Order Number 9224144

Testing nonlinearities in economic time series under moment condition failure

de Lima, Pedro José Frias, Ph.D.

The University of Wisconsin - Madison, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A dissertation entitled

TESTING NONLINEARITIES IN ECONOMIC TIME SERIES
UNDER MOMENT CONDITION FAILURE

submitted to the Graduate School of the
University of Wisconsin-Madison
in partial fulfillment of the requirements for the
degree of Doctor of Philosophy

by

PEDRO JOSE FRIAS DE LIMA

Degree to be awarded: December 19___ May 19___ August 19 92

Approved by Dissertation Readers:

Major Professor

May 27 1992

Date of Examination

Dean, Graduate School

TESTING NONLINEARITIES IN ECONOMIC TIME SERIES UNDER MOMENT CONDITION FAILURE

BY

PEDRO JOSÉ FRIAS DE LIMA

A thesis submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN-MADISON

1992

TESTING NONLINEARITIES IN ECONOMIC TIME SERIES

UNDER MOMENT CONDITION FAILURE

Pedro José Frias de Lima

Under the supervision of Professor William A. Brock

at the University of Wisconsin-Madison

Abstract

This dissertation studies nonlinearity testing in economic and financial time series.

Chapter 1 relates the topics of nonlinearity testing and moment condition failure. It has been documented in the economic and financial literature that time series data such as stock market indexes and exchange rates are generated by heavy-tailed distributions for which fourth moments are usually not finite. For such type of data, almost every test of nonlinearity available will provide wrong statistical inferences, because the asymptotic properties of those tests are established assuming that the series under study have, at least, finite fourth moments. Chapter 1 provides a survey of the moment conditions required by some widely-used tests for nonlinearity and temporal dependence, as well as a simulation study of the consequences of using these tests when the data is characterized by moment condition failure.

Chapter 2 develops a family of tests for the hypothesis of

independent and identically distributed time series. The tests make use of empirical processes theory and can be seen as a generalization of a test for nonlinearity proposed by Brock, Dechert and Scheinkman. It is shown that the tests can be used for testing the iid hypothesis regardless the existence of moments of any order.

Chapter 3 deals with the construction of statistics for the testing of the hypothesis that a stochastic process is linear. It is shown that the large sample properties of the family of tests introduced in chapter two are unaffected by the use of estimated residuals in place of the unobserved innovations. The proof of this result is obtained by generalizing results available in the statistical literature concerning U-Statistics with nondiferentiable kernels. The result is constructed by first proving this property for the BDS test; the generalization to the family of tests of Chapter 2 is also established. Therefore, these tests can be used as diagnostic tests. It is shown that the tests are robust to the non-existence of moments. The chapter concludes with a discussion of the relevance of this property for nonlinearity testing in general, as well as for the testing of the market efficiency hypothesis.

ACKNOWLEDGMENTS

I wish to thank several people for their support and assistance during the period I spent writing this dissertation. First, I would like to thank my adviser William Brock for letting me participate in his exciting agenda of economic research. His constant support and suggestions were a decisive factor in the making of this dissertation. I also would like to thank Mico Loretan for the extremely helpful discussions on econometric issues, as well as his constant encouragements. To Ken West, I want to acknowledge the usefulness of his comments.

I also wish to recognize my gratitude to all the friends who made my four years in Madison an enjoyable experience. Among them, I would like to mention Ellen Sapega, Federico Perali, Jeff Dominitz, Julliet King and Phil Ruder.

I also would like to thank the decisive financial support from the following institutions: Comissão Luso Americana/Fulbright, Junta Nacional de Investigação Científica e Tecnológica, Instituto Superior de Economia e Gestão, and the Wisconsin Alumni Research Foundation.

This list of acknowledgments would not be complete without mentioning my parents and my wife Luisa. Taking part in their lives has definitely been my biggest achievement.

To all, I want to express my deepest "obrigado".

TABLE OF CONTENTS

CHAPTER 2: A FAMILY OF TESTS FOR THE HYPOTHESIS OF I.I.D. TIME SERIES

CHAPTER 3: A FAMILY OF NONLINEARITY TESTS ROBUST TO HEAVY-TAILED DISTRIBUTED DATA

LIST OF FIGURES AND TABLES

CHAPTER 1.

CHAPTER 3

CHAPTER 1: NONLINEARITY TESTING AND MOMENT CONDITION FAILURE

1. INTRODUCTION

Nonlinear time series analysis has recently become one of the most active areas of research in Econometrics. There are several reasons why econometricians, as well as economists in general, have become interested in this type of models. We first note that most researchers have always considered linear models as some form of approximation to the true underlying model, but this use was often justified with computational considerations alone. Therefore, as progress in computing technology — as well as on the determination of the statistical properties of nonlinear methods — has been made, it would seem quite natural to use nonlinear models as a way to increase the quality of the approximations that any researcher is forced to make when trying to capture the true relationships between economic variables.

Secondly, the models that have dominated theoretical macroeconomics and finance in the last two decades are inherently nonlinear in nature. While linear approximations have been thought to work fairly well in economics, they do possess some important limitations. Consequently, economists are increasingly using models that are able to generate more interesting and complex dynamics than those generated by linear models.

The emerging preference for nonlinear models is directly reflected in the econometrics literature, which has been invaded with a growing number of new time series models. Autoregressive exponential models, bilinear models, threshold autoregressive models, and Generalized Autoregressive Conditionally Heteroskedastic (GARCH) models are just a few examples of the vast class of nonlinear models. Although this is an indication of the excitement that this new area has brought to economics, it also represents a demanding challenge for the applied researcher. In fact, those who deem nonlinearities an important characteristic of economic time series still face a most difficult task of modeling these nonlinearities. Such modeling involves choosing from many possible alternatives, while only a few (if any) rules exist for choosing between models. Furthermore, linear models have a strong theoretical support. Indeed, the Wold Decomposition Theorem states that every covariance stationary process can be expressed as an infinite order moving average process.

This simple observation implies that, from a practical point of view, nonlinear modeling is a costly task. Meanwhile, nonlinear models might represent more parsimoniously parameterized models of a given time series, with significant improvements in forecasting abilities. Thus, the simplicity and generality of linear models strongly suggests the need for tests that are indeed capable of telling a researcher whether or not linear models capture all the dynamics of the series under study. This dissertation is motivated by that simple fact and its purpose is to study the issue of nonlinearity testing. Therefore, and in order to proceed, we first define the concept of linearity that we are utilizing.

2. PRELIMINARIES

2.1 Linear processes.

To motivate the definition of linear process we first introduce the so-called Wold Decomposition Theorem. This fundamental result states that any zero-mean weakly (covariance) stationary process $\{Y_t\}$ can be represented as the sum of an infinite order moving average — MA(∞)) — process and a purely deterministic process (where purely deterministic means a process with zero variance). In formal terms:

Wold Decomposition Theorem:

Let $\{U_t\}$ be a zero mean, constant variance ($\sigma^2>0$) white noise sequence and $\{V_t\}$ a deterministic process, with $E[U_t V_s]=0$ for all $s,t \in \mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. The weakly stationary zero mean process Y_t can be expressed as

$$Y_t = \sum_{j=0}^{\infty} \psi_j U_{t-j} + V_t$$

where:

1) $\psi_0=1$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$

2) $U_t \in \mathcal{M}_t$ for each $t \in \mathbb{Z}$, where $\mathcal{M}_n = \overline{sp}\{Y_t, -\infty < t \leq n\}$ is the closed span of $\{Y_t, -\infty < t \leq n\}$, a subset of the Hilbert space $\mathcal{M} = \overline{sp}\{Y_t, t \in \mathbb{Z}\}$.

3) $V_t \in \mathcal{M}_{-\infty}$ for each $t \in \mathbb{Z}$

Proof:

See Brockwell and Davis (1987), p. 180, Theorem 5.7.1.

Most of the traditional time series analysis deals with purely non-deterministic models. For this class of models, the above decomposition is now stated as $Y_t = \sum_{j=0}^{\infty} \psi_j U_{t-j}$. From the perspective that Y_t can be obtained as a linear combination of past and present innovations it could be said that Y_t is a linear process. However, the Wold decomposition result guarantees that we may express every weakly purely deterministic stationary process in this form. Therefore, any meaningful (testable) definition of linearity must impose some additional structure (stronger than white noise) on the stochastic properties of $\{U_t\}$.

Brock and Potter (1991) defined two major categories for linear processes. Following their work, and restricting attention to purely non-deterministic processes, we say that a process is iid-linear if:

Definition 1. (IID-linear)

If $\{U_t\}$, in the Wold decomposition result, forms an iid (independent and identically distributed) sequence, then $\{Y_t\}$ is iid-linear.

Similarly, we define mds-linearity as:

Definition 2. (MDS-linear)

If $\{U_t\}$, in the Wold decomposition result, forms an mds (martingale difference sequence), then $\{Y_t\}$ is mds-linear.

A martingale difference sequence is defined as follows:

Definition 3. (White (1984), p. 56)

Let $\{U_t,\mathcal{F}_t\}_{t=1}^{\infty}$ be an adapted stochastic sequence, that is, U_t is a random scalar measurable with respect to an increasing sequence of σ-algebras $\mathcal{F}_t$. $\{U_t,\mathcal{F}_t\}$ is an mds if and only if $E[U_t|\mathcal{F}_{t-1}]=0$, for all $t\geq2$.

We are interested in linear processes that have autoregressive moving averages — ARMA(p,q) — representations, $\rho(L)y_t=\theta(L)U_t$, where L is the lag operator and $\rho(L)$ and $\theta(L)$ are, respectively, p-th order and q-th order lag polynomials. To this effect, we further assume that the sequence $\{\psi_j\}$ in the Wold decomposition is absolutely summable, and that $\rho(z) \equiv 1+\sum_{k=1}^{p}\rho_k z^k \neq 0$ for all $|z|\leq 1$, $\theta(z) \equiv 1+\sum_{k=1}^{p}\theta_k z^k \neq 0$ for all $|z|<1$, and $\rho(z)$ and $\theta(z)$ have no common factors. This guarantees that the ARMA(p,q) process is causal and invertible — see Brockwell and Davis (1987), Theorems 3.1.1 and 3.1.2.

To shed some light on the difference between the concepts of iid-linearity and mds-linearity, recall that for a zero mean process, iid $\Longrightarrow$ $E[U_t|\mathcal{F}_{t-1}]=0 \Longrightarrow E[U_tU_s]=0$, $\forall$ $t\neq s$.[1] The concept of iid-linearity can be motivated by analogy with a regression framework. If the goal is to establish the "best" relationship between two variables, y and x, then one is looking for a function f(x) such that $y_i-f(x_i)$ forms a zero mean, independent sequence. If this is the case, it means that the model does not leave any structure to "explain". This dissertation concentrates on the issue of iid-linearity testing.

However, it is perhaps more natural to define stochastic linearity as the case where the best linear predictor coincides with the minimum mean square error predictor of Y_t conditional on $\mathcal{F}_{t-1}$ — i.e., $E[Y_t|\mathcal{F}_{t-1}]$, where $\mathcal{F}_{t-1}$ is the σ-algebra generated by Y_m, $m\leq t$. This is equivalent to the condition that $E[U_t|\mathcal{F}_{t-1}]=0$, a.s., for every t — Hall and Heyde (1980, p. 183).

Evidently, many possible cases exist when iid-linearity and mds-linearity do not coincide. One such case occurs when $E[U_t^2|\mathcal{F}_{t-1}]\neq E[U_t^2]$. Note that this case is not ruled out by the Wold decomposition, which only guarantees that $E[U_t^2]$ is constant across time. However, the mds-linearity concept is compatible with this difference between conditional and unconditional higher (than the first) order moments. For example, an autoregressive process with ARCH errors is mds-linear but not iid-linear. This is a simple consequence of the fact that under iid every measurable function of U_t should be uncorrelated with any measurable function of the past U_t's, while the mds null implies only that U_t is uncorrelated with any function h of the past U_t's, such that $E|h(.)U_t| < \infty$.

At the same time, it must be noticed that if $\{Y_t\}$ is a Gaussian process, the existence of nonlinearities is ruled out by the fact that the $\{U_t\}$ in the Wold decomposition will be an uncorrelated Gaussian sequence (and therefore iid).

2.2 Nonlinearities and Heavy-Tailed Distributions

It has long been claimed that normality is not a suitable distributional assumption for some economic time series — especially for financial data. As a matter of fact, it is usually accepted that the distributions of series like stock market returns and exchange rates have heavy tails, as compared with those of normal distributions. Such distributions are said to be leptokurtic. In particular, it is known that the existence of moments for a given distribution relates closely to the rate of decay of the distribution's tails. Therefore, for heavy-tailed distributions higher order moments may not exist at all.

That moment condition failure is an important issue in economics is documented in literature following Mandelbrot's (1963) work. Mandelbrot gave some evidence that unconditional second moments might not exist for commodity price changes. More recently, Loretan and Phillips (1992) presented estimates of the maximal moment exponent, $\alpha = sup_{q>0}\{E|U|^q<\infty\}$, for some financial time series, concluding that, frequently, $2<\alpha<4$, i.e., while second moments of this data seem to be finite, fourth order (unconditional) moments do not. Furthermore, rates of foreign exchange price changes also seem to suffer from fourth moment condition failure — see Loretan (1991). On the basis of these estimates, then fourth moments are generally not finite, strongly indicating that the data are not Gaussian.

It is important to note that a considerable portion of nonlinearity testing and modeling has dealt primarily with the above-mentioned type of economic data. This raises a question, namely, how much are nonlinearity tests robust to the nonexistence of higher-order moments.

The statistical implications of this problem can be illustrated simply using the Central Limit Theorem for the sample mean. The simplest version of this theorem (due to Lindeberg and Levy, c.f. White (1984), Theorem 5.2) — covering the iid case — requires the distribution function of the data to have finite variance. If this is not the case, central limit results remain available for the so-called stable distributions, but conventional asymptotic approximations that apply the usual theorem (i.e., Gaussian Central Limit Theorem) lack any theoretical support in such cases.

More generally, the usual test methodology derives the large sample properties of some statistic on the assumption that the hypothesis to be tested is true. To derive the limiting distribution of that statistic, a central limit theorem must be invoked. Central limit theorems usually require two sets of conditions, namely on serial dependence among observations and on moment conditions (as a way to control the outlier activity). If one such set of conditions is not satisfied the assumed limiting distribution will not follow and the associated asymptotic approximation will become invalid.

3. COMPARISON OF MOMENT-REQUIREMENTS OF NONLINEARITY TESTS

We now present a survey of the moment conditions required by some widely-used tests for nonlinearity and temporal dependence. This investigation is motivated by the fact that moment condition failure frequently is a feature of some economic time series, especially financial time series, whereas many tests of nonlinearity make strong assumptions about the existence of moments. We do not aim to explain in detail how these tests work, or to discuss their small-sample properties (for such discussion see Tong (1990) and Granger and Teräsvirta (1992). Instead, the topic is testing for linearity under weak moment conditions.

3.1 Bispectrum-Based Tests

Subba Rao and Gabr (1980) have proposed a frequency domain test for the null of linearity based on the fact that if a series has a linear representation — with $E[U_t^3] < \infty$ — then the ratio

$$Y_{ij} = \frac{|f(\omega_i, \omega_j)|^2}{f(\omega_i)f(\omega_j)f(\omega_i + \omega_j)}$$

is constant, that is, it does not depend on the frequencies — ω_i and ω_j — used. Here, $f(\omega)$ and $f(\omega_1, \omega_2)$ are, respectively, the spectral density and the bispectral density function. In particular, if the process Y_t is Gaussian, the above ratio is zero.

This test, as well as the modification proposed by Hinich (1982), requires the existence of finite sixth moments of Y_t, as Subba Rao and Gabr (1980, p.147) explicitly acknowledge. This requirement reflects the fact that to establish the limiting distribution of the test statistic, one must guarantee that the bispectrum estimates are asymptotic normal. As stated in Theorem 2 in Van Ness (1966), the asymptotic normality of the bispectrum estimates is established under the hypothesis that $\{Y_t\}$ has finite unconditional sixth moments.

It is important to stress that the bispectrum based tests are directly applied to the series $\{Y_t\}$. We now turn our attention to tests that are performed on estimated residuals of a linear process.

3.2 The McLeod and Li test

This test is based on the fact that if $\{Y_t\}$ is a stationary Gaussian time series, then

$$\mathrm{Corr}(Y_t^2, Y_{t-k}^2) = \mathrm{Corr}(Y_t, Y_{t-k})^2, \quad \forall\, k$$

where corr(.,.) is the correlation coefficient. Any departures from the above result might indicate nonlinearity. McLeod and Li (1983) showed the portmanteau statistic $Q=T(T+2)\sum_{k=1}^{m} r_k^2/(T-k)$ as asymptotically distributed $\chi^2(m)$ under the assumption that $\{Y_t\}$ is an ARMA(p,q) process with independent and identically distributed innovations. Here, r_k is the sample autocorrelation of the squared fitted residuals of an ARMA(p,q) model. This asymptotic result is valid under the existence of finite eighth moment of U_t — see McLeod and Li, p. 271.

3.3 Tests of Linearity Based on Volterra Expansions

Volterra series expansions acknowledge that Y_t can be expressed as a function of past and present innovations — $Y_t=h(U_t, U_{t-1}, U_{t-2}, \ldots)$ — and that the function h is sufficiently well-behaved so that it can be expanded in a Taylor series about the origin — see Priestley (1988).

Keenan (1985) proposed a test of linearity against a second-order Volterra series expansion of Y_t. The test is based on runs an auxiliary regression which augments the linear model under test with the quadratic terms of the dependent variable. The resulting test statistic is asymptotically distributed F and is similar to Tukey's one degree of freedom test. Tsay's (1986) generalization of Keenan's test, however, has proved to have better small sample properties. Basically, both tests test whether including terms such as Y_{t-k}^2 helps improve the linear model's forecasting power. Both tests require the existence of finite fourth moments — see Tsay (1986), Theorem 1.

3.4 Lagrange Multiplier Tests

The tests presented above were derived without a specific alternative in mind. However, some linearity tests were constructed to have power against some specific departures from linearity. This class of tests, the Lagrange multiplier tests (score tests), are additionally useful because they require only estimation of the model under the null, that is, they demand no nonlinear estimation procedures when the null hypothesis is true.

Saikkonen and Luukkonen (1988) developed Lagrange multiplier tests for the null hypothesis of linearity against two particular alternatives, the bilinear model

$$y_t+\rho_1 y_{t-1}+\ldots+\rho_p y_{t-p}=U_t+\theta_1 U_{t-1}+\ldots+\theta_p U_{t-p}+\sum_{i=1}^{m}\sum_{j=1}^{k}\gamma_{ij}U_{t-j}y_{t-i}$$

and the exponential autoregressive model,

$$y_t+\rho_1 y_{t-1}+\ldots+\rho_p y_{t-p}+\{\exp(-\gamma y_{t-1}^2)-1\}\sum_{j=1}^{p}\theta_j y_{t-j}=U_t$$

The limiting chisquare distributions for the test statistics proposed in that paper are derived under the assumption that fourth and sixth moments are finite, respectively.

Luukkonen, Saikkonen and Teräsvirta (1988b) considered an LM test for another class of nonlinear models, the so-called smooth transition autoregressive model,

$$y_t+\sum_{i=1}^{p}\rho_i y_{t-i}+(\theta_0+\sum_{i=1}^{p}\theta_i y_{t-i})F(\gamma\sum_{i=1}^{p}\nu_i y_{t-i}-c)=U_t,$$

where $\gamma>0$, and $\nu_i=1$, if i=d, and zero otherwise. d is a generally unknown delay parameter, and F is a function to some smoothing conditions. To derive the limiting distribution of the test statistic, the authors assume the residual process $\{U_t\}$ has finite fourth moments.

It is worth noting that both Keenan's and Tsay's tests may be derived as Lagrange multiplier tests for some particular alternatives — see Saikkonen and Luukkonen, (1988, p. 59). Moreover, the McLeod and Li test is asymptotically equivalent to an LM test of linearity when the alternative is an ARCH model — see Luukkonen, Saikkonen and Teräsvirta (1988a).

3.5 The RESET Test

The Regression Error Specification Test (RESET) proposed by Ramsey (1969) — as well as the extension proposed by Thursby and Schmidt (1977) — is one of the most widely used tests of model misspecification.

Suppose, without loss of generality, that we are interested in testing that y_t is a *first-order autoregressive process*. The RESET procedure amounts to considering the augmented model $y_t=\rho y_{t-1}+\theta Z_t+U_t$ — where Z_t is a set of *auxiliary regressors* — and testing the hypothesis that $\theta=0$. As Z_t is usually unknown a priori, one must choose some test variables. Ramsey suggests using $Z_t=\{\hat{y}_t^2,\hat{y}_t^3,\ldots,\hat{y}_t^p\}$, where $\hat{y}_t$ are the OLS predicted values for y_t. Meanwhile Thursby and Schmidt consider powers of y_{t-1} itself and their simulation study gave evidence that their test variables enhance the power of Ramsey's test. Finally, the results also indicate that p=4 is usually a sensible choice for size and power considerations.

The procedures described above amount to running the auxiliary regression $y_t=\rho y_{t-1}+\theta Z_t+U_t$ for a given choice of Z_t. Meanwhile, the moment conditions required to derive the asymptotic distribution of the RESET-type tests depend upon the conditions required to establish the asymptotic normality of the estimators of that regression's parameters. Clearly, for $Z_t=\{y_t^2,y_t^3,\ldots,y_t^p\}$ at least 2p moments must be finite, as is the case with Ramsey's set of variables. In particular, choosing p=4 implies that finite eighth moments are required. This conclusion is supported by the fact that for p=2 the RESET test is asymptotically equivalent to Tsay's test. For this and some further considerations regarding this family of tests, see Granger and Teräsvirta (1992).

3.6 Neural Network Test

Using the theory of neural networks, Lee, White and Granger (1989) proposed a test of nonlinearity. First, it is important to stress that their concept of linearity is that of *linearity in mean with respect to the information set generated by* Z_t. In other words, their null hypothesis is characterized by

$$\Pr\{E[Y_t|Z_t]=Z_t'\theta\}=1 \text{ for some } \theta \in \mathbb{R}^k$$

Therefore, their concept of linearity lies within our definition of mds-linearity. Consequently, any measurable function of Z_t is uncorrelated with $U_t=Y_t-Z_t'\theta$. The neural network test utilizes a particular test function $h(.)$, usually a cumulative distribution function. Lee, White and Granger (1989) chose the logistic distribution. The moment conditions required to establish the test's asymptotic distribution can be inferred from the fact that Granger and Teräsvirta (1992) show that the test can be interpreted as based on a cubic Volterra expansion. Therefore, finite sixth moments should be required.

It should be noticed that the mds-linearity null hypothesis is defined as a conditional moment test, $E[U_t|\mathcal{F}_{t-1}]=0$. The law of iterated expectations implies that $E[U_t h(\mathcal{F}_{t-1})]=0$, for any function h depending on the past, provided that $E|U_t h(\mathcal{F}_{t-1})|<\infty$. That is, a conditional moment test implies that the process U_t must be uncorrelated with an infinite number of functions of the past. Therefore, it is clear that, in general, tests based on a finite number of unconditional moment conditions, as with the neural network tests, will not be consistent against every departure from the null. In this context, the tests proposed by Bierens (1990) assume particular relevance.

Bierens proposed as test function $\exp(s'\Phi(Z_t))$, where $s \in \mathbb{R}^k$ and Φ is an arbitrary Borel measurable bounded one-to-one mapping from $\mathbb{R}^k$ to $\mathbb{R}^k$. For our purposes, this test is attractive mainly because it allows us to construct a consistent test of linearity, based on a single moment condition. However, as Bierens presented results for a regression setup only, [2] we will not pursue its analysis in this paper. However, note that in his regression framework Bierens must impose only that second moments are finite.

3.7 Robinson's entropy-based test

Robinson (1991) proposes a consistent test of independence based on the nonparametric estimation of the Kullback-Leibler information criterion,

$$I(f,g)=\int f(x)\log\{f(x)/g(x)\}dx$$

To use this to test nonlinearity, we face the following two problems. First, it should be applied to the residuals of some linear model. However, while Robinson asserts it possible to extend his results to this situation, it is as yet unknown whether some distortion in the asymptotic behavior of that test statistic is in fact introduced when employing estimated residuals. Second, as Robinson (1991) notes, very strong assumptions are made to derive the test's asymptotic properties: the distribution of the errors must have a density function with compact support and bounded away from zero. As mentioned by the author, these requirements are far from corresponding to minimal sufficient conditions; in fact, they imply that all moments are finite and are extremely unlikely to be satisfied by real data. We do not know the degree to which these assumptions can be weakened.

3.8 Autocorrelation-Function Based Tests

This section concerns the testing of the so-called "random walk" behavior of stock returns. The (geometric) random walk hypothesis can be summarized as

$$R_t-R_{t-1}=\mu+U_t,$$

where $R_t \equiv \ln P_t$, P_t is the stock price at time t, and $\{U_t\}$ forms a martingale difference sequence (MDS).

The tests for market efficiency summarized below can all be expressed as functions of the autocorrelation function of R_t. The first such test is directly derived from the idea that the serial correlation of returns is a U-shaped function of the holding period. Richardson (1988) considers method of moments estimation of a testing approach first suggested by Fama and French (1988), and regresses multiperiod returns on lagged multiperiod returns. The test's large-sample distribution assumes that fourth moments are finite.

Another approach to this problem is known as the variance ratio test. Lo and MacKinlay (1988) derived this test's asymptotic distribution with finite $4+\delta$ ($\delta>0$) moments of the error term.

Finally, Durlauf (1991) constructed Cramér Von-Mises and Kolmogorov-Smirnov statistics from the autocorrelation function of the returns. Durlauf imposes the condition that the return process has finite eighth moments. It is worth noting that despite this strong requirement regarding moments, Durlauf's tests are consistent against all stationary departures from the null.

This review of some of the better known tests of temporal dependence and linearity indicates that these tests generally make strong assumptions about the existence of unconditional moments of the series under study. From our perspective, care must be taken when applying these tests to many economic time series. As already noted, there is some evidence that financial time series and exchange rates are examples of such series. Therefore, inferences are suspect when obtained by applying one of these tests to such types of data.

Chapter 2 introduces a family of tests that is robust to the nonexistence of moments. The null hypothesis of these tests will be that the data under test is iid-linear. Consequently, when applying those statistics to the test of the market-efficiency hypothesis one will too often reject the null of interest. Once again, this is a consequence of the fact that the iid-linear null is a stronger concept than the mds-linear one.

4. SIMULATIONS

In the previous section we discussed the moment conditions required to establish the asymptotic behavior of some popular nonlinearity tests. We saw that at least fourth moments are required to be finite in most cases. It is not unrealistic, then, to expect that the size and power of some of these tests will be significantly affected when applying the tests to data which do not possess finite fourth moments. Confronted with this fact, it is natural to investigate how the nonfulfillment of moment conditions affects the behavior of these tests. Ultimately, we aim to assess how misleading any inferences based on such tests can be when used for those economic time series that do not satisfy the moment requirements of the applicable central limit theorems.

For all tests it must be understood that the moment conditions stated in the previous section, while sufficient, are not necessary to establish their asymptotic properties. It is often difficult to determine minimal sufficient conditions. We therefore rely on a simulation study to determine the behavior of these tests when the associated moment conditions are not satisfied. This section provides some Monte Carlo experiments regarding the behavior of various nonlinearity tests under moment condition failure.

4.1 The McLeod-Li Test

We showed above that if $\{U_t\}$ has finite eighth moments, the McLeod-Li portmanteau test is asymptotically chisquared distributed. To assess the test's robustness when using data that violate this condition, we need to generate test data with infinite eighth moments. To this effect, we generated iid symmetric sequences $\{U_t\}$ from the Pareto family of distributions, which satisfy

$$\begin{cases} P(U>x)= \frac{1}{2}(x+1)^{-\alpha} & x>0 \\ P(U<-x)=\frac{1}{2}(x+1)^{-\alpha} & x>0 \end{cases}$$

for different values of α, the maximal moment exponent of U.

The simulations described below were constructed in order to better understand the consequences of using the asymptotic chisquared distribution as an approximation to the actual distribution of the test even when the generated data do not satisfy the moment conditions required by the applicable central limit theorem. It should be noted that when the data is generated from some specific family of distributions it is still possible to establish the asymptotic behavior of the McLeod-Li test when fourth moments, for example, are not finite.

To this purpose, assume for the moment that we observe data from a symmetric stable distribution (c.f. Ibramigov and Linnik (1971)) with maximal moment exponent α $(0\leq\alpha\leq2)$.[3] We know from the theory of sample averages for linear processes with infinite variance that for an iid sequence $\{U_t\}$ generated from this family of distributions, the asymptotic distribution of the sample autocorrelations $\hat{\rho}(h)$, h=1,2,..., is of the form

$$(T/\ln(T))^{1/\alpha}(\hat{\rho}(h)-\rho(h)) \Longrightarrow W/V$$

where W and V are independent stable distributions with characteristic exponents α and $\alpha/2$, respectively — see Brockwell and Davis, (1987, pp. 482-484). It is important to note that not only we do not have convergence to a normal distribution but also that the rate of convergence to the limiting random variable is $O_p([T/\ln(T)]^{-1/\alpha})$, faster than the usual $O_p(T^{-1/2})$. As $O_p([T/\ln(T)]^{-1/\alpha})=o_p(T^{-1/\beta})$, $\beta>\alpha$ (and $\alpha\leq2$), it follows that $T^{1/2}(\hat{\rho}(h)-\rho(h))$ converges to zero in probability, except when $\alpha=2$. (Note that when $\alpha=2$, the corresponding stable distribution is a normal distribution)

The McLeod-Li test deals with sample autocorrelations for the squared process $\{U_t^2\}$. Now, if U_t has Paretian tails with $\alpha<4$, U_t^2 lies in the normal domain of attraction of a stable distribution with stable exponent $\alpha/2$. Therefore, results available for the sample autocorrelations are also applicable to the sample autocorrelations of the squares.

We considered several values for α as well as $U_t \sim iidN(0,1)$ (that is, $\alpha=\infty$). and consider this last case for comparison purposes. Loretan (1991) describes the algorithm used to generate the symmetric iid sequences. See also the references provided there. We fixed T=1000 as our sample size and considered 1000 replications of each experiment.

Table 1.1 provides evidence that the non-existence of moments considerably affects the size of the McLeod-Li statistic. For both choices of the number of autocorrelations considered (q=20 and q=40), this relationship is revealed in two ways. We note the difference between the nominal sizes of a chisquare distribution with 20 and 40 degrees of freedom) and also the fact that the Kolmogorov-Smirnov (KS) statistics strongly reject the null that the empirical distributions are chisquare.[4] On the other hand, when the simulated data are normally distributed ($\alpha=\infty$) and q=20 , the KS statistic clearly indicates that the null of chi-square distribution cannot be rejected (the p-value is approximately 0.49). Somewhat surprisingly, according to our experiments, when q=40 not even the normal data lead one to accept that the distribution of McLeod-Li test is chisquared. It is important to bear in mind that even though we are dealing with asymptotic

Table 1.1: Size experiments on the McLeod-Li test

	q=20						
(1)	α=1.5	α=2	α=3	α=4	α=6	α=8	α=∞
0.100	0.073	0.087	0.102	0.116	0.120	0.119	0.099
0.050	0.061	0.078	0.083	0.100	0.095	0.089	0.047
0.025	0.056	0.074	0.071	0.092	0.072	0.070	0.031
0.010	0.049	0.064	0.064	0.079	0.052	0.056	0.011
0.005	0.047	0.058	0.057	0.071	0.045	0.044	0.004
MEAN	7.951	9.958	12.18	15.11	15.91	17.38	20.20
VAR	602.4	547.2	407.2	409.3	194.7	149.9	41.69
KS	25.95	23.90	19.98	16.51	12.82	9.454	0.827

(1) Rows 1 through 5 give the empirical sizes of tests whose nominal size is given by column (1), under the assumption that the test statistic is asymptotically chi-squared distributed with q degrees of freedom. The last three rows report the values of the sample mean (MEAN), the sample variance (VAR), and the value of the Kolmogorov-Smirnov (KS) statistic for the null hypothesis that the empirical distribution is $\chi^2_{(q)}$. Sample size: 1000

Table 1.1a: Size experiments on the McLeod-Li test with q=40

	q=40		
(1)	α=3	α=6	α=∞
0.100	0.119	0.146	0.129
0.050	0.104	0.118	0.075
0.025	0.093	0.089	0.041
0.010	0.081	0.070	0.020
0.005	0.077	0.059	0.010
MEAN	26.92	33.91	41.53
VAR	1047.	445.7	86.61
KS	19.15	11.60	2.501

(1) See Table 1.1. T=1000

approximations our sample size is 1000. We may expect that for samples of this size the asymptotic distribution would give a close approximation to the tests' actual distribution. Note that the error in the size estimates due to simulation has mean 0 and standard error given by $\sqrt{p(1-p)/R}$, where p is nominal size and R is number of replications. Table 1.2 reports the values of these standard errors.

Table 1.2: Standard Errors for Size Estimates

Nominal Sizes	0.100	0.050	0.025	0.010	0.005
S.E.	0.0095	0.0069	0.0049	0.0031	0.0022

S.E. reports the standard errors associated with each estimate of the tests size, as given by $\sqrt{p(1-p)/R}$ where p is nominal size and R is number of replications.

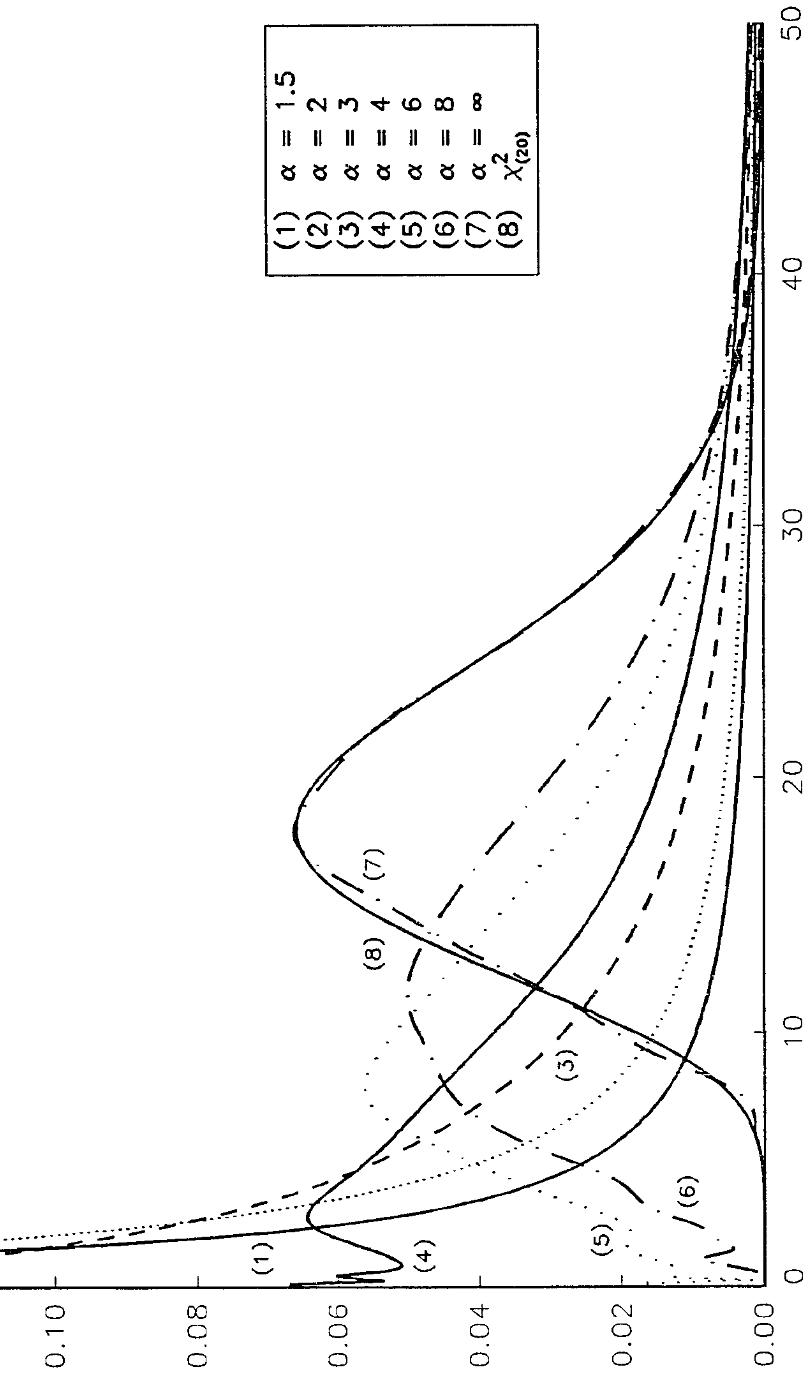
A reassuring pattern of the results is that as α increases, the empirical distribution of the tests approaches a chisquare distribution. This is evident because the means of the tests increase towards their theoretical values (20 and 40) and the dispersion of the tests becomes substantially smaller. At the same time, the KS statistic seems to be a decreasing function of α.

We can obtain a complementary perspective regarding these facts by plotting the estimated densities of the tests' statistics for the different values of α. Density estimates for $\ln(Q_i)$ — where Q is the value of the McLeod-Li statistic — were formed using the nonparametric kernel estimator, $\hat{f}_{h_n}(u)=T^{-1}\sum_{i=1}^{T} h_n^{-1} K((u-\ln(Q_i))/h_n)$. We used the Gaussian kernel $K(x)=\exp(-x^2/2)/\sqrt{2\pi}$. The bandwidth selection was $h_n=\hat{\sigma}\ T^{-1/5}$, where $\hat{\sigma}$ is the standard deviation of Q. We used the logarithmic transformation for U_t because for small values of α many observations are near zero. The transformed variable has support on $(-\infty,+\infty)$. In this way we avoid boundary problems - see Silverman (1986, Section 2.10). We obtained densities estimates for the original data by a change of variable.

It is clear from Figure 1.1 that for small values of α a strong concentration of the densities occurs around zero. This is particularly evident for values of α less than four. The analysis of the densities' plots suggests that for these values of α, the density of the tests has a pole at zero. This seems to have a simple explanation, especially if we put this observation together with the analysis, (provided previously in this chapter) of the asymptotic behavior of the McLeod-Li test with data generated by distributions that are in the normal domain of attraction of a stable distribution. For $\alpha<4$, the squared series U_t^2 are in the domain of attraction of the stable distributions. Using the analysis on rates of convergence for the sample autocorrelations described on pp. 21-22, and the fact that the McLeod-Li is a function of sample autocorrelations of the squared series, it seems clear that the McLeod-Li statistic is $o_p(1)$ for values of α less than four. The pole at the origin suggested by Figure 1.1 confirms this argument.

At the same time, Figure 1.1 also displays the occurrence of too many large values (by comparison with the corresponding chisquare

Figure 1.1: Estimated Densities of the McLeod-Li Test (q=20,T=1000)



distribution) for the test statistic. We note that this last observation casts some doubt on whether it is appropriate to use the chisquared distribution to approximate the upper tail of the empirical distribution. Using the chisquare critical values when moment conditions are not satisfied, would lead us to reject the null hypothesis too often, at least at the conventional 5% and 1% levels. We also note the slowly decaying upper tails of the tests' empirical distributions.

Although a sample size of 1000 seems large enough to guarantee that asymptotic results provide a good approximation, we decided to run the same set of experiments with T=2000. The results are summarized on Table 1.3 and do not show any significant departures from the results obtained with T=1000 — see also Table 1.3a in the Appendix to this chapter for the case q=40.

Table 1.3: Size experiments on the McLeod-Li test

	q=20						
(1)	$\alpha=1.5$	$\alpha=2$	$\alpha=3$	$\alpha=4$	$\alpha=6$	$\alpha=8$	$\alpha=\infty$
0.100	0.049	0.068	0.087	0.111	0.127	0.123	0.088
0.050	0.044	0.063	0.073	0.094	0.102	0.096	0.048
0.025	0.044	0.054	0.062	0.083	0.091	0.077	0.024
0.010	0.039	0.051	0.054	0.077	0.071	0.059	0.012
0.005	0.039	0.049	0.050	0.072	0.060	0.053	0.006
MEAN	7.025	8.530	12.04	14.62	16.68	17.74	19.87
VAR	1002.	658.7	609.8	379.7	212.0	197.6	40.29
KS	27.62	25.57	20.80	17.17	12.23	9.797	0.647

(1) - See Table 1.1. Sample Size: 2000

The fact that the empirical sizes of the tests are, in most cases, above their nominal sizes deserves some further comments. As Section 3.4 states, the McLeod-Li test is asymptotically equivalent to the Lagrange multiplier (TR^2 test) test of iid against ARCH residuals. The simulation results presented in Tables 1.1 and 1.3 clearly indicate that, for series such that $2 \le \alpha \le 8$, the McLeod-Li test is an example of a "liberal" test, because the null hypothesis is rejected too often. In other words, the alternative that the errors are ARCH will be accepted too often. This simple fact implies that when testing for ARCH disturbances in applied work one may reject the null of iid not because the errors follow an ARCH process, but simply because the iid driving process is characterized by nonexistence of finite low-order moments.

However, this conclusion may depend upon the sample size. In order to assess the impact of moment condition failure on the behavior of the test at different sample sizes, we run the same set of experiments for sample sizes 50, 100, 200, 500, 750, 1000 and 2000. Figure 1.2 plots the empirical size of the 1% McLeod-Li test for different values of α. As expected, with normal data the empirical size is always very close to the nominal size. However, for all other values of α a pattern emerges. When the sample is small (50, 100) the empirical size of the tests is below the nominal size, with the opposite happening when the sample size increases.

Figure 1.3 displays the results for the same experiment relative to the 5% size test. Although the same pattern seems to emerge, we note that for large sample sizes the difference between the nominal and empirical sizes is smaller than such differences for the 1%

Figure 1.2: Empirical Size of the 1% McLeod–Li Test (q=20)

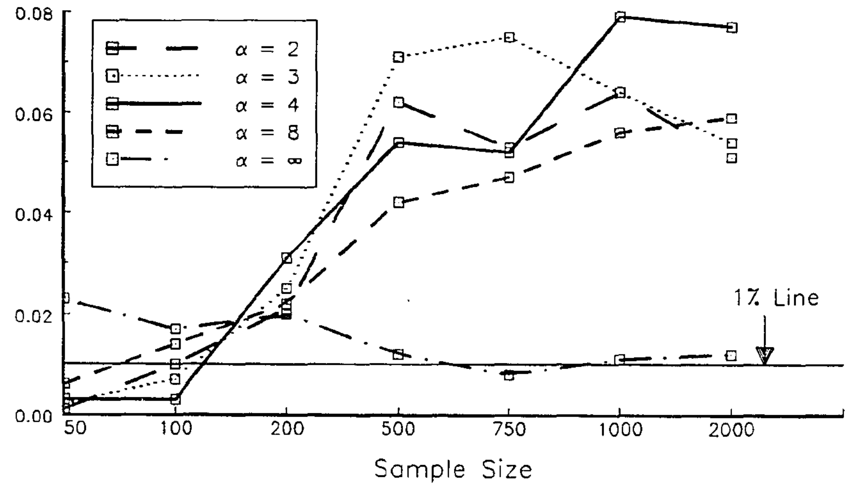
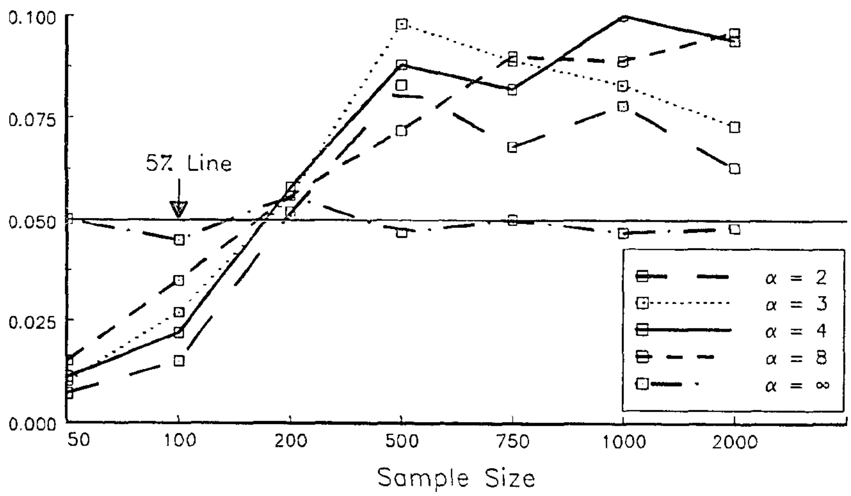


Figure 1.3: Empirical Size of the 5% McLeod–Li Test (q=20)



size test. In this last case, for values of α other than ∞ and sample sizes above or equal to 500 observations we find between four and eight times more rejections of the null of iid than we should expect.

4.1.1 Robustifying the McLeod-Li test by means of a transformation of the data

One possible way of making a test robust against moment condition failure is by "trimming" the data, that is, by transforming the data such that:

$$\tilde{U}_t = \begin{cases} U_t & \text{if } |U_t| \leq K \\ \text{sgn}(U_t)K & \text{if } |U_t| > K \end{cases}$$

where sgn(x)=1 if x>0, -1 if x<0 and 0 otherwise, and K is a given constant. In practice, one may set K according to some data-based rule. In what follows, K is an estimate of the inter-quartile range, $\xi_{3/4}-\xi_{1/4}$, where $\xi_p=\inf\{x:F(x)\geq p\}$ is the *p*th quantile of the distribution function F of U_t.

The idea behind this transformation is that if two variables are independent then any transformation of the two variables still produces two independent variables. In particular, $\tilde{U}_t$ is generated by a distribution with finite support, and consequently has finite moments of every order. Therefore, if the null is iid, this transformation seems like an easy fix for the problem of moment condition failure.

Table 1.4: Size experiments on the McLeod-Li test with transformed data

(1)	$\alpha=1.5$	$\alpha=2$	$\alpha=3$	$\alpha=4$	$\alpha=8$	$\alpha=\infty$
0.100	0.099	0.114	0.105	0.109	0.102	0.110
0.050	0.053	0.054	0.052	0.059	0.047	0.054
0.025	0.031	0.034	0.030	0.028	0.027	0.022
0.010	0.014	0.015	0.008	0.020	0.012	0.012
0.005	0.007	0.006	0.005	0.012	0.007	0.007
MEAN	19.9	20.1	20.2	20.1	19.8	19.8
VAR	41.5	44.4	41.6	44.3	41.2	41.4
KS	0.747	0.914	0.757	0.692	1.18	0.920

(1) See notes for Table 1.1

As a matter of fact, the simulations reported in Table 1.4 indicate that by applying the McLeod-Li test to $\tilde{U}_t$ instead of U_t one might expect to obtain the right sizes. However, the gain in size properties comes at the cost of reduced power against non-iid alternatives. The simple experiments reported in Table 1.5 indicate that the test's power properties are considerably weakened by this transformation. This effect is particularly strong for small values of the ARCH(1) parameter ϕ (see Table 1.5), where the test with the transformed data has problems distinguishing that series from an iid series. This is not surprising, because by cutting off some of the data's variation we loose significant information about the series when the alternative hypothesis is true. Indeed, we see no reasons to believe that this transformation guarantees that the corresponding LM [5] test will keep its asymptotically optimal properties.

Table 1.5: Power of the McLeod-Li test

	ϕ=.2		ϕ=.4		ϕ=.6	
	(1)	(2)	(1)	(2)	(1)	(2)
T=200						
1% test	0.084	0.176	0.260	0.510	0.576	0.717
5% test	0.186	0.290	0.427	0.623	0.730	0.813
T=500						
1% test	0.135	0.446	0.700	0.935	0.976	0.984
5% test	0.281	0.583	0.830	0.972	0.989	0.992
T=1000						
1% test	0.341	0.825	0.972	1.00	1.00	0.999
5% test	0.546	0.924	0.991	1.00	1.00	1.00

The data was generated by an ARCH(1) model $U_t=h_tZ_t$, where $h_t^2=\omega+\phi U_{t-1}^2$ and $Z_t\sim$ iidN(0,1). Column (1) corresponds to the values of the test for the transformed data $\tilde{U}_t$; (2) reports the values of the test for the series U_t

Furthermore, this test is usually applied to estimated residuals. Suppose now that the null of iid-linearity is true (the innovation process forms an iid sequence). In this case, the estimation process implies that the estimated residuals do not form an iid sequence, even though it should be expected that as the sample size increases the behavior of the estimated residuals' behavior more closely resembles that for an iid series. However, no theoretical statistical results likely exist to guarantee that applying the trimming transformation to estimated residuals will not affect the limiting distribution of the McLeod-Li test.

4.2 The RESET test

Section 3.5 showed that this test for misspecification requires the existence of 2p finite moments, where p is the order of the polynomial approximation to the unknown regression function. Here we consider the following experiment: we generate stationary AR(1) processes, $y_t=\rho y_{t-1}+U_t$, $|\rho|<1$, where U_t is an iid process whose distribution is of the Pareto-Levy form, for the following set of values for α: $\alpha = \{1.5,2,3,4,8,\infty\}$. For each of these six possible values of α, the autoregressive coefficient takes three different values, 0.05, 0.5 and 0.95. We considered two possible test variables. First, as Ramsey (1969) suggested, we consider the OLS-predicted values — $\hat{y}_t$ — of the regression of y_t on y_{t-1}. We run the auxiliary regression,

$$\hat{U}_t = \beta_1 y_{t-1} + \sum_{i=2}^{p} \beta_i \hat{y}_t^i + v_t$$

for two different values of p, 2 and 4. ($\hat{U}_t$ are the residuals of the OLS regression of y_t on y_{t-1} and v_t is an error term) The resulting tests are designated RAMSEY2 and RAMSEY4, respectively.

As a second choice test variable we follow Thursby and Schmidt (1977) and considered powers of the lagged endogenous variable y_{t-1}. The auxiliary regression becomes

$$\hat{U}_t = \beta_1 y_{t-1} + \sum_{i=2}^{p} \beta_i y_{t-1}^{i} + v_t$$

Again, we consider p=2 and p=4, and the corresponding test statistics are denominated THURSBY2 and THURSBY4.

Tables 1.6 and 1.7 present the results for RAMSEY2 and RAMSEY4, respectively, when the data generating process (DGP) is $y_t = .9y_{t-1} + U_t$. We obtain the test statistic by computing

$$\text{RAMSEYp} = \frac{(\sum \hat{U}_t^2 - \sum \hat{v}_t^2)/(p-1)}{\sum \hat{v}_t^2 \,/(T-1-p)}$$

Under the null of linearity, (p-1)RAMSEYp ~ $\chi^2_{(p-1)}$, provided that the moment conditions are satisfied.

The results provide some evidence that moment conditions are important for the size of both statistics. Consider first the case of p=2. For the simulations with $\alpha=8$ and $\alpha=\infty$, the respective empirical distributions display strong similarities with a chisquare distribution with one degree of freedom. This similarity is well summarized by the fact that the KS statistics do not reject the null that both tests follow such a distribution (p-values for that test are approximately 0.1 and 0.81). However, for all other values of α, including $\alpha=4$, the information in Table 1.6 seems to indicate that the asymptotic approximation provided by a $\chi^2_{(1)}$ variable is rather poor.

We find the same poor showing for the RAMSEY4 test. In this case, however, even when the innovation process has finite eighth moments, the limiting chisquare distribution does not satisfactorily approximate the test's sampling distribution. As in the McLeod-Li test, we find the dispersion of the results is smaller as α increases, while the KS statistic decreases as we approach the "critical" moment condition, 2p, from below.

In the Appendix, we include the results for the simulations when the autoregressive parameter is .5 and .05. The pattern of the results seems independent of the choice for ρ.

Table 1.6: Size experiments on RAMSEY2

DGP: $y_t = .9y_{t-1} + U_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.053	0.056	0.082	0.081	0.088	0.117
0.050	0.033	0.039	0.048	0.047	0.044	0.058
0.025	0.024	0.028	0.028	0.025	0.024	0.031
0.010	0.020	0.018	0.022	0.018	0.006	0.016
0.005	0.017	0.017	0.016	0.012	0.001	0.009
MEAN	0.821	0.737	0.896	0.935	0.923	1.066
VAR	24.28	5.211	4.421	3.307	1.661	2.621
KS	10.02	7.120	4.223	2.606	1.232	0.646

(1) - See comment to Table 1.1

Tables 1.8 and 1.9 describe results obtained for THURSBY2 and THURSBY4 from the same set of experiments. As before, fourth and eighth moments appear critical in determining the asymptotic properties of these tests. Indeed, the pattern of results is identical to those observed for the RAMSEY2 and RAMSEY4 tests.

Table 1.7: Size experiments on RAMSEY4

DGP: $y_t = .9y_{t-1} + U_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.068	0.070	0.097	0.090	0.094	0.100
0.050	0.055	0.048	0.069	0.054	0.048	0.061
0.025	0.047	0.036	0.046	0.038	0.027	0.036
0.010	0.040	0.027	0.029	0.032	0.017	0.014
0.005	0.034	0.025	0.023	0.026	0.012	0.011
MEAN	2.603	2.427	3.139	2.859	2.789	3.043
VAR	106.0	56.78	188.5	19.07	6.978	6.952
KS	14.40	11.01	7.248	5.093	2.636	0.539

(1) See notes for Table 1.1

Table 1.8: Size experiments on THURSBY2

DGP: $y_t = .9y_{t-1} + U_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.053	0.056	0.082	0.081	0.088	0.117
0.050	0.033	0.039	0.048	0.047	0.044	0.058
0.025	0.024	0.028	0.028	0.025	0.024	0.031
0.010	0.020	0.018	0.022	0.018	0.006	0.016
0.005	0.017	0.017	0.016	0.012	0.001	0.009
MEAN	0.821	0.737	0.896	0.935	0.923	1.066
VAR	24.28	5.211	4.421	3.307	1.661	2.621
KS	10.02	7.120	4.223	2.606	1.232	0.647

(1) See notes for Table 1.1

Table 1.9: Size experiments on THURSBY4

DGP: $y_t = .9y_{t-1} + U_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.068	0.070	0.097	0.090	0.094	0.100
0.050	0.055	0.048	0.069	0.054	0.048	0.061
0.025	0.047	0.036	0.046	0.038	0.027	0.036
0.010	0.040	0.027	0.029	0.032	0.017	0.014
0.005	0.034	0.025	0.023	0.026	0.012	0.011
MEAN	2.603	2.427	3.139	2.859	2.789	3.043
VAR	106.0	56.78	188.5	19.07	6.978	6.952
KS	14.40	11.01	7.248	5.093	2.636	0.539

(1) See notes for Table 1.1

5. SUMMARY

In this chapter we introduced this dissertation's topic, namely, testing nonlinearities under moment condition failure. We first discussed two related linearity concepts. It is our perspective that nonlinearities have been usually detected in those economic time series that are suspect to satisfying only a few moment conditions; i.e., in those series with maximal moment exponents most likely less than four.

We discussed the type of moment conditions required by some of the most popular nonlinearity tests and found they all require at least finite fourth moments. We provided a simulation study of the performance of the McLeod-Li and RESET tests to assess the impact on the inferences made by using these tests when moment condition failure is a problem. This study further suggests extreme care is needed when these tests are applied to data generated by heavy-tailed distributions.

One comment about the distinction between heavy-tailed distributions and moment condition failure is in order. Heavy tails do not imply moment condition failure *per se.* Indeed, one can can always argue that economic data is generated by distributions with finite support, and that most of the distributional models used in applied research — for example, the normal distribution — are just convenient approximations to the true underlying models. In this context, we may imagine models with heavy-tailed distributions but finite support.

Consequently, moments of every order would be finite.

If one believes that this is the true generating mechanism for economic time series (where heavy-tailed distributions do not coexist with moment condition failure), our concern with moment condition failure would seem misplaced. However, even if it is the case, it should be noticed that in our simulation study we have used pseudo-random numbers originated by a machine that obviously works with finite supports. That is, our randomly generated numbers all have moments of every order. From this perspective, we have essentially dealt with heavy-tailed distributions more than with moment condition failure. Therefore, it seems our analysis is more general than we first supposed.

Finally, it must be noted that we used Paretian tails type distributions to generate data characterized by moment condition failure. Our simulation findings that the McLeod-Li and the RESET tests are non-robust to the non-existence of moments may be closely related to the distributional model used. Even though we could have investigated the behavior of the tests under other types of heavy-tailed distributions, we confined our analysis to the Pareto type distributions. Such distributions provide a simple characterization of the moment condition failure and also seem to provide an adequate description of outlier activity present in economic data — See Loretan and Phillips (1992).

APPENDIX OF CHAPTER I: *Simulation Results*

Table 1.3b: Size experiments on the McLeod-Li test with q=40

	q=40		
(1)	α=3	α=6	α=∞
0.100	0.141	0.131	0.104
0.050	0.124	0.104	0.049
0.025	0.115	0.082	0.030
0.010	0.103	0.069	0.013
0.005	0.093	0.060	0.008
MEAN	26.55	34.44	39.85
VAR	1152.	657.9	83.54
KS	20.33	12.26	0.846

(1) See Table 1.3. T=2000

Table 1.10: Size experiments on RAMSEY2

DGP: $y_t = .05y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.035	0.065	0.065	0.085	0.096	0.109
0.050	0.027	0.046	0.041	0.048	0.048	0.051
0.025	0.020	0.033	0.022	0.027	0.028	0.025
0.010	0.014	0.021	0.015	0.011	0.017	0.013
0.005	0.009	0.014	0.011	0.006	0.012	0.009
MEAN	0.611	0.805	0.862	0.880	1.004	1.037
VAR	10.18	6.323	3.762	3.824	3.422	2.143
KS	9.799	6.491	3.739	3.190	1.252	1.174

(1) - The first five rows of this table give the empirical sizes of tests whose nominal size is given by column (1), under the assumption that the test statistic is asymptotically chi-squared distributed. Sample size: 1000

Table 1.11: Size experiments on RAMSEY4

DGP: $y_t = .05y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.052	0.075	0.081	0.075	0.089	0.101
0.050	0.044	0.055	0.055	0.048	0.054	0.053
0.025	0.035	0.042	0.038	0.032	0.032	0.022
0.010	0.030	0.034	0.028	0.021	0.018	0.012
0.005	0.024	0.031	0.025	0.016	0.012	0.003
MEAN	2.178	2.894	2.703	2.583	2.843	2.949
VAR	106.6	222.3	23.39	19.92	14.37	5.880
KS	14.75	10.14	6.278	5.966	3.063	0.732

(1) See Table 1.10

Table 1.12: Size experiments on RAMSEY2

DGP: $y_t = .5y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.058	0.058	0.065	0.068	0.073	0.074
0.050	0.032	0.031	0.027	0.029	0.029	0.031
0.025	0.021	0.017	0.018	0.017	0.012	0.013
0.010	0.010	0.013	0.007	0.007	0.005	0.007
0.005	0.007	0.011	0.005	0.002	0.004	0.006
MEAN	0.710	0.821	0.798	0.845	0.845	0.803
VAR	3.336	6.669	2.365	1.471	1.357	1.458
KS	5.816	4.591	2.963	1.594	1.525	2.627

(1) See Table 1.10

Table 1.13: Size experiments on RAMSEY4

DGP: $y_t = .5y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.069	0.066	0.080	0.113	0.117	0.099
0.025	0.044	0.044	0.057	0.073	0.070	0.053
0.050	0.036	0.037	0.040	0.048	0.044	0.031
0.010	0.025	0.026	0.026	0.030	0.027	0.010
0.005	0.019	0.022	0.017	0.018	0.024	0.006
MEAN	2.284	3.200	2.962	3.143	3.239	2.959
VAR	24.98	149.2	42.72	21.12	11.48	6.970
KS	9.502	7.301	4.165	1.870	0.9056	1.435

(1) See Table 1.10

Table 1.14: Size experiments on THURSBY4

DGP: $y_t = .05y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.035	0.065	0.065	0.085	0.096	0.109
0.050	0.027	0.046	0.041	0.048	0.048	0.051
0.025	0.020	0.033	0.022	0.027	0.028	0.025
0.010	0.014	0.021	0.015	0.011	0.017	0.013
0.005	0.009	0.014	0.011	0.006	0.012	0.009
MEAN	0.6110	0.8051	0.8623	0.8804	1.004	1.037
VAR	10.18	6.323	3.762	3.824	3.422	2.143
KS	9.799	6.491	3.739	3.190	1.252	1.174

(1) See Table 1.10

Table 1.15: Size experiments on THURSBY4

DGP: $y_t = .05y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.052	0.075	0.081	0.075	0.089	0.101
0.050	0.044	0.055	0.055	0.048	0.054	0.053
0.025	0.035	0.042	0.038	0.032	0.032	0.022
0.010	0.030	0.034	0.028	0.021	0.018	0.012
0.005	0.024	0.031	0.025	0.016	0.012	0.003
MEAN	2.178	2.894	2.703	2.583	2.843	2.949
VAR	106.6	222.3	23.39	19.92	14.37	5.880
KS	14.75	10.14	6.278	5.966	3.063	0.732

(1) See Table 1.10

Table 1.16: Size experiments on THURSBY2

DGP: $y_t = .5y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.058	0.058	0.065	0.068	0.073	0.074
0.050	0.032	0.031	0.027	0.029	0.029	0.031
0.025	0.021	0.017	0.018	0.017	0.012	0.013
0.010	0.010	0.013	0.007	0.007	0.005	0.007
0.005	0.007	0.011	0.005	0.002	0.004	0.006
MEAN	0.710	0.821	0.798	0.845	0.845	0.809
VAR	3.336	6.669	2.365	1.471	1.357	1.458
KS	5.816	4.591	2.963	1.594	1.525	2.627

(1) See Table 1.10

Table 1.17: Size experiments on THURSBY4

DGP: $y_t = .5y_{t-1} + \varepsilon_t$

(1)	α=1.5	α=2	α=3	α=4	α=8	α=∞
0.100	0.069	0.066	0.080	0.113	0.117	0.099
0.050	0.044	0.044	0.057	0.073	0.070	0.053
0.025	0.036	0.037	0.040	0.048	0.044	0.031
0.010	0.025	0.026	0.026	0.030	0.027	0.010
0.005	0.019	0.022	0.017	0.018	0.024	0.006
MEAN	2.284	3.200	2.962	3.143	3.239	2.959
VAR	24.98	149.2	42.72	21.12	11.48	6.970
KS	9.502	7.301	4.165	1.870	0.906	1.435

(1) See Table 1.10

Notes to Chapter 1.

[1] Strictly speaking, the definition of mds-linearity is not enough to guarantee that $E[U_t U_s]=0$, as $E[|U_t U_s|]$ might not be finite. However, the result goes through if we impose the condition that the second moment of U_t is finite, for every t.

[2] It seems that the results presented in that paper could be extended to a more general framework — Bierens (1990), p. 1444.

[3] We are using the same notation for the maximal moment exponent of a stable distributions and for the maximal moment exponent of the Paretian distribution used to generate data for our simulation study. This seemingly misleading notation is widely use in related literature.

[4] The 10%, 5% and 1% critical values for the Kolmogorov-Smirnov test are, respectively, 1.22, 1.36 and 1.64.

[5] The McLeod-Li test is asymptotically equivalent to the LM test of iid against the alternative hypothesis of existence of ARCH effects. See Section 3.4.

CHAPTER 2: A FAMILY OF TESTS FOR THE HYPOTHESIS OF IID TIME SERIES

1. INTRODUCTION

The presence of nonlinear dependence in economic time series seems to be gaining widespread acceptance among economists, as is indicated by the increasing number of published works in this area. It is not difficult to determine why this acceptance is emerging. For example, if one considers forecasting, the ability to detect and determine the nature of the nonlinear dependence is an important step towards constructing more accurate forecasts.

Most work in this area of economics is essentially theoretical. One of the reasons for this state of affairs is that the detection of nonlinearities is not yet a completely answered question. In recent years, the economics and statistics literature have proposed numerous tests of nonlinearity. A particularly attractive approach, the so-called "BDS statistic", was introduced by Brock, Dechert and Scheinkman (1987). These authors developed statistical theory for the "correlation integral" of Grassberger/Procaccia/Takens (a measure of spatial nonlinear correlation) and used this theory to formulate a test of the null hypothesis of no dependence against a large class of alternatives. Brock, Hsieh and LeBaron (1991) provide a comprehensive review of the BDS test. It is worth noting that this test has good size and power characteristics in comparison with some other popular tests for independence.

The BDS test is among the few nonlinearity tests that cannot be interpreted as a score or Lagrange Multiplier (LM) test against a nonlinear alternative — c.f. Granger and Teräsvirta (1992). This simple fact suggests both advantages and disadvantages to using this test. On one hand, the BDS test is likely to have lower power against a particular alternative hypothesis than does the corresponding LM test, which is asymptotically optimal against that alternative (cf. Engle (1984)). On the other hand, given the large number of nonlinear time series models available to researchers, it is sometimes questionable why a particular alternative may be more interesting than others. Indeed, it seems to be current practice to choose one particular model over the others more on grounds of tractability than because of any theoretically based arguments for its use. Therefore, in nonlinear time series it seems highly desirable to have tests with good statistical performances across a large set of different alternatives. Moreover, as no statistically optimal properties are associated with the BDS test, it is reasonable to think that some transformations of this test might enhance its properties.

In this chapter we develop a family of tests for the iid hypothesis. These tests are based on simple functionals of the BDS statistic and, therefore, on the correlation integral. We stress that for non-Gaussian processes the absence of serial correlation does not imply independence. Thus, a researcher who merely looks at conventional testing tools — such as the sample autocorrelation function of the series — might mistakenly fail to reject the null of iid when in fact the data generating process is nonlinear and dependent.

The proposed tests parallel a test for independence defined by Blum, Kiefer, and Rosenblatt (1961) (BKR hereafter). The BKR test is a natural Cramér-Von Mises function test based on the empirical distribution function. In its simplest form, the BKR test can be represented as $B_T=\int_0^1 N_T(x)^2 dF_T(x)$, where $F_T(x)$ is the empirical distribution function of a sample of size T from the bivariate process $\{X_t\}=\{x_{1t},x_{2t}\}$ and $N_T(x)$ is the difference between the joint empirical distribution function and the product of the marginal empirical distribution functions of $\{X_t\}_{t=1}^{T}$. The BKR test appears to be a very tough competitor against a multivariate form of the BDS test statistic — see Baek (1988). Moreover, the Cramér-Von Mises-type tests are known to have good sampling performances for different kinds of null hypotheses. Durlauf's (1991) recent work on testing the martingale hypothesis successfully applies of these types of tests in economics. So it seems natural to investigate the large sample behavior of a test closely related to B_T, but where $N_T(x)$ is a functional of the BDS statistic and, therefore, of the sample correlation integral.

The BDS test is currently applied as a diagnostic test of the correctness of an empirical model specification. This is a consequence of the fact that no distortion on the asymptotic distribution of the test is introduced when applying the test to the estimated residuals of a wide class of models. In particular, the BDS test can be used as a linearity test. However, the theorem that establishes the insensitivity of the BDS test to the use of estimated residuals in place of the true innovations, up to now, has only been proven on the basis of a smooth approximation to the indicator kernel, $\chi(x)$, the building block of the BDS statistic. — see Brock and Dechert (1988) and Brock, Hsieh and LeBaron (1991).Therefore, the result of insensitivity to the use of estimated residuals is only valid for the approximation kernels, while no formal proof has been given for the indicator kernel. In the next chapter, we present such a proof for the BDS statistic itself and its extensions to the family of tests introduced here.

The BDS statistic does have some associated drawbacks. First, some known departures from iid exist for which the BDS statistic has no power. Dechert (1988) gives one such example. Therefore, the BDS test is not consistent against all alternatives. Consequently, the more general class of tests described in this chapter will also fail to provide a globally consistent test of iid. Furthermore, while the good size and power properties of the BDS statistic generally hold only in large samples, some related tests are less demanding regarding sample size.

However, it is also the purpose of this paper to present one situation where the BDS family of tests seems to have an undisputed advantage over rival tests: this family of tests requires only very weak moment conditions in order to establish its asymptotic properties.

In the previous chapter, we presented some known results suggesting that most of the supposed nonlinear economic time series are also characterized by heavy-tailed distributions. Some recent studies — cf. Loretan (1991) and Loretan and Phillips (1992) — have found that exchange rate changes and stock market returns do not have finite fourth moments (although they seem to have finite variance). The brief survey of nonlinearity tests presented in Chapter 1 concluded that these tests are poorly suited for testing nonlinearities in data that suffer from moment condition failure.

As will be discussed later in this dissertation, the BDS test and the family of tests we describe in this paper are a natural testing device for time series where higher order moments cannot be assumed to exist. In particular, these tests may be useful when testing for market efficiency.

The family of nonlinearity tests that we introduce in this dissertation are presented in two steps. In this chapter, we present a family of tests for the null hypothesis that the data comes from an iid-process. In Chapter 3 we discuss how the test can be applied as a test for the null of iid-linearity. The remainder of this chapter is organized as follows. Section 2 presents some preliminary theory concerning the BDS test and a functional central limit theorem for U-statistics of absolutely regular processes due to Denker and Keller (1983). Section 3 develops our family of tests and discusses how to avoid the selection of two "nuisance" parameters in the BDS test. Finally, conclusions are presented. All the proofs to the results we present are found in the Appendix to this chapter.

2. PRELIMINARIES

We start by describing the BDS test. The test is based on the fact that, if $\{X_t\}$ is an iid process, then $\mathcal{C}_{\varepsilon,m} = \mathcal{C}^m_{\varepsilon,1}$, for all $\varepsilon>0$, $m=1,2,\ldots$, where:

$$\mathcal{C}_{\varepsilon,m} \equiv \lim_{T\to\infty} C_{\varepsilon,m} \tag{1}$$

$$C_{\varepsilon,m} = \frac{1}{\binom{T}{2}} \sum_{1\le s<t\le T}\sum \chi_\varepsilon(\|X^m_t - X^m_s\|) \tag{2}$$

$$X^m_t \equiv (x_t, x_{t+1}, \ldots, x_{t+m-1}) \tag{3}$$

T is the sample size, $\|.\|$ is the max-norm, and $\chi_\varepsilon(.)$ is the symmetric indicator kernel with $\chi_\varepsilon(x) = 1$ if $|x|<\varepsilon$ and 0 otherwise.

The BDS test is then derived from the following result:

Theorem 2.1 (Brock, Dechert and Scheinkman, 1987)

If $\{X_t\}$ is iid, then

$$V_{\varepsilon,m} = \sqrt{T}\,\frac{C_{\varepsilon,m} - C^m_{\varepsilon,1}}{s_{\varepsilon,m}} \xrightarrow{d} N(0,1) \qquad \forall\ \varepsilon>0\ ,\ m=2,3\ldots$$

$s_{\varepsilon,m}$ is an estimate of the asymptotic standard deviation of $\sqrt{T}(C_{\varepsilon,m} - C_{\varepsilon,1}{}^m)$ under the null of iid.

As mentioned before, there is no Lagrange Multiplier interpretation for the BDS test. This possibly leaves some room for improvements on this test.

The approach taken in this paper is motivated in part by the analogy that can be established between the BDS test and empirical distribution function, as they both build on indicator kernels. Therefore, we construct our family of tests by looking at the partial sums of the BDS test and by deriving the asymptotic properties of the empirical process that this construction generates. It might be expected that by looking at the whole path of this process (instead of only at its terminal point, as is done by the BDS statistic), the resulting family of tests will have high power in detecting local departures from iid.

The BDS statistic is a function of U-statistics. To determine convergence results for its partial sums we use a functional central limit theorem for U-statistics — U_T — due to Denker and Keller (1983). This result — Theorem 2.3 below — gives conditions under which the random elements $\xi(r)=\frac{\sqrt{T}\ r}{m\sigma}(U_{[Tr]}-\theta)$, $r\in(0,1)$, converge weakly to the standard Wiener process.

In what follows let D[0,1] define the space of functions on [0,1] that are right-continuous and have left-hand limits (cadlag functions). Denker and Keller obtained asymptotic results for U-statistics for weakly dependent processes, namely, absolutely regular processes. This concept of memory dependence of a stochastic process is defined as

<u>**Definition 2.3**</u>

$\{X_n\}_{n\in\mathbb{N}}$ *is an absolutely regular process if*

$$\beta(n)=\sup_{a\in\mathbb{N}} \mathbf{E}\,[\sup\{|P(A|\mathcal{F}_1^a)-P(A)|\,|A\in\mathcal{F}_{a+n}^{\infty}\}] \longrightarrow 0, \ \forall\ A \in \mathcal{F},$$

where the symbol $\mathcal{F}_a^b$ *denotes the sigma algebra generated by* $\{X_t\,|a\le t\le b\}$, $(1\le a\le b\le\infty)$.

This definition is introduced because, even though we are after a test of iid, there is an m-period overlap in the BDS statistic for $C_{\varepsilon,m}$. Therefore, the theory of U-statistics we use must allow for some kind of dependence in the $\{X_t^m\}$ process. Denker and Keller (1983) made the extension from the iid case to a broad class of processes that includes m-dependent processes. The absolutely regular form of dependence is sufficient for our purposes, because it includes the case of m-dependence. We note that while uniform mixing implies absolute regularity and absolute regularity implies strong mixing, this order cannot be reversed.

We will now reproduce Denker and Keller functional central limit theorem:

<u>**Theorem 2.3**</u> (Denker and Keller, 1982)

$\{X_n\}_{n\geq 1}$ *is absolutely regular with coefficients* $\beta(n)$ *satisfying* $\beta(n)^{\delta/(2+\delta)}=O(n^{-2+\varepsilon})$ *for some* $\delta>0$ *and* $\varepsilon<\frac{1}{2}, \sigma^2\neq 0$, *and* $\sup_{t_1\geq 1,\ldots,t_j\geq 1} E|h(X_{t_1},\ldots,X_{t_j})|^{2+\delta}<\infty$

If h is a non-degenerate kernel, then the D[0,1]*-valued random functions* $\xi(r)=\frac{\sqrt{T}\,r}{m\sigma}(U_{[Tr]}-\theta)$, $(0\leq r\leq 1)$, *converge weakly in* D[0,1] *to the standard Wiener process. (Here,* [] *denotes the integer part of its argument and*

$$U_{[Tr]} = \binom{T}{j}^{-1} \sum_{1\leq t_1<\ldots<t_j\leq [Tr]} h(X_{t_1},\ldots,X_{t_j})\ )$$

3. A FAMILY OF TESTS FOR IID

3.1 Functionals of the Correlation Integral

In order to present the first result of this paper, let $V_{\varepsilon,m}(r)$ denote the D[0,1]-valued random functions

$$V_{\varepsilon,m}(r)=\frac{\sqrt{T}\,r}{\sigma_m}\left[C_{[Tr],\varepsilon,m} - C^m_{[Tr],\varepsilon,1}\right] \qquad (0\leq r\leq 1) \tag{4}$$

Theorem 3.2 below establishes the weak convergence of partial sums of the BDS statistic. The proof of the theorem uses the functional central limit theorem for U-statistics (Theorem 2.3) and a

Taylor series expansion of $V_{\varepsilon,m}$ around c_m and c_1, where $c_m = E[C_{\varepsilon,m}]$. In order to determine uniform convergence of some of the terms in that expression, we need to use the following lemma on U-statistics:

Lemma 3.1

Let $\{X_t\}$ *be an iid process with distribution function* F. *Consider the parametric function* $\theta = E_F[h(X_{t_1},\dots,X_{t_j})]$ *for some kernel* $h = h(X_{t_1},\dots,X_{t_j})$. *Let* $U_T = \binom{T}{j}^{-1} \sum_{1 \le t_1 < \dots < t_j \le T} h(X_{t_1},\dots,X_{t_j})$ *be a U-statistic for* θ *such that* $\sup_{t_1 \ge 1,\dots,t_j \ge 1} E[h(X_{t_1},\dots,X_{t_j})^2] < \infty$. *Then,* $\forall\ \varepsilon > 0$,

$$\text{Prob}\left\{ \max_{j \le k \le T} k(U_k - \theta)^2 \ge \sqrt{T}\varepsilon \right\} \longrightarrow 0, \text{ as } T \to \infty$$

The next theorem is the key result for constructing the family of tests for the null hypothesis of iid. It shows that the normalized partial sums of the statistic $W(C_{\varepsilon,m}, C_{\varepsilon,1}) = C_{\varepsilon,m} - C^m_{\varepsilon,1}$ converge to a standard Wiener process.

Theorem 3.2

If $\{X_t\}$ *is an iid process with distibution function* F, *then*

a) $V_{\varepsilon,m}$ *converges weakly to the standard Wiener process.*

b) $\sigma^2_{\varepsilon,m} = 4 \{ m(m-2)C^{2m-2}(K-C^2)+K^m-C^{2m}+ 2\sum_{j=1}^{m-1} [C^{2j}(K^{m-j}-C^{2m-2j}) - mC^{2m-2j}(K-C^2)] \}$

where

$$C= E[\chi_\varepsilon(X_t,X_s)]=\int[F(x+\varepsilon)-F(x+\varepsilon)]dF(x)$$

$$K= E[\chi_\varepsilon(X_t,X_s)\chi_\varepsilon(X_s,X_r)]=\int[F(x+\varepsilon)-F(x-\varepsilon)]^2dF(x)$$

From the properties of the standard Wiener process, it follows immediately that Theorem 3.2 contains the BDS test, by fixing r=1. This fact leads us to term the tests proposed in this paper as "the BDS family of tests" because they include the BDS test itself.

It is important to notice that no moment conditions on $\{X_t\}$ are imposed in order to derive Theorem 3.2. This is a simple consequence of the fact that the correlation integral that the BDS test is built upon is an indicator kernel. Therefore, the moment conditions imposed by the functional central limit Theorem 2.2 on kernel h of the U-statistic U_T,

$$\sup_{t_1\geq 1,\ldots,t_j\geq 1} E|h(X_{t_1},\ldots,X_{t_j})|^{2+\delta}<\infty ,$$

are automatically satisfied by the indicator kernel. It follows that this family of tests can be used to test the null of iid, regardless of the existence of moments of any order for the series under scrutiny.

A simple application of the continuous mapping theorem shows that if we use a consistent estimate of σ, we can obtain a workable version of Theorem 3.2. As σ is a continuous transformation of C and K, the result depends upon the existence of consistent estimates for these last two statistics. Brock, Dechert, Scheinkman and LeBaron (1991) provide such estimates for C and K. Therefore, we have the following result:

Corollary 1 to Theorem 3.2:

Let $s_{\varepsilon,m}$ *be a consistent estimator of* $\sigma_{\varepsilon,m}$ *and let* $\hat{V}_{\varepsilon,m}$ *denote the functional obtained from* $V_{\varepsilon,m}$ *substituting* $\sigma_{\varepsilon,m}$ *for* $s_{\varepsilon,m}$. *Then* $\hat{V}_{\varepsilon,m}$ *converges weakly to the standard Wiener process.*

Having proved that $\hat{V}_{\varepsilon,m,N}$ converges to the standard Wiener process, we are now ready to develop a set of tests for the null hypothesis of iid. Setting

$$\hat{B}_{\varepsilon,m}(r)=\hat{V}_{\varepsilon,m}(r) - r\,\hat{V}_{\varepsilon,m}(1)$$

we have that $\hat{B}_{\varepsilon,m}(r)$ converges to $B(r)$, where $B(r)$ is the Brownian Bridge on $r \in [0,1]$.

It is worth noting that Theorem 3.2 and the continuous mapping theorem provide the basis for determining the asymptotic behavior of a set of "goodness of fit" statistics. These widely used statistics are based on the empirical distribution function and have the further advantage of having well defined properties. Consequently, we have the following result:

<u>**Corollary 2 to Theorem 3.2**</u>:

(Cramér-von Mises statistic) $$\int_0^1 \hat{B}_{\varepsilon,m}(r)^2 dr \Longrightarrow \int_0^1 B(r)^2 dr$$

(Anderson Darling statistic) $$\int_0^1 \frac{\hat{B}_{\varepsilon,m}(r)^2}{t(1-t)} dr \Longrightarrow \int_0^1 \frac{B(r)^2}{r(1-r)} dr$$

(Kolmogorov-Smirnov statistic) $$\sup_{0\leq r\leq 1} |\hat{B}_{\varepsilon,m,N}(r)| \Longrightarrow \sup_{0\leq r\leq 1} |\hat{B}_{\varepsilon,m,N}(r)|$$

(Kuiper statistic)

$$\sup_{0\leq s,r\leq 1} |\hat{B}_{\varepsilon,m,N}(r)-\hat{B}_{\varepsilon,m,N}(s)| \Longrightarrow \sup_{0\leq s,r\leq 1} |\hat{B}_{\varepsilon,m,N}(r)-\hat{B}_{\varepsilon,m,N}(s)|$$

For a description of the properties of these statistics, including tables of their limiting distributions, we refer the reader to Shorack and Wellner (1987).

3.2 The Choice of ε and m.

A slightly different application of this type of empirical processes may be used to generate a BDS-based test that avoids the problem of selecting the parameters ε and m. As a matter of fact, the BDS test is valid for every ε and m, and no theoretical result is available to determine which values of ε and m maximize the power of the test. Consequently, in empirical applications of the BDS test we commonly see values of the test reported for a grid of values of ε and m. As one may expect, the different values sometimes lead to different conclusions, in finite samples, regarding the validity of the null hypothesis.

As a possible solution we may look at the D[0,1] process, $G_m(\varepsilon) = C_m(\varepsilon) - C_1(\varepsilon)^{m\;1}$, where m is fixed and $\varepsilon \in \mathcal{A} \equiv \{\delta_1 \le \varepsilon \le \delta_2, 0<\delta_1<\delta_2\}$. We restrict our attention to the case where m is fixed and thus simplify the argument, while recognizing that the choice of ε seems to be far more critical for size and power considerations than does the choice of m. In any event, as m takes only a countable number of values, we may easily generalize the argument to handle the case of non-fixed m.

The suggested procedure relies on the fact that $G_m(\varepsilon)$ will converge to a Gaussian process with covariance function given by $\Gamma_m(\varepsilon_1,\varepsilon_2) = \lim_{T\to\infty} T\, E[G_m(\varepsilon_1)G_m(\varepsilon_2)]$. Proceeding informally, we have the standardized process $G^*_m(\varepsilon) = G_m(\varepsilon)/s_m(\varepsilon)$, where $s_m(\varepsilon)$ is the sample analog of $\sigma_m(\varepsilon)=\Gamma(\varepsilon,\varepsilon)$. Therefore, $G^*_m(\varepsilon)$ should converge weakly to the unit variance Gaussian process $\mathcal{G}_m(\varepsilon)$. Evidently, for each ε, $G^*_m(\varepsilon)$ is one particular realization of the BDS statistic. As a new test statistic, one could use $G^*_m = \sup_{\varepsilon\in\mathcal{A}} G^*_m(\varepsilon)$, or any of the statistics described in corollary 2 to Theorem 3.2. This would have the attractive property of not depending on the choice of ε. Hansen (1991a) discusses a method to simulate the distribution of the statistic G_m^*.

4. SUMMARY

This chapter presented a method for using correlation integrals to test whether or not a given time series is iid. The departure points were the BDS statistic and a theorem by Denker and Keller that guarantees the weak convergence of U-statistics to the standard Wiener process. Using this functional central limit theorem, along with some other results, it was possible to determine the large-sample behavior of some functionals of the BDS statistic. The next chapter shows how this family of tests can test for the null hypothesis of iid-linearity.

APPENDIX OF CHAPTER 2

Proof of Lemma 3.1

Because U_k is a U-statistic, $\{U_t-\theta, \mathcal{F}_t\}_{t\geq j}$ is a reverse martingale sequence - Serfling (1980). Hence reversing the order of the indexing set $\{j,\ldots,t\}$ we convert the above process into a forward martingale. Because convex functions of martingales are submartingales and using Theorem 1 of Chow (1960) (i.e., the submartingale extension of Hájek-Rényi inequality) we obtain the result that for every $t > j$,

$$\text{Prob}\left\{ \max_{j\leq k\leq T} k(U_k-\theta)^2 \geq \sqrt{T}\,\varepsilon \right\} \leq \frac{4}{\varepsilon\,\sqrt{T}} \left\{ jE[(U_j-\theta)^2] + \sum_{k=j+1}^{T} E[(U_k-\theta)^2] \right\} \qquad \text{(A.1)}$$

The argument that leads to (A.1) comes from Miller and Sen (1972). The rest of the proof of this proposition parallels the proof of their Lemma 2.

It is also known that $E[(U_j-\theta)^2] \leq \phi T^{-1}\zeta_j$, where $\phi < \infty$ does not depend on T (see Serfling(1980), p 183). Hence, the right-hand side of (A.1) is bounded above by

$$\frac{1}{\varepsilon\sqrt{T}}\phi\zeta_j \left\{ 1 + \sum_{k=j+1}^{T} \frac{1}{k} \right\} \leq \frac{\phi\zeta}{\sqrt{T}} \log T \longrightarrow 0 \quad \text{, as } T \to \infty \qquad \square$$

Proof of Theorem 3.2:

Let $W_{\varepsilon,m}(C_{\varepsilon,m},C_{\varepsilon,1}) = C_{\varepsilon,m}-C_{\varepsilon,1}^{m}$. A first order Taylor expansion of $W_{\varepsilon,m}(C_{\varepsilon,m},C_{\varepsilon,1})$ about (c_m,c_1) gives that

$$W_{\varepsilon,m}(C_{\varepsilon,m},C_{\varepsilon,1})=C_{\varepsilon,m}-c_m-mc_1^{m-1}(C_{\varepsilon,1}-c_1)-\frac{1}{2}\left[m(m-1)c_1^{*}(C_{\varepsilon,1}-c_1)\right]^2$$

given that $c_m=c_1^{m}$ under H_0 . ($c_1^{*} = c_1+\lambda(C_{1,N}(\varepsilon)-c_1)$, $0\le\lambda\le1$).

Using Hoeffding's (1948) projection method,

$$W_{\varepsilon,m,T}=\frac{2}{T}\sum_{t=1}^{T}(h_m(X_t^{m})-c_m) - mc_1^{m-1}\frac{2}{T}\sum_{t=1}^{T}(h_1(X_t)-c_1)+R_{\varepsilon,m,T}+R_{\varepsilon,1,T}+ \tag{A.2}$$
$$-\frac{1}{2}m(m-1)c_1^{*(m-2)}(C_{\varepsilon,1,T}-c_1)^2$$

or, in a more compact notation,

$$W_{\varepsilon,m} = \frac{2}{T}\sum g_m(X_t^{m}) + R_{\varepsilon,m}+ R_{\varepsilon,1} + A_{\varepsilon,1}$$

where:

$$g_m(X_t^{m})\equiv h_m(X_t^{m})-c_m-mc_1^{m-1}(h_1(X_t)-c_1)$$

$$A_{\varepsilon,1} \equiv -\frac{1}{2}m(m-1)c_1^{*(m-2)}(C_{\varepsilon,1}-c_1)^2$$

$$h_1(X)=E\,[\chi_{\varepsilon}(X,Y)\,|\,X=x]$$

Now consider the following functional:

$$W_{\varepsilon,m,[Tr]} = \frac{2}{T}\sum_{t=1}^{[Tr]} g_m(X_t^m) + R_{\varepsilon,m,[Tr]} + R_{\varepsilon,1,[Tr]} + A_{\varepsilon,1,[Tr]}, \quad 0\le r\le 1 \qquad (A.3)$$

Therefore, for $0\le r\le 1$,

$$V_{\varepsilon,m,T}(r) = \frac{\sqrt{T}\, r}{2\sigma}\left(W_{\varepsilon,m,[Tr]} - EW_{\varepsilon,m,[Tr]}\right) = \frac{\sqrt{T}\, r}{2\sigma} W_{\varepsilon,m,[Tr]} =$$

(A.4)

$$= \frac{\sqrt{T}\, r}{2\sigma}\,\frac{2}{T}\sum_{t=1}^{[Tr]} g_m(X_t^m) + \frac{\sqrt{T}\, r}{2\,\sigma} R_{\varepsilon,m,[Tr]} + \frac{\sqrt{T}\, r}{2\sigma} R_{\varepsilon,1,[Tr]} + \frac{\sqrt{T}\, r}{2\sigma} A_{\varepsilon,1,[Nr]}$$

By Theorem 2.2, the first term in (A.4) converges weakly in D[0,1] to the standard Wiener process. The maximal inequality of proposition 3b) in Denker and Keller (1983) shows that

$$\Pr\left\{\max_{0\le r\le 1} r\sqrt{T}\, 2^{-1}\sigma\left|R_{[Tr]}\right| \ge \xi\right\} \le$$

$$\Pr\left\{\max_{0\le r\le 1} [Tr]\left|R_{[Tr]}\right| \ge T^{-1/2}\sigma\xi\right\} = O(T^{-1/2+\varepsilon}(\log T)^3)$$

So, it remains to prove that $\dfrac{\sqrt{T}\, r\, A_{\varepsilon,m,[Tr]}}{2\sigma}$ goes to zero in probability. To show that, we will first perform some operations on

$$\sup_{0\le r\le 1}\left|\frac{\sqrt{T}\ rm(m-1)}{2\ \sigma}\ c_1^{*(m-2)}(C_{\varepsilon,1,[Tr]} - c_1)^2\right| =$$

$$\sup_{0\le r\le 1}\frac{\sqrt{T}\ rm(m-1)}{2\ \sigma}\left|c_1^*\right|\ (C_{\varepsilon,1,[Tr]} - c_1)^2$$

Given that $c_1^* = c_1 + \lambda(C_1 - c_1)$, $0\le\lambda\le 1$, and that $0\le C_1$, $c_1\le 1, m\ge 2$ we have that $\left|c_1^{*(m-2)}\right| \le 1$. Therefore,

$$\sup_{0\le r\le 1}\frac{\sqrt{T}\ rm(m-1)}{2\ \sigma}\left|c_1^*\right|^{(m-2)}(C_{\varepsilon,1,[Tr]} - c_1)^2 \le$$

$$\sup_{0\le r\le 1}\frac{\sqrt{T}\ rm(m-1)}{2\ \sigma}(C_{\varepsilon,1,[Tr]} - c_1)^2 = K \sup_{0\le r\le 1}\sqrt{T}\ r(C_{\varepsilon,1,[Tr]} - c_1)^2,$$

where $K = \frac{m(m-1)}{2\sigma}$.

Therefore,

$$\Pr\left\{\sup_{0\le r\le 1}\sqrt{T}\ r(C_{\varepsilon,1,[Tr]} - c_1)^2 \ge \xi K^{-1}\right\} \le$$

$$\Pr\left\{\sup_{0\le r\le 1}[Tr](C_{\varepsilon,1,[Tr]} - c_1)^2 \ge \xi K^{-1}\sqrt{T}\right\} \longrightarrow 0\ ,\ \text{as } T \to 0$$

by Lemma 3.1. Note that Lemma 3.1 can be applied here because, under the null, the process $\{X_T\}_{T\ge 1}$ is iid. This completes the proof of part 1 of Theorem 3.2.

With respect to $\sigma^2_{\varepsilon,m}$, one just has to notice that this is exactly the variance of the BDS statistic, by definition. □

Notes to Chapter 2.

[1] The notation introduced here is designed to show that ε is no longer fixed. Therefore, $C_m(\varepsilon)$ represents the correlation integral for the embedding dimension m, represented elsewhere in this paper as $C_{\varepsilon,m}$.

CHAPTER 3: A FAMILY OF NONLINEARITY TESTS ROBUST TO HEAVY-TAILED DISTRIBUTED DATA

1. INTRODUCTION

Chapter 2 introduced a family of tests for the null hypothesis that a time series is iid. One interesting question regards just how much the asymptotic behavior of the tests is insensitive to using estimated residuals instead of the innovations. If such a property can be established these tests can serve as diagnostic tools for those models where the invariance property holds. Particularly, it will be shown that no distortion results from applying the tests using estimated residuals of linear models. This result can be used as a basis for constructing a linearity test. Throughout the rest of this dissertation we will adhere to the convention of designating the (unobserved) residuals as "innovations", reserving the term "residuals" for the estimated ones.

Chapter 1 presented two major definitions of linearity. In particular, iid-linear processes were defined as

Definition 1 (Priestley (1988))

$\{Y_t\}$ is a stochastic iid-linear process if it can be written as

$$Y_t = \sum_{j=0}^{\infty} \psi_j U_{t-j} = \psi(L)U_t \ , \tag{1}$$

where U_t~iid, $E[U_t]=0$ and $\psi_0=1$, $\sum_j \psi_j^2 < \infty$

It will also be assumed that a linear process will have an ARMA representation. (See Chapter 1 for a discussion of the conditions under which that representation is possible.)

The linearity testing strategy proposed in this paper is as follows. First we fit a linear model, namely, a finite order ARMA process, to the data. Then, we apply one of the tests proposed in Chapter 2 to the residuals of this estimation procedure. Obviously, this testing strategy makes sense only if we can show that the noise induced by the estimation process does not change the limiting distribution of our tests.

Formally, the problem can be cast in the following terms. Consider the data generating process

$$y_t = F(Y_{t-1}, b, U_t) \tag{2}$$

where b is an unknown p×1 vector of parameters, F is C^2, $Y_t=(y_t, y_{t-1}, \ldots)$, and $\{U_t\}_{t=0}^{\infty}$, the innovation process, is iid.

Assuming that y_t and U_t are one-to-one, we can write model (2) as $U_t=G(Y_t,b)$. Now, let b_T be a $\sqrt{T}$-consistent estimator of b, where by $\sqrt{T}$-consistent estimation we mean that $\sqrt{T}(b_T-b)=O_p(1)$. Then one can get the residuals $U_{t,T}$ as $G(Y_t,b_T)$. Of course, even if $\{U_t\}$ is iid, the estimation process imposes some noise on the $\{U_{t,T}\}$ process. However, because $b_T \xrightarrow{p} b$, one might expect that, as $T \to \infty$, the sequence $\{U_{t,T}\}$ behaves like an iid series.

The problem of determining whether the family of tests introduced in Chapter 2 is invariant to the use of residuals parallels the problem of establishing the same property for the BDS statistic. This question was extensively studied by Brock, Hsieh and LeBaron (1991). In their book, an invariance property of the BDS statistic was proved for a large class of models.

The standard approach to this type of question is based on an application of the Mean Value Theorem. To illustrate this point, let $S_T(\hat{\theta})$ denote any statistic that depends upon some consistently estimated parameter, $\hat{\theta}$. To show that the asymptotic distribution of $S_T(\hat{\theta})$ is the same as the one corresponding to $S_T(\theta)$, we note that the Mean Value Theorem guarantees that

$$\sqrt{T}[S_T(\hat{\theta})-\mu(\theta)] = \sqrt{T}[S_T(\theta)-\mu(\theta)] + \sqrt{T}(\hat{\theta}-\theta)' \left[\frac{\partial}{\partial\lambda}S_T(\lambda)\bigg|_{\lambda=\theta^*}\right],$$

where $\mu(\lambda)=\lim_{T\to\infty} E[S_T(\lambda)]$, when the actual value of the parameter is θ and θ^* is a point between θ and $\hat{\theta}$. Slutsky's Theorem and the asymptotic normality of $\hat{\theta}$ guarantee that $S_T(\hat{\theta})$ and $S_T(\theta)$ are asymptotically equivalent, provided that $\lim_{T\to\infty} E\left[\frac{\partial}{\partial\gamma}S_T(\lambda)\Big|_{\lambda=\theta^*}\right]=0$.

When applying this approach to the BDS statistic, one immediately recognizes that the indicator kernel $\chi_\varepsilon(x,y)$, which is the basis for the BDS statistic, is non-differentiable. Thus, strictly speaking, the Mean Value Theorem cannot be applied. To circumvent this problem, the two proofs of the nuisance parameter theorem provided in Brock, Hsieh and Lebaron (1991) utilize a continuous approximation to the non-smooth indicator kernel $\chi_\varepsilon(x,y)$ — see also Brock and Dechert (1989). Although the validity of this approach is supported by extensive Monte Carlo experiments (also presented in the above-mentioned book), it remains desirable to provide a direct proof of the residuals' invariance property.

Additionally, as our family of tests also builds on the indicator kernel, it follows that any result concerning its invariance to the use of estimated residuals also faces the same problem. We deal with this problem by giving a two-step proof of the result that the tests introduced in chapter 2 are invariant to the use of estimated residuals. First, we present a new direct proof of this result for the BDS test. Second, we extend this finding to our general family of tests.

The proof of the first step is an extension of work by Sukhatme (1958) and Randles (1982). These papers give a set of sufficient conditions for the behavior of non-smooth U- and L-statistics to be invariant to the use of estimated parameters.

Basically, these authors prove that the expansion involved in the Mean Value Theorem remains valid if one can reverse the operations of differentiation and taking the limiting mean. Therefore, we must impose a smoothing condition on the expectation of the kernel on which a test statistic is based (e.g. the indicator kernel in the BDS case).

To address our problem, we first must consider some differences from the Sukhatme and Randles setups. Suppose we want to show that a test applied to the residuals of a model such as $y_t = G(X_t, \theta) + U_t$,where $\{U_t\}$ is iid, has the same limiting distribution as when applied to the true innovations. Our proof, as theirs, uses the concept of generalized residuals, $u_t(\lambda)$, defined as the difference between y_t and the function G when evaluated at a point λ . That is, $u_i(\lambda) = U_i + G(X_i, \theta) - G(X_i, \lambda)$. In particular, we have that $u_t(\theta) = U_t$.

The models Randles investigates are defined by the fact that the X_t variable in the G function presented above is nonstochastic, so that $u_i(\lambda)$ and $u_j(\lambda)$, $i \neq j$, are two independent random variables (although nonidentically distributed). In the present case, we concern ourselves with time series contexts mainly, where the $\{u_t(\lambda)\}$ process is temporally dependent. Therefore, Randles' proof must be generalized to handle this framework, and as such we use the the concept of strong mixing processes as a working tool.

This use, unfortunately, has a major drawback. As we are primarily concerned with constructing a test of nonlinearity, we must apply our test to the residuals of a linear filter. It is now well established that not every linear process is strong mixing. Therefore, this approach imposes some constraints on the generality of our results. Nevertheless, the subset of linear processes that are strong mixing is still quite large and the strong mixing concept is widely used in econometrics. Moreover, the assumptions that we have to make are not much stronger than those Randles impose.

The remainder of this chapter is organized as follows. First, we state the result that extends Randles' result to U-statistics with bounded kernels for data generated by a strong mixing process. We next apply this result to the BDS statistic under different data generating processes, as well as to the class of tests defined in Chapter 2. Finally, we provide a general discussion of the use of our family of tests as tests for the null of iid-linearity. It is shown that the moment requirements of our tests are minimal. This is corroborated by a small simulation study of the behavior of these tests for data generated by heavy-tailed distributions.

2. USING ESTIMATED RESIDUALS

2.1 U-statistics with Estimated Parameters

First, we need to define some notation.

Let $S_T(\theta)$ be a U-statistic with a bounded symmetric kernel, given by $h(u_1(\theta),\dots,u_m(\theta)) < B$ and $\tau_T(\theta) = E\,[h(u_1(\theta),\dots u_m(\theta))]$.

Define

$$W(u_{j_1}(\lambda),\ldots,u_{j_m}(\lambda)) =$$
$$h(u_{j_1}(\lambda),\ldots,u_{j_m}(\lambda)) - \tau_T(\lambda) - [h(u_{j_1}(\theta),\ldots,u_{j_m}(\theta)) - \tau_T(\theta)]$$

and let

$$Q_T(\lambda)\equiv\frac{\sqrt{T}}{\binom{T}{m}} \sum_j W(u_{j_1}(\lambda),\ldots,u_{j_m}(\lambda)),$$

where $\sum_j$ denotes summation over the $\binom{T}{m}$ combinations of k distinct elements $\{j_1,\ldots,j_m\}$ from $\{1,\ldots,T\}$. The kernel W represents the difference between the kernel h evaluated at two different points, λ and θ, of the residual function, which is defined below.

To establish the extension to Randles' result we use the following set of assumptions:

Assumption A1 (DGP):

$$y_t = G(\mathbf{Y}_{t-1},\theta) + U_t$$

where - $\{U_t\}$ is iid and $\mathbf{Y}_t = \{y_{t-1}, y_{t-2},\ldots,y_{t-p}\}$

- $\{y_t\}$ is a strong mixing process with mixing coefficients that satisfy the summability condition

$$\sum_{k=1}^{\infty} \alpha(k)^{1/2} < \infty$$

- G is a measurable function of $\mathbf{Y}_{t-1}$

Define the Residual Function,

$$u_t(\lambda) \equiv y_t - G(Y_{t-1},\lambda) = U_t + G(Y_{t-1},\theta) - G(Y_{t-1},\lambda)$$

$$\Leftrightarrow u_t(\lambda) = U_t + G^*(Y_{t-1},\theta,\lambda)$$

Note that

1) $u_t(\theta) = U_t$

2) $u_t(\hat{\theta}) = \hat{U}_t$, the estimated residuals.

3) $u_t(\lambda)$ is a (strong) mixing process of size γ:

G^* is a measurable function (it is the difference of two measurable functions). Now assume that Y_{t-1} is a finite-order vector (that is, p is finite). If $\{y_t\}$ is mixing such that the strong mixing coefficient $\alpha(T)$ is $O(T^{-\gamma})$, for some $\gamma>0$ then, using theorem 3.49 of White (1984, p. 47) we have that $\{u_t(\lambda)\}$ is mixing such that $\alpha(T)$ is $O(T^{-\gamma})$, for each λ.

Assumption A2:

$$E \sup_{\theta_1 \in K(\theta,d)} \left| h(u_{s_1}(\theta_1),\ldots,u_{s_m}(\theta_1)) - h(u_{s_1}(\theta),\ldots,u_{s_m}(\theta)) \right| < C_1 d$$

where C_1 is a constant and $K(\lambda,d)=\{\lambda_1 \in \mathbb{R}^p : \|\lambda_1-\lambda\| \le d\}$, $d>0$, $\|.\|$ being the max norm. That is, $K(\lambda,d)$ is the hypercube with center at λ and side equal to 2d.

Assumption A3:

$$\lim_{d\to 0} E \sup_{\theta_1 \in K(\theta,d)} |h(u_{s_1}(\theta_1),\ldots,u_{s_m}(\theta_1)) - h(u_{s_1}(\theta),\ldots,u_{s_m}(\theta))|^2 = 0$$

Note that by Lemma 2.6 of Randles (1982), our assumption A3 is automatically satisfied because we use bounded kernels only. We further assume that $\sqrt{T}$-consistent estimation of θ is possible.

Assumption A4 ($\sqrt{T}$ consistency):

$$\sqrt{T}(\hat{\theta}-\theta)=O_p(1)$$

The next result, Theorem 2.1, shows that under the set of assumptions 1-4, the limiting behavior of a bounded U-statistic with a non-differentiable kernel, is not changed when estimated parameters are present. As we show in the next section, this result can be used to prove the insensitivity of the asymptotic behavior of the BDS family of tests to the use of residuals from a large class of models.

Theorem 2.1

Under Assumptions A1-A4, $Q_T(\hat{\theta}) = \sqrt{T}\,[S_T(\hat{\theta})-\tau_T(\hat{\theta})-S_T(\theta)+\tau_T(\theta)] \xrightarrow{p} 0.$

Proof:

See Appendix A of Chapter 3.

2.2 Application of Theorem 2.1 to the BDS Family of Tests

We now show that the asymptotic distributions of the BDS statistic and of the proposed family of tests are invariant to the use of residuals. As already mentioned, the method of proof applies directly to the non-smooth indicator kernel, and avoids using a continuous function approximation to that kernel.

We first restrict our attention to the BDS statistic. The results we want to show are [1]

$$T^{1/2}\,[W(C^*(\varepsilon,m),C^*(\varepsilon,1)) - W(C(\varepsilon,m),C(\varepsilon,1))] \xrightarrow{p} 0 \qquad (3)$$

and

$$s^*(\varepsilon,m) - s(\varepsilon,m) \xrightarrow{p} 0. \qquad (4)$$

Concerning (3), trivial manipulations described in Proposition 3.2 of Appendix A show it sufficient to guarantee that $\sqrt{T}[C^*(m)-C(m)] \xrightarrow{p} 0$ and that $\sqrt{T}[C^*(1)-C(1)] \xrightarrow{p} 0$. That is, residuals do not distort the asymptotic distribution of the correlation integral at dimensions 1 and m.

Theorem 2.1 extends Randles' result to the case where the data generating process obeys a strong mixing condition. We have imposed that the studied U-statistic must be bounded. This restriction simplifies the proof considerably and fits well within the BDS statistic's framework, which builds on the indicator kernel. The proof of our result depends critically on assumptions A2 and A3. As already mentioned, Randles' (1982) lemma 2 shows that, for bounded U-statistics, A2 is sufficient for A3. Therefore, in order to show the invariance to estimated residuals property of the BDS statistic, we must show that for a given model the correlation integral (at embedding dimensions 1 and m) satisfies A2.

To complete the proof of our result, we still have to show that the estimate of the asymptotic variance, $s_T(\varepsilon,m)$, is consistent for $s(\varepsilon,m)$ when using residuals instead of the true innovations. To show that $s^*(\varepsilon,m)-s(\varepsilon,m) \xrightarrow{p} 0$, it follows by continuity of $s(\varepsilon,m)$ on C and K that we need to prove that $K_T^*-K_T \xrightarrow{p} 0$ and that $C_T^*-C_T \xrightarrow{p} 0$. As C is the V-statistic form of $C(1)^2$, it follows that we must yet show that the difference between K_T^* and K_T is asymptotically negligible. Remember that

$$K_T = \frac{1}{T^3}\sum_{r=1}^{N}\sum_{s=1}^{N}\sum_{t=1}^{N}\chi_\varepsilon(X_r,X_s)\chi_\varepsilon(X_s,X_t)$$

To attack this last problem, note that the kernel in K_T, $\chi_\varepsilon(X_r,X_s)\chi_\varepsilon(X_s,X_t)$, is not symmetric. We could easily transform it, however, into a symmetric function (cf. Serfling (1980), p. 172), giving it a V-statistic form. We do not pursue such a procedure here, for the sake of simplicity. To prove that $K_T^*-K_T \xrightarrow{p} 0$, we will show that the kernel on K_T satisfies A2. Therefore, the symmetric version of K_T will also satisfy that condition, and we can apply Theorem

2.1. The next result establishes that $K_T^*-K_T$ is $o_p(1)$.

Lemma 2.2

If $E\left[\sup_{\lambda \in K(\theta,d)} |\chi_\varepsilon(u_i(\lambda),u_j(\lambda))-\chi_\varepsilon(u_i(\theta),u_j(\theta))|\right] \leq Cd$, then $K_T^*-K_T \xrightarrow{p} 0$

Proof:

See Appendix A of Chapter 3.

This lemma implies that for all of the models such that the BDS statistic satisfies (3), we have $s^*(\varepsilon,m)-s(\varepsilon,m) \xrightarrow{p} 0$, because in that case

$$E\left[\sup_{\lambda \in K(\theta,d)} |\chi_\varepsilon(u_i(\lambda),u_j(\lambda))-\chi_\varepsilon(u_i(\theta),u_j(\theta))|\right] \leq Cd$$

This last result shows that the invariance property of the BDS test regarding using of estimated residuals depends primarily on whether assumption A.2 is satisfied. In Appendix B below, we present examples of models for which the BDS statistic satisfies that assumption. The models covered in that appendix are: linear regressions, AR(p), MA(q), ARMA(p,q), some (smooth) nonlinear autoregressions and two models of changing conditional variance: the ARCH model introduced by Engle (1982), and Nelson's (1991) EGARCH model. To achieve this, we impose two additional assumptions, A5) The distribution function of the innovations U, F(.), is absolutely continuous and differentiable.

A6) $f(.)=F'(.)\leq M^*$, that is, the density function of U is bounded.

Note that previous work on these types of models has suggested that the asymptotic distribution of the BDS statistic is not invariant to residuals obtained from ARCH models, defined as $y_t=G(Y_{t-1},\theta)U_t$, U_t~iid. Brock and Potter (1991) suggest a way to circumvent this situation:

$$N_t \equiv \ln U_t^2 = \ln y_t^2 - \ln (G(Y_{t-1},\theta))^2 ,$$

and apply the BDS test to $\hat{N}_t$. Under the null hypothesis that U_t is iid, it follows that N_t is also iid. The invariance property of the BDS statistic for models of conditional heteroskedasticity (as developed in the Appendix B to this chapter) is applicable to this logarithmic transformation only.

Some comments are in order. Not much is known about the asymptotic properties of the estimators of conditional heteroskedasticity models. However, as it is assumed in the present work that these models have iid driving innovations, the resulting processes should be stationary and ergodic — see Nelson (1990) and Bougerol and Picard (1992). In this context, it seems that our requirement that the parameters of the models must be $\sqrt{T}$-consistently estimated is not too strict.

Furthermore, all the results presented in this section are valid when the generalized residuals form a strong mixing sequence. Not much is known about the memory decay for the models analyzed in this section. For linear processes, although there are some known examples of autoregressive models that are not strong mixing (see Andrews (1984)), Chanda (1974) and Gorodetskii (1978) have shown that an important class of linear processes is strong mixing. This class is defined by linear processes with innovations that have a bounded and continuous density function. General results for models with conditional heteroskedasticity are not yet available. Hansen (1991b) establishes conditions under which the GARCH(1,1) model is near epoch dependent, a concept closely related to the notion of mixing processes.

In any event, the fact that for all the models described above Assumption A2 is satisfied, gives a strong indication that the BDS statistic requires no correction in its asymptotic distribution when used in conjunction with estimated residuals. It is important to notice that we do not present a complete characterization of the set of models for which A.2 is satisfied. However, the above analysis seems to indicate that A.2 will be verified for models that admit a representation where the error term enters the model additively.

The last result to be presented in this section of the paper extends the nuisance parameter property of the BDS statistic to the class of tests introduced in Chapter 2. To do so, we need to prove that

$$\sup_{0 \le r \le 1} |Q_{T,r}(\lambda)| \xrightarrow{p} 0, \quad \text{where}$$

$$Q_{T,r}(\lambda) \equiv \frac{\sqrt{T}\ r}{\binom{T}{m}} \sum_{1 \le j_1 < \ldots < j_m \le [Nr]} W(x_{j_1}(\lambda), \ldots, x_{j_m}(\lambda)) \equiv \sqrt{T}\ r\ U_{T,r}(\lambda)$$

Theorem 2.3

Under A1-A5, $\sup_{0 \le r \le 1} |Q_{T,r}(\lambda)| \xrightarrow{p} 0$

Proof:

See the Appendix A to Chapter 3.

Theorem 2.3 above shows that all the results proved in this chapter for the BDS statistic can be extended to the family of tests proposed in Chapter 2.

3. NONLINEARITY TESTING

The results presented in the previous section permit the use of our family of tests to examine the correction of some parametric models. More specifically, it was shown that the asymptotic behavior of those tests is invariant to the use of estimated residuals from a linear (ARMA) model. That is, this invariance property permits us to deliver the promised test of nonlinearity.

Furthermore, in Chapter 1 we discussed why most nonlinearity tests are poorly suited for testing nonlinearities in economic and financial time series. It was shown that some of these series do not have finite fourth moments, while all of the described nonlinearity tests required the existence of at least finite fourth moments. We now argue that our family of tests are appropriate for nonlinearity testing under moment condition failure. We base this claim on two types of arguments. First, we provide theoretically-based reasons. Second, we present a small simulation study.

3.1 Moment Conditions Required by the BDS Family of Tests

The BDS family of tests builds on the indicator kernel. Consequently, the moment conditions required by the theory of U-statistics are directly concerned with a bounded random variable. Specifically, Denker and Keller (1983) derive (functional) central limit theorems for U-statistics under the assumption that the kernel of the statistic under consideration has finite $2+\delta$ $(\delta>0)$ moments. The indicator kernel has finite moments of all orders. As a further consequence, the limiting behavior of the BDS family of tests as a direct test of iid (when no estimation process is involved) is independent of the existence of moments of the series under analysis.

However, when using any of the BDS family tests to detect nonlinearity, one must first remove the linear components in that series. As already mentioned, this involves estimating an ARMA process. Moreover, the nuisance parameter property described in Section 2 of this chapter requires $\sqrt{T}$-consistent estimation of the ARMA parameters.

That is, the moment conditions required by this family of tests are those required by the estimation procedure of an ARMA process. Brockwell and Davis (1987, Theorem 10.8.2) show that least squares estimation of the ARMA parameters gives the desired result if the variance of U_t is finite and if $\{U_t\}$ is iid. This implies that we can test linearity requiring only that $E\,U_t^2 < \infty$ by applying any of the tests in the BDS family. This "moment-free" property fits well into the results obtained by Loretan and Phillips (1992), who showed that although fourth moments do not seem to be finite for asset prices series, corresponding variances do appear to be finite.

Furthermore, Hannan and Kanter (1977) [3] showed that for the least squares estimator — $\hat{\rho}_i$, $i=1,2,\dots,p$ — of the parameters of a finite order autoregressive process with iid innovations whose distribution function is in the domain of attraction of a stable law of index $\alpha \in (0,2)$, we have that $\hat{\rho}_i - \rho_i = o_p(T^{-1/\gamma})$, for any $\gamma > \alpha$. This implies that, for this kind of linear processes, least squares estimators converge at a faster rate than the usual $T^{1/2}$ rate. Therefore, in what concerns the moment requirements of the BDS statistic, this shows that we can apply the BDS family of tests as a linearity test even when the innovations process do not have finite variance, — or, for this matter, when the maximal moment exponent of the innovation process is $\alpha < 2$ — provided that the we are applying the test to the residuals of an autoregressive process whose innovation process follows a distribution that is in the normal domain of attraction of a stable law with index $\alpha \in (0,2)$.

3.2 Simulation Study

In the previous section we showed our family of tests to have minimal moments requirements. In order to assess the validity of the theoretical arguments used above, we run a small Monte Carlo study similar to that presented in Chapter 1. As before, we generated iid data from a symmetric Pareto distribution

$$\begin{cases} P(U>x)=\frac{1}{2}(x+1)^{-\alpha} & x>0 \\ P(U<-x)=\frac{1}{2}(x+1)^{-\alpha} & x>0 \end{cases}$$

for α=1.5, 2, 4 and 8. $\alpha=\infty$ represents the case where $U_t \sim iidN(0,1)$. Note that $E[U_t]=0$, by construction. We fixed T=1000 as our sample size and considered 1000 replications of each experiment. Table 2.2 of Chapter 1 reports the simulation standard errors.

We present results of these simulations for two statistics from the BDS family of tests: the BDS statistic itself and the Cramer-Von Mises (CVM) type statistic $\int_0^1 \hat{B}_{\varepsilon,m}(r)^2 dr$ (See Chapter 2). We also computed the Kolmogorov-Smirnov and Kuiper statistics but the results were not satisfactory. It appears that the maximum of the empirical process on which we base our statistics behaves too wildly for the asymptotic results to provide a good approximation. This suggests bootstrapping the distribution of these statistics.

At the same time, the experiments we run with statistics based on integrals of the empirical process (CVM, Anderson and Darling, Mallows, and Watson type tests — see Shorack and Wellner (1987) for a description of these statistics — all seem to have good large sample properties. The results for this group of statistics were very similar so we simply report values concerning the CVM type test.

All the test statistics purposed in this dissertation depend on two nuisance parameters, ε and m. In Chapter 2, we suggested a method that would avoid the choice of those two parameters. However, this method implies very high computational costs. Thus, we instead chose the parameters ε and m according to the values previously found in other Monte Carlo studies — c.f. Brock, Hsieh and LeBaron (1991) to maximize the power of the BDS statistic.

For this reason we compute our statistics for two different sets of values of (ε,m), namely, $(\varepsilon,m)=(\hat{\sigma},2)$ and $(\varepsilon,m)=(1.25\hat{\sigma},2)$, where $\hat{\sigma}$ is the standard deviation of the series under test. The value of m was set equal to 2 in both experiments because the choice of m seems to have little impact on the size and power of the BDS test.

In general, the results reported in Tables 3.1 and 3.2 [4] and in Figures 3.1 to 3.4 provide strong evidence that both tests perform well in testing those situations with moment condition failure. The BDS statistic's empirical size of the is remarkably close to the nominal one in almost every case. The CVM test seems to produce empirical sizes which are slightly biased upwards. This is especially relevant in the case where $\alpha=1.5$, where both statistics have some problems. Note that in these simulations we are directly testing the null of iid, as the data was generated with no linear components. In this case, both test statistics do not require the series under test to have finite moments of every order.

Table 3.1:

Size experiments on the BDS family of tests ($m=2$, $\varepsilon=\hat{\sigma}$)

(1)		$\alpha=1.5$	$\alpha=2$	$\alpha=4$	$\alpha=8$	$\alpha=\infty$
0.100	BDS	0.111	0.090	0.109	0.101	0.110
	CVM	0.159	0.135	0.117	0.113	0.144
0.050	BDS	0.074	0.061	0.064	0.058	0.054
	CVM	0.109	0.073	0.065	0.063	0.080
0.025	BDS	0.047	0.037	0.032	0.031	0.030
	CVM	0.072	0.044	0.036	0.034	0.044
0.010	BDS	0.026	0.019	0.012	0.019	0.017
	CVM	0.052	0.018	0.015	0.021	0.017
MEAN	BDS	0.014	0.035	0.003	-0.010	0.002
	CVM	0.297	0.204	0.182	0.181	0.209
VAR	BDS	1.229	1.052	1.007	1.064	1.085
	CVM	2.261	0.049	0.025	0.028	0.032

(1) - The first eight rows give the emprical sizes of the tests designated as BDS and CVM (Cramer-Von Mises statistic), whose nominal size is given by column (1). The two last rows report the values of the sample mean (MEAN), and the sample variance (VAR). Sample size: 1000.

One possible explanation for the reduced quality of the results when $\alpha=1.5$ may lie on the choice of ε. Indeed, for $\alpha=1.5$ the generated series U_t comes from a distribution with infinite variance, so that $\hat{\sigma}$ is converging to infinity as T goes to infinity; a small

Figure 3.1: Estimated Density of the BDS Test (T=1000, $\varepsilon=\sigma$, m=2)

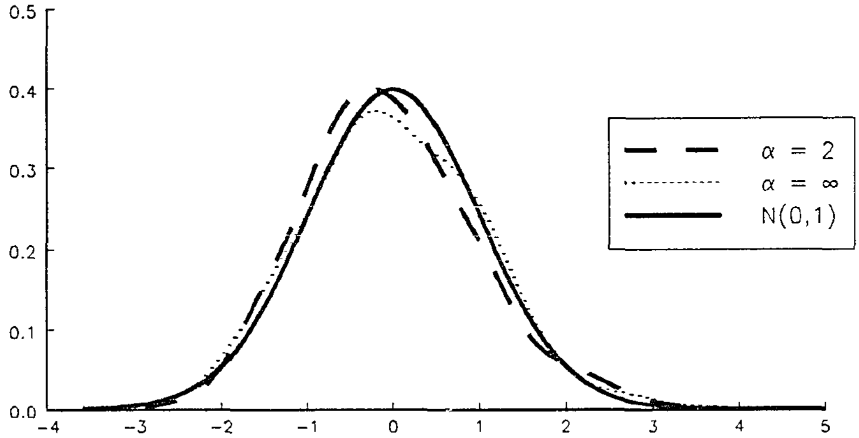


Figure 3.2: Estimated Density of the BDS Test (T=1000, $\varepsilon=1.25\sigma$, m=2)

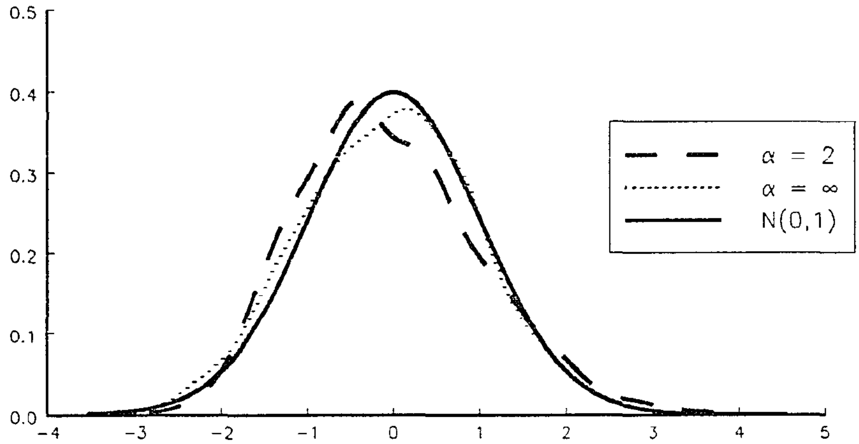


Figure 3.3: Estimated Density of the Cramer–Von Mises Test (T=1000, $\varepsilon=\sigma$, m=2)

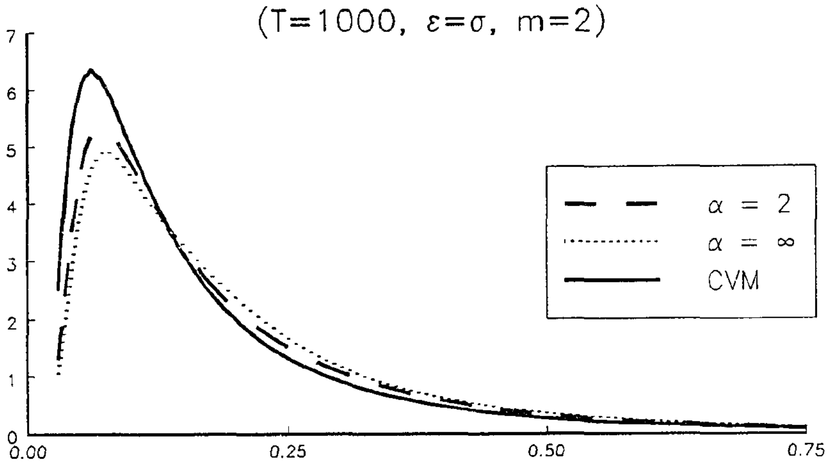
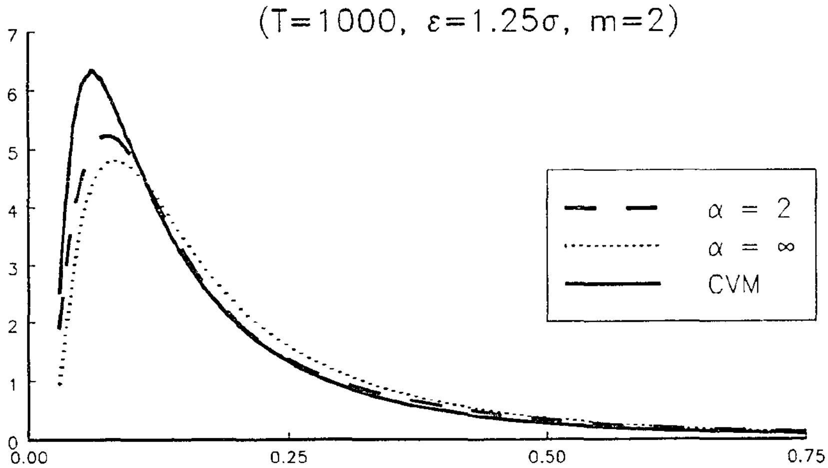


Figure 3.4: Estimated Density of the Cramer–Von Mises Test (T=1000, $\varepsilon=1.25\sigma$, m=2)



calculation [5] shows that $\hat{\sigma}$ is $O_p(T^{(2-\alpha)/2\alpha})$ if $\alpha<2$. The large values of $\hat{\sigma}$, in practice, when $\alpha<2$, will thus cause the same problems for both tests documented in our simulation experiment. This suggests that we could improve the size results when $\alpha=1.5$ by choosing ε in a different way. To do so, we could look at the interquartile range of the generated series. We decided not to run such an experiment because we are mainly interested in those series that do possess finite variance.

Table 3.2:

Size experiments on the BDS family of tests ($m=2$, $\varepsilon=1.25\sigma$)

(1)		$\alpha=1.5$	$\alpha=2$	$\alpha=4$	$\alpha=8$	$\alpha=\infty$
0.100	BDS	0.100	0.106	0.091	0.083	0.090
	CVM	0.160	0.144	0.118	0.121	0.150
0.050	BDS	0.070	0.061	0.051	0.050	0.049
	CVM	0.105	0.081	0.068	0.066	0.078
0.025	BDS	0.048	0.035	0.026	0.023	0.024
	CVM	0.069	0.050	0.038	0.038	0.047
0.010	BDS	0.033	0.015	0.008	0.012	0.012
	CVM	0.052	0.032	0.016	0.019	0.022
MEAN	BDS	-0.059	0.061	0.015	-0.063	0.038
	CVM	0.237	0.202	0.191	0.186	0.209
VAR	BDS	1.065	1.049	0.915	0.968	1.021
	CVM	0.195	0.037	0.029	0.027	0.029

(1) - See Table 3.1.

In general, results for the BDS statistic seem to be slightly better than those established for the CVM test. This is hardly surprising, as our choice of ε and m was determined by the values of ε and m that seemingly work better for the BDS statistic. We have no corresponding results available for the CVM test. This suggests that we should perform a more extensive Monte Carlo simulation of our test statistics, not only on the choice of ε and m, but also in determining the power properties of these tests. We do not present such study at this point because the central purpose of our dissertation is to test nonlinearities under moment condition failure. This same motive lead us not to use estimated residuals in this simulations (After all, the nonlinear testing strategy chosen throughout this dissertation is based on testing iid on estimated residuals). It should be noted that the simulations presented in Chapter 1 were also concerned with the performance of nonlinearity tests on the true innovations and not on the estimated residuals.

4. CONCLUSIONS

The family of tests introduced in Chapter 2 has the property that its asymptotic distribution is invariant to whether one uses residuals or unobserved innovations. This result was proved under exactly the same set of hypotheses as was the BDS statistic's nuisance parameter property. It is established by working directly with the non-smooth indicator kernel, and thus avoids the approximation approaches previously pursued for the BDS test. We extended results introduced in the statistics literature by Sukhatme (1958) and Randles (1982). It should be noticed that this method of proof could be easily generalized to the test statistic proposed by Baek (1988).

It is widely accepted that many economic and financial time series have thick tailed distribution functions. Recently, fourth moment condition failure has been documented in empirical work concerning stock market and exchange rates data. In this chapter, we proved that the BDS family of tests requires, at most, finite second moments of the series under study. This robustness to the nonexistence of finite higher order moments makes our family of tests particularly well-suited for testing nonlinearities in economic time series. This may be important, *inter alia*, in testing the efficient market hypothesis. In this case, however, we acknowledge that we may reject the null too often. Such an increased incidence reflects the fact that market efficiency implies only that the innovation sequence forms a martingale difference sequence (with respect to the past observable data) if the risk premium is time-invariant. In contrast, our family of tests are designed to test the more restrictive null of iid.

In this sense, the null hypothesis studied in this dissertation is too strong for economic and financial data. The null of mds-linearity seems more appropriate for these type of data. Moreover, from a purely statistical view point, the definition of mds-linear is also the more natural one. This is because mds-linearity implies that the best linear predictor of Y_t coincides with its best predictor, in a mean square error sense. However, consistent tests of the mds-linearity hypothesis are not easy to formulate. Further complications occur if one is concerned with tests that have minimal moment conditions requirements. We leave this topic for future research.

APPENDIX A OF CHAPTER 3: *Proofs*

Proof of Theorem 2.1

The first part of this proof follows the same arguments used by Sukhatme (1958) and Randles (1982). Remember that $Q_T(\hat{\theta})$ was constructed as a U-statistic that measures the difference of a given kernel evaluated at two points of the generalized residuals function.

$$\Pr[|Q_T(\hat{\theta})|>\xi\ ]=$$

$$\Pr[|Q_T(\hat{\theta})|>\xi\ ,\hat{\theta}\in K(\theta,B_T)\ ]+\Pr[|Q_T(\hat{\theta})|>\xi\ ,\ \hat{\theta}\notin K(\theta,B_T)\ ]\leq$$

$$\leq \Pr[\ \sup_{\lambda\in K}|Q_T(\lambda)|\ >\ \xi]+\Pr[\ \hat{\theta}\notin K(\theta,B_T)\ ],\text{ where } K=K(\theta,B_T).$$

$\sqrt{T}$-consistency (Assumption A4) implies that the second term in the last expression goes to zero. We must therefore prove that $\Pr[\sup_{\lambda\in K}|Q_T(\lambda)|>\xi]\longrightarrow 0$.

First, let δ be the mesh size of a partition of the hypercube $K(\theta,d)$ and let $N(\delta)=([1/\delta]+1)^p$ be the number of hypercubes defined by such a partition. Note that $[x]$ denotes the greatest integer contained in x. Also, let $K_1,K_2,...K_{N(\delta)}$ be some ordering of such cubes so that $\bigcup_i K_i\supset K(\theta,B_T)$. Let $\theta+b_i/\sqrt{T}$ denote the left hand vertex of K_i.

The proof's next step consists in defining the following mutually exclusive sets,

$$E_i=\{ \sup_{\lambda\in \bigcup_{j<i} K_j} |Q_T(\lambda)| \le \xi < \sup_{\lambda\in K_i} |Q_T(\lambda)|\}$$

from which it follows that

$$\Pr[\sup_{\lambda\in K}|Q_T(\lambda)| > \xi] = \Pr[\bigcup_{i\le N(\delta)} E_i] =$$

$$\Pr\left[\bigcup_{i\le N(\delta)} \{ E_i \cap [(|Q_T(\theta+b_i/\sqrt{T})| > \xi/2) \cup (|Q_T(\theta+b_i/\sqrt{T})| \le \xi/2)]\}\right] \le$$

$$\Pr\left[\bigcup_{i\le N(\delta)} |Q_T(\theta+b_i/\sqrt{T})| > \xi/2\right] + \Pr\left[\bigcup_{i\le N(\delta)} (E_i \cap |Q_T(\theta+b_i/\sqrt{T})| \le \xi/2)\right] \le$$

$$\Pr\left[\max_{i\le N(\delta)} |Q_T(\theta+b_i/\sqrt{T})| > \xi/2\right] + \sum_{i\le N(\delta)} \Pr\left[\sup_{\lambda\in K_i} |Q_T(\lambda)-Q_T(\theta+b_i/\sqrt{T})| > \xi/2\right]$$

Concerning the first term in the last expression, we have that

$$\mathrm{Var}(Q_T(\theta+b_i/\sqrt{T}))=T\binom{T}{m}^{-2} E\,[\sum_s W(u_{s_1}(\theta+b/\sqrt{T}),\ldots,u_{s_m}(\theta+b/\sqrt{T}))]^2=$$

$$T\binom{T}{m}^{-2} E[\sum_s W_s \sum_l W_l]$$

where

$W_s = W(u_{s_1}(\theta+b/\sqrt{T}),\ldots,u_{s_m}(\theta+b/\sqrt{T}))$, $W_l = W(u_{l_1}(\theta+b/\sqrt{T}),\ldots,u_{l_m}(\theta+b/\sqrt{T}))$

and $b=b_i$.

The residuals process $\{u_t(\lambda)\}$ is, as seen before, a strong mixing process of size $O(T^{-\gamma})$, provided that $\{y_t\}$ is mixing of the same size. The indicator kernel is a measurable function of the $u_t(\lambda)$'s and thus $h(u_{s_1}(\lambda),\ldots,u_{s_m}(\lambda))$ defines a strong mixing sequence. Therefore, W_s is itself strong mixing with size $O(T^{-\gamma})$ and is s_m-measurable, as $s_1<s_2<\ldots<s_m$. That is, in order to establish the mixing properties of W_s we repeatedly apply White's Theorem 3.49. Moreover, as we restrict our attention to bounded kernels, W_s is bounded, say $|W_s| < C$, and by construction $E[W_l]=0$. Using this and the fact that W_s and W_l are s_m and l_m measurable, respectively, we can apply Hall and Heyde's (1980, p. 277) Corollary A.1.,

$$|E[W_s W_l]| \le 6C(E[W_l^2])^{1/2}\alpha(|l_m - s_m|)^{1/2}, \qquad (A.5)$$

where $\alpha(m)$ is the strong mixing coefficient.

We want to show that $E[\sum_s \sum_l W_s W_l]$ goes to zero. Using (A.5),

$$E[\sum_s \sum_l W_s W_l] \le \sum_s \sum_l 6C(E[W_l^2])^{1/2}\alpha(|l_m - s_m|)^{1/2}$$

Assume, without loss of generality, that $s_m<l_m$. To prove our result, we take into account that we have $\sum_{j=k+m-1}^{T-1} \binom{j}{m-1}\binom{j-k}{m-1}$ terms where $l_m - s_m = k$. Therefore,

$$E[\sum_s \sum_l W_s W_l] = \sum_{k=0}^{T-m}\left[6C(E[W_l^2])^{1/2}\alpha(|l_m - s_m|)^{1/2}\sum_{j=k+m-1}^{T-1}\binom{j}{m-1}\binom{j-k}{m-1}\right]$$

As $E[W(e_{s_1}(\theta+b/\sqrt{T}),\ldots,e_{s_m}(\theta+b/\sqrt{T}))]^2 = O(1)$ uniformly in s, by A2 and A3

$$E\,[\sum_s \sum_l W_s W_l] = o(T^{2m-1}) \sum_{k=0}^{T-m} \alpha(k)^{1/2}$$

Putting terms together,

$$\begin{aligned} Var(Q_T(\theta+b_i/\sqrt{T})) &= T\binom{T}{m}^{-2} E[\sum_s W_s \sum_l W_l] \\ &\le T\binom{T}{m}^{-2}\left[o(T^{2m-1}) \sum_{k=0}^{T-m} \alpha(k)^{1/2}\right] = o(1) \end{aligned}$$

Therefore, $Q_T(\theta+b_i/\sqrt{T}) \xrightarrow{p} 0$. As only a finite number of i's exist, $Pr\left[\max_{i\le N(\delta)} |Q_T(\theta+b_i/\sqrt{T})| > \xi/2\right] \longrightarrow 0$.

II) We must still show that

$$\sum_{i\le N(\delta)} Pr\left[\sup_{\lambda\in K_i} |Q_T(\lambda)-Q_T(\theta+b_i/\sqrt{T})| > \xi/2\right] \longrightarrow 0.$$

Once again, this expression includes a finite number of terms only. Consequently, we just need to prove that the term inside the summation operator goes to zero. Define

$$H_j(K_i) \equiv \sup_{\lambda \in K_i} \left| W(u_{j_1}(\lambda),\ldots,u_{j_m}(\lambda)) - W(u_{j_1}(\theta+b/\sqrt{T}),\ldots,u_{j_m}(\theta+b/\sqrt{T})) \right|$$

By A2 we have that $E[H_j(K_i)] \le 2C\delta/\sqrt{T}$. Pick δ so that $2C\delta < \xi/4$. Therefore, $\sqrt{T}\binom{T}{m}^{-1}\sum_j E[H_j(K_i)] < \xi/4$ and

$$\Pr\left[\sup_{\lambda \in K_i} |Q_T(\lambda)-Q_T(\theta+b_i/\sqrt{T})| \le \xi/2\right] \le \Pr\left[\sqrt{T}\binom{T}{m}^{-1}\sum_j H_j(K_i) > \xi/2\right] \le$$

$$\Pr\left[\sqrt{T}\binom{T}{m}^{-1}\sum_j (H_j(K_i)-E[H_j(K_i)]) > \xi/4\right] \le$$

$$\mathrm{Var}\left[\sqrt{T}\binom{T}{m}^{-1}\sum_j (H_j(K_i)-E[H_j(K_i)])\right] / (\xi/4)^2,$$

by Chebyschev's inequality. We may now apply the same arguments invoked for the previous proof to show that this variance goes to zero: $H_j(K_i)$ defines a strong mixing sequence. Thus, Hall and Heyde's Corollary A2 guarantees that the summability condition (A1) imposed on the mixing coefficients is sufficient to establish our result. □

Proposition 2.1

If $\sqrt{T}[C^*(m)-C(m)] \xrightarrow{p} 0$ and $\sqrt{T}[C^*(1)-C(1)] \xrightarrow{p} 0$ then

$$\sqrt{T}[W(C^*(\varepsilon,m),C^*(\varepsilon,1))-W(C(\varepsilon,m),C(\varepsilon,1))] \xrightarrow{p} 0$$

Proof:

We start with a first order expansion of

$$W(C^*(\varepsilon,m),C^*(\varepsilon,1))-W(C(\varepsilon,m),C(\varepsilon,1)).$$

That is, as $W(x,y)=x-y^m$, expanding W around (c_m,c_1), we have that

$$W(x,y) = c_m-c_1^m+ (x-c_m) - m\bar{c}_1^{\,m-1}(y-c_1)$$

where $\bar{c}_1$ is a point on the line segment between c_1 and y.

$$\Rightarrow W(C(m),C(1)) = (C(m)-c_m) - m\bar{c}_1^{\,m-1}(C(1)-c_1)$$

$$\Rightarrow W(C^*(m),C^*(1)) = (C^*(m)-c_m) - m\bar{\bar{c}}_1^{\,m-1}(C^*(1)-c_1)$$

where $\bar{c}_1$ is a point on the line segment between c_1 and $C(1)$ and $\bar{\bar{c}}_1$ is a point on the line segment between c_1 and $C^*(1)$. Therefore,

$$W(C^*(m),C^*(1))-W(C(m),C(1))=$$

$$(C^*(m)-C(m)) + m\bar{\bar{c}}_1^{\,m-1}(C^*(1)-c_1) - m\bar{c}_1^{\,m-1}(C(1)-c_1) =$$

$$(C^*(m)-C(m)) + mc_1^{m-1}(C^*(1)-C(1)) + \qquad (A.6)$$
$$+ m(\bar{\bar{c}}_1^{\,m-1}-c_1^{m-1})(C^*(1)-c_1)-m(\bar{c}_1^{\,m-1}-c_1^{m-1})(C(1)-c_1)$$

where we obtain expression (A.6) by adding and subtracting to $W(C^*(m),C^*(1))-W(C(m),C(1))$ the terms $mc_1^{m-1}(C(1)-c_1)$ and $mc_1^{m-1}(C(1)^*-c_1)$.

We first consider the last two terms in (A.6). It follows from Theorem 1 in Denker and Keller (1983) that C(1) converges in probability to c_1, provided that the indicator kernel is non-degenerate. Therefore, because $\bar{c}_1$ is an intermediate point between c_1 and C(1), then $\bar{c}_1^{\,m-1}-c_1^{\,m-1} \xrightarrow{p} 0$. As $\sqrt{T}[C(1)-c_1)$ is $O_p(1)$, it follows by the Slutsky theorem that

$$m(\bar{c}_1^{\,m-1}-c_1^{\,m-1})\sqrt{T}[C(1)-c_1)) \xrightarrow{p} 0.$$

To prove that $m(\bar{\bar{c}}_1^{\,m-1}-c_1^{\,m-1})(C^*(1)-c_1) \xrightarrow{p} 0$, assume that $\sqrt{T}[C^*(1)-C(1)) \xrightarrow{p} 0$. This implies that the difference between $\bar{\bar{c}}_1^{\,m-1}$ and $c_1^{\,m-1}$ vanishes asymptotically. Therefore, if $\sqrt{T}[C^*(m)-C(m)]$ also converges to zero we have proved the proposition. □

Proof of Lemma 2.2:

Let $J_\varepsilon(\lambda)=\chi_\varepsilon(u_r(\lambda),u_s(\lambda))\chi_\varepsilon(u_s(\lambda),u_t(\lambda))-\chi_\varepsilon(u_r(\theta),u_s(\theta))\chi_\varepsilon(u_s(\theta),u_t(\theta))$

By Proposition B.3 in the appendix B, we have that

$$E\left[\sup_{\lambda \in K(\theta,d)} |J_\varepsilon(\lambda)|\right] \le$$

$$E\left[\sup_{\lambda \in K(\theta,d)} |\chi_\varepsilon(u_r(\lambda),u_s(\lambda))-\chi_\varepsilon(u_r(\theta),u_s(\theta))|\right] +$$

$$E\left[\sup_{\lambda \in K(\theta,d)} |\chi_\varepsilon(u_s(\lambda),u_t(\lambda))-\chi_\varepsilon(u_s(\theta),u_t(\theta))|\right]$$

Therefore, this kernel satisfies A2. This means that $\sqrt{T}(K_T^* - K_T) \xrightarrow{p} 0$, proving the result.□

Proof of Theorem 2.3

As $\sup_{0\le r\le 1} |Q_{T,r}(\lambda)| = \sup_{0\le r\le 1} |\sqrt{T}\ r\ U_{T,r}(\lambda)| \le$

$$\sup_{0\le r\le 1}\left[\sqrt{T} \sum_{1\le j_1<\ldots<j_m\le[Nr]} |W(x_{j_1}(\lambda),\ldots,x_{j_m}(\lambda))|\right] \le$$

$$\sqrt{T} \sum_{j} |W(x_{j_1}(\lambda),\ldots,x_{j_m}(\lambda))|$$

Under the assumptions of Theorem 2.1, we have that $\sup_{0\le r\le 1} |Q_{T,r}(\lambda)| \xrightarrow{p} 0$.

APPENDIX B TO CHAPTER 3: Examples of Models for which the BDS Family of Tests Satisfies Theorem 2.1

In what follows, define

$$\chi_\varepsilon(x,y) \equiv 1[|x-y|<\varepsilon] = \begin{cases} 1 \text{ if } |x-y|<\varepsilon \\ 0 \text{ otherwise} \end{cases}$$

$$J_\varepsilon(u_i(\lambda),u_j(\lambda)) \equiv \chi_\varepsilon(u_i(\lambda),u_j(\lambda)) - \chi_\varepsilon(u_i(\theta),u_j(\theta)).$$

$$H_\varepsilon(E_i(\lambda),E_j(\lambda)) \equiv \chi_\varepsilon(E_i(\lambda),E_j(\lambda)) - \chi_\varepsilon(E_i(\theta),E_j(\theta)),$$

where

$E_i(\lambda) \equiv (u_i(\lambda),\ldots,u_{i+m}(\lambda))$, so that

$$\chi_\varepsilon(E_i(\lambda),E_j(\lambda)) = \prod_{h=0}^{m} \chi_\varepsilon(u_{i+h}(\lambda),u_{j+h}(\lambda)).$$

1. Linear Regression Model

Let $y_i=X_i\beta + U_i$; X_i is a 1xp non-random vector, β is the px1 vector of coefficients, and U_i~iid, with mean zero.

$$\Longrightarrow u_i(\lambda)=U_i + X_i(\beta-\lambda)$$

Therefore, $I_\varepsilon(u_i(\lambda),u_j(\lambda)) = 1[|U_i-U_j+(X_i-X_j)(\beta-\lambda)|<\varepsilon]$. It follows that

$$\sup_{\lambda \in K(\beta,d)} |J_\varepsilon(u_i(\lambda),u_j(\lambda))| =$$

$$\sup_{\lambda \in K(\beta,d)} |1[|U_i-U_j+(X_i-X_j)(\beta-\lambda)|<\varepsilon]-1[|U_i-U_j|<\varepsilon]| =$$

$$\sup_{\lambda \in K(\beta,d)} |1[|x+a'(\beta-\lambda)|<\varepsilon]-1[|x|<\varepsilon]|, \text{ where } a=X_i-X_j \text{ and } x=U_i-U_j.$$

As shown in Proposition 1 of this appendix,

$$\sup_{\lambda \in K(\beta,d)} |1[|x+a'(\beta-\lambda)|<\varepsilon]-1[|x|<\varepsilon]| = \begin{cases} 1 \text{ if } \varepsilon-|a|d<|x|<\varepsilon+|a|d \\ 0 \text{ otherwise} \end{cases}$$

It then follows that

$$E\left[\sup_{\lambda \in K(\beta,d)} |J_\varepsilon(u_i(\lambda),u_j(\lambda))|\right] =$$

$$= F(\varepsilon+d|a|'s)-F(\varepsilon-d|a|'s)+F(-\varepsilon+d|a|'s)-F(-\varepsilon-d|a|'s)$$

where F(.) is the distribution function of U_i-U_j.

We use the mean value theorem along with assumption A6, and find that $F(\varepsilon+d|a|'s)-F(\varepsilon-d|a|'s) \le 2dM^*|a|'s$, that is,

$$F(\varepsilon+d|a|'s)-F(\varepsilon-d|a|'s) \le 2dM^* \sum_{h=1}^{p} |x_{ih}-x_{jh}|.$$

Assuming that the X's are bounded, we have that

$$E\left[\sup_{\lambda_1 \in K(\lambda,d)} |J_\varepsilon(u_i(\lambda_1),u_j(\lambda_1))|\right] \le 4dM,$$

so that A2 is satisfied. This proves that $\sqrt{T}(C^*(1)-C(1)) \xrightarrow{p} 0$. Thus, we must simply prove that $\sqrt{T}(C^*(m)-C(m)) \xrightarrow{p} 0$. Assume, without loss of generality, that m=2.

Let $H_\varepsilon(U_i(\lambda),U_j(\lambda))\equiv$

$$I_\varepsilon(u_i(\lambda),u_j(\lambda))I_\varepsilon(u_{i+1}(\lambda),u_{j+1}(\lambda))-I_\varepsilon(u_i(\beta),u_j(\beta))I_\varepsilon(u_{i+1}(\beta),u_{j+1}(\beta))$$

where $U_i(\lambda)\equiv(u_i(\lambda),u_{i+1}(\lambda))$.

$$I_\varepsilon(U_i(\lambda),U_j(\lambda)) = 1[|U_i-U_j+(X_i-X_j)(\beta-\lambda)|<\varepsilon]1[|U_{i+1}-U_{j+1}+(X_{i+1}-X_{j+1})(\beta-\lambda)|<\varepsilon] = 1[|x+a'(\beta-\lambda)|<\varepsilon]1[|y+b'(\beta-\lambda)|<\varepsilon]$$

where $a=X_i-X_j$, $x=U_i-U_j$, $b=X_{i+1}-X_{j+1}$ and $y=U_{i+1}-U_{j+1}$.

Therefore, as shown in Proposition 3 of this appendix,

$$E\left[\sup_{\lambda \in K(\beta,d)} |H_\varepsilon(E_i(\lambda),E_j(\lambda))|\right]\le 4K_1d$$

This proves that $\sqrt{T}(C^*(m)-C(m)) \xrightarrow{p} 0$. Therefore, in the framework of a linear regression model the BDS statistic's asymptotic behavior is invariant to the use of estimated residuals.

2. Nonlinear Regression

$y_i=g(X_i,\beta)+U_i$, X_i non random, U_i~iid with zero mean.

$$u_i(\lambda)=U_i+ g(X_i,\beta)-g(X_i,\lambda)$$

$$I_\varepsilon(u_i(\lambda),u_j(\lambda)) = 1[|U_i-U_j+g(X_i,\beta)-g(X_i,\lambda)-(g(X_j,\beta)-g(X_j,\lambda))|<\varepsilon]$$

Assume that $g(X_i,\lambda)$ is continuously differentiable in a neighborhood of β. Expand $g(X_i,\lambda)$ in a first-order Taylor series around β:

$$g(x,\lambda)=g(x,\beta)+\frac{\partial g(x,\lambda^*)}{\partial\lambda}(\lambda-\beta),\quad \lambda^*\in K(\beta,h)$$

Let $A(x,\gamma)=\frac{\partial g(x,\gamma)}{\partial\lambda}$. Therefore,

$$I_\varepsilon(u_i(\lambda),u_j(\lambda)) = 1[|U_i-U_j+\frac{\partial g(X_i,\lambda^*)}{\partial\lambda}(\lambda-\beta)-\frac{\partial g(X_j,\lambda^*)}{\partial\lambda}(\lambda-\beta)|<\varepsilon]=$$

$$1[|U_i-U_j+(A(X_i,\lambda^*)-A(X_j,\lambda^*))(\lambda-\beta)|<\varepsilon] = 1[|x-a'(\lambda-\beta)|<\varepsilon]$$

where $x=U_i-U_j$ and $a=A(X_j,\lambda^*)-A(X_i,\lambda^*)$.

That is, we reduced the nonlinear regression case to the simpler linear regression framework. Analogously to that case, we impose conditions A5 and A6, and that $\frac{\partial g(x,\gamma)}{\partial\lambda}$ is uniformly bounded. Under this set of conditions, we have that

$$E\left[\sup_{\lambda\in K(\beta,d)}|J_\varepsilon(u_i(\lambda),u_j(\lambda))|\right]\le 4dM.$$

Furthermore, as the Taylor series expansion reduces the nonlinear regression model case to a linear framework, it immediately follows that

$$E\left[\sup_{\lambda \in K(\beta,d)} \left|H_{\varepsilon}(U_i(\lambda),U_j(\lambda))\right|\right] \le 4K_1 d,$$

with $H_{\varepsilon}(U_i(\lambda),U_j(\lambda))|$ defined as before. Our invariance property result also holds under a nonlinear regression framework.

The two cases already considered covered non-stochastic regressors. Note that we could use Randles' proof for the two examples above, since $u_i(\lambda)$ is independent of $u_j(\lambda)$, $i \neq j$. We now extend our findings to DGPs of the type $y_t = G(Y_{t-1},\theta)+U_t$.

3. AR(q)

$$y_t = \sum_{i=1}^{q} \rho_i y_{t-i} + U_t \iff y_t = \rho(L) y_{t-1} + U_t \ , \ \{U_t\} \text{ iid}, \ E|U_t| < \infty$$

where L is the lag operator $Ly_t = y_{t-1}$ and $\rho(L)$ is a q-th order lag polynomial. The roots of $1-\rho(L)$ are assumed to be outside the unit circle. The generalized residuals are given by $u_t(\lambda) = U_t + (\rho(L)-\lambda(L))y_{t-1}$ so that

$$u_{t+p}(\lambda)-u_t(\lambda) = U_{t+p}-U_t+(\rho(L)-\lambda(L))(y_{t+p-1}-y_{t-1}) = x_{t+p}+a_{t+p-1}'(\rho-\lambda),$$

where $x_{t+p} = U_{t+p}-U_t$, $a_{t+p-1} = (y_{t+p-1}-y_{t-1},\ldots,y_{t+p-q}-y_{t-q})$, $\lambda=(\lambda_1,\ldots,\lambda_q)$ and $\rho=(\rho_1,\ldots,\rho_q)$.

Using Proposition B.1 we have that

$$E\left[\sup_{\lambda \in K(\rho,d)} |J_{\varepsilon}(u_{t+p}(\lambda),u_t(\lambda))| \;\middle|\; \mathcal{F}^{x}_{t+p-1}\right] = \qquad (1)$$

$$= F(U_t+\varepsilon+ds'|a_{t+p-1}|) - F(U_t+\varepsilon-ds'|a_{t+p-1}|) +$$

$$+ \; F(U_t-\varepsilon+ds'|a_{t+p-1}|) - F(U_t-\varepsilon-ds'+|a_{t+p-1}|$$

where F(.) is the distribution function of U, and s is a qx1 vector of ones.

By A5 and A6, together with the law of iterated expectations and the mean value theorem,

$$E[F(U_t+\varepsilon+ds'|a_{t+p-1}|) - F(U_t+\varepsilon-ds'|a_{t+p-i}|)] \leq$$

$$2dM^{*} E[\sum_{i=1}^{q} |y_{t+p-i}-y_{t-i}|] \leq 2dM^{*} \sum_{i=1}^{q} E[|y_{t+p-i}|+|y_{t-i}|] < \infty$$

using the definition of a_{t+p-1} and the fact that $E|y_t| < \infty$. This last fact results from the assumption that the AR process is causal, together with the assumption that $E|U_t| < \infty$. We can apply the same argument to the other two terms in (1), so it follows that

$$E\left[\sup_{\lambda \in K(\rho,d)} |J_{\varepsilon}(u_{t+p}(\lambda),u_t(\lambda))|\right] \leq 4dM$$

and A2 and A3 are satisfied.

To check that $E\left[\sup_{\lambda_1 \in K(\rho,d)} |H_\varepsilon(U_i(\lambda),U_j(\lambda))|\right] \le K_1 d$ consider that m=2. Then notice that

$$H_\varepsilon(U_{t+p}(\lambda),U_t(\lambda))=$$

$$1[|x_{t+p}-a_{t+p-1}'(\lambda-\rho)|<\varepsilon]1[|x_{t+p+1}-a_{t+p}'(\lambda-\rho)|<\varepsilon]-1[|x_{t+p}|<\varepsilon]1[|x_{t+p+1}|<\varepsilon]$$

where $x_{t+p+1}=U_{t+p+1}-U_{t+1}$ and $a_{t+p}=y_{t+p}-y_t$. Therefore,

$$E\left[\sup_{\lambda \in K(\rho,d)} |H_\varepsilon(U_i(\lambda),U_j(\lambda))|\right] \le$$

$$E\left[\sup_{\lambda \in K(\rho,d)} |J_\varepsilon(u_{t+p}(\lambda),u_t(\lambda))|\right] + E\left[\sup_{\lambda \in K(\rho,d)} |J_\varepsilon(u_{t+p+1}(\lambda),u_{t+1}(\lambda))|\right]$$

using Propositions B.2 and B.3. As the last two terms are bounded, so is the term on the right side of this inequality. This proves we may apply the BDS statistic to the estimated residuals of an AR(p) process, with no resulting distortion on its asymptotic behavior.

4. MA(h)

$$y_t = U_t + \sum_{i=1}^{h} \theta_i U_{t-i} \Leftrightarrow y_t=U_t+ \theta(L)U_{t-1} , \{U_t\} \text{ iid}, E|U_t| < \infty.$$

where $\theta(L)$ is an h-th order lag polynomial.

Here, $u_t(\lambda)=U_t+(\theta(L)-\lambda(L))U_{t-1}$, so that $u_{t+p}(\lambda)-u_t(\lambda)$ =

$U_{t+p}-U_t+(\theta(L)-\lambda(L))'(U_{t+p-1}-U_{t-1}) = x_{t+p}+(\theta-\lambda)'a_{t+p-1}$, where $x_{t+p}=U_{t+p}-U_t$, $a_{t+p-1}=(U_{t+p-1}-U_{t-1},\ldots,U_{t+p-h}-U_{t-h})'$, $\lambda=(\lambda_1,\ldots,\lambda_h)'$ and $\theta=(\theta_1,\ldots,\theta_h)'$.

We repeat the steps of the proof for the AR(q) case, the use of Proposition B.1, Assumptions A.5 and A.6, and the Law of Iterated Expectations to create two expressions of the type

$$E[F(U_t+\varepsilon+ds'|a_{t+p-1}|)-F(U_t+\varepsilon-ds'|a_{t+p-1}|)] \le 2dM^* \sum_{i=1}^{h} E[|U_{t+p-i}|+|U_{t-i}|]$$

As $E|U_t| < \infty$, by assumption, this proves that

$$E\left[\sup_{\lambda \in K(\rho,d)} |J_\varepsilon(u_{t+p}(\lambda),u_t(\lambda))|\right] \le 4dM$$

The case m=2 is easily derived from the proof for the AR(q) case. Therefore, applying the BDS test to the residuals of an MA(h) requires no correction on BDS statistic's asymptotic distribution. We also note that in the MA(h) case, the process $u_t(\lambda)$ is m-dependent and (trivially) strong mixing.

5. ARMA(q,h)

For notational simplicity and because the proof is almost immediate from the AR(q) and MA(h) cases, we only consider the

ARMA(1,1) process here. In this case,

$$y_t=\rho y_{t-1}+ U_t+ \theta U_{t-1},\ \{U_t\} \text{ iid},\ E|U_t| < \infty,\ |\rho| < 1$$

Let $u_t(\tilde{\lambda})=U_t+(\theta-\nu)U_{t-1}+(\rho-\lambda)y_{t-1}$, with $\tilde{\lambda}=(\lambda,\nu)'$.

Therefore,

$$u_{t+p}(\tilde{\lambda})-u_t(\tilde{\lambda}) =$$

$$U_{t+p}-U_t+(\theta-\nu)(U_{t+p-1}-U_{t-1})+(\rho-\lambda)(y_{t+p-1}-y_{t-1}) =x_{t+p}+a_{t+p-1}(\tilde{\rho}-\tilde{\lambda}),$$

where $x_{t+p}=U_{t+p}-U_t$, $a_{t+p-1}=(U_{t+p-1}-U_{t-1},y_{t+p-1}-y_{t-1})$, and $\tilde{\rho}=(\rho,\theta)'$.

Once again, we represented $u_{t+p}(\tilde{\lambda})-u_t(\tilde{\lambda})$ as $x_{t+p}+a_{t+p-1}(\tilde{\rho}-\tilde{\lambda})$ where a_{t+p-1} is t+p-1 measurable. Therefore, we can repeat all the steps in proving the AR(q) case. This shows the invariance of the asymptotic distribution of the BDS test to the use of estimated residuals from an ARMA(1,1) process.

In the cases covered so far, the method of proof utilized the representation of $u_{t+p}(\lambda)-u_t(\lambda)$ as $x_{t+p}+a_{t+p-1}(\rho-\lambda)$, where x is t+p measurable and a is t+p-1 measurable. Clearly, any model where $u_{t+p}(\lambda)-u_t(\lambda)$ can be represented in this way satisfies sufficient conditions for the invariance result of the BDS statistic that we seek, provided that $u_t(\lambda)$ defines a strong *mixing process*. From these considerations we can immediately conclude that some nonlinear processes also satisfy Theorem 2.1. As an example, we may easily show that nonlinear moving averages processes such as $\varepsilon_t=u_t+\theta u_{t-1}u_{t-2}$ are among this class of models. We turn now to some models where the representation defined above can be obtained only after some transformations.

6. Nonlinear Autoregressions

$$y_t= G(Y_{t-1},\theta) + U_t, \quad \{U_t\} \sim \text{iid}$$

$$u_t(\lambda)=U_t+ G(Y_{t-1},\theta)-G(Y_{t-1},\lambda)$$

$$\chi_\varepsilon(u_{t+p}(\lambda),u_t(\lambda)) =$$

$$1[|U_{t+p}-U_t+G(Y_{t+p-1},\theta)-G(Y_{t-p+1},\lambda)-(G(Y_{t-1},\theta)-G(Y_{t-1},\lambda))|<\varepsilon]$$

Assume that $G(Y_{t-1},\lambda)$ is continuously differentiable in a neighborhood of θ. Expand $G(Y_{t-1},\lambda)$ in a first-order Taylor series around θ:

$$G(y,\lambda)=G(y,\theta)+ \nabla G(y,\lambda^*)'(\lambda-\beta), \quad \lambda^*\in K(\beta,h)$$

where $\nabla G(y,\lambda^*)$ is the gradient vector evaluated at λ^*. Therefore,

$$\chi_\varepsilon(u_t(\lambda),u_{t+p}(\lambda)) =$$

$$1[|U_{t+p}-U_t+ (\nabla G(Y_{t+p-1},\lambda^*)-\nabla G(Y_{t-1},\lambda^*))'(\lambda-\beta)|<\varepsilon] =$$

$$= 1[|x_{t+p}-a_{t+p-1}'(\lambda-\beta)|<\varepsilon]$$

where $x_{t+p}=U_{t+p}-U_t$ and $a_{t+p-1}= \nabla G(Y_{t+p-1},\lambda^*)-\nabla G(Y_{t-1},\lambda^*)$

As for the linear case we have that

$$E\left[\sup_{\lambda \in K(\rho,d)} |J_\varepsilon(u_{t+p}(\lambda),u_t(\lambda))|\right] \le 4dM$$

provided that $E[|a_{t+p-1}|] < \infty$. Sufficient to guarantee that this condition is satisfied is the following:

$$\sup_t \sup_\lambda E[|\nabla G(Y_t,\lambda)|] < \infty$$

The proof for m=2 follows the same steps as the linear model. We have thus proved that under a nonlinear autoregression with an additive error term, the asymptotic distribution of the BDS statistic is not affected by the estimation procedure, and the subsequent use of consistent residuals in place of the true innovations.

8. Models of Conditional Heteroskedasticity

This section considers models of the type $y_t=\sigma_t u_t$, where u_t is an iid sequence with $E[u_t]=0$, $V[u_t]=1$. σ_t is a function of the past values of *y* and a vector of parameters θ. The first model we consider is the ARCH(q) specification,

$$y_t=\sigma(Y_{t-1},\theta)u_t, \qquad \sigma(Y_{t-1},\theta)^2= \omega + \sum_{i=1}^{q}\alpha_i y_{t-i}^2,$$

where $\omega > 0$ and the α_j are nonnegative. Again, $Y_{t-1}=(y_{t-1},\ldots,y_{t-q})$.

As discussed previously, the test should be performed on the estimated residuals of $N_t \equiv \ln u_t^2 = \ln y_t^2 - \ln (\sigma(Y_{t-1},\theta))^2$. Under the null hypothesis of correct specification, U_t is iid; it follows that N_t is also iid. Moreover, since we can rewrite the model as $N_t = \tilde{y}_t - \tilde{\sigma}(Y_{t-1},\theta)$, this implies that the asymptotic distribution of the BDS statistic is the same when applied to $\hat{N}_t$ or to N_t, provided that $\tilde{\sigma}$ satisfies the conditions defined for nonlinear autoregressive models. In particular, for the linear ARCH(1) model (assuming that y_t is covariance stationary, that is, $\alpha<1$),

$$\nabla\tilde{\sigma}(Y_{t-1},\theta)=1/(\omega+\alpha y_{t-1}^2)(1\ ,\ y_{t-1}^2)',\quad \omega > 0 \text{ and } \alpha \geq 0.$$

As $E|\nabla\tilde{\sigma}(Y_t,\theta)|=E[1/(\omega+\alpha y_t^2)\ ,\ y_t^2/(\omega+\alpha y_t^2)]'=E[1/(\omega+\alpha y_t^2)\ ,\ u_t^2]'$. These two expectations are finite provided that $E[u_t^2]<\infty$. Note that $1/(\omega+\alpha y_t^2) \leq 1/\omega < \infty$. Therefore, the estimated residuals $\hat{N}_t$ from an ARCH(1) model satisfy condition A2.

More generally, we may apply the logarithmic transformation used in the ARCH(1) process to a larger set of conditional heteroskedastic models. We keep in mind that this transformation is simply a way of transforming the ARCH(1) model into a model with additive noise term, thus it seems quite clear that we may also apply the logarithmic transformation to general ARCH and GARCH models. This can be shown by following the same steps that enabled the generalization of the proof from the AR(1) case to the AR(q), MA(h) and ARMA(q,h) models.

We may additionally apply the logarithmic transformation to the Exponential GARCH (EGARCH) model proposed by Nelson (1991). In this approach, the conditional variance of y_t is given by

$$\ln \sigma_t^2 = \alpha_t + \sum_{i=1}^{\infty} \beta_i g(u_{t-i}).$$

As a consequence, the variable

$$N_t \equiv \ln u_t^2 = \ln y_t^2 - \ln(\sigma(Y_{t-1}, \theta))^2$$

is a linear function of the past u_t's. Assuming as in Nelson (1991, p. 352) that $\sum_{i=1}^{\infty} \beta_i g(u_{t-i})$ has a finite order ARMA representation, the residuals $N_t(\lambda)$ look exactly like those established for the ARMA(q,h) case. Therefore, applying the BDS test to the estimated residuals of an EGARCH model should cause no distortion.

Proposition B.1

Let θ,λ,a be px1 vectors and let s be a px1 vector of ones. Define $|a|=(|a_1|,\dots,|a_p|)'$. Also, let $J_\varepsilon(x,\lambda,a)\equiv\chi_\varepsilon(x+\theta'a,\lambda'a)-\chi_\varepsilon(x,0)$. Then,

$$\sup_{\lambda\in K(\theta,d)}|J_\varepsilon(x,\lambda,a)| = \begin{cases} 1 & \text{if } \varepsilon-ds'|a|<|x|<\varepsilon+ds'|a| \\ 0 & \text{otherwise} \end{cases}$$

Proof:

1) Let $a \gg 0$.

1a) Consider first that $|x|<\varepsilon$. In this case $\sup_{\lambda\in K(\theta,d)}|J_\varepsilon(x,\lambda,a)|=1$ if $|x-a'(\lambda-\theta)|>\varepsilon \iff x > \varepsilon+a'(\lambda-\theta) \vee x < -\varepsilon+a'(\lambda-\theta)$. As $\theta_i-d < \lambda_i < \theta_i+d$, $\forall$ i=1,2,...,p, we have $x > \varepsilon-da's \vee x < -\varepsilon+da's \iff |x|>\varepsilon-ds'a$. We combine the two conditions and have

$$\sup_{\lambda\in K(\theta,d)}|J_\varepsilon(x,\lambda,a)|=1 \text{ if } \varepsilon-ds'a <|x| < \varepsilon$$

1b) We now consider the other possible case, $|x|>\varepsilon$. Then,

$$\sup_{\lambda\in K(\theta,d)}|J_\varepsilon(x,\lambda,a)|=1 \text{if } |x-a'(\lambda-\theta)|<\varepsilon.$$

That is, $\sup_{\lambda\in K(\theta,d)}|J_\varepsilon(x,\lambda,a)|=1$ if $\varepsilon < |x| < \varepsilon+da's$,

so that, for $a>0$, $\sup_{\lambda\in K(\theta,d)}|J_\varepsilon(x,\lambda,a)|=1$ if $\varepsilon-ds'a < |x| < \varepsilon+ds'a$

2) We must consider the case where $a \in \mathbb{R}^p$. However, by simply

substituting $|a|$ for a we obtain Proposition B.1.

Proposition B.2

Let $\theta,\lambda,a,b,$ be px1 vectors and let s be a px1 vector of ones. Define $|a|$ as in Proposition B.1. Let

$$H_\varepsilon(x,y,\lambda,a,b)\equiv\chi_\varepsilon(x+a'\beta,a'\lambda)\chi_\varepsilon(y+b'\beta,b'\lambda)-\chi_\varepsilon(x,0)\chi_\varepsilon(y,0).$$

Then,

$$v(x,y,\theta) = \sup_{\lambda\in K(\theta,d)} |H_\varepsilon(x,y,\lambda,a,b)| = \begin{cases} 1 & \text{if } (x,y)\in\mathcal{A} \\ 0 & \text{otherwise}\end{cases}$$

where $\mathcal{A} = \{\ (x,y):\ \varepsilon-ds'|a|<|x|<\varepsilon+ds'|a|\ \vee\ \varepsilon-ds'|b|<|y|<\varepsilon+ds'|b|\ \}$.

Proof:

To prove this, consider a>0 and b>0.

1) Suppose $|x|>\varepsilon$ and $|y|>\varepsilon$. So $v(x,y,\theta)=1$ if $|x-a'(\lambda-\theta)|<\varepsilon$ and $|y-b'(\lambda-\theta)|<\varepsilon$. As $\|\theta-\lambda\|<d$, it follows that $v(x,y,\theta)=1$ if $\varepsilon<|x|<\varepsilon+ds'a$ and $\varepsilon<|y|<\varepsilon+ds'b$

2) Now, we consider the case $|x|<\varepsilon$ and $|y|<\varepsilon$.
$v(x,y,\theta)=1$ if $|x-a'(\lambda-\beta)|>\varepsilon$ or $|y-b'(\lambda-\beta)|>\varepsilon$ (or both). Using the constraint on λ we find that $v(x,y,\theta)=1$ if: a) $\varepsilon-ds'a<|x|<\varepsilon$ and $|y|<\varepsilon$; or b) $|x|<\varepsilon$ and $\varepsilon-s'db<|y|<\varepsilon$; or c) conditions a) and b)

simultaneously.

3) $|x|<\varepsilon$ and $|y|>\varepsilon$

Proceeding as in 2) and 3), we find that $v(x,y,\theta)=1$ if $\varepsilon<|x|<\varepsilon+ds'a$ and $|y|<\varepsilon$.

4) $|x|>\varepsilon$ and $|y|<\varepsilon$

This is the mirror image of 3). Therefore, $v(x,y,\theta)=1$ if $|x|<\varepsilon$ and $\varepsilon<|y|<\varepsilon+ds'b$.

Putting 1), 2), 3) and 4) together, we find that

$$v(x,y,\theta) = \sup_{\lambda \in K(\theta,d)} |H_\varepsilon(x,y,\lambda,a,b)| = \begin{cases} 1 & \text{if } (x,y) \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{A} = \{ (x,y)\colon \varepsilon-ds'a<|x|<\varepsilon+ds'a \vee \varepsilon-ds'b<|y|<\varepsilon+ds'b \}$.

Generalizing the proof for $a,b \in \mathbb{R}^p$ we get Proposition's B.2 result.

Proposition B.3.

Let $J_\varepsilon(z,\lambda,a)$ and $H_\varepsilon(x,y,\lambda,a,b)$ be the kernels defined in Propositions B.1 and B.2, respectively. If $E\left[\sup_{\lambda \in K(\theta,d)} |J_\varepsilon(z,\lambda,a)|\right] \le K_1 d$, then

$$E\left[\sup_{\lambda \in K(\theta,d)} |H_{\varepsilon}(x,y,\lambda,a,b)|\right] \le K_2 d$$

where K_1 and K_2 are constants and z can be either x or y, two random variables.

Proof:

By Proposition B.2, we know that

$$v(x,y,\theta) = \sup_{\lambda \in K(\theta,d)} |H_{\varepsilon}(x,y,\lambda,a,b)| = \begin{cases} 1 & \text{if } (x,y) \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$$

where $\mathcal{A} = \{ (x,y): \varepsilon - ds'|a| < |x| < \varepsilon + ds'|a| \vee \varepsilon - ds'|b| < |y| < \varepsilon + ds'|b| \}$, so that

$$E\left[\sup_{\lambda \in K(\theta,d)} |H_{\varepsilon}(x,y,\lambda,a,b)|\right] = \Pr(\mathcal{A}) \le \Pr(\mathcal{A}_1 \cup \mathcal{A}_2), \text{ where}$$

$$\mathcal{A}_1 = \{(x,y): \varepsilon - ds'|a| < |x| < \varepsilon + ds'|a|\} \text{ and}$$
$$\mathcal{A}_2 = \{(x,y): \varepsilon - ds'|a| < |y| < \varepsilon + ds'|a|\}$$

Thus, $\Pr(\mathcal{A}_1 \cup \mathcal{A}_2) \le \Pr(\mathcal{A}_1) + \Pr(\mathcal{A}_2) \le 2K_1 d$,

because $\Pr(\mathcal{A}_i) = E\left[\sup_{\lambda \in K(\theta,d)} |J_{\varepsilon}(z,\lambda,a)|\right]$, $i=1,2$. Q.E.D.

Notes to Chapter 3.

1 The stars indicate that we are calculating the corresponding statistics using estimated residuals. For the definition of W(.), C(.) and s(.) see Chapter 2.

2 Note that $\sqrt{T}(V(T)-U(T)) \xrightarrow{p} 0$, so the asymptotic behavior of a U-statistic and its corresponding V-statistic is the same. See Serfling (1980), p. 206.

3 See also Davis, Knight and Liu (1992) for an account of M-estimation of linear autoregressions with infinite variance errors. This paper proves that robust estimators of the parameters of these models also converge faster than the usual rate. Moreover, M-estimation is shown to be asymptotically more efficient than least squares.

4 The simulations were carried out in the GAUSS programing language. Copies of these programs are available from the author upon request.

[5] Note that, for $\alpha<2$, the process U_t is on the normal domain of attraction of a stable distribution with index α — $U_t \in \mathcal{ND}(\alpha)$. Therefore $U_t^2 \in \mathcal{ND}(\alpha/2)$. When $\alpha>1$ and $E[U_t]=0$,

$$\hat{\sigma}^2 = \frac{1}{T} \sum_t U_t^2 = T^{2/\alpha-1} \left(T^{-2/\alpha} \sum_t U_t^2\right).$$

For the case $\alpha<2$, it follows that $2/\alpha > 1$. Therefore, we may use Ibramigov and Linnik (1971, Ch. 2) to show that $T^{-2/\alpha} \sum_t U_t^2$ is bounded in probability, so that $\hat{\sigma}^2 = O_p(T^{2/\alpha-1})$. Consequently, $\hat{\sigma} = O_p(T^{(2-\alpha)/2})$ and for $\alpha=1.5$, $\hat{\sigma} = O_p(T^{1/6})$.

References:

Andrews, D. W. K. (1984), "Non-Strong Mixing Autoregressive Processes," *Journal of Applied Probability*, Vol 21, 930-934.

Baek, E. G. (1988), *Three Essays on Nonlinearity in Economics*, Ch. 1, Ph.D. dissertation, Department of Economics, University of Wisconsin-Madison.

Bierens, H. J. (1990), "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, Vol 58, 1443-1458.

Blum, J., Kiefer, J., and M. Rosenblatt, (1961), "Distribution Free Tests of Independence Based on the Sample Distribution Function," *Annals of Mathematical Statistics*, Vol 32, 485-498.

Brock, W. A. and W. D. Dechert (1988), "A General Class of Specification Tests: The Scalar Case," in *Business and Economics Statistics Section of the Proceedings of the American Statistical Society*, 70-79.

Brock, W. A., W. D. Dechert, and J. Scheinkman (1987), "A Test for Independence Based on the Correlation Dimension," Department of Economics, University of Wisconsin, University of Houston and University of Chicago. (Revised Version, 1991: Brock, W. A., W. D. Dechert, J. Scheinkman, and B. D. LeBaron)

Brock, W. A., D. A. Hsieh, and B. D. LeBaron (1991), *A Test of Nonlinear Dynamics, Chaos and Instability: Theory and Evidence.* M.I.T Press, Cambridge.

Brock, W. A. and S. M. Potter (1991), "Nonlinear Time Series and Macroeconometrics," in Maddala, G. (ed), *Handbook of Statistics Volume 10: Econometrics,* North Holland, New York (forthcoming)

Brockwell, P. J. and R. A. Davis (1987), *Time Series: Theory and Models*, Springer-Verlag, New York.

Bougerol, P. and N. Picard (1992), "Stationarity of GARCH processes and of Some Nonegative Time Series," *Journal of Econometrics*, Vol 52, 115-127.

Chanda, K. C. (1974), "Strong Mixing Properties of Linear Stochastic Processes," *Journal of Applied Probability*, Vol 11, 401-408.

Chow, Y. S. (1960), "A Martingale Inequality and the Law of Large Numbers," *Proceedings of the American Mathematical Society*, Vol 11, 107-111.

Davis, R. A., K. Knight and J. Liu (1992), "M-estimation for Autoregressions with Infinite Variance," *Stochastic Processes and their Applications*, Vol 40, 145-180.

Dechert, W. D. (1988), "A Characterization of Independence for a Gaussian Process in Terms of The Correlation Integral," SSRI Working Paper 8812, University of Wisconsin-Madison .

Denker, M., and Keller, G. (1983), "On U-Statistics and v. Mises' Statistics for Weakly Dependent Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol 64, 617-656.

Durlauf, S. N. (1991), "Spectral Based Testing of the Martingale Hypothesis," *Journal of Econometrics*, Vol 50, 355-376.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, Vol 50, 987-1008.

Engle, R. F. (1984), "Wald, Likelihood Ratio, and Lagrange Multiplier tests in Econometrics," Ch 13 in Z. Griliches and M. D. Intriligator (eds), *Handbook of Econometrics*, Vol II, North-Holland, Amsterdam.

Fama, E. and K. French (1988), "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246-273.

Gorodetskii, V. V. (1978), "On the Strong Mixing Property for Linear Sequences," *Theory of Probability and its Applications*, Vol 22, 411-413.

Granger, C. and T. Teräsvirta, (1992), *Modeling Nonlinear Economic Relationships*, Oxford University Press, Oxford (forthcoming).

Hall P., and C. C. Heyde, (1980), *Martingale Limit Theory and its Application,* Academic Press, New York.

Hannan, E. J. and M. Kanter (1977), "Autoregressive Processes with Infinite Variance," *Journal of Applied Probability*, Vol 14, 411-415.

Hansen, B. (1991a), "The Likelihood Ratio Test Under Non-Standard Conditions: Testing the Markov Trend Model of GNP," Department of Economics, The University of Rochester

Hansen, B. (1991b), "GARCH(1,1) Processes are Near Epoch Dependent," Economics Letters, Vol 36, 181-186.

Hinich, M. (1982), "Testing for Gaussianity and Linearity of a Stationary Time Series," *Journal of Time Series Analysis*, Vol 3, 169-176.

Hoeffding, W. (1948), "A Class of Statistics with Asymptotically Normal Distribution," *Annals of Mathematical Statistics*, Vol 19, 293-325

Ibramigov, I. A., and Y. V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen.

Keenan, D. (1985), "A Tukey Nonadditivity-Type Test for Time Series Nonlinearity," *Biometrika*, Vol 72, 39-44

Lee, T., H. White, and C. Granger (1989), "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," Department of Economics, The University of California, San Diego.

Lo, A. and C. MacKinlay (1988), "Stock Markets Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, Vol 1, 41-66.

Loretan, M. (1991), *Testing Covariance Stationarity of Heavy-Tailed Economic Time Series,* Ch. 3, Ph. D. Dissertation, Yale University.

Loretan, M. and P. C. B. Phillips, (1992), "Testing the Covariance Stationarity of Heavy-Tailed Time Series: An Overview of the Theory with Applications to Several Financial Datasets," SSRI Working Paper 9208, University of Wisconsin-Madison.

Luukonen, R., P. Saikkonen, and T. Teräsvirta (1988a), "Testing Linearity in Univariate Time Series Models," *Scandinavian Journal of Statistics*, Vol 15, 161-175.

Luukonen, R., P. Saikkonen, and T. Teräsvirta (1988b), "Testing Linearity Against Smooth Transition Autoregressions," *Biometrika*, Vol 75, 491-499.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, Vol 36, 394-419.

McLeod, A. I. and W. K. Li (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-residual Autocorrelations," *Journal of Time Series Analysis*, Vol 4, 269-273.

Miller, R. G., and and P. K. Sen (1972), "Weak Convergence of U-statistics and von Mises Differentiable Statistical Functions," *Annals of Mathematical Statistics*, Vol 43, 31-41.

Nelson, D. B. (1990), "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, Vol 6, 318-334.

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, Vol. 59, 347-370.

Priestley, M. (1988), *Non-Linear and Non-Stationary Time Series Analysis*, Academic Press, New York.

Ramsey, J. B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society*, Ser. B, Vol 31, 350-371.

Randles, R. H. (1982), "On the Asymptotic Normality of Statistics with Estimated Parameters," *The Annals of Statistics*, Vol 10, 462-474.

Richardson, M. (1988), "Temporary Components of Stock Prices: A Skeptic's View," mimeo, Stanford University.

Robinson, P. M. (1991), "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, Vol 58, 437-453

Saikkonen, P. and R. Luukkonen (1988), "Lagrange Multiplier Tests for Testing Non-linearities in Time Series Models," *Scandinavian Journal of Statistics*, 15, Vol 55-58.

Serfling, R. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, New York.

Shorack, G. R. and J. A. Wellner (1987), *Empirical Processes with Applications to Statistics,* John Wiley and Sons, New York.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Subba Rao, T. and M. Gabr (1980), "A Test for Linearity of Stationary Time Series," *Journal of Time Series Analysis*, Vol 1, 145-158.

Sukhatme, B. V. (1958), "Testing the Hypothesis that Two Populations Differ Only in Location," *The Annals of Statistics*, Vol 29, 60-78

Thursby, J. G. and P. Schmidt (1977), "Some Properties of Tests for Specification Error in a Linear Regression Model," *Journal of the American Statistical Association*, Vol 72, 635-641.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical Systems Approach*, Clarendon Press, Oxford.

Tsay, R. (1986), "Nonlinearity Tests for Time Series," *Biometrika*, Vol 73, 461-466.

Van Ness, J. W. (1966), "Asymptotic Normality of Bispectral Estimates," *Annals of Mathematical Statistics*, Vol 37, 1257-1272.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, New York.