Nonparametric Test for the Form of Parametric Regression with Time Series Errors *

Lan Wang† and Ingrid Van Keilegom‡

University of Minnesota and Université catholique de Louvain

July 10, 2005

Abstract

We propose a new nonparametric method for testing the parametric form of a regression function in the presence of time series errors. The test is motivated by recent advancement in the theory of ANOVA with large number of factor levels and also utilizes a new difference-based estimation method in nonparametric regression with time-series errors proposed by Hall and Van Keilegom (2003). The test statistic is asymptotically normal under the null and local alternative hypotheses. We also propose a bootstrap method to calculate the critical values and prove its consistency. In a Monte Carlo study, we demonstrate that this bootstrap procedure has good properties for moderate sample size.

Key words: bootstrap, correlated errors, goodness-of-fit test, lack-of-fit test, nearest-neighbor windows, nonparametric regression, residual, time-series errors, trend

---

*Short title: nonparametric test for regression.

†Address for correspondence: 385 Ford Hall, School of Statistics, University of Minnesota, 224 Church Street S.E., Minneapolis, Minnesota 55455, USA. E-mail: lan@stat.umn.edu, fax: (612) 624-8868.

‡Institut de statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: vankeilegom@stat.ucl.ac.be. Financial support from the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy) is gratefully acknowledged.

1 Introduction

We consider the problem of testing the parametric form of a trend function against an omnibus alternative in the presence of time-series errors. Data (x_{im}, Y_{im}), $i = 1, \ldots, m$, are generated from the following nonparametric regression model

$$Y_{im} = g(x_{im}) + \epsilon_{im},\ \ i = 1, ..., m, \tag{1.1}$$

where Y_{im} is the response, $g(\cdot)$ is an unknown smooth regression function, $x_{1m} \leq \ldots \leq x_{mm}$ are fixed design points, and $\epsilon_{1m}, \ldots, \epsilon_{mm}$ are stationary time-series errors. The m in the subscript will be omitted from the notation when no confusion is possible.

Model (1.1) has broad applications. For instance, a large number of studies have been devoted to investigating the presence of global "warming trend" due to green house effects. The standard approach is to fit a linear trend (Woodward and Gray (1993), (1995)), where Y_i is the temperature at time i, and then proceed to test whether the slope equals zero. Rejection of this null hypothesis generally leads to the belief of the existence of linear trend. Despite its mathematical convenience, there is no special reason to believe a simple linear trend function would be suitable to model the complex climate system. Making the linearity assumption without validating its adequacy can cause serious bias in both estimation and prediction. Model (1.1) is also used frequently in economics where Y_i represents stock price, GNP growth rate, consumer price index, etc. Kim and Hart (1998) gave an example of astronomy data, where the successive maxima on the light curve of the variable star T Centaurus are studied.

In this paper, we propose a nonparametric method for testing

$$H_0 :\ \ g(x) \in S_\Theta, \tag{1.2}$$

where $S_\Theta = \{g(\cdot, \theta), \theta \in \Theta\}$ is a parametric family of functions, Θ a subset of a Euclidean space, $g(\cdot, \theta)$ is a function on $\mathcal{R}$. This allows us to check, for example, whether the linearity form of the trend function is a valid assumption. Only a very weak smoothness condition is imposed on $g(x)$. Our nonparametric test thus is designed to be powerful against a large class of alternatives. In this sense, it is called "omnibus".

When the errors ϵ_i are independent, much literature has been devoted to testing (1.2), see the manuscript of Hart (1997) for a comprehensive review. The dependent error case, although is common in application, has received relatively little discussion, see Brillinger (1989), González-Manteiga and Vilar-Fernández (1995), Bai (1996), Woodward, Bottone and Gray (1997), Kim and Hart (1998), Vogelsang (1998), Sun and Pantula (1999), Vilar-Fernández and González-Manteiga (2000). These papers, however, are restricted to the case where $g(x, \theta)$ is a linear model under the null hypothesis.

Our nonparametric testing procedure is new. It is motivated by recent advancements in the theory of analysis of variance when the number of factor levels is large, and it utilizes a new difference-based estimation method (Hall and Van Keilegom (2003)) for nonparametric regression with time-series errors. The new test generalizes the work of Wang, Akritas and Van Keilegom (2002) for independent error case. For moderately large sample size, we use a bootstrap procedure to calculate the critical values. The consistency of the bootstrap test is verified.

The paper is organized as follows. We introduce the test statistic in Section 2. In Section 3, we present the asymptotic normality of our test under null and local alternative hypotheses. The bootstrap procedure is discussed in Section 4. Numerical simulations, which demonstrate the good level and power performance of the bootstrapped test, are reported in Section 5. In Section 6, we provide the technical arguments.

2 Test Statistic

The disturbances ϵ_{im}, $i = 1, \ldots, m$, form a segment of a covariance-stationary autoregressive process of order p with zero mean, an assumption frequently satisfied in applications. More specifically, we assume $\epsilon_{im} = \sum_{j=1}^{p} \phi_j \epsilon_{i-j,m} + e_{im}$, where $\{e_{im}, -\infty < i < \infty\}$ (or simply e_i) are iid random variables with mean 0 and variance σ^2.

Instead of using directly the original data, we work with the following autoregressive transformation of (x_i, Y_i):

$$Z_i = \left(Y_{i+p} - \sum_{j=1}^{p} \widehat{\phi}_j Y_{i+p-j} \right) - \left(g(x_{i+p}, \widehat{\theta}) - \sum_{j=1}^{p} \widehat{\phi}_j g(x_{i+p-j}, \widehat{\theta}) \right), i = 1, \ldots, n, \quad (2.1)$$

with $n = m - p$, where $\widehat{\phi}_j$ is a $\sqrt{n}$-consistent estimator of ϕ_j, $j = 1, \ldots, p$, and $\widehat{\theta}$ is a $\sqrt{n}$-consistent estimator of θ under the null hypothesis, to be defined at the end of this section. We expect that the Z_i's form an approximate white noise process under H_0. This type of linear filtering has been used to estimate the parameters in the trend function $g(\cdot)$ in the time series literature. Here we propose to apply the test statistic of Wang, Akritas and Van Keilegom (2002) to (2.1) to obtain an omnibus test for (1.2). This application is new, and is different from the related work of González-Manteiga and Vilar-Fernández (1995), Kim and Hart (1998), and Vilar-Fernández and González-Manteiga (2000) who handle correlation directly through asymptotic derivations.

Motivated by recent developments in heteroscedastic ANOVA with large number of factor levels (Akritas and Papadatos (2004), Wang and Akritas (2002)), Wang, Akritas and Van Keilegom (2002) proposed a new omnibus test for the constant mean null hypothesis in (1.1) with independent errors. The basic idea is to consider each distinct covariate value x_i as a 'category' and construct a window W_i around each x_i consisting of the k_n nearest covariate values (for some k_n going to infinity). That is, they construct an artificial balanced one-way ANOVA with n categories, where the responses in the i-th category are the Y-values that correspond to the covariate values belonging to W_i. In what follows each window W_i will also be understood as the set containing the indices j of the covariate values that belong to the window around x_i, that is

$$W_i = \left\{ j : |\widehat{G}(x_j) - \widehat{G}(x_i)| \leq \frac{k_n - 1}{2n} \right\}, \tag{2.2}$$

where $\widehat{G}(x) = \frac{1}{n} \sum_{j=1}^{n} I\left(x_j \leq x\right)$. To test the null hypothesis of no effects, i.e., $g(x) = C$ where C is an unknown constant, Wang, Akritas and Van Keilegom (2002) suggested looking at $MST - MSE$, where MST is the treatment sum of squares, MSE is the error sum of squares, both computed from the hypothetical one-way ANOVA. This test statistic is thus related to the classical F-test statistic. The test is asymptotically normal under the null hypothesis and local alternatives, it is asymptotically unbiased under the null, and has been demonstrated to possess an accurate type I error rate and favorable power performance in simulations. In the next section, we show that if we apply this testing procedure to the autoregressive transformed data (2.1), the resulting test statistic is also asymptotically normal.

We now introduce some notation. Using the procedure described above, we construct a hypothetical one-way ANOVA (n cells, k_n observations per cell) from (x_i, Z_i), $i = 1, \ldots, n$. We use V_{ij}, $j = 1, \ldots, k_n$, to denote the k_n observations in the i-th category, i.e. $\{V_{i1}, \ldots, V_{ik_n}\} = \{Z_j : j \in W_i\}$. Let $\mathbf{V}$ be the $nk_n \times 1$ vector of all observations in this hypothetical one-way layout. Then our test statistic is defined as:

$$\begin{aligned} T_n &= MST - MSE \\ &= \frac{k_n}{n-1} \sum_{i=1}^{n} \left(\overline{V}_{i.} - \overline{V}_{..}\right)^2 - \frac{1}{n(k_n - 1)} \sum_{i=1}^{n} \sum_{j=1}^{k_n} \left(V_{ij} - \overline{V}_{i.}\right)^2, \end{aligned} \tag{2.3}$$

where $\overline{V}_{i.} = k_n^{-1}\sum_{j=1}^{k_n} V_{ij}$ and $\overline{V}_{..} = n^{-1}\sum_{i=1}^{n} \overline{V}_{i.}$. T_n can be expressed as a quadratic form in $\mathbf{V}$: $T_n = \mathbf{V}'\mathbf{A}\mathbf{V}$, with

$$\mathbf{A} = \frac{nk_n - 1}{n(n-1)k_n(k_n-1)} \bigoplus_{i=1}^{n} \mathbf{J}_{k_n} - \frac{1}{n(n-1)k_n}\mathbf{J}_{nk_n} - \frac{1}{n(k_n-1)}\mathbf{I}_{nk_n}, \tag{2.4}$$

where $\mathbf{I}_{k_n}$ is the k_n-dimensional identity matrix, $\mathbf{J}_{k_n} = \mathbf{1}_{k_n}\mathbf{1}'_{k_n}$ with $\mathbf{1}_{k_n}$ being the k_n-dimensional column vector of 1's, and $\bigoplus$ is Kronecker (direct) sum. The way we define the local window (2.2) enables the ith window to be symmetric around x_i, except for some of the windows at the two edges. However, it is easy to check that the windows at the two edges have asymptotically negligible influence on the distribution of the test statistic. Thus one may allow asymmetric windows at the edge, for example, a local window of size 5 around x_2 can be created by taking responses corresponding to x_i, $i = 1, 2, 3, 4, 5$.

We now discuss how to estimate θ and ϕ_j in (2.1). Any $\sqrt{n}$-consistent estimator for θ will work. When $g(x, \theta)$ is a linear function, the method of generalized least squares can be easily applied; in the nonlinear case, a $\sqrt{n}$-consistent estimator of θ was suggested by Gallant and Goebel (1976). Under some appropriate regularity conditions, $\sqrt{n}(\widehat{\theta} - \theta)$ is asymptotically normal.

For estimating ϕ_j, we employ the difference-based estimators newly proposed by Hall and Van Keilegom (2003), but any other $\sqrt{n}$-consistent estimator can be used as well (see e.g. Hart (1994)). The procedure of Hall and Van Keilegom has the advantage that it does not depend on a bandwidth parameter. This leads to some computational convenience for the bootstrap version test in Section 4, since we don't need to bring in an extra smoothing parameter. Defining $\gamma(j) = \mathrm{Cov}(\epsilon_i, \epsilon_{i-j})$ and the difference operator D_j by $(D_jY)_i = Y_i - Y_{i-j}$, $\gamma(0)$ and $\gamma(j)$ are estimated by:

$$\begin{aligned} \widehat{\gamma}(0) &= \frac{1}{m_2 - m_1 + 1} \sum_{m=m_1}^{m_2} \frac{1}{2(n-m)} \sum_{i=m+1}^{n} \{(D_mY)_i\}^2, \\ \widehat{\gamma}(j) &= \widehat{\gamma}(0) - \frac{1}{2(n-j)} \sum_{i=j+1}^{n} \{(D_jY)_i\}^2, \quad \mathit{for}\ j \geq 1. \end{aligned} \tag{2.5}$$

The rate at which m_1 and m_2 go to infinity will be specified later. Let $\mathbf{B}$ be the $p \times p$ matrix having $\widehat{\gamma}(j_1 - j_2)$ as its (j_1, j_2)th element. Using the Yule-Walker equations, the ϕ's are estimated by

$$(\widehat{\phi}_1, \ldots, \widehat{\phi}_p)' = \mathbf{B}^{-1}(\widehat{\gamma}(1), \ldots, \widehat{\gamma}(p)). \tag{2.6}$$

Under mild smoothness condition on $g(x)$, Hall and Van Keilegom proved that $\max_{1\leq j\leq p} |\widehat{\phi}_j - \phi_j| = O_p(m^{-1/2})$.

3 Large Sample Results

The following assumptions are made to derive the asymptotic distribution of T_n.

Assumption A1. The design points $x_1, x_2, ..., x_n$ on [0,1] satisfy:

$$\int_0^{x_i} r(x)dx = \frac{i}{n}, \quad i = 1, ..., n, \tag{3.1}$$

for some positive Lipschitz continuous design density $r(x)$.

Assumption A2. $g(x, \theta)$ is twice continuously differentiable with respect to x and the components of θ, and

$$\sup_{x,\theta} \left| \frac{\partial}{\partial \theta_j} g(x, \theta) \right| < \infty, \ \sup_{x,\theta} \left| \frac{\partial^2}{\partial x^2} g(x, \theta) \right| < \infty,$$

$$\sup_{x,\theta} \left| \frac{\partial^2}{\partial x \partial \theta_j} g(x, \theta) \right| < \infty, \ \sup_{x,\theta} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} g(x, \theta) \right| < \infty,$$

for $j, k = 1, \ldots, \dim(\Theta)$.

Assumption A3. The estimator $\widehat{\theta}$ satisfies $\widehat{\theta} - \theta = O_p(n^{-1/2})$ under H_0.

Assumption A4. $\{e_{im}, -\infty < i < \infty\}$ are iid variables with mean 0, variance σ^2 and finite fourth moment; $\epsilon_{im} = \sum_{j=1}^{p} \phi_j \epsilon_{i-j,m} + e_{im}$ is a stationary and causal autoregressive process of order p.

Assumption A5. $m_1 \leq m_2$, $m_1/\log n \to \infty$ and $m_2 = O(n^{1/2})$.

The asymptotic distribution of the test statistic of Wang, Akritas and Van Keilegom (2002) applied to (2.1) turns out to have the same form as for the independent case.

Theorem 3.1 *Assume conditions A1-A5. If $n \to \infty$ and $k_n \to \infty$ such that $n^{-1}k_n^{3/2} \to 0$, then under H_0,*

$$\left(\frac{n}{k_n}\right)^{1/2} T_n \to N\left(0, \frac{4\sigma^4}{3}\right)$$

in distribution.

Remark 1. Note that the asymptotic distribution does not depend on the covariance structure of the error process. This is because the effect of the covariance structure has been taken out when transforming the responses (see (2.1)). The variance σ^2 can be estimated by applying Rice's (1984) estimator to the autoregression transformed data (2.1):

$$\widehat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=2}^{n} (Z_i - Z_{i-1})^2. \tag{3.2}$$

Originally, this estimator was proposed for independent observations. The Z_i's mimic a white noise sequence and it is easily shown that $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$, using the consistency of $\widehat{\theta}$ and $\widehat{\phi}_j$.

Remark 2. Our test statistic can detect local alternatives converging to the null at rate $(nk_n)^{-1/4}$. We state below the asymptotic normality for the case $g(\cdot)$ is linear and $p = 1$, i.e., $\epsilon_i = \phi\epsilon_{i-1} + e_i$. The local alternative sequences are:

$$g(x) = a + bx + (nk_n)^{-1/4} l(x), \tag{3.3}$$

where $l(x)$ is Lipschitz continuous, and satisfies $\int l(x)r(x)dx = \int xl(x)r(x)dx = 0$ and $\int l^2(x)r(x)dx < \infty$. Assume that the conditions of Theorem 3.1 hold, then under (3.3), we can show

$$\left(\frac{n}{k_n}\right)^{1/2} T_n \to N\left((1-\phi)^2 \int l^2(x)r(x)dx, \frac{4}{3}\sigma^4\right)$$

in distribution. The proof is analogous to that in Wang, Akritas and Van Keilegom (2002). The same proof applies when $g(\cdot)$ has some general parametric form but the bias term will be different.

Remark 3. The above asymptotic results also hold in the stochastic design setting where the design points $X_1, \ldots, X_n$ are ordered values of independent and identically distributed random variables whose density is $r(x)$. In the random design case, we need the additional assumption that the X_i's and the random errors e_i's are independent.

4 A Bootstrap-based Testing Procedure

As has been observed by several authors (González-Manteiga and Vilar-Fernández (1995), Woodward, Bottone and Gray (1997), Kim and Hart (1998), Vilar-Fernández and González-Manteiga (2000)), lack-of-fit tests in regression with time-series errors that are based on asymptotic distributions tend to overreject in finite sample situations, especially when the autoregressive coefficients are positive and close to 1. To avoid size distortion, we propose to bootstrap the test statistic. More specifically, we adopt the residual bootstrap approach for autoregressions. Consider the following estimator for the noise e_i:

$$\widehat{e_i} = \left(Y_i - \sum_{j=1}^{p} \widehat{\phi}_j Y_{i-j} \right) - \left(g(x_i, \widehat{\theta}) - \sum_{j=1}^{p} \widehat{\phi}_j g(x_{i-j}, \widehat{\theta}) \right)$$

for $i = p+1, ..., m$ (note that $\widehat{e_i} = Z_{i-p}$), where the $\widehat{\phi}$'s are the difference-based estimators of Hall and Van Keilegom (2003), and $\widehat{\theta}$ is a $\sqrt{n}$-consistent estimator of θ under the null hypothesis. Bickel and Freedman (1983) suggested using the scaled residual $\left(\frac{m}{m-1}\right)^{1/2} \widehat{e_i}$ (an ad hoc action) because the residuals tend to be smaller than the true errors, see also Li and Maddala (1996). Denote the centered $\widehat{e_i}$'s by $\widehat{e_i^c}$ and the empirical distribution function of these $\widehat{e_i^c}$'s by

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=p+1}^{m} I(\widehat{e_i^c} \leq x). \tag{4.1}$$

From $\widehat{F}_n(x)$ we obtain a sample of independent and identically distributed random variables e_i^*, $i = p+1, \ldots, m$, and construct the error series:

$$\epsilon_i^* = \sum_{j=1}^{p} \widehat{\phi}_j \epsilon_{i-j}^* + e_i^*, \quad \text{for} \quad i = p+1, \ldots, m.$$

(We may take $\epsilon_i^* = \widehat{\epsilon_i}$ or 0 for $i \leq p$. We allow a burn-in period to make the effect of the initial value negligible.) The bootstrap sample (x_i, Y_i^*), $i = 1, \ldots, n$, is obtained by letting $Y_i^* = g(x_i, \widehat{\theta}) + \epsilon_i^*$, for $i = 1, \ldots, m$. Let $Z_i^*, \widehat{\theta}^*, \widehat{\phi}_j^*$ and T_n^* be calculated from the bootstrap sample in the same way as $Z_i, \widehat{\theta}, \widehat{\phi}_j$ and T_n were calculated from the original observations. Repeat the above steps a large number of times. The critical value of the test is obtained as the upper-α quantile of the bootstrap distribution of $n^{1/2}k_n^{-1/2}T_n^*$. The null hypothesis is rejected if $n^{1/2}k_n^{-1/2}T_n$ is greater than this critical value.

This residual-based scheme for generating bootstrap samples from a stationary time series process was first used in Freedman and Peters (1984) and then by many other authors, including Efron and Tibshirani (1986), Kreiss and Franke (1992), Hjellvik and Tjøstheim (1995), Kim and Hart (1998) and Vilar-Fernández and González-Manteiga (2000). Our bootstrap-based test is appropriate in the sense that it gives the asymptotically correct α-level under the null hypothesis.

Theorem 4.1 *Assume A1-A5 and assume in addition that* $\widehat{\theta}^* - \widehat{\theta} = O_{p^*}(n^{-1/2})$*. If* $n \to \infty$ *and* $k_n \to \infty$ *such that* $n^{-1}k_n^{3/2} \to 0$*, then under* H_0*,*

$$\sup_t |P^*(n^{1/2}k_n^{-1/2}T_n^* \leq t) - P(n^{1/2}k_n^{-1/2}T_n \leq t)| = o_P(1),$$

where P^* *denotes the probability measure conditional on the original sample.*

We assume the order of autoregression p is known. Kreiss (1997) has shown that the above residual bootstrap also works in the case where the order of the fitted autoregressive model depends on the data, for instance, it may be estimated by a certain order selection procedure.

Note that the same bootstrap testing procedure can also be applied to the random design setting discussed in Remark 3 of Section 3. In that case we replace the x_i's by a bootstrap sample from their empirical distribution, independent of the e_i^*'s.

5 Numerical Results

In this section, we summarize the results from a Monte Carlo study, which sheds light on the finite sample size and power behavior of our test statistic. The data (x_i, Y_i), $i = 1, \ldots, m$, are generated from (1.1). We assume that the errors follow a stationary AR(1) process: $\epsilon_i = \phi\epsilon_{i-1} + e_i$, where e_i, $i = 1, \ldots, m$, constitute an independent and identically distributed sequence of normal random variables with zero mean and variance σ^2. We test the no effect null hypothesis $g(x) = C$. The design points x_i are equal-distant on [0,1].

The bootstrap test is computed as described in the previous section. We generate 500 samples of observations, for each sample 500 bootstrap samples are produced. The autoregressive coefficient is estimated using the estimator of Hall and Van Keilegom (2003). Following their advice, we take $m_1 = n^{0.1}$ and $m_2 = n^{0.5}$. The level of the test is 0.05.

We first investigate the empirical size of our bootstrap test. To reveal the seriousness of the size distortion when the correlation is ignored, we include the results for the nonparametric test of Wang, Akritas and Van Keilegom (2002), which assumes that the data are independent. We also make a comparison with the test of Vogelsang (1998). Vogelsang's test is based on partial sum regressions. It does not require estimation of the serial correlation coefficient and can handle unit root.

Data under the null hypothesis are generated by setting $g(x) = 0$. Table 1 gives the observed size of the bootstrap test when $\phi = -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$, and $\sigma = 0.5$, for sample size $n = m - 1 = 100$ or 200. For the new test and the test assuming independence, different local window sizes k_n are considered.

Put Table 1 about here

We observe that for $\phi \leq 0.6$ both the bootstrap test and Vogelsang's test give quite accurate type I error rates, while for the test assuming independence the size is much too conservative or much too liberal. When $n = 100$, the empirical size is a little liberal for $\phi = 0.6$ but it is close to the desired level when $n = 200$. We also tried $\sigma = 0.2$ and $\sigma = 1$ and found similar results (not reported).

For ϕ equal to 0.8 (positive and strong correlation), even the bootstrap test becomes quite liberal, the empirical level is roughly 0.20 for $n = 100$ and 0.12 for $n = 200$. Increasing the sample size to 400 does not help much. This phenomenon has also been observed by González-Manteiga and Vilar-Fernández (1995), Woodward, Bottone and Gray (1997), Kim and Hart (1998) and Vilar-Fernández and González-Manteiga (2000). Woodward, Bottone and Gray (1997) conjectured that it occurs because the usual estimate (such as the ordinary least squares estimator or the maximum likelihood estimator) of ϕ is biased towards zero. This is indeed the case here. For example, when $n = 100$, using $m_1 = n^{0.1} = 1.6$ and $m_2 = n^{0.5} = 10$ results in an estimator with mean 0.6754 and variance 0.0046 based on 500 runs when the true value of the autocorrelation coefficient is 0.8. In this situation a more accurate estimator can be obtained by using a larger m_1. In the previous example, when we use $m_1 = 8$ and $m_2 = 10$, the estimator of ϕ in 500 runs has mean 0.7472 and variance 0.0053. With this improved estimate of ϕ, the observed levels become 0.056, 0.062 and 0.064 for $k_n = 5$, 7, 9, respectively. An intuitive understanding of why a larger m_1 works better for the strong positive correlation case can be obtained by taking a closer look at (2.5). In the definition of $\hat{\gamma}(0)$, $(n-m)^{-1}\sum_{i=m+1}^{n}\{(D_mY)_i\}^2$ is supposed to approximate $2Var(\epsilon_i)$. For large m, $(D_mY)_i = Y_i - Y_{i-m}$ is not a good approximation of $\epsilon_i - \epsilon_{i-m}$. So m should not be chosen too large. On the other hand, when m is too small, ${(D_mY)_i}^2$ is more or less $\epsilon_i^2 + \epsilon_{i-m}^2 - 2\epsilon_i\epsilon_{i-m}$. For the first two terms, the average (over all i and m) will tend to the variance of ϵ_i. The third term should tend to zero, which is only true when m tends to infinity. This explains why we take m larger than a certain m_1 and less than a certain m_2. When the correlation is strong, the third term can be quite big for small m. In that case the lower bound m_1 should be chosen sufficiently large.

For $n = 200$, Table 1 also displays the empirical levels for the test based on the asymptotic normality given in Theorem 3.1. The observed level is close to the specified level and somewhat conservative when $\phi < 0.6$ and deteriorates rapidly when ϕ becomes larger. Although the computation is faster compared to the bootstrap test, we do not recommend using critical values computed from asymptotic normality due to its unstable behavior. Quite a few authors have observed that the asymptotic theory for the smoothing-based test does not work very well for moderate sample size and have recommended using the bootstrap method as an alternative. In addition to the references given at the beginning of Section 4, Hjellvik and Tjøstheim ((1995), p.355) commented in a similar setting that "It is very difficult to use the asymptotic theory with any degree of confidence...".

For the power analysis, we consider two types of alternatives: linear alternative $g(x) = 1+2x$ and sinusoidal alternative $g(x) = \cos(2x)$. Simulation results are displayed in Table 2, for $\sigma = 1$, $\phi = 0, 0.2, 0.4, 0.6, 0.8$ and sample size $n = 100$. For $\phi = 0.8$, as discussed in the previous paragraph, $m_1 = 8$ and $m_2 = 10$ are used to estimate ϕ. Vogelsang's test is more powerful in the first case since it is designed to detect linear deviation from a constant mean. However, Vogelsang's test has no power to detect the sinusoidal alternative while our bootstrap based nonparametric test gives good power in this case.

Note that in both Table 1 and 2, the choice of k_n does not have a large influence on the level of the bootstrap test. Choosing a bandwidth to maximize the power of the smoothing-based test is still an open problem. See §6.4 of Hart (1997) for some related discussion. Azzalini and Bowman (1991) suggest calculating the p-value for several different choices of the smoothing parameter, and call the plot of P-values versus the smoothing parameter a "significant trace".

Put Table 2 about here

6 Proofs

We give the proof of Theorem 3.1 and Theorem 4.1 for the case $\dim(\Theta) = 1$ and $p = 1$, i.e., $\epsilon_i = \phi\epsilon_{i-1} + e_i$, $i = 1, \ldots, n$, for some $|\phi| < 1$. The case $\dim(\Theta) > 1$ and $p > 1$ can be proved similarly with slightly more complex notation.

Proof of Theorem 3.1. We have

$$\begin{aligned}
Z_i &= (Y_{i+1} - \widehat{\phi}Y_i) - \Big(g(x_{i+1}, \widehat{\theta}) - \widehat{\phi}g(x_i, \widehat{\theta})\Big) \\
&= (Y_{i+1} - g(x_{i+1}, \theta)) + (g(x_{i+1}, \theta) - g(x_{i+1}, \widehat{\theta})) - \phi(Y_i - g(x_i, \theta)) \\
&\quad -\phi(g(x_i, \theta) - g(x_i, \widehat{\theta})) + (\phi - \widehat{\phi})(Y_i - g(x_i, \theta)) + (\phi - \widehat{\phi})(g(x_i, \theta) - g(x_i, \widehat{\theta})) \\
&= e_{i+1} + (g(x_{i+1}, \theta) - g(x_{i+1}, \widehat{\theta})) - \phi(g(x_i, \theta) - g(x_i, \widehat{\theta})) \\
&\quad +(\phi - \widehat{\phi})(Y_i - g(x_i, \theta)) + (\phi - \widehat{\phi})(g(x_i, \theta) - g(x_i, \widehat{\theta})), \qquad (6.1)
\end{aligned}$$

for $i = 1, \ldots, n$. Let $\eta_1(x_i) = g(x_{i+1}, \theta) - g(x_{i+1}, \widehat{\theta})$, $\eta_2(x_i) = \phi(g(x_i, \theta) - g(x_i, \widehat{\theta}))$, $\eta_3(x_i) = (\phi - \widehat{\phi})(Y_i - g(x_i, \theta))$ and $\eta_4(x_i) = (\phi - \widehat{\phi})(g(x_i, \theta) - g(x_i, \widehat{\theta}))$. Let $\boldsymbol{\eta}_1$ be the $nk_n \times 1$ vector of all observations in the hypothetical one-way ANOVA constructed from $(x_i, \eta_1(x_i))$, $i = 1, \ldots, n$. Similarly, we define $\boldsymbol{\eta}_2$, $\boldsymbol{\eta}_3$, $\boldsymbol{\eta}_4$, $\mathbf{e}$ and $\mathbf{Z}$. By (6.1), $\mathbf{Z} = \mathbf{e} + \boldsymbol{\eta}_1 - \boldsymbol{\eta}_2 + \boldsymbol{\eta}_3 + \boldsymbol{\eta}_4$. Thus our test statistic T_n, which is equal to $\mathbf{Z}'\mathbf{A}\mathbf{Z}$ in this new notation, can be decomposed as

$$\begin{aligned}
T_n &= \mathbf{e}'\mathbf{A}\mathbf{e} + 2\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_1 - 2\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_2 + 2\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_3 + 2\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_4 + \boldsymbol{\eta}_1'\mathbf{A}\boldsymbol{\eta}_1 - 2\boldsymbol{\eta}_1'\mathbf{A}\boldsymbol{\eta}_2 + 2\boldsymbol{\eta}_1'\mathbf{A}\boldsymbol{\eta}_3 \\
&\quad +2\boldsymbol{\eta}_1'\mathbf{A}\boldsymbol{\eta}_4 + \boldsymbol{\eta}_2'\mathbf{A}\boldsymbol{\eta}_2 - 2\boldsymbol{\eta}_2'\mathbf{A}\boldsymbol{\eta}_3 - 2\boldsymbol{\eta}_2'\mathbf{A}\boldsymbol{\eta}_4 + \boldsymbol{\eta}_3'\mathbf{A}\boldsymbol{\eta}_3 + 2\boldsymbol{\eta}_3'\mathbf{A}\boldsymbol{\eta}_4 + \boldsymbol{\eta}_4'\mathbf{A}\boldsymbol{\eta}_4. \qquad (6.2)
\end{aligned}$$

Since $\{e_i\}$ are iid, by the result of Wang, Akritas and Van Keilegom (2002), $n^{1/2}{k_n}^{-1/2}\mathbf{e}'\mathbf{A}\mathbf{e} \xrightarrow{d} N\left(0, \frac{4\sigma^2}{3}\right)$, where $\sigma^2 = Var(e_i)$. We need to show that all the other terms in (6.2) are asymptotically negligible. As representative examples, we show

$$n^{1/2}k_n^{-1/2}\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_1 \xrightarrow{p} 0, \quad n^{1/2}k_n^{-1/2}\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_3 \xrightarrow{p} 0, \quad n^{1/2}k_n^{-1/2}\boldsymbol{\eta}_3'\mathbf{A}\boldsymbol{\eta}_3 \xrightarrow{p} 0. \qquad (6.3)$$

We first show that $n^{1/2}k_n^{-1/2}\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_1 \xrightarrow{p} 0$.

$$\begin{aligned}
&\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_1 \\
&= \frac{nk_n - 1}{n(n-1)k_n(k_n-1)} \sum_{i=1}^n \left[\sum_{j=1}^n \eta_1(x_j) I(j \in W_i)\right] \left[\sum_{k=1}^n e_{k+1} I(k \in W_i)\right] \\
&\quad - \frac{1}{n(n-1)k_n} \left[k_n \sum_{i=1}^n \eta_1(x_i)\right] \left[k_n \sum_{k=1}^n e_{k+1}\right] - \frac{k_n}{n(k_n-1)} \sum_{i=1}^n \eta_1(x_i) e_{i+1}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{nk_n - 1}{n(n-1)(k_n-1)} \sum_{i=1}^{n} \eta_1(x_i) \left[\sum_{k=1}^{n} e_{k+1} I(k \in W_i)\right] \\
&\quad - \frac{k_n}{n(n-1)} \left[\sum_{i=1}^{n} \eta_1(x_i)\right] \left[\sum_{k=1}^{n} e_{k+1}\right] - \frac{k_n}{n(k_n-1)} \sum_{i=1}^{n} \eta_1(x_i) e_{i+1} \\
&\quad + \frac{nk_n - 1}{n(n-1)(k_n-1)} O_p(n^{-3/2} k_n) \sum_{i=1}^{n} \left|\sum_{k=1}^{n} e_{k+1} I(k \in W_i)\right|, \qquad (6.4)
\end{aligned}
$$

since $|\eta_1(x_i) - \eta_1(x_j)| \leq \sup_{\theta,x} \left|\frac{\partial^2}{\partial\theta\partial x} g(x,\theta)\right| |\widehat{\theta} - \theta| \, |x_{i+1} - x_{j+1}| = O_p(n^{-3/2})$ uniformly in $1 \leq i \leq n$ and $j \in W_i$. It follows that (6.4) is

$$
\begin{aligned}
&\frac{(nk_n - 1)k_n}{n(n-1)(k_n-1)} \sum_{k=1}^{n} \eta_1(x_k) e_{k+1} - \frac{k_n}{n(n-1)} \left[\sum_{i=1}^{n} \eta_1(x_i)\right] \left[\sum_{k=1}^{n} e_{k+1}\right] \\
&- \frac{k_n}{n(k_n-1)} \sum_{i=1}^{n} \eta_1(x_i) e_{i+1} + \frac{nk_n - 1}{n(n-1)(k_n-1)} O_p(n^{-3/2} k_n^2) \sum_{k=1}^{n} |e_{k+1}| \\
= \; & \frac{k_n}{n-1} \sum_{k=1}^{n} \left(\eta_1(x_k) - n^{-1} \sum_{i=1}^{n} \eta_1(x_i)\right) e_{k+1} + \frac{nk_n - 1}{n(n-1)(k_n-1)} O_p(n^{-3/2} k_n^2) \sum_{k=1}^{n} |e_{k+1}| \\
= \; & \frac{k_n}{n-1} (\theta - \widehat{\theta}) \sum_{k=1}^{n} \left(\frac{\partial}{\partial\theta} g(x_{k+1}, \theta) - n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} g(x_{i+1}, \theta)\right) e_{k+1} \\
&+ \frac{k_n}{2(n-1)} (\theta - \widehat{\theta})^2 \sum_{k=1}^{n} \left(\frac{\partial^2}{\partial\theta^2} g(x_{k+1}, \tilde{\theta}) - n^{-1} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} g(x_{i+1}, \tilde{\theta})\right) e_{k+1} + O_p(n^{-3/2} k_n^2) \\
= \; & O_p(n^{-1} k_n) + O_p(n^{-3/2} k_n^2),
\end{aligned}
$$

where $\tilde{\theta}$ is between θ and $\widehat{\theta}$, and where the last step follows from the Central Limit Theorem for triangular arrays (use Liapunov's condition and the fact that $E|e_i|^4 < \infty$). Thus $n^{1/2} k_n^{-1/2} \mathbf{e}' \mathbf{A} \boldsymbol{\eta}_1 = O_p(n^{-1/2} k_n^{1/2}) + O_p(n^{-1} k_n^{3/2}) = o_p(1)$. We now look at $\mathbf{e}' \mathbf{A} \boldsymbol{\eta}_3$.

$$
\begin{aligned}
&\mathbf{e}' \mathbf{A} \boldsymbol{\eta}_3 \\
&= \frac{nk_n - 1}{n(n-1)k_n(k_n-1)} (\phi - \widehat{\phi}) \sum_{i=1}^{n} \left[\sum_{j=1}^{n} \epsilon_j I(j \in W_i)\right] \left[\sum_{k=1}^{n} e_{k+1} I(k \in W_i)\right] \\
&\quad - \frac{k_n}{n(n-1)} (\phi - \widehat{\phi}) \left[\sum_{i=1}^{n} \epsilon_i\right] \left[\sum_{k=1}^{n} e_{k+1}\right] - \frac{k_n}{n(k_n-1)} (\phi - \widehat{\phi}) \sum_{i=1}^{n} \epsilon_i e_{i+1} \\
&= D_1 - D_2 - D_3,
\end{aligned}
$$

where the definition of D_i, $i = 1, 2, 3$, should be clear from the context. Evaluate D_3 first. Since $\epsilon_j = \phi \epsilon_{j-1} + e_j$, and since the e_i's $(-\infty < i < \infty)$ are independent, e_{i+1} is independent of ϵ_j for $j \le i$. Hence,

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} \epsilon_i e_{i+1}\right)^2 &= \sum_{i=1}^{n} E(\epsilon_i^2 e_{i+1}^2) + 2E\left(\sum_{i<i'} \epsilon_i \epsilon_{i'} e_{i+1} e_{i'+1}\right) \\
&= \frac{n\sigma^4}{1-\phi^2} + 0 = O(n).
\end{aligned}
$$

Thus $n^{1/2}k_n^{-1/2}D_3 = n^{1/2}k_n^{-1/2}O_p(n^{-1}) = O_p(n^{-1/2}k_n^{-1/2}) = o_p(1)$. Now check D_2. Note that $\sum_{k=1}^n e_{k+1} = O_p(n^{1/2})$ and $\sum_{i=1}^n \epsilon_i = O_p(n)$, and so $D_2 = O_p(n^{-1}k_n) = o_p(n^{-1/2}k_n^{1/2})$. Finally consider D_1. We first evaluate the order of $\sum_{j,k}\sum_{j',k'} E(\epsilon_j \epsilon_{j'} e_{k+1} e_{k'+1})$. Consider three cases for the four indices in the sum: (1) (j, j', k, k') contains exactly two pairs, the order of such terms is $O(n^2)$; (2) (j, j', k, k') contains exactly one pair, the order of such terms is $O(n^3)$; (3) there is no pair in (j, j', k, k'). For case (3),

$$
\begin{aligned}
& \sum_{j,j',k,k' \text{ are distinct}}^{n} E(\epsilon_j \epsilon_{j'} e_{k+1} e_{k'+1}) \\
= \;& 4\sum_{j<j'} \sum_{k<k',j,j',k,k' \text{ are distinct}} E(\epsilon_j \epsilon_{j'} e_{k+1} e_{k'+1}) \\
= \;& 4\sum_{k<k'<j<j'} E(\epsilon_j \epsilon_{j'} e_{k+1} e_{k'+1}) + 4\sum_{k<j<k'<j'} E(\epsilon_j \epsilon_{j'} e_{k+1} e_{k'+1}) \\
& +4\sum_{j<k<k'<j'} E(\epsilon_j \epsilon_{j'} e_{k+1} e_{k'+1}) \\
= \;& Q_1 + Q_2 + Q_3,
\end{aligned}
$$

where the definition of Q_i, for $i = 1, 2, 3$, is clear from the context. Look at Q_1 first. Since for any k, $\epsilon_j = \phi^{j-k}\epsilon_k + \sum_{\ell=1}^{j-k} \phi^{j-k-\ell} e_{k+\ell}$, $\epsilon_{j'} = \phi^{j'-k}\epsilon_k + \sum_{\ell=1}^{j'-k} \phi^{j'-k-\ell} e_{k+\ell}$, the only terms in the product $\epsilon_j \epsilon_{j'} e_{k+1} e_{k'+1}$ with nonzero expectation are $\phi^{j-k-1}e_{k+1}^2 \phi^{j'-k'-1} e_{k'+1}^2$

and $\phi^{j-k'-1}e_{k'+1}^2\phi^{j'-k-1}e_{k+1}^2$. Thus,

$$
\begin{aligned}
& Q_1 \\
= \;& O\left(\sum_{k<k'<j<j'}^{n} \phi^{j+j'-k-k'-2}\right) \\
= \;& O\left(\sum_{j'=4}^{n}\sum_{j=3}^{j'-1}\sum_{k'=2}^{j-1}\sum_{k=1}^{k'-1} \phi^{j+j'-k-k'-2}\right) \\
= \;& O\left(\sum_{j'=4}^{n}\sum_{j=3}^{j'-1}\sum_{k'=2}^{j-1} \phi^{j+j'-k'-2}\frac{\phi^{-(k'-1)}-1}{1-\phi}\right) \qquad (6.5)\\
= \;& O\left(\sum_{j'=4}^{n}\sum_{j=3}^{j'-1}\sum_{k'=2}^{j-1} \phi^{j+j'-2k'-1}\right) \\
= \;& O\left(\sum_{j'=4}^{n}\sum_{j=3}^{j'-1} \phi^{j+j'-1}\frac{\phi^{-2(j-2)}-1}{\phi^2-\phi^4}\right) \\
= \;& O(n^2),
\end{aligned}
$$

since $|\phi^{j+j'-1}\frac{\phi^{-2(j-2)}-1}{\phi^2-\phi^4}| \leq 1$ is bounded. Q_2 and Q_3 can be evaluated similarly. Similarly as above, we can show

$$
E\left[\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j,k}\sum_{j',k'} \epsilon_j\epsilon_{j'}e_{k+1}e_{k'+1}I(j,k\in W_i)I(j',k'\in W_{i'})\right] = O(n^2k_n^2),
$$

hence $n^{1/2}k_n^{-1/2}D_1 = O_p(k_n^{-1/2}) = o_p(1)$. We thus obtain $n^{1/2}k_n^{-1/2}\mathbf{e}'\mathbf{A}\boldsymbol{\eta}_3 \xrightarrow{p} 0$. At last, we evaluate $\boldsymbol{\eta}_3'\mathbf{A}\boldsymbol{\eta}_3$.

$$
\begin{aligned}
& \boldsymbol{\eta}_3'\mathbf{A}\boldsymbol{\eta}_3 \\
= \;& \frac{nk_n-1}{n(n-1)k_n(k_n-1)}(\phi-\widehat{\phi})^2\sum_{i=1}^{n}\left[\sum_{j=1}^{n}\epsilon_j I(j\in W_i)\right]^2 \\
& -\frac{k_n}{n(n-1)}(\phi-\widehat{\phi})^2\left[\sum_{i=1}^{n}\epsilon_i\right]^2 - \frac{k_n}{n(k_n-1)}(\phi-\widehat{\phi})^2\sum_{i=1}^{n}\epsilon_i^2 \\
= \;& (\phi-\widehat{\phi})^2(E_1-E_2-E_3) = O_p(n^{-1})(E_1-E_2-E_3),
\end{aligned}
$$

where the definition of E_1, E_2 and E_3 is clear from the context. Look at E_3 first : $E_3 \geq 0$ and

$$\begin{aligned} E(E_3) &= \frac{k_n}{n(k_n - 1)} \sum_{i=1}^{n} E(\epsilon_i^2) \\ &= O(n^{-1}) \frac{n\sigma^2}{1 - \phi^2} = O(1), \end{aligned}$$

thus $n^{1/2} k_n^{-1/2} O_p(n^{-1}) E_3 = o_p(1)$. Next, $E_2 \geq 0$ and $E(E_2) = O(n^{-2} k_n n^2) = O(k_n)$. So $n^{1/2} k_n^{-1/2} O_p(n^{-1}) E_2 = O_p(n^{-1/2} k_n^{1/2}) = o_p(1)$. Now check E_1, we have $E_1 \geq 0$ and $E(E_1) = O(n^{-1} k_n^{-1} n k_n^2) = O(k_n)$, and so $O_p(n^{-1}) E_1 = O_p(n^{-1} k_n) = o_p(n^{-1/2} k_n^{1/2})$. The derivation for the remaining terms in (6.2) is similar or easier than the derivation for the three terms we have considered. The result follows. □

To prove Theorem 4.1, we use the following two lemmas.

Lemma 6.1 *Let F be the cumulative distribution function of e_i. For any $k \geq 1$ and $\widehat{F}_n$ defined in (4.1)*

$$d_k(\widehat{F}_n, F) \to 0, \ \textit{in probability under } H_0,$$

where d_k denotes Mallow's distance (see Shao and Tu (1995), pp. 73-74).

Proof. Denote by F_n the usual empirical distribution function based on the unobserved $e_2, \ldots, e_m$. By Bickel and Freedman ((1981), Lemma 8.4),

$$d_k(F_n, F) \to 0, \text{as } n \to \infty, \text{ almost everywhere.} \tag{6.6}$$

Let J be the Laplace distribution on $\{2, \ldots, m\}$, that is, $J = j$ with probability $\frac{1}{n}$, $j = 2, \ldots, m$. We define random variables U_1 and V_1 with marginals F_n and $\widehat{F}_n$ according to $U_1 = e_J$ and $V_1 = \widehat{e}_J$ respectively. We have

$$d_k^k(F_n, \widehat{F}_n) = \inf_{U,V} E(U - V)^k \leq E(U_1 - V_1)^k = \frac{1}{n} \sum_{j=2}^{m} (e_j - \widehat{e}_j)^k,$$

where the infimum is taken over all random variables U and V that have distribution F_n and $\widehat{F}_n$ respectively. Under H_0, $e_j = (Y_j - \phi Y_{j-1}) - (g(x_j, \theta) - \phi g(x_{j-1}, \theta))$, $\widehat{e}_j = (Y_j - \widehat{\phi} Y_{j-1}) - (g(x_j, \widehat{\theta}) - \widehat{\phi} g(x_{j-1}, \widehat{\theta}))$, and so we have

$$\begin{aligned} e_j - \widehat{e}_j &= (\widehat{\phi} - \phi) Y_{j-1} + (g(x_j, \widehat{\theta}) - g(x_j, \theta)) \\ &\quad - (\widehat{\phi} - \phi) g(x_{j-1}, \widehat{\theta}) - \phi(g(x_{j-1}, \widehat{\theta}) - g(x_{j-1}, \theta)). \end{aligned}$$

From the consistency of $\widehat{\theta}$ and $\widehat{\phi}$ (see Hall and Van Keilegom (2003) for the latter), it is easy to show that $\frac{1}{n}\sum_{j=2}^{m}(e_j - \widehat{e}_j)^k = o_p(1)$ and thus $d_k(F_n, \widehat{F}_n) = o_p(1)$. Combined with (6.6), we finish the proof. □

Lemma 6.2 *Under the conditions of Theorem 4.1,* $\max_{1\le j\le p} |\widehat{\phi}_j^* - \widehat{\phi}_j| = O_{p^*}(n^{-1/2})$.

Proof. We show that $\sup_{0\le j\le n} |\widehat{\gamma}^*(j) - \widehat{\gamma}(j)| = O_{p^*}(n^{-1/2})$, where $\widehat{\gamma}^*(j)$ is obtained by replacing Y by Y^* in the definition of $\widehat{\gamma}(j)$. The result then follows immediately. For this, we closely follow the proof of (2.11) in Hall and Van Keilegom (2003), up to the point where it is shown that $\widehat{\gamma}(j) = \gamma(j) + \Delta_j + o_p(n^{-1/2})$ and $\Delta_j = O_p(n^{-1/2})$. The proof of their (4.1)-(4.3) and of the rate of Δ_j follows from calculations similar to the ones in (6.5). In the bootstrap world the analogue of this derivation continues to hold provided $\sup_{1\le j\le p} |\widehat{\phi}_j - \phi_j| = o_p(1)$ and $E^*|e_j^*|^k \xrightarrow{p} E|e_j|^k$ for $k = 1, 2, 3, 4$. The former assertion follows from Hall and Van Keilegom (2003), the latter follows from Lemma 6.1 in combination with Lemma 8.3 in Bickel and Freedman (1981). □

Proof of Theorem 4.1. We have

$$\begin{aligned} T_n^* &= (\mathbf{e}^*)'\mathbf{A}\mathbf{e}^* + 2(\mathbf{e}^*)'\mathbf{A}\boldsymbol{\eta}_1^* - 2(\mathbf{e}^*)'\mathbf{A}\boldsymbol{\eta}_2^* + 2(\mathbf{e}^*)'\mathbf{A}\boldsymbol{\eta}_3^* + 2(\mathbf{e}^*)'\mathbf{A}\boldsymbol{\eta}_4^* + (\boldsymbol{\eta}_1^*)'\mathbf{A}\boldsymbol{\eta}_1^* \\ &\quad -2(\boldsymbol{\eta}_1^*)'\mathbf{A}\boldsymbol{\eta}_2^* + 2(\boldsymbol{\eta}_1^*)'\mathbf{A}\boldsymbol{\eta}_3^* + 2(\boldsymbol{\eta}_1^*)'\mathbf{A}\boldsymbol{\eta}_4^* + (\boldsymbol{\eta}_2^*)'\mathbf{A}\boldsymbol{\eta}_2^* - 2(\boldsymbol{\eta}_2^*)'\mathbf{A}\boldsymbol{\eta}_3^* - 2(\boldsymbol{\eta}_2^*)'\mathbf{A}\boldsymbol{\eta}_4^* \\ &\quad +(\boldsymbol{\eta}_3^*)'\mathbf{A}\boldsymbol{\eta}_3^* + 2(\boldsymbol{\eta}_3^*)'\mathbf{A}\boldsymbol{\eta}_4^* + (\boldsymbol{\eta}_4^*)'\mathbf{A}\boldsymbol{\eta}_4^*, \end{aligned} \tag{6.7}$$

where $\eta_1^*(x_i) = g(x_{i+1}, \widehat{\theta}) - g(x_{i+1}, \widehat{\theta}^*)$, $\eta_2^*(x_i) = \widehat{\phi}(g(x_i, \widehat{\theta}) - g(x_i, \widehat{\theta}^*))$, $\eta_3^*(x_i) = (\widehat{\phi} - \widehat{\phi}^*)(Y_i^* - g(x_i, \widehat{\theta}))$, $\eta_4^*(x_i) = (\widehat{\phi} - \widehat{\phi}^*)(g(x_i, \widehat{\theta}) - g(x_i, \widehat{\theta}^*))$, and $\boldsymbol{\eta}_j^*$ $(j = 1, 2, 3, 4)$ (respectively $\mathbf{e}^*$) is the $nk_n \times 1$ vector of all observations in the hypothetical one-way ANOVA constructed from $(x_i, \eta_j^*(x_i))$ (respectively (x_i, e_i^*)), $i = 1, \dots, n$. Similarly as in the proof of Theorem 2.2 in Wang, Akritas and Van Keilegom (2002), we can show that $n^{1/2}k_n^{-1/2}(V^*)^{-1/2}(\mathbf{e}^*)'\mathbf{A}\mathbf{e}^* \xrightarrow{d^*} N(0, 1)$, where

$$V^* = \frac{2}{n(k_n - 1)^2 k_n}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j\ne l} Var^*(e_j^*)Var^*(e_l^*)I(j, l \in W_{i_1} \cap W_{i_2}).$$

By Lemma 8.3 in Bickel and Freedman (1981) and Lemma 6.1, $\text{Var}^*(e_j^*) \xrightarrow{p} \sigma^2$. It is thus easy to show that $V^* \xrightarrow{p} \frac{4}{3}\sigma^4$. It remains to show that all the other terms in (6.7) are $o_{p^*}(1)$. First note that the consistency of $\widehat{\theta}^*$ and $\widehat{\phi}_j^*$ follows from the statement of the theorem and Lemma 6.2 respectively. The proof is now similar to that of Theorem 3.1 (for the verification of Liapunov's condition, use the fact that $E^*|e_j^*|^4 \xrightarrow{p} E|e_j|^4 < \infty$, and to derive (6.5) recall $\widehat{\phi} - \phi \xrightarrow{p} 0$ and $\text{Var}^*(e_j^*) \xrightarrow{p} \sigma^2$). This finishes the proof. □

References

[1] Akritas, M. G. and Papadatos, N. (2004). Heteroscedastic one-way ANOVA and lack-of-fit test. *J. Amer. Statist. Assoc.* **99**, 368-382.

[2] Azzalini, A. and Bowman, A. W. (1991). Nonparametric methods for repeated measurements. In *Nonparametric Functional Estimation and Related Topics*, (Ed. by G. Roussas), pp. 377-387. NATA ASI Series C, Mathematical and Physical Sciences. Kluwer Academic Publishers, Dordrecht.

[3] Bai, J. (1996). Testing for parametric constancy in linear regressions: an empirical distribution function approach. *Econometrica* **64,** 597–622.

[4] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9,** 1196–1217.

[5] Bickel, P.J. and Freedman, D.A. (1983). Bootstrapping regression models with many parameters. In: *A Festschrift for Erich Lehmann.* (Ed. by Bickel, P., Doksum, K. and Hodges, J. L.), pp. 28–48. Wadsworth, Belmont, California.

[6] Brillinger, D.R. (1989). Consistent detection of a monotonic trend superposed on a stationary time series. *Biometrika* **76,** 23–30.

[7] Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* **1,** 54–77.

[8] Freedman, D.A. and Peters, S.C. (1984). Bootstrapping a regression equation: some empirical results. *J. Amer. Statist. Assoc.* **79,** 97–106.

[9] Gallant, A.R. and Goebel, J.J. (1976). Nonlinear regression with autocorrelated errors. *J. Amer. Statist. Assoc.* **71,** 961–967.

[10] González-Manteiga, W. and Vilar-Fernández, J.M. (1995). Testing linear regression models using non parametric regression estimators when errors are non-independent. *Comput. Statist. Data Anal.* **20,** 521–541.

[11] Hall, P. and Van Keilegom, I. (2003). Using difference-based methods for inference in nonparametric regression with time-series errors. *J. Roy. Statist. Soc. Ser. B* **65,** 443–456.

[12] Hart, J. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. *J. Roy. Statist. Soc. Ser. B* **56,** 529–542.

[13] Hart, J. (1997). *Nonparametric Smoothing and Lack-of-fit Test*. Springer-Verlag, New York.

[14] Hjellvik, V. and Tjøstheim, D. (1995). Nonparametric tests of linearity for time series. *Biometrika* **82,** 351-368.

[15] Kim, J.H. and Hart, J.D. (1998). Test for change in a mean function when the data are dependent. *J. Time Ser. Anal.* **19,** 399–424.

[16] Kreiss, J.-P. (1997). Asymptotical properties of residual bootstrap for autoregressions. *Preprint*, Institut für Mathematische Stochastik, Technische Universität Braunschweig, Germany.

[17] Kreiss, J.-P. and Franke, J. (1992). Bootstrapping stationary autoregressive moving-average models. *J. Time Ser. Anal.* **13,** 297–317.

[18] Li, H. and Maddala, G.S. (1996). Bootstrapping time series models (and discussions). *Econometric Reviews* **15,** 115–158.

[19] Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12,** 1215–1230.

[20] Shao, J. and Tu, DS. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.

[21] Sun, H. and Pantula, S. (1999). Testing for trends in correlated data. *Statist. Probab. Lett.* **41,** 87–95.

[22] Vilar-Fernández, J.M. and González-Manteiga, W. (2000). Resampling for checking linear regression models via non-parametric regression estimation. *Comput. Statist. Data Anal.* **35,** 211–231.

[23] Vogelsang, T.J. (1998). Trend function hypothesis testing in the presence of serial correlation. *Econometrica* **66,** 123–148.

[24] Wang, L. and Akritas, M. G. (2002). Two-way heteroscedastic ANOVA when the number of levels is large. *Submitted.*

[25] Wang, L. Akritas, M. G. and Van Keilegom, I. (2002). Nonparametric goodness-of-fit test for heteroscedastic regression models. *Submitted.*

[26] Woodward, W.A. Bottone, S. and Gray, H.L. (1997). Improved tests for trend in time series data. *Journal of Agricultural, Biological, and Environmental Statistics* **4,** 403-416.

[27] Woodward, W.A. and Gray, H.L. (1993). Global warming and the problem of testing for trend in time series data. *J. Climate* **6,** 953-962.

[28] Woodward, W.A. and Gray, H.L. (1995). Selecting a model for detecting the presence of a trend. *J. Climate* **8,** 1929-1937.

Table 1: Empirical level of several tests for sample size 100 and 200 ($\alpha = 0.05$).

		sample size $n = 100$								
		ϕ								
Test	k_n	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8
New	5	0.054	0.066	0.048	0.054	0.062	0.050	0.064	0.088	0.2140
	7	0.050	0.058	0.050	0.046	0.068	0.046	0.072	0.084	0.2020
	9	0.048	0.056	0.048	0.044	0.066	0.054	0.070	0.092	0.1900
Indep.	5	0.000	0.000	0.000	0.002	0.065	0.537	0.954	0.996	1.000
	7	0.000	0.000	0.000	0.002	0.050	0.437	0.892	0.992	1.000
	9	0.000	0.000	0.000	0.003	0.048	0.377	0.828	0.986	1.000
Vogelsang		0.064								

		sample size $n = 200$								
		ϕ								
Test	k_n	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8
New	7	0.048	0.056	0.062	0.040	0.050	0.062	0.050	0.066	0.132
	9	0.048	0.056	0.064	0.048	0.040	0.060	0.048	0.058	0.116
	11	0.046	0.054	0.066	0.060	0.050	0.060	0.048	0.066	0.108
Indep.	7	0.000	0.000	0.000	0.002	0.070	0.644	0.992	1.000	1.000
	9	0.000	0.000	0.000	0.002	0.060	0.560	0.972	1.000	1.000
	11	0.000	0.000	0.000	0.002	0.056	0.514	0.950	1.000	1.000
Asy.	7	0.024	0.026	0.032	0.032	0.036	0.026	0.054	0.180	0.612
	9	0.026	0.022	0.034	0.036	0.040	0.036	0.066	0.192	0.566
	11	0.028	0.022	0.036	0.042	0.036	0.044	0.076	0.188	0.542
Vogelsang		0.047								

Note: "New" is the bootstrap test proposed in this paper; "Indep." is the test of Wang, Akritas and Van Keilegom (2002) for independent data; "Asy." is the test based on asymptotic normality and "Vogelsang" is the test of Vogelsang (1998).

Table 2: Power of the bootstrap test and the test of Vogelsang for $n = 100$ and $\alpha = 0.05$.

$g(x) = 1 + 2x$

		ϕ				
Test	k_n	0.0	0.2	0.4	0.6	0.8
New	5	0.975	0.858	0.618	0.498	0.158
	7	0.980	0.840	0.605	0.438	0.163
	9	0.983	0.843	0.590	0.433	0.142
Vogelsang		0.812				

$g(x) = \cos(2x)$

		ϕ				
Test	k_n	0.0	0.2	0.4	0.6	0.8
New	5	0.863	0.690	0.455	0.303	0.130
	7	0.863	0.668	0.435	0.275	0.110
	9	0.860	0.680	0.438	0.260	0.110
Vogelsang		0.000				

Note: "New" is the bootstrap test proposed in this paper; "Vogelsang" is the test of Vogelsang (1998).