Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

Working Paper nº 07/05

On the Asymptotic Power Properties of Specification Tests for Dynamic Parametric Regressions

J. Carlos Escanciano

Facultad de Ciencias Económicas y Empresariales
Universidad de Navarra

ON THE ASYMPTOTIC POWER PROPERTIES OF SPECIFICATION TESTS FOR DYNAMIC PARAMETRIC REGRESSIONS

J. Carlos Escanciano

Working Paper No.07/05
May 2005
JEL No. C12, C14, C52.

ABSTRACT

Economic theories in dynamic contexts usually impose certain restrictions on the conditional mean of the underlying economic variables. Omnibus specification tests are the primary tools to test such restrictions when there is no information on the possible alternative. In this paper we study in detail the power properties of a large class of omnibus specification tests for parametric conditional means under time series processes. We show that all omnibus specification tests have a preference for a finite-dimensional space of alternatives (usually unknown to the practitioner) and we characterize such space for Cramér-von Mises (CvM) tests. This fact motivates the use of optimal tests against such preferred spaces instead of the omnibus tests. We proposed new asymptotically optimal directional and smooth tests that are optimally designed for cases in which a finite-dimensional space of alternatives is in mind. The new proposed optimal procedures are asymptotically distribution-free and are valid under weak assumptions on the underlying data generating process. In particular, they are valid under possibly time varying higher conditional moments of unknown form, e.g., conditional heteroskedasticity. A Monte Carlo experiment shows that previous asymptotic results provide good approximations in small sample sizes. Finally, an application of our theory to test the martingale difference hypothesis of some exchange rates provides new information on the rejection of omnibus tests and illustrates the relevance of our results for practitioners.

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 for helpful comments. Parts of this paper were written while I was visiting the Cowles Foundation at Yale University, whose hospitality is gratefully acknowledged. Research funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON and by the Universidad de Navarra reference number 16037001.

ON THE ASYMPTOTIC POWER PROPERTIES OF SPECIFICATION TESTS FOR DYNAMIC PARAMETRIC REGRESSIONS*

By J. Carlos Escanciano†

Universidad de Navarra

May 31, 2005

Abstract

Economic theories in dynamic contexts usually impose certain restrictions on the conditional mean of the underlying economic variables. Omnibus specification tests are the primary tools to test such restrictions when there is no information on the possible alternative. In this paper we study in detail the power properties of a large class of omnibus specification tests for parametric conditional means under time series processes. We show that all omnibus specification tests have a preference for a finite-dimensional space of alternatives (usually unknown to the practitioner) and we characterize such space for Cramér-von Mises (CvM) tests. This fact motivates the use of optimal tests against such preferred spaces instead of the omnibus tests. We proposed new asymptotically optimal directional and smooth tests that are optimally designed for cases in which a finite-dimensional space of alternatives is in mind. The new proposed optimal procedures are asymptotically distribution-free and are valid under weak assumptions on the underlying data generating process. In particular, they are valid under possibly time varying higher conditional moments of unknown form, e.g., conditional heteroskedasticity. A Monte Carlo experiment shows that previous asymptotic results provide good approximations in small sample sizes. Finally, an application of our theory to test the martingale difference hypothesis of some exchange rates provides new information on the rejection of omnibus tests and illustrates the relevance of our results for practitioners.

*Research funded by the Spanish Ministry of Education and Science reference number SEJ2004-04583/ECON and by the Universidad de Navarra reference number 16037001.

JEL Classification: C12, C14, C52.

Key words and phrases. Conditional mean, nonlinear time series, power function, omnibus tests, smooth tests, directional tests.

†Corresponding address: Universidad de Navarra, Facultad de Económicas, Edificio Biblioteca (Entrada Este), Pamplona, 31080, Navarra, Spain, e-mail: jescanci@unav.es. 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 for helpful comments. Parts of this paper were written while I was visiting the Cowles Foundation at Yale University, whose hospitality is gratefully acknowledged.

1. INTRODUCTION

Economic theories in dynamic contexts usually impose certain restrictions on the conditional mean function of the underlying economic variables. Omnibus specification tests are the primary tools to test such restrictions when there is no information on the possible alternative. They are intended to have some power against all alternatives. However, in econometric applications practitioners are sometimes interested in knowing if the rejection of onmibus tests has been caused by departures in the direction of some specific alternatives. For instance, in a simple linear regression model the econometrician might not be worried about a misspecification of the linear model as long as the errors are uncorrelated with the regressors. When particular alternatives are in mind optimal tests are possible. The first main purpose of this paper is to proposed optimal tests when there is only one specific alternative in mind and when there is a finite set of them (more than one). We call these optimal procedures *optimal directional test* and *optimal smooth test*, respectively. Our second target is to study in detail the asymptotic power properties of omnibus specification tests for dynamic regressions. As a result of this study, we show that all omnibus tests have a preference for a finite-dimensional space of alternatives. Apart from this "preferred" space the power of the omnibus test is almost flat. This fact motivates the use of smooth tests against the preferred space instead of the omnibus test. Directional and smooth tests are not only useful because they focus their power on the desired alternatives, but also because they provide information on an alternative model in the case of rejection, see Rayner and Best (1989). The asymptotic power properties of omnibus tests and the design of directional and smooth tests in the context of classical goodness-of-fit tests for distributions functions are now well-developed and have been a large field of study since the initial work by Pearson (1900). This fact contrasts with that of the specification tests for conditional mean functions, or more generally, with the literature on conditional moment restrictions, where there have been few works focused on these problems. The main purpose of this paper is to help to fill this gap.

More concretely, we consider the so-called integrated-based tests that provide a large family of omnibus specification tests for dynamic regression models, see Bierens (1982, 1984, 1990), de Jong (1996), Stute (1997), Bierens and Ploberger (1997), Koul and Stute (1999), Whang (2001), Domínguez and Lobato (2003) or Escanciano (2004a), among many others. In the integrated approach tests are based on a general class of residual marked processes (RMP). All tests considered in this paper, omnibus, smooth and directional, are continuous functionals of these RMP. Therefore, we show that the RMP are the building-blocks for a unified theory of a large class of specification tests for parametric conditional means with different power properties and different purposes.

We first study in some detail the asymptotic local power function (ALPF) of the omnibus integrated-based tests. Omnibus tests are capable of detecting every misspecification asymptotically, i.e., they are consistent. But such as assertion is only useful, if one knows which types of deviations can be detected with a reasonable sample size, and for which other alternatives its power is rather poor. In addition, since there may be several competing omnibus tests with different power properties that are usually unknown a priory, the practitioner faces the problem of which test to use. To overcome these two problems, we define asymptotic local relative efficiency (ALRE) measures between different tests that can be used for comparison purposes. These efficiency measures may help to practitioners to choose the best test when a particular direction is in mind and to check which deviations are well detected and which are bad detected for a specific integrated-based test. We show that all the omnibus integrated-based tests have reasonable power only against a set of alternatives belonging to a finite-dimensional subset. Apart form this space the power is almost flat. We characterize such "preferred" space for the Cramér-von Mises (CvM) tests and propose a candidate for it in the case of a general omnibus integrated-based tests.

In the second part of the paper we propose optimal directional and smooth tests in the context of specification tests for dynamic regressions. These optimal procedures are very convenient when particular alternatives are in mind. By applying the smooth methodology to the preferred space of the omnibus tests we obtain smooth versions of the omnibus tests that are optimal against such preferred space, and therefore, compare very well with the omnibus tests. Contrary to the omnibus tests, these smooth versions are asymptotically distribution-free, so critical values can be tabulated. To compute the smooth versions of the omnibus tests we need estimations of the principal components of the RMP. We provide such estimations and show their consistency. These estimations are also useful for computing the ALRE.

Since the fundamental work by Pearson (1900) there has been a large body of statistical literature devoted to the study of the goodness-of-fit tests for distributions functions and the power properties of such tests. In this framework the asymptotic behavior of the well-known CvM test has been investigated in Anderson and Darling (1952), Durbin and Knott (1972) and Neuhaus (1976), among others. Hájek and Šidák (1967) and Milbrodt and Strasser (1990) studied the one-sided and two-sided Kolmogorov-Smirnov (KS) test, respectively. See also Janssen (1995). It is well-known that the CvM and KS tests are omnibus. The literature on smooth tests began with the seminal work by Neyman (1937). See Rayner and Best (1989) for a monograph on smooth tests for distributions. Neyman's (1937) test has been studied and generalized by numerous authors, see, e.g., Kallenberg and Ledwina (1997) and references therein.

The literature on the asymptotic power properties of tests and the design of directional and smooth tests in the context of specification tests for regressions is scarce. Stute (1997) proposed omnibus, smooth and directional tests for regression models using the nonparametric principal components of the underlying RMP, see also Stute, Thies and Zhu (1998). The smooth tests considered by these authors are smooth versions of the CvM test for univariate regressions. Fan and Huang (2001) consider data-driven smooth tests using Fourier transforms for linear models with Gaussian errors, extending previous work by Fan (1996) to regressions. These works assume independent and identically distributed (*iid*) observations. In a time series framework, Bierens and Ploberger (1997) study the power properties of some integrated-based tests under conditional homoscedasticity. However, they restricted the analysis to the CvM tests and their main interest was to prove the asymptotic admissibility of the CvM test. Our results extend these works in several aspects. We consider a much larger class of directional and smooth tests, not only smooth versions of the CvM tests but also smooth tests against any finite set of alternatives. Furthermore, even for the smooth versions of the CvM tests our proposal uses new estimators of the principal components different and more general than those considered in Stute (1997). Our study of the power properties of omnibus tests considers a general continuous functional, including but not restricting to CvM tests. In particular, our analysis covers KS-type functionals. Also, here we are concerned with the development of measures for comparing different tests and the computation of such measures in practice. Finally, our assumptions are very weak; they are valid for time series processes with multivariate regressors and under higher conditional moments of unknown form, in particular under conditional heteroskedasticity. Note that this is very important for econometric applications. We would like to stress at this point that the arguments used in our theory are not exclusive of the specification tests for conditional means and that they hold for more general conditional moment restrictions under additional mild assumptions. However, to make the exposition simpler, we have restricted ourselves to specification tests for time series regressions.

The paper is organized as follows. In Section 2 we review the integrated methodology for specification tests of regression functions and we introduce the assumptions. In Section 3 we study some analytical properties of the ALPF of the integrated-based tests as a function of the distance and the direction to the null. We find the directions of maximum power for the CvM tests. We show that all omnibus integrated-based tests have a preference for a finite-dimensional space of alternatives. We characterize such space for CvM tests. We also compute the slope of the ALPF of general functionals, that allows us to define an ALRE concept very useful for comparing different tests. We define a large class of optimal directional tests in Section 4. Section 5 is devoted to the design of smooth tests against a finite-set of alternatives and for the smooth versions of CvM tests. In Section 6 we propose new estimators of the principal components of the RMP and show their consistency. These estimations are necessary to put some previous theory into practice. A Monte Carlo experiment in Section 7 shows that the previous asymptotic theory is an acceptable approximation for finite samples. Finally, an empirical application to some exchange rates in Section 8 highlights the merits of our approach and illustrates the relevance of our results for practitioners. Proofs are deferred to Section 9.

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. $\overline{\mathbb{R}}^d$ denotes the extended d-dimensional Euclidean space, i.e., $\overline{\mathbb{R}}^d = [-\infty, \infty]^d$. In what follows, Π_c will denote a compact subset of $\Pi \subseteq \overline{\mathbb{R}}^q$, and let $\ell^\infty(\Pi)$ be the space of all complex-valued functions that are uniformly bounded on Π_c, for all compact subsets of $\Pi \subseteq \overline{\mathbb{R}}^q$. Let $\Longrightarrow$ 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). Note that if Π is compact (e.g., $\overline{\mathbb{R}}^q$), then $\Longrightarrow$ reduces to the classical weak convergence concept of Hoffmann-Jørgensen (see Chapter 1.5 in 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 limits are taken as the sample size $n \to \infty$.

2. INTEGRATED-BASED TESTS FOR MODEL CHECKS

To begin with, let us consider the dependent variable $Y_t \in \mathbb{R}$, and the information set at time $t-1$, $I_{t-1} \in \mathbb{R}^d$, $d \in \mathbb{N}$, say, that is given by $I_{t-1} = (W'_{t-1}, Z'_{t-1})'$, where $Z_{t-1} \in \mathbb{R}^m$, $m \in \mathbb{N}$, is a m-dimensional observable random variable (r.v) and $W_{t-1} = (Y_{t-1}, ..., Y_{t-s}) \in \mathbb{R}^s$, so $d = s + m$. We shall assume throughout the paper that $\{(Y_t, I'_{t-1})' : t = 0, \pm 1, \pm 2, ...\}$ is a strictly stationary and ergodic time series process defined on the probability space $(\Omega, \mathcal{F}, P)$ and such that Y_t is P-integrable. Under the assumed conditions, we can write the tautological expression

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

where $f(z) := E[Y_t \mid I_{t-1} = z]$, $z \in \mathbb{R}^d$, is the conditional mean function of Y_t given the information set I_{t-1} and $\varepsilon_t := Y_t - E[Y_t \mid I_{t-1}]$. Then, in parametric time series models one assumes the existence of a parametric family of functions $\mathcal{M} = \{f(\cdot, \theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ and proceeds to test the hypothesis $f \in \mathcal{M}$. 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 $\mathcal{M}$ include linear and nonlinear autoregressive models, such as Markov-switching, exponential or threshold autoregressive models among many others, see, e.g., Fan and Yao (2003). We say that $f(I_{t-1}, \theta)$ is correctly specified for $f(I_{t-1})$ when there exists some θ_0 in $\Theta \subset \mathbb{R}^p$ such that $f(I_{t-1}, \theta_0) = f(I_{t-1})$ almost surely (a.s.). The correct specification of the conditional mean is important in order to avoid wrong conclusions in statistical inferences based on the parametric model $f(I_{t-1}, \theta_0)$. Our target is then to test the hypothesis that $f \in \mathcal{M}$, i.e.,

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

against the nonparametric alternatives

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

or against the local alternatives

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

where $a \in \mathcal{A}$, and $\mathcal{A}$ is the space of all measurable functions $a(\cdot) : \mathbb{R}^d \longrightarrow \mathbb{R}$ that are P-measurable, with zero mean, bounded variance and satisfy $P(a(I_{t-1}) = 0) < 1$. In the local alternatives (1), c represents the distance from the alternative to H_0 and a the direction of the alternative.

Let us define the parametric error $e_t(\theta) := Y_t - f(I_{t-1}, \theta)$, $t \in \mathbb{Z}$. It is easy to see that H_0 is tantamount to

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

The literature on testing the correct specification of regression models is huge. A partial list of works can be found in Escanciano (2004a). This extensive literature can be divided in two approaches. The first class of tests uses nonparametric smoothing estimations of $E[e_t(\theta_0) \mid I_{t-1}]$ and proceeds to test condition (2), see Wooldridge (1992), Yatchew (1992), Horowitz and Härdle (1994) or Zheng (1996), to mention a few. 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 a review of the local approach when $d = 1$.

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

$$E[e_t(\theta_0) \mid I_{t-1}] = 0 \ a.s. \Longleftrightarrow E[e_t(\theta_0) w(I_{t-1}, x)] = 0, \text{ almost everywhere } (a.e.) \text{ in } \Pi \subseteq \overline{\mathbb{R}}^q, \tag{3}$$

where $\Pi \subseteq \overline{\mathbb{R}}^q$, $q \in \mathbb{N}$, is a properly chosen space and the parametric family $\{w(\cdot, x) : x \in \Pi\}$ is such that the equivalence (3) 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 (3) 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. See Fan and Li (2000) for a comparison between the integrated and local approaches.

Since the initial work by Bierens (1982) there has been a large body of literature using the integrated approach. Bierens (1982) considered the exponential weight function $w(I_{t-1}, x) = \exp(ix' I_{t-1})$

in (3), where $i = \sqrt{-1}$ denotes the imaginary unit, whereas Stute (1997) used the indicator function $w(I_{t-1}, x) = 1(I_{t-1} \leq x)$. Bierens and Ploberger (1997) proposed a general class of weight functions including $w(I_{t-1}, x) = \sin(x' I_{t-1})$ or $w(I_{t-1}, x) = 1/(1 + \exp(c - x' I_{t-1}))$ with $c \in \mathbb{R}$, $c \neq 0$, among many others. Recently, Escanciano (2004a) has considered $w(I_{t-1}, x) = 1(\beta' I_{t-1} \leq u)$, with $x = (\beta', u)' \in \Pi_{pro} = \mathbb{S}^d \times [-\infty, \infty]$, where $\mathbb{S}^d$ is the unit ball in $\mathbb{R}^d$, i.e., $\mathbb{S}^d = \{\beta \in \mathbb{R}^d : |\beta| = 1\}$, as a combination of Bierens-Stute weights. See Stinchcombe and White (1998) for other families $\{w(\cdot, x) : x \in \Pi\}$. Note that different families w deliver different power properties of the integrated-based tests. However, a question that remains unsolved is which weight function is the optimal, in the sense of asymptotic power properties of the associated integrated test, for testing H_0. The results of Bierens and Ploberger (1997) show that there does not exist an optimal weight function uniformly over the whole space of alternatives $\mathcal{A}$.

In view of a sample $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$, the standardized sample version of $E[e_t(\theta_0) w(I_{t-1}, x)]$ is given by the residual marked empirical process

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

where θ_n is a $\sqrt{n}$-consistent estimator for θ_0. Because of (3), test statistics are based on a norm of $R^1_{n,w}$, say $\Gamma(R^1_{n,w})$. The most used norms are the CvM and KS functionals

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

and

$$KS_{n,w} := \sup_{x \in \Pi_c} \left| R^1_{n,w}(x) \right|,$$

respectively, where $\Psi(x)$ is an integrating function satisfying some mild conditions. The integrated-based tests reject the null hypothesis H_0 for "large" values of $\Gamma(R^1_{n,w})$.

Now, we discuss the asymptotic null and local distribution for the test based on $\Gamma(R^1_{n,w})$. To derive the asymptotic theory we consider the following assumptions. First, let us 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 $g(I_{t-1}, \theta_0) := (\partial / \partial \theta') f(I_{t-1}, \theta_0)$. Let $\mathcal{F}_t := \sigma(I'_t, I'_{t-1}, ..., I'_0)$ be the σ-field generated by the information set obtained up to time t. Let $F(\cdot)$ be the joint cumulative distribution function (cdf) of (Y_t, I'_{t-1}), and let $F_Y(\cdot)$ and $F_I(\cdot)$ be their marginal distributions, respectively. Let $\sigma^2(y)$ be the conditional error variance, i.e., $\sigma^2(y) := E[\varepsilon_t^2 \mid I_{t-1} = y]$. Given two points x_1 and x_2 on $\overline{\mathbb{R}}^q$, the bracket $[x_1, x_2]$ is the set of all points x with $x_1 \leq x \leq x_2$. An ε-bracket is a bracket $[x_1, x_2]$ with $d_w(x_1, x_2) < \varepsilon$. The bracketing number $N(\Pi_c, d_w, \varepsilon)$ is the minimum number of ε-brackets needed to cover Π_c.

Assumption A1:

A1(a): $\{(Y_t, I'_{t-1})' : t = 0, \pm 1, \pm 2, ...\}$ is a strictly stationary and ergodic process with $E|Y_1| < C$.

A1(b): $E[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ a.s. for all $t \geq 1$, $E\left|\varepsilon_1\right|^2 < C$ and $\sigma^2(y) \geq C > 0$ for all $y \in \overline{\mathbb{R}}^d$.

A1(c): $d_w(x_1, x_2)$ is continuous on $\Pi_c \times \Pi_c$, for any compact subset $\Pi_c \subset \Pi \subseteq \overline{\mathbb{R}}^q$.

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 $|g(I_{t-1,}\theta)| \leq M(I_{t-1})$, $\forall \theta \in \Theta$, such that $M(I_{t-1})$ is $F_I(\cdot)$-integrable.

Assumption A3: The parametric space Θ is compact in $\mathbb{R}^p$. The true parameter θ_0 belongs to the interior of Θ. The estimator θ_n satisfies the following asymptotic expansion under H_0

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

where $k(\cdot)$ is such that $L(\theta_0) = E[\varepsilon_1^2 k(I_0, \theta_0) k'(I_0, \theta_0)]$ exists and is positive definite. Whereas under $H_{A,n}(c)$

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

where $\xi_a = cE[a(I_0)k(I_0, \theta_0)]$.

Assumption A4:

A4(a): The weighting function $w(\cdot)$ is such that the equivalence in (3) holds. For any compact set Π_c of Π, $w(I_{t-1}, x)$ is uniformly bounded (a.s.) on Π_c, satisfies

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

and the uniform law of large numbers (ULLN)

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

holds whenever $\{(\xi_t, X_t')', \, t = 0, \pm 1, ...\}$ is a strictly stationary and ergodic process with $E\left|\xi_1\right| < C$.

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

Conditions A1 to A4 are considered in Escanciano (2004a) and are discussed in detail there. Note that Assumption A1 is very mild and allows for conditional higher moments of unknown form, such as conditional heteroskedasticity or time varying conditional kurtosis. In A3(b) we assume that the estimator θ_n satisfies a Bahadur linear representation under the null and under local alternatives. This condition is satisfied for a large class of estimators resulting from a martingale estimating equation, see Heyde (1997). In particular, it is satisfied under mild conditions by the nonlinear conditional least squares estimator (NLSE) with $k(I_{t-1}, \theta) = A^{-1}(\theta) g(I_{t-1}, \theta)$, where $A(\theta) = E[g(I_{0,}\theta) g'(I_{0,}\theta)]$, see Tjøstheim (1986). Assumption A4(a) restrict the "size" of the family

$\{w(\cdot,x) : x \in \Pi\}$. Escanciano (2004a) shows that A4(a) holds for all weight functions w considered in the literature. We are now in position to establish the asymptotic distribution of $R^1_{n,w}$ under the null and local alternatives. To this end, let us define the functions $G_w(x) \equiv G_w(x,\theta_0) := E[g(I_0,\theta_0)w(I_0,x)]$ and $\phi_w(s,x,\theta_0) \equiv \phi_w(s,x) := w(s,x) - G'_w(x,\theta_0)k(s,\theta_0)$, $x \in \Pi \subseteq \overline{\mathbb{R}}^q$, $s \in \mathbb{R}^d$.

THEOREM 1: (*Escanciano 2004a, Theorem 3) Under the null hypothesis H_0 and Assumptions A1-A3 and A4(a), uniformly in x on compacta*

$$R^1_{n,w}(x) = n^{-1/2}\sum_{t=1}^{n}\varepsilon_t\phi_w(I_{t-1},x,\theta_0) + o_P(1).$$

Furthermore

$$R^1_{n,w}(\cdot) \Longrightarrow R^1_{\infty,w}(\cdot),$$

where $R^1_{\infty,w}$ is a zero mean Gaussian process with covariance function

$$K_w(x_1,x_2) = E[\varepsilon_1^2\phi_w(I_0,x_1,\theta_0)\phi^c_w(I_0,x_2,\theta_0)].$$

THEOREM 2: *(Escanciano 2004a, Theorem 5) Under the local alternatives $H_{A,n}(c)$, Assumptions A1-A3 and A4(a)*

$$R^1_{n,w} \Longrightarrow R^1_{\infty,w} + cD_{w,a},$$

where $R^1_{\infty,w}$ is the process defined in Theorem 1 and $D_{w,a}(\cdot) = E[a(I_0)\phi_w(I_0,\cdot,\theta_0)]$.

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

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

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

Whereas under the assumptions of Theorem 2

$$\Gamma(R^1_{n,w}) \xrightarrow{d} \Gamma(R^1_{\infty,w} + cD_{w,a}).$$

To end this section, we shall find conditions that guarantee that the test based on $\Gamma(R^1_{w,n})$ is asymptotically unbiased. Let us define the asymptotic local power function (ALPF) as

$$\Pi_{w,\Gamma}(\alpha,c,a) := \underset{n\to\infty}{Lim} P\left(\Gamma(R^1_{n,w}) \geq c_{\Gamma,\alpha} \mid H_{A,n}(c)\right),$$

where $a \in \mathcal{A}$, $c \in \mathbb{R}$ and where $c_{\Gamma,\alpha}$ is such that $\underset{n\to\infty}{Lim} P\left(\Gamma(R^1_{n,w}) \geq c_{\Gamma,\alpha} \mid H_0\right) = \alpha$. Then, we find conditions under which $\Pi_{w,\Gamma}(\alpha, c, a) > \alpha$ holds for $c \neq 0$. Theorem 2 yields that for continuous functionals $\Gamma(\cdot)$

$$\Pi_{w,\Gamma}(\alpha, c, a) = \mathbb{P}_0(\Gamma(R^1_{\infty,w} + cD_{w,a}) > c_{\Gamma,\alpha}),$$

where $\mathbb{P}_0$ the probability measure associated to $R^1_{\infty,w}$ under the null hypothesis. If $\Gamma(\cdot)$ is an even functional, we have that $\Pi_{w,\Gamma}(\alpha, c, -a) = \Pi_{w,\Gamma}(\alpha, c, a)$. Anderson's Lemma, cf. Anderson (1955), yields that $\Pi_{w,\Gamma}(\alpha, c, a)$ is a nondecreasing function of $|c|$. Furthermore, it can be shown, see Theorem 6 below, that the derivative of $\Pi_{w,\Gamma}(\alpha, c, a)$ at $c = 0$ is zero and that the second derivative is strictly positive provided that $D_{w,a} \neq 0$. These arguments show that for $c \neq 0$, $\Pi_{w,\Gamma}(\alpha, c, a) > \alpha$ holds. Obviously, if $\Pi_{w,\Gamma}(\alpha, c, a) > \alpha$ then $D_a \neq 0$. Therefore, the test based on a continuous even functional Γ is unbiased if and only if $D_a \neq 0$ with positive measure. Note that the latter condition is true if and only if $a(I_{t-1}) \neq Cg(I_{t-1}, \theta_0)$ with positive probability. To our knowledge this result has not been established previously in the literature under such generality.

3. ASYMPTOTIC LOCAL POWER FUNCTION OF OMNIBUS TESTS

In this section we study in some detail the asymptotic local power properties of the integrated-based tests for testing H_0 against $H_{A,n}(c)$, that is, we study the ALPF $\Pi_{w,\Gamma}(\alpha, c, a)$ as a function of $a \in \mathcal{A}$ and $c \in \mathbb{R}$. In particular, we are interested in the analytical behavior of $\Pi_{w,\Gamma}(\alpha, c, a)$ for fixed α and a, as a function of c, and for fixed α and c, as a function of a. We shall start studying $\Pi_{w,\Gamma}(\alpha, c, a)$ as a function of a. In what follows, the subscript w in some quantities that depend on the weighting family chosen, as well as on Ψ, will be dropped whenever there is no confusion.

3.1 Asymptotic local power function as a function of the direction.

In this section we are interested in studying $\Pi_{w,\Gamma}(\alpha, c, a)$ as a function of the direction $a \in \mathcal{A}$. For simplicity we shall start with the CvM tests. That is, we are now concerned with the ALPF

$$\Pi_{w,\Psi}(\alpha, c, a) := \underset{n\to\infty}{Lim} P\left(CvM_{n,w} \geq c_\alpha \mid H_{A,n}(c)\right),$$

where c_α is such that $\underset{n\to\infty}{Lim} P\left(CvM_{n,w} \geq c_\alpha \mid H_0\right) = \alpha$. We need some further notation. Let $H_1 := L_2(\Pi, \Psi)$ be the Hilbert space of all Ψ-square integrable complex-valued functions on Π, furnished with the inner-product

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

and the induced norm $\|h\|_{H_1} = \langle h, h\rangle^{1/2}_{H_1}$. H_1 is endowed with the natural Borel σ-field induced by the norm $\|\cdot\|_{H_1}$; see, e.g., Chapter VI in Parthasarathy (1967) for random variables (r.v's) with values on Hilbert spaces. Similarly, we define $H_2 := L_2(\mathbb{R}^d, G)$, where $G(dy) := \sigma^2(y)F_I(dy)$, $\langle \cdot, \cdot \rangle_{H_2}$ and $\|\cdot\|_{H_2}$. Here, we restrict the directions to $a \in \mathcal{A} \cap H_2$. Note that $R^1_{n,w}$ can be viewed as a random element with values in H_1 instead of $\ell^\infty(\Pi)$. In fact, $CvM_{n,w} = \left\| R^1_{n,w} \right\|^2_{H_1}$.

As a mapping in H_1, $R^1_{\infty,w}$ 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

$$\begin{aligned} C_w(h)(x) &= E[\left\langle R^1_{\infty,w}, h \right\rangle_{H_1} R^1_{\infty,w}(x)] \\ &= E[\varepsilon_1^2 \left\langle \phi_w(I_0, \cdot), h \right\rangle_{H_1} \phi_w(I_0, x)] \qquad h \in H_1. \end{aligned}$$

Under our assumptions, the covariance operator C_w has the singular decomposition $C_w = L^*_w \circ L_w$, where $\circ$ stands for composition of operators, $L_w : H_1 \longrightarrow H_2$ is the compact linear operator given by

$$L_w h(s) = \left\langle \phi_w(s, \cdot), h \right\rangle_{H_1} \qquad s \in \mathbb{R}^d, h \in H_1$$

and $L^*_w : H_2 \longrightarrow H_1$ is defined by

$$L^*_w a(x) = \left\langle \phi_w(\cdot, x), a \right\rangle_{H_2} \qquad x \in \Pi, h \in H_1.$$

L^*_w is the adjoint (dual) operator of L_w and therefore, they satisfy

$$\langle a, L_w h \rangle_{H_2} = \langle L^*_w a, h \rangle_{H_1}.$$

The singular decomposition of C_w plays a crucial role in the power properties of $\Pi_{w,\Psi}(\alpha, c, a)$. Let H^0_1 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 $\{\lambda_{i,w}, \varphi_{i,w}\}_{i=1}^\infty$ is a complete sequence of eigenelements of it, i.e., $\{\lambda_{i,w}\}_{i=1}^\infty$ are real-valued and positive, and the corresponding eigenfunctions $\{\varphi_{i,w}\}_{i=1}^\infty$ form a complete orthonormal basis for H^1_1. Hence any H^1_1-valued random element has a Fourier expansion in terms of $\{\varphi_{i,w}\}_{i=1}^\infty$. In particular, we have the so-called Kac-Siegert representations (in distribution)

$$R^1_{n,w} = \sum_{i=1}^\infty \sqrt{\lambda_{i,w}} \epsilon_{n,i} \varphi_{i,w},$$

$$R^1_{\infty,w} = \sum_{i=1}^\infty \sqrt{\lambda_{i,w}} \epsilon_i \varphi_{i,w},$$

where $\epsilon_i := \lambda_{i,w}^{-1/2} \left\langle R^1_{\infty,w}, \varphi_{i,w} \right\rangle_{H_1}$ and $\epsilon_{n,i} := \lambda_{i,w}^{-1/2} \left\langle R^1_{n,w}, \varphi_{i,w} \right\rangle_{H_1}$. Note that by Theorem 1, $\{\epsilon_i\}_{i=1}^\infty$ are *iid*. $N(0,1)$ r.v's and $\{\epsilon_{ni}\}_{i=1}^\infty$ are, at least, uncorrelated with unit variance. Then, Parseval's identity yields

$$CvM_{\infty,w} = \sum_{i=1}^\infty \lambda_{i,w} \epsilon_i^2. \tag{4}$$

Therefore, the asymptotic null distribution of $CvM_{n,w}$ can be expressed as a weighted sum of independent χ^2_1 r.v's with weights depending on the data generating process (DGP). As we shall see, the principal components $\{\epsilon_i\}_{i=1}^{\infty}$ play a central role in the power properties of the CvM tests. Although the CvM tests are consistent against all alternatives in H_A, in practice they are 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, cf. Theorem 3 below. This can be immediately seen from (4), 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.

Again, by Parseval's identity we have that for each $h \in H_1$

$$P(\left\|R_{\infty,w}^1 + h\right\|_{H_1}^2 \geq c_\alpha) = P\left\{\sum_{i=1}^{\infty}\lambda_{i,w}(\epsilon_i + \rho_i)^2 + \rho_0^2 \geq c_\alpha\right\},$$

with $\rho_i = \lambda_{i,w}^{-1/2}\left\langle h, \varphi_{i,w}\right\rangle_{H_1}$, $i = 1, 2, ...$, and $\rho_0^2 = \|h\|_{H_1}^2 - \sum_{i=1}^{\infty}\rho_i^2$. Accordingly, Theorem 2 yields that

$$\Pi_{w,\Psi}(\alpha, c, a) = P(\left\|R_{\infty,w}^1 + cD_{w,a}\right\|_{H_1}^2 \geq c_\alpha).$$

First, note that for directions a such that $D_{w,a} = 0$ a.e. the power $\Pi_w(\alpha, c, a)$ is minimum, i.e., $\Pi_w(\alpha, c, a) = \alpha$. From (3), this is the case if and only if $a(I_{t-1}) = Cg(I_{t-1}, \theta_0)$ a.s. for some $C \in \mathbb{R}$.

By definition of L_w we have that $\{\psi_{i,w}\}_{i=1}^{\infty}$, defined by $\psi_{i,w} := \lambda_{i,w}^{-1/2} L_w\varphi_{i,w}$, forms a complete orthonormal system of $\overline{L}_w(H_1)$, the closure of the image of H_1 by L_w. Then, we are now in position to establish the first main result of the paper. We find the directions of maximum local power of the CvM tests. The analogous result for goodness-of-fit tests of distributions functions was proved in Neuhaus (1976, Theorem 2.2). Intuitively, given the orthonormality of $\{\psi_{i,w}\}_{i=1}^{\infty}$ and the equality $D_{w,a} = L_w^{*}\sigma^{-2}(\cdot)a(\cdot)$ we have that the direction with maximum local power is the maximizer of $\rho_1 = \lambda_{1,w}^{-1/2}\left\langle D_{w,a}, \varphi_{1,w}\right\rangle_{H_1} = \left\langle \sigma^{-2}(\cdot)a, \psi_{1,w}\right\rangle_{H_2}$, which is $\sigma^2(\cdot)\psi_{1,w}(\cdot)$.

THEOREM 3: *Assume A1 to A4. Then, the limiting power* $\Pi_{w,\Psi}(\alpha, c, a)$ *of the CvM test has for the vectors* $\psi_{i,w}$, $i \geq 1$, *and every* $c \geq 0$ *the properties*

$$\max\{\Pi_{w,\Psi}(\alpha, c, a) : a \in \overline{L}_w(H_1) \cap \mathcal{A}, \|a\|_{H_2} = 1\} = \Pi_{w,\Psi}(\alpha, c, \sigma^2(\cdot)\psi_{1,w}(\cdot)),$$

$$\Pi_{w,\Psi}(\alpha, c, \sigma^2(\cdot)\psi_{i,w}(\cdot)) \leq \Pi_{w,\Psi}(\alpha, c, \sigma^2(\cdot)\psi_{j,w}(\cdot)) \qquad \textit{for } 1 \leq j \leq i,$$

$$\lim_{i\to\infty}\Pi_{w,\Psi}(\alpha, c, \sigma^2(\cdot)\psi_{i,w}(\cdot)) = \alpha.$$

Furthermore, for each value $\beta \in (\alpha, \Pi_{w,\Psi}(\alpha, c, \sigma^2(\cdot)\psi_{1,w}(\cdot)))$, *there exists a direction* $a \in \overline{L}_w(H_1)$ *such that* $\beta = \Pi_{w,\Psi}(\alpha, c, a)$.

The theorem shows that there is one direction, namely $\sigma^2(\cdot)\psi_{1,w}(\cdot)$, with the highest asymptotic local power that is possible. In each other direction, the power is smaller, and for bad directions, the power is about α. Note that the best direction $\sigma^2(s)\psi_{1,w}(s) = \sigma^2(s)\lambda_1^{-1/2}\left\langle \phi_w(s, \cdot), \varphi_{i,w}\right\rangle_{H_1}$

depends on the family $\{w, \Psi\}$ and, in general, on the true model and DGP. In Section 6 we provide estimations for $\{\psi_{i,w}(\cdot)\}$, see (11).

EXAMPLE 1: Consider an *iid* sequence of r.v's $\{\varepsilon_t\}_{t=1}^n$ distributed as F_ε with $E\varepsilon_1^2 = 1$. Define $Y_t = \varepsilon_t$ and $I_{t-1} = Y_{t-1} = \varepsilon_{t-1}$. Then, $f(I_{t-1}) = 0$ a.s. Consider the model $f(I_{t-1}, \theta_0) = \theta_0$, with θ_0 unknown. It can be shown that the process

$$R_{n,ind}(t) = n^{-1/2}\sum_{t=1}^{n}(\varepsilon_t - \overline{\varepsilon}_n)1(F_\varepsilon(I_{t-1}) \leq t) \qquad t \in [0,1],$$

with $\overline{\varepsilon}_n = n^{-1}\sum_{i=1}^{n}\varepsilon_t$, converges weakly to a standard Brownian bridge $B(t)$ on $[0,1]$. The eigenelements of the covariance operator associated to B are, for $t \in [0,1]$ and $i \geq 1$,

$$\varphi_i(t) = \sqrt{2}\sin(i\pi t) \qquad \lambda_i = \frac{1}{i^2\pi^2}.$$

Then, the direction of maximum power of the CvM test based on $R_{n,ind}$ with the integrating function $d\Psi(t) = dt$ is given by $a^*(Y_{t-1}) = \psi_1(Y_{t-1}) = -\sqrt{2}\cos(F_\varepsilon(Y_{t-1})\pi/2)$. If $\varepsilon_t \sim U[0,1]$, then $a^*(Y_{t-1}) = -\sqrt{2}\cos(Y_{t-1}\pi/2)$ and the CvM test based on $R_{n,ind}$ is specialist in detecting low frequency alternatives, i.e., alternatives that do not oscillate very much.

We now return to the problem of studying $\Pi_{w,\Gamma}(\alpha, c, a)$ as a function of a for a general continuous functional Γ. A consequence of Theorem 3 above is that for CvM tests the power function $\Pi_{w,\Gamma}(\alpha, c, a)$ is flat on balls of alternatives except for alternatives coming from the finite-dimensional subspace generated by $\{\sigma^2(\cdot)\psi_{1,w}(\cdot), ..., \sigma^2(\cdot)\psi_{m,w}(\cdot)\}$ for a sufficiently large $m \in \mathbb{N}$. An extension of this result to a general functional Γ is proved in the following theorem which is based on Theorem 2.1 of Janssen (2000). Let $V^\perp \subset H_2$ denote the orthogonal complement of the linear subspace V of H_2.

THEOREM 4: *Assume A1 to A4. Let Γ be any continuous functional and $\alpha \in (0,1)$. For each $\varepsilon > 0$ and $K > 0$ there exists a linear subspace $V \subset H_2$ of finite dimension with*

$$\sup\{\Pi_{w,\Gamma}(\alpha, c, a) : a \in V^\perp, \|a\|_{H_2} \leq K\} \leq \varepsilon.$$

Morever the following upper bound:

$$\dim(V) - 1 \leq \varepsilon^{-1}\alpha(1-\alpha)\left(\exp(K^2) - 1\right)$$

holds for the dimension of V.

A consequence of Theorem 4 is that any integrated-based test has a preference for a finite-dimensional space of alternatives. For CvM tests this space is given by the space generated by $\{\sigma^2(\cdot)\psi_{1,w}(\cdot), ..., \sigma^2(\cdot)\psi_{m,w}(\cdot)\}$ for m large enough. For other functionals Γ this finite dimensional set is much more difficult to characterize, see Theorem 6 below for a possible candidate.

3.2 Asymptotic local power function as a function of the distance to the null.

We now study $\Pi_{w,\Gamma}(\alpha, c, a)$ for a fixed direction a and varying in c, $c \geq 0$. More specifically, we are interested in the analytical behavior of $\Pi_{w,\Gamma}(\alpha, c, a)$ for small and large c's. The first theorem considers the case $c \to \infty$. This result extends Theorem 4 of Bierens and Ploberger (1997) to a general functional Γ and the conditional heteroskedastic case. Note that this extension is far from being trivial. In fact, their proof depends crucially on the structure of the CvM test and the homoscedastic assumption. Theorem 4 of Bierens and Ploberger (1997) relies on the principal component decomposition of the CvM tests. For a general functional such a spectral representation is no longer available. However, we shall show in this section that a similar analysis is still possible using the likelihood ratio of the limit process $R^1_{w,\infty}$ under the null and under local alternatives (see 6). Henceforth we assume the normalization $E[a^2(I_{t-1})\sigma^{-2}(I_{t-1})] = 1$.

Theorem 5: *Assume A1 to A4. For any continuous functional* Γ, *for all* $\alpha \in (0,1)$ *and* $a \in \overline{L}_w(H_1) \cap \mathcal{A}$ *with* $E[a^2(I_{t-1})\sigma^{-2}(I_{t-1})] = 1$, *it holds that*

$$\lim_{c\to\infty} c^{-2} \ln(1 - \Pi_{w,\Gamma}(\alpha, c, a)) = -\frac{1}{2}.$$

Theorem 5 implies that if the test has nontrivial local power, that is, if $a(I_{t-1}) \neq Cg(I_{t-1}, \theta_0)$ with positive probability, then $\Pi_w(\alpha, c, a)$ approaches 1 at an exponential rate as $c \to \infty$. See Bierens and Ploberger (1997) for further implications of this result. For the case in which $c \to 0$ we have the next theorem.

Theorem 6: *Under the assumptions of Theorem 5*

$$\Pi_{w,,\Gamma}(\alpha, c, a) = \alpha + \frac{c^2}{2} A_{w,\Gamma}(\alpha, a) + o(c^2), \tag{5}$$

where

$$A_{w,\Gamma}(\alpha, a) = \sum_{i=1}^{\infty} \mu_i \left\langle a, \sigma^{-2}(\cdot) a_i \right\rangle^2_{H_2},$$

with $\{a_i\}_{i=1}^{\infty} \subset H_2$ *a suitable orthonormal system and the positive sequence* $\mu_i \downarrow 0$.

The coefficient of c^2 in (5) constitutes the curvature of the ALPF at the origin of the the integrated test based on $\Gamma(R^1_{n,w})$. Since in the case of an arbitrary (unknown) unconditional variance σ^2 the test is based on $R^1_{n,w}(x)/\sigma_n$ rather than $R^1_{n,w}(x)$, with σ_n a consistent estimate for σ, we have to replace c^2 in (5) by c^2/σ^2. This features the loss of power if the noise variance increases. In Section 6 we propose a bootstrap approximation for computing $A_{w,\Gamma}(\alpha, a)$ for general functionals Γ. Stute (1997) found a similar expansion to (5) for a CvM test for testing linear regressions under *iid* data. Thus, Theorem 6 extends Stute's (1997) expansion to a general continuous functional Γ and a time series framework. In Theorem 6 the sequence $\{\mu_i, a_i\}_{i=1}^{\infty}$ depends on the functional Γ used. In the CvM case this sequence is related to the spectral representation of the covariance operator C_w by the relations, $\forall i \geq 1$,

$$\mu_i = 1 - \alpha - \int \epsilon_i^2 1(\left\|R^1_{\infty,w}\right\|^2_{H_1} \leq c_\alpha) d\mathbb{P}_0$$

and

$$a_i = \psi_{i,w}(\cdot).$$

This allows us to conjecture that the role played by $\psi_{i,w}(\cdot)$ in Theorem 3 is played by a_i for a general functional Γ. In particular, the candidate for V in Theorem 4 for general functional Γ is of the form $\{\sigma^2(\cdot)a_1(\cdot), ..., \sigma^2(\cdot)a_m(\cdot)\}$ for large enough m. However, we are only able to prove formally such claim locally ($c \to 0$).

Theorems 5 and 6 provide two different methods for comparing two tests based on (w_1, Γ_1) and (w_2, Γ_2) in the direction a. The first method consists in comparing the level points of the tests. The level point of the test based on (w, Γ) is the smallest distance $|c|$ from the null hypothesis where the power $\beta \in (\alpha, 1)$ is attained, namely

$$lp_{w,\Gamma}(\beta, a) := \inf\{|c| : \Pi_{w,\Gamma}(\alpha, c, a) \geq \beta\}.$$

The quotient

$$\frac{lp_{w_1,\Gamma_1}(\beta, a)}{lp_{w_2,\Gamma_2}(\beta, a)}$$

provides a way to compare the two tests. The second method of comparison uses the expansion (5) of Theorem 6. The slopes $A_{w,\Gamma}(\alpha, a)$ can be used to define asymptotic relative efficiencies for comparing the asymptotic power behavior of different tests. We define the asymptotic local relative efficiency (ALRE) measure between the tests based on (w_1, Γ_1) and (w_2, Γ_2) as

$$ALRE(\alpha, a, (w_1, \Gamma_1), (w_2, \Gamma_2)) = \frac{A_{w_1,\Gamma_1}(\alpha, a)}{A_{w_2,\Gamma_2}(\alpha, a)}.$$

Both kind of measures have been proposed and used in the literature of goodness-of-fit tests for distributions functions, see, for instance, Neuhaus (1976) or Janssen (2000). The measure based on slopes is more operative than the measure based on level points due to its local character. Neuhaus (1976) and Milbrodt and Strasser (1990) provide numerical methods for computing $A_{w,\Gamma}(\alpha, a)$ for CvM and KS functionals, respectively. In this paper we extend the use of these measures to goodness-of-fit tests for time series regressions and we propose much simpler bootstrap approximations of $A_{w,\Gamma}(\alpha, a)$ for general continuous functionals in Section 6.

4. OPTIMAL DIRECTIONAL TESTS

In this section we shall construct a large class of asymptotically optimal directional tests for testing H_0 against $H_{A,n}(c : c \neq 0)$. All the tests we consider are continuous functionals of the RMP $R^1_{w,n}$.

First, we shall employ the principal components of $R^1_{n,w}$ in order to construct asymptotically optimal directional tests. To establish the asymptotic theory of these optimal inference procedures we need to consider estimation and consistency results of such principal components of the RMP $R^1_{n,w}$, which is postponed to Section 6. After this initial approach, we shall generalize the class of directional tests to a larger class of tests based on combinations of generalized orthogonal components.

We have seen that for directions a such that $D_{w,a} \neq 0$ in a subset of positive 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 $R^1_{\infty,w}$. Therefore, tests for H_0 against $H_{A,n}$ can be viewed as tests for $\widetilde{H}_0 : E\left[R^1_{\infty,w}\right] = 0$ against $\widetilde{H}_{A,n} : E\left[R^1_{\infty,w}(\cdot)\right] = cD_{w,a}(\cdot)$. In a fundamental work, Grenander (1952) generalized the optimal Neyman-Pearson theory to this framework. 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. As was previously commented, under the null $\epsilon_i \sim iid\ N(0,1)$, whereas under the local alternatives $\epsilon_i \sim iid\ N(c\delta_i, 1)$, with $\delta_i = \lambda_{i,w}^{-1/2} \left\langle D_{w,a}, \varphi_{i,w} \right\rangle_{H_1}$. The likelihood ratio of $(\epsilon_1, ..., \epsilon_m)$, $m \in \mathbb{N}$, under the null and under the local alternative is then

$$\frac{d\mathbb{P}^m_{1a}}{d\mathbb{P}^m_0} := \exp\left(c\sum_{i=1}^m \epsilon_1 \delta_i - \frac{1}{2}c^2 \sum_{i=1}^m \delta_i^2\right).$$

The condition $E[a^2(I_{t-1})\sigma^{-2}(I_{t-1})] = 1$ and Bessel's inequality imply that

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

Therefore, using the results of Grenander (1952, p. 215) we have that the last display ensures that the distribution of $R^1_{\infty,w}$ under the alternatives $H_{A,n}$, $\mathbb{P}_{1a}$ say, is absolutely continuous with respect to the distribution of $R^1_{\infty,w}$ under the null, $\mathbb{P}_0$, that is, the likelihood ratio $d\mathbb{P}^m_{1a}/d\mathbb{P}^m_0$ is well-defined as $m \to \infty$. The limit being

$$\frac{d\mathbb{P}_{1a}}{d\mathbb{P}_0}(h) = \exp\left(cZ_a(h) - \frac{1}{2}c^2\right) \qquad h \in H_1, \tag{6}$$

where $Z_a(h) = \sum_{i=1}^\infty \lambda_{i,w}^{-1} \left\langle h, \varphi_{i,w} \right\rangle_{H_1} \left\langle D_{w,a}, \varphi_{i,w} \right\rangle_{H_1}$. Thus, 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 $\{\left|Z_a(R^1_{n,w})\right| \geq z_{\alpha/2}\}$, 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 estimations $\{(\lambda_{n,i,w}, \varphi_{n,i,w}) : 1 \leq i \leq n\}$, which will be defined in Section 6. For a finite sample size n, the (approximated) Neyman-Pearson α-level test for H_0 against $H_{A,n}(c : c \neq 0)$ has critical region

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

where

$$\widehat{Z}_{a,m}(R^1_{n,w}) = \sum_{i=1}^{m} \frac{\widehat{\rho}_i \widehat{\epsilon}_{n,i}}{\widehat{\gamma}}, \tag{7}$$

$\widehat{\epsilon}_{n,i}$ and $\widehat{\rho}_i$ are estimations of the principal component ϵ_i and components of the shift ρ_i, respectively, given by

$$\widehat{\epsilon}_{n,i} = (\lambda_{n,i,w})^{-1/2} \left\langle R^1_{n,w}, \varphi_{n,i,w} \right\rangle_{H_1} \qquad 1 \le i \le m,$$

$$\widehat{\rho}_i = (\lambda_{n,i,w})^{-1/2} \left\langle \widehat{D}_{w,a}, \varphi_{n,i,w} \right\rangle_{H_1} \qquad 1 \le i \le m,$$

and the shift is estimated by

$$\widehat{D}_{w,a}(x) = \frac{1}{n} \sum_{t=1}^{n} a(I_{t-1}) \phi_w(I_{t-1}, x, \theta_n),$$

where $\phi_w(s, x, \theta_n) = w(s, x) - \widehat{G}'_w(x, \theta_n) k(s, \theta_n)$, $\widehat{G}_w(x, \theta_n) = n^{-1} \sum_{t=1}^{n} g(I_{t-1}, \theta_n) w(I_{t-1}, x)$, $\widehat{\gamma}^2 = \sum_{i=1}^{m} \widehat{\rho}_i^2$, and m is a user-chosen parameter, usually small because of the weights $\lambda_{i,w}$. The asymptotic theory for these optimal directional tests will be given in Section 6.

The asymptotic local power function of the optimal directional test for testing H_0 against $H_{A,n}(c : c \neq 0)$ is independent of the direction a, and it is given by

$$\Pi(\alpha, c) = 1 - \Phi(c + z_{\alpha/2}) + \Phi(c - z_{\alpha/2}),$$

where $\Phi(\cdot)$ is the cdf of a standard $N(0, 1)$. Simple algebra shows that the slope in the local representation (5) of the optimal directional test is

$$A_d(\alpha, a) = 2 z_{\alpha/2} \phi(z_{\alpha/2}),$$

where ϕ is the density of $\Phi(\cdot)$. Then, we can define the asymptotic local efficiency (ALE) of a test based on (w, Γ) as

$$ALE(\alpha, a, w, \Gamma) = \frac{A_{w,\Gamma}(\alpha, a)}{A_d(\alpha, a)},$$

which satisfies $0 \le ALE(\alpha, a, w, \Gamma) \le 1$ due to the optimality of the directional test.

In econometrics the simplest and well-known specification tests are those based on lack of correlation between the residuals and the regressors, see, for instance, the well-known Ramsey-RESET-type tests, cf. Ramsey (1969). These tests are based on rejecting the null hypothesis when

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\theta_n) w(I_{t-1}),$$

is large, for suitable transformations w of the information set. Tests based on correlations have power whenever $E\left[e_t(\theta_*) w(I_{t-1})\right] \neq 0$, where θ_* is the probabilistic limit of θ_n under the alternative H_A. Therefore, within this class of tests the optimal is the one that uses the transformation w^* solving (under the normalization $E[e_t^2(\theta_*)] = 1$) the optimization problem

$$\max_{w,\ E[w^2(I_{t-1})]=1} \left| E\left[e_t(\theta_*) w(I_{t-1})\right] \right|^2 .$$

Under our notation this is equivalent to

$$\max_{w,\ \|w\|_{H_2}=1} \left\langle \sigma^{-1}(\cdot)m(\cdot,\theta_*), w\right\rangle_{H_2}^2,$$

where $m(\cdot,\theta_*) := E[e_t(\theta_*) \mid I_{t-1}]$ is normalized such that $\left\|\sigma^{-1}(\cdot)m(\cdot,\theta_*)\right\|_{H_2} = 1$. The solution of the previous optimization problem is attained at $w^*(\cdot) = \sigma^{-1}(\cdot)m(\cdot,\theta_*)$. Therefore, among the tests based on correlations the optimal for testing H_0 against $H_{A,n}(c : c \neq 0)$ is that with critical region

$$\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\theta_n)a(I_{t-1})\sigma^{-2}(I_{t-1})\right| > c_\alpha,$$

for some suitable choice of c_α. Moreover, it can be shown that the latter test is an asymptotically optimal directional test, that is, it is the asymptotic uniformly most powerful test for testing H_0 against $H_{A,n}(c : c \neq 0)$. To prove this result note that, as expected, if $\sigma^{-2}a \in \overline{L}_w(H_1)$

$$\begin{aligned} Z_a(R^1_{n,w}) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\theta_0)\sum_{i=1}^{\infty}\lambda_{i,w}^{-1}\left\langle \phi_w(I_{t-1},\cdot),\varphi_{i,w}\right\rangle_{H_1}\left\langle L_w^*\sigma^{-2}a,\varphi_{i,w}\right\rangle_{H_1} + o_P(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\theta_0)\sum_{i=1}^{\infty}\psi_{i,w}\left\langle \sigma^{-2}a,\psi_{i,w}\right\rangle_{H_2} + o_P(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\theta_0)a(I_{t-1})\sigma^{-2}(I_{t-1}) + o_P(1). \end{aligned} \tag{8}$$

Now we generalize the previous directional tests to a larger class of optimal directional tests. Note that $Z_a(R^1_n)$ can be written as a linear combination of the principal components $\{\epsilon_{n,i}\}$ as

$$Z_a(R^1_{n,w}) = \sum_{i=1}^{\infty}\rho_i\epsilon_{n,i},$$

$$\epsilon_{n,i} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\theta_n)d_{i,w}(I_{t-1})$$

and

$$d_{i,w}(I_{t-1}) = \lambda_{i,w}^{-1/2}\int w(I_{t-1},x)\varphi^c_{i,w}(x)\Psi(dx).$$

In fact, we shall show that there exists an infinite number of optimal directional tests constructed in such a way. We can call these tests asymptotically optimal directional tests based on linear combinations of generalized orthogonal components. These tests are similar in spirit to those proposed in the goodness-of-fit tests literature for distributions by Schoenfeld (1977, 1980). Let $\{d_1, d_2, ...\}$ and orthonormal basis of $\overline{L}_w(H_1)$. Then, we define the generalized component as

$$\widehat{\delta}_i := \frac{1}{\sqrt{n}}\sum_{t=1}^{n} e_t(\theta_n)d_i(I_{t-1}),$$

and we consider the test statistic based on linear combinations of such components

$$\sum_{i=1}^{m} b_i\widehat{\delta}_i,$$

where $\{b_i\}$ is any sequence such that $\sum_{i=1}^{\infty} b_i^2 < \infty$. We have the following result.

THEOREM 7: *Let $m \to \infty$ as $n \to \infty$. Then, the following holds:*

(i) *Under the local alternatives $H_{A,n}(c : c \neq 0)$ the asymptotic distribution of $c\sum_{i=1}^{m} b_i\widehat{\delta}_i$ is that of a normal r.v with mean $\sum_{i=1}^{\infty} b_i a_i$ and variance $\sum_{i=1}^{\infty} b_i^2$, where*

$$a_i := \int a(x) d_i(x) F_I(dx).$$

(ii) *The test with critical region $\left|\sum_{i=1}^{m} a_i\widehat{\delta}_i\right| > c_\alpha$, where c_α is such that $\lim_{n \longrightarrow \infty} P(\left|\sum_{i=1}^{m} a_i\widehat{\delta}_i\right| \geq c_\alpha \mid H_0) = \alpha$, is the asymptotic uniformly most powerful test for testing H_0 against $H_{A,n}(c : c \neq 0)$.*

REMARK 1: Note that the test based on $Z_a(R_{n,w}^1)$ is a particular test based on linear combinations of generalized orthogonal components. More concretely, it is the one associated to the orthonormal basis $\{d_1, d_2, ...\} = \{\psi_1, \psi_2, ...\}$.

5. OPTIMAL SMOOTH TESTS

Omnibus tests are designed for cases in which the practitioner does not know the alternative at hand and she/he needs of consistent tests. However, sometimes she/he is interested in some particular alternatives and in tests that direct their power against such desired alternatives. For instance, in a dynamic regression model the econometrician might not be worried about a misspecification as long as some (of all) parameters in the regression are identified. Smooth tests represent a compromise between directional and omnibus tests. Optimal smooth tests are specially design to detect, in an optimal way, a finite number of specific alternatives. On the other hand, we have seen before that each omnibus test has a preference for a finite-dimensional space of alternatives. Apart from this space, the power function is almost flat. Such a preferred space is usually unknown to the practitioner. However, we have shown in previous sections how the practitioner can analyze the omnibus test to get some knowledge about its preferences (see Section 3). Because the omnibus tests are in fact concentrated on the preferred space, it seems natural to consider optimal tests against such finite-dimensional space instead of applying the omnibus test. The latter fact provides further motivation for the use of smooth tests in econometrics. In this section we consider the general problem of how to construct an optimal test when a finite-set of alternatives are in mind, that is, how to construct optimal smooth tests. We call the latter tests *optimal smooth tests* because they are in spirit similar to Neyman's (1937) smooth test for densities. Like with the directional tests, we are interested in constructing a large class of such optimal smooth tests as functionals of the RMP

$R^1_{n,w}$. Consider the functionals of the form

$$\|z\|^2_{N,m} = \sum_{i=1}^{m} \left(\int_{\Pi} h_i(x) z^c(x) \Psi(dx) \right)^2 ,$$

where $m \in \mathbb{N}$, and $\{h_1, ..., h_m\} \subset H_1$. We shall discuss in detail the asymptotic local power function of the tests based on $\left\|R^1_{n,w}\right\|^2_{N,m}$. Simple algebra shows that

$$\left\|R^1_{n,w}\right\|^2_{N,m} = \sum_{i=1}^{m} \left(\int_{\Pi} h_i(x) \left(R^1_{n,w}(x)\right)^c \Psi(dx) \right)^2 = \sum_{i=1}^{m} \widehat{\delta}^2_{i,w},$$

where now

$$\widehat{\delta}_{i,w} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\theta_n) \int_{\Pi} h_i(x) w^c(I_{t-1}, x) \Psi(dx) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\theta_n) d_i(I_{t-1}), \tag{9}$$

and where the d_i's are implicitly defined. Tests based on $\left\|R^1_{n,w}\right\|^2_{N,m}$ are called smooth tests. In the case that $\{d_1, d_2, ..., d_m\}$ is an orthonormal system in $\overline{L}_w(H_1)$ we call the test based on $\left\|R^1_{n,w}\right\|^2_{N,m}$ an optimal smooth test. Smooth tests are a compromise between the omnibus tests and directional tests because they have only power against those alternatives for which $E[e_t(\theta_*) d_i(I_{t-1})] \neq 0$ for some i, $1 \leq i \leq m$, where θ_* is the probabilistic limit of θ_n under the alternative.

It is easy to show that the asymptotic null distribution of $\left\|R^1_{n,w}\right\|^2_{N,m}$ is, in general, that of a weighted sum of dependent χ^2_1 r.v's. To obtain a χ^2_m asymptotic null distribution is sufficient to choose $\{d_1, d_2, ..., d_m\}$ as an orthonormal system in $\overline{L}_w(H_1)$. That is, an optimal smooth tests has χ^2_m asymptotic null distribution. Many specification tests considered in the econometric literature are smooth tests. The best well-known examples are the specification tests proposed by Ramsey (1969), Hausman (1978) or the overidentified restrictions tests of the generalized method of moments (GMM) literature, see Hansen (1982). Other important example is the Pormanteau-type tests, cf. Box y Pierce (1970). The test for overidentified restrictions based on $\{d_1, d_2, ..., d_m\}$ with $m > p$ rejects the null hypothesis for large values of

$$\widehat{\delta}'_w C_{n,w} \widehat{\delta}_w$$

where $\widehat{\delta}_w = (\widehat{\delta}_{1,w}, ..., \widehat{\delta}_{m,w})'$, θ_n is the GMM estimator and $C_{n,w}$ is a suitably chosen positive definite weighting matrix. Under the orthonormality assumption of $\{d_1, d_2, ..., d_m\}$ it turns out that the optimal choice of $C_{n,w}$ is the identity matrix and in that case $\widehat{\delta}'_w C_{n,w} \widehat{\delta}_w$ is an optimal smooth test. For the optimality of smooth tests in the context of goodness-of-fit tests for distributions functions see Neyman (1937).

One of the possible choices of $\{h_1, ..., h_m\}$ is $\{\lambda^{-1/2}_{1,ind} \varphi_{1,ind}, ..., \lambda^{-1/2}_{m,ind} \varphi_{m,ind}\}$ which corresponds with Stute's (1997) smooth version of the CvM test for testing the correct specification of a linear regression with *iid* observations. It is worth to note that the terminology for smooth tests that we are using here is much more general than that considered in Stute (1997) in the sense that we are concerned with general finite sets of alternatives, not only those that correspond with the preferred space of the CvM test. In addition, even for the choice $\{h_1, ..., h_m\} = \{\lambda_{1,ind}^{-1/2}\varphi_{1,ind}, ..., \lambda_{m,ind}^{-1/2}\varphi_{m,ind}\}$ our construction of the smooth test differs from that of Stute (1997) because we consider different nonparametric estimations of the directions $\{\lambda_{1,ind}^{-1/2}\varphi_{1,ind}, ..., \lambda_{m,ind}^{-1/2}\varphi_{m,ind}\}$, see Section 6 for details. Note also that our smooth versions of the CvM tests are valid for general dynamic regressions under possibly conditional heteroskedasticity of unknown form and general RMP $R_{n,w}^1$. The limit distribution of $\left\|R_{n,w}^1\right\|_{N,m}^2$ for the latter choice is

$$\left\|R_{\infty,w}^1\right\|_{N,m}^2 = \sum_{i=1}^{m}\epsilon_i^2.$$

Therefore, we observe that, contrary to the CvM tests, these smooth tests may be able to detect "high-frequency" alternatives that are heavily downweighted by $\lambda_{i,w}$ in (4), see in Section 7 the empirical application to the Canadian Dollar exchange rate for a revealing example. Smooth versions of other functionals are possible by using as $\{d_1, ..., d_m\}$ the functions $\{a_1, ..., a_m\}$ of Theorem 6.

We know apply previous theory to the Neyman's functional. It is easy to prove that if $\{d_1, d_2, ..., d_m\}$ are orthonormal in $\overline{L}_w(H_1)$, the asymptotic local power function of test based on $\left\|R_{n,w}^1\right\|_{N,m}^2$ is

$$\Pi_N(\alpha, c, a) = 1 - \chi_{m,des}^2(c_\alpha^m),$$

where c_α^m is the α-quantile of a χ_m^2 distribution and *des* is the noncentrality parameter of the χ_m^2 given by

$$des = \sum_{j=1}^{m}\left\langle h_j, D_{w,a}\right\rangle_{H_1}^2.$$

On the other hand, analogously to the proof of Theorem 6 it can be shown that as $c \to 0$

$$\Pi_N(\alpha, c, a) = \alpha + \lambda^{(m)}\frac{des}{2}c^2 + o(c^2),$$

where the coefficients $\lambda^{(m)}$ are $\lambda^{(m)} = \chi_m^2(c_\alpha^m) - \chi_{m+2}^2(c_\alpha^m)$, see Lemma 2.7 in Milbrodt and Strasser (1990).

The name "optimal" for optimal smooth tests is partly justified by the following arguments. Assume the class of alternative models

$$E[e_t(\theta_0) \mid I_{t-1}] = \delta_1 d_1(I_{t-1}) + \delta_2 d_2(I_{t-1}) + \cdots + \delta_m d_m(I_{t-1}), \tag{10}$$

where $\{d_1, d_2, ..., d_m\}$ is an orthonormal system in $\overline{L}_w(H_1)$. Suppose that we would like to test

$$H_0^m : \delta_1 = \delta_2 = \cdots = \delta_m = 0$$

against

$$H_1^m : \delta_j \neq 0 \text{ for some } j = 1, ..., m.$$

Then, the Lagrange Multiplier (LM) test constructed with a quasi-likelihood function (that is, assuming that $e_t(\theta_0) \mid I_{t-1} \sim N(0, \sigma^2(I_{t-1}))$) is asymptotically equivalent to the optimal smooth test $\left\| R_{n,w}^1 \right\|_{N,m}^2$. Moreover, each component $\widehat{\delta}_{i,w}^2$ is a LM test for testing $H_{0,h}^m : \delta_h = 0$ against $H_{1,h}^m : \delta_h \neq 0$, $h = 1, ..., m$. Because $\{\widehat{\delta}_{i,w}\}_{i=1}^m$ are asymptotically independent, this shows that each $\widehat{\delta}_{i,w}^2$ is a detector for δ_i and no other. Then, the optimality of the smooth tests follows from the well-known optimality properties of LM tests.

Two remarks at this point are important. First, in the general case in which the practitioner is interested in a particular set $\{a_1, ..., a_m\} \subset H_2$ of alternatives, the natural candidate for $\{d_1, ..., d_m\}$ is the Gram-Schmidt orthonormalization of $\{\widetilde{a}_1, ..., \widetilde{a}_m\}$, where $\{\widetilde{a}_1, ..., \widetilde{a}_m\}$ is the orthogonal projection of $\{a_1, ..., a_m\}$ onto $\overline{L}_w(H_1)$. And secondly, an attractive feature of optimal smooth tests is that when H_1^m is rejected, $E[Y_t \mid I_{t-1}] = f(I_{t-1}, \theta_0) + \widehat{\delta}_{1,w} d_{1,w}(I_{t-1}) + \widehat{\delta}_{2,w} d_{2,w}(I_{t-1}) + \cdots + \widehat{\delta}_{m,w} d_{m,w}(I_{t-1})$ provides an alternative model for the conditional mean $f(I_{t-1})$. In this sense, smooth tests are more informative than omnibus tests when the null hypothesis is rejected, see our application to exchange rates in Section 7.

An important problem is the choice of the parameter m in $\left\| R_{n,w}^1 \right\|_{N,m}^2$. A large body of literature in the goodness-of-fit testing for distributions has considered data-driven Neyman's smooth tests, i.e., has considered the case in which m is chosen from the data, see, e.g., Eubank and LaRiccia (1992), Ledwina (1994), Fan (1996), Inglot and Ledwina (1996) or Kallenberg and Ledwina (1997), among others. Some of these works allow for $m \to \infty$ as $n \to \infty$. Similar ideas can be considered in our framework but this is beyond the scope of this paper. Interestingly enough, in the case of $m \to \infty$ as $n \to \infty$ the optimal smooth test is equivalent to a L_2-distance test based on a series expansion estimator of $E[e_t(\theta_0) \mid I_{t-1}]$ using the basis $\{d_1, d_2, ...\}$. Optimal smooth tests become local tests in the latter case, cf. Hong and White (1995).

6. ESTIMATION OF THE PRINCIPAL COMPONENTS

In this section we are concerned with the estimation of the eigenelements $\{(\lambda_{i,w}, \varphi_{i,w}) : i = 1, 2, ...\}$ of C_w. These estimators are important in order to estimate the directions of maximum power, to estimate the ALRE measures and to develop directional and optimal smooth tests related to the CvM tests. Note that the empirical counterpart of C_w under the null hypothesis is given by

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

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 $\phi_w(I_{t-1},\cdot)$, $t=1,...,n$). 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 $\varphi_{n,i,w}$, $1\leq i\leq n$, be an eigenvalue and eigenfunction of $C_{n,w}$, respectively. The eigenfunction $\varphi_{n,i,w}$ necessarily has the form $n^{-1}\sum_{t=1}^{n}\beta_{i,t}\phi_w(I_{t-1},\cdot)$, for some coefficients $\beta_{i,t}$, $t=1,...,n$, and the equation to solve becomes

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

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

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

with $a_{ts}=\int\limits_{\Pi}e_t^2(\theta_n)\phi_w(I_{t-1},x)\phi_w^c(I_{s-1},x)\Psi(dx)$. The solutions $\beta_i=(\beta_{i,1},...,\beta_{i,n})'$ 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 us 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

$$\left\|\varphi_{n,i,w}-\widetilde{\varphi}_{i,w}\right\|_{H_1}\leq c_i\left\|C_{n,w}-C_w\right\|,\qquad i\geq 1,$$

where c_i is a real number that depends only on $\lambda_{i,w}$ and $\widetilde{\varphi}_{i,w}=sgn\left(\langle\varphi_{n,i,w},\varphi_{i,w}\rangle_{H_1}\right)\varphi_{i,w}$ (sgn is the sign function, i.e., $sgn(x)=1(x>0)-1(x<0)$). The last inequalities and Theorem 8 imply the consistency of the estimated eigenelements. Given the consistency of $\{(\lambda_{n,i,w},\varphi_{n,i,w}):1\leq i\leq n\}$ it is not difficult to show the following corollary. For the optimal smooth tests based on $\{h_1,...,h_m\}=\{\lambda_{1,w}^{-1/2}\varphi_{1,w},...,\lambda_{m,w}^{-1/2}\varphi_{m,w}\}$ the situation is analogous, and then, it is omitted.

Corollary 2: *(Asymptotic theory of directional tests) Under the assumptions of Theorem 1 for a fixed $m\in\mathbb{N}$*

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

Whereas under the assumptions of Theorem 2

$$\widehat{Z}_{a,m}(R^1_{n,w}) \longrightarrow N(\mu,1),$$

where

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

With the estimation of the principal components we can approximate the slope of the ALPF of CvM tests at the origin, i.e., $A_{w,\Gamma}(\alpha,a)$ in (5), by

$$A^*_{m,w}(\alpha,a) = \sum_{i=1}^{m} \lambda_{n,i,w}^{-1} \rho_{n,i,w}^2 \zeta_i^*,$$

where $\rho_{n,i,w} = \lambda_{n,i,w}^{-1/2} \left\langle \widehat{D}_{w,a}, \varphi_{n,i,w} \right\rangle_{H_1}$ and ζ_i^* is a bootstrap approximation of $\zeta_i := 1-\alpha-\int \epsilon_i^2 1(\left\|R^1_{\infty,w}\right\|_{H_1}^2 \leq c_\alpha) d\mathbb{P}_0$ given by

$$\zeta_i^* = 1-\alpha-\frac{1}{B}\sum_{j=1}^{B}\left(\widehat{\epsilon}^j_{n,i}\right)^2 1(\left\|R^{1,j}_{\infty,w}\right\|_{H_1}^2 \leq c^*_\alpha),$$

where $\widehat{\epsilon}^j_{n,i} = \left(\lambda^j_{n,i,w}\right)^{-1/2} \left\langle R^{1,j}_{n,w}, \varphi^j_{n,i,w} \right\rangle_{H_1}$, $R^{1,j}_{n,w}$ for $j=1,...,B$ are fixed design wild bootstrap realizations of $R^1_{n,w}$ and c^*_α is the bootstrap critical value, see Escanciano (2004a) for details on this bootstrap approximation. Alternatively ζ_i can be calculated numerically using expansions of distributions of quadratic forms in Gaussian variables into series of central χ^2-distribution functions or Laguerre polynomials, see Jonhson and Kotz (1970), Chapter VI for details.

For other functionals the situation is much more involved. In the proof of our Theorem 6 we show that $\{\mu_i, a_i\}_{i=1}^{\infty}$ appearing in the slope $A_{w,\Gamma}(\alpha,a)$ in (5) are the eigenelements of a Hilbert-Schmidt operator T_Γ from $\overline{L}_w(H_1)$ to $\overline{L}_w(H_1)$ given by

$$T_\Gamma a = \sum_{i=1}^{\infty} \widetilde{\mu}_i \left\langle a, \psi_{i,w} \right\rangle_{H_2} \psi_{i,w},$$

where

$$\widetilde{\mu}_i := \left\langle \psi_{i,w}(\cdot), T_\Gamma \psi_{i,w}(\cdot) \right\rangle_{H_2} = 2^{-1/2}\left[\int \left(Z^2_{\psi_{i,w}}(R^1_{\infty,w})-1\right)\left(1(\Gamma(R^1_{\infty,w}) \leq c_\alpha)-\alpha\right) d\mathbb{P}_0\right].$$

Under conditional homoscedasticity, $Z^2_{\psi_{i,w}}(R^1_{\infty,w})$ reduce to ϵ_i^2. In the latter case a simple bootstrap approximation of $\widetilde{\mu}_i$ is given by

$$\widetilde{\mu}_i^* = 2^{-1/2}\left(\frac{1}{B}\sum_{j=1}^{B}\left\{\left(\widehat{\epsilon}^j_{n,i}\right)^2-1\right\}\left\{1(\Gamma(R^{1,j}_{\infty,w}) \leq c^*_{\alpha,\Gamma})-\alpha\right\}\right).$$

A simple approximation of T_Γ is then

$$T_{m,\Gamma} a(\cdot) = \sum_{i=1}^{m} \widetilde{\mu}_i^* \left(\frac{1}{n}\sum_{t=1}^{n} a(I_{t-1})\psi_{n,i,w}(I_{t-1})\right)\psi_{n,i,w}(\cdot),$$

where $\psi_{n,i,w}(\cdot)$ is an estimator of $\psi_{i,w}(\cdot)$ given by

$$\psi_{n,i,w} = \lambda_{n,i,w}^{-1/2} \int_{\Pi} \phi_w(I_{t-1}, x) \varphi_{n,i,w}^c(x) \Psi(dx). \tag{11}$$

The estimator $T_{m,\Gamma}$ of T_Γ is new in the literature. T_Γ plays a crucial role in the asymptotic local power properties of general continuous functionals Γ. Note that as the operator $C_{n,w}$, $T_{m,\Gamma}$ has a finite dimensional closed range (that is spanned by the functions $\psi_{n,i,w}(\cdot)$, $i = 1, ..., m$). Therefore, the number of eigenvalues and eigenfunctions of $T_{m,\Gamma}$ is finite and bounded by m, and they can be computed by solving a linear system. For the general heteroskedastic case the situation is the same but with a more involve bootstrap approximation for $\widetilde{\mu}_i$. We do not discuss this further for the sake of space.

7. MONTE CARLO SIMULATIONS

In this section we put some of the previous theory into practice. We compare in terms of local power the omnibus tests with the optimal smooth and directional tests against some alternatives. We also study how performs in finite samples the estimation of the principal components and the estimation of the directions of maximum asymptotic local power for omnibus tests.

7.1 Empirical local power properties of tests.

We briefly describe our simulation setup. Let $I_{t-1,P} = (Y_{t-1}, ..., Y_{t-P})$ the information set at time $t-1$. We denote by $PCvM_{n,P}$ the Cramér-von Mises test based on $1(\beta' I_{t-1,P} \leq u)$. 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

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

and

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

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

Bierens (1982) proposed to use $w(I_{t-1}, x) = \exp(i I_{t-1}' x)$ as the weighting function in (3) and considered the Cramér-von Mises test statistic

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

where

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

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

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

Escanciano (2004a) considers a multivariate bootstrap version of the RMP

$$R^1_{n,ind,P}(x) = \frac{1}{\widehat{\sigma}_e\sqrt{n}}\sum_{t=1}^{n} e_t(\theta_n)1(I_{t-1} \leq x),$$

used in Koul and Stute (1999) for $P = 1$. His CvM and KS tests statistics are given by

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

and

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

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

We consider the optimal directional tests given in (7) and based on

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

and the optimal smooth tests

$$\widehat{S}_{a,m}(R^1_{n,w}) = \sum_{i=1}^{m}\widehat{\epsilon}^2_{n,i},$$

for $R^1_{n,pro,P}$, $R^1_{n,\exp,P}$ and $R^1_{n,ind,P}$, that is, for $w(I_{t-1},x) = 1(\beta' I_{t-1} \leq u)$, $w(I_{t-1},x) = \exp(ix'I_{t-1})$, $w(I_{t-1},x) = 1(I_{t-1} \leq x)$, respectively.

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. For the wild bootstrap approximation of omnibus tests we employ a sequence $\{V_t\}$ of *iid* Bernouilli variates given by $P(V_t = 0.5(1-\sqrt{5})) = (1+\sqrt{5})/2\sqrt{5}$ and $P(V_t = 0.5(1+\sqrt{5})) = 1-(1+\sqrt{5})/2\sqrt{5}$, see Escanciano (2004a) for details of the bootstrap approximation.

We consider the null model of no-effect, or as it is known in the econometric literature, the tests for the martingale difference hypothesis (MDH). The MDH is central in many areas of economics and finance, see, e.g., the market efficiency hypothesis or asset pricing theory. In the sequel ε_t is a sequence of *iid.* $N(0,1)$. The null model is that of a martingale difference sequence

$$E[Y_t \mid Y_{t-1}, Y_{t-2}, ...] = 0 \ a.s.$$

We examine the adequacy of this model under the following data generating processes (DGP):

1. A strong white noise model: $Y_t = \varepsilon_t$.

2. An autoregressive of order one local alternative model (LAR(1)): $Y_{t,n} = n^{-1/2} Y_{t-1} + \varepsilon_t$.

3. A nonlinear autoregressive local alternative model (LSIN): $Y_{t,n} = n^{-1/2} 3\sin(0.7\pi Y_{t-2}) + \varepsilon_t$.

4. An autoregressive of order two local alternative model (LAR(2)): $Y_{t,n} = n^{-1/2}(0.6Y_{t-1} - 0.9Y_{t-2}) + \varepsilon_t$.

For models 1 and 2 we consider $P = 1$, whereas for models 3 and 4 we take $P = 2$. The sample sizes considered are $n = 50$, 100 and 200. For the smooth and directional tests we choose $m = 3$. The critical values for smooth and directional tests against models 2 to 4 are size-corrected and are based on 5000 replications of model 1. We report the rejection probabilities (RP) for these models and the test statistics $PCvM_{n,P}$, $CvM_{n,\exp,P}$, $CvM_{n,ind,P}$, $KS_{n,ind,P}$, $\widehat{Z}_{a,m}(R^1_{n,pro,P})$, $\widehat{Z}_{a,m}(R^1_{n,\exp,P})$, $\widehat{Z}_{a,m}(R^1_{n,ind,P})$, $\widehat{S}_{a,m}(R^1_{n,pro,P})$, $\widehat{S}_{a,m}(R^1_{n,\exp,P})$ and $\widehat{S}_{a,m}(R^1_{n,ind,P})$ in Table 1 and Table 2.

Table 1 shows that the empirical size properties of tests are good even for as small sample sizes as $n = 50$. Only $\widehat{S}_{a,3}(R^1_{n,\exp,1})$ presents some underrejection for model 1. For model 2, we observe that the directional tests outperform the rest of the tests, as expected, and that smooth tests present a better empirical local power against this alternative than omnibus tests. One reason that may explain the last fact is that the weights of the CvM tests undervalue the contribution of the second and third component (e.g., for the indicator case the true weights are $0.10132, .02533, 0.01125$) whereas the smooth tests take into account such contributions. Within each class of tests (omnibus, smooth and directional) all tests perform similarly, that is, there is not much difference between indicator and exponential based tests.

Please insert Table 1 about here

In Table 2 we present the RP for models 3 and 4. The directional tests present the best empirical local properties against these two local alternatives. For model 3 the exponential-based omnibus test is the best among the omnibus tests. The smooth tests are, in general, comparable to omnibus tests for this alternative. For model 4, the test based on projections $R^1_{n,pro,P}$ is the best among omnibus tests, and again the smooth tests are comparable to omnibus tests. Summarizing, for these alternatives we see that, as expected, the optimal directional tests are superior in terms of empirical local power, and, in general, the smooth tests present similar empirical local power properties to the omnibus tests. In some cases, the smooth tests are better than omnibus tests, cf. model 2.

Please insert Table 2 about here

7.2 Estimation of the eigenfunctions and directions of maximum power.

In this subsection we shall investigate the properties of the proposed estimators for the eigenelements $\{(\lambda_{i,w}, \varphi_{i,w}) : i = 1, 2, ...\}$ of C_w. In this section we make use of the fact that we know the true eigenelements of the covariance operator of the null limit process of

$$R_{n,ind}(t) = n^{-1/2}\sum_{t=1}^{n}(\varepsilon_t - \overline{\varepsilon}_n)1(\varepsilon_{t-1} \leq x) \qquad x \in \mathbb{R},$$

with $\overline{\varepsilon}_n = n^{-1}\sum_{i=1}^{n}\varepsilon_t$. The eigenelements are given by

$$\varphi_i(x) = \sqrt{2}\sin\left(i\pi F_\varepsilon(x)\right) \qquad \lambda_i = \frac{1}{i^2\pi^2} \qquad i \geq 1, \tag{12}$$

see Example 1. Then, we can compare the estimated eigenelements with the true ones. To that end, we define the integrated mean square error (IMSE) for the estimator $\varphi_{n,i,ind} = n^{-1}\sum_{t=1}^{n}\beta_{i,t}\phi_{ind}(I_{t-1}, \cdot)$ proposed in Section 6

$$IMSE_{n,i} := \int_{\mathbb{R}} MSE_{n,i}(x)F_\varepsilon(x),$$

where $MSE_{n,i}(x)$ is the mean square error (MSE) for $\varphi_{n,i,ind}(x)$ as an estimator of $\varphi_i(x)$ for a fixed $x \in \mathbb{R}$, that is, $MSE_{n,i}(x) = E[(\varphi_{n,i,w}(x) - \varphi_i(x))^2]$. $IMSE_{n,i}$ is approximated using the empirical distribution function of $\{\varepsilon_t\}_{t=1}^{n}$, say $F_n(x)$, i.e.,

$$IMSE_{n,i} = \int_{\Pi} MSE_{n,i}(x)F_n(x) = n^{-1}\sum_{t=1}^{n} MSE_{n,i}(\varepsilon_t).$$

We have made a simple experiment with 1000 simulations of model 1 in previous section. We have computed the $IMSE_{n,i}$ for the estimators $\{\varphi_{n,i,ind} : i = 1, 2, ...\}$ proposed in Section 6 as estimators of φ_i in (12). We show the $IMSE_{n,i}$ in Table 3 for samples sizes $n = 50, 100$ and 200 and $i = 1, 2, 3$ and 4. We also plot the estimated eigenfunctions and the true eigenfunctions in Figure 1. From these results we conclude that the proposed estimators perform quite well in finite samples, even for small sample sizes as $n = 50$. The $IMSE_{n,i}$ for $\varphi_{n,i,ind}$ decrease with the sample size, as expected, and also increase in i. That is, high components have larger MSE than low components. The results for the eigenvalues are similar, and then, they are not reported.

Please insert Table 3 and Figure 1 about here.

With the estimators of the eigenfunctions $\varphi_{n,i,w}$ we can construct estimations of the directions of maximum power of the CvM test based on w and Ψ as $\psi_{n,1,w}(I_{t-1})\sigma_n^2(I_{t-1})$, where $\sigma_n^2(y)$ is a consistent nonparametric estimator of $\sigma^2(y)$, for instance, a Nadaraya-Watson estimator, and

$$\begin{aligned}\psi_{n,1,w} &= \lambda_{n,1,w}^{-1/2}\int_{\Pi}\phi_w(I_{t-1},x)\varphi_{n,1,w}^c\Psi(dx)\\ &= \lambda_{n,1,w}^{-1/2}n^{-1}\sum_{s=1}^{n}\beta_{1,s}\int_{\Pi}\phi_w(I_{t-1},x)\phi_w^c(I_{s-1},x)\Psi(dx).\end{aligned}$$

We plot the first and second directions of maximum local power for $CVM_{n,\exp,P}$ and $CVM_{n,ind,P}$ for the previous examples assuming that $\sigma^2(y)=\sigma^2=1$, so the directions are given by $\psi_{n,1,w}$ and $\psi_{n,2,w}$, respectively.

Please insert Figure 2 about here.

8. TESTING THE MDH OF EXCHANGE RATES

In this section we examine the martingale properties of some exchange rates returns studied previously by Fong and Ouliaris (1995) or Escanciano and Velasco (2003), among others. Also recently, Hong and Lee (2003) have studied the MDH properties of a related data set. The main objective of this section is to show the ability of our new proposed smooth tests to find the information that provides the omnibus (CvM) tests when the null hypothesis is rejected or accepted. The data set consists in four 760 weekly exchange rate returns on the Canadian Dollar (Can), the German Deutschmark (Dm), the French Franc (Fr) and the Japanese Yen (¥), from August 14, 1974 to March 29, 1989. The empirical results are reported in Tables 4 and 5.

Please insert Tables 4 and 5 about here.

We use the same implementation as in the Monte Carlo experiments and we show the empirical p-values. For $P=1$, all tests reject the MDH at 10% for all data sets and at 5% for Dm, Fr and ¥. For $P=3$ the conclusions are similar. The results for the Canadian dollar are contradictory with the statistics CvM_P and KS_P. Note that for the Canadian dollar the smooth test $\widehat{S}_{a,3}(R_{n,ind,3}^1)$ is able to reject the MDH. A detailed study of this case may explain the contradictory results for CvM_P with this exchange rate. In Table 6 we have presented the first three components of the RMP $R_{n,ind,3}^1$ for all data sets. We observe that for the Can data the first and second components are small whereas the third component is very large. This may explain the inability of the CvM_P to detect such alternative. On the other hand, the smooth test based on $\widehat{S}_{a,3}(R_{n,ind,3}^1)$ is able to detect such component and rejects the MDH for this data set. This kind of behavior of indicator-based tests have been found in other simulations studies, specially under nonlinear alternatives, see Escanciano (2004a,b), and may be explained by similar reasons. Therefore, this application shows that our new smooth tests are able to find nonlinear dependence in the conditional mean of these exchange rates, in agreement with some previous studies (see Escanciano and Velasco, 2003, and Hong and Lee, 2003). The nonlinearity in the conditional mean suggest that additional effort has to be dedicated to investigate the form of such nonlinearity before modelling the conditional variance. Unlike omnibus tests, smooth tests are able to provide an alternative model in the case of rejection. As an application of this principle, we consider in Figure 3 the alternative model provided by the smooth tests $\widehat{S}_{a,3}(R^1_{n,ind,1})$ for $P=1$ and for all data sets. The alternative model is given by $E[Y_t \mid Y_{t-1}] = \widehat{\delta}_{1,ind}\psi_{n,1,ind}(Y_{t-1}) + \widehat{\delta}_{2,ind}\psi_{n,2,ind}(Y_{t-1}) + \widehat{\delta}_{3,ind}\psi_{n,3,ind}(Y_{t-1})$, see Section 5 for details. We observe the there are two different patterns in the predicted models, one for the Can data and the other for the remaining exchange rates. These differences are due to the different frequency character of these alternatives of the MDH.

Please insert Table 6 and Figure 3 about here.

To conclude, throughout this simulations and the empirical application we have shown that smooth tests provide useful alternative modelling specification tests to omnibus tests because of their good properties. We have shown that the smooth versions of the omnibus tests present similar empirical local power that omnibus tests, with the additional properties of being robust to higher-frequency alternatives (do not downweight moderate components), being asymptotically distribution-free, avoiding resampling methods that are computationally intensive, and providing alternative models to the one specified when the null hypothesis is rejected. Therefore, due to these properties and others shown in the paper we conclude that smooth tests can play a valuable role in time series modeling.

9. PROOFS

Proof of Theorem 3: Define $\rho_i = \langle a, \psi_{i,w} \rangle_{H_2}$ and consider the directions

$$b_j = \left(\sum_{i=1}^{j} \rho_i^2\right)^{1/2} \psi_{1,w} + \sum_{i=j+1}^{\infty} \rho_i \psi_{i,w}, \qquad j \geq 1.$$

Note that $a = b_1$. We first show that

$$\Pi(\alpha, c, \sigma^2 b_1) \leq \Pi(\alpha, c, \sigma^2 b_2). \tag{13}$$

To this end, write

$$\Pi(\alpha, c, \sigma^2 b_1) = P\left\{\lambda_1(\epsilon_1 + \rho_1)^2 + \lambda_2(\epsilon_2 + \rho_2)^2 \geq c_\alpha - \sum_{i=3}^{\infty} \lambda_i(\epsilon_i + \rho_i)^2\right\},$$

$$\Pi(\alpha, c, \sigma^2 b_2) = P\left\{\lambda_1(\epsilon_1 + \rho)^2 \geq c_\alpha - \sum_{i=3}^{\infty} \lambda_i(\epsilon_i + \rho_i)^2\right\},$$

where $\rho = (\rho_1^2 + \rho_2^2)^{1/2}$. If we denote by $\rho_1 = \rho\cos\phi$ y $\rho_1 = \rho\sin\phi$, we have by the symmetry of ϵ_1 and ϵ_2 that

$$\Pi(\alpha, c, b_1\sigma^2) = P\left\{\lambda_1(\epsilon_1 - \rho\cos\phi)^2 + \lambda_2(\epsilon_2 - \rho\sin\phi)^2 \geq c_\alpha - \sum_{i=3}^{\infty}\lambda_i(\epsilon_i + \rho_i)^2\right\}.$$

Applying Proposition 2.1 of Neuhaus (1976) we obtain (13). Similarly, it can be shown

$$\Pi(\alpha, c, \sigma^2 b_j) \leq \Pi(\alpha, c, \sigma^2 b_{j+1}) \qquad \forall j \geq 2.$$

Using that $\|b_j - \psi_1\|_{H_1}^2 \to 0$ as $j \to \infty$, we obtain the first statement of the theorem. Also, for $h > j$

$$\Pi(\alpha, c, \sigma^2\psi_j) = P\left\{\lambda_j(\epsilon_j + \rho_j)^2 + \lambda_h\epsilon_h^2 \geq c_\alpha - \sum_{i=1, i\neq j,h}^{\infty}\lambda_i\epsilon_i^2\right\},$$

$$\Pi(\alpha, c, \sigma^2\psi_h) = P\left\{\lambda_h(\epsilon_h + \rho_h)^2 + \lambda_j\epsilon_j^2 \geq c_\alpha - \sum_{i=1, i\neq j,h}^{\infty}\lambda_i\epsilon_i^2\right\}.$$

Thus, applying the same argument as before and Proposition 2.1 of Neuhaus (1976) we obtain the second statement of the Theorem. The last two statements follow exactly as in Theorem 2.2 of Neuhaus (1976) and then, they are omitted. ■

Proof of Theorem 4: The proof follows exactly the same lines as the proof of Theorem 2.1 in Janssen (2000), and then, it is omitted. ■

Proof of Theorem 5: From (6) we have

$$1 - \Pi_\Gamma(\alpha, c, a) = \int\limits_{\{\Gamma(R^1_{\infty,w}) \leq c_\alpha\}} \exp\left(cZ_a(R^1_{\infty,w}) - \frac{1}{2}c^2\right) d\mathbb{P}_0.$$

Denote $B = \{z \in H_1 : \Gamma(z) \leq c_\alpha\}$. Let us define $\widetilde{B} = Z_a(B)$ and denote by $\widetilde{B}_*$ and $\widetilde{B}^*$ its minimal and maximal measurable majorant, see pg. 7 in VW. Therefore,

$$\begin{aligned} 1 - \Pi_\Gamma(\alpha, c, a) &\geq \int\limits_{\widetilde{B}_*} \exp\left(cz - \frac{1}{2}c^2\right)\frac{1}{\sqrt{2\pi}}\exp(-0.5z^2)dz \\ &= \frac{1}{\sqrt{2\pi}}\int\limits_{\widetilde{B}_*}\exp(-0.5\,(z-c)^2)dz. \end{aligned}$$

By Jensen's inequality

$$\lim_{c\longrightarrow\infty} c^{-2}\ln(1 - \Pi_\Gamma(\alpha, c, a)) \geq -\frac{1}{2}.$$

As for the other inequality, use that $|Z_a(h)| \leq C\,\|h\|_{H_1}^2$, and therefore, by the continuity of $\Gamma(\cdot)$ we have that $\widetilde{B}^*$ is contained in a symmetric compact interval of $\mathbb{R}$, $[-M, M]$, $M > 0$, and for $M > 0$ and $c - M > 0$ or $c + M < 0$ it holds that $\phi(|c| + M) \leq \Phi(c + M) - \Phi(c - M) \leq \phi(|c| - M)$, where Φ and ϕ are the cdf and the density of a standard normal r.v, respectively. Conclude taking logarithms, dividing by c^2 and taking limits. ■

Proof of Theorem 6: Let us define the symmetric bilinear form

$$B_\Gamma : (h_1, h_2) \longrightarrow \int\limits_{\{\Gamma(R^1_{\infty,w})>c_\alpha\}} Z_{h_1}(R^1_{\infty,w}) Z_{h_2}(R^1_{\infty,w}) d\mathbb{P}_0 - \alpha \langle h_1, h_2 \rangle_{H_1} \qquad h_1, h_2 \in H_1.$$

From a Taylor expansion it is clear that $A_{w,\Gamma}(\alpha, h) = B_\Gamma(h, h)$, $h \in H_2$. B_Γ is continuous and positive semidefinite. Thus, there is a bounded, symmetric and positive semidefinite operator T_Γ from H_2 to H_2 such that

$$A_{w,\Gamma}(\alpha, h) = \langle T_\Gamma h, h \rangle_{H_2}.$$

Moreover, by Lemma 4.1 in Milbrodt y Strasser (1990) T_Γ is compact. See Theorem 2.1 in Jansen (1995) for an alternative proof. As a consequence there is a spectral representation

$$T_\Gamma = \sum_{i=1}^{\infty} \mu_i \langle \cdot, a_i \rangle^2_{H_2},$$

where $\{a_i\}_{i=1}^\infty \subset H_2$ is a orthonormal system and the sequence $\mu_i \downarrow 0$. The theorem follows from the last display. ■

Proof of Theorem 7: First, we prove (i). Fix $m \in \mathbb{N}$. Since $\{d_1, d_2, ...\}$ is an orthonormal basis of H_2^1 (orthogonal to $g(I_{t-1}, \theta_0)$) we have that $\widehat{\delta}_i = \widetilde{\delta}_i + o_P(1)$, $i = 1, ..., m$, where

$$\widetilde{\delta}_i := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\theta_0) d_i(I_{t-1}).$$

The Central Limit Theorem for stationary and ergodic martingales difference sequences of Billingsley (1961) yields that the vector $\widetilde{\delta}^m = (\widetilde{\delta}_1, ..., \widetilde{\delta}_m)'$ converges under the local alternatives to a multivatiate normal random vector with mean vector $a^m = (\widetilde{a}_1, ..., \widetilde{a}_m)'$ and identity variance-covariance matrix. Using Theorem 4.2 of Billingsley (1968) and that $\sum_{i=1}^\infty b_i^2 < \infty$ part (i) is proved. As for part (ii), since $m \to \infty$ as $n \to \infty$ and

$$\sum_{i=1}^{m} a_i \widehat{\delta}_i := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t(\theta_0) g(I_{t-1}) + o_P(1),$$

with

$$g(I_{t-1}) := \sum_{i=1}^{\infty} a_i d_i(I_{t-1}) = \sigma^{-2}(I_{t-1}) a(I_{t-1}),$$

and where the last equality is because $\{d_1, d_2, ...\}$ is an orthonormal basis of H_2^1 and $\sigma^{-2} a \in H_2^1$. Then, from (8)

$$\sum_{i=1}^{m} a_i \widehat{\delta}_i = Z_a(R_n) + o_P(1).$$

Therefore, part (ii) follows from the optimality of $Z_a(R_n)$. ■

Proof of Theorem 8: The result follows from the inequality

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

where

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

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

REFERENCES

Anderson, T.W. and Darling, D.A. (1952) Asymptotic theory of certain "Goodness of fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193-212.

Anderson, T.W. (1955) The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematicl Society* **6**, 170-176.

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 funcional form. *Ecomometrica* **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.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley, New York.

Bickel, P.J. and Ritov, Y. (1992) Testing for goodness of t: a new approach. In Nonparametric Statistics and Related Topics, Ed. A.K.Md.E. Saleh, pp.51-7. North-Holland, New York.

Box, G. and Pierce, D. (1970) Distribution of residual autocorrelations in autorregressive integrated moving average time series models. *Journal of American Statistical Association* **65**, 1509-1527.

de jong, R.M. (1996) The Bierens' tets under data dependence. *Journal of Econometrics* **72**, 1-32.

Domínguez, M. and Lobato, I.N. (2003) A consistent test for the martingale difference hypothesis. *Econometric Reviews* **22**, 351-377.

Durbin, J. and Knott, M. (1972) Components of the Cramer-von Mises statistics I. *Journal of the Royal Statistical Society Series B* **34**, 290-307.

Escanciano, J.C. (2004a) Model Checks Using Residual Marked Empirical Processes. Working Paper 13/04 Facultad de Económicas y Empresariales, Universidad de Navarra.

Escanciano, J.C. (2004b) Goodness-of-fit tests for linear and non-linear time series models. Second revision in *Journal of American Statistical Association.*

Escanciano, J.C. and Velasco, C. (2003) Generalized spectral tests for the martingale difference hypothesis. Forthcoming in *Journal of Econometrics*.

Eubank, R. and Hart, J. (1992) Testing goodness-of-fit in regression via order selection criteria. *Annals of Statistics* **20,** 1412-1425.

Eubank, R.L. and LaRiccia, V.M. (1992) Asymptotic comparison of Cramer-von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Annals of Statistics* **20**, 2071-2086.

Fan, J. (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association* **91**, 674-688.

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. and Li, Q. (2000) Consistent model specification tests: kernel-based tests versus Bierens'

ICM tests. *Econometric Theory* **16**, 1016-1041.

Fong, W. M. and Ouliaris, S. (1995) Spectral tests of the martingale hypothesis for exchange rates. *Journal of Applied Econometrics* **10**, 255-271.

Grenander, U. (1952) Stochastic processes and statistical inference. *Arkiv for Matematik* **1**, 195-277.

Hájek, J. and Šidák, Z. (1967) *Theory of rank tests.* Academic Press, New York.

Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regresion fits. *Annals of Statistics* **21**, 1926-1974.

Hart, J.D. (1997) *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer Verlag, New-York.

Hansen, L. (1982) Large sample properties of generalised methods of moments estimators. *Econometrica* **50**, 1029-1054.

Hausman, J.A. (1978) Specification Tests in Econometrics. *Econometrica* **46**, 1251-1272.

Heyde, C.C. (1997) *Quasi-Likelihood and Its Applications.* Springer, Berlin.

Hong, Y. and Lee, T.H. (2003) Inference on predictability of foreign exchange rate changes via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics* **85**, 1048-1062.

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.

Inglot, T. and Ledwina, T. (1996) Asymptotic optimality of data-driven Neyman's tests for uniformity. *Annals of Statistics* **24**, 1982-2019.

Janssen, A. (1995) Principal component decomposition of non-parametric tests. *Probability Theory and Related Fields* **101**, 193-209.

Janssen, A. (2000) Global power functions of goodness of fit tests. *Annals of Statistics* **28**, 239-253.

Jonhson, N.L. and Kotz, S. (1970) *Continuous Univariate Distributions.* Vol. 2. Houghton Miffin, Boston, MA.

Kallenberg, W.C.M. and Ledwina, T. (1997) Data-Driven smooth tests when the hypothesis is composite. *Journal of the American Statistical Association* **92**, 1094-1104.

Koul, H.L. and Stute W. (1999) Nonparametric model checks for time series. *Annals of Statistics* **27**, 204-236.

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.

Neuhaus, G., (1976) Asymptotic power properties of the Cramér-von Mises test under contiguous alternatives. *Journal of Multivariate Analysis* **6**, 95-110.

Neyman, J. (1937) Smooth test for goodness of fit. *Scandinavian Aktuarietidskr.* **20**, 149-199.

Parthasarathy, K.R. (1967) *Probability Measures on Metric Spaces.* New York. Academic Press.

Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series* **5**, 50:157–175.

Ramsey, J.B. (1969) Test for specification errors in classical linear least squares regression analysis. *Journal of the Royal Statistical Society Series B* **31**, 350-371.

Rayner, J. C. W., and Best, D. J. (1989) *Smooth Tests of Goodness of Fit.* London: Oxford University Press.

Schoenfeld, D. A. (1977) Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramer-von Mises statistic. *Annals of Statistics* **5**, 1017-1026.

Schoenfeld, D. A. (1980) Tests based on linear combinations of the orthogonal components of the Cramer-von Mises statistic when parameters are estimated. *Annals of Statistics* **8**, 1017-1022.

Stinchcombe, M., and White, H. (1998) Consistent specification testing with nuisance parameters present only ùnder the alternative. *Econometric Theory* **14**, 295-325.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics* **25**, 613-641.

Stute, W., Thies, S., and Zhu, L.X. (1998) Model checks for regression: an innovation process approach. *Annals of Statistics* **26**, 1916-1934.

Tjøstheim, D. (1986) Estimation in nonlinear time series models. *Stochastic Processes and Their Applications* **21**, 251-273.

Wooldridge, J. (1992) A test for functional form against nonparametric alternatives. *Econometric Theory* **8,** 452-475.

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.

van der Vaart, A.W., and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes.* Springer. New York.

Table 1. Local Power of 5% Tests for the MDH.

P = 1, m=3		**WN**			**LAR(1)**	
n	**50**	**100**	**200**	**50**	**100**	**200**
$CvM_{n,ind,1}$	5.2	4.9	4.3	5.2	4.9	4.3
$CvM_{n,\exp,1}$	5.0	4.6	4.5	5.0	4.6	4.5
KS_1	5.6	5.6	5.3	5.6	5.6	5.2
$\widehat{Z}_{a,3}(R^1_{n,ind,1})$	4.2	4.2	5.1	23.7	25.3	26.8
$\widehat{Z}_{a,3}(R^1_{n,\exp,1})$	5.4	4.3	4.9	21.6	23.2	23.7
$\widehat{S}_{a,3}(R^1_{n,ind,1})$	3.0	5.0	4.5	9.9	8.2	10.0
$\widehat{S}_{a,3}(R^1_{n,\exp,1})$	1.6	3.4	3.4	7.4	10.8	10.9

Table 2. Local Power of 5% Tests for the MDH.

P = 2, m=3		**LSIN**			**LAR(2)**	
n	**50**	**100**	**200**	**50**	**100**	**200**
$PCVM_{n,2}$	11.0	12.2	10.9	12.8	12.6	13.8
$CvM_{n,ind,2}$	9.4	10.9	10.8	7.8	5.9	7.3
$CvM_{n,\exp,2}$	16.3	18.3	18.0	10.8	10.5	12.7
KS_2	11.5	14.3	15.2	9.8	8.7	8.3
$\widehat{Z}_{a,3}(R^1_{n,pro,2})$	26.1	24.4	23.5	15.8	18.5	16.3
$\widehat{Z}_{a,3}(R^1_{n,ind,2})$	33.7	29.9	25.7	16.0	15.4	14.3
$\widehat{Z}_{a,3}(R^1_{n,\exp,2})$	27.5	29.0	31.9	17.3	21.0	25.8
$\widehat{S}_{a,3}(R^1_{n,pro,2})$	10.4	13.1	10.3	12.1	11.3	11.6
$\widehat{S}_{a,3}(R^1_{n,ind,2})$	12.1	14.5	12.2	5.8	7.7	6.9
$\widehat{S}_{a,3}(R^1_{n,\exp,2})$	13.2	13.1	11.2	10.2	10.5	8.7

Table 3. IMSE for model 1.

$IMSE_{n,i}$	**n=50**	**n=100**	**n=200**
i=1	0.06604	0.03419	0.01638
i=2	0.23881	0.13022	0.06433
i=3	0.75030	0.27731	0.13991
i=4	1.74966	0.78480	0.26303

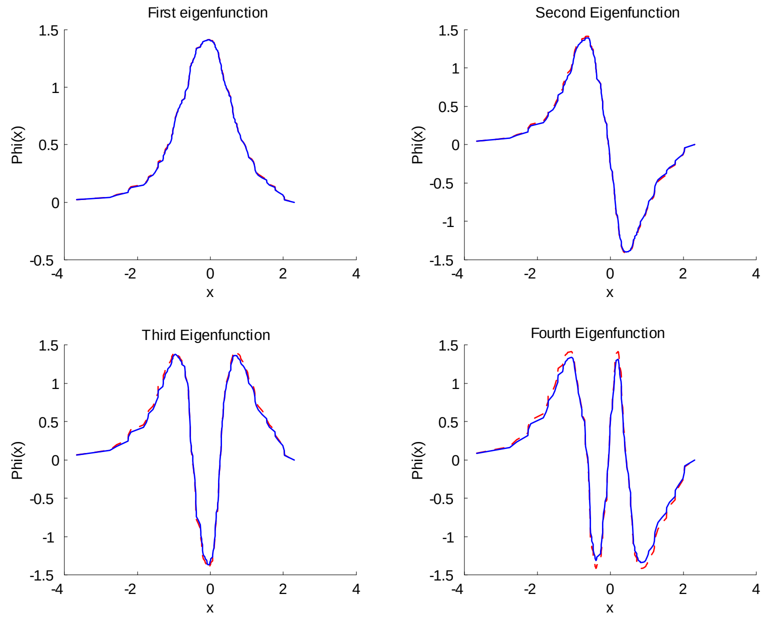

Fig1. Estimated eigenfunctions (dash line) and true eigenfunctions (solid line) for model 1.

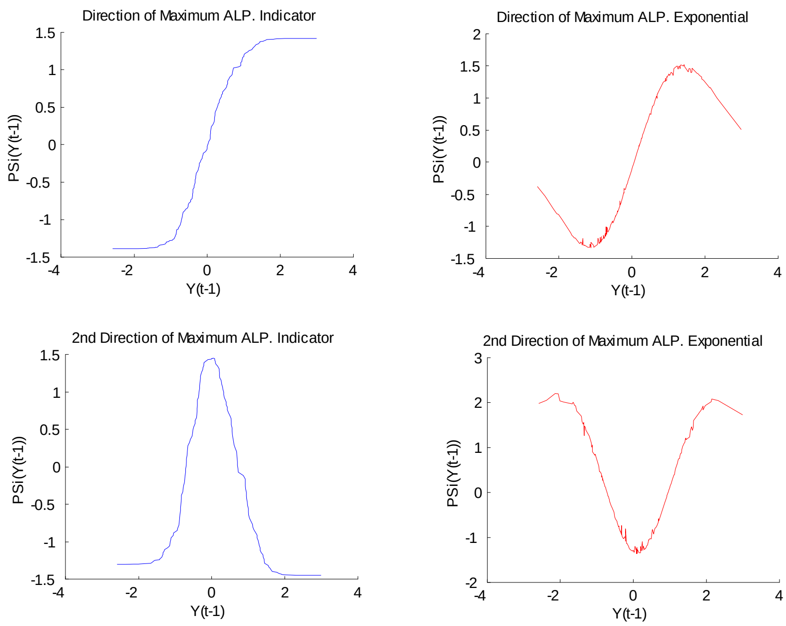

Fig.2. Estimated directions of Maximum ALPF for indicator and exponential functions.

Table 4. P-values for the Exchange Rates Returns.

$n=760, P=1$	**Can**	**Dm**	**Fr**	¥
$CvM_{n,ind,1}$	0.016	0.000	0.010	0.003
$CvM_{n,\exp,1}$	0.040	0.003	0.023	0.000
KS_1	0.053	0.000	0.013	0.006
$\widehat{S}_{a,3}(R^1_{n,ind,1})$	0.088	0.010	0.040	0.007
$\widehat{S}_{a,3}(R^1_{n,\exp,1})$	0.099	0.005	0.011	0.002

Table 5. P-values for the Exchange Rates Returns.

$n=760, P=3$	**Can**	**Dm**	**Fr**	¥
$PCVM_{n,3}$	0.013	0.000	0.010	0.000
$CvM_{n,ind,3}$	0.143	0.003	0.063	0.000
$CvM_{n,\exp,3}$	0.006	0.000	0.000	0.003
KS_3	0.390	0.000	0.060	0.000
$\widehat{S}_{a,3}(R^1_{n,pro,3})$	0.024	0.010	0.096	0.004
$\widehat{S}_{a,3}(R^1_{n,ind,3})$	0.014	0.015	0.075	0.003
$\widehat{S}_{a,3}(R^1_{n,\exp,3})$	0.003	0.000	0.000	0.000

Table 6. Individual components

$n=760, P=3$	**Can**	**Dm**	**Fr**	¥
$\widehat{\delta}^2_{1,ind}$	1.4813	10.065**	6.0308*	10.2919**
$\widehat{\delta}^2_{2,ind}$	1.3471	0.3422	0.8651	2.5410
$\widehat{\delta}^2_{3,ind}$	7.704**	0.0002	0.0000	1.0942

Note: *Significative at 5%. **Significative at 1%.

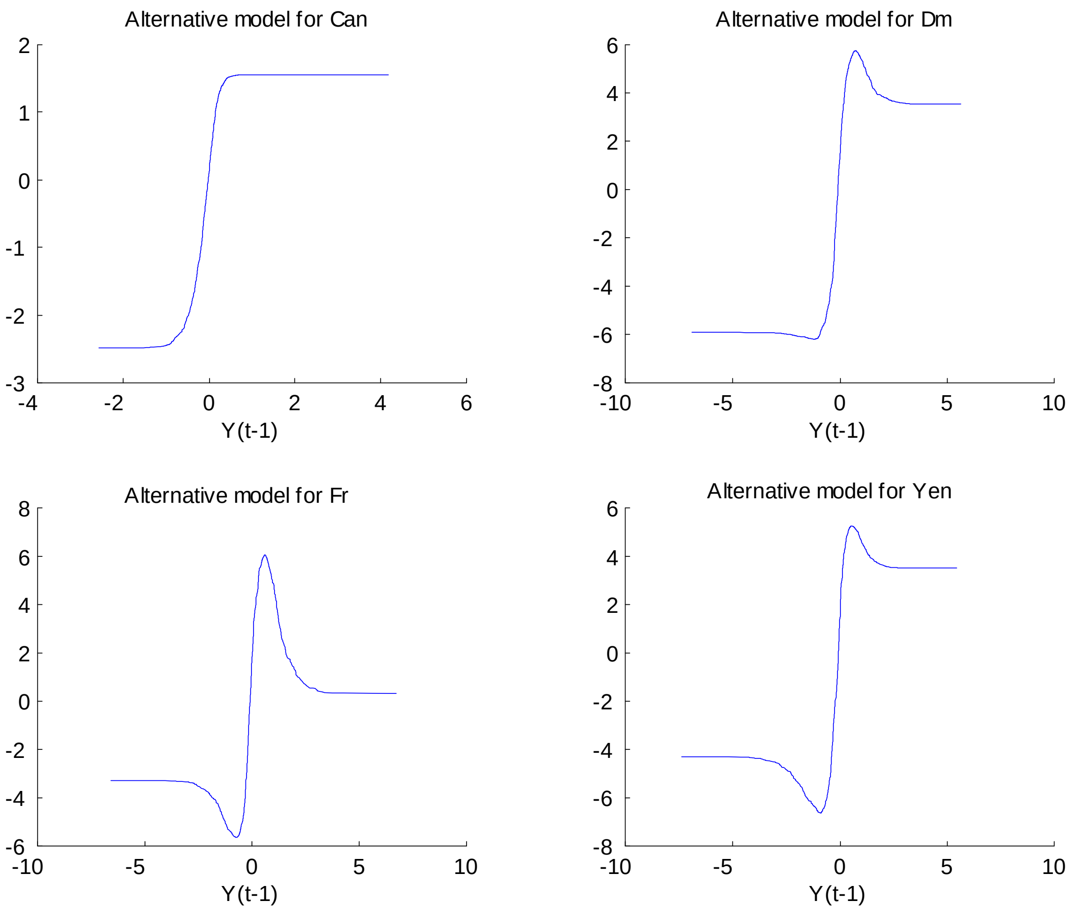

Fig.3. Alternative models for exchange rates.