Joint Significance Testing of the Functional Components in the Nonlinear ARCH Model

by

M. Levine and T.J. Li
Purdue University

Technical Report #08-02

Department of Statistics
Purdue University

August 2008

Joint Significance Testing of the Functional Components in the Nonlinear ARCH Model

Michael Levine [*], Tony J. Li [†‡]

Abstract

Volatility modeling has been an important area of research in nonlinear time series and financial econometrics in the last two decades, generating enormous interest from statisticians, econometricians and financial practitioners alike. The limited reach of the more traditional parametric volatility models became clear relatively early and a number of more data-driven nonparametric models had been proposed. Such models are properly classified as nonlinear time series models and, as usual, the question of model selection becomes very important. This usually involves testing joint statistical significance of a subset of the functional components of such a model.

Levine and Li (2007) suggested a flexible nonlinear ARCH model that has linear and interactive components in both conditional mean and variance (volatility) functions and provided a convenient estimation scheme for the functional components of this mode. This model was introduced to provide a more flexible and data-driven alternative to the so-called generalized nonlinear ARCH (GANARCH) model of Kim and Linton (2004). However, the question of whether the inclusion of interactive components is always beneficial had been left out. Here, the authors suggest a large-sample testing scheme that tests the joint significance of all of the interactive components based on the data. Based on the test results, the interactive components are either included in the model or are left out, thus using the more traditional GAM (generalized additive model) to explain the data behavior. The performance of the test in finite samples is studied using simulation. It is shown that the test has high power for samples of the size commonly encountered in empirical finance applications. The testing procedure depends on several

[*]Corresponding Author. Address: 250 N. University Street, Department of Statistics, College of Science, Purdue University, West Lafayette, IN 47907. Phone: 765-496-7571. Fax: 765-494-0558. E-mail: mlevins@stat.purdue.edu.

[†]Department of Statistics, Purdue University, West Lafayette, IN 47907.

[‡]Keywords and phrases: volatility modeling, autoregression, interactive components, joint significance testing

"tuning" parameters whose selection is also discussed and data-driven selection procedures for them is suggested. Finally, the procedure is also applied to the volatility model of the currency exchange rate data using the real dataset.

1 Introduction

Volatility modeling has been a very active research area in financial econometrics and empirical finance in the last two decades; it arose from the need to model and predict the volatility (conditional variance) of financial data, such as stock returns, currency exchange rates and others. As a first step, parametric volatility models were considered, beginning with the famous ARCH/GARCH framework (see Engle (1982) and Bollerslev (1986)). The common problem of all parametric models is misspecification, since in practice the choice of this or that functional form is, to a great extent, subjective.

A possible alternative consists of using nonparametric volatility models that assume very little known information about the volatility structure. Effectively, these models allow the data " to speak for themselves." One of the most general nonparametric models that can be considered is the nonparametric ARCH model

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

where $d > 0$ and ε_t are independent identically distributed random variables with mean zero and variance 1. Without imposing some additional functional structure on this model the famous problem of the"curse of dimensionality" limits its usefulness rather severely. Also, it is quite difficult to visualize and interpret any multidimensional surface with the number of dimensions more than 2.

The classical approach to handling the curse of dimensionality problem is to consider the so-called GAMs (generalized additive models) which were the subject of the defining monograph by Hastie and Tibshirani (1990). The classical form of GAM assumes that

$$m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha})$$

while the conditional variance function is constant. It is also quite natural to impose a similar

structure on the conditional variance function by assuming that

$$v(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha}).$$

In this paper we assume that the number of lags involved is the same for conditional mean and variance functions for the sake of simplicity. The component estimation method proposed will, however, also work when the number of lags is different as well. The attractiveness of generalized additive models stems in part from the fact that (see Stone (1985, 1986)) the optimal rate of convergence of the functional component estimators is the same as in one-dimensional nonparametric regression. Nevertheless, GAM structure is often not very flexible. A number of ways to generalize it has been considered in the literature; some of the examples in volatility modeling context are the multiplicative model of Yang, Härdle and Nielsen (1998) and the generalized additive nonlinear ARCH (GANARCH) model of Kim and Linton (2004). A conceptually very simple generalization was proposed in Levine and Li (2007) where, in addition to "main effects" m_α and v_α, $1 \leq \alpha \leq d$, the "interactive components" $m_{\alpha\beta}$ and $v_{\alpha\beta}$ are also included. Thus, the model becomes

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

$$\text{where } m(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_m + \sum_{\alpha=1}^{d} m_\alpha(y_{t-\alpha}) + \sum_{1\leq\alpha<\beta\leq d} m_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta})$$

$$v(y_{t-1}, y_{t-2}, \cdots, y_{t-d}) = C_v + \sum_{\alpha=1}^{d} v_\alpha(y_{t-\alpha}) + \sum_{1\leq\alpha<\beta\leq d} v_{\alpha\beta}(y_{t-\alpha}, y_{t-\beta}).$$

Thus, both conditional mean and conditional variance functions are allowed to have both additive and two-way interactive components. The model is very much data-driven since the number of interactive components to be included in the model can be determined from the data; if, for example, the asymptotic distribution of the interactive component estimators is known, it is easy to construct the large-sample test that would check statistical significance of these components and allow us to decide whether a particular component should or should not be included in the model (2).

Because of high multicollinearity between the lagged values of y_t, it is of little interest to test the significance of an individual interactive component. Instead, multiple hypothesis testing should

be of main interest here. More precisely, the goal should be to test

$$H_0^m : m_{\alpha\beta} = 0, \text{ for all } 1 \le \alpha < \beta \le d \tag{3}$$
$$H_1^m : m_{\alpha_0\beta_0} \neq 0, \text{ for at least one } 1 \le \alpha_0 < \beta_0 \le d$$

and the analogous multiple hypotheses for components of the conditional variance function

$$H_0^v : v_{\alpha\beta} = 0, \text{ for all } 1 \le \alpha < \beta \le d \tag{4}$$
$$H_1^v : v_{\alpha_0\beta_0} \neq 0, \text{ for at least one } 1 \le \alpha_0 < \beta_0 \le d$$

in the model (2).

Testing of interactive components in the nonlinear regression models has been a research topic for some time, particularly in econometrics; as an example, one can mention Sperlich,Tjöstheim and Yang (2002) and references therein. To the best of our knowledge, there has been no work regarding the joint testing of the nonlinear autoregression components. The current research aims to begin filling this void.

2 A Conditional Mean Test

This section is dedicated to developing the linearity test for the conditional mean function in the model (2). Such a test is aimed at testing the joint significance of the non-additive components of the function $m(y_{t-1}, \ldots, y_{t-d})$. Note that the model (2) is the nonlinear autoregression. If the interactive components are found to be jointly insignificant, the data can be successfully modeled using the classical GAM model using well-known methods such as backfitting or marginal integration.

Significance testing for the interactive components of the function $m(y_{t-1}, \ldots, y_{t-d})$ is an important problem in the context of model selection. Individual testing of pairwise interaction components had been a subject of interest earlier, mostly in the setting that assumes that the explanatory variable vector $X = (X_1, \ldots, X_d)'$ represents a sequence of independent and identically distributed vectors of explanatory variables. Such a setting will be called an iid setting for short. The iid setting was considered earlier in Sperlich, Tjöstheim and Yang (2002); a similar problem in the nonparametric regression on the lattice had been considered by Derbort, Dette and Munk (2002). Hjellvik, Yao and Tjöstheim (1998) considered the linearity testing in a very general context that

includes the nonlinear autoregressive processes of the type (2); however, their approach allows only for the individual pairwise interaction testing only. They note that "...Following a suggestion by the referee, in principle, the joint limit distribution of the statistics...can be derived" but the idea is not pursued any further. The individual testing for each pairwise interaction may be of rather limited importance in the time series context since the explanatory variables $y_{t-1}, \ldots, y_{t-d}$ are obviously strongly correlated. It is entirely possible that individual pairwise components $m_{ij}(y_{t-i}, y_{t-j})$ are not statistically significant but their joint effect is significant.

Our goal is to construct an F-type test that would perform the multiple hypotheses testing for all of the components $m_{ij}(\cdot, \cdot)$. For simplicity, the model of order two

$$y_t = m(y_{t-1}, y_{t-2}) + v^{1/2}(y_{t-1}, y_{t-2})\varepsilon_t \tag{5}$$

is used to illustrate the idea. It is assumed that 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. The goal is to test whether the conditional mean function $m(\cdot, \cdot)$ can be represented as $m_1(y_1) + m_2(y_2)$ or if it also has an interactive component. Thus, the null hypothesis is

$$H_0 : y_t = c_m + m_1(y_{t-1}) + m_2(y_{t-2}) + v^{1/2}(y_{t-1}, y_{t-2})\varepsilon_t$$

while the alternative is

$$H_1 : y_t = c_m + m_1(y_{t-1}) + m_2(y_{t-2}) + m_{12}(y_{t-1}, y_{t-2}) + v^{1/2}(y_{t-1}, y_{t-2})\varepsilon_t$$

Chen, Liu and Tsay (1995) discussed a similar problem earlier; they assumed, however, that the conditional variance function $v \equiv 1$. This assumption does not always hold in practically important nonparametric models.

The procedure suggested here can be viewed as the two-way unbalanced analysis of variance (ANOVA) with unequal variances. In the traditional ANOVA setting, coping with unequal variances is rather difficult. It has been established in the past (see, e.g. Bishop (1976)) 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 such a setting was proposed by Bishop and Dudewicz (1978). That procedure requires additional sampling at the second stage. Chen and Chen (1998) made an improvement by proposing a single stage sampling procedure that tests the

null hypotheses in ANOVA models under heteroscedasticity. This procedure is computationally much more effective than the earlier procedure by Bishop and Dudewicz (1978) and it does not require additional sampling of the data. This last requirement is particularly important in the time series setting where, unlike when dealing with experimental data, it is impossible to obtain replicates.

Our procedure is based on the one by Chen and Chen (1998). It can be described as follows.

- Choose a shrinking factor $0 < \delta < 1$ and a positive integer m. Partition the shrunken data range $\delta(y_{max} - y_{min})$ into m equal intervals (a_{i-1}, a_i), $i = 1, \ldots, m$. 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.

- The overall number of observations is denoted N. For $t = 3, \cdots, N$, the observation y_t is classified in the (i, j)-th cell if $y_{t-1} \in (a_{i-1}, a_i)$ and $y_{t-2} \in (a_{j-1}, a_j)$. The ANOVA-style notation $X_{ijk} = y_t$ can be used 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 shrunken range, y_t is dropped from further consideration. As a result, the number of observations used in the eventual ANOVA scheme is $n < N$. The number of observations in the (i, j) th cell is denoted n_{ij}. Clearly, $\sum_{i,j=1}^{m} n_{ij} = n$.

- The model (5) can be viewed as

$$X_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \tag{6}$$

  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 σ_{ij}^2, $0 < \sigma_{ij}^2 < \infty$ and unknown. Since the number of possible "treatments" in this scheme is not predetermined, the model can be viewed as the two-way random effect model. 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 (6)

The above procedure is based on the idea that, when functions $m_1(\cdot)$, $m_2(\cdot)$ and $m_{12}(\cdot, \cdot)$ are sufficiently smooth, the observations in the same cell have roughly the same mean and conditional

variance values. This argument is used asymptotically when the length of the interval goes to zero as $m \to \infty$. Additionally, under the strong mixing condition, the observations in a cell behave like independent observations. Therefore, testing the null hypothesis H_0 is, indeed, analogous to the two-way unbalanced design ANOVA with unequal variances. Such a procedure is performed in the following way.

- Within each cell, the first $n_{ij} - 1$ observations are used to compute the regular sample mean and variance:

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

- The weights of the observations in cell (i, j) are computed as

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

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

For each cell (i, j), the weighted sample mean is defined as

$$\tilde{X}_{ij\cdot} = \sum_{k=1}^{n_{ij}} W_{ijk} X_{ijk} \tag{7}$$

where

$$W_{ijk} = \begin{cases} U_{ij} & \text{for } 1 \le k \le n_{ij} - 1 \\ V_{ij} & \text{for } k = n_{ij} \end{cases}$$

Note that thus defined weights are indeed properly normalized; it is easy to verify that $\sum_{k=1}^{n_{ij}} W_{ijk} = (n_{ij} - 1)U_{ij} + V_{ij} = 1$ and $\sum_{k=1}^{n_{ij}} W_{ijk}^2 = (n_{ij} - 1)U_{ij}^2 + V_{ij}^2 = \frac{S_{\max}^2}{n_{ij} S_{ij}^2}$. Next, partial

means

$$\tilde{X}_{i\cdot\cdot} = \frac{1}{m}\sum_{j=1}^{m} \tilde{X}_{ij\cdot}$$
$$\tilde{X}_{\cdot j\cdot} = \frac{1}{m}\sum_{i=1}^{m} \tilde{X}_{ij\cdot}$$
$$\tilde{X}_{\cdots} = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} \tilde{X}_{ij\cdot}$$

are computed, using the weighted sample mean (7).

- The F-type test statistic is defined as

$$F_m = \sum_{i=1}^{m}\sum_{j=1}^{m}\left(\frac{\tilde{X}_{ij\cdot} - \tilde{X}_{i\cdot\cdot} - \tilde{X}_{\cdot j\cdot} + \tilde{X}_{\cdots}}{S^2_{max}/\sqrt{n_{ij}}}\right)^2 \tag{8}$$

It can be shown that for large sample sizes, as $N \to \infty$, $n_{ij} \to \infty$ for all i, j and the number of intervals $m \to \infty$ as well, the distribution of F_m under the null hypothesis H_0 behaves approximately as $\chi^2_{(m-1)^2}$. The more precise statement is as follows.

Proposition 2.1 *Consider the additive-interactive model* (5) *where the process* y_t *has a stationary density* $p(\cdot)$ *and* $m(y_{t-1}, y_{t-2}) = m_1(y_{t-1}) + m_2(y_{t-2})$. *For simplicity, we assume that the process* y_t *is defined on the compact support. Assume that* ε_t *are independent normal random variables with mean zero and variance* 1. *Suppose* $m_1(\cdot)$ *and* $m_2(\cdot)$ *are continuous differentiable functions with bounded first derivatives while the conditional variance function* $v(\cdot,\cdot)$ *is bounded. Suppose that the number of intervals used in the test is* $m = O(N^{1/3+\delta})$ *for some small* $\delta > 0$. *Then, as* $N \to \infty$, *the test statistic* F_m *follows asymptotically the* $\chi^2_{(m-1)^2}$ *distribution.*

An immediate consequence is that, at a confidence level of $100(1-\alpha)\%$, we can approximately reject H_0 if F_m exceeds the upper α quantile of $\chi^2_{(m-1)^2}$.

Proof See Appendix.

Remarks.

1. The Chi-Square distribution of the test statistic under the null hypothesis H_0 is true in the asymptotic sense, i.e. as the sample size N and all of n_{ij}'s in ANOVA approach infinity. It may not be a very good approximation when the initial sample size N is small. The simulation

study in this paper shows that for a sample size $N = 5000$ the proposed test procedure works quite well.

2. The shrinking factor δ plays an important and interesting role in the testing approach. It seems rather intuitive, at first sight, that with larger δ comes larger effective sample size and the higher power of the test. However, it is not as straightforward as it may seem. The problem is that numbers of observations in different cells do not always grow in sync but rather proportionally to the marginal density of the stationary process y_t, earlier denoted $p(y_t)$. Because of that, in particular, the number of observations in the boundary cells of the $m \times m$ layout that are near extremes $y_{\max}$ and $y_{\min}$ can decrease with the increase in δ. This, in turn, means that the power of the test may not grow monotonically as δ increases. The simulation study in the next chapter illustrates this point in greater detail.

3. The issue of choosing the number of partition intervals m is quite complicated. It is clear from the proof of the Proposition that choosing larger m results in the more precise χ^2 approximation of the true distribution of the test statistic (8) under the null hypothesis. However, large m has its own downside - the power of the test maybe much less than desired. Thus, the choice of m should reflect this trade-off. In the next section, a data-driven parameter selection procedure for m and δ will be suggested.

4. It is easy to show that the statement of the Proposition remains true if the process ε_t is not iid but rather constitutes a martingale difference sequence. Note that the model (5) can be rewritten as

$$(y_t - m(y_{t-1}, y_{t-2}))^2 = v(y_{t-1}, y_{t-2}) + v(y_{t-1}, y_{t-2})(\varepsilon_t^2 - 1) \tag{9}$$

and it can be verified in a straightforward way that $v(y_{t-1}, y_{t-2})(\varepsilon_t^2 - 1)$ is a martingale difference sequence. This suggests a simple procedure for testing the joint significance of the interactive components of the conditional variance function $v(\cdot)$. As a first step, a consistent estimator of the mean function $\hat{m}(\cdot)$ should be selected. That can be, for instance, the local instrumental variable estimator described earlier. Then, the ANOVA type algorithm described earlier should be applied to squared residuals $\left\{(y_t - \hat{m}(y_{t-1}, y_{t-2}))^2\right\}_{t=3}^{N}$.

3 A Simulation Study

This section is dedicated to how the proposed test works with simulated data. The data that is considered is generated by the process

$$y_t = 0.5sin(0.5y_{t-1}) - 0.5sin(0.5y_{t-2}) - 0.5atan(0.5y_{t-1})atan(0.5y_{t-2}) \quad (10)$$
$$+ \sqrt{1.5 + 0.5sin(0.5y_{t-1}) - 0.5sin(0.5y_{t-2}) - 0.5atan(0.5y_{t-1})atan(0.5y_{t-2})}\varepsilon_t$$

where $\varepsilon_t \sim N(0,1)$. The choice of conditional mean and variance functions is meant to ensure the strict stationarity of the process; the exact conditions are stated in Lu and Jiang (2001) and referred to in Levine and Li (2007). The sample sizes N=500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000 and 10000 are used. As an illustration, Table (1) shows the number of observations in the shrunken range $N(\delta)$ and the number of observations that end up being used in ANOVA approximation n when $\delta = 0.5$ for each of the choices of N. Both $N(\delta)$ and n are showed as absolute numbers and percentages of the initial sample size N. The number of intervals chosen is $m = 3$. Clearly, $n \leq N(\delta) \leq N$ for any $0 < \delta < 1$. Note that the eventual loss of data due to the selection procedure employed is not very large. In the worst case scenario, for a relatively low sample size of $N = 500$, 75% of the data are eventually retained for the ANOVA analysis. For all of the sample sizes that are larger than 500, the loss does not fall below 80%. Typical distributions of observations among the ANOVA cells for $m = 3$, $\delta = 0.5$ and initial sample sizes $N = 500$, 5000 and 10000, respectively, are shown in Tables (2), (3) and (4). Note that the choice of $m = 3$ implies the 3×3 design. It is instructive to estimate the type I error. It is done here for sample sizes ranging from $N = 500$ to $N = 1000$, assuming that $m = 3$ and $\delta = 0.5$. In order to do this, a sequence of observations is generated by the process (10) under the null hypothesis H_0. In other words, the observations are now generated by the process

$$y_t = 0.5sin(0.5y_{t-1}) - 0.5sin(0.5y_{t-2})$$
$$+ \sqrt{1.5 + 0.5sin(0.5y_{t-1}) - 0.5sin(0.5y_{t-2}) - 0.5atan(0.5y_{t-1})atan(0.5y_{t-2})}\varepsilon_t$$

with $\varepsilon_t \sim N(0,1)$.

Table (5) shows the percentage of rejections of H_0 with 500 replications for each sample size. Note that the type I error assumes the specified level once sample size reaches the value of 5000.

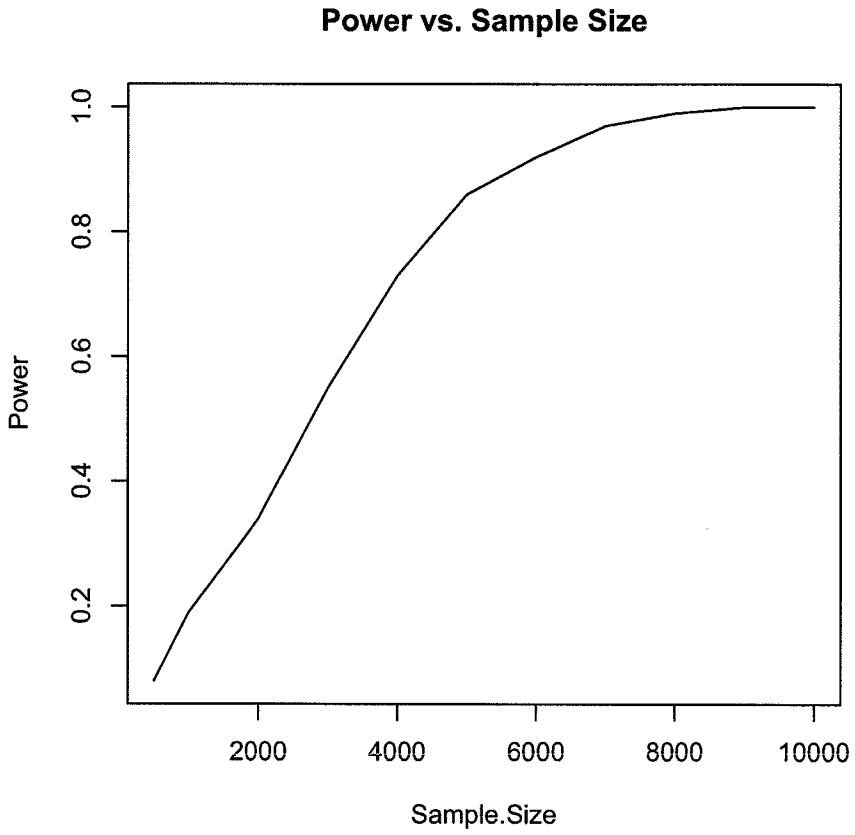

Figure 1: Power against Sample Size

N	N_δ	$\frac{N_\delta}{N}$	n	$\frac{n}{N}$
500	436	(87%)	376	(75%)
1000	939	(94%)	879	(88%)
2000	1856	(93%)	1722	(86%)
3000	2767	(92%)	2555	(85%)
4000	3707	(93%)	3433	(86%)
5000	4686	(94%)	4390	(88%)
6000	5611	(94%)	5240	(87%)
7000	6460	(92%)	5968	(85%)
8000	7506	(94%)	7046	(88%)
9000	8327	(93%)	7702	(86%)
10000	9366	(94%)	9077	(91%)

Table 1: Number of Observations

To see whether this value is a possible "threshold" , it is useful to study also the power of the test. The power of the proposed test is evaluated using observations generated by the process (10) under different values of δ. As a reminder, the null hypothesis is the one of no interactive components in the conditional mean:

$$H_0 : m(y_{t-1}, y_{t-2}) = 0.5sin(0.5y_{t-1}) - 0.5sin(0.5y_{t-2})$$

The important issue is how the choice of δ influences the power of the test. For a fixed value of m and changing values of δ, the Table (6) shows the evolution of the power of the test. The sample sizes considered are from $N = 1000$ to $N = 7000$ with the step of 1000. The shrunken range is split into $m = 3$ intervals and the number of replications is 1000. For each of the sample sizes considered, all of $\delta = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ options are examined.

Note that the power of the test seems to grow fairly quickly with the sample size as long as δ is reasonably large. For sample sizes greater than 5000 and $\delta \geq 0.5$, the power exceeds 0.9 which is often good enough for practical purposes. The growth of the power of the test with the sample size when $\delta = 0.5$ and $m = 3$ is illustrated in the Figure (1). This sample size magnitude is very reasonable in many applications, for example, in those stemming from the empirical finance

Table 2: Distribution of Observations among the ANOVA Cells of a typical realization for $m = 3$. The initial sample size is N=500

	1	2	3
1	49	51	24
2	55	81	36
3	21	42	17

Table 3: Distribution of Observations among the ANOVA Cells of a typical realization for $m = 3$. The initial sample size is N=5000

	1	2	3
1	405	581	239
2	661	1082	435
3	195	526	266

Table 4: Distribution of Observations among the ANOVA Cells of a typical realization for $m = 3$. The initial sample size is $N = 10000$

	1	2	3
1	722	1242	365
2	1379	2668	911
3	290	1029	471

Table 5: Type I error in 500 Realizations Under H_0

Sample Size	500	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
Type I Error	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05

Table 6: The relationship between the power and the shrinking factor δ

δ	$N = 1000$	$N = 2000$	$N = 3000$	$N = 4000$	$N = 5000$	$N = 6000$	$N = 7000$
0.1	0.2136	0.0619	0.0341	0.0410	0.0400	0.0600	0.0900
0.2	0.0644	0.0370	0.0440	0.0510	0.0560	0.0800	0.1000
0.3	0.0592	0.0701	0.1340	0.1620	0.2400	0.2900	0.3900
0.4	0.1116	0.20704	0.3430	0.4640	0.6380	0.7800	0.8100
0.5	0.1635	0.3574	0.5916	0.7518	0.8669	0.9500	0.9800
0.6	0.2584	0.5050	0.6901	0.8516	0.9288	0.9900	1.0000
0.7	0.3447	0.5563	0.7185	0.8541	0.9276	0.9495	0.9800
0.8	0.4189	0.5941	0.7427	0.8496	0.9109	0.9400	0.9694
0.9	0.4917	0.6222	0.7479	0.8402	0.8760	0.9175	0.9588

problems. Thus, the test seems to exhibit good performance in practice.

It is easy to see that the power does not increase monotonically with δ. More specifically, for small sample sizes there seems to be a jump in the power of the test when the δ is small; for example, if $N = 1000$, the power is only 0.0644 for $\delta = 0.2$ but reaches 0.2136 for $\delta = 0.1$. The combination of the small δ and small initial sample size means that, for any stationary distribution that is not very skewed, a large part of all observations is being preserved after shrinking operation and becoming "trapped" in a short interval of length $\delta(y_{\max} - y_{\min})$. Because the interval is short, the stationary density of the process y_t can be expected to be relatively constant; therefore, numbers of observations in $m = 3$ equal intervals can be expected to be close to each other, thus making the power of the test rise.

This effect does not work for larger sample sizes where the loss of observations due to lower δ seems to predominate. However, for larger values of N, the growth in δ does not always mean automatic increase of the test power. The problem is that the growth of the number of observations available for ANOVA analysis does not occur monotonically across all of the cells. In the Table (6), this tapering-off of the power of the test is clearly seen when $\delta \geq 0.6$. It is reasonable to expect that this phenomenon should depend on the stationary distribution of the process y_t: the more symmetric this distribution is, the less pronounced it should be. On the contrary, with the

strongly skewed stationary distribution, the number of boundary observations that are removed from consideration when δ is small can be quite large. The natural outcome of this is that cells that cover the center of the range may have relatively few observations in them. The logical outcome, thus, would be the decreasing power of the test which is clearly seen for large δ and sample sizes that are greater than 4000.

As mentioned earlier, m should be chosen small enough to keep the power of the test high and large enough for the χ^2-approximation of the test statistic to be appropriate. It can be said that, as is the case with the choice of δ, there exists a trade-off between these two goals which should be reflected in the choice of m. The following ad-hoc procedure for selection of m is suggested.

If δ is chosen, for each integer m, there are m^2 cells considered. Let us denote the number of observations falling in the (i,j)th cell n_{ij}. Then, in order to achieve the necessary trade-off between the local power of the test and the necessary degree of χ^2 approximation for the test statistic, it seems sensible to choose the value of m that maximizes the product $m^2 \min_{i,j} n_{ij}$ over a grid in m. The first term in the product $m^2 \min_{i,j} n_{ij}$ product favors larger m while the second favors a smaller one. To make it more precise, the optimal value of m for a fixed value of $\delta = \delta_0$ can be defined as

$$m_{opt} = argmax_m \left\{ m^2 \min(n_{ij} | \delta = \delta_0) \right\}, \quad 2 \leq m \leq N^{1/2} \tag{11}$$

Note that there is no need to search over any m that is larger than $N^{1/2}$ since it automatically leads to at least one cell having no observations at all and thus $\min(n_{ij}|\delta = \delta_0) = 0$, regardless of the choice of δ_0.

The performance of the method is illustrated by using the model (5) to generate data with sample sizes $N = 5000$, $N = 10000$ and $N = 20000$, respectively. For each sample size, 100 realizations are generated and it is assumed that $\delta = 0.5$. The optimal m is computed using the ad-hoc procedure described in (11) as the average of 100 possible choices of m that correspond to 100 simulations conducted for each sample size. The resulting optimal choices of m are 4, 5 and 7 for sample sizes of 5000, 10000 and 20000, respectively. In order to illustrate the choice made, the power and the type I error of the test are calculated for several different values of m for each choice of the sample size and shown in Tables (7)-(9).

Overall, the procedure seems to perform reasonably well; more specifically, it is clear that it works better for larger sample sizes $N = 10000$ and especially $N = 20000$ than for a smaller sample size $N = 5000$. Analogous calculations for other sample sizes that lie between 5000 and 10000 as

Table 7: Simulated power and type I error: N=5000

m	power	type I error
2	0.93	0.05
3	0.86	0.05
4	0.80	0.05
5	0.64	0.05

Table 8: Simulated power and type I error: N=10000

m	power	type I error
2	1	0.08
3	1	0.05
4	1	0.04
5	0.95	0.05
6	0.90	0.05
7	0.66	0.04

Table 9: Simulated power and type I error: N=20000

m	power	type I error
2	1	0.15
3	1	0.12
4	1	0.11
5	1	0.10
6	1	0.08
7	0.99	0.04
8	0.93	0.05
9	0.85	0.05
10	0.79	0.05

well as between 10000 and 20000 were also performed; all of them support this conclusion. They are not given here for reasons of brevity but are available upon request.

4 A Real Data Example

In this section the proposed testing method is applied to the data set of the deutschmark/US dollar exchange rates from January 2, 1980 to May 23 1986. The data set contains the total of 1603 observations. The same data was analyzed in Yang, Härdle and Nielsen (1999) where the nonparametric autoregressive model with multiplicative volatility and additive mean was fit to it. The data are plotted in Figure (2).

As a first step, the generalized additive-interactive volatility model of order 3 is fit to the data set described. The detailed description of the local instrumental variable (LIVE) method used to fit the model as well as the fitting procedure itself can be found in Levine and Li (2007) and, therefore, only the general outline is given here. One-dimensional Gaussian and product Gaussian kernels are used to estimate univariate and multivariate density functions, respectively. Local linear regressions are used to estimate the additive and interactive components of the mean and volatility functions. Optimal bandwidths are selected using the cross-validation method. Levine and Li (2007) showed that the estimated additive components of the mean functions seem to be close to zero which is not the case with the additive volatility components. The latter finding showed the asymmetric influence of the past values of the exchange rate on its current volatility.

The analysis of Levine and Li (2007) also showed that the volatility function also contains what looks like rather substantial interactive components with interesting interpretation. To confirm or reject the presence of interactive components in both mean and variance functionas, the formal testing is performed.

In this case, it is not advisable to select large values of δ since the squared residuals from the mean fit are much smaller in absolute value compared to the original data points y_t. Because of that, choosing larger values of δ results in the loss of a very substantial percentage of "observations" (i.e., squared residuals). To illustrate this point, it is instructive to compute the percentages of squared residuals remaining to be used in the ANOVA scheme for several values of δ and a fixed ms, say $m = 3$. For the overall sample size $N = 1603$, $\delta = 0.5$ results in 641 remaining squared residuals

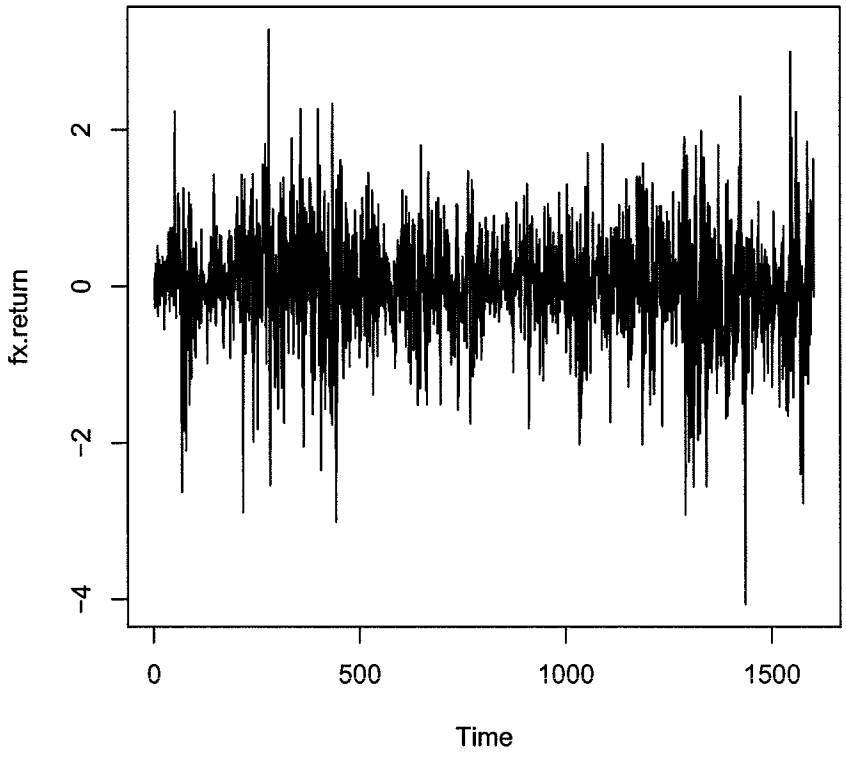

Figure 2: Plot of the daily return of deutschmark/US dollar, from Jan. 2 1980 to May 23 1986

available for ANOVA analysis (40% of the original number), $\delta = 0.6$ results in 585 remaining which makes up 36% of the original number and for $\delta = 0.7$ it is 510 - 32% of the original number. This suggests that choosing values of δ larger than 0.5 may not be advisable. For the purpose of this analysis, the value $\delta = 0.5$ is chosen; this value is used for testing both mean and variance function components. The entire data range is split into $m = 2$ intervals; the choice of small m is due to the relatively small initial sample size $N = 1063$. The shrunken range contains $n = 1556$ observations that constitute 97.1% of the original observations. Two tests are performed using both the original observations from the shrunken range (to test for the presence of interactive mean components) and the squared residuals from the mean function fit (to test for the presence of interactive variance components). The P-values for the tests on mean and volatility functions are 0.322 and 0.021, respectively. This suggests that at a 95% confidence level and even after Bonferroni adjustment, the interactive components in the mean functions are not significant while those in the volatility functions are. Note that this is not the trivial result: the presence of interactive volatility components makes it possible to draw specific conclusions about the influence of past lags of the currency exchange rate on the current volatility which is impossible to do using methods of Yang, Härdle and Nielsen (1999); see Levine and Li (2007) for detailed discussion of significance of interactive components. Also, unlike Yang, Härdle and Nielsen (1999), local linear regression estimators are used instead of local constant estimators, thus providing automatic boundary bias correction.

5 Discussion and Conclusions

The approach to testing the joint significance of the interaction components proposed here is convenient in practice and is conceptually rather simple. It can be viewed in the framework of the random effect model ANOVA with the resulting χ^2 test performing very well for reasonably large samples. It is particularly important that the test can accommodate the joint testing of interactive components in mean under the assumption of heteroskedasticity - this is new and, to the best of our knowledge, has not been done before. As mentioned earlier, the proposed test can also be easily generalized to testing the presence of interactive components in the conditional variance (volatility) function as well.

The power of the test is higher for small values of the number of intervals m used; however, the procedure as discussed here does not take into account the true stationary density of the process y_t. A rather obvious way to improve the power of the test is to partition the shrunken range into unequal size intervals in accordance with the estimated stationary density of the process y_t. Such a procedure would be particularly useful for relatively small sample sizes where, as can be seen from the Table (6), there is some room for improvement.

The entire exposition in this paper was restricted to the case of $d = 2$ for the sake of simplicity. The generalization to the case where there are nonparametric interactive components of order $d > 2$ is conceptually straightforward; it requires the generalization of the testing framework of Chen and Chen (1998) to the case of k-way ANOVA under heteroskedasticity when $k > 2$. Note that this will result in highly computationally intensive procedure.

References

[1] Heteroscedastic ANOVA, MANOVA and multiple comparisons. Unpublished PhD thesis, The Ohio State University

[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] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics, 31, 307-327

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

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

[6] Derbort, S., Dette, H. and Munk, A. (2002). A Test for Additivity in Nonparametric Regression. Annals of the Institute of Statistical Mathematics, 54(1), 60-82

[7] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, Econometrica, Vol. 50, 4, 987-1007

[8] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* Chapman & Hall, London

[9] Hjellvik, V., Yao, Q. and Tjöstheim, D. (1998) Linearity testing using local polynomial approximation. Journal of Statistical Planning and Inference, 68(2),295-321

[10] Kim, W. and Linton, O.B. (2004) The live method for generalized additive volatility models. Econometric theory, 20(6), 1094-1139

[11] Levine, M. and Li, J. (2007). Local instrumental variable (LIVE) method for the generalized additive-interactive nonlinear volatility model. Technical report.

[12] Lu Z., and Jiang, Z. (2001). $L1$ geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. Statistics and Probability Letters, 51, 121-130

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

[14] Stone, C.J. (1985). Additive regression and other nonparametric models. Annals of Statistics, 13, 685-705.

[15] Stone, C.J. (1986) The dimensionality reduction principle for generalized additive models. Annals of Statistics, 14, 592-606.

[16] Yang, L., Härdle, W. and Nielsen, J.P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. Journal of Time Series Analysis 20, 5, 579-604.

6 Appendix

Proof: Denote c_{ij} the 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 (5) can be represented as $X_{ijk} = f_{ijk} + \varepsilon_{ijk}$ 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}; the number of observations falling in c_{ij} is denoted n_{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

$$\begin{aligned} X_{ijk} &= \{f_{ijk} - f_{ij.}\} + \{f_{ij.} + \varepsilon_{ijk}\} \\ &= \{f_{ijk} - f_{ij.}\} + x_{ijk} \end{aligned} \tag{12}$$

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 $x_{ijk} = f_{ij.} + \varepsilon_{ijk}$.

Following the decomposition (12), we obtain easily the following:

$$\begin{aligned} \tilde{X}_{ij.} &= \sum_{k=1}^{n_{ij}} W_{ijk} X_{ijk} = \sum_{k=1}^{n_{ij}} W_{ijk} x_{ijk} + \sum_{k=1}^{n_{ij}} W_{ijk}(f_{ijk} - f_{ij.}) = \tilde{x}_{ij.} + \tilde{f}_{ij.} \\ \tilde{X}_{i..} &= \frac{1}{m}\sum_{j=1}^{m} \tilde{X}_{ij.} = \frac{1}{m}\sum_{j=1}^{m} \tilde{x}_{ij.} + \frac{1}{m}\sum_{j=1}^{m} \tilde{f}_{ij.} = \tilde{x}_{i..} + \tilde{f}_{i..} \\ \tilde{X}_{.j.} &= \frac{1}{m}\sum_{i=1}^{m} \tilde{X}_{ij.} = \frac{1}{m}\sum_{i=1}^{m} \tilde{x}_{ij.} + \frac{1}{m}\sum_{i=1}^{m} \tilde{f}_{ij.} = \tilde{x}_{.j.} + \tilde{f}_{.j.} \\ \tilde{X}_{...} &= \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} \tilde{X}_{ij.} = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} \tilde{x}_{ij.} + \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} \tilde{f}_{ij.} = \tilde{x}_{...} + \tilde{f}_{...} \end{aligned}$$

It is easy to verify that the sample variance of each cell can be expressed as

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

The F-type statistic F_m, first introduced in (8), can then be split into three parts:

$$\begin{aligned} F_m &= \sum_{i,j=1}^{m} \left(\frac{\tilde{X}_{ij.} - \tilde{X}_{i..} - \tilde{X}_{.j.} + \tilde{X}_{...}}{S_{max}/\sqrt{n_{ij}}}\right)^2 \\ &= \sum_{i,j=1}^{m} \left(\frac{\tilde{x}_{ij.} - \tilde{x}_{i..} - \tilde{x}_{.j.} + \tilde{x}_{...}}{S_{max}/\sqrt{n_{ij}}}\right)^2 \qquad (13) \\ &+ \sum_{i,j=1}^{m} \left(\frac{\tilde{f}_{ij.} - \tilde{f}_{i..} - \tilde{f}_{.j.} + \tilde{f}_{...}}{S_{max}/\sqrt{n_{ij}}}\right)^2 \qquad (14) \\ &+ \sum_{i,j=1}^{m} \left(\frac{\tilde{x}_{ij.} - \tilde{x}_{i..} - \tilde{x}_{.j.} + \tilde{x}_{...}}{S_{max}/\sqrt{n_{ij}}}\right)\left(\frac{\tilde{f}_{ij.} - \tilde{f}_{i..} - \tilde{f}_{.j.} + \tilde{f}_{...}}{S_{max}/\sqrt{n_{ij}}}\right) \qquad (15) \end{aligned}$$

Note that (13) is approximately $\chi^2_{(m-1)^2}$ distributed when all of $n_{ij} \to \infty-$ this follows immediately from Chen and Chen (1998). Thus, in order to prove the statement, it is necessary to show that both (14) and (15) are negligible as $n_{ij} \to \infty$ and $m \to \infty$.

Let us denote I_c the indicator function of the 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

$$\frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}}(f_{ijk} - f_{ij.})^2 = O_p\{v_{ij}(f)\} = O_p(1/m^3) \tag{16}$$

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

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

consequently,

$$\sum_{i,j=1}^{m}\left(\frac{\tilde{f}_{ij.}}{S_{max}/\sqrt{n_{ij}}}\right)^2 = O_p\left(\frac{n}{m^3}\right). \tag{17}$$

Similar algebra shows that the other three squared terms as well as the cross-product terms in (14) are all $O_p\left(\frac{n}{m^3}\right)$. Thus, (14) is negligible as $n \to \infty$ and $m = O(n^{1/3} + \delta)$ for any small $\delta > 0$. This suggests that m must be large enough for χ^2 approximation to be precise.

This conclusion is further confirmed by analyzing the cross-product term (15). Note that this term is a sum of approximately independent components; the latter is due to the fact that observations across cells are approximately independent if the strict stationarity of y_t is assumed. Choosing as an example the first cross-product term $\sum_{i,j=1}^{m}\left(\frac{\tilde{x}_{ij.}}{S_{max}/\sqrt{n_{ij}}}\frac{\tilde{f}_{ij.}}{S_{max}/\sqrt{n_{ij}}}\right)$, It is easy to show that its order is $O_p\left(\left(\frac{n}{m^3}\right)^{1/2}\right)$ and the same is true for the rest of the crossproduct terms. As

a consequence, it is possible to say that, based on (13)-(15), the test statistic F_m is distributed as $\chi^2_{(m-1)^2}$ up to the term of the order $O_p\left(\left(\frac{n}{m^3}\right)^{1/2}\right)$; thus, the larger m is, the more precise the χ^2 approximation of the test statistic F_m is. Formally, $F_m = \chi^2_{(m-1)^2} + O_p\left(\left(\frac{n}{m^3}\right)^{1/2}\right)$.