

Testing for neglected nonlinearity in regression models: A collection of new tests based on the theory of random fields^a

Christian M. Dahl
Department of Economics
Purdue University

Gloria González-Rivera
Department of Economics
University of California, Riverside

Abstract

Within the framework of the flexible regression model approach proposed by Hamilton (2000), we demonstrate the use of the theory of random fields as a building block for a battery of new tests for neglected nonlinearity. The common denominator to all the tests is that they are TR^2 versions of the Lagrange multiplier statistic. The proposed tests are free of unidentified nuisance parameters, contrary to Hamilton's and the popular Neural Network tests, where the number of unidentified nuisance parameters under the null hypothesis increases with the dimensionality of the model. Furthermore, the proposed tests are robust to the specification of the variance-covariance function of the random field, and their computation is simple compared to the statistic proposed by Hamilton, since they are based on auxiliary regressions. These advantages are the result of 1) studying alternative representations of the random field governing the outcome of the conditional mean in Hamilton's regression model, and 2) using local approximations to the variance-covariance function based on Taylor's expansion. The main modification to the random field consists of switching from the Euclidean distance measure to the Minkowski distance measure. A Monte Carlo simulation study suggests that our statistic based on the Minkowski distance broadly matches the power properties of the statistic suggested by Hamilton. The tests match the simplicity of the

^aThe paper has benefited from numerous discussions and comments from Peter Storgaard. We would also like to thank Niels Haldrup, Svend Hylleberg, Alvaro Escribano, Esben Høg, Peter Jensen and Allan Würtz and seminar participants at the University of Aarhus, University of California, Riverside, and the Econometric Society Meetings in Boston. The second author acknowledges the financial support of the Academic Senate of the University of California, Riverside.

popular Tsay's statistic and have superior power properties compared to those of the Tsay's test. This suggests that treating the unobserved nonlinear function as random rather than deterministic is a fruitful approach for the identification of nonlinear components in parametric regression models.

JEL classification: C12; C15; C22

Keywords: Flexible regression models; Random fields; Neglected nonlinearity

1 Introduction

In a recent paper Hamilton (2000) proposes a parametric approach to estimating nonlinear relations that shares the flexibility of nonparametric models. He studies a model of the form $y_t = f(x_t) + \varepsilon_t$ where the functional form $f(x_t)$ is unknown and is assumed to be the outcome of a Gaussian random field with a simple moving average representation. As a by-product, Hamilton derives a new test for nonlinearity based on the Lagrange multiplier principle. The proposed statistic depends on a set of nuisance parameters that are only identified under the alternative hypothesis. On computing the statistic, the researcher has to have some priors on the magnitude of the unidentified parameters. In small samples, dealing with the unidentified nuisance parameter problem by assuming full knowledge of the parameterized stochastic process that determines the random field, may have adverse effects on the power properties of the test, e.g. Hansen (1996b). Nevertheless, simulation studies, based on low dimensional linear and nonlinear models, indicate that the Hamilton's test appears to be very powerful in comparison to existing tests such as Ramsey's (1969) RESET test, Tsay's (1986) test, and the Neural Network test proposed by White (1989). However, in Hamilton's setting as well as in the Neural Network test, increasing the dimensionality of the model amounts to an increase in the number of unidentified parameters leaving the performance of this type of tests in large models an open question. Furthermore, Hamilton's test is based on a particular specification of the variance-covariance function that uniquely determines the characteristics of a zero mean random field. A priori, one should expect that different parameterizations of the variance-covariance function gives rise to different tests with different properties from those of Hamilton's statistic. These concerns motivate a search for new tests that are free of unidentified nuisance parameters, regardless of the dimensionality of the system, as well as robust to the specification of the variance-covariance function.

In this paper, within the Hamilton's framework, we construct a battery of tests for neglected nonlinearity that are characterized, mainly by 1) being free of unidentified nuisance parameters, 2) robust to the specification of the variance-covariance function associated with the random field, and 3) easiness in computation. All the proposed tests are written in a TR^2 version of the Lagrange multiplier statistics. By making modifications to the random field governing the outcome of conditional mean function in Hamilton's regression model and by using local approximations, based on Taylor expansions, we are able to provide a TR^2 test that actually is free of unidentified nuisance parameters. The main modification to the random field consists of switching from the Euclidean distance measure to the Minkowski distance measure and to consider random fields that may not have a simple moving average representation. The local approximation approach is related to Lukkonen, Saikkonen and Teräsvirta (1988) who use the method to solve the identification problem that arises in Lagrange multiplier statistics for testing linearity against a STAR alternative, and to the work of Teräsvirta, Lin and Granger (1993) that suggest a version of the Neural Network test free of unidentified nuisance parameters under the null. Relative to Hamilton's statistic, the new battery of statistics based on auxiliary regressions, has a clear advantage with respect to computational simplicity and can actually be calculated within all statistical packages/environments that can perform simple ordinary least square regressions.

We also show that there is an alternative way of testing the null of linearity in Hamilton's flexible regression model to the one advocated by Hamilton. This alternative approach seems particularly promising because the number of unidentified parameters under the null hypothesis equals one regardless of the dimension of the model. Nevertheless, we provide a TR^2 version of the Lagrange multiplier statistics that is also free of unidentified nuisance parameters under the null.

We perform a Monte Carlo study to assess the size and power properties of the proposed new tests. All together, the evidence from the simulation studies seems to suggest that the test statistics, based on the Minkowski distance, that are free of identification problems and do not require full knowledge of the stochastic process that drives the random field, fully matches the power properties of the statistic suggested by Hamilton (2000). In addition, the tests matches the simplicity of the popular Tsay statistic but has a clear advantage over the Tsay test in terms of far better power properties. This may suggest that treating the unobserved nonlinear function as random rather than deterministic seems a particularly fruitful approach for identification of nonlinear components in parametric regression models.

The paper is organized as follows. In section 2, we deal with preliminaries, offering a summary of the basic theory of random fields, we describe Hamilton's flexible regression model and we explain the testing problem. In section 3, we consider alternative representations of the random field. In section 4, we present the battery of Lagrange multiplier tests. Section 5 contains the results of the Monte Carlo study, and in section 6, we conclude.

2 Preliminaries

In this section, we shortly describe some of the main tools needed to derive the new tests for neglected nonlinearity. We begin with a brief description of some of the main results that we use from the theory of random fields or random functions. We continue with a brief introduction to Hamilton's (2000) flexible regression model approach and, in particular, we discuss the assumptions underlying the derivation of the moving average spatial random field governing the outcome of the functional form of the conditional mean function in the regression model.

2.1 Basic definitions and properties of random fields

The purpose of this section is to present some of the most important results of the general theory of random fields that are crucial to the understanding of the new Lagrange multiplier statistics for neglected nonlinearity that we propose. For a more detailed account of the theory of random fields, see Yaglom (1962) and Adler (1981) and the references herein.

Let us begin by explicitly defining the statistical properties governing the random field $\eta(\mathbf{x})$

Definition 1 Let $A \subset \mathbb{R}^k$ be a set. A random field is defined as a function $\eta(\cdot; \mathbf{x}) : \omega \in A \rightarrow \mathbb{R}^m$ such that $\eta(\cdot; \mathbf{x})$ is a random variable for each $\mathbf{x} \in A$. A random field is also denoted $\eta(\mathbf{x})$; $\mathbf{x} \in A$. If $k = 1$; $\eta(\mathbf{x})$ is a random process. For $m > 1$; $\eta(\mathbf{x})$ is a vector random field. In this paper, we focus in scalar random fields: $\eta(\cdot; \mathbf{x}) : \omega \in A \rightarrow \mathbb{R}^1$.

For fixed ω we say that the function $\eta(\mathbf{x})$; $\mathbf{x} \in A$; is a realization of the random field or a sample function.

For any T elements $\mathbf{x}_i \in A$; $i = 1, \dots, T$, the probability distribution function of the T dimensional random variable

$$\eta(\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_T) = (\eta(\mathbf{x}_1); \eta(\mathbf{x}_2); \dots; \eta(\mathbf{x}_T)) \quad (1)$$

is given by

$$F_{x_1; x_2; \dots; x_T}(c_1; c_2; \dots; c_T) = P(\xi(x_1) < c_1; \xi(x_2) < c_2; \dots; \xi(x_T) < c_T) \quad (2)$$

The T dimensional random variable $\xi(x_1; x_2; \dots; x_T)$ will be fully specified once the associated multidimensional distribution function given by (2) is known. In particular, if $\xi(x); x \in A$ is a system of random variables with finite dimensional Gaussian distributions, then the scalar random field is said to be Gaussian, and it is completely determined once we specify the mean function and the covariance function

$$\begin{aligned} \mu(x) &= E\xi(x) \\ B(x; y) &= E(\xi(x) - \mu(x))(\xi(y) - \mu(y)) \end{aligned}$$

2.1.1 Homogeneous or stationary scalar random fields

Let again $A \subset \mathbb{R}^k$ be a set. The scalar random field $\xi(x); x \in A \subset \mathbb{R}^k$ will be called homogeneous or stationary in the strict sense if all of its finite dimensional functions (2) defining $\xi(x)$ are invariant with respect to the group A of shifts in \mathbb{R}^k , i.e. if

$$F_{x_1+\zeta; x_2+\zeta; \dots; x_T+\zeta}(c_1; c_2; \dots; c_T) = F_{x_1; x_2; \dots; x_T}(c_1; c_2; \dots; c_T) \quad (3)$$

for any $T; x_1; x_2; \dots; x_T$; and $\zeta \in \mathbb{R}^k$: In particular, for a random function being strictly homogeneous or strictly stationary this implies that all the one dimensional distribution functions satisfying (3) have to be identical and consequently $\xi(x) \stackrel{d}{=} \xi(x + \zeta)$ for any $\zeta \in \mathbb{R}^k$:

A random field $\xi(x); x \in A \subset \mathbb{R}^k$ satisfying $E\xi^2(x) < \infty$ is called homogeneous in the wide sense if its mathematical expectation $\mu(x)$ and the covariance function $B(x; y)$ are invariant with respect to the group of shifts ζ in \mathbb{R}^k ; that is $\mu(x) = \mu(x + \zeta)$ and the $B(x; y) = B(x + \zeta; y + \zeta)$ for any $x; y; \zeta \in \mathbb{R}^k$:

Theorem 1 If a random field $\xi(x); x \in \mathbb{R}^k \subset \mathbb{R}^1$ is homogeneous or stationary then $E\xi(x) = \mu(x) = \mu$ and the covariance function $B(x; y) = B(x - y)$ depends only on the difference $x - y$

Without loss of generality it will usually be assumed that $\mu = 0$:

If a random field that is homogeneous in strict sense possesses a second order moment, it is homogeneous in the wide sense too. Hence, in the case of Gaussian scalar random fields both concept of homogeneity coincide.

2.1.2 Homogeneous and isotropic scalar random fields

A homogeneous random field $\{X(\mathbf{x}) : \mathbf{x} \in A \subset \mathbb{R}^k\}$ in a Euclidean space is said to be isotropic if its correlation function depends only on the Euclidean measure of distance between the two vector-arguments upon which it is defined. Formally, this property writes

Theorem 2 If a random field $\{X(\mathbf{x}) : \mathbf{x} \in A \subset \mathbb{R}^k\}$ is homogeneous and isotropic then $E\{X(\mathbf{x})\} = \mu$ and the correlation function $B(\mathbf{x}; \mathbf{y}) = B(\|\mathbf{x} - \mathbf{y}\|)$ depends only on $\|\mathbf{x} - \mathbf{y}\|$ where $\|\cdot\|$ denotes the Euclidean measure of distance.

2.1.3 Continuity and differentiability of scalar random fields

A random field $\{X(\mathbf{x}) : \mathbf{x} \in A\}$, having a finite second order moment is said to be mean square continuous at point \mathbf{x}_0 if $E\{[X(\mathbf{x}_0 + \mathbf{h}) - X(\mathbf{x}_0)]^2\} \rightarrow 0$ as $\|\mathbf{h}\| \rightarrow 0$. If this relation holds for any $\mathbf{x}_0 \in A$, the field $\{X(\mathbf{x}) : \mathbf{x} \in A\}$ is called mean square continuous on the set A . In order for a random field $\{X(\mathbf{x}) : \mathbf{x} \in A\}$ having $E\{X^2(\mathbf{x})\} < \infty$ to be mean square continuous it is necessary and sufficient that the function $\text{cov}\{X(\mathbf{x}); X(\mathbf{y})\}$ is continuous along the diagonal $f(\mathbf{x}; \mathbf{y}) \in A \times A : \mathbf{x} = \mathbf{y}$:

Theorem 3 A random field $\{X(\mathbf{x}) : \mathbf{x} \in A \subset \mathbb{R}^k\}$ is continuous in mean square at the point $\mathbf{x}^0 \in \mathbb{R}^k$ if and only if its covariance function $B(\mathbf{x}; \mathbf{y})$ is continuous at the point $\mathbf{x} = \mathbf{y} = \mathbf{x}^0$: If $B(\mathbf{x}; \mathbf{y})$ is continuous at every diagonal point $\mathbf{x} = \mathbf{y}$ then it is everywhere continuous.

We now turn to consider differentiability in mean square, for which we assume that the random field $\{X(\mathbf{x}) : \mathbf{x} \in A\}$ is real valued. The following results can be established

Theorem 4 Let \mathbf{e}_i be a vector whose i th element is 1 with all others being zero. If the derivative $\frac{\partial^2 B(\mathbf{x}; \mathbf{y})}{\partial x_i \partial y_i}$ exists and is finite at the point $(\mathbf{x}; \mathbf{x}) \in \mathbb{R}^{2k}$ then the limit

$$\mu_i(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{E\{[X(\mathbf{x} + h\mathbf{e}_i) - X(\mathbf{x})]^2\}}{h} \quad (4)$$

exists and $\mu_i(\mathbf{x})$ is called the mean square derivative of $X(\mathbf{x})$ at \mathbf{x} : If this exists for each $\mathbf{x} \in \mathbb{R}^k$ then $X(\mathbf{x})$ is said to possess a mean square derivative. The covariance function of $\mu_i(\mathbf{x})$ is then given by $\frac{\partial^2 B(\mathbf{x}; \mathbf{y})}{\partial x_i \partial y_i}$:

It is interesting to note that mean square continuity and mean square differentiability of a random field is determined by the ordinary continuity and differentiability of the corresponding covariance function.

Notice that when the basic process $\mu(x)$ is homogeneous the conditions ensuring mean square continuity and the existence of mean square derivatives become particularly simple. As an example, $\mu(x)$ will be mean square continuous at s if its covariance function is continuous at $(s; s)$. But because of homogeneity this is equivalent to demanding that $B(x)$ will be continuous at $s; s = 0$: Hence if $B(x)$ is continuous at the origin, $\mu(x)$ will be mean square continuous for all $x \in \mathbb{R}^k$: Similarly, if the $2k$ th-order partial derivative $\frac{\partial^{2k} B(x)}{\partial x_{i_1} \dots \partial x_{i_k}}$ exists and is finite at the point $x = 0$ the k th order partial derivative of $\mu_{i_1 \dots i_k}(x)$ exists for all $x \in \mathbb{R}^k$ as a mean square limit.

2.2 Hamilton's flexible regression model approach

Hamilton (2000) suggests representing the nonlinear component in a general regression model by an isotropic Gaussian random field. Let $y_t \in \mathbb{R}^1$ and $x_t \in \mathbb{R}^k$ for $t = 1; \dots; T$. The general model suggested by Hamilton (2000) can be written as

$$y_t = \mu_0 + x_t^0 + \lambda m(g \cdot x_t) + \epsilon_t \quad (5)$$

where ϵ_t is a sequence of independent $N(0; \sigma^2)$ -distributed error terms, the parameter λ represents the contribution of the nonlinear part to the conditional mean, the $k \in 1$ vector g drives the curvature of the conditional mean, the symbol \cdot denotes element-by-element multiplication, and $m(z)$; for any choice of z ; represents a realization from a random field that can be described as follows. Define a grid in \mathbb{R}^k by the nodes $f(x_{ij})$ where $x_{ij} = x_{i;j} + \phi$ for $i = 1; \dots; k$ and $j = 1; \dots; N$. Define the set A_N that consists of the N^k distinct points in \mathbb{R}^k covered by this grid

$$A_N = \{f(x_{ij_1}; x_{ij_2}; \dots; x_{ij_k}); j_1 = 1; \dots; N; \dots; j_k = 1; \dots; N\} \quad (6)$$

For each $x \in A_N$ associate $e(x) \sim N(0; 1)$ such that $e(x)$ is independent of $e(w)$ for $x \neq w$. Finally, define the set $B_N(x) \subseteq A_N$ as

$$B_N(x) = \{w \in A_N : (x_i - w_i)^2 \leq 1\} \quad (7)$$

It follows that $B(x)$ is the set of all points in A_N whose Euclidean distance from x is less than or equal to unity. We will now consider the scalar $m_N(x)$

associated with every point x in A_N given by the expression

$$m_N(x) = n_N(x)^{-\frac{1}{2}} \sum_{w \in B_N(x)} e(w) \quad (8)$$

where $n_N(x)$ denotes the number of points in $B_N(x)$: A random field generated as a sequence of uncorrelated random variables, as $m_N(x)$ in (8), is said to have a moving average representation, e.g. Yaglom (1962, p.31). Furthermore, by assuming $e(\cdot)$ to be independently and Gaussian distributed with zero mean implies that the random field is uniquely determined by its limiting covariance function as we saw in the previous section. The motivation for considering random fields generated as a sequence of uncorrelated random variables is that the associated covariance function often can be derived explicitly from the moving average representation of the random field and, by construction, is non-negative definite. From equation (8) it is straightforward to determine the limiting properties of the moving average random field in general, and the limiting covariance function in particular. In the limit, the partition Φ becomes arbitrarily fine and $m_N(x) \rightarrow m(x)$; where $m(x)$ now represents a continuous valued k -dimensional random field with the following well defined statistical properties¹

$$m(x) \gg N(0; 1) \quad (9)$$

Furthermore, for any x and z defined on \mathbb{R}^k ; the correlation between $m(x)$ and $m(z)$ is given by the ratio of the volume of the overlap of k -dimensional unit spheroids centered at x and z to the volume of a single k -dimensional unit spheroid. Hence, if the Euclidean distance between x and z is greater than 2; the correlation between $m(x)$ and $m(z)$ will be equal to zero. If we let $H_k^e(x; z) = E(m(x)m(z))$ then the covariance function $H_k^e(x; z)$ is given by

$$H_k^e(x; z) = \begin{cases} \frac{1}{2} G_{k-1}(h; 1) = G_{k-1}(0; 1) & \text{if } h \leq 1 \\ 0 & \text{if } h > 1 \end{cases} \quad (10)$$

$$G_k(h; r) = \int_h^r (r^2 - w^2)^{k-2} dw$$

$$h = \frac{1}{2} d_{L_2}(g^{-1}x; g^{-1}z)$$

where $d_{L_2}(v; w)$ is a real-valued function that maps the Euclidean distance between $v \in \mathbb{R}^k$ and $w \in \mathbb{R}^k$ onto $[-1, 1]$: The expression for H_k^e turns out

¹For a formal proof, see Hamilton (2000)

to be a rather complex function particular in the cases of k being an even number. Hamilton (2000) provides the general formulae for any number of regressors. In the geostatistical literature, this covariance is widely known as the spherical covariance function, and was derived by alternative methods by Matheron (1973).²

Example 1 Let $x_t \in \mathbb{R}^k$ for $t = 1, 2, \dots, T$ be observations on the explanatory variables in the model given equation (5). Furthermore, denote the spherical covariance function suggested by Hamilton (2000) as $E(m(x_t)m(x_s)) = H_k^e(x_t; x_s)$. In the case of $k = 1, 2, 3$; $H_k^e(x_t; x_s)$ will be given by

$$H_1^e(x_t; x_s) = \begin{cases} \frac{1}{2} (1 - h_{ts}) & \text{if } h_{ts} \leq 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} \quad (14)$$

$$H_2^e(x_t; x_s) = \begin{cases} \frac{1}{2} (1 - \frac{2}{3}[h_{ts}(1 - \frac{1}{2}h_{ts}^2) + \sin^{-1}(h_{ts})]) & \text{if } h_{ts} \leq 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} \quad (15)$$

$$H_3^e(x_t; x_s) = \begin{cases} \frac{1}{2} (1 - \frac{3}{2}h_{ts} + \frac{1}{2}h_{ts}^3) & \text{if } h_{ts} \leq 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases} \quad (16)$$

where $h_{ts} = \frac{1}{2}d_{L_2}(g^{-1}x_t; g^{-1}x_s)$ and $d_{L_2}(v; w)$ is defined as previously.

²Within the area of geostatistics, Matheron (1973) suggests an alternative method to compute the covariance function for any arbitrary value k , and provides a method to verify that the spherical covariance function actually is nonnegative definite. In particular, he shows that if the $C(h)$ is isotropic such that there exists a function $C_k(h)$ on \mathbb{R}_+ then, by the so called turning bands operator, the permissible covariance function $C_k(h)$ will be given by

$$C_k(h) = 2 \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k-1}{2})} \int_0^1 C_1(uh)(1-u^2)^{\frac{1}{2}(k-3)} du \quad (11)$$

where Γ is the gamma function. If we define C_1^a as follows

$$C_1^a(s) = \begin{cases} \frac{1}{2} (1 - s) & \text{if } s \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

then

$$C_k^a(h) = 2 \frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k-1}{2})} \int_h^1 (1-uh)(1-u^2)^{\frac{1}{2}(k-3)} du \quad (13)$$

for $k > 1$. From this formula, it can be verified that $C_k^a(h) = H_k^e(h)$ for all values of $k > 1$: From (11) and the moving average representation of the random field, it is possible to generate a very large class of stationary random fields. Other procedures to build permissible covariance functions by use of spectral methods are described in Christakos (1984).

2.3 Testing for nonlinearities

Consider the model given in equation (5). It is easy to observe that a test for neglected nonlinearity will be subject to a nuisance parameter problem, where a set of parameters are identified only under the alternative hypothesis. There are two different ways to specify the null hypothesis of linearity.

Case 1. If the null hypothesis is written as $H_0 : \sigma^2 = 0$; the parameter vector g is unidentified under the null.

Case 2. If the null hypothesis is written as $H_0 : g_1 = g_2 = \dots = g_k = 0$; the parameter σ is also unidentified under the null. Furthermore, in this second case, the stochastic process under the null becomes non-ergodic. To see this, consider the model (5) under the null

$$y_t = \mu_0 + x_t^0 + \sigma m(0) + \epsilon_t$$

$$m(0) \gg N(0; 1)$$

It becomes apparent that the model will be linear on x_t , but y_t will be non-ergodic since

$$E(y_t y_{t-s}) = \begin{cases} \frac{1}{2} \sigma^2 + \frac{3}{4} \sigma^2 & \text{for } s = 0 \\ \sigma^2 & \text{for } s > 0 \end{cases}$$

even under the strict assumption that x_t is deterministic. Ergodicity is a critical assumption for the law of large numbers to hold. This could imply that a test for nonlinearity based on the parameter vector g may not have a well defined asymptotic distribution under the null.

We propose a solution to the nuisance parameter problem in Case 1 and Case 2 that is based on alternative representations of the random field $m(x)$, which we will discuss in the next section.

3 Alternative representations of the random field $m(x)$

In this section, we present different representations of the random field. Our main concern is to assure that the covariance function associated with a random field is non-negative definite, hence permissible. We maintain the assumptions of Gaussianity and homogeneity of the random field while exploring the role played by the isotropy assumption.

3.1 A non-isotropic random field

Hamilton's framework is developed by using a moving average representation of the random field under the Euclidean distance or L_2 norm. We propose an alternative moving average representation based on the Minkowski distance or L_1 norm as a device to simplify the tests of neglected nonlinearity. The general setup follows Hamilton (2000) but instead of the set $B_N(x)$; consider the set $B_N^\alpha(x)$ defined as

$$B_N^\alpha(x) = \{w \in A_N : |x_i - w_j| \leq 1\}$$

where $\mathbf{1}$ is a k -dimensional vector consisting of ones in each entry and let $m_N^\alpha(x)$ be given by the moving average representation

$$m_N^\alpha(x) = n_N^\alpha(x)^{-\frac{1}{2}} \sum_{w \in B_N^\alpha(x)} e(w) \quad (17)$$

where $n_N^\alpha(x)$ now denotes the number of points in $B_N^\alpha(x)$: In the limit, as the partition becomes arbitrarily fine we have that $m_N^\alpha(x) \rightarrow m^\alpha(x)$ for $\alpha > 0$; where $m^\alpha(x)$ is a continuously valued k -dimensional random field distributed as

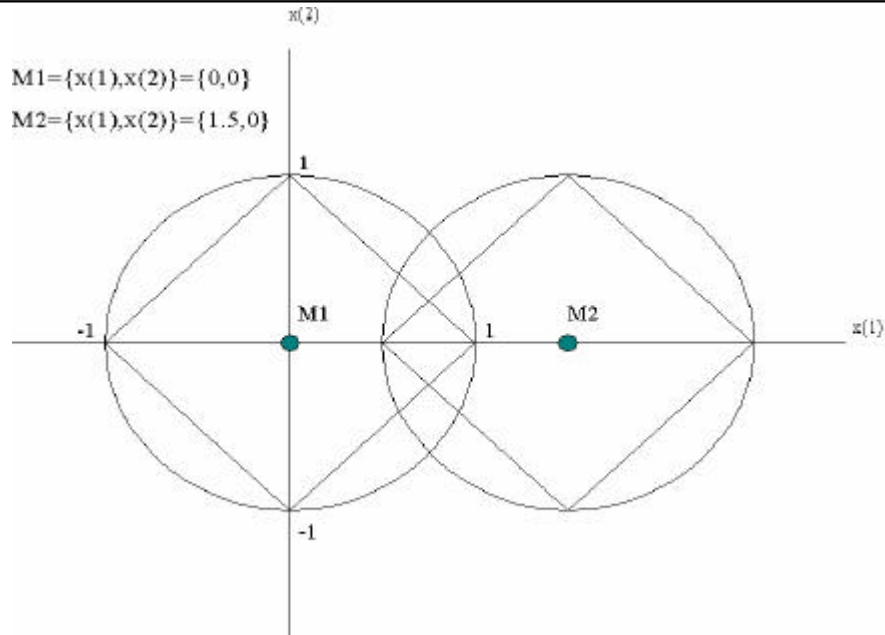
$$m^\alpha(x) \sim N(0; 1)$$

which is identical to the distribution $m(x)$: However, the correlation between $m^\alpha(x)$ and $m^\alpha(z)$ for any arbitrary $x \neq z$ in \mathbb{R}^k will be different. The correlation between $m^\alpha(x)$ and $m^\alpha(z)$ will be given by the volume of relative overlap between k -dimensional orthogons centered at x and z relative to the volume of a k -dimensional unit orthogon. The volume of overlap, and hence the correlation between $m^\alpha(x)$ and $m^\alpha(z)$ will be equal to zero if the Minkowski distance between x and z is greater than or equal to two. We will denote this particular covariance function $H_k^\alpha(x; z)$:

Following pure geometric arguments, we obtain expressions for $H_k^\alpha(x; z) = E(m^\alpha(x)m^\alpha(z))$ in the case of $k = 1; 2; 3$:

Example 2 Let $x_t = (x_{1t}; x_{2t}; \dots; x_{kt}) \in \mathbb{R}^k$ for $t = 1; 2; \dots; T$ be observations on the explanatory variables in the model given in equation (5) where $m(x_t)$ is replaced by $m^\alpha(x_t)$. Furthermore, denote the associated covariance function as $E(m^\alpha(x_t)m^\alpha(x_s)) = H_k^\alpha(x_t; x_s)$. Finally, define $d_{L_1}(v; w) = \sum_{i=1}^k |v_i - w_i|$ as a real-valued function mapping the distance between the vector values

Figure 1: Grid of values for $x(i_1; i_2)$ for $k = 2$ and $\Phi_{1N} = \Phi_{2N}$ and illustration of nodes whose values for $e(x)$ get averaged to determine $E(m_N(0;0)m_N(1;5;0))$ - given by the intersection of the unit circles - and $E(m_N^a(0;0)m_N^a(1;5;0))$ - given by the intersection of the two othogons.



$(z; w) \leq 2^{-k}$ as the Minkowski L_1 norm such that $d_{L_1}(v; w) = \sum |v_i - w_i|$. In the case of $k = 1; 2; 3$; $H_k^a(x_t; x_s)$ will be given by

$$H_1^a(x_t; x_s) = \begin{cases} \frac{1}{2} (1 - h_{ts}^+) & \text{if } h_{ts}^+ \leq 1 \\ 0 & \text{if } h_{ts}^+ > 1 \end{cases}$$

$$H_2^a(x_t; x_s) = \begin{cases} \frac{1}{2} (1 - h_{ts}^+)^2 + (1 - h_{ts}^+) \min\{jx_{1t} - x_{1s}; jx_{2t} - x_{2s}\} & \text{if } h_{ts}^+ \leq 1 \\ 0 & \text{if } h_{ts}^+ > 1 \end{cases}$$

$$H_3^a(x_t; x_s) = \begin{cases} \frac{8}{3} (1 - h_{ts}^+)^3 + \frac{3}{2} (1 - h_{ts}^+)^2 \{jx_{2t} - x_{2s} + jx_{3t} - x_{3s}\} + \frac{1}{2} (1 - h_{ts}^+) \{jx_{2t} - x_{2s}\} \{jx_{3t} - x_{3s}\} & \text{if } h_{ts}^+ \leq 1 \\ 0 & \text{if } h_{ts}^+ > 1 \end{cases}$$

respectively, where $h_{ts}^+ = \frac{1}{2} d_{L_1}(g^- x_t; g^- x_s)$. For $k = 3$; we need $jx_{1t} - x_{1s}$, $jx_{2t} - x_{2s}$ and $jx_{3t} - x_{3s}$.

Notice, that in the case of $k = 1$ the L_2 norm and the L_1 norm will provide identical covariance functions. Furthermore, the expression for

$H_2^a(x_t; x_s)$ is highly simplified in comparison to the expression for $H_2^g(x_t; x_s)$: The reason is that, in the case of $k = 2$; the volume of relative overlap will be equal to the area of a rectangle which will have relatively simple mathematical expressions. Figure 1 illustrates the differences between the sets $B_N(x)$ and $B_N^a(x)$ on calculating the covariance function.

The disadvantage of a random field with a moving average representation in L_1 is that the field is not isotropic. Consider two orthogons centered at points x and y , with the property that x and y are equidistant in L_1 from the origin. It is easy to see that the volume of relative overlap between the orthogon centered at the origin and the orthogon centered at x is different from the relative overlap between the orthogon centered at the origin and the orthogon centered at y : In other words, the location of the points matter, as it can be read from the formulas of $H_k^a(x_t; x_s)$ in the previous example.

Consider now the following homogeneous covariance function $C_k^a(x; z) = C_k^a(h^+)$ for $x; z \in \mathbb{R}^k$, that depends only on the distance measure $h^+ = \frac{1}{2}d_{L_1}(x; z)$ such that

$$C_k^a(h^+) = \begin{cases} \frac{1}{2} (1 - h^+)^k & \text{if } h^+ \leq 1 \\ 0 & \text{if } h^+ > 1 \end{cases} \quad (18)$$

for $k \geq 1$: Notice that for all k ; $C_k^a(h^+)$ equals the first term in $H_k^a(x; z)$; and does not depend on location: Two questions are pertinent. First, is (18) a permissible covariance function? Second, if it is, what is the random field associated with it? To answer the first question, we need to check the necessary and sufficient conditions of permissibility outlined in Yaglom (1962) and Christakos (1984, 1992). To answer the second question, we need the theorem of Khinchin (1934).

According to Yaglom (1962) and Christakos (1984) any permissible covariance function - associated with a general homogeneous Gaussian random field such that $C_k(x; z) = C_k(x - z)$ - must satisfy the following necessary conditions in the time domain

$$\begin{aligned} C_k(0) &> 0 \\ |C_k(x - z)| &\leq C_k(0) \\ C_k(x - z) &= C_k(z - (x - z)) \\ \lim_{|x - z| \rightarrow \infty} \frac{C_k(x - z)}{|x - z|^{(1-k)/2}} &= 0 \end{aligned}$$

for $(x - z) \in \mathbb{R}^k$: Sufficient conditions for permissibility for isotropic processes are spelled out in Christakos (1992) for $k = 1; 2; 3$:³

³Notice that if a random field is isotropic, it is also homogenous.

$$\begin{aligned}
& C_k^0(d(x; z))|_{d(x; z)=0} < 0 \\
& C_k^{00}(d(x; z)) \geq 0 \text{ in } <^1 \\
& \int_{d(x; z)}^u \frac{1}{u^2} \frac{1}{d(x; z)^2} dC_k^{00}(d(x; z)) \geq 0 \text{ in } <^2 \\
& C_k^{00}(d(x; z)) \int_{d(x; z)} C_k^{000}(d(x; z)) \geq 0 \text{ in } <^3
\end{aligned}$$

From these conditions on permissible covariance functions, it is apparent that they are very nicely behaved functions. The proposed covariance function (18) clearly satisfies the previous conditions. These characteristics will become important in the forthcoming sections when we discuss Taylor's approximations of the covariance function $C_k(x; z)$. Due to its pure simplicity, $C_k^a(h^+)$ will serve as an important tool for implementation of our proposed tests for neglected nonlinearity.

If the random field has a moving average representation given by equation (17), and we use the simple $C_k^a(h^+)$ in (18) instead of $H_k^a(x; z)$; we would underestimate the covariance function. However, the theorem of Khinchin (1934) tells us that if it is not possible to construct a simple covariance function on the L_1 based on a moving average representation of the random field, we can start by specifying a simple covariance function on L_1 . If the covariance function is permissible, the theorem states that there is a uniquely determined Gaussian and homogeneous random function associated with it.

Based on the methodology used by Hamilton (2000) and the theorem of Khinchin (1934) we are able to show - by Theorem 5 and 6 - that $C_k^a(h^+)$ uniquely determines the properties of a particularly homogeneous and Gaussian random field.

Theorem 5 Let $x \in \mathbb{R}^k$ and $\bar{z} \in \mathbb{R}^k$ and let $m^a(x)$ and $m^a(\bar{z})$ be random field generated as the limit of equation (17) as $\Phi \rightarrow 0$ evaluated at fixed points $x = (x_1; x_2; \dots; x_k)$ and $\bar{z} = (\bar{z}_1; \bar{z}_2; \dots; \bar{z}_k) \in \mathbb{R}^k$ where $h^+ = \frac{1}{2}d_{L_1}(x; \bar{z})$. Then

$$E(m^a(x)m^a(\bar{z})) = C_k^a(h^+)$$

where

$$C_k^a(h^+) = \begin{cases} \frac{1}{2} (1 - h^+)^k & \text{if } h^+ \leq 1 \\ 0 & \text{if } h^+ > 1 \end{cases}$$

implying that $C_k^a(h^+)$ is a permissible covariance function for any value of $h^+ \in \mathbb{R}_+$

Theorem 6 (Khinchin, 1934) Let $x \in \mathbb{R}^k$ and $z \in \mathbb{R}^k$ and let $C_k^{\alpha}(h^+)$ be the permissible covariance function given by Theorem 5. Since $C_k^{\alpha}(h^+)$ only depends on $h^+ = \frac{1}{2}d_{L_1}(x; z) \in \mathbb{R}_+$ there exists a homogeneous and unique Gaussian random field $m(x)$ ($m(x) \in \mathcal{M}^{\alpha}(x)$) such that

$$E(m(x)m(z)) = C_k^{\alpha}(h^+)$$

The proofs of Theorem 5 and Theorem 6, are given in the appendix.

3.2 The concept of structure functions

In the previous section, we showed that testing for linearity in model (5) with the null hypothesis

$$H_0 : g_1 = g_2 = \dots = g_k = 0$$

produces a non-ergodic stochastic process under the null. We will provide a simple modification of the specification of the function $m(x)$ that preserves the ergodicity of y_t under the null as well as under the alternative. Consider the modified unrestricted model

$$y_t = x_t^0 + \beta m(g(x_t)) + \varepsilon_t$$

where

$$m(x) = m(x) + m(0)$$

Notice that $m(0) = 0$; and the model under the null becomes

$$y_t = x_t^0 + \varepsilon_t$$

restoring the ergodicity of y_t under the null hypothesis, provided that x_t and ε_t are stationary. We need to specify the covariance structure of $m(x)$: Since we will restrict ourselves to the class of Gaussian random fields we know that $m(x)$ - the sum of two Gaussian random fields - also will be Gaussian, i.e.

$$m(x) \gg N(0; \mathcal{C}_k(x; z))$$

where we consider $\mathcal{C}_k(x; z)$ the covariance function that uniquely determines the random field $m(x)$: It is given by $\mathcal{C}_k(x; z) = E(m(x)m(z))$ as

$$\begin{aligned} \mathcal{C}_k(x; z) &= E(m(x) + m(0))(m(z) + m(0)) \\ &= E(m(x)m(z)) + E(m(0)m(0)) + \\ &\quad E(m(x)m(0)) + E(m(z)m(0)) \end{aligned}$$

By writing

$$\begin{aligned}
 E(m(x)m(z)) &= C_k(x; z) \\
 E(m(0)m(0)) &= C_k(0) \\
 E(m(x)m(0)) &= C_k(x) \\
 E(m(z)m(0)) &= C_k(z)
 \end{aligned}$$

the general expression for the covariance function becomes

$$\mathfrak{C}_k(x; z) = C_k(0) + C_k(x; z) - C_k(x) - C_k(z) \quad (19)$$

$\mathfrak{C}_k(x; z)$ is called the structure function by Yaglom (1962, p.87) and was first introduced by Kolmogorov (1941a, 1941b) and applied in statistical turbulence theory. Structure functions are particularly important when the random field is not stationary but the increments of the random field are stationary. However, in the case of the random function being stationary, the structure function given by $\mathfrak{C}_k(x; z)$ is still a perfectly valid covariance function provided that $C_k(x; z)$ is a valid covariance function, e.g. Yaglom (1962, p.88). This property will be important to have in mind when deriving the Lagrange multiplier statistics corresponding to the null $g_1 = g_2 = \dots = g_k = 0$:

4 Lagrange multiplier tests for neglected nonlinearity

In this section we will derive various Lagrange multiplier tests for neglected nonlinearity. We will consider two classes of tests. The first class of tests - denoted ψ -tests - are derived under the null $H_0 : \psi^2 = 0$; while the second class of tests - denoted g -tests - are derived under the null $H_0 : g_1 = g_2 = \dots = g_k = 0$. We introduce a ψ -test for neglected nonlinearity that does not require full knowledge of the parametric specification of the covariance function and in particular, it does not require any prior knowledge about the numerical values of each element in the g vector. Analogously, we introduce a g -test for neglected nonlinearity that does not depend on the values of ψ : It should be emphasized that since the test statistics are all Lagrange multiplier test statistics they are characterized by being the locally most powerful tests asymptotically.

4.1 The likelihood function

The concentrated log likelihood function corresponding to the model given by equation (5) is

$$\begin{aligned} \ell(y; X; \beta; g; \Sigma) &= -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln | \Sigma^2 C_k + \Sigma^2 I_T | \\ &\quad - \frac{1}{2} [y - X \mathbf{b}_T(\beta; g; \Sigma)]' (\Sigma^2 C_k + \Sigma^2 I_T)^{-1} [y - X \mathbf{b}_T(\beta; g; \Sigma)] \\ \mathbf{b}_T(\beta; g; \Sigma) &= [X' (\Sigma^2 C_k + \Sigma^2 I_T)^{-1} X]^{-1} [X' (\Sigma^2 C_k + \Sigma^2 I_T)^{-1} y] \end{aligned} \quad (20)$$

where C_k denotes the covariance function associated with the random field, which will equal $H_k^e(d_{L_2}(g^{-1}x_t; g^{-1}x_s))$ when we refer to Hamilton's spherical covariance function given by equation (10), and where

$$\begin{aligned} y &= (y_1; y_2; \dots; y_T)' \\ X &= \begin{pmatrix} 1 & x_1^0 & \dots & 1 \\ 1 & x_2^0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_T^0 & \dots & 1 \end{pmatrix} \\ \beta &= (\beta_1; \beta_2; \dots; \beta_T)' \\ \Sigma &= (\Sigma_1; \Sigma_2; \dots; \Sigma_T)' \end{aligned}$$

This likelihood function is the basis for the Lagrange multiplier tests for neglected nonlinearity.

We construct a Lagrange multiplier test for a generic null hypothesis for a parameter vector μ , $H_0: \mu = \mu_0$; where $\mu = (\mu_1; \mu_2)'$ with μ_2 being the maximum likelihood estimate of μ_2 under the null, and μ_1 the parameter μ_1 specified under the null. The Lagrange multiplier statistic is given by

$$LM = T^{-1} s(\hat{\mu})' I^{-1}(\mu_0) s(\hat{\mu}) \quad (21)$$

where $s(\mu)$ denotes the score function, and $I^{-1}(\mu)$ the inverse of the information matrix, both evaluated under the null. We consider two different estimators of the information matrix both consistent under the null. The first estimator, denoted I_H ; is based on the Hessian of the log likelihood function, and the second, denoted I_{OP} ; is based on the outer-product of the scores. Using the I_{OP} estimator, we construct the TR^2 version of the Lagrange

multiplier, where the R^2 is the uncentered coefficient of determination of a particular least square regression. Under the usual regularity conditions, the Lagrange multiplier statistics will be $\hat{A}^2(q)$ -distributed where q equals the number of restrictions under the null. When a subset of the parameters in μ_2 is not identified under the null it is not possible to obtain consistent maximum estimates of this subset of parameters under the null. In this situation, we say that the testing problem is suffering from unidentified nuisance parameters. Unidentified nuisance parameters makes the evaluation of $s(\mu)$ at $\hat{\mu}$; where $\hat{\mu}$ is a consistent estimate of μ under the null, impossible. Methods to solve the problem of unidentified nuisance parameters depend to some extent on the particular model under consideration, hence we will deal with the problems in details after having specified the models we wish to consider under the null as well as under the alternative.

4.2 Tests based on known covariance functions

Hamilton (2000) derives the χ^2 -test for neglected nonlinearity based on the I_H estimator of the information matrix and with $C_k(x_t; x_s) = H_k^e(d_{L_2}(g^{-1}x_t; g^{-1}x_s))$ in the log likelihood function. Let us denote Hamilton's test statistics χ^2_{H} : Under the null hypothesis, the parameter vector g is not identified. To deal with this identification issue, Hamilton suggests fixing the parameter g to a given value and proceed to derive the Lagrange multiplier test, which then follows a standard asymptotic distribution. Heuristically, fixing g will be identical to a derivation of the Lagrange multiplier statistics assuming complete knowledge of the covariance matrix associated with the random field. Our first aim will be to derive the TR^2 version of the test statistics conditional on full knowledge of the parameterized covariance function. Recall the log-likelihood function (20). Keeping g fixed and defining $\Sigma = \sigma^2 C_k + \frac{1}{4} I_T$, the score functions are equal to

$$S(\sigma^2)j_g = -\frac{1}{2} \text{tr}(-\sigma^{-1} C_k(I_T - \sigma^{-1} 22^0))$$

$$S(\frac{1}{4})j_g = -\frac{1}{2} \text{tr}(-\sigma^{-1} (I_T - \sigma^{-1} 22^0))$$

Evaluating the score functions under the null of linearity ($H_0 : \sigma^2 = 0 \Rightarrow \Sigma = \frac{1}{4} I_T$)⁴ gives

⁴ Though this null is on the boundary of the parameter space, the properties of the LM test are not altered, see Godfrey (1988) pg. 92-98.

$$S(\cdot^2)_{j_s^2=0:g} = i \frac{1}{2^{3/4^2}} \text{tr}(C_k(I_T i \frac{22^0}{3/4^2}))$$

$$S(3/4^2)_{j_s^2=0:g} = i \frac{1}{2^{3/4^2}} \text{tr}(I_T(I_T i \frac{22^0}{3/4^2}))$$

Using the following property of the trace $\text{tr}(AB) = ((A^v)^0(B^0)^v)$; where the superscript v denotes the vec-operator, the score functions are written as

$$S(\cdot^2)_{j_s^2=0:g} = i \frac{1}{2^{3/4^2}} \mathfrak{X}_1^0 u \quad (22)$$

$$S(3/4^2)_{j_s^2=0:g} = i \frac{1}{2^{3/4^2}} \mathfrak{X}_2^0 u \quad (23)$$

where

$$\mathfrak{X}_1 = C_k^v \quad (24)$$

$$\mathfrak{X}_2 = I_T^v$$

$$u = (I_T i \frac{22^0}{3/4^2})^v$$

The TR^2 version of the Lagrange multiplier statistics - which we will denote \mathcal{E}_{OP} - is obtained by the following procedure

1. Estimate the model under the null and obtain an estimate of the residuals and of the variance of the residuals, i.e.

$$\mathbf{b} = y - X(X^0X)^{-1}X^0y$$

$$\mathbf{b}^2 = T^{-1} \mathbf{b}^0 \mathbf{b}$$

2. Obtain an OLS estimate of v - denoted \mathbf{b} - based on the auxiliary regression

$$\mathbf{b} = \beta_1 \mathfrak{X}_1 + \beta_2 \mathfrak{X}_2 + \epsilon$$

where $\mathbf{b} = (I_T i \frac{22^0}{3/4^2})^v$ and the regressors \mathfrak{X}_1 and \mathfrak{X}_2 are defined as above.

3. Obtain the uncentered R^2 from this regression as

$$R^2 = \frac{SSR_0 - SSR(g)}{SSR_0}$$

$$SSR_0 = \mathbf{b}^0 \mathbf{b}$$

$$SSR_1(g) = \mathbf{b}^0 \mathbf{b}$$

4. Calculate the Lagrange multiplier statistics as

$$\hat{L}_{OP}^E(g) = T^2 R^2 \gg \hat{A}^2(1)$$

In order to stress that also the TR^2 version of the Lagrange multiplier statistics depends crucially on the choice of a fully specified covariance matrix we write it as $\hat{L}_{OP}^E(g)$: Notice the computational simplicity of the statistics - where only the calculation of the vector \mathcal{X}_1 may seem a bit tedious, but as we will now illustrate by some examples it will in general be easy to obtain. Finally, notice that since there are T^2 observations in the auxiliary regression, R^2 should be multiplied by T^2 instead of T in order to make the $\hat{L}_{OP}^E(g)$ statistics \hat{A}^2 -distributed asymptotically.

Example 3 Computation of the vector \mathcal{X}_1 when it is assumed that the covariance function associated with the unobserved random function equals Hamilton's spherical covariance function. For simplicity consider the case of only one regressor, i.e. $k = 1$. In this case \mathcal{X}_1 will equal

$$\begin{aligned} \mathcal{X}_1 &= (1 - I(gz))^{-1} I(gz - 1) \\ z &= \left(\frac{1}{2} \sum_{i=1}^T x_i^0 (x_i^0)^0 \right)^v \\ x &= (x_{11}; x_{21}; \dots; x_{T1})' \end{aligned}$$

where $I(w = 1)$ denotes the indicator function that equals unity if the condition $w = 1$ is satisfied, and zero otherwise. The vector where all entries equal unity is given by $\mathbf{1}$: Having obtained the expression for \mathcal{X}_1 the computation of the $\hat{L}_{OP}^E(g)$ statistics turns out to be straightforward once a proper value of $g \in \mathbb{R}_+^1$ is specified. When working with the spherical covariance function we suggest choosing g as in Hamilton (2000)⁵.

Example 4 Computation of the vector \mathcal{X}_1 when the covariance function is of the multiplicative Ornstein-Uhlenbeck type. In the case of k regressors \mathcal{X}_1 will equal

$$\begin{aligned} \mathcal{X}_1 &= \exp(i g^0 r^a) \\ r^a &= (z_1^0; z_2^0; \dots; z_k^0) g^0 \\ z_1 &= \left(\frac{1}{2} \sum_{i=1}^T x_{i1}^0 (x_{i1}^0)^0 \right)^v \\ x_1 &= (x_{11}; x_{21}; \dots; x_{T1})' \\ l &= (1; \dots; k) \end{aligned}$$

⁵ The size and power properties of Hamilton's (2000) original statistics for this particular choice of g are analyzed in Dahl (1999) and turned out to be very encouraging.

Again the choice of $g \in \mathbb{R}_+^k$ will be crucial to the entries in \mathcal{X}_1 and consequently for the $\int_{OP}^E(g)$ statistics.

4.3 \int_{OP}^E tests based on unknown covariance functions

In this section we suggest an alternative method of testing the null hypothesis of linearity where an exact parameterization of the covariance function is not needed in order to obtain the statistics. The approach we suggest is based on the substitution of $\mathcal{X}_1 = C_k^V$ for a Taylor approximation of \mathcal{X}_1 in the auxiliary regression. Our approach can be viewed as a further generalization of the ideas in Luukkonen et al. (1988) where the problem of interest is testing linearity against a smooth transition autoregressive model. By using the Taylor approximation we are able to capture the characteristics of a broader class of unknown but continuous and twice differentiable real-valued nonlinear functions C_k^V . Since in most situations, the fully parameterized covariance function will be unobservable, any specification will constitute only an approximation of the true covariance function. One can argue that we model this information problem explicitly. The class of models we consider is given by

$$y_t = x_t^0 + \int_{OP}^E m(g - x_t) + \varepsilon_t$$

where $m(x)$ now is assumed to be a smooth - continuous and differentiable - homogenous and Gaussian random field given by

$$\begin{aligned} m(x) &\gg N(0; C_k(x; z)) \\ C_k(x; z) &= E(m(x)m(z)) \end{aligned}$$

Due to the assumption of homogeneity (Theorem 1) we can write the covariance function $C_k(x; z) = C_k(x; z)$ as a function in k variables instead of $2 \times k$ variables. We define $r_{xz} = r(x; z) = x; z$, where $r_{xz} \in \mathbb{R}^{2 \times k}$. Furthermore, if we assume that the random function is smooth, the covariance function will be differential everywhere according to Theorems 3 and 4. Hence, it is possible to approximate it by a Taylor's expansion. This approximation can be viewed either as an approximation to the random field (the conditional mean of y_t) or to the log likelihood function. The use of a Taylor's expansion amounts to giving up information about the structure of the model under the alternative hypothesis in order to circumvent the identification problem. This suggests that there is a natural trade-off - between size and power properties - when using a fully known covariance function or using a Taylor's expansion around an unknown class of covariance functions in the

auxiliary regression. This type of argument follows very closely Teräsvirta (1998) where he performs a Taylor approximation of the logistic function in order to deal with the problem of unidentified nuisance parameters in a Lagrange multiplier type test for linearity against the STAR alternative. Since all covariance functions associated with smooth and homogeneous random fields also are smooth and bounded - according to the necessary conditions of Christakos (1984) - we expect low order Taylor approximations to be working very well. As stated earlier, the covariance function should satisfy the two following conditions locally: $C_k(r_{xz}) = C_k(j r_{xz})$, and $jC_k(r_{xz})j \rightarrow C_k(0)$: However, it is easy to show that if $C_k(r_{xz})$ satisfies these two conditions, in general they will not be satisfied by the Taylor approximation of $C_k(r_{xz})$ around $r_{xz0} \in \mathbb{R}^k$. It is possible to show that if we restrict the class of random fields under considerations to include only those where the associated covariance function is a function of $|r_{xz}| \in \mathbb{R}_+^k$; i.e. $C_k(r_{xz}) = C_k(|r_{xz}|)$; then a linearization of $C_k(|r_{xz}|)$ around $|r_{xz0}| \in \mathbb{R}_+^k$ maintains the above properties. Some examples of permissible covariance functions expressed solely in terms of $r_{xz}^a = |r_{xz}|$ are

- 2 The extension of the so-called tent covariance function $x; z \in \mathbb{R}^k; 1 \leq k \leq 3$ given by Example 2
- 2 The simplified extension of the so-called tent covariance function $x; z \in \mathbb{R}^k; k \geq 1$ given by Theorem 5
- 2 The multiplicative Ornstein-Uhlenbeck covariance function, for $x; z \in \mathbb{R}^k; k \geq 1; q = f1; 2g; fa_i \geq 0 _ ig$

$$C_k(r_{xz}^a) = \prod_{i=1}^k \exp(-a_i |x_i - z_i|^q)$$

$$= \exp(-\sum_{i=1}^k a_i |x_i - z_i|^q)$$

- 2 For $x; z \in \mathbb{R}^k; k \geq 1$

$$C_k(r_{xz}^a) = \exp(-a^2 d_{L_2}(x; z)^2)$$

- 2 For $x; z \in \mathbb{R}^k; k \geq 1; b > 0$

$$C_k(r_{xz}^a) = f1 + \exp(-a^2 d_{L_2}(x; z)^2)^{i^b}$$

² The damped oscillation covariance function, for $x; z \in \mathbb{R}^k; k \geq 1; f a_i \geq 0; \omega_i > 0; b > 0; f c_i \geq 0; \omega_i > 0$

$$C_k(r_{xz}^a) = b^a \exp\left(-\sum_{i=1}^k a_i |x_i - z_i|\right) \cos\left(\sum_{i=1}^k c_i |x_i - z_i|\right)$$

In order to simplify notation let $h_{ts}^+ = \frac{1}{2} d_{L_1}(g - x_t; g - x_s) = \frac{1}{2} r_{ts}^{a0} g$ for $g \in \mathbb{R}^k$.⁶ Furthermore, assume that C_k is a symmetric real-valued function of two vectors $x_t; x_s \in \mathbb{R}^k$ such that

$$C_k(x_t; x_s) = C_k(h_{ts}^+)$$

Consider a second order Taylor approximation of $C_k(h_{ts}^+)$ around \bar{h}^+ and denote the approximation of C_k by D_k . Then D_k will be defined as

$$D_k(h_{ts}^+) = C_k(\bar{h}^+) + (h_{ts}^+ - \bar{h}^+) C_k^{\prime}(\bar{h}^+) + \frac{1}{2} (h_{ts}^+ - \bar{h}^+)^2 C_k^{\prime\prime}(\bar{h}^+)$$

The linearized auxiliary regression then becomes

$$\mathbf{b} = \beta_1 D_k(h^+)^v + \beta_2 \mathbf{x}_2 + \mathbf{A}_a$$

or written more explicitly as

$$\mathbf{b}_{ts} = \beta_0 + h_{ts}^+ \beta_1 + (h_{ts}^+)^2 \beta_2 + \beta_3 \mathbf{x}_{2;ts} + v_{a;ts} \quad (25)$$

for $t; s = 1; \dots; T$ where

$$\begin{aligned} \beta_0 &= \beta_1 C_k(\bar{h}^+) + \bar{h}^+ C_k^{\prime}(\bar{h}^+) + \frac{1}{2} (\bar{h}^+)^2 C_k^{\prime\prime}(\bar{h}^+) g \\ \beta_1 &= \beta_1 C_k^{\prime}(\bar{h}^+) + \bar{h}^+ C_k^{\prime\prime}(\bar{h}^+) g \\ \beta_2 &= \beta_1 \frac{1}{2} C_k^{\prime\prime}(\bar{h}^+) g \end{aligned}$$

Notice that in order to compute h_{ts}^+ and to perform the regression given by equation (25) a prior of g must be used. However, as long as h_{ts}^+ is linear in g this can be avoided by substituting h_{ts}^+ with $r_{ts}^{a0} g$ in equation (25). Consequently, g will simply be treated as a part of the parameter space and will be estimated as the minimizer of $\sum_{t,s} v_{a;ts}^2$ by OLS. In the exact test, $\beta_1 = 0$ implies that it is not possible to reject the null hypothesis. This still

⁶ This simplification will not lead to any loss in generality in the following approximations of \mathbf{x}_1 defined by equation (24)

holds when the test is based on the approximated Lagrange multiplier test since $\bar{r}_1 = \bar{0}$, $\bar{r}_0 = \bar{r}_1 = \bar{r}_2 = 0$. By considering an approximation and by using $h_{ts}^+ = \frac{1}{2}r_{ts}^0 g, k(1 + \frac{1}{2}(1+k))$ regressors have been added to the auxiliary regression, and the Lagrange multiplier test now has $(1+k) \times (1 + \frac{1}{2}k)$ degrees of freedom and will be $\hat{A}_{(1+k) \times (1 + \frac{1}{2}k)}^2$ -distributed. We will denote the test statistics based on the Taylor approximation \hat{A}_{OP}^A : To summarize the results, the TR^2 version of the Lagrange multiplier statistics aimed at testing the null hypothesis of linearity against the alternative - where a smooth random function enters the conditional mean - can be obtained by the following procedure

1. Estimate the model under the null and obtain an estimate of the residuals and of the variance of the residuals, i.e.

$$\mathbf{b} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\hat{\sigma}^2 = T^{-1} \mathbf{b}'\mathbf{b}$$

2. Obtain an OLS estimate of $v_{a;ts}$ - denoted $b_{a;ts}$ - based on the auxiliary regression

$$b_{ts} = \bar{r}_0 + \sum_{i=1}^k g_i r_{ts;i}^a + \sum_{i=1}^k \sum_{j=1}^k g_i g_j r_{ts;i}^a r_{ts;j}^a + \bar{r}_3 \mathbf{r}_{2;ts} + v_{a;ts}$$

$$r_{ts;i}^a = \frac{1}{2} d_{L_1}(X_{it}; X_{is})$$

$$\mathbf{x}_t = (X_{1t}; X_{2t}; \dots; X_{kt})'$$

$$t;s = 1; 2; \dots; T$$

where $\mathbf{b} = (\mathbf{I}_T \otimes \frac{\mathbf{e}\mathbf{e}'}{2})^v$ and the regressor \mathbf{r}_2 is defined as previously. Notice, that b_{ts} and $\mathbf{r}_{2;ts}$ refer to the $(t \times s)$ th entry/row in \mathbf{b} and \mathbf{r}_2 respectively,

3. Obtain the uncentered R^2 from this regression as

$$R^2 = \frac{SSR_0 - SSR_1}{SSR_0}$$

$$SSR_0 = \sum_{t,s} \mathbf{X}_t' \mathbf{X}_s \mathbf{b}_{ts}^2$$

$$SSR_1 = \sum_{t,s} \mathbf{X}_t' \mathbf{X}_s \mathbf{b}_{a;ts}^2$$

4. Calculate the Lagrange multiplier statistics as

$$\hat{A}_{OP} = T^2 R^2 \gg \hat{A}^2((1+k) \alpha (1 + \frac{1}{2}k))$$

4.4 g-tests based on unknown structure functions

In this section, we restrict attention to the class of smooth random fields that is at least two times differentiable in g . We derive the g-tests for neglected nonlinearity based on the structure functions associated with homogeneous and Gaussian random field that were introduced in section 3.2, i.e. with C_k replaced by \mathfrak{C}_k in the log likelihood function. As for the χ^2 -tests we get around the identification problems by using an approximation of the structure function. The class of smooth random function under consideration allow us to do so. In particular, we let \mathfrak{C}_k be a symmetric real-valued function of two vectors $x_t; x_s \in \mathbb{R}^k$ such that

$$\mathfrak{C}_k(x_t; x_s) = \mathfrak{C}_k(d(x_t; x_s; g))$$

where $d(x; z; \mu)$ is a general parameterized real-valued function that maps "the distance" between two points x and z in \mathbb{R}^k onto \mathbb{R}^1 : For convenience we write $d_{ts} \equiv d(x_t; x_s; g)$: Consider a second order Taylor's approximation of $\mathfrak{C}_k(d_{ts})$ around \bar{d} : The second order expansion denoted $J_k(d_{ts})$ writes

$$J_k(d_{ts}) = \mathfrak{C}_k(\bar{d}) + (d_{ts} - \bar{d}) \mathfrak{C}_k'(\bar{d}) + \frac{1}{2} (d_{ts} - \bar{d})^2 \mathfrak{C}_k''(\bar{d})$$

For simplicity let us define d_{ts} as

$$d_{ts} = \frac{1}{2} [\|x_t - x_s\| - \|x_t\| - \|x_s\|] g$$

we get

$$\frac{\partial J_k(d_{ts})}{\partial g} = \frac{1}{2} [\|x_t - x_s\| - \|x_t\| - \|x_s\|] \mathfrak{C}_k'(\bar{d}) + (d_{ts} - \bar{d}) \mathfrak{C}_k''(\bar{d}) g$$

Naturally, a variety of alternative specifications of d_{ts} are possible. The only restriction in this setup is that d_{ts} must be linear in g . However, this condition is not very restrictive since we can always obtain it by approximation. Notice that since $d(x_t; x_s; 0) = 0$ for all $x_t; x_s$ we obtain

$$\frac{\partial J_k(d_{ts})}{\partial g} \Big|_{g=0} = \frac{1}{2} [\|x_t - x_s\| - \|x_t\| - \|x_s\|] \mathfrak{C}_k'(\bar{d}) + \bar{d} \mathfrak{C}_k''(\bar{d}) g$$

Furthermore, since our primary interest is in properties of $J_k(d_{ts})$ in a close neighborhood of $g = 0$ it would be natural to also set $\bar{d} = 0$ implying that

$$\frac{\partial J_k(d_{ts})}{\partial g^0} \Big|_{g=0} = \frac{1}{2} [j_{xt} \quad x_{sj} \quad j_{xtj} \quad j_{x_{sj}}] \begin{matrix} (1 \times k) \\ (1 \times 1) \end{matrix} e_k^0(0)$$

where

$$\frac{\partial J_k(d_{ts})}{\partial g^0} \Big|_{g=0} = \begin{matrix} 2 \\ \frac{\partial J_k(d_{ts})}{\partial g_1} \Big|_{g=0}; \end{matrix} \begin{matrix} 3_0 \\ \frac{\partial J_k(d_{ts})}{\partial g_2} \Big|_{g=0}; \end{matrix} \dots; \begin{matrix} 3_0 \\ \frac{\partial J_k(d_{ts})}{\partial g_k} \Big|_{g=0} \end{matrix}$$

Let $\frac{\partial J_k(d_{ts})}{\partial g_i} \Big|_{g=0}$ be the $\langle t; s \rangle$ entry in $\frac{\partial J_k}{\partial g_i} \Big|_{g=0}$: Since $\frac{\partial J_k}{\partial g_i} \Big|_{g=0} = \frac{1}{2} J_k + \frac{3}{4} I_T$; and by keeping $\frac{\partial J_k}{\partial g_i} \Big|_{g=0}$ fixed, the score functions are

$$S(g_i) \Big|_{g=0} = \frac{1}{2} \text{tr} \left(\frac{\partial J_k}{\partial g_i} \Big|_{g=0} (I_T - \frac{1}{2} I_T) \right); \quad i = 1; 2; \dots; k$$

$$S(\frac{3}{4}) \Big|_{g=0} = \frac{1}{2} \text{tr} \left(\frac{\partial J_k}{\partial g_i} \Big|_{g=0} (I_T - \frac{1}{2} I_T) \right)$$

Evaluating the score function under the null of linearity ($H_0 : g = 0$), we have

$$S(g_i) \Big|_{g=0} = \frac{1}{2} \text{tr} \left(\frac{\partial J_k}{\partial g_i} \Big|_{g=0} (I_T - \frac{1}{2} I_T) \right); \quad i = 1; 2; \dots; k$$

$$S(\frac{3}{4}) \Big|_{g=0} = \frac{1}{2} \text{tr} \left(\frac{\partial J_k}{\partial g_i} \Big|_{g=0} (I_T - \frac{1}{2} I_T) \right)$$

Notice that $\frac{\partial J_k}{\partial g_i} \Big|_{g=0} = J_k(0) = 0$ by construction in order to avoid the non-ergodicity problem, implying that $\frac{\partial J_k}{\partial g_i} \Big|_{g=0} = \frac{3}{4} I_T$:

As in the previous section, we can write the score functions in the form

$$S(g_i) \Big|_{g=0} = \frac{1}{2} \text{tr} \left(\frac{\partial J_k}{\partial g_i} \Big|_{g=0} u \right); \quad i = 1; 2; \dots; k \quad (26)$$

$$S(\frac{3}{4}) \Big|_{g=0} = \frac{1}{2} \text{tr} \left(\frac{\partial J_k}{\partial g_i} \Big|_{g=0} u \right) \quad (27)$$

where

$$\begin{aligned} \mathbf{x}_i &= \left(\frac{\partial J_k}{\partial g_i} \Big|_{g=0} \right)^v; \quad i = 1; 2; \dots; k \\ \mathbf{x}_{k+1} &= \mathbf{1}_T^v \\ \mathbf{u} &= \left(\mathbf{1}_T \quad i \quad \frac{22^0}{34^2} \right)^v \end{aligned}$$

The scores (26) and (27) allow us to compute the g-tests as TR^2 statistics, which we will denote g_{OP}^A . The construction of the test statistics follows the procedures already outlined implying that step 2 should be replaced by

- 2.a Obtain an OLS estimate of $\mathbf{e}_{a;ts}$ - denoted $\mathbf{b}_{a;ts}$ - based on the auxiliary regression

$$\begin{aligned} \mathbf{b}_{ts} &= \mathbf{X} \mathbf{e}_{i;ts;i} + \mathbf{e}_{k+1} \mathbf{x}_{k+1;ts} + \mathbf{e}_{a;ts} \\ \mathbf{e}_{ts;i} &= d(\mathbf{x}_{it}; \mathbf{x}_{is}; 1) \\ \mathbf{x}_t &= \mathbf{f}(\mathbf{x}_{1t}; \mathbf{x}_{2t}; \dots; \mathbf{x}_{kt}) \\ t; s &= 1; 2; \dots; T \end{aligned}$$

where $\mathbf{b} = \left(\mathbf{1}_T \quad i \quad \frac{22^0}{34^2} \right)^v$ and the regressor \mathbf{x}_{k+1} are defined as previously. Again, \mathbf{b}_{ts} and $\mathbf{x}_{k+1;ts}$ refer to the $(t \times s)$ th entry/row in \mathbf{b} and \mathbf{x}_{k+1} respectively,

The test statistics is computed by using $\mathbf{b}_{a;ts}$ instead of $\mathbf{b}_{a;ts}$ in the calculation of SSR_1 : Otherwise, the computation of the statistics will be identical to the procedure outlined above, i.e. step 4 becomes

- 4.a Calculate the Lagrange multiplier statistics as

$$g_{OP}^A = T^2 R^2 \gg \hat{A}^2(k) \quad (28)$$

Notice that the unidentified nuisance parameter γ_j is included and is proportional to the $\gamma_j; j = 1; 2; \dots; k + 1$ parameters in the auxiliary regression. Hence, no prior specification of the value of γ_j is needed in order to compute g_{OP}^A : In order to secure power under the alternative we suggest to include the term $\sum_{i=1}^k \sum_{j=1}^k \mathbf{e}_{ij} \mathbf{e}_{ts;i} \mathbf{e}_{ts;j}$ in the auxiliary regression although $\mathbf{e}_{ij}; 8i; j$ equals zero under the null given by $H_0 : g_1 = g_2 = \dots = g_k = 0$: In that case the augmented auxiliary regression becomes

$$b_{ts} = \sum_{i=1}^k e_{i,ts}^a + \sum_{i=1}^k \sum_{j=i}^k e_{ij,ts}^a e_{ts,j}^a + e_{k+1,ts}^a + e_{a,ts}$$

and the Lagrange multiplier test statistics given by equation (28) will be $\hat{A}^2((1+k) \times (1 + \frac{1}{2}k) - 1)$ distributed.

5 Simulation studies on size and power properties

In this section, we perform a Monte Carlo simulation study to analyze the small sample size and power properties of the various statistics that we propose. The setup of the Monte Carlo experiment follows closely Dahl (1999).

5.1 The design of the Monte Carlo experiment

We examine the size and power properties of the tests for neglected nonlinearity by considering three blocks of linear and nonlinear dynamic models. All the chosen models have been used in previous studies dealing with testing of linearity. The use of these “benchmark” models allows us to make comparisons with earlier studies. The models included in Table 1 were originally used by Lee et al. (1993). The models of Table 2 have been more extensively studied by Keenan (1985), Tsay(1986), Ashley, Patterson and Hinich (1986), Chan and Tong (1986), and Lee et al. (1993). Finally, the models in Table 3 have been studied by Teräsvirta et al. (1993).

The five models contained in Table 1 are all characterized by being simple dynamic univariate models, where the dynamic part is represented by one lag of the endogenous variable only. The models are all stationary. The models included are the autoregressive model (AR), the bilinear model (BL) of Granger and Anderson (1978), the threshold autoregressive model (TAR) of Tong (1983), the sign autoregressive model (SGN), and the nonlinear autoregressive model (NAR). The exact parameterization of the models is given in Table 1. We also consider two bivariate representations where we do not impose any dynamic structure. We consider a squared relation, which we denote SQ, and an exponential relation, denoted EXP. The parameterization of the bivariate models is also shown in Table 1.

The models in Table 2 are characterized by having a much richer dynamic structure compared to those in Table 1. Model1 is an MA(2) representation

Table 1: Block 1 models and bivariate models

AR	$y_t = 0.6y_{t-1} + \varepsilon_t$
BL	$y_t = 0.7y_{t-1} + 0.2y_{t-2} + \varepsilon_t$
TAR	$y_t = 0.9y_{t-1}I_{(y_{t-1} < 1)} + 0.3y_{t-1}I_{(y_{t-1} > 1)} + \varepsilon_t$
SGN	$y_t = \theta_{(y_{t-1} > 1)} + \theta_{(y_{t-1} < 1)} + \varepsilon_t$
NAR	$y_t = (0.7 + 0.3I_{(y_{t-1} > 1)})y_{t-1} + \varepsilon_t$
SQ	$y_t = x_t^2 + e_t$ $x_t = 0.6x_{t-1} + \varepsilon_t$ $e_t \gg N(0, \frac{3}{4}); \frac{3}{4} = 1; 25; 400$
EXP	$y_t = \exp(x_t) + e_t$ $x_t = 0.6x_{t-1} + \varepsilon_t$ $e_t \gg N(0, \frac{3}{4}); \frac{3}{4} = 1; 25; 400$

and Model2 is a heteroskedastic MA(2). These two models together with Model4 - an AR(2) model - are all linear models. They are included primarily to evaluate the nominal size of the nonlinearity tests and their ability to distinguish between dynamic misspecification and misspecification due to nonlinearity in the conditional mean. Model3, Model5 and Model6 are the truly nonlinear models in Table 2. Model3 is a nonlinear MA(2). Model5 and Model6 belong to the family of bilinear models. Model5 is a bilinear autoregressive model, while Model6 is a bilinear autoregressive moving average model.

The first model in Table 3 is the logistic smooth transition autoregressive model (LSTAR). Its properties are discussed in detail in Teräsvirta (1990). The second model is a special case of the exponential smooth transition autoregressive model (ESTAR) known as the exponential autoregressive model of Haggan and Ozaki (1981). The NN and BN models are univariate and bivariate neural network models, respectively.

Throughout $\varepsilon_t \gg N(0; 1)$ is a white noise series. In Table 1, the information set for the univariate models and bivariate models is $\{y_{t-1}\}$ and $\{x_t, y_{t-1}\}$, respectively. The information set for the models contained in the Table 2 and Table 3 - except the BN model - equals $\{y_{t-1}; y_{t-2}\}$. For the BN model

Table 2: Block 2 models

Model1 $y_t = \alpha_1 + 0.4y_{t-1} + 0.3y_{t-2}$

Model2 $y_t = \alpha_1 + 0.4y_{t-1} + 0.3y_{t-2} + 0.5y_{t-2}^2$

Model3 $y_t = \alpha_1 + 0.3y_{t-1} + 0.2y_{t-2} + 0.4y_{t-1}^2 + 0.25y_{t-2}^2$

Model4 $y_t = 0.4y_{t-1} + 0.3y_{t-2} + \alpha_1$

Model5 $y_t = 0.4y_{t-1} + 0.3y_{t-2} + 0.5y_{t-1}^2 + \alpha_1$

Model6 $y_t = 0.4y_{t-1} + 0.3y_{t-2} + 0.5y_{t-1}^2 + 0.8y_{t-1} + \alpha_1$

the information set contains $f_{y_{t-1}}; x_t g$:

Table 3: Block 3 models

LSTAR $y_t = 1.8y_{t-1} + 1.06y_{t-2} + (0.02 + 0.9y_{t-1} + 0.795y_{t-2})F(y_{t-1}) + v_t$
 $F(y_{t-1}) = [1 + \exp(-100(y_{t-1} - 0.02))]^{-1}$
 $v_t \gg N(0; \frac{3}{4}^2); \frac{3}{4}^2 = 0.02^2$

ESTAR $y_t = 1.8y_{t-1} + 1.06y_{t-2} + (0.9y_{t-1} + 0.795y_{t-2})F(y_{t-1}) + v_t$
 $F(y_{t-1}) = 1 - \exp(-400(y_{t-1})^2)$
 $v_t \gg N(0; \frac{3}{4}^2); \frac{3}{4}^2 = 0.01^2$

NN $y_t = \alpha_1 + [1 + \exp(-100(y_{t-1} - 0.8y_{t-2}))]^{-1} + [1 + \exp(-100(y_{t-1} + 0.8y_{t-2}))]^{-1} + v_t$
 $v_t \gg N(0; \frac{3}{4}^2); \frac{3}{4}^2 = 0.05^2$

BN $y_t = \alpha_1 + [1 + \exp(-100(y_{t-1} - x_t))]^{-1} + [1 + \exp(-100(y_{t-1} + x_t))]^{-1} + v_t$
 $x_t = 0.8x_{t-1} + u_t$
 $v_t \gg N(0; \frac{3}{4}^2); \frac{3}{4}^2 = 0.05^2; u_t \gg N(0; \frac{3}{4}_u^2); \frac{3}{4}_u^2 = 0.05^2$

5.2 Results on size and power properties

The results from the simulation study on the size properties are reported in Table 4. Let $\hat{L}_H(g)$ denote the Hamilton's Lagrange Multiplier statistic based on the Hessian representation of the information matrix and on the

spherical variance-covariance matrix, and $\lambda_{OP}^E(g)$ the Lagrange multiplier test based on the variance-covariance matrix proposed in Theorem 5. λ_{OP}^A ; and g_{OP}^A are defined in the previous section.

In the case of two regressors - Table 4 - the λ_{OP}^A , $\lambda_{OH}^E(g)$ and g_{OP}^A tests have the best size-properties, with simulated critical values very close to the nominal for samples of size 200. The test $\lambda_{OP}^E(g)$ seems to have a smaller simulated size than the nominal size, rejecting the null hypothesis too often.

In order to make a fair power comparison across the various tests, we use the simulated critical values reported in Table 4 instead of the asymptotic critical values.

In addition to the tests already introduced, we also consider the Tsay's test. The reason is that the Tsay's test statistics, like the λ_{OP}^A and g_{OP}^A statistics, is free of unidentified nuisance parameters under the null, and is derived, based on a second order Taylor approximation of an unknown - but in contrast to the statistics suggested in this paper - deterministic functional form. Furthermore, like the λ_{OP}^A and g_{OP}^A statistics, the TSAY statistic has a TR^2 form and is computed by using an auxiliary regression very similar to the one outlined above. Due to the simplicity of the TSAY test and its relatively good power properties, it is included in most statistical computer packages. λ_{OP}^A and g_{OP}^A are statistics just as simple to obtain computationally.

Table 4: Critical values (5%) based on AR(2) model. The first number equals the simulated critical value. The number in parantheses in the second row is the asymptotic critical value. The number in brackets denotes the "asymptotic" size of the statistics when based on the simulated critical values (equals the area under the asymptotic distribution to the right of the simulated critical value). The results are based on 1000 replications

Test	T=50	T=100	T=200
	5%	5%	5%
$\hat{s}_H^E(g)$	3.47 (3.84) [0.062]	3.41 (3.84) [0.065]	3.58 (3.84) [0.058]
$\hat{s}_{OP}^E(g)$	6.17 (3.84) [0.013]	5.92 (3.84) [0.015]	5.32 (3.84) [0.021]
\hat{s}_{OP}^A	14.86 (12.59) [0.021]	12.94 (12.59) [0.044]	12.13 (12.59) [0.059]
g_{OP}^A	8.58 (12.59) [0.127]	9.23 (12.59) [0.100]	10.57 (12.59) [0.060]

In Table 5, we observe that all tests are extremely powerful when the non-linearity does not included lagged endogenous variables, such as the SQ and EXP models. For bilinear models, the g_{OP}^A test is the most powerful; for TAR and SGN models, the \hat{s}_H^E and \hat{s}_{OP}^A tests are the best performers; and for NAR models the power of the Tsay test is comparable to the \hat{s}_{OP}^A test. In Table 6, where the models have higher dynamics, the g_{OP}^A and the Tsay tests are the best competitors overall. We notice that both tests are relative good statistics to detect conditional heteroscedasticity (Model 2). They have identical performance when nonlinear MA(2) models are considered, and, together with the \hat{s}_{OP}^A and \hat{s}_H^E , they are able to detect bilinearity in AR and MA models. In Table 7, the Tsay and \hat{s}_{OP}^A tests are the most powerful when the model is LSTAR; when we consider ESTAR and NN models, there is a superior advantage of the \hat{s}_i tests over the g-tests and the Tsay test.

Table 5: Power vs. sample size. Power using 5% simulated critical values. Replications = 1000, sample size = 50, 100, 200. The number in parentheses equals the rejection frequency when the asymptotic critical value is used instead of the simulated

Test	AR	BL	TAR	SGN	NAR	SQ	EXP
$\hat{\rho}_H^E(g)$							
T=50	5.2	12.7	63.3	75.5	8.4	100.0	96.6
T=100	4.3	19.4	93.1	98.1	11.9	100.0	99.8
T=200	3.8	24.4	99.8	100.0	22.4	100.0	100.0
$\hat{\rho}_{OP}^E(g)$							
T=50	5.0 (5.1)	3.5	34.8	43.4	4.8	99.8	91.9
T=100	5.0 (3.6)	5.1	81.4	90.7	6.4	100.0	99.9
T=200	5.0 (3.4)	8.9	99.6	100.0	9.7	100.0	100.0
$\hat{\rho}_{OP}^A$							
T=50	5.0 (4.2)	19.4	50.6	67.1	8.6	100.0	98.6
T=100	5.0 (3.6)	31.0	89.6	97.0	16.0	100.0	100.0
T=200	5.0 (3.5)	42.6	99.8	100.0	26.1	100.0	100.0
g_{OP}^A							
T=50	5.0 (0.6)	37.9	4.6	6.4	8.4	100.0	96.7
T=100	5.0 (0.8)	57.1	15.7	32.2	11.9	100.0	100.0
T=200	5.0 (1.5)	76.5	68.4	95.1	14.6	100.0	100.0
TSAY							
T=50	5.0 (6.1)	23.9	9.5	19.6	12.6	100.0	98.8
T=100	5.0 (4.7)	33.7	6.1	17.8	15.4	100.0	99.9
T=200	5.0 (5.1)	40.8	5.8	17.4	21.6	100.0	100.0

Overall, three clear results emerge. First, the g_{OP}^A tests seem a promising alternative to existing tests when the alternative could be a bilinear type model, mainly when most tests for linearity have difficulties identifying nonlinearities of this type, e.g. Lee et al. (1993). Second, the $\hat{\rho}_{OP}^A$ tests, which do not require the explicit knowledge of the functional form of the covariance matrix, emerge as a powerful alternative across models. Third, the $\hat{\rho}$ -tests are superior on detecting neural network specifications. In addition, the $\hat{\rho}$ and g -tests match the simplicity of the popular Tsay statistic.

Table 6: Power vs. sample size. Power using 5% simulated critical values. Replications = 1000, sample size = 50, 100, 200. The number in parentheses equals the rejection frequency when the asymptotic critical value is used instead of the simulated

Test	Model1	Model2	Model3	Model4	Model5	Model6
$\beta_{H}^E(g)$						
T=50	5.2	7.0	25.2	4.8	51.8	37.0
T=100	5.2	8.4	53.7	5.5	91.7	74.9
T=200	3.8	13.0	87.4	3.9	100.0	98.9
$\beta_{OP}^E(g)$						
T=50	5.4	4.5	6.5	5.0 (18.5)	14.0	5.3
T=100	6.2	4.1	18.1	5.0 (15.5)	56.3	28.2
T=200	5.3	4.1	50.0	5.0 (15.8)	97.1	79.4
β_{OP}^A						
T=50	5.5	7.4	46.0	5.0 (6.1)	59.5	30.9
T=100	5.8	11.0	80.6	5.0 (5.5)	94.7	63.7
T=200	5.5	15.4	99.2	5.0 (6.4)	100.0	90.4
g_{OP}^A						
T=50	7.3	13.2	57.8	5.0 (3.0)	47.3	44.8
T=100	6.2	16.7	85.1	5.0 (3.7)	86.4	64.8
T=200	5.8	19.2	99.5	5.0 (4.7)	99.7	86.5
TSAY						
T=50	5.1	9.8	53.3	5.0 (5.0)	76.2	56.9
T=100	7.0	15.9	85.5	5.0 (4.9)	95.5	77.8
T=200	5.9	19.0	99.3	5.0 (5.5)	98.7	90.8

Table 7: Power vs. sample size. Power using 5% simulated critical values. Replications = 1000, sample size = 50, 100, 200.

	LSTAR	ESTAR	NN	BN
		$\hat{g}_H^E(g)$		
T=50	33.8	20.7	61.3	61.0
T=100	71.0	47.2	96.6	96.5
T=200	98.7	82.2	100.0	100.0
		$\hat{g}_{OP}^E(g)$		
T=50	4.7	5.3	30.1	29.1
T=100	30.4	21.0	87.1	85.1
T=200	87.9	61.8	100.0	100.0
		\hat{g}_{OP}^A		
T=50	44.7	14.6	23.3	28.1
T=100	78.9	28.8	65.3	67.8
T=200	98.6	56.3	97.1	97.9
		g_{OP}^A		
T=50	25.9	8.7	9.1	17.1
T=100	51.0	14.6	15.0	37.2
T=200	87.5	27.8	26.7	71.8
		TSAY		
T=50	65.0	9.4	9.6	17.6
T=100	93.8	9.7	11.5	15.9
T=200	100.0	8.9	13.6	15.4

6 Conclusion

In this paper, we have offered a battery of tests for neglected nonlinearities within the framework of a flexible regression model proposed by Hamilton (2000). We have classified the tests in two categories: \hat{g} -tests and g -tests: The common denominator to both is that they can be characterized as TR^2 versions of the Lagrange multiplier statistic. There are mainly two advantages of our tests compared to Hamilton's statistic and other more popular tests for nonlinearities, such as the Neural Network test and the Tsay's test.

First, the proposed tests are free of unidentified nuisance parameters under the null, contrary to Hamilton's and the Neural Network test, where the number of unidentified nuisance parameters under the null increases with the dimensionality of the model. Furthermore, we propose tests for which the exact parameterization of the variance-covariance function is not needed. Secondly, the computation of our tests is extremely simple compared to the statistic proposed by Hamilton, since they are based on auxiliary regressions.

Simple tests that are free of unidentified nuisance parameters are the result of 1) studying alternative representations of the random field governing the outcome of the conditional mean in Hamilton's regression model, and 2) using local approximations to the variance-covariance function based on Taylor's expansion.

We have considered switching from the Euclidean distance measure to the Minkowski distance, and we have measured the random field in deviations from a random field centered at zero. The payoff of this strategy has been confirmed in an extensive Monte Carlo study. The simulation evidence suggests that for samples larger than 50, the test statistic based on the random field defined on the Minkowski distance broadly matches the power properties of the more complicated statistic suggested by Hamilton (2000). The test matches the simplicity of the popular Tsay's statistic and has superior power properties when the true model is TAR, SGN, ESTAR, NN and BN, compared to those of Tsay's test.

The theory of random fields suggests that if we choose to use the exact and parameterized non-differentiable covariance function as suggested by Hamilton (2000) - or some other non-differentiable covariance function - it will enable us to construct a test statistic that will have asymptotic power against a certain class of differentiable as well as non-differentiable random functions. The weakness of this approach is how to choose a proper covariance function as well as fixing properly the unidentified parameter values properly under the null hypothesis of linearity. Alternatively, if we choose to work with a second order Taylor approximation of the covariance function - implicitly assuming that it is second order mean square differentiable - then the test statistic will have mainly asymptotic power in the direction of detecting neglected nonlinear random functions that are differentiable in mean square. This suggests that we should consider the tests developed in this paper complementary rather than competing alternatives to the test procedure suggested by Hamilton, since the statistics will potentially have power in different directions.

References

- Adler, R. J. (1981), *The Geometry of Random Fields*, Wiley, New York.
- Ashley, R., Patterson, D. & Hinich, M. (1986), 'A diagnostic test for non-linear serial dependence in time series ...tting errors', *Journal of Time Series Analysis* 7, 165–178.
- Chan, W. S. & Tong, H. (1986), 'On tests for nonlinearity in time series models', *Journal of Forecasting* 5, 217–228.
- Christakos, G. (1984), 'On the problem of permissible covariance and variogram models', *Water Resources Research* 20(2), 251–265.
- Christakos, G. (1992), *Random Field Models in Earth Sciences*, Academic Press Inc., San Diego, CA.
- Dahl, C. M. (1999), 'An investigation of tests for linearity and the accuracy of flexible nonlinear inference'. Unpublished manuscript, Department of Economics, University of Aarhus.
- Davies, R. B. (1977), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* 64, 247–254.
- Godfrey, L. (1988), *Misspecification Tests in Econometrics. The Lagrange Multiplier Principle and Other Approaches*, Cambridge University Press, New York, NY.
- Granger, C. W. J. & Andersen, A. (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoech und Ruprecht, Gottingen.
- Haggan, V. & Ozaki, T. (1981), 'Modelling nonlinear random vibrations using an amplitude dependent autoregressive time series model', *Biometrika* 68, 189–196.
- Hamilton, J. D. (2000), 'A parametric approach to flexible nonlinear inference', *Econometrica* (forthcoming) .
- Hansen, B. E. (1996b), 'Inference when a nuisance parameter is not identified under the null hypothesis', *Econometrica* 64, 413–430.
- Keenan, D. M. (1985), 'A tukey nonadditivity test for time series nonlinearity', *Biometrika* 72, 39–44.

- Khinchin, A. Y. (1934), 'Korrelationstheorie der stationären stochastischen prozesse', *Math. Ann.* 109, 604.
- Kolmogorov, A. N. (1941a), 'Dissipation of energy in locally isotropic turbulence', *Dokl. Akad. Nauk SSSR* .
- Kolmogorov, A. N. (1941b), 'The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers', *Dokl. Akad. Nauk SSSR* .
- Lee, T.-H., White, H. & Granger, C. W. J. (1993), 'Testing for neglected nonlinearity in time series models', *Journal of Econometrics* 56, 269–290.
- Luukkonen, R., Saikkonen, P. & Terasvirta, T. (1988), 'Testing linearity against smooth transition autoregressive models', *Biometrika* .
- Matheron, G. (1973), 'The intrinsic random functions and their applications', *Advances in applied probability* 5, 439–468.
- Ramsey, J. B. (1969), 'Tests for specification errors in classical linear least square regression analysis', *Journal of the Royal Statistical Society Series B*, 31, 350–371.
- Terasvirta, T. (1990), 'Specification, estimation and evaluation of smooth transition autoregressive models'. Discussion paper 90-39, Department of Economics, University of California, San Diego.
- Terasvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in A. Ullah & D. E. A. Giles, eds, 'Handbook of Applied Economic Statistics', Dekker.
- Terasvirta, T., Lin, C. F. & Granger, C. W. J. (1993), 'Power of the neural network linearity test', *Journal of Time Series Analysis* 14, 209–220.
- Tong, H. (1983), *Threshold Models in Nonlinear Time Series Analysis*, Springer-Verlag, New York, NY.
- Tsay, R. S. (1986), 'Nonlinearity test for time series', *Biometrika* 73, 461–466.
- White, H. (1989), An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks, in 'Proceedings of the international joint conference on neural networks', IEEE Press, New York, NY, Washington, DC, pp. 451–455.

Yaglom, A. M. (1962), An Introduction to the Theory of Stationary Random Functions, Prentice-Hall, Englewood Cliffs, N.J.

7 Appendix

Lemma 1 Let r and h be scalars such that $r \geq h \geq 0$: Define

$$G_{k-1}(h; r) = \int_h^r (r - |z|)^{k-1} dz$$

then

$$G_{k-1}(h; r) = (r - h) \frac{(k-1)}{k} G_{k-2}(h; r) \quad (29)$$

with

$$G_0(h; r) = r - h$$

Proof. Solving the integral: Case a) $k = 1$

$$\begin{aligned} G_0(h; r) &= \int_h^r (r - |z|)^0 dz \\ &= [z]_h^r = r - h \end{aligned}$$

since the bounds of integration are always non-negative implying that $\int_h^r (r - |z|)^{k-1} dz = \int_h^r (r - z)^{k-1} dz$: Case b) $k = 2$

$$\begin{aligned} G_1(h; r) &= \int_h^r (r - |z|)^1 dz \\ &= \int_h^r (r - z) dz \\ &= \left[r z - \frac{1}{2} z^2 \right]_h^r \\ &= \frac{1}{2} (r - h)^2 \end{aligned}$$

Case c) $k = 3$

$$\begin{aligned} G_2(h; r) &= \int_h^r (r - |z|)^2 dz \\ &= \int_h^r (r - z)^2 dz \\ &= \left[r z - \frac{1}{2} z^2 + \frac{1}{3} z^3 \right]_h^r \\ &= \frac{1}{3} (r - h)^3 \end{aligned}$$

In general

$$G_{k-1}(h; r) = \frac{1}{k} (r - h)^k$$

so the result (29) follows.

■

Lemma 2

$$G_{k_i-1}(0; r) = \frac{r^k}{k} = r^k G_{k_i-1}(0; 1)$$

Lemma 3 Let $x = (x_1; \dots; x_k)^0$ $2 < k$ and define the k_i dimensional orthogon of radius r

$$A_k(r) = \{x \mid \sum_{i=1}^k x_i^2 \leq r^2\}$$

Let $V_k(r)$ denote the volume of $A_k(r)$: Then

$$V_k(r) = r^k V_k(1) \tag{30}$$

where

$$V_1(1) = 2 \tag{31}$$

$$V_k(1) = 2 V_{k_i-1}(1) G_{k_i-1}(0; 1) \tag{32}$$

Proof. Case a) $k = 1$: (length of line)

$$V_1(r) = 2r$$

$$V_1(1) = 2$$

proving (31). Case b) $k = 2$: (area of a square)

$$V_2(r) = 2r^2$$

$$V_2(1) = 2$$

Case c) $k = 3$: (volume of octahedron)

$$V_3(r) = \frac{4}{3} r^3$$

$$V_3(1) = \frac{4}{3}$$

by induction

$$V_k(r) = r^k V_k(1)$$

proving (30): For any x_k^a such that $\sum_{j=1}^k x_j^a < r$ the values x such that $x \in A_k(r)$ and $x_k = x_k^a$ can be written as

$$\begin{aligned} A_k(r) &= \{x \mid \sum_{j=1}^k x_j^2 \leq r^2 \text{ and } x_k = x_k^a\} \\ &= \{x \mid \sum_{j=1}^{k-1} x_j^2 \leq r^2 - x_k^a{}^2 \text{ and } x_k = x_k^a\} \end{aligned}$$

Given x_k^a this is a $(k-1)$ dimensional polyhedron with a circumscribed spheroid of radius $r - jx_k^a$; denoted $A_{k-1}(r - jx_k^a)$: The volume of this polyhedron in the interval $[x_k^a; x_k^a + dx_k]$ is

$$V_{k-1}(r - jx_k^a)dx_k^a$$

so the volume of $A_k(r)$ over $x_k^a \in [r; 0]$ is

$$V_k(r) = 2 \int_0^r V_{k-1}(r - jx_k^a)dx_k$$

from (30)

$$V_{k-1}(r - jx_k^a) = (r - jx_k^a)^{k-1}V_{k-1}(1)$$

and substituting back

$$\begin{aligned} V_k(r) &= 2V_{k-1}(1) \int_0^r (r - jx_k^a)^{k-1}dx_k \\ &= 2V_{k-1}(1)G_{k-1}(0; r) \\ &= 2V_{k-1}(1)r^k G_{k-1}(0; 1) \end{aligned}$$

proving (32).
■

Lemma 4 Consider

$$\begin{aligned} A_k(1) &= \{x_j \mid jx_1 + jx_2 + \dots + jx_k \leq 1\} \\ B_k(1) &= \{x_j \mid jx_1 + jx_2 + \dots + jx_k \leq 2h\} \end{aligned}$$

so $A_k(1)$ is a octahedron centered at $(0; \dots; 0)$ with a circumscribed spheroid of radius 1. $B_k(1)$ is a octahedron centered at $(0; \dots; 2h)$ with a circumscribed spheroid of radius 1. Consider

$$J_k(h) = A_k(1) \setminus B_k(1)$$

the volume of $J_k(h)$ is

$$\begin{aligned} &\int_{x \in J_k(h)} \\ &= 2V_{k-1}(1)G_{k-1}(h; 1) \end{aligned}$$

Proof. The proof follows Hamilton (2000) but with $G_{k_i-1}(h; 1)$ defined as

$$G_{k_i-1}(h; 1) = \frac{(1 - h)^k}{k}$$

Then

$$\begin{aligned} \int_{\mathbf{z} \in \mathcal{Z}_k(h)} \int_{\mathbf{z} \in \mathcal{Z}_k(1)} V_{k_i-1}(1 - h) \prod_{j=1}^k dx_k \\ &= 2V_{k_i-1}(1) \int_{\mathbf{z} \in \mathcal{Z}_k(1)} (1 - h)^{k_i-1} dx_k \\ &= 2V_{k_i-1}(1) G_{k_i-1}(h; 1) \end{aligned}$$

■

Theorem 5 Let $\mathbf{x} \in \mathcal{Z}_k$ and $\mathbf{z} \in \mathcal{Z}_k$ and let $m^a(\mathbf{x})$ and $m^a(\mathbf{z})$ be random fields generated as the limit of equation (17) as $\Phi \rightarrow 0$ evaluated at fixed points $\mathbf{x} = (x_1, x_2, \dots, x_k)$ and $\mathbf{z} = (z_1, z_2, \dots, z_k)$ where $h^+ = \frac{1}{2}d_{L_1}(\mathbf{x}; \mathbf{z})$. Then

$$E(m^a(\mathbf{x})m^a(\mathbf{z})) = C_k^a(h^+)$$

where

$$C_k^a(h^+) = \begin{cases} \frac{1}{2} (1 - h^+)^k & \text{if } h^+ \leq 1 \\ 0 & \text{if } h^+ > 1 \end{cases}$$

implying that $C_k^a(h^+)$ is a permissible covariance function for any value of $h^+ \in [0, 1]$.

Proof. The covariance of $m^a(\mathbf{x})$ and $m^a(\mathbf{z})$ is given by the ratio of the volume of the overlap of two k -dimensional orthogons with a circumscribed unit spheroid centered at $(0; 0; \dots; 0)$ and $(0; 0; \dots; 2h^+)$, given by

$$2V_{k_i-1}(1)G_{k_i-1}(h^+; 1)$$

- as shown in Lemma (4) - to the volume of a single unit k -dimensional orthogon $V_k(1)$. The normalization with respect to $V_k(1)$ is not critical but is done to obtain ones along the diagonal of the covariance matrix. The volume of relative overlap is then given by

$$\begin{aligned} C_k^a(h^+) &= \frac{2V_{k_i-1}(1)G_{k_i-1}(h^+; 1)}{2V_{k_i-1}(1)G_{k_i-1}(0; 1)} = \frac{G_{k_i-1}(h^+; 1)}{G_{k_i-1}(0; 1)} = (1 - h^+)^k \text{ if } h^+ \leq 1 \\ C_k^a(h^+) &= 0 \text{ otherwise} \end{aligned}$$

Since, $C_k^{\alpha}(h^+)$ is derived explicitly from the moving average representation of a random field evaluated at x and z it is naturally a permissible covariance function and furthermore it will by construction only depend on h^+ .

■

Theorem 6 (Khinchin, 1934) Let $x \in \mathbb{R}^k$ and $z \in \mathbb{R}^k$ and let $C_k^{\alpha}(h^+)$ be the permissible covariance function given by Theorem 5. Since $C_k^{\alpha}(h^+)$ only depends on $h^+ = \frac{1}{2}d_{L_1}(x; z) \in \mathbb{R}_+$ there exists a homogeneous and unique Gaussian random field $m(x)$ ($m(x) \in m^{\alpha}(x)$) such that

$$E(m(x)m(z)) = C_k^{\alpha}(h^+)$$

Proof. see Khinchin (1934).

■