A simple additivity test for conditionally heteroscedastic nonlinear autoregression

Michael Levine [*], [†‡]Jinguang (Tony) Li[§]

Abstract

In this article we propose a test for additivity of a nonlinear conditionally heteroscedastic autoregressive model. A test is based on the unequal variance unbalanced design ANOVA scheme. Asymptotic distribution of the test statistic is derived and the test performance in finite samples is studied using simulation. To the best of our knowledge, this is the first additivity test for a conditionally heteroscedastic time series model.

1 Introduction

Historically, nonlinear nonparametric time series models acquired significant popularity when it became clear that it was hard to choose an appropriate class of model in a real application study. For a good example of such a situation see, for example, Lingjaerde et al (2001). "Letting the data speak for itself" became an important principle suggesting the choice of nonparametric models in real life applications. One of the most popular of these models is the autoregressive model.

A general nonlinear autoregressive model for a time series y_t is

$$y_t = f(y_{t-k_1}, \dots, y_{t-k_p}) + \varepsilon_t \tag{1}$$

[*]Corresponding author. Address: 250 N. University St., Purdue University, West Lafayette, IN 47907-2066. E-mail: mlevins@purdue.edu. Phone: 765-496-7571. Fax: 765-494-0558.

[†]The work of M. Levine was partially supported by the NSF-DMS grant # 0805748. This manuscript was finished during M. Levin's sabbatical leave at SAMSI. M. Levine would like to thank SAMSI for its hospitality and support.

[‡]Keywords and phrases: nonlinear autoregressive model, additivity test, heteroscedasticity, unbalanced ANOVA with unequal variances

[§]Address: 250 N. University St., Purdue University, West Lafayette, IN 47907-2066.

where k_i, $i = 1, \ldots, p$ and p are positive integers, $f(\cdot)$ is p-dimensional real-valued function and ε_t is a white noise series with mean zero and finite variance. Such a model was first considered in Jones (1978). Robinson (1983) also considered its estimation when y_t is stationary and strongly mixing. However, estimating this model in its full generality immediately leads to the curse of dimensionality. Moreover, even if the sufficient amount of data is available, interpretation of such a model is also very difficult. This is why some simplifying assumptions on the structure of the conditional mean function $f(\cdot)$ are commonly made. The most common of them is the additivity assumption. The resulting generalized additive model assumes that $f(y_{t-k_1}, \ldots, y_{t-k_p}) = f_1(y_{t-k_1}) + \ldots + f_p(y_{t-k_p})$ for a set of smooth functions $f_i(\cdot)$, $i = 1, \ldots, p$. In the time series context, the additivity assumption was considered in detail in Chen and Tsay (1993). Such a model is convenient in practice since it only requires nonparametric estimation of functions of one-dimensional argument. It is also easy to interpret in practice. Another common choice, though not investigated in this paper, is modeling conditional mean as a piecewise linear function; a combination of piecewise linear conditional mean function and the classical parametric ARCH structure in the conditional variance was called a "second generation" model in Tong (1990).

In practice, however, one must be cautious before assuming the additive structure of a conditional mean function; preferably, there should be a test that would verify whether the additive structure may, indeed, be true, based on the data. Chen, Liu and Tsay (1995) suggested several tests for that purpose. Chan, Kristoffersen and Stenseth (2003) suggested a Lagrange multiplier test for additivity based on the Bürmann expansion of a conditional mean function. Both of these approaches assume that the conditional variance of the time series y_t is unknown but constant.

In practice, the requirement of a constant conditional variance can be quite restrictive. Conditionally heteroscedastic nonlinear time series models are widely used today in many areas of application, particularly in econometrics. More specifically, general models that incorporate both nonlinear autoregression and nonlinear conditional variance function have been investigated by a number of authors. The bibliography is quite extensive; we will mention some, for example, Li and Li (1996), Liu, Li and Li (1997), Lu (1998), Ling (1999), Hwang and Woo (2001), Lu and Jiang (2001), and Lanne and Saikkonen (2005). However, to the best of our knowledge, there are no tests that are capable of checking for additivity of a nonlinear time series with an unknown non-constant conditional variance. One of the methods considered in Chen, Liu and Tsay (1995) is a conditional mean test that is essentially a likelihood ratio based test. Such a test can also be viewed as based on a two-way unbalanced ANOVA. The method performs very well when the conditional variance is constant; however, it does not seem likely to perform well, in common with other F-type tests, if the homoscedasticity assumption is violated. We design a modified version of the conditional mean test that is capable of handling a conditionally heteroscedastic time series y_t. The test that we suggest is based on an unbalanced design ANOVA with unequal variances scheme first considered by Bishop and Dudewicz (1978). The proposed testing procedure is described in detail in Chapter (2). In the same chapter, we derive the asymptotic result for the proposed test-statistic. We do not specify the exact structure of the variance function in this paper since our main interest lies in additivity testing for the mean function. In practice, variance function can take on a number of possible functional forms as long as sufficient conditions for test statistic convergence to the asymptotic distribution are satisfied. Chapter (3) illustrates numeric properties of the test using a number of suitable models.

2 A Conditional Mean Test

In this section we extend the idea of conditional mean test introduced in Chen, Liu and Tsay (1995) to nonlinear autoregression with conditionally heteroscedastic errors. Such a test would aim to uncover the evidence of possible non-additivity in the conditional mean function when the process itself is conditionally heteroscedastic.

Testing for possible presence of non-additive terms in nonlinear models has been a research topic for some time; as an example, one can mention Sperlich, Yang and Tjöstheim (2002) and references therein. Most of this work has been done in the iid data context; the nonlinear *auto*regression has received relatively little attention. For simplicity, we use the model of order two to illustrate the idea; however, it can be applied to a model of any order $d > 2$. Consider the following nonlinear autoregressive model of order two:

$$y_t = m(y_{t-1}, y_{t-2}) + v^{1/2}(y_{t-1}, y_{t-2})\varepsilon_t \tag{2}$$

where ε_t is a series of iid random variables with mean zero and variance 1 while $m(\cdot,\cdot)$ and $v(\cdot,\cdot)$ are smooth bivariate functions. We will also use the notation $\nu_t \equiv v^{1/2}(y_{t-1}, y_{t-2})\varepsilon_t$. Let $\mathcal{F}_t$ be the $\sigma-$ algebra generated by $\{y_s, s \leq t\}$. We are testing whether the conditional mean function $m(y_{t-1}, y_{t-2}) = E\{y_t|\mathcal{F}_t\}$ can be represented in the additive form or not. Let us use the notation $m(y_1, y_2)$ for the value of the mean function m at an arbitrary point y_1, y_2 in its domain. Then, the null hypothesis is

$$H_0 : m(y_1, y_2) = m_1(y_1) + m_2(y_2)$$

while the alternative is that the conditional mean function is an unstructured $m(y_1, y_2)$. Chen, Liu and Tsay (1995) discussed a similar problem earlier; they assumed, however, that the conditional variance function $v \equiv 1$. Chan, Kristoffersen and Stenseth (2003) proposed a Lagrange multiplier test for additivity based on the Bürmann expansion of a conditional mean function. That test was also designed for the homoscedastic case. The homoscedasticity assumption does not commonly hold in the nonlinear time series and, therefore, the extension to the conditionally heteroscedastic case is needed.

The procedure we suggest can be viewed as a two-way unbalanced ANOVA with unequal variances. This is a fairly difficult problem that has not been extensively studied before. It has been established in the past that a classical F-test is not robust to violation of equal variance assumption, especially when the design itself is unbalanced. A two-stage testing procedure for the two-way unbalanced ANOVA with unequal variances has been proposed by Bishop and Dudewicz (1978). That procedure requires additional samples at the second stage. Chen and Chen (1998) made an improvement by proposing a single stage sampling procedure. This procedure is computationally much more effective than the earlier procedure by Bishop and Dudewicz (1978) and it does not require additional sampling.

We begin with the brief description of the procedure of Chen, Liu and Tsay (1995). Let us assume that N observations had been generated from the model (2). Then, the procedure can be described as follows.

1. Choose a shrinking factor $0 < \delta < 1$ and a positive integer m. Partition the reduced data range $\delta(y_{max} - y_{min})$ into m equal intervals (a_i, a_{i+1}), $i = 0, \ldots, m-1$. The points a_i are defined as $a_i = y_{min} + (1-\delta)(y_{max} - y_{min})/2 + i\delta(y_{max} - y_{min})/m$ for $i = 0, \cdots, m$. The data range has to be shrunk to avoid the boundary bias problem common to many nonparametric smoothing procedures.

2. For $t = 3, \cdots, N$, we classify y_t into the (i, j)-th cell if $y_{t-1} \in (a_{i-1}, a_i)$ and $y_{t-2} \in (a_{j-1}, a_j)$. We denote $X_{ijk} = y_t$, where the third subscript k is used to distinguish different observations in the same cell. If y_{t-1} or y_{t-2} are outside the reduced range, we drop y_t from further consideration. The number of observations in each cell is n_{ij} and the total number of observations is $\sum_{i=1}^m \sum_{j=1}^m n_{ij} = n < N$.

3. When the conditional variances are constant, the model (2) can be viewed as
$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \tag{3}$$
where $i = 1, \cdots, m$, $j = 1, \cdots, m$, $k = 1, 2, \cdots, n_{ij}$, ε_{ijk} are independent random variables with mean zero and variance 1. We assume that identifiability conditions $\sum_{i=1}^{m} \alpha_i = \sum_{j=1}^{m} \beta_j = \sum_{i=1}^{m} \alpha\beta_{ij} = \sum_{j=1}^{m} \alpha\beta_{ij} = 0$ are satisfied. Then, the two-way analysis of variance procedure is carried out to obtain an F statistic for testing the null hypothesis $H_0 : \alpha\beta_{ij} = 0$ for all i and j in the model (3)

The above procedure is based on the heuristic idea that, when functions $m_1(\cdot)$, $m_2(\cdot)$ and $m(\cdot, \cdot)$ are sufficiently smooth, the observations in the same cell have roughly the same conditional mean values. This argument is true in large samples as the number of observations $N \to \infty$ and the number of intervals grows at the rate smaller than N but sufficiently large to ensure that the limiting distribution of the test statistic is non-degenerate. Additionally, under the strong mixing condition, the observations in a cell behave as if they were approximately independent. Therefore, testing the null hypothesis H_0 can be carried out using the usual two-way analysis of variance. When the conditional variance function is not constant, that algorithm cannot be used. We modify the procedure of Chen, Liu and Tsay (1995) to handle conditional heteroscedasticity in the following way. After grouping observations y_t generated by model (2) in accordance with steps (1)-(2) of the algorithm of Chen, Liu and Tsay (1995), we can view the setting as an approximate unbalanced two-way ANOVA model with unequal variances

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \tag{4}$$

$i, j = 1, \ldots, m$ and $k = 1, \ldots, n_{ij}$; in the above, ε_{ijk} are independent random variables with zero mean and unknown finite variances $0 < \sigma_{ij}^2 < \infty$. This is based on the heuristic idea that, when functions $m_1(\cdot)$, $m_2(\cdot)$, $m(\cdot, \cdot)$ and $v(\cdot, \cdot)$ are sufficiently smooth, the observations in the same cell have roughly the same conditional mean and conditional variance values. Yet again, this argument is true in large samples as the number of observations $N \to \infty$ and the number of intervals grows at the rate smaller than N but sufficiently large to ensure that the limiting distribution of the test statistic is non-degenerate. The approximate independence is implied by a strong mixing condition that will be described in detail in the main result statement. Now we can describe the procedure in detail.

1. We use the first $n_{ij}-1$ observations within each cell to compute the sample mean and sample variance as

$$\bar{X}_{ij} = \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}-1} X_{ijk}$$
$$S_{ij}^2 = \frac{1}{n_{ij}-2}\sum_{k=1}^{n_{ij}-1} \left(X_{ijk}-\bar{X}_{ij}\right)^2$$

2. We compute the weights of the observations in a cell (i,j) as follows:

$$U_{ij} = \frac{1}{n_{ij}} + \frac{1}{n_{ij}}\sqrt{\frac{1}{n_{ij}-1}\left(\frac{S_{max}^2}{S_{ij}^2}-1\right)}$$
$$V_{ij} = \frac{1}{n_{ij}} - \frac{1}{n_{ij}}\sqrt{(n_{ij}-1)\left(\frac{S_{max}^2}{S_{ij}^2}-1\right)}$$

   where $S_{max}^2 = \max\{S_{11}^2, S_{12}^2, \cdots, S_{mm}^2\}$.

3. For each cell (i,j), we compute the weighted sample mean as

$$\tilde{X}_{ij\cdot} = \sum_{k=1}^{n_{ij}} W_{ijk} X_{ijk}$$

   where $W_{ijk} = U_{ij}$ if $1 \le k \le n_{ij}-1$ and $W_{ijk} = V_{ij}$ if $k = n_{ij}$. Note that thusly defined weights are normalized: $\sum_{k=1}^{n_{ij}} W_{ijk} = 1$ and $\sum_{k=1}^{n_{ij}} W_{ijk}^2 = \frac{S_{max}^2}{n_{ij}S_{ij}^2}$

4. We compute $\tilde{X}_{i\cdot\cdot} = \frac{1}{m}\sum_{j=1}^{m}\tilde{X}_{ij\cdot}$, $\tilde{X}_{\cdot j\cdot} = \frac{1}{m}\sum_{i=1}^{m}\tilde{X}_{ij\cdot}$ and $\tilde{X}_{\cdot\cdot\cdot} = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\tilde{X}_{ij\cdot}$

5. The test statistic

$$F_m = \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{\tilde{X}_{ij\cdot}-\tilde{X}_{i\cdot\cdot}-\tilde{X}_{\cdot j\cdot}+\tilde{X}_{\cdot\cdot\cdot}}{S_{max}^2/\sqrt{n_{ij}}}\right)^2 \tag{5}$$

   is used to test the modified null hypothesis $\tilde{H}_0 : \alpha\beta_{ij} \equiv 0$.

This procedure can also be directly generalized to the general autoregressive model of order k

$$y_t = m(y_{t-1}, \dots, y_{t-k}) + v^{1/2}(y_{t-1}, \dots, y_{t-k})\varepsilon_t. \qquad (6)$$

Such a generalization entails using a k-way unbalanced ANOVA with unequal variances. In practice, this may become prohibitively expensive computationally for large k. Another generalization, however, is not nearly as straightforward as increasing the number of lags present: the proposed method cannot be applied directly when the conditional variance function depends on the lags of y_t different than those the conditional mean function depends upon; in other words, it will not work if observations y_t are generated by the model $y_t = m(y_{t-k_1}, y_{t-k_2}) + v^{1/2}(y_{t-k_3}, y_{t-k_4})\varepsilon_t$ if the set of double indices (k_1, k_2) is not the same as (k_3, k_4). This is an interesting and potentially important topic for future research.

The following asymptotic result about the test statistic (5) is proved in the Appendix

Theorem 2.1 *Suppose the process y_t generated by the model (2) has a stationary density $p(\cdot)$ and is strongly mixing; moreover, we assume that $E|y_t|^\delta < \infty$ and the mixing coefficients α_j satisfy $\sum_{j=1}^{\infty} \alpha(j)^{1-2/\delta} < \infty$ for some constant $\delta > 2$. We also assume that the process y_t satisfies the positivity condition of Besag (1974): under the stationarity assumption, for any two points y_1, y_2 the joint density $p(y_1, y_2) > 0$ if and only if both $p(y_1) > 0$ and $p(y_2) > 0$. Assume that ε_t are independent random variables with mean zero and variance 1. Let the functions $m_1(\cdot)$ and $m_2(\cdot)$ be continuous differentiable functions with bounded first derivatives while the conditional variance function $v(\cdot, \cdot)$ is bounded from below: there exists $\eta > 0$ such that $v(y_1, y_2) \geq \eta > 0$ for any pair (y_1, y_2). Then, as $N \to \infty$ and the number of intervals $m = m_N$ goes to infinity at the rate higher than $N^{1/3}$ (meaning that $m > CN^{1/3+\rho}$ for some $C > 0$ and $\rho > 0$) the limiting distribution of statistic $F_m \equiv F_{m_N}$ is $\chi^2_{(m-1)^2}$ under the null hypothesis $m(y_{t-1}, y_{t-2}) = m_1(y_{t-1}) + m_2(y_{t-2})$; this means that, the absolute difference between the distribution function of F_{m_N} and $\chi^2_{(m-1)^2}$ distribution function goes to zero as $N \to \infty$ and $m \to \infty$ at the rate of at least $m > CN^{1/3+\rho}$.*

Remarks

1. Theorem (2.1) does not assume much more than what is needed for the result of Chen, Liu and Tsay (1995) to hold. First, in addition to strong mixing property of the process y_t, we require that the mixing

coefficients $\alpha(j)$ satisfy $\sum_{j=1}^{\infty} \alpha(j)^{1-2/\delta} < \infty$ for some $\delta > 2$. That assumption is slightly stronger than simply requiring convergence of the series $\sum_{j=1}^{\infty} \alpha(j)$ and, essentially, imposes some limitations on the convergence rate of the coefficients $\alpha(j) \to 0$ as $j \to \infty$. Second, we require that the conditional variance function $v(\cdot, \cdot)$ be bounded away from zero which is a standard technical assumption. Note that, unlike Chen, Liu and Tsay (1995), we not only obtain the asymptotic distribution of the proposed test statistic as the sample size $N \to \infty$ and the number of intervals $m \to \infty$ but also explicitly characterize the lower bound of the rate at which m has to increase as a fraction of N in order for the test statistic to converge to a non-degenerate distribution under the null hypothesis.

2. The above Theorem suggests that we should reject $\tilde{H}_0$ at a confidence level of $100(1-\alpha)\%$ if F_m exceeds the upper α quantile of $\chi^2_{(m-1)^2}$ for sufficiently large values of N. However, note that in practice the deviation from the asymptotic chi-square distribution of the test statistic is substantial if some n_{ij} increase much slower then others as $N \to \infty$. The number of observations in each cell n_{ij} depends on the window size that is, in turn, controlled by the choice of m. It is possible that n_{ij} is quite small in some cells even for large sample sizes if the chosen m is too large; this happens because observations y_t, $t = 1, \ldots, N$ are not independent. If this happens, selecting large m may interfere with the convergence of the test statistic to its asymptotic distribution under the null hypothesis.

3. Choosing the number of intervals m is a difficult task. This is an exciting topic for future research; for now, we suggest the following simple algorithm that can provide a rough estimate. Note that the test statistic F_m attempts, in effect, to estimate the magnitude of the interactive term $m_{1,2}(y_{t-1}, y_{t-2})$ to see if it has to be included in the model or not. In order to do that, the "main effects" representing additive terms $m_1(y_{t-1})$ and $m_2(y_{t-2})$ have to be subtracted from the cell means first. From the nonparametric function estimation viewpoint, this is similar to estimating additive terms first, subtracting the resulting estimates from observations to obtain the residuals and then estimating interactive term based on these residuals. This suggests the following simple algorithm:

   (a) Estimate the optimal bandwidth for the local linear regression fit of the additive term $m_1(y_{t-1})$ or $m_2(y_{t-2})$. The best approach here is, probably, the K- fold crossvalidation with a relatively small K.

(b) Compute the range of observations y_t

(c) Estimate the optimal number of intervals m as the (rounded to get an integer) ratio of the data range $y_{max} - y_{min}$ to the optimal bandwidth defined in (3a)

(d) Repeat the previous three steps a certain number of times while recording the resulting value of m every time. Choose a representative value of m (such as the median) to be an approximate estimate of the true m.

This algorithm effectively assumes that the same bandwidth is used to fit both $m_1(y_{t-1})$ and $m_2(y_{t-2})$ which stems from using the same m to classify both y_{t-1} and y_{t-2}.

4. Continuing the discussion started in Remark (2), we conjecture that it may be better to partition the reduced range into unequal windows As noted earlier in the remark (3), a more comprehensive solution of this problem may be splitting the reduced data range $\delta(y_{max} - y_{min})$ into unequal intervals whose length is inversely proportional to the stationary density of the process y_t. This will ensure that, in areas where the data is sparse, longer intervals are selected and the number of observations in all of the m^2 cells stays approximately equal. Finding a satisfactory procedure to select m is a very interesting topic for future research.

5. Note that the asymptotic distribution of the test statistic does not depend on the conditional variance function $v(y_1, y_2)$; therefore, the test has the level α regardless of how the conditional variance function looks like. In particular, even if the conditional variance function is the function of only one of the lags used to define conditional mean (e.g. $v \equiv v(y_{t-k})$ with k either 1 or 2), the test still retains its level α.

3 A Simulation Study

For the purpose of Type I error study, we consider the following three additive models with nonconstant variance:

$$y_t = 0.5y_{t-1} + sin(y_{t-2}) + (1 + |\sin(y_{t-1})|)\left(1 + \sqrt{|y_{t-2}|}\right)\varepsilon_t \tag{7}$$

$$y_t = 0.5y_{t-1} + \left(0.5 + \left[0.5y^+_{t-1} + 0.5y^+_{t-2} + 0.5y^-_{t-1} + 0.5y^-_{t-2}\right]\right)^{1/2}\varepsilon_t \tag{8}$$

$$y_t = 0.8y_{t-1} - 0.3y_{t-2} + (1 + |\sin(y_{t-1})|)\left(1 + \sqrt{|y_{t-2}|}\right)\varepsilon_t \tag{9}$$

The first model is an additive AR-ARCH model with multiplicative conditional variance structure. It is very similar to example (1) of Lu and Jiang (2001) except that our model only uses the first two lags of Y_t. The second model is an example of so-called $\beta-$ ARCH(p) model introduced in Guegan and Diebolt (1994) with $p = 2$ and $\beta = 0.5$; we use notation $Y^+ = \max(Y, 0)$ and $Y^- = \max(-Y, 0)$. Note that the conditional mean function in this case depends on Y_{t-1} only since the coefficient of Y_{t-2} is equal to zero. The third model is a linear AR-ARCH model with a multiplicative conditional variance structure that is the same as in (7).

It is well known that, when applying nonparametric methods to nonlinear time series, the sample size has to be relatively large. We use the sample size $N = 2000$ and apply the proposed test to 300 realizations of each model. The errors ε_t are standard normal $N(0, 1)$. In practice, as noted before, the number of observations in cells along and close to the main diagonal is much larger than that in the peripheral cells. Moreover, some cells close to the periphery of the ANOVA table end up empty. The larger the number of intervals m selected, the more obvious this phenomenon becomes. As noted before in the Remark (2), this can impede convergence of the test statistic to the asymptotic $\chi^2_{(m-1)^2}$ distribution because the number of observations in some peripheral cells n_{ij} may not go to infinity as the overall sample size $n \to \infty$. In order to prevent this from happening and keep the number of degrees of freedom correct, some small adjustments are needed. First, we remove rows and/or columns that consist of empty cells only. Second, if there are any cells remaining that have zero or a very small number of observations, we treat all of these cells as empty and employ a simple imputation procedure. Specifically, sample means/sample standard deviations in those cells are presumed to be average sample means/sample standard deviations across the entire ANOVA table ("grand" mean/standard deviation). For purposes of this study we treat cells containing less than 40 observations (2% of the sample size) as "empty".

Table 1: Percentiles of asymptotic p-values of the proposed test under the null hypothesis

	Model (7)		Model (8)		Model (9)	
Probability	$m = 5$	$m = 7$	$m = 5$	$m = 7$	$m = 5$	$m = 7$
0.01	0.003	0.016	0.022	0.013	0.021	0.007
0.05	0.048	0.080	0.049	0.076	0.068	0.048
0.10	0.106	0.130	0.083	0.131	0.133	0.125
0.25	0.343	0.340	0.308	0.311	0.327	0.330
0.50	0.623	0.628	0.645	0.628	0.605	0.581
0.75	0.825	0.868	0.871	0.852	0.831	0.818
0.90	0.923	0.955	0.958	0.961	0.952	0.954
0.95	0.967	0.981	0.981	0.979	0.967	0.978
0.99	0.994	0.999	0.994	0.999	0.995	0.998

Another interesting topic is the choice of the parameter δ. Of course, one tries to remove as little data as possible in practice while at the same time avoiding boundary problems common to all nonparametric procedures. The following observation may help in making this decision. Let the range be denoted $R = y_{max} - y_{min}$. Then the reduced range is δR and it is split into m intervals of equal length $\frac{\delta R}{m}$. It is reasonable to assume that equal length intervals are removed at both ends of the data range; denote that length x. Therefore, we have $\delta R + 2x = R$ and, consequently, $x = \frac{R(1-\delta)}{2}$. Since it makes no sense to remove the area (at either end) that is longer than any of the m intervals, we assume that $x = \frac{R(1-\delta)}{2} < \frac{\delta R}{m}$ or $\frac{1-\delta}{\delta} < \frac{2}{m}$. The function $\frac{1-\delta}{\delta}$ is monotonically decreasing in δ; therefore, one selects δ no smaller than the solution of the equation $\frac{1-\delta}{\delta} = \frac{2}{m}$. That solution is $\delta^*(m) = \frac{m}{m+2}$ and it can be viewed as the lower bound on δ.

In this study, we use $\delta = 0.8$ and two possible choices of $m : m = 5$ and $m = 7$. For $m = 5$, $\delta^*(5) \approx 0.7143$ and for $m = 7$ $\delta^*(7) \approx 0.7778$; since 0.8 is clearly above both of these lower bounds we believe we haven't lost data unnecessarily. For each realization, we compute the value of our statistic and the corresponding p-value with respect to the asymptotic distribution under the null hypothesis of no interactive term present. The Table (1) shows the percentiles of asymptotic p- values for each of the three models. The results are comparable for both choices of m and seem to be relatively good. None of the models demonstrates results considerably worse than others.

Table 2: Percentages of rejection by our test under various alternative hypotheses and window sizes

	Model (10)		Model (11)		Model (12)	
Significance level	$m = 5$	$m = 7$	$m = 5$	$m = 7$	$m = 5$	$m = 7$
0.10	0.663	0.810	0.670	0.710	0.650	0.770
0.05	0.597	0.757	0.600	0.647	0.590	0.683
0.01	0.473	0.667	0.447	0.557	0.463	0.600

The second simulation considers the following three models:

$$y_t = 0.9y_{t-1}\sin(y_{t-2}) + (1 + |\sin(y_{t-1})|)\left(1 + \sqrt{|y_{t-2}|}\right)\varepsilon_t \tag{10}$$

$$\begin{aligned} y_t &= (0.5y_{t-1} - 0.4y_{t-2})I(y_{t-1} < 0) + (0.5y_{t-1} + 0.3y_{t-2})I(y_{t-1} \geq 0) \\ &\quad + (1 + |\sin(y_{t-1})|)\left(1 + \sqrt{|y_{t-2}|}\right)\varepsilon_t \end{aligned} \tag{11}$$

$$y_t = 2\exp(-0.1y_{t-1}^2)y_{t-1} - \exp(-0.1y_{t-1}^2)y_{t-2} + (0.05 + 0.5y_{t-1}^2 + 0.5y_{t-2}^2)\varepsilon_t \tag{12}$$

These models are used to study the power properties of the proposed test. Model (10) is a functional coefficient AR(2) model with a multiplicative ARCH(2) term. A similar model with constant conditional variance had been used before in Chen and Tsay (1993) and Chen, Liu and Tsay (1995). The model (11) is a threshold autoregressive model (TAR) with the discontinuous conditional mean function. It has been chosen, in part, to see if discontinuities in conditional mean impact the performance of the test to any considerable extent. It has the same multiplicative ARCH(2) conditional variance structure as the previous model (10). The model (12) is an exponential autoregressive model with an additive conditional variance structure. The homoscedastic versions of the last two models were used in Chen, Liu and Tsay (1995). The Table (2) shows percentages of rejection by our test under different significance levels for models (10)-(12).

It can be clearly seen that the test has fairly high power for the choices of m used. Chen, Liu and Tsay (1995) note that the power of their conditional mean test seems to rise with the increase in m (i.e. decrease in the "window" size) except when the mean function is discontinuous. We do not observe this effect in our test; even in the case of model (11) the power of test is higher for $m = 7$ then for $m = 5$. One must note, however, that the magnitude of that power increase seems to be less for (11) then for either (10) or (12).

We also tried to use the simple method described in (3) to obtain suggested number of intervals m. We chose $K = 5$ for the K-fold crossvalida-

Table 3: Percentages of rejection by the test of Chen, Liu and Tsay (1995) for models (10)-(12)

Significance level	Model (10)	Model (11)	Model (12)
0.10	0.093	0.097	0.120
0.05	0.047	0.033	0.083
0.01	0.007	0.007	0.023

tion that is used in the algorithm. After rounding, the values obtained were $m_1 = 4$ for the model (10), $m_2 = 5$ for the model (11) and $m_3 = 3$ for the model (12). While the first two values seem fairly reasonable, the last one is somewhat off the mark. It is possible that the reason for that is the strongly nonlinear nature of the conditional mean in the model (12); it is the feature that makes the choice of K for the K-fold crossvalidation more difficult than usual. The problem of selecting optimal m requires extensive future research.

Note that the test we propose is, indeed, necessary because the older additivity test of Chen, Liu and Tsay (1995) cannot handle the heteroscedastic processes adequately. To illustrate that, we again consider a set of models (10)-(12) and apply the test of Chen, Liu and Tsay (1995) to these models. This is equivalent to ignoring conditional heteroscedasticity and using the test designed for conditionally homoscedastic processes. We use $m = 5$ and $\delta = 0.8$ again. The number of simulations is 300 and the sample size is 2000. The results are collected in the Table (3). It is quite clear that the earlier test cannot handle conditionally homoscedastic processes - percentages of rejection of (incorrect) null hypothesis are very small.

Finally, we also want to test for possible loss of efficiency when the true conditional variance of the process y_t is constant. To this end, we use the models (10)-(12) with the conditional variance now being set to 1. The same testing procedure is again used; the value of statistic (5) is computed for each realization and the p-value with respect to the asymptotic $\chi^2_{(m-1)^2}$ distribution is computed. The results are collected in the Table (4). The differences in the power of the test under the two scenarios seem to be very small. It is inconsistent for the three models shown: for the first model a small loss of efficiency is observed, while for the other two homoscedasticity seems to result in a slightly higher test power. Since magnitudes of these differences are very small, the loss of efficiency does not seem to be of major practical concern.

Table 4: Percentages of rejection by our test when the process is homoscedastic

	Model (10)		Model (11)		Model (12)	
Significance level	$m = 5$	$m = 7$	$m = 5$	$m = 7$	$m = 5$	$m = 7$
0.10	0.633	0.720	0.640	0.800	0.700	0.773
0.05	0.567	0.663	0.567	0.763	0.607	0.717
0.01	0.417	0.600	0.457	0.667	0.473	0.623

4 Acknowledgements

Authors would like to use this opportunity to thank Prof. Rong Chen for his valuable comments.

5 Appendix

Proof: If the support of p is not a closed interval, there exists an interval $[a, b]$ such that for any $y \in [a, b]$ $p(y) > 0$ and $p([a, b])$ is bounded away from zero. We denote N^* the number of observations $(y_{t-1}, y_{t-2}) \in [a, b] \times [a, b]$. Due to the positivity condition, $N^* = O(N)$ as $n \to \infty$. This allows us to assume that, for all practical purposes, the stationary density is supported on a closed interval; to simplify notation, we will use N throughout the proof regardless of the true support of the stationary density $p(\cdot)$. Denote c_{ij} a small square with vertices (a_{i-1}, a_{j-1}), (a_i, a_{j-1}), (a_{i-1}, a_j) and (a_i, a_j); also, let $c_i = (a_{i-1}, a_i)$ and $c_j = (a_{j-1}, a_j)$. This square will play the role of the ANOVA cell in the following analysis. The model (2) can be represented as $X_{ijk} = f_{ijk} + \nu_{ijk}$ if $(y_{t-1}, y_{t-2}) \in c_{ij}$ where $X_{ijk} \equiv y_t$ and the index k is used to distinguish the observations falling in the same cell c_{ij}. Let n_{ij} be the number of observations falling in c_{ij}. Under the null hypothesis, $f_{ijk} = c + m_1(y_{t-1}) + m_2(y_{t-2})$. Note that X_{ijk} can be decomposed as

$$X_{ijk} = \{f_{ijk} - f_{ij.}\} + \{f_{ij.} + \nu_{ijk}\} = \{f_{ijk} - f_{ij.}\} + x_{ijk} \tag{13}$$

where $f_{ij.} = E\left[I_{c_{ij}}\{c + m_1(Y_{t-1}) + m_2(Y_{t-2})\}\right]$ is the true cell mean, and $x_{ijk} \equiv f_{ij.} + \nu_{ijk}$. Following the decomposition (13), we obtain the following expressions for various weighted means:

$$\tilde{X}_{ij.} = \sum_{k=1}^{n_{ij}} W_{ijk} X_{ijk} = \sum_{k=1}^{n_{ij}} W_{ijk} x_{ijk} + \sum_{k=1}^{n_{ij}} W_{ijk} (f_{ijk} - f_{ij.}) = \tilde{x}_{ij.} + \tilde{f}_{ij.}$$

$$\tilde{X}_{i..} = \frac{1}{m} \sum_{j=1}^{m} \tilde{X}_{ij.} = \frac{1}{m} \sum_{j=1}^{m} \tilde{x}_{ij.} + \frac{1}{m} \sum_{j=1}^{m} \tilde{f}_{ij.} = \tilde{x}_{i..} + \tilde{f}_{i..}$$

$$\tilde{X}_{.j.} = \frac{1}{m} \sum_{i=1}^{m} \tilde{X}_{ij.} = \frac{1}{m} \sum_{i=1}^{m} \tilde{x}_{ij.} + \frac{1}{m} \sum_{i=1}^{m} \tilde{f}_{ij.} = \tilde{x}_{.j.} + \tilde{f}_{.j.}$$

$$\tilde{X}_{...} = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \tilde{X}_{ij.} = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \tilde{x}_{ij.} + \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \tilde{f}_{ij.} = \tilde{x}_{...} + \tilde{f}_{...}$$

The test statistic F_m, first introduced in (5), can then be split into three parts:

$$\begin{aligned} F_m &= \sum_{i,j=1}^{m} \left(\frac{\tilde{X}_{ij.} - \tilde{X}_{i..} - \tilde{X}_{.j.} + \tilde{X}_{...}}{S_{max}/\sqrt{n_{ij}}} \right)^2 \\ &= \sum_{i,j=1}^{m} \left(\frac{\tilde{x}_{ij.} - \tilde{x}_{i..} - \tilde{x}_{.j.} + \tilde{x}_{...}}{S_{max}/\sqrt{n_{ij}}} \right)^2 && (14) \\ &+ \sum_{i,j=1}^{m} \left(\frac{\tilde{f}_{ij.} - \tilde{f}_{i..} - \tilde{f}_{.j.} + \tilde{f}_{...}}{S_{max}/\sqrt{n_{ij}}} \right)^2 && (15) \\ &+ \sum_{i,j=1}^{m} \left(\frac{\tilde{x}_{ij.} - \tilde{x}_{i..} - \tilde{x}_{.j.} + \tilde{x}_{...}}{S_{max}/\sqrt{n_{ij}}} \right) \left(\frac{\tilde{f}_{ij.} - \tilde{f}_{i..} - \tilde{f}_{.j.} + \tilde{f}_{...}}{S_{max}/\sqrt{n_{ij}}} \right) && (16) \end{aligned}$$

Recall the weighted average $\tilde{X}_{ij.} = \sum_{k=1}^{n_{ij}} W_{ijk} X_{ijk}$. Conditional on knowing sample variances S_{ij}^2, one obtains that, approximately up to the higher order term,

$$\tilde{X}_{ij.} \sim N \left(\mu + \alpha_i + \beta_j + \alpha\beta_{ij}, \sigma_{ij}^2 \sum_{k=1}^{n_{ij}} W_{ijk}^2 \right)$$

due to the central limit theorem (CLT) for strictly stationary strongly mixing processes. This CLT is given in, among other sources, Fan and Yao (2003), p. 75; it can be used here due to the strong mixing condition in the statement of the Theorem. That statement can also be reformulated to say that

$$t_{ij} = \frac{\tilde{X}_{ij.} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij})}{\sqrt{S_{ij}^2 \sum_{k=1}^{n_{ij}} W_{ijk}^2}}$$

is conditionally approximately normally distributed with mean zero and variance $\frac{\sigma_{ij}^2}{S_{ij}^2}$. Now we are interested in the *unconditional* distribution of t_{ij}. It can be shown that the joint distribution of t_{ij} for all $i, j = 1, \ldots, m$ is, up to the higher order term, a product of m^2 density functions of random variables where each has $t_{n_{ij}-2}$ density function; therefore, t_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, m$ are approximately independently distributed as $t_{n_{ij}-2}$. The proof of that fact is almost a verbatim repetition of the analogous proof by Bishop and Dudewicz (1978) and is omitted here for brevity. Note also that each t_{ij} can be represented as

$$t_{ij} = \frac{X_{ijk} - (\mu + \alpha_i + \beta_j + \alpha\beta_{ij})}{S_{\max}/\sqrt{n_{ij}}}$$

since $S_{ij}^2 \sum_{k=1}^{n_{ij}} W_{ijk}^2 = \frac{S_{\max}^2}{n_{ij}}$. Let us denote weighted averages of t_{ij} $t_{i.} = \frac{\sqrt{n_{ij}}}{m} \sum_{j=1}^{m} \frac{t_{ij}}{\sqrt{n_{ij}}}$, $t_{.j} = \frac{\sqrt{n_{ij}}}{m} \sum_{i=1}^{m} \frac{t_{ij}}{\sqrt{n_{ij}}}$, and $t_{..} = \frac{\sqrt{n_{ij}}}{m^2} \sum_{i,j=1}^{m} \frac{t_{ij}}{\sqrt{n_{ij}}}$. Consequently, the first term in the expansion of the test statistic F_m (14) can be represented as

$$\sum_{i,j=1}^{m} \left(t_{ij} - t_{i.} - t_{.j} + t_{..} + \frac{\alpha\beta_{ij}}{S_{\max}} \right)^2 .$$

Under the null hypothesis H_0, it becomes

$$T_m = \sum_{i,j=1}^{m} (t_{ij} - t_{i.} - t_{.j} + t_{..})^2$$

It can be shown that the difference between the distribution function of T_{m_N} and the $\chi^2_{(m-1)^2}$ distribution function goes to zero as the cell size $n_{ij} \to \infty$; for details, see Bishop and Dudewicz (1978). In practice this means, of course, that T_m can be approximated (when sample sizes n_{ij} are large) by $\chi^2_{(m-1)^2}$.

In the remainder of our argument, we intend to show the negligibility of both (15) and (16) as $n_{ij} \to \infty$ and $m \to \infty$. Let I_c be the indicator function of an arbitrary set c and let $L_i = sup_x\{|m_i'(x)|\}$. Denote $p_0 = \sup_S p(x)$ where S is the (compact) support of the stationary density $p(\cdot)$. Since y_t, y_{t-1} and y_{t-2} all have the same marginal stationary distribution $p(\cdot)$,

$$\begin{aligned} v_{ij}(f) &= var\left[I_{c_{ij}}\{c + m_1(Y_{t-1}) + m_2(Y_{t-2})\}\right] \\ &\leq max_{i,j}\left[var\{m_1(Y_t)I_{c_i}\} + var\{m_2(Y_t)I_{c_j}\}\right] \leq 2p_0(L_1^2 + L_2^2)/m^3 \end{aligned}$$

where variance is taken with respect to the stationary density $p(\cdot)$. Hence, under H_0, we have

$$\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}(f_{ijk}-f_{ij.})^2 = O_p\{v_{ij}(f)\} = O_p(1/m^3) \tag{17}$$

By using Cauchy-Schwarz inequality, it is easy to show that

$$\begin{aligned}
\tilde{f}_{ij.}^2 &= \left(\sum_{k=1}^{n_{ij}} W_{ijk}(f_{ijk}-f_{ij.})\right)^2 \\
&\le \left(\sum_{k=1}^{n_{ij}} W_{ijk}^2\right)\left(\sum_{k=1}^{n_{ij}}(f_{ijk}-f_{ij.})^2\right) \\
&= O_p\left(\frac{1}{n_{ij}}\right)\left(\sum_{k=1}^{n_{ij}}(f_{ijk}-f_{ij.})^2\right) \\
&= O_p\left(\frac{1}{m^3}\right);
\end{aligned}$$

consequently,

$$\sum_{i,j=1}^{m}\left(\frac{\tilde{f}_{ij.}}{S_{max}/\sqrt{n_{ij}}}\right)^2 = O_p\left(\frac{N}{m^3}\right). \tag{18}$$

Similar algebra shows that the other three squared terms as well as the cross-product terms in (15) are all $O_p\left(\frac{N}{m^3}\right)$. Thus, (15) is negligible as $N \to \infty$ and there exist $C > 0$ and $\rho > 0$ such that $m > CN^{1/3+\rho}$.

As a final step, consider an arbitrary cross product from (16), say, the first one. Let us denote $\tilde{\nu}_{ij.} = \sum_{k=1}^{n_{ij}} W_{ijk}\nu_{ijk}$. Clearly, $\tilde{x}_{ij.} = f_{ij.} + \tilde{\nu}_{ij.}$. Then, the first crossproduct becomes

$$\sum_{i,j=1}^{m}\frac{\tilde{x}_{ij.}}{S_{max}/\sqrt{n_{ij}}}\frac{\tilde{f}_{ij.}}{S_{max}/\sqrt{n_{ij}}} = \sum_{i,j=1}^{m}\frac{f_{ij.}\tilde{f}_{ij.}}{S_{max}^2/n_{ij}} + \frac{\tilde{\nu}_{ij.}\tilde{f}_{ij.}}{S_{max}^2/n_{ij}}$$

It is easy to show that this term is of the order $O_p\left(\left(\frac{N}{m^{3/2}}\right)\right)$ and the same is true for the rest of the crossproduct terms. As a consequence, it is possible to say that, based on (14)-(16), the test statistic $F_m \equiv F_{m_N}$ has the limiting $\chi^2_{(m-1)^2}$ distribution; in other words, the absolute difference between the distribution function of F_{m_N} and the $\chi^2_{(m-1)^2}$ distribution function goes to zero as long as $N \to \infty$ and $m \to \infty$; moreover, m has to increase at the rate less then N but $m > CN^{1/3+\rho}$ for some $C > 0$ and $\rho > 0$.

References

[1] Besag, J. (1974) "Spatial interaction and the statistical analysis of lattice system (with discussion)," J. R. Statist.Soc. B 36, 192-236

[2] Bishop, T.A. and Dudewicz E.J. (1978) "Exact analysis of variance with unequal variances: test procedures and tables," Technometrics 20, 4, 419-430.

[3] Chan, K.S., Kristoffersen, A.B., and Stenseth, N. Chr. (2003) " Bürmann expansion and test for additivity," Biometrika, 90, 1, 209-222

[4] Chen, S. and Chen, H. (1998) "Single-stage analysis of variance under heteroscedasticity," Communications in Statistics - Simulation and Computation 27, 3, 641-666.

[5] Chen, R., Liu J. and Tsay, R. (1995) "Additive tests for nonlinear autoregression," Biometrika 82, 2, 369-383.

[6] Chen, R. and Tsay, R.S. (1993) " Nonlinear additive ARX models," J. Am. Statist. Assoc. 88, 955-67

[7] Hwang, S.Y. and Woo, M.-J. (2001) "Threshold ARCH(1) processes: asymptotic inference", Statistics and Probability Letters, 53, 11-20

[8] Jones, D.A. (1978) "Nonlinear autoregressive process," Proc. R. Soc. Lond. A 360, 71-95

[9] Lanne, M. and Saikkonen, P. (2005) "Nonlinear GARCH models for highly persistent volatility", Econometrics, 8, 251-276

[10] Li, W.C. and Li, W.K. (1996) " On a double threshold autoregressive heteroscedastic time series model," J. Appl. Econometrics 11, 253-274

[11] Liu, J., Li, W.K. and Li, C.W. (1997) " On a threshold autoregression with conditional heteroscedastic variance," J. Statist. Plann. Inference 62, 279-300

[12] Ling, S. (1999) "On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model", Journal of Applied Probability, 36, 688-705

[13] Lingjaerde et al. (2001), "Exploring non- linearities in the stage-specific density - dependent structure of experimental blowfly population using non-parametric additive modelling", Ecology 82, 2645-58

[14] Lu, Z. (1998) "On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic error term," Statist. Sinica 8, 1205-1217

[15] Lu, Z. and Jiang, Z. (2001) "L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term", Statistics and Probability Letters, 51, 121-130

[16] Robinson, P.M. (1983) "Non-parametric estimation for time series models," J. Time Ser. Anal. 4, 185-208

[17] Sperlich, S., Yang, L., and Tjöstheim, D. (2002): "Nonparametric Estimation and Testing of Interaction in Additive Models ," Econometric Theory 18, 2, 197-251

[18] Tong, H. (1990) Non-linear Time Series Analysis: A Dynamical System Approach", London:Oxford University Press