ELSEVIER

Journal of Econometrics 114 (2003) 141–164

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Testing for neglected nonlinearity in regression models based on the theory of random fields

Christian M. Dahl[a], Gloria González-Rivera[b,*]

[a] *Department of Economics, Purdue University, West Lafayette, IND47907, USA*
[b] *Department of Economics, University of California, Riverside, CA 92521-0427, USA*

Received 20 December 2000; received in revised form 5 June 2002; accepted 14 September 2002

Abstract

Within a flexible regression model (J.D. Hamilton, Econometrica 69 (3) (2001) 537) we offer a battery of new Lagrange multiplier statistics that circumvent the problem of unidentified nuisance parameters under the null hypothesis of linearity and that are robust to the specification of the covariance function that defines the random field. These advantages are the result of (i) switching from the L_2 to the L_1 norm; and (ii) assuming that the random field is sufficiently smooth for its covariance function to be locally approximated by a high order Taylor expansion. A Monte Carlo simulation suggests that our statistics have superior power performance on detecting bilinear, neural network, and smooth transition autoregressive specifications. We also provide an application to the Industrial Production Index of sixteen OECD countries.

JEL classification: C12; C15; C22

Keywords: Flexible regression; Random field; Neglected nonlinearity

1. Introduction

Hamilton (2001) proposed a parametric approach to estimating nonlinear relations that shares the flexibility of nonparametric models. He studied a model of the form $y_t = \mu(x_t) + \varepsilon_t$ where the functional form $\mu(x_t)$ 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

* Corresponding author. Tel.: +1-909-827-1590; fax: +1-909-787-5685.
E-mail addresses: dahlc@mgmt.purdue.edu (C.M. Dahl), gloria.gonzalez@ucr.edu (G. González-Rivera).

0304-4076/03/$ - see front matter
PII: S0304-4076(02)00222-1

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 (1996). Simulation studies as in Dahl (2002), based on low dimensional linear and nonlinear models, indicate that the Hamilton's test seems to be powerful in comparison to existing linearity tests. However, in Hamilton's test as well as in the neural network test (White, 1989), 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 give 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 of the random field.

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, and (2) robust to the specification of the variance-covariance function associated with the random field. All the proposed tests are Lagrange multiplier statistics, of which we present the TR^2 version. To achieve our goals, we modify the Hamilton approach in two directions. First, we specify the random field in the L_1 norm instead of the L_2 norm. The main advantage of the L_1 norm is that this distance measure is a linear function of the nuisance parameters, in contrast to the L_2 norm which is a nonlinear function. Second, we consider random fields that may not have a simple moving average representation. Logically, we proceed in an opposite fashion to Hamilton. Whereas Hamilton first proposed a moving average representation of the random field and, secondly, he derived its corresponding variance-covariance function, we, in our approach, first propose a covariance function, and secondly we inquire whether there is random field associated with it. We show that the proposed covariance function is general enough to accommodate a broad class of covariance functions. We need to assume that the random field is smooth, in this case, the covariance function is differentiable and it can be approximated reasonable well by a high order Taylor expansion. The local approximation approach is related to Luukkonen et al. (1988) who used 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 et al. (1993) that suggested a version of the neural network test free of unidentified nuisance parameters under the null.

We also show that there is an alternative way of formulating the null hypothesis of linearity to the one advocated by Hamilton. This alternative approach seems particularly promising because the number of unidentified parameters under the null is equal to one regardless of the dimension of the model. Nevertheless, we also provide a Lagrange multiplier statistic that circumvents the problem 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 proposed statistics not only fully match the power properties of the statistic suggested by Hamilton (2001) but also have superior performance on detecting bilinear, neural network, and smooth transition autoregressive specifications. In addition, we compare our tests with the popular Tsay statistic that is also based on a Taylor approximation of an unknown but deterministic functional form. Our tests seem to have better power properties than the Tsay test. This may suggest that treating the unobserved nonlinear function as random rather than deterministic seems a fruitful approach for identification of nonlinear components in parametric regression models.

We show the implementation of our tests with an empirical application concerning the potential nonlinearity of the industrial production index. This example serves to illustrate the perverse effect of a large number of unidentified nuisance parameters. In large models, tests that assume full knowledge of the parameterized random field (a priori fixing the nuisance parameters) tend to loose power rather quickly.

The paper is organized as follows. In Section 2, we deal with preliminaries offering a summary of the Hamilton's flexible regression model, which helps to frame our contribution, and we present the testing problem. In Section 3, we consider alternative representations of the random field and introduce the concept of structure functions. In Section 4, we present the battery of Lagrange multiplier tests. Section 5 contains the results of the Monte Carlo study. In Section 6, we offer an empirical application of the proposed tests to detect nonlinearities in the index of industrial production of the OECD countries. Finally, in Section 7, we conclude.

2. Preliminaries

In this section, we begin with a brief introduction to Hamilton's (2001) flexible regression model, which is the foundation of the test statistics that we propose in the forthcoming sections, and we introduce the testing problem. We discuss the assumptions underlying the derivation of the moving average spatial random field that governs the functional form of the conditional mean in a regression model. For a more detailed analysis of the theory of random fields, see Yaglom (1962, 1987) and Adler (1981) and the references herein.

2.1. Hamilton's flexible regression model approach

Hamilton (2001) suggested representing the nonlinear component in a general regression model by a homogeneous and isotropic Gaussian scalar random field. A scalar random field is defined as a function $m(\omega, \boldsymbol{x}) : \Omega \times A \to \mathbb{R}$ such that $m(\omega, \boldsymbol{x})$ is a random variable for each $\boldsymbol{x} \in A$ where $A \subseteq \mathbb{R}^k$. A random field is also denoted as $m(\boldsymbol{x})$. If $m(\boldsymbol{x})$ 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 by its mean function $\mu(\boldsymbol{x}) = \mathrm{E}[m(\boldsymbol{x})]$ and its covariance function with typical element $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{E}[(m(\boldsymbol{x}) - \mu(\boldsymbol{x}))(m(\boldsymbol{z}) - \mu(\boldsymbol{z}))]$ for any $\boldsymbol{x}, \boldsymbol{z} \in A$. The random field is said to

be homogeneous or stationary if $\mu(\boldsymbol{x})=\mu$ and the covariance function depends only on the difference vector $\boldsymbol{x}-\boldsymbol{z}$ and we should write $\boldsymbol{C}(\boldsymbol{x},\boldsymbol{z})=\boldsymbol{C}(\boldsymbol{x}-\boldsymbol{z})$. Furthermore, the random field is said to be isotropic if the covariance function depends on $d(\boldsymbol{x},\boldsymbol{z})$, where $d(\cdot)$ is a scalar measure of distance. In this situation we write $\boldsymbol{C}(\boldsymbol{x},\boldsymbol{z})=\boldsymbol{C}(d(\boldsymbol{x},\boldsymbol{z}))$.

The specification suggested by Hamilton (2001) can be represented as

$$y_t=\beta_0+\boldsymbol{x}_t'\boldsymbol{\beta}_1+\lambda m(\boldsymbol{g}\odot\boldsymbol{x}_t)+\varepsilon_t, \tag{1}$$

for $y_t\in\mathbb{R}$ and $\boldsymbol{x}_t\in\mathbb{R}^k$, both stationary and ergodic processes. The conditional mean has a linear component given by $\beta_0+\boldsymbol{x}_t'\boldsymbol{\beta}_1$ and a nonlinear component given by $\lambda m(\boldsymbol{g}\odot\boldsymbol{x}_t)$, where $m(\boldsymbol{z})$, for any choice of $\boldsymbol{z}$, represents a realization of a Gaussian and homogenous random field (to be described shortly); $\boldsymbol{x}_t$ could be predetermined or exogenous and is independent of $m(\cdot)$, and ε_t is a sequence of independent and identically distributed $\mathrm{N}(0,\sigma^2)$ variates independent of both $m(\cdot)$ and $\boldsymbol{x}_t$ as well as of lagged values of $\boldsymbol{x}_t$. The scalar parameter λ represents the contribution of the nonlinear part to the conditional mean, the vector $\boldsymbol{g}\in\mathbb{R}^k_{0,+}$ drives the curvature of the conditional mean, and the symbol $\odot$ denotes element-by-element multiplication. Hamilton specification is also closely related to the cubic spline smoother of Wahba (1978) as discussed in Dahl (2002).

The random field analyzed in Hamilton (2001) is generated as a sequence of uncorrelated random variables. A short description of the construction of this random field follows. Define a grid in $\mathbb{R}^k$ by the nodes $\{\boldsymbol{x}(i_1,i_2,\ldots,i_k)\}$ where the index $i_j\in\{1,\ldots,N\}$ for $j=1,\ldots,k$. Define the set A_N that consists of the N^k distinct points in $\mathbb{R}^k$ covered by this grid. For each $\boldsymbol{x}\in A_N$, associate a random variable $e(\boldsymbol{x})\sim\mathrm{N}(0,1)$ such that $e(\boldsymbol{x})$ is independent of $e(\boldsymbol{z})$ for $\boldsymbol{x}\neq\boldsymbol{z}$. Define the set $B_N(\boldsymbol{x})\subset A_N$ of all points in A_N such that the Euclidean distance from $\boldsymbol{x}$ is less or equal to one

$$B_N(\boldsymbol{x})=\{\boldsymbol{z}\in A_N:(\boldsymbol{x}-\boldsymbol{z})'(\boldsymbol{x}-\boldsymbol{z})\leqslant 1\}. \tag{2}$$

Consider the scalar $m_N(\boldsymbol{x})$ associated with every point $\boldsymbol{x}$ in A_N

$$m_N(\boldsymbol{x})=n_N(\boldsymbol{x})^{-1/2}\sum_{\boldsymbol{z}\in B_N(\boldsymbol{x})}e(\boldsymbol{z}), \tag{3}$$

where $n_N(\boldsymbol{x})$ denotes the number of points in $B_N(\boldsymbol{x})$. A random field generated as $m_N(\boldsymbol{x})$ in (3), is said to have a moving average representation, e.g. Yaglom (1962, pp. 31). If the grid partition becomes arbitrarily fine, in the limit we have that $m_N(\boldsymbol{x})\to m(\boldsymbol{x})$, where $m(\boldsymbol{x})$ represents a continuous valued k-dimensional random field such that for any $\boldsymbol{x}$, the scalar $m(\boldsymbol{x})\sim\mathrm{N}(0,1)$.

Let $\boldsymbol{H}_k$ be the covariance (correlation) function of the random field $m(\cdot)$ with typical element defined as $\boldsymbol{H}_k(\boldsymbol{x},\boldsymbol{z})=\mathrm{E}[m(\boldsymbol{x})m(\boldsymbol{z})]$. Hamilton (2001) proved that the covariance function depends solely upon the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{z}$, rendering the random field isotropic. For any $\boldsymbol{x}$ and $\boldsymbol{z}\in\mathbb{R}^k$, the correlation between $m(\boldsymbol{x})$ and $m(\boldsymbol{z})$ is given by the ratio of the volume of the overlap of k-dimensional unit spheroids centered at $\boldsymbol{x}$ and $\boldsymbol{z}$ to the volume of a single k-dimensional unit spheroid. If the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{z}$ is greater than two, the correlation between $m(\boldsymbol{x})$

and $m(\boldsymbol{z})$ will be equal to zero. The general expression of the correlation function is

$$\boldsymbol{H}_k(h) = \begin{cases} G_{k-1}(h,1)/G_{k-1}(0,1) & \text{if } h \leqslant 1, \\ 0 & \text{if } h > 1, \end{cases}$$

$$G_k(h,r) = \int_h^r (r^2 - w^2)^{k/2}\,\mathrm{d}w, \tag{4}$$

where $h \equiv \frac{1}{2}d_{L_2}(\boldsymbol{x},\boldsymbol{z})$, and $d_{L_2}(\boldsymbol{x},\boldsymbol{z}) \equiv [(\boldsymbol{x}-\boldsymbol{z})'(\boldsymbol{x}-\boldsymbol{z})]^{1/2}$ is the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{z}$.[1] In the geostatistical literature, the covariance (4) is widely known as the *spherical covariance function*.[2]

2.2. The testing problem

Consider the model given in Eq. (1). The contribution of the nonlinear component to the conditional mean is driven by the parameter λ and/or by the parameter vector $\boldsymbol{g}$. 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 alternative approaches to specify the null hypothesis of linearity: (i) If the null hypothesis is written as $\mathrm{H}_0 : \lambda^2 = 0$, the parameter vector $\boldsymbol{g}$ is unidentified under the null and the number of unidentified parameters increases with the dimensionality of the model. For this case, Hamilton (2001) proposed an LM test where the vector $\boldsymbol{g}$ is fixed to the mean of its prior distribution. (ii) If the null hypothesis is written as $\mathrm{H}_0 : \boldsymbol{g} = \boldsymbol{0}_k$, the parameter λ becomes unidentified under the null. In this case, the number of unidentified parameters remains equal to one whenever the dimensionality of the model increases. Under this null, the stochastic process becomes nonergodic. For instance, consider model (1) that, under the null, becomes $y_t = \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \lambda m(\boldsymbol{0}_k) + \varepsilon_t$, where $m(\boldsymbol{0}_k) \sim \mathrm{N}(0,1)$. It is apparent that the model will be linear on $\boldsymbol{x}_t$, but y_t will be nonergodic, since even in the simplest case where $\boldsymbol{x}_t$ is deterministic, we have that $\mathrm{cov}(y_t y_{t-s}) = \lambda^2$ for any $s > 0$. 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 $\boldsymbol{g}$ may not have a well defined asymptotic distribution under the null.

We propose a solution to the nuisance parameter problem in (i) and (ii) that is based on alternative representations of the random field $m(\boldsymbol{x})$, which we discuss in the next section.

3. Alternative representations of the random field

We investigate two representations of a random field following two different approaches. We maintain the assumptions of Gaussianity and homogeneity of the random

[1] For a formal proof, see Theorem 2.2 in Hamilton (2001).

[2] Within the area of geostatistics, Matheron (1973) suggested an alternative method based on the turning bands operator to compute the covariance function for any arbitrary value k. Other procedures based on spectral methods to build permissible covariance functions are described in Christakos (1992).

field while exploring the role played by the isotropy assumption. In the first approach, we construct a random field based on the L_1 norm and we derive its covariance function. We show that the change of norm affect the isotropic properties of the random field. In the second approach, we propose a potential covariance function and we inquire whether there is a random field associated with it. In this approach, our main concern is to assure that the proposed function is non-negative definite and, hence, it can be considered a permissible covariance function.

3.1. A nonisotropic 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 investigate an alternative moving average representation based on the Minkowski distance or L_1 norm. The advantage of the L_1 norm is that this distance measure is a linear function of the unidentified nuisance parameters and that will simplify the tests of neglected nonlinearity, as it will be seen in the forthcoming sections. The general setup follows Hamilton (2001) but instead of the set $B_N(\boldsymbol{x})$, we consider the set $B_N^*(\boldsymbol{x})$ defined as all points in A_N such that the L_1 distance from $\boldsymbol{x}$ is less or equal to one

$$B_N^*(\boldsymbol{x}) = \{\boldsymbol{z} \in A_N : |\boldsymbol{x} - \boldsymbol{z}|'\boldsymbol{\iota} \leqslant 1\}.$$

Let $m_N^*(\boldsymbol{x})$ be a random field with the following moving average representation:

$$m_N^*(\boldsymbol{x}) = n_N^*(\boldsymbol{x})^{-1/2} \sum_{\boldsymbol{z} \in B_N^*(\boldsymbol{x})} e(\boldsymbol{z}), \tag{5}$$

where $n_N^*(\boldsymbol{x})$ denotes the number of points in $B_N^*(\boldsymbol{x})$. As before, if the partition becomes arbitrarily fine, in the limit, we have that $m_N^*(\boldsymbol{x}) \to m^*(\boldsymbol{x})$ where $m^*(\boldsymbol{x})$ is a continuously valued k-dimensional random field, i.e. $m^*(\boldsymbol{x}) \sim \mathrm{N}(0,1)$, which is identical to the distribution of $m(\boldsymbol{x})$. However, the correlation between $m^*(\boldsymbol{x})$ and $m^*(\boldsymbol{z})$ for any arbitrary $\boldsymbol{x} \neq \boldsymbol{z}$ will be different. The correlation between $m^*(\boldsymbol{x})$ and $m^*(\boldsymbol{z})$ is given by the ratio of the volume of the overlap of k-dimensional unit orthogons centered at $\boldsymbol{x}$ and $\boldsymbol{z}$ to the volume of a k-dimensional unit orthogon. The volume of the overlap, and hence the correlation between $m^*(\boldsymbol{x})$ and $m^*(\boldsymbol{z})$ will be equal to zero if the L_1 distance between $\boldsymbol{x}$ and $\boldsymbol{z}$ is greater than or equal to two. We denote this particular covariance function $\boldsymbol{H}_k^*(\boldsymbol{x},\boldsymbol{z})$. Following pure geometric arguments, we obtain expressions for $\boldsymbol{H}_k^*(\boldsymbol{x},\boldsymbol{z}) = \mathrm{E}[m^*(\boldsymbol{x})m^*(\boldsymbol{z})]$ for the case of $k = 1, 2$.

Example 1. Let $d_{L_1}(\boldsymbol{x}_t, \boldsymbol{x}_s) = |\boldsymbol{x}_t - \boldsymbol{x}_s|'\boldsymbol{\iota}$. Consider the covariance function $\boldsymbol{H}_k^*$, with typical element $\boldsymbol{H}_k^*(\boldsymbol{x}_t, \boldsymbol{x}_s) = \mathrm{E}[m^*(\boldsymbol{x}_t)m^*(\boldsymbol{x}_s)]$. For $k = 1, 2$

$$\boldsymbol{H}_1^*(\boldsymbol{x}_t, \boldsymbol{x}_s) = \begin{cases} 1 - h_{ts}^* & \text{if } h_{ts}^* \leqslant 1, \\ 0 & \text{if } h_{ts}^* > 1, \end{cases}$$

$$\boldsymbol{H}_2^*(\boldsymbol{x}_t, \boldsymbol{x}_s) = \begin{cases} (1 - h_{ts}^*)^2 + (1 - h_{ts}^*)\min\{|x_{1t} - x_{1s}|, |x_{2t} - x_{2s}|\} & \text{if } h_{ts}^* \leqslant 1, \\ 0 & \text{if } h_{ts}^* > 1, \end{cases}$$

where $h_{ts}^* = \frac{1}{2} d_{L_1}(\boldsymbol{x}_t, \boldsymbol{x}_s)$.

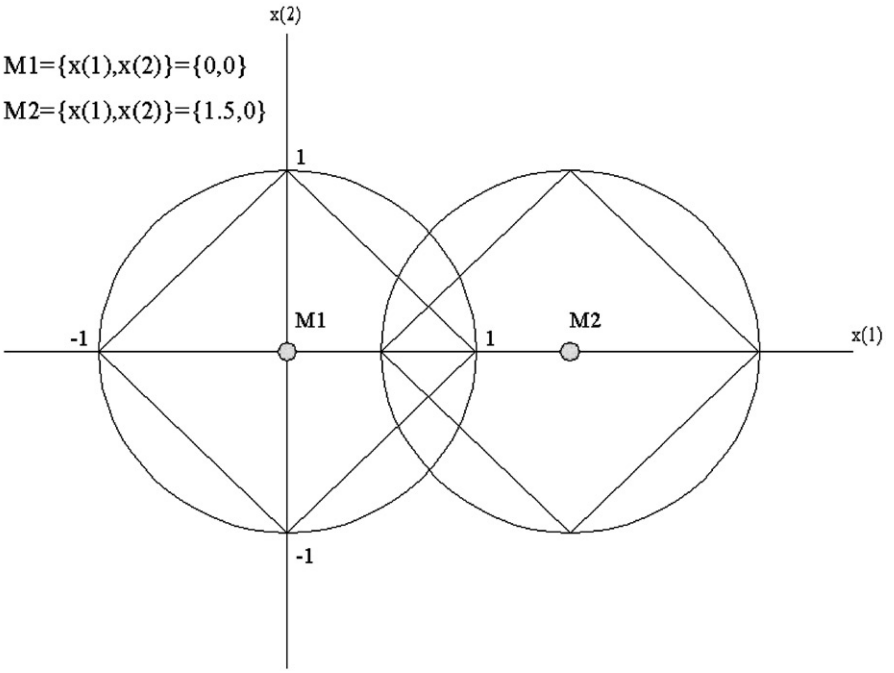

Fig. 1. Illustration of the sets $B_N(\boldsymbol{x})$ and $B_N^*(\boldsymbol{x})$ as the basis for calculating the spherical covariance function $\mathrm{E}(m_N(0,0)m_N(1.5,0))$—given by the intersection of the unit circles—and the L_1 norm covariance function $\mathrm{E}(m_N^*(0,0)m_N^*(1.5,0))$—given by the intersection of the two othogons.

Fig. 1 illustrates the difference between the sets $B_N(\boldsymbol{x})$ and $B_N^*(\boldsymbol{x})$ on calculating the spherical covariance function and the L_1 norm-based covariance function. Notice, that in the case of $k=1$ the L_2 norm and the L_1 norm will provide identical covariance functions. In the case of $k=2$, the volume of the relative overlap based on the L_1 norm is equal to the area of a rectangle.

The disadvantage of a random field with a moving average representation in L_1 is that the field is not isotropic. Consider three orthogons centered at points $\boldsymbol{x}$ and $\boldsymbol{z}$, and at the origin, with the property that $\boldsymbol{x}$ and $\boldsymbol{z}$ are equidistant in L_1 from the origin. Calculate the covariance between the orthogon centered at $\boldsymbol{x}$ and the one centered at the origin; and the covariance between the orthogon centered at $\boldsymbol{z}$ and the one centered at the origin. It is easy to see that the volume of the overlap between the orthogon centered at the origin and the orthogon centered at $\boldsymbol{x}$ is different from the overlap between the orthogon centered at the origin and the orthogon centered at $\boldsymbol{z}$, even though the L_1 distance from $\boldsymbol{x}$ to the origin is the same as the L_1 distance from point $\boldsymbol{z}$ to the origin. In other words, the location of the points matters, as it can be read from the formulas of $\boldsymbol{H}_k^*(\boldsymbol{x}_t,\boldsymbol{x}_s)$ in the previous example. If the random field is not isotropic, the calculation of its covariance function could be very cumbersome when the number of regressors is large. However, the functional form of the leading term of

the covariance function in Example 1 provides a starting point to construct a simple permissible covariance function of an isotropic random field.

3.2. Permissible covariance functions

Consider the following homogeneous and isotropic function $\boldsymbol{C}_k(\boldsymbol{x},\boldsymbol{z})=\boldsymbol{C}_k(r)$ for $\boldsymbol{x},\boldsymbol{z}\in\mathbb{R}^k$, that depends only on a distance measure (defined either in L_1 norm or L_2 norm) $r=\frac{1}{2}d(\boldsymbol{x},\boldsymbol{z})$ such that

$$\boldsymbol{C}_k(r)=\begin{cases}(1-r)^{\delta} & \text{if } r\leqslant 1,\\ 0 & \text{if } r>1,\end{cases} \tag{6}$$

for $\delta\geqslant 1$. Note that if $r=h^*$ and $\delta=k$, then it can be shown that $\boldsymbol{C}_k(r)$ equals the first term of $\boldsymbol{H}_k^*(\boldsymbol{x},\boldsymbol{z})$, for all k. Two questions are pertinent. First, is (6) a permissible covariance function? Second, if it is, is there a random field associated with it?

A covariance function $\boldsymbol{C}_k$ is said to be permissible if and only if it satisfies the positive semidefiniteness condition, that is $\boldsymbol{q}'\boldsymbol{C}_k(\boldsymbol{x},\boldsymbol{z})\boldsymbol{q}\geqslant 0$ for all $\boldsymbol{q}\neq\boldsymbol{0}_T$. This condition for permissibility is impossible to check in practice but there are permissibility criteria that guarantee the positive semidefiniteness property. According to Yaglom (1962, 1987) and Christakos (1992) any permissible covariance function associated with a general homogeneous Gaussian random field such that $\boldsymbol{C}_k(\boldsymbol{x},\boldsymbol{z})=\boldsymbol{C}_k(\boldsymbol{x}-\boldsymbol{z})$ must satisfy the following necessary conditions in the time domain:

$$\begin{aligned}&\boldsymbol{C}_k(\boldsymbol{0}_k)>0,\\ &|\boldsymbol{C}_k(\boldsymbol{x}-\boldsymbol{z})|\leqslant\boldsymbol{C}_k(\boldsymbol{0}_k),\\ &\boldsymbol{C}_k(\boldsymbol{x}-\boldsymbol{z})=\boldsymbol{C}_k(-(\boldsymbol{x}-\boldsymbol{z})),\\ &\lim_{d(\boldsymbol{x},\boldsymbol{z})\to\infty}\frac{\boldsymbol{C}_k(\boldsymbol{x}-\boldsymbol{z})}{d(\boldsymbol{x},\boldsymbol{z})^{(1-k)/2}}=0.\end{aligned} \tag{7}$$

For isotropic random fields, the norm in which the covariance function is defined plays a crucial role. A function that is a permissible covariance function in the Euclidean geometry may not be in another geometry such that defined in the L_1 norm. For isotropic fields defined in the L_2 norm, Christakos (1992, Chapter 7) provided a set of sufficient conditions for permissibility of a covariance function for $k=1,2,3$. Define $d\equiv d_{L_2}(\boldsymbol{x},\boldsymbol{z})$, sufficient conditions for permissibility are:

$$\begin{aligned}&\mathrm{d}\boldsymbol{C}_k(d)|_{d=0}<0,\\ &\mathrm{d}^2\boldsymbol{C}_r(d)\geqslant 0 \quad \text{in } \mathbb{R},\\ &\int_d^{\infty}\frac{u}{\sqrt{u^2-d^2}}\,\mathrm{d}^3\boldsymbol{C}_k(d)\geqslant 0 \quad \text{in } \mathbb{R}^2,\\ &\mathrm{d}^2\boldsymbol{C}_k(d)-d\times[\mathrm{d}^3\boldsymbol{C}_k(d)]\geqslant 0 \quad \text{in } \mathbb{R}^3.\end{aligned} \tag{8}$$

These conditions impose some degree of smoothness on the covariance functions. In the L_2 norm, the function (6) for which $r=h\equiv\frac{1}{2}d_{L_2}(\boldsymbol{x},\boldsymbol{z})$ and $\delta=k$ satisfies the necessary conditions (7) and the sufficient conditions (8).

The function (6) can be viewed as a particular case of the function studied in Zastavnyi (2000). He provided necessary and sufficient conditions for the function (6) to be positive semidefinite in different geometries. In the L_2 norm, (6) is positive semidefinite if and only if $\delta \geqslant (k+1)/2$; and in the L_1 norm, (6) is positive semidefinite if and only if $\delta \geqslant 2k-1$. Hence, the function (6) for $r = h^* \equiv \frac{1}{2} d_{L_1}(\boldsymbol{x}, \boldsymbol{z})$

$$\boldsymbol{C}_k^*(h^*) = \begin{cases} (1-h^*)^{2k} & \text{if } h^* \leqslant 1, \\ 0 & \text{if } h^* > 1, \end{cases} \tag{9}$$

is a permissible covariance function. The smoothness of the function (9) together with its simplicity will be important characteristics in the forthcoming sections when we discuss Taylor approximations of the covariance function $\boldsymbol{C}_k(\boldsymbol{x}, \boldsymbol{z})$ and when we implement our proposed tests for neglected nonlinearity.

An important question is whether there is a random field that can be associated with (9). The answer is positive in the light of the Khinchin's (1934) theorem and Bochner's (1959) theorem. The basic argument is that the class of functions which are covariance functions of homogenous random fields coincides with the class of positive semidefinite functions. Hence, (9) being a positive semidefinite function must be the covariance function of a homogenous random field.

3.3. The concept of structure functions

In the previous section, we showed that testing for linearity in model (1) with the null hypothesis $\mathrm{H}_0 : \boldsymbol{g} = \boldsymbol{0}_k$ produces a nonergodic stochastic process under the null. We will provide a simple modification of the specification of the function $m(\boldsymbol{x})$ that preserves the ergodicity of y_t under the null. We write the modified unrestricted model as $y_t = \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \lambda \tilde{m}(\boldsymbol{g} \odot \boldsymbol{x}_t) + \varepsilon_t$, where $\tilde{m}(\boldsymbol{x}) = m(\boldsymbol{x}) - m(\boldsymbol{0}_k)$. Notice that $\tilde{m}(\boldsymbol{0}_k) = 0$, and the model under the null becomes $y_t = \beta_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \varepsilon_t$ restoring the ergodicity of y_t under the null hypothesis, provided that $\boldsymbol{x}_t$ and ε_t are stationary and ergodic. We need to specify the covariance structure of $\tilde{m}(\boldsymbol{x})$. Since we restrict ourselves to the class of Gaussian random fields, $\tilde{m}(\boldsymbol{x})$—the sum of two Gaussian random fields—will also be Gaussian. Let $\tilde{\boldsymbol{C}}_k$ be the covariance function that uniquely determines the random field $\tilde{m}(\boldsymbol{x})$. The typical element in $\tilde{\boldsymbol{C}}_k$ is defined as $\tilde{\boldsymbol{C}}_k(\boldsymbol{x}, \boldsymbol{z}) = \mathrm{E}[\tilde{m}(\boldsymbol{x})\tilde{m}(\boldsymbol{z})] = \mathrm{E}\{m(\boldsymbol{x}) - m(\boldsymbol{0}_k)\}\{m(\boldsymbol{z}) - m(\boldsymbol{0}_k)\}$, hence the covariance function can be written as $\tilde{\boldsymbol{C}}_k(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{C}_k(\boldsymbol{x}, \boldsymbol{z}) + \boldsymbol{C}_k(\boldsymbol{0}_k, \boldsymbol{0}_k) - \boldsymbol{C}_k(\boldsymbol{x}, \boldsymbol{0}_k) - \boldsymbol{C}_k(\boldsymbol{0}_k, \boldsymbol{z})$. Yaglom (1962, p. 87) names $\tilde{\boldsymbol{C}}_k$ the structure function. In the case of homogenous random fields, the structure function $\tilde{\boldsymbol{C}}_k$ is a permissible covariance function provided that $\boldsymbol{C}_k$ is permissible, e.g. Yaglom (1962, p. 88). This property will be important when we derive the Lagrange multiplier statistics corresponding to $\mathrm{H}_0 : \boldsymbol{g} = \boldsymbol{0}_k$.

4. Lagrange multiplier tests for neglected nonlinearity

We consider two types of Lagrange multiplier tests for neglected nonlinearity. The first type of tests are derived under $\mathrm{H}_0 : \lambda^2 = 0$, and denoted λ-tests. The second type of tests are derived under $\mathrm{H}_0 : \boldsymbol{g} = \boldsymbol{0}_k$, and denoted g-tests. We introduce a λ-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 of the numerical values of each element in the $\boldsymbol{g}$ vector. Analogously, we introduce a g-test for neglected nonlinearity that does not depend on the values of λ. It should be emphasized that since the test statistics are all Lagrange multiplier test statistics, they are characterized by being locally most powerful tests.

4.1. The likelihood function

Consider the model given by Eq. (1). We can write $\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X\beta}, \lambda^2 \boldsymbol{C}_k + \sigma^2 \boldsymbol{I}_T)$ where $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)'$, $\boldsymbol{X}_1 = (\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_T')'$, $\boldsymbol{X} = (\boldsymbol{\iota} : \boldsymbol{X}_1)$, $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)'$ and σ^2 is the variance of ε_t. $\boldsymbol{C}_k$ is a generic covariance function associated with the random field, which could be equal to the spherical covariance function in (4), or to that of the anisotropic field in Example 1, or to that of the isotropic field in (9). The log-likelihood function corresponding to this model is

$$\ell(\boldsymbol{\beta}, \lambda^2, \boldsymbol{g}, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\lambda^2 \boldsymbol{C}_k + \sigma^2 \boldsymbol{I}_T| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\lambda^2 \boldsymbol{C}_k + \sigma^2 \boldsymbol{I}_T)^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}), \tag{10}$$

which 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 ϑ, $\mathrm{H}_0 : \vartheta = \tilde{\vartheta}$, where $\tilde{\vartheta} = (\vartheta_1^{0\prime}, \tilde{\vartheta}_2')'$ with $\tilde{\vartheta}_2$ being the maximum likelihood estimate of ϑ_2 under the null, and ϑ_1^0 the parameter ϑ_1 specified under the null. The Lagrange multiplier statistic is given by $LM = \boldsymbol{s}(\tilde{\vartheta})' \boldsymbol{\mathcal{I}}^{-1}(\tilde{\vartheta}) \boldsymbol{s}(\tilde{\vartheta})$, where $\boldsymbol{s}(\tilde{\vartheta})$ denotes the score function, and $\boldsymbol{\mathcal{I}}^{-1}(\tilde{\vartheta})$ 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 $\boldsymbol{\mathcal{I}}_{\mathrm{H}}$, is based on the Hessian of the log likelihood function, and the second, denoted $\boldsymbol{\mathcal{I}}_{\mathrm{OP}}$, is based on the outer-product of the score. Using the $\boldsymbol{\mathcal{I}}_{\mathrm{OP}}$ 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 squares regression. Under the usual regularity conditions, the Lagrange multiplier statistics will be $\chi^2(q)$-distributed where q equals the number of restrictions under the null. When a subset of the parameters in ϑ_2 is not identified under the null, it is not possible to obtain consistent maximum likelihood estimates of this subset of parameters under the null. In this situation, we say that the testing problem is suffering from unidentified nuisance parameters. Methods to solve the problem of unidentified nuisance parameters depend to some extent on the particular model under consideration. In the next subsections, we deal with this problem in detail after having specified the model under the null as well as under the alternative hypothesis.

4.2. Tests based on known covariance functions

Hamilton (2001) derived the λ-test for neglected nonlinearity based on the $\boldsymbol{\mathcal{I}}_{\mathrm{H}}$ estimator of the information matrix and with $\boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{x}_s) = \boldsymbol{H}_k(h)$ for $h = \frac{1}{2} d_{L_2}(\boldsymbol{x}_t, \boldsymbol{x}_s)$.

Let us denote Hamilton's test statistic $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$. To deal with the identification issues, Hamilton suggested fixing $\boldsymbol{g}$ to a given value and proceed to derive the Lagrange multiplier test, which then follows a standard asymptotic distribution. Heuristically, fixing $\boldsymbol{g}$ is identical to a derivation of the Lagrange multiplier statistic assuming complete knowledge of the covariance matrix associated with the random field. Our first aim is to derive the TR^2 version of the test statistic conditional on full knowledge of the parameterized covariance function. Recall the log-likelihood function (10). Keeping $\boldsymbol{g}$ fixed, defining $\boldsymbol{\Omega} = \lambda^2 \boldsymbol{C}_k + \sigma^2 \boldsymbol{I}_T$, and evaluating the score functions under the null of linearity ($\mathrm{H}_0 : \lambda^2 = 0 \Rightarrow \boldsymbol{\Omega} = \sigma^2 \boldsymbol{I}_T$), we can write [3]

$$s(\lambda^2)|_{\lambda^2=0,\boldsymbol{g}} \equiv \left.\frac{\partial \ell(\boldsymbol{\beta}, \lambda^2, \boldsymbol{g}, \sigma^2)}{\partial \lambda^2}\right|_{\lambda^2=0,\boldsymbol{g}} = -\frac{1}{2\sigma^2} \operatorname{tr}\left[\boldsymbol{C}_k \left(\boldsymbol{I}_T - \frac{\varepsilon\, \varepsilon'}{\sigma^2}\right)\right],$$

$$s(\sigma^2)|_{\lambda^2=0,\boldsymbol{g}} \equiv \left.\frac{\partial \ell(\boldsymbol{\beta}, \lambda^2, \boldsymbol{g}, \sigma^2)}{\partial \sigma^2}\right|_{\lambda^2=0,\boldsymbol{g}} = -\frac{1}{2\sigma^2} \operatorname{tr}\left[\boldsymbol{I}_T \left(\boldsymbol{I}_T - \frac{\varepsilon\varepsilon'}{\sigma^2}\right)\right],$$

where $s(.)|_{\lambda^2=0,\boldsymbol{g}}$ means the score function evaluated under the null hypothesis $\lambda^2 = 0$ for $\boldsymbol{g}$ fixed. Using the following property of the trace $\operatorname{tr}(\mathbf{AB}) = \operatorname{vec}(\mathbf{A})' \operatorname{vec}(\mathbf{B}')$, the score functions can be written as

$$s(\lambda^2)|_{\lambda^2=0,\boldsymbol{g}} = -\frac{1}{2\sigma^2}\, \tilde{\boldsymbol{x}}_1' \boldsymbol{u}, \tag{11}$$

$$s(\sigma^2)|_{\lambda^2=0,\boldsymbol{g}} = -\frac{1}{2\sigma^2}\, \tilde{\boldsymbol{x}}_2' \boldsymbol{u}, \tag{12}$$

where $\tilde{\boldsymbol{x}}_1 = \operatorname{vec}(\boldsymbol{C}_k)$, $\tilde{\boldsymbol{x}}_2 = \operatorname{vec}(\boldsymbol{I}_T)$, and $\boldsymbol{u} = \operatorname{vec}(\boldsymbol{I}_T - \varepsilon\, \varepsilon'/\sigma^2)$. Letting $\tilde{\boldsymbol{x}} = (\tilde{\boldsymbol{x}}_1 : \tilde{\boldsymbol{x}}_2)$, the Lagrange multiplier test is $\frac{1}{2}\boldsymbol{u}'\tilde{\boldsymbol{x}}(\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{x}}'\boldsymbol{u}$. Since $\boldsymbol{u}'\boldsymbol{u}/T^2 \xrightarrow{\mathrm{p}} 1$, the TR^2 version of the Lagrange multiplier statistic, which we denote $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$ is

$$\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g}) = \frac{T^2}{2} \frac{\boldsymbol{u}'\tilde{\boldsymbol{x}}(\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}})^{-1}\tilde{\boldsymbol{x}}'\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{u}} \sim \chi^2(1).$$

The statistic is easily obtained by the following procedure: (1) Estimate the model under the null and compute $\hat{\varepsilon} = \boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y})$ and $\hat{\sigma}^2 = T^{-1}\hat{\varepsilon}'\hat{\varepsilon}$. (2) Obtain the least squares estimate of $\boldsymbol{v}$—denoted $\hat{\boldsymbol{v}}$—from the auxiliary regression $\hat{\boldsymbol{u}} = \phi_1 \tilde{\boldsymbol{x}}_1 + \phi_2 \tilde{\boldsymbol{x}}_2 + \boldsymbol{v}$, using $\hat{\boldsymbol{u}} = \operatorname{vec}(\boldsymbol{I}_T - \hat{\varepsilon}\hat{\varepsilon}'/\hat{\sigma}^2)$. (3) Obtain the uncentered R^2 as $R^2 = 1 - \hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}/\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$. (4) Finally, the Lagrange multiplier statistic is given as $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g}) = \frac{1}{2}T^2R^2$.

In order to stress that the above version of the Lagrange multiplier statistic depends crucially on the choice of a fully specified covariance matrix we write it as $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$. Notice that since there are T^2 observations in the auxiliary regression, the R^2 should be multiplied by T^2 instead of T in order to make the $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$ statistic asymptotically $\chi^2(1)$-distributed.

4.3. λ-tests based on unknown covariance functions

In this section, we propose a new statistic to test the null hypothesis of linearity that circumvents the problem of unidentified nuisance parameters and does not depend

[3] Though this null is on the boundary of the parameter space, the properties of the LM test are not altered, see Godfrey (1988, pp. 92–98).

on the exact parameterization of the covariance function. The approach we suggest is based on the substitution of $\tilde{\boldsymbol{x}}_1 = \text{vec}(\boldsymbol{C}_k)$ for a Taylor approximation of $\tilde{\boldsymbol{x}}_1$ 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 differentiable real-valued nonlinear functions $\text{vec}(\boldsymbol{C}_k)$. 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 that we consider is as in (1) where the random field $m(\boldsymbol{x})$ now is assumed to be a smooth—continuous and differentiable—homogenous and Gaussian.

Consider model (1) with covariance function given by (9), i.e. $\boldsymbol{C}_k^*(h_{ts}^*) = (1 - h_{ts}^*)^{2k} 1_{(h_{ts}^* \leqslant 1)}$, where $h_{ts}^* \equiv \frac{1}{2} d_{L_1}(\boldsymbol{g} \odot \boldsymbol{x}_t, \boldsymbol{g} \odot \boldsymbol{x}_s) = \frac{1}{2} \boldsymbol{r}_{ts}' \boldsymbol{g}$, and $\boldsymbol{r}_{ts} = \{|x_{t1} - x_{s1}|, |x_{t2} - x_{s2}|, \ldots, |x_{tk} - x_{sk}|\}'$. Since $(1-h^*)^{2k} = \sum_{j=0}^{2k} \binom{2k}{j} h^{*j}(-1)^j$ we can rewrite the auxiliary regression as

$$\begin{aligned}\hat{\boldsymbol{u}}_{ts} &= \phi_1 \tilde{\boldsymbol{x}}_{ts,1} + \phi_2 \tilde{\boldsymbol{x}}_{ts,2} + \boldsymbol{v}_{ts} \\ &= \phi_1 \left[\sum_{j=0}^{2k} \binom{2k}{j} h_{ts}^{*j} (-1)^j \right] 1_{(h_{ts}^* \leqslant 1)} + \phi_2 \tilde{\boldsymbol{x}}_{ts,2} + \boldsymbol{v}_{ts}\end{aligned}$$

and proceed to calculate the test statistic as in Section 4.2. However, we can construct a test that does not depend on the nuisance parameters $\boldsymbol{g}$. The key elements are: (1) the choice of the norm, and (2) the approximation of the indicator function by a smooth function. In our case, the L_1 norm depends linearly on the parameters $\boldsymbol{g}$. The indicator function $1_{(h_{ts}^* \leqslant 1)}$ can be approximated by a differentiable function like the logistic function, i.e. $1_{(h_{ts}^* \leqslant 1)} \approx (1 + \exp(-\gamma(1 - h_{ts}^*)))^{-1}$ for fixed $\gamma \gg 0$. A second order Taylor expansion of the logistic function around an average value of h_{ts}^* together with the computation of the powers of h^*, that is h^{*j}, give rise to the following auxiliary regression

$$\begin{aligned}\hat{u}_{ts} = {} & \bar{\phi}_0 + \bar{\phi}_1 \sum_{i=1}^{k} g_i r_{ts,i} + \bar{\phi}_2 \sum_{i=1}^{k} \sum_{j \geqslant i}^{k} g_i g_j r_{ts,i} r_{ts,j} \\ & + \bar{\phi}_3 \sum_{i=1}^{k} \sum_{j \geqslant i}^{k} \sum_{l \geqslant j}^{k} g_i g_j g_l r_{ts,i} r_{ts,j} r_{ts,l} + \cdots \\ & + \bar{\phi}_{2k+2} \sum_{i=1}^{k} \sum_{j \geqslant i}^{k} \cdots \sum_{m}^{k} g_i g_j \cdots g_m r_{ts,i} r_{ts,j} \cdots r_{ts,m} + \phi_2 \tilde{x}_{2,ts} + v_{ts}, \end{aligned} \tag{13}$$

where $\bar{\phi}_j$ is directly proportional to ϕ_1, that is $\bar{\phi}_j = c_j \phi_1$ with c_j being the proportionality parameter. The subindex ts attached to the vectors $\hat{\boldsymbol{u}}, \tilde{\boldsymbol{x}}_2$, and $\boldsymbol{v}$ means the tsth entry/row in the respective vector for $t, s = 1, 2, \ldots, T$, and g_i and $r_{ts,i}$ denote the ith entry in the vectors $\boldsymbol{g}$ and $\boldsymbol{r}_{ts}$, respectively. The second order expansion of the logistic

function linearizes the auxiliary regression at the expense of increasing the number of regressors. Notice that now we can proceed to estimate the auxiliary regression (13) by OLS, treating the nuisance vector $\boldsymbol{g}$ as part of the parameter space. In the exact test of Section 4.2, $\phi_1 = 0$ implied that the null hypothesis of linearity cannot be rejected. Considering the regression (13), $\phi_1 = 0$ implies that $\bar{\phi}_j = 0$, $j = 0, 1, 2, \ldots, 2k+2$. Hence, the regression (13), where $\sum_{j=1}^{2k+2} \binom{k+j-1}{k-1}$ regressors have been added, is the basis to compute a Lagrange multiplier test that is free of nuisance parameters but, in this case, the new test denoted $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ will be χ^2-distributed with $1 + \sum_{j=1}^{2k+2} \binom{k+j-1}{k-1}$ degrees of freedom.

In most situations the econometrician will not know the parametric form of the covariance function and any particular choice will be an approximation to the true function. However, we argue that if we assume the covariance function (9), we are modeling a very broad class of covariance functions. We need to assume that the homogenous random function is smooth, in this case the covariance function will be differentiable everywhere, and an approximation by a Taylor expansion will be possible. 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 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 of the test when using a fully known covariance function or using a Taylor expansion in the auxiliary regression. This type of argument follows very closely Teräsvirta (1998) where he performed 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.

Consider a higher order Taylor approximation of $\boldsymbol{C}_k(h_{ts}^*)$ around $\bar{h}^*$ and denote the approximation of $\boldsymbol{C}_k$ by $\boldsymbol{D}_k$ with typical element given as,

$$\boldsymbol{D}_k(h_{ts}^*) = \boldsymbol{C}_k(\bar{h}^*) + \sum_{j=2}^{n} \frac{1}{(j-1)!} (h_{ts}^* - \bar{h}^*)^{j-1} \mathrm{d}^{j-1} \boldsymbol{C}_k(\bar{h}^*).$$

The linearized auxiliary regression then becomes $\hat{\boldsymbol{u}} = \phi_1 \mathrm{vec}(\boldsymbol{D}_k) + \phi_2 \tilde{\boldsymbol{x}}_2 + \boldsymbol{v}$ or written more explicitly

$$\hat{u}_{ts} = \bar{\phi}_0 + \bar{\phi}_1 h_{ts}^* + \bar{\phi}_2 h_{ts}^{*2} + \bar{\phi}_3 h_{ts}^{*3} + \cdots + \bar{\phi}_{n-1} h_{ts}^{*(n-1)} + \phi_2 \tilde{x}_{2,ts} + v_{ts}, \tag{14}$$

for $t, s = 1, \ldots, T$, where

$$\bar{\phi}_0 = \phi_1 \left\{ \boldsymbol{C}_k(\bar{h}^*) - \bar{h}^* \mathrm{d}\boldsymbol{C}_k(\bar{h}^*) \pm \cdots \pm \frac{1}{(n-1)!} \bar{h}^{*(n-1)} \mathrm{d}^{n-1} \boldsymbol{C}_k(\bar{h}^*) \right\},$$

$$\bar{\phi}_1 = \phi_1 \left\{ \mathrm{d}\boldsymbol{C}_k(\bar{h}^*) - \bar{h}^* \mathrm{d}^2 \boldsymbol{C}_k(\bar{h}^*) \pm \cdots \pm \frac{1}{(n-2)!} \bar{h}^{*(n-2)} \mathrm{d}^{n-1} \boldsymbol{C}_k(\bar{h}^*) \right\},$$

$$\bar{\phi}_2 = \phi_1 \left\{ \frac{1}{2} \mathrm{d}^2 \boldsymbol{C}_k(\bar{h}^*) - \frac{1}{2} \bar{h}^* \mathrm{d}^3 \boldsymbol{C}_k(\bar{h}^*) \pm \cdots \pm \frac{1}{(n-3)!2} \bar{h}^{*(n-3)} \mathrm{d}^{n-1} \boldsymbol{C}_k(\bar{h}^*) \right\},$$

$$\cdots$$

$$\bar{\phi}_{n-1} = \phi_1 \frac{1}{(n-1)!} \mathrm{d}^{n-1} \boldsymbol{C}_k(\bar{h}^*).$$

Notice that the auxiliary regression (14) is equivalent to the auxiliary regression (13) once the powers of h_{ts}^* are calculated and $n-1=2k+2$. The quantity $SSR_1=\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}$ would be identical when OLS is performed in (14) and in (13). Consequently, the $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ test is also a test for nonlinearity when the covariance function is unknown but it can be approximated reasonably well by a high order Taylor expansion.

Some examples of permissible covariance functions (Yaglom 1962, 1987) whose characteristics can be captured by the $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ test are

- $\boldsymbol{C}_k(\boldsymbol{r}_{ts}) = \exp(-\sum_{i=1}^k a_i |x_{ti} - x_{si}|^q)$ for $\boldsymbol{x}_t, \boldsymbol{x}_s \in \mathbb{R}^k$, $k \geqslant 1$, $q = \{1,2\}$, $\{a_i \geqslant 0, \forall i\}$. This is the multiplicative Ornstein–Uhlenbeck covariance function.
- $\boldsymbol{C}_k(\boldsymbol{r}_{ts}) = \exp(-a^2 d_{L_2}(\boldsymbol{x}_t, \boldsymbol{x}_s)^2)$, for $\boldsymbol{x}_t, \boldsymbol{x}_s \in \mathbb{R}^k$, $k \geqslant 1$.
- $\boldsymbol{C}_k(\boldsymbol{r}_{ts}) = \{1 + \exp(-a^2 d_{L_2}(\boldsymbol{x}_t, \boldsymbol{x}_s)^2)^{-b}$, for $\boldsymbol{x}_t, \boldsymbol{x}_s \in \mathbb{R}^k$, $k \geqslant 1$, $b > 0$.
- $\boldsymbol{C}_k(\boldsymbol{r}_{ts}) = b * \exp(-\sum_{i=1}^k a_i |x_{ti} - x_{si}|) \cos(\sum_{i=1}^k c_i |x_{ti} - x_{si}|)$, for $\boldsymbol{x}_t, \boldsymbol{x}_s \in \mathbb{R}^k$, $k \geqslant 1$, $\{a_i \geqslant 0, \forall i\}$, $b > 0$, $\{c_i \geqslant 0, \forall i\}$. This is the damped oscillation covariance function.

4.4. g-tests

Consider model (1) with a random field whose structure function is given by $\tilde{\boldsymbol{C}}_k(\boldsymbol{x}_t, \boldsymbol{x}_s) = \boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{x}_s) + \boldsymbol{C}_k(\boldsymbol{0}_k, \boldsymbol{0}_k) - \boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{0}_k) - \boldsymbol{C}_k(\boldsymbol{0}_k, \boldsymbol{x}_s)$, where $\boldsymbol{C}_k(\boldsymbol{x}_t, \boldsymbol{x}_s)$ has the parametric form of (9), with $h_{ts}^* \equiv \frac{1}{2} d_{L_1}(\boldsymbol{g} \odot \boldsymbol{x}_t, \boldsymbol{g} \odot \boldsymbol{x}_s) = \frac{1}{2} \boldsymbol{r}_{ts}' \boldsymbol{g}$. The likelihood function is as (10) where $\boldsymbol{C}_k$ is replaced by $\tilde{\boldsymbol{C}}_k$ such that $\boldsymbol{\Omega} = \lambda^2 \tilde{\boldsymbol{C}}_k + \sigma^2 \boldsymbol{I}_T$. We proceed with the derivation of the LM test as in Section 4.2. In the case of the g-tests, the λ parameter is unidentified under the null hypothesis and on calculating the score function we need to keep λ fixed. Evaluating the score function under the null of linearity $\mathrm{H}_0 : \boldsymbol{g} = 0$, keeping λ fixed, and $\boldsymbol{\Omega}|_{\boldsymbol{g}=0} = \sigma^2 \boldsymbol{I}_T$, we can write the score functions in vectorized form as

$$s(g_i)|_{\lambda^2, \boldsymbol{g}=0} = -\frac{\lambda^2}{2\sigma^2} \tilde{\boldsymbol{x}}_i' \boldsymbol{u}, \quad i = 1, 2, \ldots, k, \tag{15}$$

$$s(\sigma^2)|_{\lambda^2, \boldsymbol{g}=0} = -\frac{1}{2\sigma^2} \tilde{\boldsymbol{x}}_{k+1}' \boldsymbol{u}, \tag{16}$$

where $\tilde{\boldsymbol{x}}_i = \partial \operatorname{vec}(\tilde{\boldsymbol{C}}_k)/\partial g_i|_{g=0}$, for $i=1,2,\ldots,k$, $\tilde{\boldsymbol{x}}_{k+1} = \operatorname{vec}(\boldsymbol{I}_T)$, and $\boldsymbol{u} = \operatorname{vec}(\boldsymbol{I}_T - \varepsilon\varepsilon'/\sigma^2)$.[4]

With the scores (15) and (16), we compute the g-test as a TR^2 statistic, which is free of the nuisance parameter λ. We denote such a test by g_{OP}. The construction of the test statistic follows the procedure already outlined in Sections 4.2 and 4.3. After

[4] To calculate $\tilde{\boldsymbol{x}}_i$, the indicator function has been substituted for a logistic function, i.e. $1_{(h_{ts}^* \leqslant 1)} \approx (1 + \exp(-\gamma(1 - h_{ts}^*))^{-1}$ for fixed $\gamma \gg 0$.

having obtained $\hat{\varepsilon}$ and $\hat{\sigma}^2$ compute the uncentered R^2 from the auxiliary regression $\hat{u}_{ts} = \sum_{i=1}^{k} \tilde{\phi}_i \tilde{r}_{ts,i} + \tilde{\phi}_{k+1} \tilde{x}_{k+1,ts} + \tilde{v}_{ts}$, where $\tilde{r}_{ts,i} = -k(|x_{ti} - x_{si}| - |x_{ti}| - |x_{si}|)$, for $t,s = 1,2,\ldots,T$. The Lagrange multiplier statistic is then given as $g_{\mathrm{OP}} = \frac{1}{2} T^2 R^2 \sim \chi^2(k)$.

Notice that the g-test does not depend on the unidentified nuisance parameter λ. In order to increase power under the alternative hypothesis, the auxiliary regression can be augmented with higher powers and cross products of $\tilde{r}_{ts}$, thereby increasing the number of degrees of freedom of the asymptotic distribution of the test.

5. Simulation studies on size and power properties

We perform a Monte Carlo simulation study to analyze the small sample size and power properties of the various statistics that we propose.[5] The set up of the simulation follows closely that of Dahl (2002).

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. These benchmark models allow to make comparisons with earlier studies. Their parametric representations are presented in Table 1. The models included in Block 1 were originally used by Lee et al. (1993). The models in Block 2 have been more extensively studied by Keenan (1985), Tsay (1986), Ashley et al. (1986), Chan and Tong (1986), and Lee et al. (1993). The models in Block 3 have been studied by Teräsvirta et al. (1993).

The first five models in Block 1 are all characterized by being simple dynamic univariate models, where the dynamic part is represented by one lag of the endogenous variable. The models are all stationary. The models included are the autoregressive model (AR), the bilinear model (BL) of Granger and Andersen (1978), the threshold autoregressive model (TAR) of Tong (1983), the sign autoregressive model (SGN), and the nonlinear autoregressive model (NAR). We also consider two bivariate representations where we do not impose any dynamic structure. We consider a squared relation (SQ), and an exponential relation (EXP). The models in Block 2 are characterized by having a richer dynamic structure than those in Block 1. Model 1 is an MA(2) representation, Model 2 is a heteroskedastic MA(2), and Model 4 is an AR(2). These three models are linear and they are included primarily to evaluate the nominal size of the tests of linearity and their ability to distinguish between dynamic misspecification and misspecification due to nonlinearity in the conditional mean. Model 3, Model 5 and Model 6 are the truly nonlinear models in Block 2. Model 3 is a nonlinear MA(2), Model 5 is a bilinear autoregressive model, and Model 6 is a bilinear autoregressive moving average model. In Blocks 1 and 2, ε_t is a white noise series distributed $N(0,1)$. In Block 3, the first model is the logistic smooth transition autoregressive model (LSTAR) in Teräsvirta (1994). The second model is a special case of

[5] Our GAUSS programs can be downloaded, free of charge, from our websites.

Table 1
Models considered in the Monte Carlos study

Block 1	
AR	$y_t = 0.6y_{t-1} + \varepsilon_t$
BL	$y_t = 0.7y_{t-1}\varepsilon_{t-2} + \varepsilon_t$
TAR	$y_t = 0.9y_{t-1}1_{(\lvert y_{t-1}\rvert \leqslant 1)} - 0.3y_{t-1}1_{(\lvert y_{t-1}\rvert > 1)} + \varepsilon_t$
SGN	$y_t = 1_{(y_{t-1}>1)} - 1_{(y_{t-1}<1)} + \varepsilon_t$
NAR	$y_t = (0.7\lvert y_{t-1}\rvert)/(\lvert y_{t-1}\rvert + 2) + \varepsilon_t$
SQ	$y_t = x_t^2 + e_t$ where $x_t = 0.6x_{t-1} + \varepsilon_t$
	$e_t \sim \mathrm{N}(0, \sigma^2)$, $\sigma^2 = 1, 25, 400$
EXP	$y_t = \exp(x_t) + e_t$ where $x_t = 0.6x_{t-1} + \varepsilon_t$
	$e_t \sim \mathrm{N}(0, \sigma^2)$, $\sigma^2 = 1, 25, 400$
Block 2	
Model 1	$y_t = \varepsilon_t - 0.4\varepsilon_{t-1} + 0.3\varepsilon_{t-2}$
Model 2	$y_t = \varepsilon_t - 0.4\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + 0.5\varepsilon_t\varepsilon_{t-2}$
Model 3	$y_t = \varepsilon_t - 0.3\varepsilon_{t-1} + 0.2\varepsilon_{t-2} + 0.4\varepsilon_{t-1}\varepsilon_{t-2} - 0.25\varepsilon_{t-2}^2$
Model 4	$y_t = 0.4y_{t-1} - 0.3y_{t-2} + \varepsilon_t$
Model 5	$y_t = 0.4y_{t-1} - 0.3y_{t-2} + 0.5y_{t-1}\varepsilon_{t-1} + \varepsilon_t$
Model 6	$y_t = 0.4y_{t-1} - 0.3y_{t-2} + 0.5y_{t-1}\varepsilon_{t-1} + 0.8\varepsilon_{t-1} + \varepsilon_t$
Block 3	
LSTAR	$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2})F(y_{t-1}) + v_t$
	$F(y_{t-1}) = [1 + \exp(-100(y_{t-1} - 0.02))]^{-1}$
	$v_t \sim \mathrm{N}(0, \sigma^2)$, $\sigma^2 = 0.02^2$
ESTAR	$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (-0.9y_{t-1} + 0.795y_{t-2})F(y_{t-1}) + v_t$
	$F(y_{t-1}) = 1 - \exp(-4000(y_{t-1})^2)$
	$v_t \sim \mathrm{N}(0, \sigma^2)$, $\sigma^2 = 0.01^2$
NN	$y_t = -1 + [1 + \exp(-100(y_{t-1} - 0.8y_{t-2}))]^{-1}$
	$+ [1 + \exp(-100(y_{t-1} + 0.8y_{t-2}))]^{-1} + v_t$
	$v_t \sim \mathrm{N}(0, \sigma^2)$, $\sigma^2 = 0.05^2$
BN	$y_t = -1 + [1 + \exp(-100(y_{t-1} - x_t))]^{-1} + [1 + \exp(-100(y_{t-1} + x_t))]^{-1} + v_t$
	$x_t = 0.8x_{t-1} + u_t$
	$v_t \sim \mathrm{N}(0, \sigma^2)$, $\sigma^2 = 0.05^2$, $u_t \sim \mathrm{N}(0, \sigma_u^2)$, $\sigma_u^2 = 0.05^2$

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.

5.2. *Results on size and power properties*

In Tables 2–5, we report the results of the simulation study. We compare the Hamilton statistic with the three types of Lagrange multiplier tests proposed in Section 4. Let $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$ denote the Hamilton's Lagrange multiplier statistic based on the Hessian representation of the information matrix and on the spherical variance-covariance matrix; $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$ be the test based on the outer product of the score and on the variance-covariance function (9); $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ be the test based on a higher order Taylor approximation to the variance-covariance function; and g_{OP} be the statistic proposed in Section 4.4.

Table 2
Critical values at 5% significance level and asymptotic size of the test statistics for an AR(2) (Model 4)

Test	$T=50$	$T=100$	$T=200$
$\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$	3.073	3.289	4.038
	(3.84)	(3.84)	(3.84)
	[0.079]	[0.069]	[0.045]
$\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$	3.621	3.335	3.070
	(3.84)	(3.84)	(3.84)
	[0.057]	[0.067]	[0.079]
$\lambda_{\mathrm{OP}}^{\mathrm{A}}$	23.434	24.163	25.196
	(24.99)	(24.99)	(24.99)
	[0.075]	[0.062]	[0.047]
g_{OP}	16.015	20.582	18.501
	(23.69)	(23.69)	(23.69)
	[0.312]	[0.113]	[0.185]

The first number is the simulated critical value. The number in parenthesis is the asymptotic critical value. The number in brackets is the asymptotic size of the statistic based on the simulated critical value. Number of replications is 1000. The tests $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$ and $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$ have one degree of freedom. The test $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ has 15 degrees of freedom including up to four powers of $\boldsymbol{r}_{ts}$. The test g_{OP} has 14 degrees of freedom because the auxiliary regression has been augmented to include up to four powers of $\tilde{\boldsymbol{r}}_{ts}$.

For comparison reasons, we also consider the Tsay's linearity test (Tsay, 1986) that has good power properties and is based on a second order Taylor approximation of an unknown, but in contrast to the statistics proposed in this paper, deterministic functional form. Briefly sketched, the Tsay test is computed by the following procedure: (1) Estimate the model under the null and compute $\hat{\varepsilon}=\boldsymbol{y}-(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y})$ and $SSR_0=\hat{\varepsilon}'\hat{\varepsilon}$. (2) Obtain the least squares residuals of $\boldsymbol{v}$—denoted $\hat{\boldsymbol{v}}$—from the auxiliary regression of $\hat{\varepsilon}$ on $\boldsymbol{X}$ and on m auxiliary regressors given by powers and cross products of the regressors in $\boldsymbol{X}$; for instance, if we have two regressors in $\boldsymbol{X}$, say y_{t-1} and y_{t-2}, the auxiliary regression will consist of regressing $\hat{\varepsilon}$ on $y_{t-1}, y_{t-2}, y_{t-1}^2, y_{t-1}y_{t-2}$, and y_{t-2}^2. Calculate $SSR=\hat{\boldsymbol{v}}'\hat{\boldsymbol{v}}$. (3) Compute the F-statistic $F=[(T-k-m)/m](SSR_0-SSR)/SSR$ that is distributed as an $F(m,T-k-m)$ where k is the number of regressors considered in $\boldsymbol{X}$.

In Table 2, we report the critical values and the asymptotic size of the proposed statistics when the model under the null is an AR(2). The specification is that of Model 4 in Table 1. The λ-tests have a better size than the g-tests. The $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$ and $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ have a size very close to the nominal size of 5% for sample sizes of 200. The g-test is undersized; the simulated critical value is smaller than the asymptotic critical value, which implies that the test will reject less often the true null hypothesis. However, the bootstrapped size is very close to 5%, as we show in the forthcoming tables.

In Tables 3–5, we report the bootstrapped rejection frequencies of the statistics. We observe that the bootstrap procedure increases the size of the g-tests substantially. For instance, in Table 4, for Model 4, which is an AR(2), the bootstrapped size is 4.9% for

Table 3
Power of the test statistics. Rejection frequencies for bootstrapped p-values at the 5% significance level. 1000 replications and 100 resamples

Test	AR	BL	TAR	SGN	NAR	SQ	EXP
$\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$							
$T=50$	5.2	10.7	57.8	70.5	7.7	100.0	96.3
$T=100$	4.4	19.0	93.7	97.9	11.1	100.0	100.0
$T=200$	4.5	24.9	99.9	100.0	16.7	100.0	100.0
$\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$							
$T=50$	5.0	3.5	39.6	51.3	5.0	99.2	91.0
$T=100$	4.7	6.5	86.9	93.0	7.3	100.0	99.9
$T=200$	6.3	8.2	99.8	100.0	9.6	100.0	100.0
$\lambda_{\mathrm{OP}}^{\mathrm{A}}$							
$T=50$	4.5	23.7	52.5	66.0	6.6	100.0	98.3
$T=100$	3.6	37.7	90.9	96.6	9.8	100.0	100.0
$T=200$	4.7	53.0	99.9	100.0	16.8	100.0	100.0
g_{OP}							
$T=50$	3.9	31.5	9.9	45.8	5.5	100.0	97.0
$T=100$	2.6	49.8	31.8	92.2	7.7	100.0	100.0
$T=200$	3.9	70.6	78.8	100.0	11.1	100.0	100.0
Tsay							
$T=50$	3.1	16.8	3.1	11.1	10.3	100.0	98.9
$T=100$	3.2	31.2	5.7	13.9	12.6	100.0	100.0
$T=200$	4.6	39.9	4.7	14.5	21.2	100.0	100.0

sample sizes of 200. The Monte Carlo bootstrap procedure that we have implemented consists of the following steps: (1) Estimate the model under the null hypothesis of linearity, i.e. $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\varepsilon$, obtain $\{\hat{\boldsymbol{\beta}}', \hat{\sigma}^2\}$ and compute the test statistic of interest, say $\mathscr{T}=\mathscr{T}(\boldsymbol{y},\boldsymbol{X})$. (2) Resample $\{\hat{\varepsilon}\}$ from the distribution $\mathrm{N}(0,\hat{\sigma}^2)$ and construct $\hat{\boldsymbol{y}}$ based on the resampled residuals. (3) Compute $\mathscr{T}_1^*=\mathscr{T}_1^*(\hat{\boldsymbol{y}},\boldsymbol{X})$. (4) Repeat steps 2 and 3 up to R times and obtain $\{\mathscr{T}_r^*\}_{r=2}^R$. (5) The bootstrapped p-value is equal to $\Sigma_{r=1}^R 1(\mathscr{T}_r^* \geqslant \mathscr{T})/R$. For each Monte Carlo replication, we compute the bootstrapped p-value of step 5, and the bootstrapped rejection frequency is the percentage number of replications in which the bootstrapped p-value is less than 5%.

In Table 3, we report the bootstrapped rejection frequencies for the models in Block 1. We observe that all tests are extremely powerful when the nonlinearity does not included lagged endogenous variables, such as the SQ and EXP models. For bilinear models, the g_{OP} test is the most powerful, for instance, for a sample size of 200, the power of g_{OP} is 71% while the Tsay test has 40% power, and the Hamilton statistic 25%. For TAR and SGN models, the $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$ and $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ tests are the best performers with 100% power for samples of 200. For NAR models, the tests have, in general, low power; the Tsay test, with 20% power for a sample size of 200, is comparable

Table 4
Power of the test statistics. Rejection frequencies for bootstrapped p-values at the 5% significance level. 1000 replications and 100 resamples

Test	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$						
$T=50$	5.5	6.5	36.6	4.1	60.2	40.4
$T=100$	5.1	8.3	68.4	4.8	94.0	74.2
$T=200$	5.7	13.8	93.6	4.5	99.9	97.7
$\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$						
$T=50$	4.7	5.4	7.0	4.9	13.9	7.0
$T=100$	4.8	4.3	15.5	6.0	50.7	28.5
$T=200$	5.3	6.8	48.0	5.1	95.3	79.0
$\lambda_{\mathrm{OP}}^{\mathrm{A}}$						
$T=50$	4.2	5.1	39.4	3.5	67.9	46.2
$T=100$	4.2	10.4	72.9	4.8	97.2	78.9
$T=200$	5.4	15.0	96.8	4.7	99.9	97.9
g_{OP}						
$T=50$	5.2	15.4	47.1	3.5	49.9	56.3
$T=100$	3.6	22.3	79.8	3.4	89.8	78.6
$T=200$	4.9	28.6	98.4	4.7	99.8	95.9
Tsay						
$T=50$	3.6	10.5	48.8	3.5	74.6	54.5
$T=100$	5.2	11.6	82.9	4.2	94.1	76.9
$T=200$	6.5	20.4	98.8	5.2	99.0	90.6

to the $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$ and $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ tests. In Table 4, we consider the power of the tests for the models in Block 2, which exhibit higher dynamics than those of Block 1. Models 1 and 4 are linear models and the bootstrapped power is close to the nominal size of 5%, as we should expect. Model 2 is also linear but has heteroscedastic errors. In this case, the g_{OP} and the Tsay tests seem to reject linearity more often than the remaining tests $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$, $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$, and $\lambda_{\mathrm{OP}}^{\mathrm{A}}$. In particular, $\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$ seems to discriminate very well between nonlinearity and conditional heteroscedasticity, for instance, for a sample size of 200, the power is around 7%, which is close to the nominal size of 5%. The g_{OP} and the Tsay tests are the best performers on detecting the nonlinear MA specification of Model 3 with almost 100% power. These two tests together with $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ and $\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$, are able to detect bilinear ARMA models (Models 5 and 6) very strongly. In Table 5, we report the bootstrapped power for models in Block 3. The five tests are powerful to detect LSTAR specifications, with the Tsay and $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ tests being the best performers. However, when the model is the ESTAR, the $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ test outperforms, by a wide margin, the remaining tests; for instance, for a sample size of 200, the power of $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ is 94% while the Tsay test has 10% power and the Hamilton statistic 71%. For the NN models, there is a superior performance of the λ-tests over the Tsay test. The former have 100% power while the Tsay test has between 10% and 20% power.

Table 5
Power of the test statistics. Rejection frequencies for bootstrapped p-values at the 5% significance level. 1000 replications and 100 resamples

Test	LSTAR	ESTAR	NN	BN
$\lambda_{\mathrm{H}}^{\mathrm{E}}(\boldsymbol{g})$				
$T=50$	27.7	16.8	54.3	54.3
$T=100$	63.8	33.8	89.4	90.5
$T=200$	97.2	71.3	99.9	99.9
$\lambda_{\mathrm{OP}}^{\mathrm{E}}(\boldsymbol{g})$				
$T=50$	8.2	7.0	34.7	28.8
$T=100$	39.1	19.5	84.3	83.2
$T=200$	90.7	59.5	99.9	100.0
$\lambda_{\mathrm{OP}}^{\mathrm{A}}$				
$T=50$	55.0	26.2	41.6	47.3
$T=100$	93.9	58.2	84.5	92.0
$T=200$	99.9	94.1	99.9	100.0
g_{OP}				
$T=50$	31.5	11.5	14.8	19.8
$T=100$	64.8	22.9	41.7	52.4
$T=200$	96.9	55.6	87.9	93.1
Tsay				
$T=50$	62.1	7.9	9.0	13.8
$T=100$	93.7	7.4	9.7	15.8
$T=200$	99.9	9.5	9.9	16.8

Overall, three clear results emerge. First, the g_{OP} tests seem a promising alternative to existing tests when the alternative is a bilinear type model, most tests have difficulties identifying nonlinearities of this type, e.g. Lee et al. (1993). Second, the $\lambda_{\mathrm{OP}}^{\mathrm{A}}$ 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 λ-tests are superior on detecting neural network specifications. In addition, the λ and g-tests match the simplicity of the Tsay statistic.

6. Empirical illustration

We apply the proposed tests for neglected nonlinearity to the industrial production (IP) index of sixteen OECD countries and Europe, the latter is an aggregate of European countries that are OECD members. The IP is considered a leading indicator of the business cycle. The data is seasonally adjusted with quarterly frequency from 1960-Q1 to 2001-Q3, for a total of 167 observations. We analyze the quarterly growth rate of the index. The results of the linearity testing are summarized in Table 6.

Table 6
Bootstrapped p-values of the tests for neglected nonlinearity for the quarterly growth rate of the Industrial Production Index. 1960Q1–2001Q3. Seasonally adjusted data. 1000 resamples

Country	Lags	$\lambda^{E}_{H}(\boldsymbol{g})$	$\lambda^{E}_{OP}(\boldsymbol{g})$	λ^{A}_{OP}	g_{OP}	Diagnosis
USA	8	0.109	0.822	0.018	0.016	Nonlinear
Japan	8	0.506	0.703	0.223	0.188	Linear
Austria	5	0.176	0.284	0.344	0.248	Linear
Belgium	6	0.336	0.600	0.275	0.393	Linear
Finland	5	0.417	0.870	0.014	0.001	Nonlinear
France	5	0.102	0.911	0.001	0.001	Nonlinear
Germany	5	0.317	0.981	0.004	0.009	Nonlinear
Greece	5	0.577	0.730	0.898	0.469	Linear
Ireland	5	0.824	0.418	0.366	0.026	?
Italy	5	0.061	0.878	0.021	0.001	Nonlinear
Holland	5	0.731	0.847	0.196	0.407	Linear
Norway	5	0.016	0.082	0.002	0.002	Nonlinear
Portugal	5	0.695	0.567	0.238	0.001	?
Spain	5	0.533	0.510	0.044	0.050	Nonlinear
Sweden	5	0.230	0.723	0.500	0.675	Linear
G.Britain	5	0.412	0.679	0.002	0.003	Nonlinear
Europe	5	0.655	0.674	0.767	0.969	Linear

We need to choose the number of lags in the linear and in the nonlinear components of the model. For the linear component, the Akaike Information Criterium selects between one (Finland, France, and Great Britain) and eight (USA and Japan) lags. Conservatively, a moderate number of lags is recommended to guard against dynamic misspecification. We use the same number of lags in the nonlinear part of the model. When we have a large number of regressors in the nonlinear part of the model, the implementation of the auxiliary regression required in the test λ^{A}_{OP}, and described in Section 4.3, may be difficult because, for moderate sample sizes, we run out of degrees of freedom. In this case, we constrain the auxiliary regression to second order terms, or alternatively, one may remove the terms that involve cross-products and leave the high power terms. For all four tests, we have bootstrapped their distribution and reported the bootstrapped p-value, as described in Section 5.2. The most striking feature of the results presented in Table 6 is the different diagnosis provided by the tests $\lambda^{E}_{H}(\boldsymbol{g})$ and $\lambda^{E}_{OP}(\boldsymbol{g})$ in one hand, and the tests λ^{A}_{OP} and g_{OP} on the other. In sixteen out of the all seventeen series, $\lambda^{E}_{H}(\boldsymbol{g})$ and $\lambda^{E}_{OP}(\boldsymbol{g})$ fail to reject linearity. This is a reflection of the nuisance parameter problem that we argue in this paper. When the number of nuisance parameters grows, the tests $\lambda^{E}_{H}(\boldsymbol{g})$ and $\lambda^{E}_{OP}(\boldsymbol{g})$ tend to loose power; on the contrary, the tests λ^{A}_{OP} and g_{OP} are immune to the number of nuisance parameters and have higher power. The tests λ^{A}_{OP} and g_{OP} deliver the same diagnosis in fifteen series, the exception being Ireland and Portugal where λ^{A}_{OP} fails to reject linearity and g_{OP} rejects it. In the seventeen series, we find nonlinearity in eight, we fail to reject linearity in seven, and we find mixed results in two. A similar data set was used in Teräsvirta and Anderson (1992), where the authors tested the null hypothesis of linearity against LSTAR and

ESTAR specifications. Our results, although not strictly comparable because of the time period and the seasonality adjustment, are roughly in agreement with those of Teräsvirta and Anderson (1992). In general, they reject linearity in favor of a LSTAR specification more often than we do, as for instance, in the cases of Japan, Austria, Belgium, and Sweden. However, we have shown in our Monte Carlo study that our test λ_{OP}^{A} has power of almost 100% for sample sizes of 200 against either LSTAR or ESTAR models. A reason that can explain the different diagnosis may be the potential misspecification of the number of lags in the nonlinear component of the model.

7. Conclusion

Within a flexible regression model (Hamilton, 2001), we have offered a battery of Lagrange multiplier tests for neglected nonlinearities. We have classified the tests in two categories: λ-tests and g-tests. The statistic proposed by Hamilton is a λ-test where, under the null hypothesis of linearity, the number of unidentified nuisance parameters increases with the dimensionality of the model and where the covariance function that defines the random field is spherical, which is based on the L_2 norm. The presence of unidentified nuisance parameters may have adverse consequences on the power of the test, and the choice of a particular covariance function may result on a test statistic that has power against a certain class of random functions. The tests that we have proposed aim to generalize and complement the Hamilton statistic. They are free of unidentified nuisance parameters under the null hypothesis of linearity and they are able to capture a broad class of covariance functions. These advantages are the result of two new directions. First, we have switched from the L_2 norm to the L_1 norm, and second, we have assumed that the random field is sufficiently smooth for its covariance function to be locally approximated by a higher order Taylor expansion.

The payoff of this strategy has been confirmed in an extensive Monte Carlo study. The simulation evidence suggests that for small and moderate sample sizes, the proposed statistics not only match the power properties of Hamilton test but also have superior power against bilinear models, neural network, and smooth transition autoregressive specifications. We have also compared our tests with the Tsay's statistic since this is also based on a Taylor expansion of an unknown but deterministic functional form. Compared to the Tsay's test, the λ-tests and g-tests have superior power properties against specifications such as threshold autoregressive, sign autoregressive, exponential smooth transition, and neural networks.

Finally, we have offered an application of our tests to detect nonlinearities in the Industrial Production Index of sixteen OECD countries. Teräsvirta and Anderson (1992) studied the same index, testing linearity against exponential and logistic smooth transition autoregressive models. Our tests are ideal candidates to assess linearity given their superior performance to detect smooth transition dynamics. Broadly speaking, our results agree with those of Teräsvirta and Anderson, although they found more nonlinearity than we did; for instance, in Japan, Austria, Belgium, and Sweden, they rejected a linear model in favor of a logistic smooth transition, but our tests seem to indicate that linearity cannot be rejected. This empirical application also illustrates that a large

number of nuisance parameters have a pernicious effect on the power of those tests that fix the unidentified parameters to certain values.

Acknowledgements

We have benefited from numerous discussions and comments from Jim Hamilton and 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, University of California, Davis, and the Econometric Society Meetings in Boston. We are also grateful to four anonymous referees and the Associate Editor for their constructive comments. All errors remain our responsibility. González-Rivera acknowledges the financial support of the Academic Senate of the University of California, Riverside. The scientific notation follows the standard proposed by Abadir and Magnus (2002).

References

Abadir, K., Magnus, J., 2002. Notation in econometrics: a proposal for a standard. Econometrics Journal 5, 76–90.

Adler, R.J., 1981. The Geometry of Random Fields. Wiley, New York.

Ashley, R., Patterson, D., Hinich, M., 1986. A diagnostic test for nonlinear serial dependence in time series fitting errors. Journal of Time Series Analysis 7, 165–178.

Bochner, S., 1959. Lectures on Fourier Integrals. Princeton University Press, New Jersey.

Chan, W.S., Tong, H., 1986. On tests for nonlinearity in time series models. Journal of Forecasting 5, 217–228.

Christakos, G., 1992. Random Field Models in Earth Sciences. Academic Press Inc., San Diego, CA.

Dahl, C., 2002. An investigation of tests for linearity and the accuracy of likelihood based inference using random fields. Econometrics Journal 1, 1–25.

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., 2001. A parametric approach to flexible nonlinear inference. Econometrica 69 (3), 537–573.

Hansen, B.E., 1996. 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. Biometrica 72, 39–44.

Khinchin, A.Y., 1934. Korrelationstheorie der stationaren stochastischen Prozesse. Math. Ann. 109, 604–615.

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., Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75, 491–499.

Matheron, G., 1973. The intrinsic random functions and their applications. Advances in Applied Probability 5, 439–468.

Teräsvirta, T., 1994. Specification, estimation and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89 (425), 208–218.

Teräsvirta, T., 1998. Modelling economic relationships with smooth transition regressions. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York.

Teräsvirta, T., Anderson, H., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. Journal of Applied Econometrics 7, 119–139.

Teräsvirta, 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, New York, NY.

Tsay, R.S., 1986. Nonlinearity test for time series. Biometrika 73, 461–466.

Wahba, G., 1978. Improper priors, spline smoothing and the problem of guarding against model errors in regression. Journal of the Royal Statistical Society Series B 40, 364–372.

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, NJ.

Yaglom, A.M., 1987. Correlation Theory of Stationary and Related Random Functions. Springer, New York, NY.

Zastavnyi, V.P., 2000. On positive definiteness of some functions. Journal of Multivariate Analysis 73, 55–81.