $D_{w,a}$ is a linear operator of a , i.e., $D_{w,a}(\cdot) = La(\cdot)$, where L is a linear operator from $H_2 = L_2(\mathbb{R}^d, F_I)$, the Hilbert space of all square F_I -integrable functions, to H_1 . Let denote by $\langle \cdot, \cdot \rangle_{H_2}$ y $\|\cdot\|_{H_2}$ the inner product and norm on H_2 , respectively. Also, L^* represents the adjoint operator of L , which by definition satisfies

$$\langle La, h \rangle_{H_1} = \langle a, L^*h \rangle_{H_2} \quad \forall h \in H_1, \forall a \in H_2.$$

Then, $E[a^2(I_{t-1})] = 1$ and Bessel's inequality imply that

$$\sum_{i=1}^{\infty} \lambda_{i,w}^{-1} \langle D_{w,a}, \varphi_{i,w} \rangle_{H_1}^2 = \sum_{i=1}^{\infty} \lambda_{i,w}^{-1} \langle a, L^* \varphi_{i,w} \rangle_{H_2}^2 \leq E[a^2(I_{t-1})] = 1.$$

Therefore, from Grenander (1952, p. 215) we have that the distribution of $R_{\infty,w}^1$ under the alternatives $H_{A,n}$, \mathbb{P}_{1a} say, is absolutely continuous with respect to the distribution of $R_{\infty,w}^1$ under the null, \mathbb{P}_0 . The Radom-Nikodym derivative equals

$$\frac{d\mathbb{P}_{1a}}{d\mathbb{P}_0}(h) = \exp \left(cZ_a(h) - \frac{1}{2}c^2 \right) \quad h \in H_1, \quad (19)$$

where $Z_a(h) = \sum_{i=1}^{\infty} \lambda_{i,w}^{-1} \langle h, \varphi_{i,w} \rangle_{H_1} \langle D_{w,a}, \varphi_{i,w} \rangle_{H_1}$. So, by the Neyman-Pearson's Lemma we obtain that the asymptotic optimal directional test for testing H_0 against $H_{A,n}(c : c \neq 0)$ has critical region $\{|Z_a(R_{n,w}^1)| \geq z_{\alpha/2}\}$, where z_{α} is the α -quantile of the standard $N(0, 1)$ -distribution.

Note that in the general case, the eigenfunctions $\varphi_{i,w}(\cdot)$ and eigenvalues $\lambda_{i,w}$ are unknown, and therefore, have to be estimated from the sample $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$. Here, we consider previous estimations $\{(\lambda_{n,i,w}, \varphi_{n,i,w}) : 1 \leq i \leq n\}$ to approximate the optimal directional test. We have that, for a finite sample size n , the (approximated) Neyman-Pearson α -level test for (2) against (18) has critical region

$$\left| \widehat{Z}_{a,m}(R_{n,w}^1) \right| \geq z_{\alpha/2},$$

where

$$\widehat{Z}_{a,m}(R_{n,w}^1) = \sum_{i=1}^m \frac{\widehat{\tau}_{i,w} \widehat{\epsilon}_{n,i,w}}{\widehat{\gamma}}$$

$$\widehat{\epsilon}_{n,i,w} = (\lambda_{n,i,w})^{-1/2} \langle R_{n,w}^1, \varphi_{n,i,w} \rangle_{H_1} \quad 1 \leq i \leq m,$$

$$\widehat{\tau}_{i,w} = (\lambda_{n,i,w})^{-1/2} \langle \widehat{D}_{w,a}, \varphi_{n,i,w} \rangle_{H_1} \quad 1 \leq i \leq m,$$

$$\widehat{D}_{w,a}(x) = \frac{1}{n} \sum_{t=1}^n a(I_{t-1}) w(I_{t-1}, x) - \widehat{G}'_w(x) \widehat{\xi}_a,$$

$$\widehat{G}_w(x) = \frac{1}{n} \sum_{t=1}^n g_t(\theta_n) w(I_{t-1}, x),$$

$$\widehat{\gamma}^2 = \sum_{i=1}^m \widehat{\tau}_{i,w}^2,$$

and $\widehat{\xi}_a$ and m are, respectively, a \sqrt{n} -consistent estimator for ξ_a and a user-chosen parameter, usually small because of the weights $\lambda_{i,w}$. Given the consistency of $\{(\lambda_{n,i,w}, \varphi_{n,i,w}) : 1 \leq i \leq n\}$ and Theorem 3 it is not difficult to show that, in distribution

$$\widehat{Z}_{a,m}(R_{n,w}^1) \longrightarrow N(0, 1) \text{ under } H_0,$$

whereas

$$\widehat{Z}_{a,m}(R_{n,w}^1) \longrightarrow N(\mu, 1) \text{ under } H_{A,n}(c),$$

with

$$\mu = \sum_{i=1}^m \lambda_{i,w}^{-1} \langle D_{w,a}, \varphi_{i,w} \rangle_{H_1}^2.$$

Now, we shall discuss about the efficient estimation of parameters using the minimum distance criterium (11) and the relationship between this optimal estimator and a MLE. Carrasco and Florens (2000) have shown in a i.i.d setup that optimal estimation under a similar context to (11) is possible and requires the use of C_w^{-1} , which is the counterpart of the inverse of the covariance matrix in the finite-dimensional framework. But, note that C_w is a compact operator, and therefore, is not invertible on the full reference space. It is possible to solve this problem by applying standard regularization operators techniques, see e.g. Section 15.5 in Kress (1999). Inverting C_w is equivalent to find the solution Φ of the Fredholm equation of the first kind

$$C_w \Phi = h \tag{20}$$

for a given $h \in H_1$. This equation is typically an ill-posed problem in contrast to the well posed problems, see, e.g., Groetsch (1993). Picard's Criterion, cf. Theorem 15.18 in Kress (1999), tell us that if $\sum_{i=1}^{\infty} \lambda_{i,w}^{-2} \langle h, \varphi_{i,w} \rangle_{H_1}^2 < \infty$, we can define the mapping C_w^{-1} as

$$C_w^{-1}h := \sum_{i=1}^{\infty} \frac{\langle h, \varphi_{i,w} \rangle_{H_1}}{\lambda_{i,w}} \varphi_{i,w}, \quad h \in H_1^1 + H_1^0,$$

that can be understood as a Moore-Penrose generalized inverse or a least squares solution of (20). Similarly, we may define the square root $C_w^{-1/2}$ of C_w^{-1} by

$$C_w^{-1/2}h := \sum_{i=1}^{\infty} \frac{\langle h, \varphi_{i,w} \rangle_{H_1}}{\lambda_{i,w}^{1/2}} \varphi_{i,w}, \quad h \in \mathcal{D}(C_w^{-1/2}),$$

where

$$\mathcal{D}(C_w^{-1/2}) := \left\{ h \in H_1 : \sum_{i=1}^{\infty} \frac{\langle h, \varphi_{i,w} \rangle_{H_1}^2}{\lambda_{i,w}} < \infty \right\},$$

with the convention that $0/0 = 0$. Then, based on the Tikhonov's regularization method, Carrasco and Florens (2000) propose the estimator for $C_w^{-1/2}$

$$C_{n,w}^{-1/2}h := \sum_{i=1}^n \frac{\lambda_{n,i,w}^{1/2}}{(\lambda_{n,i,w}^2 + \alpha_n)^{1/2}} \langle h, \varphi_{n,i,w} \rangle_{H_1} \varphi_{n,i,w},$$

where α_n is a sequence that converges to zero at a certain rate that will be specified below. The regularization parameter α_n is used to discard the least informative principal components, i.e., those associated to smallest eigenvalues. It can be selected by cross-validation methods, see e.g. Hansen (1992). Then, in a i.i.d setup and with our notation, Carrasco and Florens (2000) show that under $E[e_t^2(\theta)] < \infty$ and $E[g_t^2(\theta)] < \infty$, for any $\theta \in \Theta$, and that $n\alpha_n^3$ goes to infinity as $n \rightarrow \infty$ and $\alpha_n \rightarrow 0$, the estimator defined by

$$\theta_{n,e} := \arg \min_{\theta \in \Theta} \left\| C_{n,w}^{-1/2} n^{-1/2} R_{n,w}^1(\theta) \right\|_{H_1}^2$$

is asymptotically efficient. The proof in the time series case follows exactly the same steps but need of weak dependence assumptions (or a CLT for U-statistics) and is not considered here for the sake of space. Now, it is easy to show that the objective function of the asymptotically efficient estimator reduces to

$$\left\| C_{n,w}^{-1/2} n^{-1/2} R_{n,w}^1(\theta) \right\|_{H_1}^2 = \sum_{i=1}^n \frac{\lambda_{n,i,w}}{\lambda_{n,i,w}^2 + \alpha_n} \left\langle n^{-1/2} R_{n,w}^1(\theta), \varphi_{n,i,w} \right\rangle_{H_1}^2,$$

and therefore, $\theta_{n,e}$ is asymptotically equivalent to a MLE defined through the n first principal components $\{\epsilon_{i,w} : 1 \leq i \leq n\}$, i.e., through the likelihood (19) for testing $\tilde{H}_0 : E [R_{\infty,w}^1(\cdot, \theta)] \equiv 0$ against $\tilde{H}_A : E [R_{\infty,w}^1(\cdot, \theta)] \equiv E[e_t(\theta)w(I_{t-1}, \cdot)]$.

5. 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 some goodness-of-fit tests for the regression function 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 $w(I_{t-1}, x) = 1(\beta' I_{t-1} \leq u)$, $x = (\beta', u)'$. More concretely, we shall compare the extension to time series of the Cramér-von Mises test of Escanciano (2004a), the multivariate extension of Koul and Stute's (1999) tests and a version of the Bierens' (1982) test.

We briefly describe our simulation setup. Let $I_{t-1,P} = (Y_{t-1}, \dots, Y_{t-P})$ be the information set at time $t - 1$. In the simulations we consider the values $P = 3, 5$ and 7 .

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 (2004a) proposed the CvM test

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

where

$$R_{n,pro,P}^1(\beta, u) := \frac{1}{\hat{\sigma}_e \sqrt{n}} \sum_{t=1}^n e_t(\theta_n) 1(\beta' I_{t-1,P} \leq u)$$

and

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

For a simple algorithm to compute $PCVM_{n,P}$ see Appendix B in Escanciano (2004a).

Bierens (1982) proposed to use $w(I_{t-1}, x) = \exp(iI'_{t-1}x)$ as the weighting function in (4) and considered the Cramér-von Mises test statistic

$$CvM_{n,\exp,P} := \int_{\Pi} |R_{n,\exp,P}^1(x)|^2 \Psi(dx),$$

where

$$R_{n,\text{exp},P}^1(x) := \frac{1}{\widehat{\sigma}_\varepsilon \sqrt{n}} \sum_{t=1}^n e_t(\theta_n) \exp(ix' I_{t-1,P}),$$

and with $\Psi(dx)$ a suitable chosen integrating function. In order that $CvM_{n,\text{exp},P}$ 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 P -variate r.v.. In that case, $CvM_{n,\text{exp},P}$ simplifies to

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

Koul and Stute (1999) have considered a model diagnostic test for an autoregressive regression of order one, i.e., $P = 1$. Our previous theory provides the asymptotic theory for the non-transformed Koul and Stute's (1999) tests for the multivariate case. We denote by CvM_P and KS_P their Cramér-von Mises and Kolmogorov-Smirnov statistics, respectively. These statistics are based on the multivariate integrated regression function and are given by

$$CvM_P = \frac{1}{\widehat{\sigma}_\varepsilon^2 n^2} \sum_{j=1}^n \left[\sum_{t=1}^n e_t(\theta_n) 1(I_{t-1,P} \leq I_{j-1,P}) \right]^2,$$

$$KS_P = \max_{1 \leq i \leq n} \left| \frac{1}{\widehat{\sigma}_\varepsilon \sqrt{n}} \sum_{t=1}^n e_t(\theta_n) 1(I_{t-1,P} \leq I_{i-1,P}) \right|.$$

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

We consider the FDWB approximation for all the test statistics. Our Theorem 7 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 to this data set 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:

$$Y_t = 1.05 + 1.41Y_{t-1} - 0.77Y_{t-2} + \varepsilon_t. \tag{21}$$

2. AR(2) model with heteroskedasticity (ARHET):

$$Y_t = 1.05 + 1.41Y_{t-1} - 0.77Y_{t-2} + h_t \varepsilon_t,$$

$$h_t^2 = 0.1 + 0.1Y_{t-1}^2.$$

3. 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. TAR(2) model: $Y_t = \begin{cases} 0.62 + 1.25Y_{t-1} - 0.43Y_{t-2} + \varepsilon_t, & \text{if } Y_{t-2} \leq 3.25, \\ 2.25 + 1.52Y_{t-1} - 1.24Y_{t-2} + \varepsilon_t, & \text{if } Y_{t-2} > 3.25. \end{cases}$

6. 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 in the literature 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 1000 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. Random numbers were generated using IMSL ggnml subroutine. We employ a sequence $\{V_t\}$ of i.i.d Bernoulli variates given in (14) and (15).

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 in the heteroskedastic case $CvM_{n,\text{exp},P}$ presents some small size distortion (underrejection).

Please, insert Table 1 about here.

In Table 2 we report the empirical power against the AR(3) and ARMA(2,2) processes. It increases with the sample size n for all test statistics, as expected. It is shown that the Cramér-von Mises test $PCvM_{n,P}$ has the best empirical power in all cases. The empirical power for $CvM_{n,\text{exp},P}$ is low for these alternatives and, in general, less than CvM_P and KS_P , specially for large P . In Table 3 we show the RP for the TAR and EXPAR models. Again, the test statistic $PCvM_{n,P}$ has more empirical power in almost all cases. $CvM_{n,\text{exp},P}$ has good empirical power properties for these models, in particular, it overtakes CvM_P and KS_P . For the last two models the test statistics CvM_P and KS_P decrease in empirical power as the lag parameter P increases, possibly due to the problem of the curse of dimensionality, this is the case in general when nonlinear models are considered, see Escanciano (2004a).

Please, insert Tables 2 and 3 about here.

Now, we present an application of previous goodness-of-fit tests to the well-known Canadian lynx data set. This data set consists in 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 this data set see Tong (1990, pp. 357-418) and references therein. This data set has become a benchmark series to test new statistical methodology for time series analysis. The first time series model built on this particular data set was probably that of Moran (1953). Moran fitted the linear AR(2) model given in (21) 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 $PCvM_{n,P}$, $CvM_{n,\text{exp},P}$, CvM_P and KS_P in Table 4.

Please, insert Table 4 about here.

The AR(2) specification is rejected at 5% with $PCvM_{n,P}$ and $CvM_{n,\text{exp},P}$ for all values of P , whereas CvM_P and KS_P fails to reject for large values of P ($P = 6$ and 10). So, this specification is not satisfactory, this fact was realized by many authors, including Moran (1953). Now, we consider two further specifications for this data set, both considered in Tong (1990). First, we consider the TAR(2,2,2) model given in DGP 5. Again, we consider $P = 2, 4, 6$ and 10 and we report the empirical p-values for the test statistics in Table 5.

Please, insert Table 5 about here.

For $P = 2$ all the test statistics fail to reject the TAR(2,2,2) specification, whereas for $P = 4$ and 10, $PCvM_{n,P}$ and $CvM_{n,\text{exp},P}$ reject it at 5% level. The test statistics CvM_P and KS_P are in agreement with the TAR(2,2,2) model for $P = 4$ and 10, but again, this may be due to the curse of dimensionality and not because of a well 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.

Please, insert Table 6 about here.

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

Finally, we comment on some extensions for future research. Similar theory to that presented here is expected to hold for goodness-of-fit tests for conditional distributions functions, joint model checks for the conditional mean and conditional variance, covering AR-ARCH models, see Engle (1982), and other moment conditional restrictions, such as conditional symmetry. These topics are currently being investigated.

Acknowledgments. 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. Any errors are solely mine. This work is part of my PhD. Thesis at Universidad Carlos III de Madrid. Research funded by the Spanish Dirección General de Enseñanza Superior (DGES) reference number BEC2001-1270.

6. PROOFS

First, let consider the next lemma which corresponds to Theorems 1.5.4 and 1.5.6 of VW.

LEMMA A1. Let \mathcal{T} be a non-empty set. For every $n \in \mathbb{N}$ let $(\Omega_n, \mathcal{F}_n, P_n)$ be a probability space, and X_n be a mapping from Ω_n to $\ell^\infty(\mathcal{T})$. Consider the following statements:

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

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

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), then there exists a version \tilde{X} of X such that $X_n \implies \tilde{X}$ 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 previous lemma holds. To this end, fix a compact subset $\Pi_c \subset \Pi$, and using W2 we can choose a nested sequence of finite partitions $\mathcal{P}_q = \{B_{qk}; 1 \leq k \leq N_q\}$ of Π_c , for every $q \in \mathbb{N}$, $q \geq 1$, such that

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

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

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

To simplify notation define $M_t^n(x) = n^{-1/2} \varepsilon_t w(X_t, x)$. Then, by the previous lemma, see also the proof of Theorem 2.5.6 of VW, is sufficient to prove that for every $\varepsilon, \eta > 0$ there exists a $q_0 \in \mathbb{N}$ such that

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

where $\|\cdot\|_{\Pi_c}$ denotes the uniform norm on Π_c . To this end, fix any q_0 for a while, and let define the quantities for each fixed n and large $q \geq q_0$

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

and the events

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

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

and

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

Now, similarly to VW p. 131, we decompose

$$\begin{aligned} M_t^n(x) - M_t^n(\pi_{q_0}x) &= (M_t^n(x) - M_t^n(\pi_{q_0}x))D_{t, q_0}^n \\ &+ \sum_{q=q_0+1}^{\infty} (M_t^n(x) - M_t^n(\pi_q x))D_{t, q}^n \\ &+ \sum_{q=q_0+1}^{\infty} (M_t^n(\pi_q x) - M_t^n(\pi_{q-1}x))C_{t, q}^n. \end{aligned} \quad (22)$$

On the other hand, by (2)

$$\begin{aligned} 0 &= E[(M_t^n(x) - M_t^n(\pi_{q_0}x))D_{t, q_0}^n | X_t] \\ &+ \sum_{q=q_0+1}^{\infty} E[(M_t^n(x) - M_t^n(\pi_q x))D_{t, q}^n | X_t] + \sum_{q=q_0+1}^{\infty} E[(M_t^n(\pi_q x) - M_t^n(\pi_{q-1}x))C_{t, q}^n | X_t]. \end{aligned} \quad (23)$$

Now, by (22) and (??)

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

where

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

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

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

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

and

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

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

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

Hereafter, we perform estimations for terms I_1 , I_2 , II_3 , II_2 and III . First, from $\Delta_t^n(B_q x) \leq 2 \|M_t^n(x)\|_{\Pi_c}$, we have that

$$\Delta_t^n(B_{q_0} x) D_{t,q_0}^n \leq C n^{-1/2} |\varepsilon_t| 1(|\varepsilon_t| > C\sqrt{n}a_{q_0}) \quad \text{identically.}$$

Thus, using the last displayed, (2) 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

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

Then, for the estimation of II_2 , we see that

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

As for II_3 , since

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

and

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

it follows from the Freedman's (1975) inequality, which plays here the same roll 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

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

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

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

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

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

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

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

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

PROOF OF COROLLARY 1. We start with $1(X_t \leq x)$. The continuity of $d_{ind}(x_1, x_2)$ guarantees that we can form a partition $\mathcal{B}_\varepsilon = \{H_k : 1 \leq k \leq N_\varepsilon\}$ of Π_{ind} in ε -brackets $H_k = [x_k, y_k]$, i.e., $\{H_k : 1 \leq k \leq N_\varepsilon\}$ covers Π_{ind} , $x_k \leq y_k$, and $d_{ind}(x_k, y_k) \leq \varepsilon$. Then, condition (8) follows from the construction of the partition, the monotonicity of $1(X_t \leq x)$ and the Ergodic Theorem, jointly with a Glivenko-Cantelli's argument to show the uniformity in $\varepsilon > 0$. On the other hand, (7) follows because the class of functions $\{1(X_t \leq x) : x \in \Pi_{ind}\}$ is a Vapnik Cervonenkis class (VC), and therefore, it satisfies a uniform entropy condition, see VW. As for $w(X_t, x) = 1(\beta' X_t \leq u)$, note that

$$\begin{aligned}
& \sup_{x_1, x_2 \in H_k} |1(\beta'_1 X_t \leq u_1) - 1(\beta'_2 X_t \leq u_2)| \\
\leq & \sup_{x_1, x_2 \in H_k} |1(\beta'_1 X_t \leq u_1) - 1(\beta'_1 X_t \leq u_2)| \\
& + \sup_{x_1, x_2 \in H_k} |1(\beta'_1 X_t \leq u_2) - 1(\beta'_2 X_t \leq u_2)| \\
\leq & \sup_{x_1, x_2 \in H_k} |1(\beta'_1 X_t \leq \overline{u_k}(\beta_1)) - 1(\beta'_1 X_t \leq \underline{u_k}(\beta_1))| \\
& + \sup_{x_1, x_2 \in H_k} |1(\underline{\beta}'(u_2) X_t \leq u_2) - 1(\overline{\beta}'(u_2) X_t \leq u_2)|,
\end{aligned}$$

where $\{H_k : 1 \leq k \leq N_\varepsilon\}$ are the ε -closed balls constructed with the semimetric d_{pro} , and for fixed β_1 , $\underline{u_k}(\beta_1)$ and $\overline{u_k}(\beta_1)$ are the minimum and the maximum of the u 's such that $(u, \beta_1) \in H_k$, respectively, and for fixed u_2 , $\underline{\beta}'(u_2)$ and $\overline{\beta}'(u_2)$ are the directions that maximizes the angle among the β 's such that $(u_2, \beta) \in H_k$. Note, that these quantities exist because H_k is closed. By the same argument, the sup in last inequality is achieved in H_k . Therefore, conditions (7) and (8) follows from the same arguments as in $1(X_t \leq x)$. \square

PROOF OF THEOREM 2. It follows from Theorem 1. \square

PROOF OF THEOREM 3. Applying the classical mean value theorem argument we have

$$\begin{aligned}
R_{n,w}^1(x) &= R_{n,w}(x) - n^{-1/2} \sum_{t=1}^n \{f(I_{t-1}, \theta_n) - f(I_{t-1}, \theta_0)\} w(I_{t-1}, x) \\
&= R_{n,w}(x) - I - II - III
\end{aligned}$$

where

$$I = n^{1/2}(\theta_n - \theta_0)' \frac{1}{n} \sum_{t=1}^n \{g(I_{t-1}, \theta_{ni}) - g(I_{t-1}, \theta_0)\} w(I_{t-1}, x),$$

$$II = n^{1/2}(\theta_n - \theta_0)' \frac{1}{n} \sum_{t=1}^n [g(I_{t-1}, \theta_0) w(I_{t-1}, x) - G_w(x, \theta_0)]$$

and

$$III = n^{1/2}(\theta_n - \theta_0)' G_w(x, \theta_0),$$

and where θ_{ni} satisfies $|\theta_{ni} - \theta_0| \leq |\theta_n - \theta_0|$ a.s.. By A1-A3, A4(a) and the uniform law of large numbers (ULLN) of Jennrich (1969, Theorem 2) 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). \square

PROOF OF THEOREM 4: For the consistency (i), it is sufficient to show that

$$\int_{\Pi} n^{-1} |R_{n,w}^1(x, \theta)|^2 \Psi(dx) \longrightarrow \int_{\Pi} |E[e_t(\theta) w(I_{t-1}, x)]|^2 \Psi(dx) \text{ a.s., uniformly in } \theta.$$

This result holds applying the Continuous Mapping Theorem, since

$$n^{-1/2}R_{n,w}^1(x, \theta) \longrightarrow E[e_t(\theta)w(I_{t-1}, x)] \text{ a.s., uniformly in } (x, \theta) \in \Pi \times \Theta,$$

which follows from A4(a), A1' and from Theorem 2 in Jennrich (1969). As to the asymptotic normality (ii), for simplicity we only consider the case of real-valued weighting function $w(I_{t-1}, x)$, the complex-valued case follows exactly the same steps but with some cumbersome notation. The first order conditions of the minimization problem are

$$\int_{\Pi} n^{-1}R_{n,w}^1(x, \theta_n) \frac{\partial R_{n,w}^1(x, \theta_n)}{\partial \theta} \Psi(dx) = 0.$$

A2 and the Mean Value Theorem imply

$$\int_{\Pi} n^{-1}R_{n,w}^1(x, \theta_0) \frac{\partial R_{n,w}^1(x, \theta_n)}{\partial \theta} \Psi(dx) + C_n(\theta_n - \theta_0) = 0,$$

with

$$C_n = \int_{\Pi} n^{-1} \frac{\partial R_{n,w}^1(x, \theta_n)}{\partial \theta} \frac{\partial R_{n,w}^1(x, \theta^*)}{\partial \theta} \Psi(dx),$$

and where the vector θ^* is such that there exists a (random) $\lambda \in [0, 1]$, with $\theta^* = \lambda\theta_0 + (1 - \lambda)\theta_n$ (a.s.). Therefore, for n sufficiently large

$$\sqrt{n}(\theta_n - \theta_0) = -C_n^{-1} \int_{\Pi} R_{n,w}^1(x, \theta_0) n^{-1/2} \frac{\partial R_{n,w}^1(x, \theta_n)}{\partial \theta} \Psi(dx).$$

Now, note that, again, by A1, A2, A4(a) and from Theorem 2 in Jennrich (1969) we have that

$$n^{-1/2} \frac{\partial R_{n,w}^1(x, \theta)}{\partial \theta} \longrightarrow E[g_t(\theta)w(I_{t-1}, x)] \text{ a.s., uniformly in } (x, \theta) \in \Pi \times \Theta.$$

Then, the result follows from Slutsky's Theorem and Theorem 3. \square

PROOF OF THEOREM 5: From A4(a), under A1-A3

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

uniformly in $x \in \Pi_c$. \square

PROOF OF THEOREM 6: Under the local alternatives (12) write

$$\begin{aligned} R_{n,w}^1(x) &= n^{-1/2} \sum_{t=1}^n \left\{ f(I_{t-1}, \theta_0) + \frac{a(I_{t-1})}{n^{1/2}} + \varepsilon_t - f(I_{t-1}, \theta_n) \right\} w(I_{t-1}, x) \\ &: = R_{n,w}(x) + A_1 + A_2, \end{aligned} \tag{24}$$

with

$$A_1 = n^{-1/2} \sum_{t=1}^n \{f(I_{t-1}, \theta_0) - f(I_{t-1}, \theta_n)\} w(I_{t-1}, x)$$

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

$$\left| A_1 + n^{1/2}(\theta_n - \theta_0)' G_w(x, \theta_0) \right| = o_P(1)$$

uniformly in $x \in \Pi_c$. On the other hand, A4(a) 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 equations and (24), the theorem holds by A3' and Theorem 3. \square

PROOF OF THEOREM 7. We need to show that the process $R_{n,w}^{1*}$ (conditionally on the sample) has the same asymptotic finite dimensional distributions that the process $R_{n,w}^1$, with θ_* replacing θ_0 , and that $R_{n,w}^{1*}$ is asymptotically tight, both with probability one. Then, write similarly to Theorem 3

$$\begin{aligned} R_{n,w}^{1*}(x) &= n^{-1/2} \sum_{t=1}^n e_t^*(\theta_n) w(I_{t-1}, x) - n^{-1/2} \sum_{t=1}^n \{f(I_{t-1}, \theta_n^*) - f(I_{t-1}, \theta_n)\} w(I_{t-1}, x) \\ &= R_{n,w}^*(x) - I^* - II^* - III^*, \end{aligned}$$

where

$$I^* = n^{1/2}(\theta_n^* - \theta_n)' \frac{1}{n} \sum_{t=1}^n \{g(I_{t-1}, \theta_{ni}^*) - g(I_{t-1}, \theta_n)\} w(I_{t-1}, x),$$

$$II^* = n^{1/2}(\theta_n^* - \theta_n)' \frac{1}{n} \sum_{t=1}^n [g(I_{t-1}, \theta_n) w(I_{t-1}, x) - G_w(x, \theta_*)]$$

and

$$III^* = n^{1/2}(\theta_n^* - \theta_n)' G_w(x, \theta_*),$$

and where θ_{ni}^* satisfies $|\theta_{ni}^* - \theta_n| \leq |\theta_n^* - \theta_n|$ a.s. (conditionally on the sample). Under our assumptions is easy to show that conditionally on the sample, $I^* = o_P(1)$ and $II^* = o_P(1)$ uniformly in $x \in \Pi_c$, with probability one. So, uniformly in $x \in \Pi_c$

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

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 Theorem 2.5.2 in VW. The proof is finished. \square

PROOF OF THEOREM 8: The result follows from the inequality

$$\|C_{w,n} - C_w\| \leq \int_{\Pi} \int_{\Pi} \left[\frac{1}{n} \sum_{t=1}^n k(Y_t, I_{t-1}, x, y, \theta_n) - E[e_t^2(\theta_0)w(I_{t-1}, x)w^c(I_{t-1}, y)] \right]^2 \Psi(dx)\Psi(dy),$$

where

$$k(Y_t, I_{t-1}, x, y, \theta_n) = e_t^2(\theta_n)w(I_{t-1}, x)w^c(I_{t-1}, y),$$

and using a mean value argument, A1-A4 and the Ergodic Theorem. \square

REFERENCES

- ANDERSON, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceed. Amer. Mathem. Soc.* **6** 170-176.
- BAE, J. AND LEVENTAL, S. (1995). Uniform CLT for Markov chains and its invariance principle: a martingale approach. *Journal of Theoretical Probability* **8** 549-570.
- BIERENS, H. J. (1982). Consistent model specification tests. *Journal of Econometrics* **20** 105-134.
- BIERENS, H. J. (1984). Model specification testing of time series regressions. *Journal of 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. *Proceedings of the American Mathematical Society* **12** 788-792.
- CARRASCO, M. AND FLORENS, J.P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* **16** 797-834.
- 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. *Journal of Econometrics* **72** 1-32.
- DEHLING, H., AND PHILIPP, W. (2002). Empirical process techniques for dependent data. In *Empirical processes techniques for dependent data*. H. Dehling, T. Mikosch and M. Sorensen (Eds.). Birkhauser Boston, p. 1-115, 2002.
- ENGLE, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50** 987-1008.
- ESCANCIANO, J. C. (2004a). A consistent diagnostic test for regression models using projections. Preprint.
- ESCANCIANO, J. C. (2004b). Specification test of econometric time series models. PhD. Dissertation. Universidad Carlos III de Madrid.
- ESCANCIANO, J. C. AND VELASCO, C. (2003). Generalized spectral tests for the martingale difference hypothesis. Under revision in *Journal of Econometrics*.

- 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 nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85** 387-392.
- FAN, J. AND HUANG, L. (2001). Goodness-of-fit tests for parametric regression models. *Journal of the American Statistical Association* **96** 640-652.
- FAN, J. AND YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.
- FAN, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters. *Econometric Theory* **14** 604-621.
- FAN, Y. AND LI, Q. (1996). Consistent model specification tests: omitted variables, parametric and semiparametric functional forms. *Econometrica* **64** 865-890.
- FAN, Y. AND LI, Q. (2000). Consistent model specification tests: kernel-based tests versus Bierens' ICM tests. *Econometric Theory* **16** 1016-1041.
- FRANKE, J., KREISS, J. P., AND MAMMEN, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8**(1) 1-37.
- FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100-118.
- GINÉ, E., AND J. ZINN (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851-869.
- GRENANDER, U. (1952). Stochastic processes and statistical inference. *Arkiv for Matematik* **1** 195-277.
- GROETSCH, C. (1993). *Inverse Problems in Mathematical Sciences*. Wiesbaden: Vieweg.
- 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.
- HANSEN, P.C. (1992). Numerical tools for analysis and solution of Fredholm integral equations of the first kind. *Inverse Problems* **8** 849-872.
- 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. *Econometric Theory* **10** 821-848.
- JENNRICH, R.I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Mathem. Statist.* **40** 633-643.
- KOUL, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*. 2nd ed, Lecture

- Notes in Statistics, Vol. 166, Springer.
- KOUL, H.L. AND NI, P. (2004). Minimum distance regression model checking. *Journal of Statistical Planning and Inference* **119** 109-144.
- KOUL, H.L., AND STUTE W. (1999). Nonparametric model checks for time series. *Ann. Statist.* **27** 204-236.
- KREISS, J.P., NEUMANN, M.H., AND YAO, Q. (1998). Bootstrap tests for simple structures in nonparametric time series regression. Preprint.
- KRESS, R. (1999). *Linear Integral Equations*. Springer.
- LEVENTAL, S. (1989). A uniform CLT for uniformly bounded families of martingale differences. *Journal of Theoretical Probability* **2** 271-287.
- LI, Q. (1999). Consistent model specification test for time series econometric models. *Journal of Econometrics* **92** 101-147.
- LI, Q. AND WANG, S. (1998). A simple consistent bootstrap test for a parametric regression functional form. *Journal of Econometrics* **87** 145-165.
- LI, Q., HSIAO, C. AND ZINN, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics* **112** 295-325.
- LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d models. *Ann. Statist.* **16** 1696-1708.
- MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255-285.
- MILBRODT, H. AND STRASSER H. (1990). On the asymptotic power of the two-sided Kolmogorov-Smirnov test. *Journal of Statistical Planning and Inference* **26** 1-23.
- MORAN, P.A.P. (1953). The statistical analysis of the Canadian lynx cycle, I: structure and prediction. *Aust. J. Zool.* **1** 163-173.
- NEUHAUS, G., (1976). Asymptotic power properties of the Cramér-von Mises test under contiguous alternatives. *J. Multiv. Anal.* **6** 95-110.
- NEUMANN, M.H. AND KREISS, J.P. (1998). Regression-type inference in nonparametric autoregression. *Ann. Statist.* **26** 1570-1613.
- NISHIYAMA, Y. (2000). Weak convergence of some classes of martingales with jumps. *Ann. Probab.* **28** 685-712.
- OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897-919.
- PARTHASARATHY, K.R. (1967). *Probability Measures on Metric Spaces*. New York. Academic Press.

- POLITIS, D. AND ROMANO, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica* **4** 461-476.
- STINCHCOMBE, M., AND WHITE, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric 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. *Scandinavian Journal of Statistics* **29** 535-545.
- TAYLOR, A. E. Y LAY, D. C. (1980). *Introduction to Functional Analysis*. Wiley, New York.
- TONG, H. (1990). *Nonlinear Time Series: A Dynamic system Approach*. Clarendon Press: Oxford.
- TRIPATHI, G. Y KITAMURA, Y. (2003). Testing conditional moment restrictions. *Ann. Statist.* **31** 2059-2095.
- 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. Mathem. Statist.* **25** 131-138.
- WOOLDRIDGE, J. (1992). A test for functional form against nonparametric alternatives. *Econometric 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. *Econometric Theory* **8** 435-451.
- ZHENG, X. (1996). A consistent test of functional form via nonparametric estimation technique. *Journal of Econometrics* **75** 263-289.

FACULTAD DE ECONÓMICAS
 UNIVERSIDAD DE NAVARRA
 EDIFICIO BIBLIOTECA (ENTRADA ESTE)
 PAMPLONA, 31080, NAVARRA, SPAIN
 E-MAIL: jescanci@unav.es

TABLE 1. *Size of 5% Tests.*

$n=100$	AR(2)			ARHET		
	P	3	5	7	3	5
PCvM $_{n,P}$	4.6	5.2	5.1	5.3	4.3	4.2
CvM $_{n,exp,P}$	4.4	5.4	5.4	3.2	2.1	2.0
CvM $_P$	5.2	5.0	4.0	5.5	5.1	5.1
KS $_P$	4.7	5.4	5.1	6.0	7.4	5.5

TABLE 2. *Power of 5% Tests.*

P		AR(3)			ARMA(2,2)		
		3	5	7	3	5	7
$n=100$	PCvM $_{n,P}$	61.9	56.4	40.7	30.7	46.7	34.7
	CvM $_{n,exp,P}$	25.0	18.2	14.2	11.3	10.6	9.1
	CvM $_P$	22.1	33.2	25.9	15.2	25.4	12.9
	KS $_P$	22.0	26.3	17.7	13.7	21.9	12.4
$n=200$	PCvM $_{n,P}$	91.2	88.4	76.7	61.2	80.7	71.2
	CvM $_{n,exp,P}$	45.9	32.3	21.8	16.4	12.6	10.8
	CvM $_P$	34.3	54.7	46.1	24.1	49.1	24.4
	KS $_P$	36.5	48.4	36.5	21.6	44.9	20.4

TABLE 3. *Power of 5% Tests.*

P		TAR			EXPAR		
		3	5	7	3	5	7
$n=100$	PCvM $_{n,P}$	76.2	40.0	27.9	81.2	57.1	47.9
	CvM $_{n,exp,P}$	52.2	31.3	19.4	73.1	65.7	62.6
	CvM $_P$	60.9	4.9	9.0	72.7	11.5	6.5
	KS $_P$	40.4	7.4	11.5	53.3	4.7	0.0
$n=200$	PCvM $_{n,P}$	98.2	79.3	57.4	98.9	94.4	80.2
	CvM $_{n,exp,P}$	89.0	69.4	44.1	97.8	96.6	92.6
	CvM $_P$	90.9	15.2	16.1	97.3	34.4	4.4
	KS $_P$	73.6	22.0	27.6	89.3	12.8	0.6

TABLE 4. *p*-values for the Canadian lynx data*AR(2) model*

<i>P</i>	2	4	6	10
PCvM _{<i>n,P</i>}	0.000	0.000	0.042	0.000
CVM _{<i>n,exp,P</i>}	0.002	0.000	0.016	0.000
CvM _{<i>P</i>}	0.000	0.004	0.066	0.090
KS _{<i>P</i>}	0.000	0.008	0.196	0.090

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

<i>P</i>	2	4	6	10
PCvM _{<i>n,P</i>}	0.322	0.008	0.086	0.012
CVM _{<i>n,exp,P</i>}	0.180	0.004	0.094	0.016
CvM _{<i>P</i>}	0.174	0.002	0.098	0.114
KS _{<i>P</i>}	0.408	0.022	0.182	0.140

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

<i>P</i>	7	8	9	10
PCvM _{<i>n,P</i>}	0.990	0.974	0.958	0.966
CVM _{<i>n,exp,P</i>}	0.966	0.874	0.834	0.824
CvM _{<i>P</i>}	0.674	0.178	0.476	0.600
KS _{<i>P</i>}	0.342	0.132	0.390	0.486