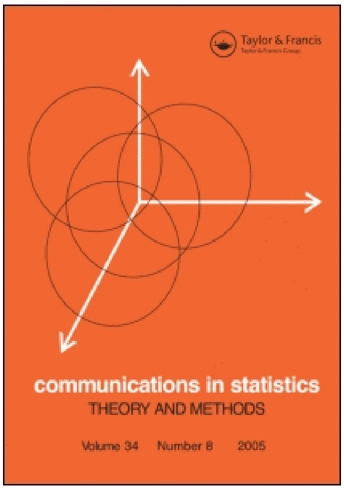


**Communications in Statistics - Theory and Methods**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

# Falsifying ARCH/GARCH Models Using Bispectral Based Tests

Melvin Hinich [a]
[a] Applied Research Laboratories, The University of Texas at Austin, Austin, Texas, USA

Online Publication Date: 01 January 2009

*Communications in Statistics—Theory and Methods*, 38: 529–541, 2009

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920802245741





# Falsifying ARCH/GARCH Models Using Bispectral Based Tests

MELVIN HINICH

Applied Research Laboratories, The University of Texas at Austin, Austin, Texas, USA

*This article shows that the Hinich (1982) bispectrum test for gaussianity and the Hinich and Rothman (1998) test for time reversibility can be used to falsify the null hypothesis that an autoregressive conditionally heteroskedastic model (ARCH) or its generalization (GARCH) generates nonlinear behavior in the variance of an observed time series. The term "falsify" means that the null hypothesis can be rejected with a given size using a nonparametric test based on the bispectrum where the data is trimmed to control the sizes. Rejecting the null hypothesis implies that the ARCH or GARCH model that is estimated from the data is not a complete statistical description of the dependence structure in the variance of the process.*

**Keywords** ARCH; Bispectrum; GARCH; Reversibility; Trimming.

**Mathematics Subject Classification** Primary 62M10; Secondary 62M07, 62M15.

## 1. Introduction

Linear and nonlinear time series models have been widely employed in the literature to explain the dynamics of financial time series. Since its introduction 24 years ago, the applications of Autoregressive Conditional Heteroskedasticity (ARCH) model introduced by Engle (1982) or its extension Generalized Autoregressive Conditional Heteroskedasticity (GARCH) by Bollerslev (1986) in finance have become commonplace (for a survey see Bollerslev et al., 1992). This class of models relaxes the assumption of the classical linear regression model that the variance of the disturbance term is conditionally as well as unconditionally constant.

Let $\{x(t_n)\}$ denote an equally spaced sampled time series from a stationary random process $\{x(t)\}$ where $t_n = n\tau$. A zero mean ARCH/GARCH model for this time series is of the form $x(t_n) = \sqrt{h(t_n)}e(t_n)$ where $\{e(t_n)\}$ a zero mean pure white noise process (i.i.d.) is and $\{h(t_n)\}$ is a positive valued autoregressive moving average process whose inputs are lagged $e^2(t_n)$. For example, an ARCH($q$) model

Received March 26, 2008; Accepted June 3, 2008
Address correspondence to Melvin Hinich, Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, USA; E-mail: hinich@mail.utexas.edu

is of the form $h(t_n) = \alpha_0 + \sum_{k=1}^{q} \alpha_k e^2(t_{n-k})$ and a GARCH($p, q$) model is $h(t_n) = \alpha_0 + \sum_{k=1}^{q} \alpha_k e^2(t_{n-k}) + \sum_{j=1}^{p} \beta_j h(t_{n-j})$. In most cases it is assumed that the $e(t_n)$ have a normal (gaussian) distribution but sometimes the assumed distribution is a Student's $t$.

Since the GARCH generalization the number of empirical and theoretical developments in the field has exploded, with rapid development of applications and variants. This popularity is evidenced by the incorporation of GARCH estimation into major software packages (for reviews of GARCH software see Brooks, 1997; McCullough and Renfro, 1999).

The ARCH/GARCH type of nonlinear time series model is claimed to be the model of a special type of nonlinearity in the data generating process known as multiplicative nonlinearity, or nonlinear-in-variance, in which nonlinearity affects the process through its variance (Hsieh, 1989). Although these models have been heralded as an accurate description of a number of important characteristics of financial data, as Hall et al. (1989) noted, the ARCH parameterization of the conditional variance does not have any solid grounding in economic theory, but represents "a convenient and parsimonious representation of the data."

Given the importance of these models in econometric time series it is important to be able to use a nonparametric statistical tool to falsify them. If it turns out that the ARCH/GARCH models lack a certain statistical property that has not been exploited then the time series community may create new nonlinear models that more accurately captures the complexity of the nonlinearity inherent in high-frequency market data.

The bispectrum is defined in the next section. Since the $e^2(t_n)$ are independently distributed it will now be shown that all the bispectral values of $\{x(t_n)\}$ are zero as long as the distribution of the $e(t_n)$ is *symmetric*. The Hinich test for gaussianity is really a test of the null hypothesis that the bispectrum is zero for all bifrequencies and thus, if the Hinich test rejects the null hypothesis, then the ARCH/GARCH specification is falsified for any set of model parameters. This point was first advanced by Brock (1987) in an unpublished article.

It will also be shown that the bispectrum of any ARCH or GARCH process is a real constant (its imaginary part is zero) for any distribution of $e(t_n)$ with finite moments. The Hinich-Rothman test of time reversibility is really a test for the null hypothesis that the imaginary part of the bispectrum is zero for all bifrequencies. Thus, if the Hinich–Rothman test rejects this null hypothesis the ARCH/GARCH specification is falsified in a nonparametric manner.

The asymptotic properties of these two tests are valid for ARCH/GARCH models that have finite moments but the fat tails of especially the GARCH processes produce false rejections for moderate and even large sample sizes. The definition of the bispectrum is given in Sec. 2. Section 3 covers the estimation of the bispectrum and the large sample properties of the test statistics. Section 4 presents a discussion of the use of data trimming to control the sizes of the tests. Simulations are presented in Sec. 5 to support validity of trimming to obtain proper test sizes for sample sizes common for high frequency financial data. Section 6 presents applications of the methods to five intraday NYSE stock rates of return series. An alternative nonlinear model that may yield better fits to the complex dynamics of asset returns is presented in Sec. 7.

## 2. The Bispectrum

The bispectrum of a bandlimited random process $\{x(t_n)\}$ is

$$B_x(f_1, f_2) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} b_x(r, s) \exp[-i2\pi(f_1 t_r + f_2 t_s)], \tag{2.1}$$

where $b_x(r, s) = Ex(t_n)x(t_{n+r})x(t_{n+s})$ is called the bicorrelation for lags $r$ and $s$. The set of positive support of the bispectrum is the triangle $\Omega = \{0 < f_1 < f_o/2, f_2 < f_1, f_1 + f_2 < f_o/2\}$, where $f_o$ is the highest frequency component of the process (Hinich and Messer, 1995), which is called the band limit in the signal processing literature.

The conditional product $E[x(t_n)x(t_{n+r})x(t_{n+s}) \mid e(t_m), m < n + s]$ is zero for all $0 < r \leq s$ and any ARCH or GARCH model since the $e(t_n)$'s are independently distributed, implying that $b_x(r, s) = 0$ for all $r \neq 0$ and $s \neq 0$. Thus, $B_x(f_1, f_2) = E[h^{3/2}(t_n)e^3(t_n)]$. The bispectra of any ARCH or GARCH process is a *real constant* for all bifrequencies, and thus if the Hinich–Rothman test rejects the null hypothesis that the imaginary part of the bispectrum is zero for all bifrequencies, then the process can not be ARCH/GARCH.

In addition, if $Ee^3(t_n) = 0$ for a given ARCH/GARCH model then its bispectrum is zero for all bifrequencies. With the zero skewness assumption for the $e(t_n)$'s then if the Hinich zero bispectrum test rejects the null hypothesis then it rejects the ARCH/GARCH model.

The bispectrum estimation method and the test statistics are presented in the next section.

## 3. Bispectral Estimation

The spectrum and bispectrum can be estimated using conventional nonparametric methods (Hinich and Clay, 1968). I prefer to use the frame averaging spectrum estimation method to illuminate the statistical issues of bispectrum normalization but the results will hold for any method that yields estimates that have similar asymptotic properties to the frame averaging method. Details of estimating the bispectrum from a sample of discrete-time observations of the process and the sampling properties of the estimate are presented in Brillinger (1965), Hinich (1982), Brockett et al. (1988), Hinich and Patterson (1989, 1992), and Hinich and Wolinsky (1988, 2005).

Consider a sample $\{x(t_1), \ldots, x(t_N)\}$ where $t_k = k\delta$. This sample is partitioned into $P = [N/L]$ non overlapping frames of length $L\delta$ where the last frame is deleted if it has less than $L$ observations. To simplify notation, normalize the time unit by setting $\delta = 1$. The resolution bandwidth is then $f_1 = \frac{1}{L}$. If $P = N/L$ then the last undersized frame is not used to estimate the bispectrum.

The $p$th frame is $\{x_p(1), \ldots, x_p(L)\} = \{x((p-1)L + 1), \ldots, x(pL)\}$. The discrete Fourier transform of the $p$th frame is $X_p(k) = \sum_{t=1}^{L} x_p(t) \exp(-i2\pi \frac{kt}{L})$ and the periodogram of the $p$th frame is $\frac{1}{L}|X_p(k)|^2 = \frac{1}{L} X_p(k) X_p(-k)$. Since $N \approx LP$, the frame-averaged estimate of the spectrum at frequency $f_k = \frac{k}{L}$ is

$$\widehat{S}(f_k) = \frac{1}{N} \sum_{p=1}^{P} |X_p(k)|^2. \tag{3.1}$$

Then $E[\widehat{S}(f_k)] = S(f_k) + O\left(\frac{1}{L}\right)$ where the error term of order $1/L$ is due to the frame windowing of the spectrum. The variance of the estimate for large values of $L$ and $P$ is $\frac{1}{P}S^2(f_k)$.

The frame-averaged estimate of the bispectrum at the bifrequencies $(f_{k_1}, f_{k_2})$ is

$$\widehat{B}(f_{k_1}, f_{k_2}) = \frac{1}{N}\sum_{p=1}^{P} X_p(k_1)X_p(k_2)X_p(-k_1-k_2). \tag{3.2}$$

Then $E[\widehat{B}(f_{k_1}, f_{k_2})] = B(f_{k_1}, f_{k_2}) + O\left(\frac{1}{L}\right)$ and the variance for large $L$ and $P$ is $\frac{L}{P}S(f_{k_1})S(f_{k_2})S(f_{k_1}+f_{k_2})$.

The normalization of the estimated bispectrum is

$$\widehat{\Gamma}(f_{k_1}, f_{k_2}) = \frac{\widehat{B}(f_{k_1}, f_{k_2})}{\sqrt{\widehat{S}(f_{k_1})\widehat{S}(f_{k_2})\widehat{S}(f_{k_1}+f_{k_2})}}. \tag{3.3}$$

This normalization standardizes the variance of the bispectrum estimate using the estimated variance in place of the true variance. Let $\Gamma(f_{k_1}, f_{k_2}) = [S(f_{k_1})S(f_{k_2})S(f_{k_1} + f_{k_2})]^{-\frac{1}{2}}B(f_{k_1}, f_{k_2})$. Let $L = N^{\varepsilon}$, where the bandwidth parameter $e$ is in the interval $0 < \varepsilon < 0.5$. Then the real and imaginary parts of each $\sqrt{2N^{-1+2\varepsilon}}[\widehat{\Gamma}(f_{k_1}, f_{k_2}) - \Gamma(f_{k_1}, f_{k_2})]$ are asymptotically independent Normal variates with zero means and unit variances as $N \to \infty$ (Hinich, 1982). Moreover, the $\widehat{\Gamma}(f_{k_1}, f_{k_2})$ are asymptotically independently distributed across the principal domain of the bifrequencies. The smaller the value of $\varepsilon$, the fewer the number of bifrequencies and thus the smaller the power of the tests but the larger the number of frames, which implies a faster convergence to the asymptotic sampling properties.

Suppose that the null hypothesis is that the imaginary part of the bispectrum is zero and thus $\operatorname{Im}\Gamma(f_{k_1}, f_{k_2}) = 0$ for all bifrequencies, which is true for ARCH/GARCH models. The Hinich–Rothman *TR* test statistic is the sum $S_{TR}$ of $2N^{-1+2\varepsilon}[\operatorname{Im}\widehat{\Gamma}(f_{k_1}, f_{k_2})]^2$ over the $L^2/16$ bifrequencies in the support set. This distribution of this sum is approximately central chi-squared with $M = L^2/16$ degrees of freedom for large $N$.

Now suppose that the null hypothesis is that the bispectrum is zero and thus $\Gamma(f_{k_1}, f_{k_2}) = 0$ for all bifrequencies, which is true for ARCH/GARCH models with $Ee^3(t_n) = 0$ as it is true for a gaussian process. The Hinich (1982) test statistic is the sum $S_0$ of $2N^{-1+2\varepsilon}|\widehat{\Gamma}(f_{k_1}, f_{k_2})|^2$ over the $L^2/16$ bifrequencies and its large sample distribution is central chi-squared with $2M$ degrees of freedom. For both tests, if the null hypothesis is false then the statistics have non central chi-square distributions and the tests are one sided.

The Hinich bispectrum based procedure transforms the tests statistic using the cumulative distribution function of the test statistics under the appropriate null hypothesis. For example, the Hinich zero bispectrum test statistic is $Y_0 = F_{2M}(S_0)$ where $F_{2M}(s)$ is the cumulative distribution function (cdf) of a central chi-square density with $2M$ degree of freedom. Thus, $Y_0$ has a uniform (0, 1) distribution under the null hypothesis. The null hypothesis is rejected if the $p$-value $p = 1 - F_{unif}(Y_0)$, where $F_{unif}$ is the cdf of the uniform (0, 1) distribution, is deemed to small by the analyst.

The Hinich–Rothman TR test statistic is $Y_{\mathrm{TR}} = F_M(S_{\mathrm{TR}})$. The null hypothesis is rejected if $p = 1 - F_{\mathrm{unif}}(Y_{\mathrm{TR}})$ is deemed too small.

The sampling properties of the Hinich bispectral based tests of normality and linearity as well as the Hinich and Rothman TR test are large sample results based on the asymptotic normal distribution's mean and variance. The validity of any asymptotic result for a finite sample is always an issue in statistics. The rate of convergence to normality depends on the size of the cumulants of the observed process.

All data is finite since all measurements have an upper bound to their magnitudes. If the data is leptokurtic as is the case for stock returns and exchange rates, then the kurtosis is large. Trimming the tails of the empirical distribution of the data is an effective statistical method to limit the size of the kurtosis and higher-order cumulants in order to get a more rapid convergence to the asymptotic distribution. Trimming of time series in order to improve the validity of the use of the asymptotic properties of the test is discussed next.



## 4. Trimming Time Series Data

Trimming data to make sample means less sensitive to outliers has been used in applied statistics for years. Trimming is a simple data transformation that make statistics based on the trimmed sample more normally distributed. Transforming data is a technique with a long pedigree; it can be dated back at least to Galton (1879) and McAlister (1879) at the dawn of modern statistics. Subsequently, Edgeworth (1898) and Johnson (1949), among others, have contributed to our understanding of this technique for examining data.

Suppose one wants to trim the upper and lower $\kappa/2\%$ values of the sample $\{x(t_1), \ldots, x(t_N)\}$. Order the data and find the $\kappa/200$ quantile $x_{\kappa/200}$ and the $1 - \kappa/200$ quantile $x_{1-\kappa/200}$ of the order statistics. Then set all sample values less than the $\kappa/200$ quantile to $x_{\kappa/200}$ and set all sample values greater than $1 - \kappa/200$ quantile to $x_{1-\kappa/200}$. The remaining $(100 - \kappa)\%$ data values are not transformed.

Assume that the $e(t_n)$'s have a symmetric density. Then the bicorrelations of the trimmed sample are *zero* for all lags using the same conditioning argument as was used to show that the bicorrelations of a ARCH/GARCH are zero. Thus, the imaginary part of the bispectrum of the trimmed sample is zero for all bifrequencies. Also, the real part is zero for all bifrequencies since $Ee^3(t_n) = 0$. Simulations will be presented to determining conservative trimming levels for two GARCH(1, 1) models and two ARCH(1) models previously used in the econometric literature.

## 5. Simulations Using GARCH(1, 1) and ARCH(1) Models

Two GARCH(1, 1) models and two ARCH(1) models were used in the simulations. Brooks et al. (2001) used the following GARCH(1,1) model parameters as a benchmark for their comparison of econometric software packages: $\alpha_0 = 0.0107613$, $\alpha_1 = 0.153134$, $\beta_1 = 0.805974$. This model is designated GARCH 1 in Tables 1–7. The simulations are designed to learn how much trimming is needed to get sizes that match the size derived from the asymptotic sampling theory.

The second GARCH(1, 1) model is in Example 2 of Horowitz et al. (2006), where $\alpha_0 = 0.001$, $\alpha_1 = 0.05$, $\beta_1 = 0.9$ using the notation above. This model is called GARCH 2.

**Table 1**
Zero bispectrum test sizes, Gaussian model, $c = 0.4$, $N = 50{,}000$

| Trimming level (%) | GARCH 1 | | GARCH 2 | |
|---|---|---|---|---|
| | 1% size | 5% size | 1% size | 5% size |
| $\kappa = 2$ | 27.78 | 51.38 | 22.19 | 45.03 |
| $\kappa = 5$ | 7.04 | 20.70 | 5.56 | 17.46 |
| $\kappa := 10$ | 1.12 | 5.45 | 0.87 | 4.58 |
| $\kappa := 15$ | 0.30 | 1.90 | 0.25 | 1.61 |
| $\kappa := 20$ | 0.10 | 0.83 | 0.08 | 0.74 |

**Table 2**
Time reversibility test sizes, Gaussian model, $c = 0.4$, $N = 50{,}000$

| Trimming level (%) | GARCH 1 | | GARCH 2 | |
|---|---|---|---|---|
| | 1% size | 5% size | 1% size | 5% size |
| $\kappa = 2$ | 14.83 | 33.40 | 11.85 | 29.20 |
| $\kappa = 5$ | 4.43 | 14.91 | 3.63 | 12.83 |
| $\kappa := 10$ | 1.18 | 5.42 | 1.00 | 4.69 |
| $\kappa := 15$ | 0.47 | 2.53 | 0.44 | 2.33 |
| $\kappa := 20$ | 0.21 | 0.1.53 | 0.18 | 1.43 |

**Table 3**
Zero bispectrum test, double tailed exponential $c = 0.4$, $N = 50{,}000$

| Trimming level (%) | GARCH 1 | | GARCH 2 | |
|---|---|---|---|---|
| | 1% size | 5% size | 1% size | 5% size |
| $\kappa := 10$ | 3.80 | 13.33 | 3.16 | 11.77 |
| $\kappa := 15$ | 0.66 | 3.49 | 0.56 | 3.06 |
| $\kappa := 20$ | 0.19 | 1.22 | 0.16 | 1.10 |

**Table 4**
Time reversibility test sizes, double tailed exponential $c = 0.4$, $N = 50{,}000$

| Trimming level (%) | GARCH 1 | | GARCH 2 | |
|---|---|---|---|---|
| | 1% size | 5% size | 1% size | 5% size |
| $\kappa = 10$ | 2.67 | 10.24 | 2.33 | 9.19 |
| $\kappa = 15$ | 0.78 | 3.83 | 0.67 | 3.51 |
| $\kappa := 20$ | 0.30 | 1.86 | 0.29 | 1.76 |

**Table 5**
Both test sizes, student $t(6)$ model, $c = 0.4$, $N = 50{,}000$

| Tests | GARCH 1 1% size | GARCH 2 5% size |
|---|---|---|
| Bispectrum = 0 | | |
| $T(6) - \kappa = 10$ | 2.43 | 9.39 |
| $T(6) - \kappa = 15$ | 0.46 | 2.69 |
| Time reverse | | |
| $T(6) - \kappa = 10$ | 1.92 | 7.90 |
| $T(6) - \kappa = 15$ | 0.58 | 3.28 |



**Table 6**
5% & 10% trimming levels for both tests $c = 0.4$, $N = 50{,}000$

| | ARCH 1 | | ARCH 2 | |
|---|---|---|---|---|
| Tests | 1% size | 5% size | 1% size | 5% size |
| Bispectrum = 0 | | | | |
| Normal $- \kappa = 5$ | 0.84 | 4.34 | 0.54 | 3.15 |
| Normal $- \kappa = 10$ | 0.21 | 1.39 | 0.16 | 1.15 |
| Time reverse | | | | |
| Normal $- \kappa = 5$ | 0.98 | 4.78 | 0.73 | 3.80 |
| Normal $- \kappa = 10$ | 0.39 | 2.14 | 0.34 | 1.86 |
| Bispectrum = 0 | | | | |
| $T(6) - \kappa = 5$ | 2.23 | 8.75 | 1.41 | 6.31 |
| $T(6) - \kappa = 10$ | 0.30 | 2.08 | 0.24 | 1.63 |
| Time reverse | | | | |
| $T(6) - \kappa = 5$ | 1.70 | 7.59 | 1.25 | 5.91 |
| $T(6) - \kappa = 10$ | 0.47 | 2.61 | 0.36 | 2.22 |

**Table 7**
$p$-values for both tests

| Tests | BIR | EOG | FE | IMN | NCC |
|---|---|---|---|---|---|
| Bispectrum = 0 | | | | | |
| $\kappa = 15$ | 0.000000 | 0.016148 | 0.000655 | 0.005348 | 0.053120 |
| $\kappa = 25$ | 0.000000 | 0.751091 | 0.070074 | 0.079690 | 0.479115 |
| Time reverse | | | | | |
| $\kappa = 15$ | 0.000086 | 0.031828 | 0.056489 | 0.384946 | 0.037025 |
| $\kappa = 25$ | 0.000850 | 0.462694 | 0.218435 | 0.517544 | 0.529166 |

Using a number of different trimming levels simulations were run using these models with 50,000 replications and a sample size of $N = 50{,}000$. The distribution of the pseudorandom pure white $\{e(t_n)\}$ used to generating the simulations are: gaussian (normal), double-tailed exponential, and Student $t$ with six degrees of freedom. The 1% and 5% sizes of the two test statistics were estimated from the 50,000 replications computed from trimmed series using different trimming levels in order to determine how close they were to the large sample sizes for the two test statistics.

The two values of the bandwidth parameter used to estimate the bispectra are $\varepsilon = 0.33$ and $\varepsilon = 0.4$. The differences between the quantiles computed for the two bandwidth parameters were small and so only the $\varepsilon = 0.4$ will be shown.

Table 1 presents the size results for the zero bispectrum test using the five trimming levels $\kappa = 2\%$, 5%, 10%, 15%, and 20% for the GARCH 1 and GARCH 2 models with normal pure white $\{e(t_n)\}$. Table 2 presents the results for the TR test using the same setup.

Note that the sizes are too large for the trimming levers of $\kappa = 2$ and 5%. The sizes would be even larger for no trimming. The sizes are controlled for a 10% data trimming level and the tests become conservative for a 15% trimming level.

Tables 3 and 4 present the size results for the two tests using the trimming levels $\kappa = 10$, 15, and 20% for the GARCH 1 and GARCH 2 models with double-tailed exponential pure white $\{e(t_n)\}$. The kurtosis of the exponential is six and thus a double tailed exponential density had fatter tails than is generally believed. For these time series the sizes are also conservative for a 15% data trimming level.

Table 5 presents the results for the two tests using the trimming levels $\kappa = 10$ and 15% for the GARCH 1 model where the $\{e(t_n)\}$ have a Student $t$ with six degrees of freedom. Once again the 15% trimming level yields conservative sizes.

To test several trimming levels for ARCH models two ARCH(1)'s used by Becker and Hurn (2004) were simulated. The ARCH 1 parameters are $\alpha_0 = 1$ and $\alpha_1 = 0.8$. The ARCH 2 parameters are $\alpha_0 = 1$ and $\alpha_1 = 0.5$. The size results for the trimming levels $\kappa = 5$ and 10% and the normal and Student $t(6)$ $\{e(t_n)\}$ are presented in Table 6. For both ARCH models a $\kappa = 5\%$ trimming is sufficient to yield proper sizes for the tests for the normal $\{e(t_n)\}$. For the Student $t(6)$ case, the 10% trimming yields conservative sizes but the 5% trimming is not sufficient to keep the sizes at or below their target values.

Trimming at the 15% level makes the two tests conservative for all the models used but will reduce their power of the test to falsify the ARCH/GARCH hypothesis. It is important to keep in mind that not rejecting the null hypothesis is *not* the same as accepting it. If the tests reject the null after trimming then one has confidence that the null is rejected. One approach to justifying this confidence is to fit an ARCH or GARCH model assuming a symmetric density for the $\{e(t_n)\}$ and then using my simulation program to determine the proper trimming level.

The power of the bispectrum test for trimmed output of a new type of nonlinear model is presented in the next section.

## 6. Intraday Stock Returns Example

The two bispectrum based tests were applied to five intraday stock returns for the period January 4, 1999–December 29, 2000 randomly selected set of stock data given to me by Professor Doug Patterson. This is an interesting period because it was the peak of the "dot com" bubble. The stocks are Birmingham Steel (BIR), EOG Resources (EOG), First Energy (FE), Imation (IMN), and National City Corporation (NCC).

The original tick strike prices were adjusted for dividends and stock splits and smoothed using a Tukey filter to removed low-frequency and trend components. Then each series was skip sampled to produce a unaliased aggregated price for each ten minutes for each trading day. The rates of return were constructed from the actual trade prices corrected for splits by a method that produced unaliased ten minute averages for each trading day. There are 36 such aggregated rates for each trading day yielding 19,656 points.

The NYSE trading day is six hours long. Each frame has 720 10-minute returns and thus the frame length is twenty trading days.

The two bispectrum based tests presented in this article are not the only tests that have been used to falsify ARCH or GARCH models with symmetrically distributed innovation. Brooks and Hinich (1998) presented a test using a binary transformation of the data that is called "hard clipping" in the signal processing literature. Each return is transformed to a one if it is positive or zero and a minus one if it is negative. A symmetric GARCH model implies that transformed returns will be a Bernoulli process. This idea was first presented by Hinich and Patterson in an article given in the NSF-NBER Time Series Workshop in 1995 and revised in 2001. A pdf file with the name "Episodic-stocks" is on my webpage in the folder Statistics. A shortened version edited book by Belongia and Brinner referenced below.

In the 2001 article, we also present a sequence of the Engle LM test for ARCH/GARCH computed for seven segments of 214 days of high-frequency rates of return for a number of NYSE. The IBM and General Motors rejected the null at the 1% level for 5 of the 7 segments and the other stocks rejected less often. There was no pattern to the subgroup rejections. If the returns were well modeled by an ARCH model for the whole period, the LM test would reject for each segment for at least the 5% level but that was not so.

To replicate the episodic nature of the LM test I computed the LM test with 10 lags for 27 non overlapping frames of 10-min samples of the rates of returns for the period January 4, 1999–December 29, 2000 of the following NYSE stocks.

The LM statistic used has a central chi-square distribution with ten degrees-of-freedom for the null that the series is white noise. My program called T23 transforms the chi-square statistic to a uniform (0,1) variate under the null using the cumulative distribution function of a $\chi_{10}$. The results are shown in Figs. 1–3. For Birmingham Steel the null hypothesis is rejected at the 1% level for 16 of the 27 frames and only 19 at the 5% level. For First Energy the null is rejected at the 1% level for 14 frames and 19 at the 5% level. For Imation the null is rejected at the 1% level for 11 frames and only 12 frames at the 5% level.

Whereas the hard clipping test is immune to distributional assumptions of ARCH and GARCH, other than the requirement that the $e(t_n)$ are symmetric the use of the asymptotic properties of the LM test is suspect. The two bispectral based tests using trimming that is presented in this article provides a more general test of the validity of the model structure.

The bispectrum tests applied to the five intraday data sets used a frame length of $L = 72$ implying an exponent value $c = 0.43$. Thus, the number of full frames is 273.

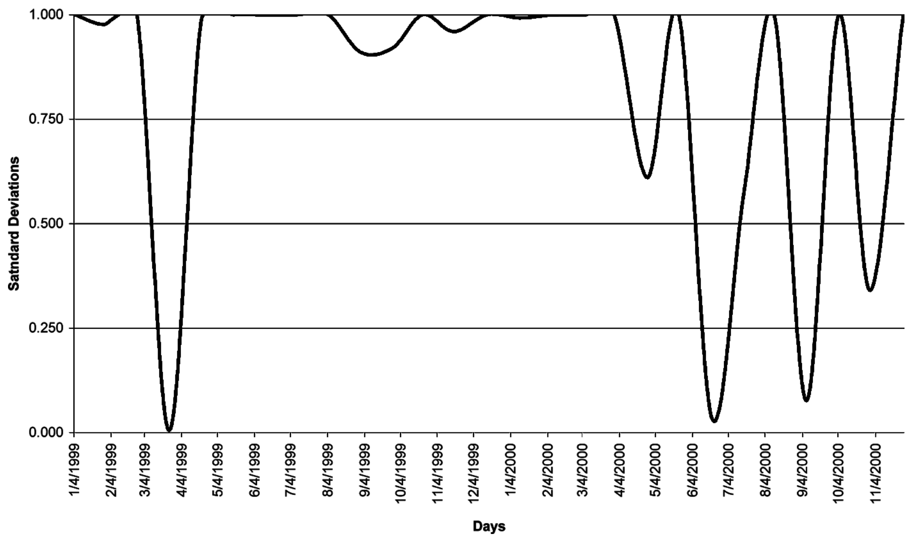


**Figure 1.** LM(10) statistics for 27 frames of 720 10-min–10-min Birmingham steel ticks January 4, 1999–December 29, 2000.



The $p$-values of the test were computed to six decimal places. The $p$-values for the two test using the trimming levels of $\kappa = 15$ and $\kappa = 25\%$. The results are shown in Table 7. Both tests on the BIR data were highly significant for both trimming levels. The zero bispectrum test for FE and IMN were highly significant for $\kappa = 15\%$ but at the higher trimming level the tests were not significant at the 5% level. Both tests were significant for the EOG data but not for the $\kappa = 25\%$ trimming.

But at least one bispectral $p$-value was highly significant for the FE, EOG, IMN, and NDD bispectra for $\kappa = 25\%$ trimming. The number of biperiods

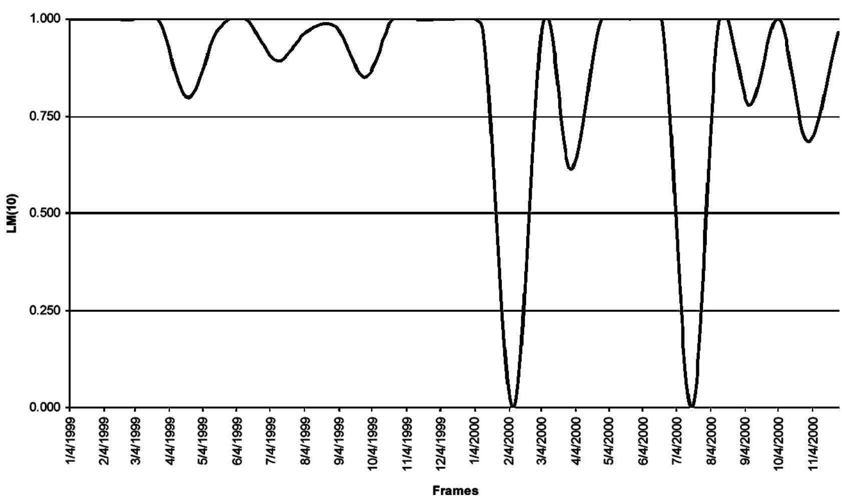


**Figure 2.** LM(10) statistics for 27 frames of 720 10-min first energy ticks January 4, 1999–December 29, 2000.

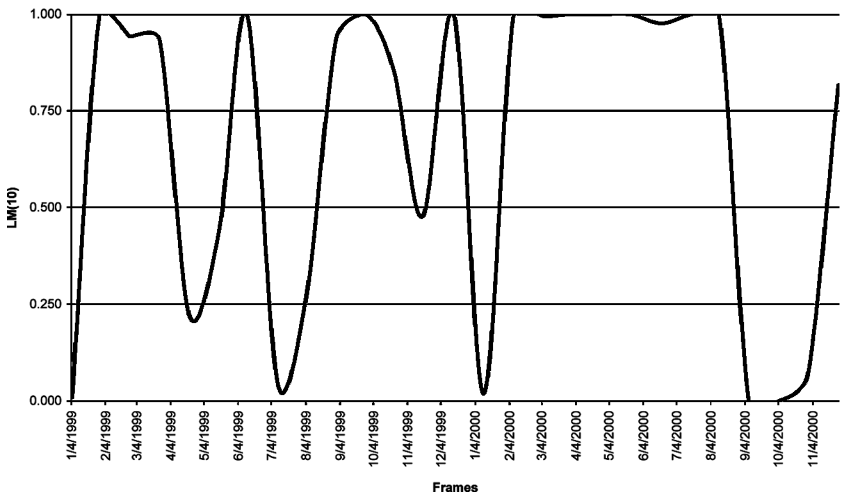


**Figure 3.** LM(10) statistics for 27 frames of 720 10-min imation ticks January 4, 1999–December 29, 2000.



(bifrequencies) in the support set is 325. The estimates are approximately gaussian and independently distributed across the bifrequencies points.

For the 25% trimming the EOG bispectrum has a $p$-value of $1.95 \times 10^{-3}$ for the biperiod (14, 8) which is very unlikely for a set of 325 independent values. The IMN bispectrum has $p$-values of $0.49 \times 10^{-3}$ for (14.4, 18) and $1.42 \times 10^{-3}$ for (14.4, 14.4). The NDD bispectrum has $p$-values of $3.64 \times 10^{-3}$ for the biperiod (14.4, 12) and $4.26 \times 10^{-3}$ for (14.4, 10.29). Thus, the null hypothesis of ARCH/GARCH is rejected at this higher level of trimming. Perhaps the trimming level is not high enough but the simulations presented in this article suggests that the results call into question the adequacy of the model class for the data analyzed.

The next section presents an alternative nonlinear model that may better explain the episodic nonlinear nature of financial rates of returns.

## 7. An Alternative Nonlinear Model

The following nonlinear AR(2) model is a discrete-time version of the continuous time Duffing equation (Jordan and Smith, 1987, Ch. 5, Sec. 6), which is a nonlinear model of vibrations around an equilibrium.

$$x(t_n) + a_1(x(t_{n-2}))x(t_{n-1}) + a_2(x(t_{n-2}))x(t_{n-2}) = \sigma u(t_n), \tag{7.1}$$

where

$$\begin{aligned} a_2(x(t_{n-2})) &= \exp(-c(1 + \delta_c x^2(t_{n-2}) + \delta_c^2 x^2(t_{n-3}) + \delta_c^3 x^2(t_{n-4}) + \delta_c^4 x^2(t_{n-5}))), \\ a_1(x(t_{n-2})) &= -2a_2(x(t_{n-2}))\cos 2\pi f(1 + \delta_f x^2(t_{n-2}) + \delta_f^2 x^2(t_{n-3}) \\ &\quad + \delta_f^3 x^2(t_{n-4}) + \delta_f^4 x^2(t_{n-5})) \end{aligned} \tag{7.2}$$

If the perturbation parameter $\delta_c$ is small relative to the damping coefficient $c$ and the other perturbation parameter $\delta_f$ is small, then the sample paths of the model

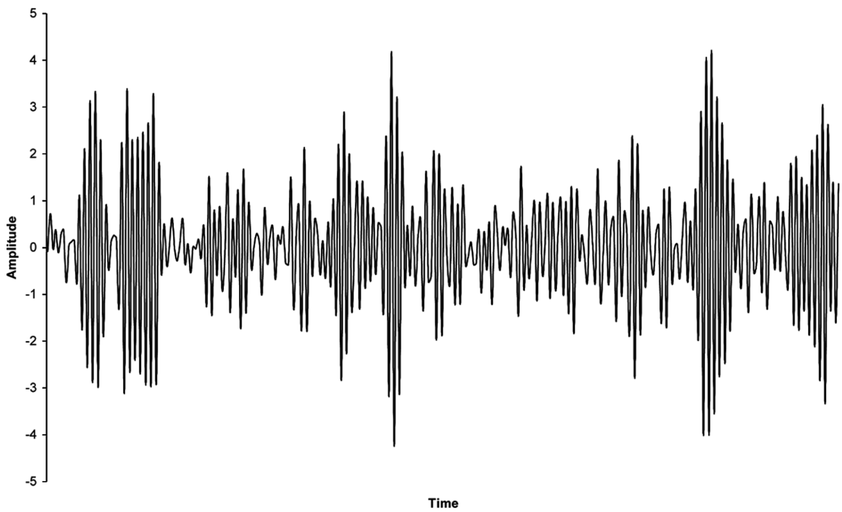


**Figure 4.** Nonlinear AR(2) model with $c = 0.2$, $f = 0.4$, $\delta_c = 0.005$, $\delta_t = 0.002$, $\sigma = 0.5$.

are quasi periodic with a seeming random variation in amplitude and phase. An example is given in Fig. 4 for $N = 300$ with the parameter values $c = 0.2$, $f = 0.4$, $\delta_c = 0.05$, $\delta_f = 0.02$, and $\sigma = 0.5$.

## 8. Conclusion

The methods presented in this article are to designed to aid an empiricist to determine if the ARCH/GARCH model specification is supported by data analysis. This article is not an attack on the specification.

Data trimming has been shown to control the distortion of fat-tailed time series on the sizes of the zero bispectrum tests that can be used to falsify ARCH/GARCH models in a nonparametric fashion. The results presented in this article show that the trimming level depends on the nature of the model used. How is a user to determine the proper trimming level to use the bispectrum based tests to see if a fitted ARCH or GARCH is falsified? One way is to use my simulation program SIMARCH that I used for the simulations used for this article. The program source code and its executable are in the folder CUMSPEC on my webpage. The program BISPEC is also available on my website (www.gov.utexas.edu/hinich).

## References

Becker, R., Hurn, A. S. (2004). Testing for nonlinearity in mean in the presence of heteroskedasticity, School of Economics and Finance, Queensland University of Technology.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31:307–327.

Bollerslev, T., Chou, R. Y., Kroner, K. F. (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *J. Econometrics* 52(5):5–59.

Brillinger, D. R. (1965). An introduction to polyspectra. *Ann. Mathemat. Statist.* 36:1351–1374.

Brock, W. A. (1987). *Nonlinearity in Finance and Economics*. Department of Economics, The University of Wisconsin.

Brockett, P. L., Hinich, M. J., Patterson, D. M. (1988). Bispectral based tests for detection of gaussianity and linearity in time series. *J. Amer. Statist. Assoc.* 83(403):657–664

Brooks, C. (1997). GARCH modeling in finance: a review of the software options. *Economic J.* 107:1271–1276.

Brooks, C., Hinich, M. J. (1998). Episodic nonstationarity in exchange rates. *App. Econ. Lett.* 5:719–722.

Brooks, C., Burke, S. P., Persand, G. (2001). Benchmarks and the accuracy of GARCH model estimation. *Int. J. Forecasting* 17:45–56.

Edgeworth, F. Y. (1898). On the representation of statistics by mathematical formulae. *J. Roy. Statist. Soc.* 61:670–700.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1007.

Galton, F. (1879). The geometric mean in vital and social statistics. *Proc. Roy. Soc. London* 29:365–367.

Hall, S. G., Miles, D. K., Taylor, M. P. (1989). Modeling asset prices with time-varying betas. *The Manchester School* 57(4):340–356.

Hinich, M. J. (1982). Testing for gaussianity and linearity of a stationary time series. *J. Time Ser. Anal.* 3:169–176.

Hinich, M. J., Clay, C. S. (1968). The application of the discrete fourier transform in the estimation of power spectra, coherence and bispectra of geophysical data. *Reviews of Geophysics* 6(3):347–363.

Hinich, M. J., Wolinsky, M. (1988). A test for aliasing using bispectral analysis. *J. Amer. Statist. Assoc.* 83(402):499–502.

Hinich, M. J., Patterson, D. M. (1989). Evidence of nonlinearity in the trade-by-trade stock market return generating process. In: Barnett, W., Geweke, J., Shell, K., eds. *Economic Complexity, Chaos, Sunspots, Bubbles, and Nonlinearity*. Ch. 16, Cambridge: Cambridge University Press, pp. 383–409.

Hinich, M. J., Patterson, D. M. (1992). A new diagnostic test of model inadequacy which uses the martingale difference criterion. *J. Time Ser. Anal.* 13(3):233–252.

Hinich, M. J., Dalle Molle, J. (1995). Trispectral analysis of stationary random time series. *J. Acoustical Soc. Amer.* 97(5):2963–2978.

Hinich, M. J., Messer, H. (1995). On the principal domain of the discrete bispectrum of a stationary signal. *IEEE Trans. Signal Process.* 43(9):2130–2134.

Hinich, M. J., Rothman, P. (1998). Frequency-domain test of time reversibly. *Macroeconomic Dyn.* 2(1):72–88.

Hinich, M. J., Patterson, D. M. (2005). Detecting epochs of transient dependence in white noise. In: Binner, J., Belongia, M., eds. *Money, Measurement and Computation*. London: Palgrave.

Hinich, M. J., Wolinsky, M. (2005). Normalizing bispectra. *J. Statist. Plann. Infer.* 130:405–411.

Hinich, M. J., Lim, K. P., Liew, V. K. S. (2005). Statistical inadequacy of GARCH models for Asian stock markets: evidence and implications. *J. Emerging Market Finance* 4(3):263–279.

Horowitz, J. L., Lobato, I. N., Nankervis, J. C., Savin, N. E. (2006). Bootstrapping the box–pierce *Q* test: a robust test of uncorrelatedness. *J. Econometrics* 133(2):841–862.

Hsieh, D. A. (1989). Testing for nonlinear dependence in daily foreign exchange rates. *J. Business* 62:339–368.

Johnson, N. L. (1949). Systems of frequency curves generated by means of translation. *Biometrika* 36:149–176.

Jordan, D. W., Smith, P. (1987). *Nonlinear Ordinary Differential Equations*. Oxford: Oxford University Press.

McAlister, D. (1879). The law of the geometric mean. *Proc. Roy. Soc. London* 29:367–376.

McCullough, B. D., Renfro, C. G. (1999). Benchmarks and software standards: a case study of GARCH procedures. *J. Econ. Social Measurement* 25:59–71.