Studies in Nonlinear Dynamics & Econometrics

Volume 4, *Issue* 4 2000 *Article* 1

Neural Network Test and Nonparametric Kernel Test for Neglected Nonlinearity in Regression Models

Tae-Hwy Lee
University of California, Riverside

http://www.bepress.com/snde

ISSN: 1081-1826

This volume was previously published by MIT Press.

Neural Network Test and Nonparametric Kernel Test for Neglected Nonlinearity in Regression Models

Tae-Hwy Lee
Department of Economics
University of California, Riverside
taelee@mail.ucr.edu

Abstract. *This article considers two conditional moment tests for neglected nonlinearity in regression models and examines their finite sample performance. The two tests are the nonparametric kernel test by Li and Wang (1998) and Zheng (1996) and the neural network test of White (1989). The article examines an asymptotic test, a naive bootstrap test, and a wild bootstrap test for weakly dependent time series and independent data.*

Keywords. asymptotic test, conditional bootstrap, naive bootstrap, recursive bootstrap, wild bootstrap

Acknowledgments. I would like to thank Qi Li for his GAUSS program, two anoymous referees and Aman Ullah for comments, and University of California Regents Faculty Fellowship and Faculty Development Awards for research support.

1 Introduction

This article explores the issues in testing for functional forms, especially for neglected nonlinearity in parametric linear models. Many articles have appeared in the recent literature that deal with the issues of how to carry out various specification tests in parametric regression models. To construct the tests, various methods are used to estimate the alternative models. For example, Fan and Li (1996), Li and Wang (1998), Zheng (1996), and Bradley and McClelland (1996) use local constant kernel regression; Hjellvik, Yao, and Tjøstheim (1998) use local polynomial kernel regression; Cai, Fan, and Yao (2000) and Matsuda (1999) use nonparametric functional coefficient models; Poggi and Portier (1997) use a functional autoregressive model; White (1989) uses neural network models; Eubank and Spiegelman (1990) use spline regression; Hong and White (1995) use series regression; Stengos and Sun (1998) use wavelet methods; and Hamilton (2001) uses a flexible parametric regression model.

There are also many articles that compare different approaches in testing for linearity. For example, Lee, White, and Granger (1993), Teräsvirta, Lin, and Granger (1993), and Teräsvirta (1996) examine the neural network test of White (1989) and many other tests. Dahl (1999) and Dahl and González-Rivera (2000) study Hamilton's (forthcoming) test and compare it with various tests, including the neural network test. Blake and Kapetanios (1999, 2000) extend the neural network test using a radial basis function for the neural network activation function instead of using the typical logistic function employed in Lee, White, and Granger (1993).[1]

[1]For radial basis functions, see, e.g., Campbell, Lo, and MacKinlay 1997 (p. 517).

Lee and Ullah (forthcoming-a, forthcoming-b) examine the tests of Li and Wang (1998), Zheng (1996), Ullah (1985), Cai, Fan, and Yao (2000), Härdle and Mammen (1993), and Aït-Sahalia, Bickel, and Stoker (1994). Fan and Li (forthcoming) compare the tests of Li and Wang (1998), Zheng (1996), and Bierens (1990). Whang (2000) generalizes the Kolmogorov-Smirnov and Cramer–von Mises tests to the regression framework and compares them with the tests of Härdle and Mammen (1993) and Bierens and Ploberger (1997). Hjellvik and Tjøstheim (1995, 1996) propose tests based on nonparametric estimates of conditional mean and variances and compare them with a number of tests, such as the bispectrum test and the Brock, Deckert, and Scheinkman (BDS) test.

This article investigates and compares the kernel-based test of Li and Wang (1998) and Zheng (1996) (hereafter, LWZ) and the neural network test (hereafter, NN). Both LWZ and NN tests are conditional moment tests whose null hypothesis consists of conditional moment conditions that hold if the linear model is correctly specified for the conditional mean. The two tests differ in the choice of "test functions" that are to be checked for their correlation with the residuals from the linear regression model. This article examines asymptotic tests, the naive bootstrap test, and the wild bootstrap test for weakly dependent time series and independent series. The bootstrap for time series data is implemented in two ways (termed the conditional bootstrap and the recursive bootstrap, as discussed in section 3.2) and perhaps surprisingly, we find that the conditional bootstrap is more reliable than the recursive bootstrap. The size performance of these tests under the presence of conditional heteroskedasticity (of GARCH form) is also examined.

The plan of the article is as follows. In Section 2, based on nonparametric kernel regression and neural network models, the LWZ test and NN test are discussed. In Section 3, the bootstrap procedures and their performance for these tests are examined via a Monte Carlo experiment. Section 4 gives conclusions.

2 Testing for Linearity

Let $\{\boldsymbol{Z}_t\}_{t=1}^n$ be a stochastic process, and partition $\boldsymbol{Z}_t$ as $\boldsymbol{Z}_t = (y_t\ \boldsymbol{x}_t)$, where y_t is a scalar and $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tk})$. $\boldsymbol{x}_t$ may (but need not necessarily) contain a constant and lagged values of y_t. Consider the regression model

$$y_t = m(\boldsymbol{x}_t) + \varepsilon_t \tag{1}$$

where $m(\boldsymbol{x}_t) \equiv E(y_t \mid \boldsymbol{x}_t)$ is the true but unknown regression function and ε_t is the error term such that $E(\varepsilon_t \mid \boldsymbol{x}_t) = 0$ by construction. To test for a parametric model $g(\boldsymbol{x}_t, \boldsymbol{\beta})$ we consider

$$H_0\ :\ m(\boldsymbol{x}_t) = g(\boldsymbol{x}_t, \boldsymbol{\beta}^*) \text{ almost everywhere (a.e.) for some } \boldsymbol{\beta}^* \in \mathbb{R}^k \tag{2}$$

$$H_1\ :\ m(\boldsymbol{x}_t) \neq g(\boldsymbol{x}_t, \boldsymbol{\beta}) \text{ on a set with positive measure for all } \boldsymbol{\beta} \in \mathbb{R}^k \tag{3}$$

In particular, if we are to test for neglected nonlinearity in the regression models, set $g(\boldsymbol{x}_t, \boldsymbol{\beta}) = \boldsymbol{x}_t\boldsymbol{\beta}$. Then under H_0, the process $\{y_t\}$ is linear in mean, conditional on x_t, that is,

$$H_0\!:\ m(\boldsymbol{x}_t) = \boldsymbol{x}_t\boldsymbol{\beta}^* \text{ a.e. for some } \boldsymbol{\beta}^* \in \mathbb{R}^k \tag{4}$$

The alternative of interest is the negation of the null, that is,

$$H_1\!:\ m(\boldsymbol{x}_t) \neq \boldsymbol{x}_t\boldsymbol{\beta} \text{ on a set with positive measure for all } \boldsymbol{\beta} \in \mathbb{R}^k \tag{5}$$

When the alternative is true, a linear model is said to suffer from "neglected nonlinearity" (Lee, White, and Granger 1993).

If a linear model is capable of an exact representation of the unknown function $m(\boldsymbol{x}_t)$, then there exists a vector $\boldsymbol{\beta}^*$ such that Equation (4) holds, which implies

$$E(\varepsilon_t^* \mid \boldsymbol{x}_t) = 0 \text{ a.e.} \tag{6}$$

where $\varepsilon_t^* = y_t - \boldsymbol{x}_t\boldsymbol{\beta}^*$. By the law of iterated expectations ε_t^* is uncorrelated with any measurable functions of $\boldsymbol{x}_t$, say $h(\boldsymbol{x}_t)$. That is,

$$E[h(\boldsymbol{x}_t)\varepsilon_t^*] = 0 \tag{7}$$

Depending on how we use these moment conditions and the function $h(\cdot)$, various specification tests may be considered. The specification tests based on these moment conditions, so-called conditional moment tests, have been studied by Newey (1985), Tauchen (1985), White (1987, 1994), Bierens (1990), Bierens and Ploberger (1997), and Stinchcombe and White (1998), among others. The neural network test exploits Equation (7), with $h(\cdot)$ being chosen as the neural network hidden unit activation function. LWZ's nonparametric kernel test utilizes Equation (7), with $h(\cdot)$ being chosen as $E(\varepsilon_t^* \mid \boldsymbol{x}_t)f(\boldsymbol{x}_t)$, where $f(\boldsymbol{x}_t)$ is the density of $\boldsymbol{x}_t$. We now turn to details of these two tests.

2.1 Nonparametric kernel test

If H_0 is true, i.e., $g(\boldsymbol{x}_t, \boldsymbol{\beta}) = \boldsymbol{x}_t\boldsymbol{\beta}$ is a correctly specified family of parametric regression functions, one can construct a consistent least-squares (LS) estimator of $m(\boldsymbol{x}_t)$ given by $\boldsymbol{x}_t\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is the LS estimator of the parameter $\boldsymbol{\beta}$, obtained by minimizing $\sum \varepsilon_t^2 = \sum (y_t - \boldsymbol{x}_t\boldsymbol{\beta})^2$ with respect to $\boldsymbol{\beta}$. The LS estimator is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, where X is an $n \times k$ matrix with $\boldsymbol{x}_t$ in its tth row. If H_0 is not true, then an alternative model is to use the nonparametric regression estimation of the unknown $m(\boldsymbol{x}_t)$. This article considers nonparametric kernel regression and neural network regression.

A kernel estimator is a local LS (LLS) estimator obtained by minimizing $\sum \varepsilon_t^2 K(\frac{\boldsymbol{x}_t - \boldsymbol{x}}{h})$ where $\varepsilon_t = y_t - g(\boldsymbol{x}_t, \boldsymbol{\beta})$, $K_t = K(\frac{\boldsymbol{x}_t - \boldsymbol{x}}{h})$ is a decreasing function of the distances of the regressor vector $\boldsymbol{x}_t$ from the point $\boldsymbol{x} = (x_1, \ldots, x_k)$, and $h > 0$ is the width of the window that determines how rapidly the weights decrease as the distance of $\boldsymbol{x}_t$ from $\boldsymbol{x}$ increases. For example, when $g(\boldsymbol{x}_t, \boldsymbol{\beta}) = \boldsymbol{x}_t\boldsymbol{\beta}(\boldsymbol{x})$, an explicit expression of the LLS estimator of $\boldsymbol{\beta}$ is

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{x}) = (\boldsymbol{X}'K(\boldsymbol{x})\boldsymbol{X})^{-1}\boldsymbol{X}'K(\boldsymbol{x})\boldsymbol{y} \tag{8}$$

where $K(\boldsymbol{x})$ is the diagonal matrix with the diagonal elements $(K(\frac{\boldsymbol{x}_t - \boldsymbol{x}}{h}))$, $t = 1, \ldots, n$. The estimator $\tilde{\boldsymbol{\beta}}(x)$ is the local linear LS (LLLS) or simply the local linear (LL) estimator. (For more details, see Fan and Gijbels 1996 and Pagan and Ullah 1999.)

As $E(\varepsilon_t^* \mid \boldsymbol{x}_t) = 0$ a.e. under the null Equation (4), by the law of iterated expectations,

$$E[(\varepsilon_t^* E(\varepsilon_t^* \mid \boldsymbol{x}_t)] = E[E(\varepsilon_t^* \mid \boldsymbol{x}_t)^2] = 0 \tag{9}$$

if H_0 is true. Li and Wang (1998) and Zheng (1996) proposed a conditional moment test based on the density-weighted version of Equation (9) to avoid the random denominator problem that arises in nonparametric estimation: construct the test based on $E[\varepsilon_t^* E(\varepsilon_t^* \mid \boldsymbol{x}_t)f(\boldsymbol{x}_t)]$, where $f(\boldsymbol{x}_t)$ is the density function of $\boldsymbol{x}_t$. This is estimated by

$$\begin{aligned} L' &= \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t E(\hat{\varepsilon}_t \mid \boldsymbol{x}_t)\hat{f}(\boldsymbol{x}_t) \\ &= \frac{1}{n(n-1)h^k}\sum_{t=1}^{n}\sum_{t'=1, t'\neq t}^{n} \hat{\varepsilon}_t\hat{\varepsilon}_{t'}K_{t't} \end{aligned} \tag{10}$$

where $\hat{\varepsilon}_t = y_t - \boldsymbol{x}_t\hat{\boldsymbol{\beta}}$, $E(\hat{\varepsilon}_t \mid \boldsymbol{x}_t) = \sum_{t'\neq t}\hat{\varepsilon}_{t'}K_{t't}/\sum_{t'\neq t}K_{t't}$ and $\hat{f}(\boldsymbol{x}_t) = [(n-1)h^k]^{-1}\sum_{t'\neq t}K_{t't}$ is the kernel density estimator; $K_{t't} = K(\frac{\boldsymbol{x}_{t'} - \boldsymbol{x}_t}{h})$. Under the assumptions stated in Li 1999 (p. 107), the asymptotic test statistic is then given by

$$L = nh^{k/2}\frac{L'}{\hat{\sigma}} \xrightarrow{d} N(0, 1) \tag{11}$$

where $\hat{\sigma}^2 = 2(n(n-1)h^k)^{-1}\sum_t\sum_{t'\neq t}\hat{\varepsilon}_t^2\hat{\varepsilon}_{t'}^2K_{t't}^2$ is a consistent estimator of the asymptotic variance of $nh^{k/2}L'$. See Zheng 1996, Fan and Li 1996, and Li and Wang 1998 for details.

2.2 Neural network test

Another alternative model we consider is an augmented single-hidden-layer feedforward neural network model in which network output y_t is determined, given input $\boldsymbol{x}_t$, as

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta} + \sum_{j=1}^{q} \delta_j \psi(\boldsymbol{x}_t\boldsymbol{\gamma}_j) + \varepsilon_t \tag{12}$$

where $\boldsymbol{\beta}$ is a conformable column vector of connection strength from the input layer to the output layer; $\boldsymbol{\gamma}_j$ is a conformable column vector of connection strength from the input layer to the hidden units, $j = 1, \ldots, q$; δ_j is the (scalar) connection strength from the hidden unit j to the output unit, $j = 1, \ldots, q$; and ψ is a squashing function (e.g., the logistic squasher) or a radial basis function. Input units x send signals to intermediate hidden units, then each hidden unit produces an activation ψ that then sends signals toward the output unit. The integer q denotes the number of hidden units added to the affine (linear) network. When $q = 0$, we have a two-layer *affine* network $y_t = \boldsymbol{x}_t\boldsymbol{\beta} + \varepsilon_t$. Hornik, Stinchcombe, and White (1989) show that a neural network is a nonlinear flexible functional form being capable of approximating any Borel-measurable function to any desired level of accuracy provided sufficiently many hidden units are available.

White (1989) developed a test for neglected nonlinearity likely to have power against a range of alternatives based on neural network models. (See also Lee, White, and Granger 1993 and Teräsvirta 1996 on the neural network test and its comparison with other specification tests.) The neural network test is based on a test function $h(\boldsymbol{x}_t)$ chosen as the activations of "phantom" hidden units $\psi(\boldsymbol{x}_t\boldsymbol{\Gamma}_j),\ j = 1, \ldots, q$, where $\boldsymbol{\Gamma}_j$ are random column vectors independent of $\boldsymbol{x}_t$. That is,

$$E[\psi(\boldsymbol{x}_t\boldsymbol{\Gamma}_j)\varepsilon_t^* \mid \boldsymbol{\Gamma}_j] = E[\psi(\boldsymbol{x}_t\boldsymbol{\Gamma}_j)\varepsilon_t^*] = 0, \quad j = 1, \ldots, q \tag{13}$$

under H_0, so that

$$E(\boldsymbol{\Psi}_t\varepsilon_t^*) = 0, \tag{14}$$

where $\boldsymbol{\Psi}_t = (\psi(\boldsymbol{x}_t\boldsymbol{\Gamma}_1), \ldots, \psi(\boldsymbol{x}_t\boldsymbol{\Gamma}_q))'$ is a phantom hidden unit activation vector. Evidence of correlation of ε_t^* with $\boldsymbol{\Psi}_t$ is evidence against the null hypothesis that y_t is linear in mean. If correlation exists, augmenting the linear network by including an additional hidden unit with activations $\psi(\boldsymbol{x}_t\boldsymbol{\Gamma}_j)$ would permit an improvement in network performance. Thus the tests are based on sample correlation of affine network errors with phantom hidden unit activations,

$$n^{-1}\sum_{t=1}^{n} \boldsymbol{\Psi}_t\hat{\varepsilon}_t = n^{-1}\sum_{t=1}^{n} \boldsymbol{\Psi}_t(y_t - \boldsymbol{x}_t\hat{\boldsymbol{\beta}}) \tag{15}$$

Under suitable regularity conditions it follows from the central limit theorem that $n^{-1/2}\sum_{t=1}^{n} \boldsymbol{\Psi}_t\hat{\varepsilon}_t \xrightarrow{d} N(0, W^*)$ as $n \to \infty$, and if one has a consistent estimator for its asymptotic covariance matrix, say $\hat{W}_n$, then an asymptotic chi-square statistic can be formed as

$$\left(n^{-1/2}\sum_{t=1}^{n} \boldsymbol{\Psi}_t\hat{\varepsilon}_t\right)' \hat{W}_n^{-1} \left(n^{-1/2}\sum_{t=1}^{n} \boldsymbol{\Psi}_t\hat{\varepsilon}_t\right) \xrightarrow{d} \chi^2(q). \tag{16}$$

Elements of $\boldsymbol{\Psi}_t$ tend to be collinear with X_t and with themselves, and computation of $\hat{W}_n$ can be tedious. Thus we conduct a test on $q^* < q$ principal components of $\boldsymbol{\Psi}_t$ not collinear with x_t, denoted $\boldsymbol{\Psi}_t^*$, and employ the equivalent test statistic that avoids explicit computation of $\hat{W}_n$, denoted N_{q,q^*},

$$N_{q,q^*} \equiv nR^2 \xrightarrow{d} \chi^2(q^*), \tag{17}$$

where R^2 is uncentered squared multiple correlation from a standard linear regression of $\hat{\varepsilon}_t$ on $\mathbf{\Psi}_t^*$ and $\boldsymbol{x}_t$. This test is to determine whether or not there exists some advantage to be gained by adding hidden units to the affine network.

It should be noted that the asymptotic equivalence of Equations (16) and (17) holds under the conditional homoskedasticity, $E(\varepsilon_t^* \mid \boldsymbol{x}_t) = \sigma^2$. Under the presence of conditional heteroskedasticity such as ARCH, N_{q,q^*} will not be $\chi^2(q^*)$-distributed. To resolve the problem in that case, we can use either Equation (16) with $\hat{W}_n$ being estimated robust to the conditional heteroskedasticity (White 1980 and Andrews 1991) or Equation (17) with the empirical null distribution of the statistic computed by a bootstrap procedure that is robust to the conditional heteroskedasticity. The latter is used in this article, specifically, the wild bootstrap.

3 Monte Carlo

The goal of this article is to examine the finite sample properties of these tests, especially with the empirical null distributions being generated by the bootstrap method. The LWZ test (denoted as L) and the NN test (denoted as N_{q,q^*}) are considered, for both of which both the naive bootstrap (Efron 1979) and the wild bootstrap (Wu 1986, Liu 1988) are used.

3.1 Data-generating processes

To generate data we use the following models, all of which have been employed in the related literature. (See Granger and Teräsvirta 1993 and Tong 1990.) There are four blocks. The error term ε_t below is always i.i.d. $N(0, 1)$. $\mathbf{1}(\cdot)$ is an indicator function that takes a value of one if its argument is true and zero otherwise. All data-generating processes (DGPs) below fulfill the regularity and moment conditions for the investigated testing procedures. (See Li 1999 [p. 107] for the LWZ tests and White 1994 [chapter 9] for the NN tests or other parametric conditional moment tests.)

Block 1 (Lee, White, and Granger 1993 and Teräsvirta 1996)

DGP 1.1 *Linear AR*

$$y_t = 0.6y_{t-1} + \varepsilon_t,$$

DGP 1.2 *Linear AR with GARCH*

$$\begin{aligned} y_t &= 0.6y_{t-1} + \varepsilon_t \\ h_t &\equiv E(\varepsilon_t^2 \mid y_{t-1}) = 0.01 + 0.3\,\varepsilon_{t-1}^2 + 0.69\,h_{t-1} \end{aligned}$$

DGP 1.3 *Bilinear*

$$y_t = 0.7y_{t-1}\varepsilon_{t-2} + \varepsilon_t$$

DGP 1.4 *Threshold autoregressive*

$$y_t = 0.9y_{t-1}\mathbf{1}(|y_{t-1}| \le 1) - 0.3y_{t-1}\mathbf{1}(|y_{t-1}| > 1) + \varepsilon_t$$

DGP 1.5 *Sign nonlinear autoregressive*

$$y_t = \mathbf{1}(y_{t-1} > 0) - \mathbf{1}(y_{t-1} < 0) + \varepsilon_t$$

DGP 1.6 *Rational nonlinear autoregressive*

$$y_t = \frac{0.7|y_{t-1}|}{|y_{t-1}| + 2} + \varepsilon_t$$

http://www.bepress.com/snde/vol4/iss4/art1

Block 2 (Lee, White, and Granger 1993)
DGP 2.1 *Linear MA(2)*

$$y_t = \varepsilon_t - 0.4\varepsilon_{t-1} + 0.3\varepsilon_{t-2}$$

DGP 2.2 *Heteroskedastic MA(2)*

$$y_t = \varepsilon_t - 0.4\varepsilon_{t-1} + 0.3\varepsilon_{t-2} + 0.5\varepsilon_t\varepsilon_{t-2}$$

DGP 2.3 *Nonlinear MA*

$$y_t = \varepsilon_t - 0.3\varepsilon_{t-1} + 0.2\varepsilon_{t-2} + 0.4\varepsilon_{t-1}\varepsilon_{t-2} - 0.25\varepsilon_{t-2}^2$$

DGP 2.4 *Linear AR(2)*

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

DGP 2.5 *Bilinear AR*

$$y_t = 0.4y_{t-1} - 0.3y_{t-2} + 0.5y_{t-1}\varepsilon_{t-1} + \varepsilon_t$$

DGP 2.6 *Bilinear ARMA*

$$y_t = 0.4y_{t-1} - 0.3y_{t-2} + 0.5y_{t-1}\varepsilon_{t-1} + 0.8\varepsilon_{t-1} + \varepsilon_t$$

Note that the forecastable part of DGP 2.2 is linear and the final term introduces heteroskedasticity.

Block 3 (Lee, White, and Granger 1993)
DGP 3.1 *Square*

$$y_t = z_t^2 + \sigma\varepsilon_t$$

DGP 3.2 *Exponential*

$$y_t = \exp(z_t) + \sigma\varepsilon_t$$

These are bivariate models where $\sigma = 5,\ z_t = 0.6z_{t-1} + e_t,\ e_t \sim N(0, 1)$, and e_t, ε_t are independent.

Block 4 (Zheng 1996)
Four models with $\boldsymbol{x}_t = (x_{t1}\ x_{t2})'$ are considered. Let z_{t1} and z_{t2} be independently drawn from $N(0, 1)$. Two regressors x_{t1} and x_{t2} are defined as $x_{t1} = z_{t1}$ and $x_{t2} = (z_{t1} + z_{t2})/\sqrt{2}$.
DGP 4.1 *Linear*

$$y_t = 1 + x_{t1} + x_{t2} + \varepsilon_t$$

DGP 4.2 *Quadratic*

$$y_t = 1 + x_{t1} + x_{t2} + x_{t1}x_{t2} + \varepsilon_t$$

DGP 4.3 *Concave*

$$y_t = (1 + x_{t1} + x_{t2})^{1/3} + \varepsilon_t$$

DGP 4.4 *Convex*

$$y_t = (1 + x_{t1} + x_{t2})^{5/3} + \varepsilon_t$$

3.2 Simulation design

For the simulations, the information set is $\boldsymbol{x}_t = y_{t-1}$ for Block 1, $\boldsymbol{x}_t = (y_{t-1}\ y_{t-2})'$ for Block 2, $\boldsymbol{x}_t = z_t$ for Block 3, and $\boldsymbol{x}_t = (x_{t1}\ x_{t2})'$ for Block 4.

For N_{q,q^*}, the logistic squasher $\psi = [1 + \exp(-\boldsymbol{x}'\boldsymbol{\gamma})]^{-1}$ is used, with $\boldsymbol{\gamma}$ being generated from the uniform distribution on $[-2, 2]$ and y_t, $\boldsymbol{x}_t$ being rescaled onto $[0, 1]$. A number of additional hidden units to the affine network q $(= 10, 20)$ are used. q^* $(= 1, 3, 5)$ largest principal components (excluding the first principal component) of these are chosen. The results are reported for $(q, q^*) = (10, 1), (10, 3), (20, 3)$, and $(20, 5)$.

For L, as in Li and Wang 1998 (p. 154), we use a standard normal kernel. Note that $\boldsymbol{x}_t$ is a $1 \times k$ vector, and $k = 1$ for Blocks 1, 3 and $k = 2$ for Blocks 2, 4. Thus the smoothing parameter h is chosen as $h_i = c\hat{\sigma}_i n^{-1/5}$ $(i = 1)$ for Blocks 1 and 3, and $h_i = c\hat{\sigma}_i n^{-1/6}$ $(i = 1, 2)$ for Blocks 2 and 4, where $\hat{\sigma}_i$ is the sample standard deviation of ith element of $\boldsymbol{x}$. Four values of $c(= 0.1, 0.5, 1$, and $2)$ are used, and the corresponding estimated rejection probability is denoted as L_c. In computing L_c, h^k, shown in Equations (10) and (11), is replaced with $\prod_{i=1}^{k} h_i$.

Test statistics are denoted as N^i_{q,q^*} and L^i_c, with the superscripts $i = A, B, W$ referring to the methods of obtaining the null distributions of the test statistics; asymptotics $(i = A)$, naive bootstrap $(i = B)$, and wild bootstrap $(i = W)$. Monte Carlo experiments are conducted with 500 bootstrap resamples and 1,000 Monte Carlo replications.

Let T_n be a statistic (either N or L) computed using the sample $\{y_t\ \boldsymbol{x}_t\ \hat{\varepsilon}_t\}_{t=1}^n$. The following steps are taken to compute the p-values of the naive and wild bootstrap test statistics.

1. Generate the bootstrap residuals $\{\varepsilon_t^*\}$ from $\hat{\varepsilon}_t = y_t - \boldsymbol{x}_t\hat{\boldsymbol{\beta}}$:
   a. For naive bootstrap, $\{\varepsilon_t^*\}$ is obtained from random resampling of $\{\hat{\varepsilon}_t\}$ with replacement.
   b. For wild bootstrap, $\varepsilon_t^* = a\hat{\varepsilon}_t$ with probability $r = (\sqrt{5}+1)/2\sqrt{5}$ and $\varepsilon_t^* = b\hat{\varepsilon}_t$ with probability $1 - r$ $(t = 1, \ldots, n)$, where $a = -(\sqrt{5}-1)/2$ and $b = (\sqrt{5}+1)/2$. (See Li and Wang 1998 [pp. 150–151].)
2. Generate the bootstrap sample $\{y_t^*\ \boldsymbol{x}_t^*\ \varepsilon_t^*\}_{t=1}^n$:
   a. When $\boldsymbol{x}_t$ is exogenous (Blocks 3, 4), then $\boldsymbol{x}_t^* = \boldsymbol{x}_t$ and $y_t^* \equiv \boldsymbol{x}_t\hat{\boldsymbol{\beta}} + \varepsilon_t^*$ $(t = 1, \ldots, n)$.
   b. When $\boldsymbol{x}_t$ is lagged dependent variables (Blocks 1, 2), we generate the bootstrap sample in two different ways.
      i. Generate $y_t^* \equiv \boldsymbol{x}_t\hat{\boldsymbol{\beta}} + \varepsilon_t^*$ $(t = 1, \ldots, n)$, conditioning on $\boldsymbol{x}_t^* = \boldsymbol{x}_t$. This is equivalent to treating $\boldsymbol{x}_t$ as exogenous. I call this procedure "the *conditional* bootstrap."
      ii. Generate initial values of y_t^* for $t = 1, \ldots, k$, from $N(\bar{y}, \hat{\sigma}_Y^2)$, and then get $y_t^* \equiv \boldsymbol{x}_t^*\hat{\boldsymbol{\beta}} + \varepsilon_t^*$ recursively for $t = k+1, \ldots, n$. $\bar{y}$ and $\hat{\sigma}_Y^2$ are unconditional sample mean and variance of y. I call this procedure "the *recursive* bootstrap."
3. Using the bootstrap sample $\{y_t^*\ \boldsymbol{x}_t^*\ \varepsilon_t^*\}_{t=1}^n$, calculate the bootstrap test statistic T_n^*.
4. Repeat the above steps B times. I use $B = 500$. The bootstrap p-value of T_n is the relative frequency of the event $\{T_n^* \geq T_n\}$ in the B bootstrap resamples.

3.3 Results

For weakly dependent processes in Blocks 1 and 2, the results of the conditional bootstrap are presented in Tables 1 and 2 and the results of the recursive bootstrap are presented in Table 3. For Blocks 3 and 4 where x_t is exogenous, there is no need to distinguish the conditional and recursive bootstrap procedures, and the results are presented in Tables 1 and 2.

Table 1 gives the estimated size of the tests for the five DGPs that are linear in conditional mean. The 95% asymptotic confidence interval of the estimated size is (0.036, 0.064) at 5% nominal level of significance, and

Table 1
Size

Panel A. Size of NN test at 5% nominal level of significance													
DGP	n	$N^A_{10,1}$	$N^B_{10,1}$	$N^W_{10,1}$	$N^A_{10,3}$	$N^B_{10,3}$	$N^W_{10,3}$	$N^A_{20,3}$	$N^B_{20,3}$	$N^W_{20,3}$	$N^A_{20,5}$	$N^B_{20,5}$	$N^W_{20,5}$
1.1	50	35	32	40	40	40	46	40	42	45	39	41	43
	100	40	40	47	34	30	38	31	33	39	42	41	40
	200	46	44	49	53	52	55	51	51	50	44	42	50
1.2	50	60	58	32	131	129	40	126	119	40	165	161	35
	100	105	104	46	201	187	39	193	172	40	249	225	31
	200	172	166	49	276	258	40	269	257	37	365	333	39
2.1	50	45	39	38	48	42	57	50	43	50	45	42	42
	100	51	50	57	52	48	57	59	53	49	39	41	42
	200	51	48	56	51	50	50	50	49	56	47	42	45
2.4	50	42	41	42	48	33	49	44	40	50	47	49	49
	100	38	35	41	55	50	53	43	43	47	38	39	43
	200	61	59	57	64	63	56	57	57	56	48	47	50
4.1	50	46	40	43	45	45	41	60	60	58	59	55	56
	100	52	49	53	52	52	56	49	43	53	47	44	50
	200	54	58	55	51	51	50	49	51	41	44	43	46
Panel B. Size of NN test at 10% nominal level of significance													
DGP	n	$N^A_{10,1}$	$N^B_{10,1}$	$N^W_{10,1}$	$N^A_{10,3}$	$N^B_{10,3}$	$N^W_{10,3}$	$N^A_{20,3}$	$N^B_{20,3}$	$N^W_{20,3}$	$N^A_{20,5}$	$N^B_{20,5}$	$N^W_{20,5}$
1.1	50	89	86	94	95	93	100	88	81	95	84	79	94
	100	87	90	90	77	77	86	81	82	82	84	89	86
	200	79	84	91	97	96	108	105	101	97	94	93	103
1.2	50	108	107	79	210	190	83	196	187	90	240	222	84
	100	165	161	89	279	273	93	272	266	96	343	328	87
	200	247	256	108	366	359	101	360	357	92	467	455	88
2.1	50	92	81	91	101	85	105	101	90	105	89	82	91
	100	108	110	112	93	91	89	102	96	94	95	91	97
	200	112	116	118	110	105	111	118	117	111	92	89	100
2.4	50	93	82	92	112	99	108	112	104	107	104	98	109
	100	88	88	93	97	93	102	86	82	94	92	90	96
	200	105	102	113	114	111	116	102	105	108	100	99	99
4.1	50	106	96	106	97	89	91	116	98	108	128	113	113
	100	92	91	90	116	113	104	102	103	100	116	113	111
	200	91	91	90	90	89	94	105	101	98	93	88	101

(*continued*)

(0.081, 0.119) at 10% nominal level of significance, since if the true size is s (e.g., $s = 0.05, 0.10$) the estimated size follows the asymptotic normal distribution with mean s and variance $s(1-s)/1{,}000$ with 1,000 Monte Carlo replications. We observe the following size behavior of the two tests under the null:

1. For DGP 1.1, 2.1, 2.4, and 4.1, where the conditional variance of y_t is constant, both the naive and wild bootstrap procedures give similar size behavior for the NN test and for the LWZ test.

2. The asymptotic NN test (N^A_{q,q^*}) performs very well even at the small sample size $n = 50$ and is as good as the bootstrap tests (N^B_{q,q^*} and N^W_{q,q^*}). The size of N^i_{q,q^*} ($i = A, B, W$) is not sensitive to q and q^*.

3. For the LWZ test, both bootstrap tests L^B_c and L^W_c perform very well even at the small sample size $n = 50$ and better than the asymptotic test L^A_c. The size of L^i_c ($i = B, W$) is not sensitive to c, but the size of L^A_c is.

4. The asymptotic LWZ test (L^A_c) does not perform well even at the larger sample size $n = 200$. Its size performance is better with smaller values of c, as explained by Li (1999, p. 118), who shows that the rate at which the test converges to the standard normal limiting distribution depends on c (and thus on b), and a smaller c will lead to a smaller error in the normal approximation. (But as noted above, the bootstrap tests L^i_c ($i = B, W$) have adequate size for all c in the range considered.)

5. Turning to DGP 1.2, where y_t is conditionally heteroskedastic, the size distortion is severe for the naive bootstrap tests, N^B_{q,q^*} and L^B_c, and generally gets worse as n increases, because the naive bootstrap does

Table 1
(*continued*)

Panel C. Size of LWZ test at 5% nominal level of significance

DGP	n	$L^A_{0.1}$	$L^B_{0.1}$	$L^W_{0.1}$	$L^A_{0.5}$	$L^B_{0.5}$	$L^W_{0.5}$	$L^A_{1.0}$	$L^B_{1.0}$	$L^W_{1.0}$	$L^A_{2.0}$	$L^B_{2.0}$	$L^W_{2.0}$
1.1	50	47	57	59	13	50	46	1	46	51	0	40	40
	100	31	41	42	15	41	43	6	51	42	0	41	43
	200	38	43	43	27	57	58	8	49	52	2	50	43
1.2	50	47	59	51	31	73	50	10	81	45	0	70	45
	100	46	55	53	34	70	47	24	92	54	4	116	56
	200	62	74	59	46	100	50	43	125	50	19	169	56
2.1	50	26	52	31	20	39	37	6	41	43	0	46	51
	100	30	46	37	26	48	55	11	49	43	0	48	48
	200	44	53	48	26	55	53	9	50	45	0	38	41
2.4	50	21	47	32	19	45	44	7	44	39	0	47	52
	100	46	54	48	38	62	58	10	55	48	1	52	51
	200	57	58	61	34	61	66	18	52	52	0	52	51
4.1	50	26	53	31	31	60	61	10	65	61	1	55	53
	100	32	44	38	27	45	45	13	46	47	0	59	54
	200	51	60	60	30	55	59	14	54	51	2	43	45

Panel D. Size of LWZ test at 10% nominal level of significance

DGP	n	$L^A_{0.1}$	$L^B_{0.1}$	$L^W_{0.1}$	$L^A_{0.5}$	$L^B_{0.5}$	$L^W_{0.5}$	$L^A_{1.0}$	$L^B_{1.0}$	$L^W_{1.0}$	$L^A_{2.0}$	$L^B_{2.0}$	$L^W_{2.0}$
1.1	50	89	121	115	33	103	97	5	102	102	0	78	94
	100	60	91	88	29	86	88	11	94	95	0	91	98
	200	65	101	95	50	97	99	20	105	102	3	90	91
1.2	50	80	102	87	48	134	95	17	145	104	0	130	94
	100	78	112	94	51	135	93	40	147	101	6	176	100
	200	103	134	107	76	179	109	60	201	100	29	241	104
2.1	50	99	107	79	36	98	97	13	94	98	1	89	117
	100	95	96	92	51	100	106	17	111	107	0	98	91
	200	100	102	104	61	102	99	15	101	99	1	83	90
2.4	50	102	108	76	47	95	90	14	96	95	0	98	103
	100	100	102	94	62	108	112	20	108	103	3	101	110
	200	105	112	112	73	119	115	27	104	104	2	104	105
4.1	50	95	103	91	51	122	115	18	110	104	1	100	99
	100	83	86	79	43	86	90	20	109	109	0	112	113
	200	117	120	120	60	110	113	26	106	107	4	102	102

Note: Test statistics are denoted as N^i_{q,q^*} and L^i_c, with superscripts $i = A, B, W$ referring to the methods of obtaining the null distributions of the test statistics using the asymptotics (A), naive bootstrap (B), and wild bootstrap (W). Number of bootstrap resamples = 500, and number of Monte Carlo replications = 1,000. The numbers of rejections out of 1,000 replications are reported. The 95% asymptotic confidence interval of the estimated size is (36, 64) at 5% nominal level of significance and (81, 119) at 10% nominal level of significance. DGP 1.2 is a linear model with GARCH errors.

not preserve the conditional heteroskedasticity in resampling. The effect of the conditional heteroskedasticity can be removed using the wild bootstrap, which preserves the heteroskedasticity in resampling. The result shows that the tests with the wild bootstrap procedure generally have an adequate size for DGP 1.2 for both NN and LWZ tests.

6. Also, it can be noted that the asymptotic NN test N^A_{q,q^*} is not robust to the presence of conditional heteroskedasticity, because Equation (17) is obtained from Equation (16) under conditional homoskedasticity, as noted in Section 2.2. On the other hand, the asymptotic normality of Equation (11) for the LWZ test does not require conditional homoskedasticity as long as some moment conditions are satisfied (see Li 1999, p. 107), and thus the size of L^A_c may not be affected by the presence of GARCH. But as mentioned above, L^A_c is very sensitive to c.

Table 2 presents the power of the tests N_{q,q^*} and L_c at the 5% level. (The power results at 10% are available from the author but are not reported here for space reasons.) As the results obtained can be considerably influenced by the choice of nonlinear models, we try to include as many different types of nonlinear models

Table 2
Power

Panel A. Power of NN test at 5% nominal level of significance

DGP	n	$N^A_{10,1}$	$N^B_{10,1}$	$N^W_{10,1}$	$N^A_{10,3}$	$N^B_{10,3}$	$N^W_{10,3}$	$N^A_{20,3}$	$N^B_{20,3}$	$N^W_{20,3}$	$N^A_{20,5}$	$N^B_{20,5}$	$N^W_{20,5}$
1.3	50	143	135	56	184	170	43	170	163	34	268	257	39
	100	188	190	54	291	280	57	291	285	61	407	388	46
1.4	50	79	79	88	448	446	481	461	450	494	480	482	505
	100	84	83	100	785	778	798	790	785	812	879	880	877
1.5	50	140	137	130	643	643	618	637	624	619	592	595	547
	100	205	209	196	930	928	933	946	940	947	951	949	926
1.6	50	97	93	89	63	64	55	61	65	60	54	56	49
	100	145	137	148	101	97	98	95	93	95	73	74	69
2.2	50	65	56	57	85	78	71	64	59	56	100	89	75
	100	79	71	67	120	107	78	106	103	82	113	107	81
2.3	50	147	132	124	358	335	257	399	366	297	375	354	262
	100	189	181	144	636	626	534	696	685	606	710	705	584
2.5	50	163	141	130	477	460	338	458	447	351	679	663	507
	100	249	241	183	715	707	581	697	689	562	934	927	819
2.6	50	142	131	109	284	255	184	279	253	181	424	402	271
	100	202	200	138	491	477	339	495	479	353	685	669	500
3.1	50	649	632	543	554	540	362	545	543	372	460	456	254
	100	914	904	868	855	849	713	850	848	739	796	798	546
3.2	50	461	452	374	461	460	287	459	459	307	437	441	193
	100	738	732	628	765	763	559	762	756	577	737	732	366
4.2	50	560	538	396	976	968	893	987	987	930	987	988	901
	100	652	646	492	996	996	983	1,000	1,000	995	1,000	1,000	994
4.3	50	129	120	112	210	202	175	211	197	188	208	200	189
	100	195	191	182	397	389	366	391	377	367	397	389	356
4.4	50	652	642	477	997	996	980	1,000	1,000	998	1,000	1,000	1,000
	100	746	734	565	998	998	989	1,000	1,000	1,000	1,000	1,000	1,000

Panel B. Power of LWZ test at 5% nominal level of significance

DGP	n	$L^A_{0.1}$	$L^B_{0.1}$	$L^W_{0.1}$	$L^A_{0.5}$	$L^B_{0.5}$	$L^W_{0.5}$	$L^A_{1.0}$	$L^B_{1.0}$	$L^W_{1.0}$	$L^A_{2.0}$	$L^B_{2.0}$	$L^W_{2.0}$
1.3	50	69	80	44	48	118	47	31	153	45	4	168	43
	100	69	83	54	91	158	55	72	215	51	30	269	61
1.4	50	320	356	355	435	580	588	247	602	612	2	376	424
	100	660	701	687	872	923	920	780	937	939	82	829	844
1.5	50	404	452	439	523	718	662	293	716	683	6	530	504
	100	840	864	865	942	977	973	864	978	971	192	929	927
1.6	50	46	59	54	28	69	64	11	74	75	0	76	87
	100	40	53	58	37	84	84	20	114	106	3	131	135
2.2	50	35	62	44	37	67	59	12	72	73	0	81	70
	100	48	55	49	47	75	64	33	91	83	2	115	93
2.3	50	31	71	45	94	170	139	99	287	230	33	431	346
	100	75	91	73	179	253	224	266	493	425	183	743	636
2.5	50	79	109	73	290	383	318	262	556	433	10	589	446
	100	142	162	139	643	728	647	753	916	815	337	962	810
2.6	50	76	96	60	222	330	222	126	420	237	3	362	177
	100	155	173	128	515	623	481	469	752	513	75	728	364
3.1	50	123	147	134	195	322	269	198	456	378	62	584	505
	100	207	248	237	443	572	537	492	723	676	365	865	810
3.2	50	132	152	125	190	289	231	187	375	306	108	464	386
	100	227	254	222	374	488	415	411	600	508	326	717	611
4.2	50	102	182	107	745	829	695	884	968	870	859	995	955
	100	343	378	315	978	986	968	999	999	993	999	1,000	997
4.3	50	27	61	36	69	129	118	59	219	213	9	293	272
	100	51	67	57	146	229	223	186	395	388	73	509	515
4.4	50	538	627	491	1,000	1,000	992	1,000	1,000	997	1,000	1,000	1,000
	100	946	964	930	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Note: See *Note* to Table 1 for explanation of notation and abbreviations.

as possible. We observe the following power behavior of the two tests:

1. Neither test is uniformly superior to the other in terms of power.
2. For quite a few DGPs the power of the naive bootstrap NN test (N^B_{q,q^*}) is greater than the power of the wild bootstrap NN test (N^W_{q,q^*}). Those DGPs are DGP 1.2, 1.3, 2.2, 2.3, 2.5, and 2.6, all of which are either bilinear processes or nonlinear moving-average processes. As noted by Bera and Higgins (1997) and Weiss (1986), these processes are conditionally heteroskedastic or exhibit apparent heteroskedastic structure. So the use of the wild bootstrap could absorb some of these nonlinearities and thus could have an adverse impact on the power of the statistics. Similarly, for the LWZ test, power is greater when the naive bootstrap is used compared to the wild bootstrap, because the test with the naive bootstrap procedure not only may have power to detect nonlinearity in conditional mean but also is not robust to the presence of conditional heteroskedasticity or seemingly similar heteroskedastic structures of bilinear and nonlinear moving-average processes.
3. Although the size of N^i_{q,q^*} $(i = A, B, W)$ is not sensitive to q and q^*, the power of N^i_{q,q^*} $(i = A, B, W)$ is affected by the choice of q^* (but not by the choice of q). The results show that although $N_{10,3}$ is as powerful as $N_{20,3}$ when the same $q^* = 3$ is used, $N_{10,1}$ is less powerful than $N_{10,3}$, and $N_{20,3}$ is also sometimes less powerful than $N_{20,5}$. The power is substantially reduced if too small a number of the principal components of the neural network activation functions are used in the NN test. Hence, small values of q^* should be avoided in practice, and larger values of q^* are recommended, as long as the collinearity/singularity in computing N_{q,q^*} in Equation (17) may be avoided.
4. As shown in Equation (11), the LWZ test diverges to $+\infty$ at the rate of $nh^{k/2}$, and thus the higher values of h and c will make the test more powerful. But because of the severe downward size distortion of the asymptotic test L^A with higher values of c, we actually observe that the power of the asymptotic LWZ test (L^A) may be weaker for higher values of c. Thus increasing the bandwidth factor c up to a value of two reduces both the type-I and type-II errors for the asymptotic test L^A, as noted in Li 1999. The power of the bootstrap test L^B and L^W generally increases with higher values of c, however, because the size of the bootstrap LWZ tests (L^B and L^W) is very good for all four values of c considered.

Table 3 presents the size and power performances of the recursive bootstrap tests, whereas Tables 1 and 2 present those performances of the conditional bootstrap tests. Comparing these two bootstrap procedures for the weakly dependent time series (Blocks 1, 2) provides us with very useful information about using the bootstrap for time series models:

1. The size of the conditional bootstrap test is better than that of the recursive bootstrap test. The use of the conditional bootstrap benefits the LWZ test much more than the NN test. The size of the conditional bootstrap LWZ test is not sensitive to c, whereas the size of the recursive bootstrap LWZ test is quite sensitive to c. Hence, even for the time series, use of the conditional bootstrap is recommended, treating the lagged dependent variables as exogenous instead of bootstrapping them recursively from the estimated models.
2. The power performance of both bootstrap procedures is rather similar.

4 Conclusions

This article has considered two conditional moment tests for neglected nonlinearity in time series regression models and the finite sample performance. Both the naive bootstrap and the wild bootstrap are used to generate the critical values, together with asymptotic distributions. For parametric models, Davidson and MacKinnon (1999) show that the size distortion of a bootstrap test is at least of an order $n^{-1/2}$ smaller than that of the corresponding asymptotic test. For nonparametric models, h also enters to the order of refinement. Li and

Table 3
Recursive bootstrap for blocks 1, 2

Panel A. Size of NN and LWZ tests at 5% nominal level of significance

DGP	n	$N^B_{10,3}$	$N^W_{10,3}$	$N^B_{20,3}$	$N^W_{20,3}$	$L^B_{0.5}$	$L^W_{0.5}$	$L^B_{1.0}$	$L^W_{1.0}$	$L^B_{2.0}$	$L^W_{2.0}$
1.1	50	47	47	43	50	27	30	18	15	2	2
	100	48	45	45	48	29	27	20	18	9	12
	200	55	52	53	51	36	36	26	23	9	10
1.2	50	140	55	141	56	51	38	52	24	26	6
	100	178	39	190	38	71	50	70	37	49	20
	200	309	59	314	49	91	53	118	47	133	37
2.1	50	26	27	32	26	28	27	22	18	2	2
	100	22	24	31	38	30	35	28	26	7	6
	200	42	35	40	46	32	45	18	20	3	7
2.4	50	38	35	39	40	25	26	19	16	3	4
	100	36	40	47	44	30	29	18	21	6	6
	200	38	38	45	51	41	37	31	30	6	4

Panel B. Size of NN and LWZ tests at 10% nominal level of significance

DGP	n	$N^B_{10,3}$	$N^W_{10,3}$	$N^B_{20,3}$	$N^W_{20,3}$	$L^B_{0.5}$	$L^W_{0.5}$	$L^B_{1.0}$	$L^W_{1.0}$	$L^B_{2.0}$	$L^W_{2.0}$
1.1	50	97	99	102	99	63	63	42	44	10	13
	100	118	114	109	109	75	74	52	48	23	24
	200	113	114	111	109	85	82	63	59	33	30
1.2	50	220	120	233	132	105	79	90	58	47	22
	100	279	103	287	115	123	91	118	70	93	38
	200	424	129	427	128	163	101	185	103	188	81
2.1	50	72	71	62	64	75	74	52	44	13	11
	100	59	71	77	82	72	70	54	54	17	14
	200	77	88	87	96	87	91	49	58	21	22
2.4	50	85	91	77	87	61	58	35	35	13	12
	100	87	84	87	95	68	66	38	39	18	15
	200	86	87	96	93	85	83	65	63	18	17

Panel C. Power of NN and LWZ tests at 5% nominal level of significance

DGP	n	$N^B_{10,3}$	$N^W_{10,3}$	$N^B_{20,3}$	$N^W_{20,3}$	$L^B_{0.5}$	$L^W_{0.5}$	$L^B_{1.0}$	$L^W_{1.0}$	$L^B_{2.0}$	$L^W_{2.0}$
1.3	50	166	56	162	49	69	40	84	31	82	15
	100	273	76	276	86	121	44	162	52	179	42
1.4	50	404	412	416	419	482	461	410	398	69	74
	100	781	797	785	812	902	908	884	881	527	546
1.5	50	658	639	685	672	629	621	529	534	145	143
	100	962	965	971	968	971	972	955	960	678	710
1.6	50	94	86	92	83	62	61	52	47	21	26
	100	92	87	95	85	57	57	60	57	36	39
2.2	50	50	45	50	46	60	48	27	25	10	9
	100	71	68	74	67	60	50	63	50	27	19
2.3	50	284	231	285	255	101	91	141	125	157	128
	100	609	553	657	601	211	205	359	346	479	419
2.5	50	453	390	440	358	344	309	425	362	255	155
	100	717	624	710	610	680	651	855	798	802	652
2.6	50	250	200	239	193	242	184	253	153	124	34
	100	458	342	392	283	540	478	657	474	449	171

Note: See *Note* to Table 1 for explanation of notation and abbreviations.

Wang (1998) show that if the distribution of L admits an Edgeworth expansion, then the bootstrap distribution approximates the null distribution of L, improving over the asymptotic approximation. The failure of the first-order asymptotics for nonparametric tests is well known; see the discussion and some Monte Carlo findings reported, for example, in the survey of Tjøstheim (1999, sections 2.5–2.7). This motivates the use of bootstrap. The bootstrap tests L^B and L^W are indeed more accurate than the asymptotic test L^A, as confirmed in the simulation presented in this article. The asymptotic NN test N^A_{q,q^*} performs very well, whereas the asymptotic LWZ test L^A_c is sensitive to c. The bootstrap is very useful for the L test. The bootstrap LWZ tests (L^B_c, L^W_c) work really well and are robust to the choice of c. A particularly useful result is that the performance of the conditional bootstrap is much more reliable than that of the recursive bootstrap even for time series models.

References

Aït-Sahalia, Y., P. J. Bickel, and T. M. Stoker (1994). "Goodness-of-fit tests for regression using kernel models." Unpublished manuscript. Princeton: Princeton University, Berkeley: University of California at Berkeley, and Cambridge: Massachusetts Institute of Technology.

Andrews, D. W. K. (1991). "Heteroskedasticity and autocorrelation consistent matric estimation." *Econometrica*, 59(3): 817–858.

Bera, A. K., and M. L. Higgins (1997). "ARCH and bilinearity as competing models for nonlinear dependence." *Journal of Business and Economic Statistics* 15: 43–50.

Bierens, H. J. (1990). "A consistent conditional moment test of functional form." *Econometrica*, 58: 1443–1458.

Bierens, H. J., and W. Ploberger (1997). "Asymptotic theory of integrated conditional moment tests." *Econometrica*, 65: 1129–1151.

Blake, A. P., and G. Kapetanios (1999). "A radial based function artificial neural network test for neglected nonlinearity." London: National Institute of Economic and Social Research.

Blake, A. P., and G. Kapetanios (2000). "A radial basis function artificial neural network test for ARCH." *Economics Letters*, 69: 15–23.

Bradley, R., and R. McClelland (1996). "A kernel test for neglected nonlinearity." *Studies in Nonlinear Dynamics and Econometrics*, 1(2): 119–130.

Cai, Z., J. Fan, and Q. Yao (2000). "Functional-coefficient regression models for nonlinear time series." *Journal of the American Statistical Association*, 95: 941–956.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Dahl, C. M. (1999). "An investigation of tests for linearity and the accuracy of flexible nonlinear inference." Unpublished manuscript. Purdue, Ind.: Purdue University.

Dahl, C. M., and G. González-Rivera (2000). "Testing for neglected nonlinearity in regression models: A collection of new tests." Unpublished manuscript. Purdue, Ind.: Purdue University, and Riverside: University of California at Riverside.

Davidson, R., and J. G. MacKinnon (1999). "The size distortion of bootstrap tests." *Econometric Theory*, 15(3): 361–376.

Efron, B. (1979). "Bootstrapping methods: Another look at the jackknife." *Annals of Statistics*, 7: 1–26.

Eubank, R. L., and C. H. Spiegelman (1990). "Testing the goodness-of-fit of the linear models via non-parametric regression techniques." *Journal of the American Statistical Association*, 85: 387–392.

Fan, J., and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fan, Y., and Q. Li (1996). "Consistent model specification tests: Omitted variables and semiparametric functional forms." *Econometrica*, 64: 865–890.

Fan, Y., and Q. Li (forthcoming). "Consistent model specification tests: Nonparametric versus Bierens' tests." *Econometric Theory*.

Granger, C. W. J., and T. Teräsvirta (1993). *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.

Hamilton, J. D. (2001). "A parametric approach to flexible nonlinear inference." *Econometrica*, 69(3): 537–573.

Härdle, W., and E. Mammen (1993). "Comparing nonparametric versus parametric regression fits." *Annals of Statistics*, 21: 1926–1947.

Hjellvik, V., and D. Tjøstheim (1995). "Nonparametric tests of linearity for time series." *Biometrika*, 82(2): 351–368.

Hjellvik, V., and D. Tjøstheim (1996). "Nonparametric statistics for testing of linearity and serial independence." *Journal of Nonparametric Statistics*, 6: 223–251.

Hjellvik, V., Q. Yao, and D. Tjøstheim (1998). "Linearity testing using local polynomial approximation." *Journal of Statistical Planning and Inference*, 68: 295–321.

Hong, Y., and H. White (1995). "Consistent specification testing via nonparametric series regression." *Econometrica*, 63: 1133–1160.

Hornik, K., M. Stinchcombe, and H. White (1989). “Multi-layer feedforward networks are universal approximators.” *Neural Network*, 2: 359–366.

Lee, T.-H., and A. Ullah (forthcoming-a). “Nonparametric bootstrap tests for neglected nonlinearity in time series regression models.” *Journal of Nonparametric Statistics*.

Lee, T.-H., and A. Ullah (forthcoming-b). “Nonparametric bootstrap specification testing in econometric model.” Submitted to *Computer-Aided Econometrics*, D. Giles, ed., New York: Marcel Dekker.

Lee, T.-H., H. White, and C. W. J. Granger (1993). “Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests.” *Journal of Econometrics*, 56: 269–290.

Li, Q. (1999). “Consistent model specification tests for time series econometric models.” *Journal of Econometrics*, 92(1): 101–147.

Li, Q., and S. Wang (1998). “A simple consistent bootstrap test for a parametric regression function.” *Journal of Econometrics*, 87: 145–165.

Liu, R. Y. (1988). “Bootstrap procedures under some non-iid models.” *Annals of Statistics*, 16: 1697–1708.

Matsuda, Y. (1999). “A test of linearity against functional-coefficient autoregressive models.” *Communications in Statistics, Theory and Method*, 28(11): 2539–2551.

Newey, W. K. (1985). “Maximum likelihood specification testing and conditional moment tests.” *Econometrica*, 53: 1047–1070.

Pagan, A. R., and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Poggi, J.-M., and B. Portier (1997). “A test of linearity for functional autoregressive models.” *Journal of Time Series Analysis*, 18(6): 615–639.

Stengos, T., and Y. Sun (1998). “Consistent model specification test for a regression function based on nonparametric wavelet estimation.” Unpublished manuscript. Guelph, Ont.: University of Guelph.

Stinchcombe, M., and H. White (1998). “Consistent specification testing with nuisance parameters present only under the alternative.” *Econometric Theory*, 14: 295–325.

Tauchen, G. (1985). “Diagnostic testing and evaluation of maximum likelihood models.” *Journal of Econometrics*, 30: 415–443.

Teräsvirta, T. (1996). “Power properties of linearity tests for time series.” *Studies in Nonlinear Dynamics and Econometrics*, 1(1): 3–10.

Teräsvirta, T., C.-F. Lin, and C. W. J. Granger (1993). “Power of the neural network linearity test.” *Journal of Time Series Analysis*, 14(2): 209–220.

Tjøstheim, D. (1999). “Nonparametric specification procedures for time series.” In S. Ghosh (ed.), *Asymptotics, Nonparametrics, and Time Series: A Tribute to M.L. Puri*. New York: Marcel Dekker, pp. 149–199.

Tong, H. (1990). *Nonlinear Time Series: A Dynamic System Approach*. Oxford: Clarendon.

Ullah, A. (1985). “Specification analysis of econometric models.” *Journal of Quantitative Economics*, 1: 187–209.

Weiss, A. A. (1986). “ARCH and bilinear time series models: Comparison and combination.” *Journal of Business and Economic Statistics*, 4: 59–70.

Whang, Y.-J. (2000). “Consistent bootstrap tests of parametric regression functions.” *Journal of Econometrics*, 98: 27–46.

White, H. (1980). “A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity.” *Econometrica*, 48: 817–838.

White, H. (1987). “Specification testing in dynamic models.” In T. F. Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, vol. 1. New York: Cambridge University Press, pp. 1–58.

White, H. (1989). “An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks.” *Proceedings of the International Joint Conference on Neural Networks*, 2: 451–455.

White, H. (1994). *Estimation, Inference, and Specification Analysis*. Cambridge: Cambridge University Press.

Wu, C. F. J. (1986). “Jackknife, bootstrap, and other resampling methods in regression analysis.” *Annals of Statistics*, 14: 1261–1350.

Zheng, J. X. (1996). “A consistent test of functional form via nonparametric estimation techniques.” *Journal of Econometrics*, 75: 263–289.