A Simple and General Test for White Noise

Ignacio N. Lobato and Carlos Velasco
Centro de Investigación Económica
Instituto Tecnológico Autónomo de México (ITAM)
Av. Camino Sta. Teresa 930, México D.F. 10700, Mexico
e-mail:ilobato@itam.mx
Departamento de Econometría y Estadística
Universidad Carlos III de Madrid
Avda. de la Universidad 30, 28911 Leganés (Madrid), Spain.
e-mail:cavelas@est-econ.uc3m.es

PRELIMINARY VERSION. DO NOT QUOTE.

Abstract

This article considers testing that a time series is uncorrelated when it possibly exhibits some form of dependence. Contrary to the currently employed tests that require selecting arbitrary user-chosen numbers to compute the associated tests statistics, we consider a test statistic that is very simple to use because it does not require any user chosen number and because its asymptotic null distribution is standard under general weak dependent conditions, and hence, asymptotic critical values are readily available. We consider the case of testing that the raw data is white noise, and also consider the case of applying the test to the residuals of an ARMA model. Finally, we also study finite sample performance.

Keywords: Gaussianity, nonparametric, autocorrelation, periodogram, bootstrap, nonlinear dependence.

PRELIMINARY VERSION. DO NOT QUOTE.

1 Introduction

One of the statistical problems that historically has attracted more attention in statistics and econometrics with time series data has been testing for lack of serial correlation. Early references include Yule (1926), Bartlett (1955), Grenander and Rosenblatt (1957) or Durbin and Watson (1950). Most of this literature assumed Gaussianity and, hence, identified lack of serial correlation with independence. In general, an uncorrelated series is not necessarily independent since dependence can be reflected in other aspects of the joint distribution such as higher order moments. The distinction between independence and lack of correlation has been stressed recently. In fact, during the last years a variety of models designed to reflect nonlinear dependence has been studied in the econometrics literature. For instance, in empirical finance, ARCH and bilinear models have been widely studied, see Bera and Higgins (1993, 1997) and Weiss (1986) for a comparison. Following the time series literature we will use the term white noise to denote an uncorrelated series that can present some form of dependence.

Tests for white noise have been proposed both in the time domain and in the frequency domain. Next, we briefly review both approaches. In the time domain the most popular test (apart from the Durbin-Watson which is designed to test for lack of first order serial correlation using regression residuals) has been the Box-Pierce Q_p test. The Q_p test is designed for testing that the first p autocorrelations of a series (possibly residuals) are zero. Note that under the assumption of independence, the asymptotic covariance matrix of the first p autocorrelations is the identity matrix, justifying the χ^2_p asymptotic null distribution of the Q_p test statistic. However, when the series present some kind of nonlinear dependence, such as conditional heteroskedasticity, this asymptotic null covariance matrix is no longer the identity. Hence, for this general case the Q_p test is invalid and the literature has proposed the following two modifications of the Q_p test. The first one is to modify the Q_p statistic by introducing a consistent estimator of the asymptotic null covariance matrix of the sample autocorrelations, so that the modified Q_p statistic retains the χ^2_p asymptotic null distribution. The main drawback of this approach is that in order to estimate consistently this matrix a bandwidth number has to be introduced (Lobato, Nankervis and Savin (2002)). The second modification has been studied by Horowitz, Lobato, Nankervis and Savin (2002) who employ a bootstrap procedure to estimate consistently the asymptotic null distribution of the Q_p test for the general case. This solution presents a similar problem, though, namely the researcher has to choose arbitrarily a block length number. In addition, since in the time domain the null hypothesis states that all the autocovariances (not just the first p) of the considered process are zero, the previous tests present the additional problem of selecting the number p. Hence, the practical problem for these tests is that statistical inference can be sensitive to the two arbitrarily chosen numbers, namely, the order of the serial correlation tested, and the bandwidth or block length. Hong (1996) has proposed a consistent test in the time domain for the general null hypothesis for the case of regression residuals. However, notice that Hong restricted to the independent case and introduced a bandwidth number in order to handle the fact that the null hypothesis implies an infinite number of autocovariances. Hong and Lee (2003) have extended Hong's procedure to allow for conditional heteroskedasticity, but their framework still restricts the sample autocorrelations to be asymptotically independent. Francq, Roy and Zakoïan (2003) have considered goodness-of-fit tests for ARMA models with uncorrelated errors, but they need to introduce a bandwidth whose selection is not addressed.

In the frequency domain the null hypothesis is stated in terms of the spectral density instead of the autocorrelations. Hence, the problem of selecting p does not appear. The null hypothesis implies that the spectral density is constant and the most common statistics have been based on the standardized cumulative periodogram (see Bartlett (1955), Grenander and Rosenblatt (1957), Durlauf (1991) or Deo (2000)). Under the assumption of independence, Bartlett and Grenander and Rosenblatt showed that the standardized cumulative periodogram converges to the Brownian bridge. Durlauf further showed that this result still holds when the independence assumption is relaxed to martingale difference sequence (MDS) with conditional homoskedasticity. For the MDS case with conditional heteroskedasticity (and some moment conditions), Deo slightly modified this statistic so that the standardized cumulative periodogram retained the convergence to the Brownian bridge. Notice that in this setup there is no need of introducing any user-chosen number since under the stated assumptions (see condition A in Deo (2000, p. 293) the autocorrelations are asymptotically independent. As Deo comments, his assumption (vii) is the main responsible for the diagonality of the asymptotic null covariance matrix of the sample autocorrelations. However, for many common models, such as GARCH models with asymmetric innovations, EGARCH models and bilinear models, Deo's condition (vii) does not hold and the autocorrelations are not asymptotically independent under the null hypothesis. Hence, for the general case, Deo's test is not asymptotically valid.

For the general case, the frequency domain is still useful, though. In fact, Chen and Romano (1999) have proposed to employ the bootstrap to derive asympotically valid procedures for testing for white noise in the general case. A main obstacle of Chen and Romano's procedure is that, similar to the bootstrap procedure commented above, the proposed bootstrap requires the selection of a block length.

Summarizing, testing for white noise presents two challenging features. The first aspect is that the null hypothesis implies that an infinite number of autocorrelations are zero. This feature has been addressed successfully in the frequency domain under severe restrictions on the dependence structure of the process. The second feature is that the null hypothesis allows the white noise series to present some form of dependence beyond the second moments. This dependence entails that the asymptotic null covariance matrix of the sample autocorrelations is not diagonal, so that it has n^2 non-zero terms, where n is the sample size (contrary to Durlauf (1991) and Deo (2000) who consider a diagonal matrix, and hence, it has only n non-zero elements). This aspect has been handled by introducing some arbitrary user-chosen numbers whose selection complicates statistical inference.

The purpose of this paper is to introduce a simple test for white noise under general weak dependent assumptions. The proposed test statistic has already been considered in Milhøj (1981) as a goodness of fit test. However, Milhøj did not address our problem. He studied goodness of fit for linear processes with i.i.d. innovations and provided an informal analysis. We view Milhøj's test statistic as a classical Cramer-von Mises (CVM) statistic and consider briefly the corresponding Kolmogorov-Smirnov (KS) test. This KS test is related to Fisher's test for hidden periodicities, Fisher (1929). However, the KS test presents a slow rate of convergence to the asymptotic null distribution. This is a serious theoretical drawback because it implies that the test is not able to detect local alternatives that tend to the null at the parametric rate. In addition, the KS test present a worse performance in finite samples. Hence, the practical recommendation is the use of the CVM test. We stress that both tests are straightforward to use since the test statistics are very simple functions of the periodogram and, furthermore, their asymptotic null distributions are standard and therefore, the user does not need to provide any arbitrary number such as a bandwidth or a block length to estimate these asymptotic null distributions.

The plan of the paper is the following. Section 2 introduces notation and the tests, Section 3 provides the asymptotic theory, Section 4 addresses the case of testing for white noise for residuals of ARMA models. Section 5 study the finite sample performance, Section

6 reports an empirical example and Section 7 concludes. The technical material is in the Appendix.

2 Framework and test statistic

Let y_t be a weakly dependent strictly stationary time series. Denote its mean by μ, its centered moments of order k by $\mu_k = E(x_t - \mu)^k$, its autocovariance of order k by $\gamma_k = E[(y_t - \mu)(y_{t+k} - \mu)]$, its autocorrelation of order k by $\rho_k = \gamma_k/\gamma_0$ and define the spectral density $f(\lambda)$ by

$$\gamma_k = \int_\Pi f(\lambda)\exp(ik\lambda)d\lambda \qquad k = 0, 1, 2, \ldots$$

where $\Pi = [-\pi, \pi]$. The sample mean is denoted by $\overline{y}$, the sample autocovariance is $\widehat{\gamma}_k = n^{-1}\sum_{t=1}^{n-|k|}(y_t - \overline{y})(y_{t+|k|} - \overline{y})$, the sample autocorrelation of order k is $\widehat{\rho}_k = \widehat{\gamma}_k/\widehat{\gamma}_0$ and the periodogram is $I(\lambda) = |w(\lambda)|^2$ where $w(\lambda) = n^{-1/2}\sum_{t=1}^{n} x_t \exp(it\lambda)$. Also, call $T = [2^{-1}(n-1)]$. In addition, we denote the q-th order cumulant of $y_1, y_{1+j_1}, \ldots, y_{1+j_{q-1}}$ as $\kappa_q(j_1, \ldots, j_{q-1})$ and the marginal cumulant of order q as $\kappa_q = \kappa_q(0, \ldots, 0)$. Finally, the q-th order cumulant spectral density is denoted by $f_q(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} \in \Pi^{q-1}$ and $\Pi = [-\pi, \pi]$, see expression (2.6.2) in Brillinger (1981, p.25).

The null hypothesis of interest is that $\gamma_k = 0$ for all $k = 1, 2, \ldots$ The alternative hypothesis is the negation of the null, that is, there exists some k such that $\gamma_k \neq 0$. Equivalently, in terms of the spectral density, the null hypothesis states that $f(\lambda) = \gamma_0/2\pi$ for $\lambda \subset \Pi$, and the alternative is that there exists some interval $\Omega \subset \Pi$ such that $f(\lambda) \neq \gamma_0/2\pi$ for $\lambda \subset \Omega$. Note that assumption A in the next section imposes that $f(\lambda)$ is continuous and smooth under the null hypothesis.

The classical approach in the frequency domain involves the standardized cumulative periodogram, that is,

$$Z_n(\lambda) = \sqrt{T}\left(\frac{\sum_{j=1}^{[\lambda T/\pi]} I(\lambda_j)}{\sum_{j=1}^{T} I(\lambda_j)} - \frac{\lambda}{\pi}\right).$$

Based on $Z_n(\lambda)$, the two classical tests statistics are the Kolmogorov-Smirnov

$$\max_{j=1,\ldots,T} |Z_n(\lambda_j)|,$$

and the Cramer von Mises

$$\frac{1}{T}\sum_{j=1,}^{T} Z_n(\lambda_j)^2.$$

These tests statistics have been commonly employed (see Bartlett (1955) and Grenander and Rosenblatt(1957)) because when the series y_t is not only white noise but also independent, it can be shown that the process $Z_n(\lambda)$ converges weakly in $D[0,\pi]$ (the space of cadlag functions in $D[0,\pi]$) to the Brownian bridge process (see Tanaka (1996, pp.39-40) for the definition of Brownian bridge and see Dahlhaus (1985) for the convergence result). Hence, asymptotic critical values are readily available for the independent case. In fact, Durlauf (1991) has shown that the independence assumption can be relaxed to MDS with conditional heteroskedasticity.

However, under general weak dependent assumptions (see Dahlhaus (1985)) the asymptotic null distribution of the process $Z_n(\lambda)$ is no longer the Brownian bridge but, in fact, it converges weakly in $D[0,\pi]$ to a zero mean Gaussian process with covariance given by

$$\frac{\pi G(\pi)}{F(\pi)^2}\left\{\frac{G(\lambda\wedge\mu)}{G(\pi)}+\frac{F(\lambda)F(\mu)}{F(\pi)^2}-\frac{F(\lambda)G(\mu)}{F(\pi)G(\pi)}-\frac{F(\mu)G(\lambda)}{F(\pi)G(\pi)}\right.$$
$$\left.+\frac{F_4(\lambda,\mu)}{G(\pi)}+\frac{F_4(\pi,\pi)}{G(\pi)}\frac{F(\lambda)F(\mu)}{F(\pi)^2}-\frac{F_4(\mu,\pi)}{G(\pi)}\frac{F(\lambda)}{F(\pi)}-\frac{F_4(\lambda,\pi)}{G(\pi)}\frac{F(\mu)}{F(\pi)}\right\},$$

where $F(\lambda)$ denotes the spectral distribution function, $F(\lambda)=\int_0^\lambda f(\omega)d\omega$,

$$G(\lambda)=\int_0^\lambda f(\omega)^2 d\omega,$$

and

$$F_4(\lambda,\mu)=\int_0^\lambda\int_0^\mu f_4(\omega,-\omega,-\theta)d\omega d\theta. \tag{1}$$

The important message from the previous complicated covariance is that the asymptotic null distribution depends on the nature of the data generating process of y_t. Since no asymptotic critical values are available, Chen and Romano (1999, p.628) propose to estimate the asymptotic distribution by means of either the block bootstrap or the subsampling technique. Unfortunately, these bootstrap procedures require the selection of some arbitrary number and in our framework no theory is available about their optimal selection. Alternative bootstrap procedures which do not require the selection of a user-chosen number such as resampling the periodogram as in Franke and Hardle (1992) or in Dahlhaus and Janas (1996) will not estimate consistently the asymptotic null distribution because of the fourth cumulant terms.

Hence, in this paper we consider an alternative approach. Rewrite the null hypothesis as

$$\frac{2\pi f(\lambda)}{\gamma_0}-1=0 \text{ for } \lambda\subset\Pi,$$

and write down the sample analog of the left hand side

$$\frac{I(\lambda)}{\widehat{\gamma}_0} - 1$$

(given our definition of the periodogram, note that $EI(\lambda) = 2\pi f(\lambda)$ see Brockwell and Davies (1991), for instance). Now, in order to derive consistent tests, our proposed test statistic is based on applying the Cramer von Mises functional to $I(\lambda)/\widehat{\gamma}_0 - 1$, leading to the following test statistics

$$M_n = \frac{1}{T}\sum_{j=1}^{T}\left(\frac{I(\lambda_j)}{\widehat{\gamma}_0} - 1\right)^2 = \frac{p_{2n}}{p_{1n}^2} - 1,$$

where $p_{in=}T^{-1}\sum_{j=1}^{T} I(\lambda_j)^i$, $i = 1, 2$, using that $\widehat{\gamma}_0 = p_{1n}$.

Milhøj (1981) has already employed M_n as a general goodness of fit test statistic for time series. Milhøj informally justified the use of this statistic for testing the adequacy of linear time series models, but since he identified white noise with i.i.d. (see p. 177), he neither addressed our problem nor provided a formal analysis of it, that is, he did not derive the asymptotic properties of M_n under general dependence conditions.

Deo and Chen (2000) have proposed a continuous version of M_n and study its properties for Gaussian processes.

An alternative to M_n would be to employ as test statistic one based on the Kolmogorov-Smirnov functional applied to $I(\lambda)/\widehat{\gamma}_0 - 1$, leading to the following test statistic

$$T_n = \max_{j=1,\ldots,T}\left|\frac{I(\lambda_j)}{\widehat{\gamma}_0} - 1\right|.$$

This statistic is related to Fisher's statistic

$$T_n^S = \max_{j=1,\ldots,T}\frac{I(\lambda_j)}{\widehat{\gamma}_0}.$$

Fisher (1929) introduced it to test that y_t is a Gaussian white noise against the alternative that y_t contains an added sinusoidal deterministic component of an unspecified frequency. Similarly to T_n, this statistic is in the KS spirit. However, notice that T_n^S is not consistent for our null hypothesis *since a sharp trough can not be detected.*

We will not pursue the use of T_n as a test because it could be shown that under the null hypothesis and regularity conditions, $(T_n + 1 - \log T) \to_d \Lambda$ where Λ denotes the Gumbel distribution. Hence, this test presents a slow rate of convergence to the asymptotic null distribution. This is a serious theoretical drawback because it implies that the test is not able to detect local alternatives that tend to the null at the parametric rate. In addition, this test present a worse performance in finite samples, as we will see in Section 5.

We finish this section by describing briefly the corresponding test statistics in the time domain. The continuous version of M_n is proportional (or equivalent?) to (see Milhøj (1981/5, p.178) and Deo and Chen (2000, p.162))

$$M_n = \sum_{j=0}^{n-1} r_j^2,$$

whereas the corresponding KS statistic would be $\max_j |r_j|$. We will not pursue these statistics since deriving the asymptotic properties in the time domain is especially challenging given that r_j are not asymptotically independent under the null hypothesis.

3 Asymptotic Theory

We state the assumption for deriving the asymptotic null distribution of M_n.

Assumption 1. *The process x_t satisfies $Ex_t^8 < \infty$, and for $q = 2, 3, \ldots, 8$*

$$\sum_{j_1=-\infty}^{\infty} \cdots \sum_{j_{q-1}=-\infty}^{\infty} |\kappa_q(j_1, \ldots, j_{q-1})| < \infty, \tag{2}$$

$$\sum_{j=1}^{\infty} \left[E \left| \left(E(x_0 - \mu)^4 \mid \Im_{-j} \right) - \mu_4 \right|^2 \right]^{1/2} < \infty, \tag{3}$$

where $\Im_{-j}$ denotes the σ-field generated by x_t, $t \leq -j$, and,

$$E[(x_0 - \mu)^4 - \mu_4]^2 + 2 \sum_{j=1}^{\infty} E \left([(x_0 - \mu)^4 - \mu_4][(x_j - \mu)^4 - \mu_4] \right) > 0. \tag{4}$$

Assumption 1 is a weak dependent assumption that implies that the higher order spectral densities up to the eighth order are bounded and continuous. Notice that the weak condition imposed by *Assumption* 1 allows for time series where the asymptotic null covariance matrix of the sample autocorrelations is not diagonal. Instead of summability of cumulants, we could have employed alternative mixing conditions allowing for nondiagonality of the asymptotic covariance matrix of the sample autocorrelations under the null hypothesis.

Theorem 1. Under Assumption 1 and the null hypothesis

$$\sqrt{T}\left(M_n - 1\right) \to_d N\left(0, 4\right).$$

The proof is omitted since it is a straightforward application of the delta method to the vector $(p_{1n}, p_{2n})'$ that under Assumption 1 obeys

$$\sqrt{T}\begin{pmatrix} p_{1n} - (2\pi)^{-1}\sigma^2 \\ p_{2n} - (2\pi^2)^{-1}\sigma^4 \end{pmatrix} \to_d N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{4\pi^2}\begin{pmatrix} \sigma^4 + \pi f_4 & 2\sigma^6 + 2\sigma^2\pi f_4 \\ 2\sigma^6 + 2\sigma^2\pi f_4 & 5\sigma^8 + 4\sigma^4\pi f_4 \end{pmatrix}\right) \quad (5)$$

where $f_4 = F_4(\pi, \pi)$, and F_4 has been defined in (1). The previous CLT (5) can be derived using Theorem 3 in Rosenblatt (1985)

4 Testing for white noise in regression residuals

In the previous sections, we have considered the case of testing that raw data are white noise. In this section, we will see that our test can also be employed as an specificication test for ARMA models. In particular the considered null hypothesis is that the selected orders for the AR and MA components are correct. As we will see next, this null can be tested consistently by applying our test to the residuals of the assumed ARMA model.

Franq et al. (2003) have also consider goodness-of-fit tests for ARMA models with uncorrelated errors. However, they did not state clearly their null hypothesis, they introduce a bandwith

Milhøj (1981) has considered using the same statistic for testing the correct specification of linear models. However, Milhøj provided a heuristic analysis and assumed i.i.d. innovations with all their moments finite. Obviously, these are very restrictive assumptions that would rule out most economic and financial applications, sucha as ARMA models with arch-type innovations.

Hong and Lee (2003)

We do not consider the case of testing for the correct specification of nonlinear time series models because our test statistic only uses the information contained in the autocovariances, hence the only misspecifications it can detect are misspecifications that are reflected in non-zero autocovariances. Since many different nonlinear processes possess the same autocovariance structure (see Tong (1994) for many examples, such as bilinear and arch-type processes), any test based exclusively on autocovariances has trivial power to detect an alternative model with the same autocovariance structure.

Let introduce some notation. Let say that the stochatic process y_t satisfies an ARMA(p, q)

model, when for all $t \in \mathbb{Z}$,

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-i}$$

where ε_t is a white noise process that satisfies Assumption 1 in the previous section and the polynomials $\phi(L) = 1 - \phi_1 L - ... - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + ... + \theta_q L^q$ have all their roots outside the unit circle and no roots in common.

Then, introduce the following

Assumption 2. The stochastic process y_t is strictly stationary and satisfies the ARMA model with $p = p_v$ and $q = q_v$. (and some moment and weak dependence conditions for ε_t such as assumption 1).

We assume that consistent estimators of the parameters $\phi' s$ and $\theta' s$ are available. For instance least squares estimators (see reference?). Let call $\widehat{\varepsilon}_t$ to the residuals of this fitted model.

Consider testing the null hypothesis

$$H_0 : p = p_v \text{ and } q = q_v$$

where (p, q) are the orders of the ARMA model that the researcher chooses. Then, denote M_n^r the M_n test applied to $\widehat{\varepsilon}_t$.

Theorem 2. Under the null hypothesis and assumption 2

$$\sqrt{T} \left(M_n^r - 1 \right) \rightarrow_d N \left(0, 4 \right).$$

5 Finite Sample Behavior

This section considers the finite sample behavior of the proposed tests. We consider experiments under both the null and under the alternative hypothesis. Under the null hypothesis we compare the performance of our tests with two of the tests considered in Durlauf (1991) and Deo (2000). As commented in the introduction, Durlauf and Deo developed testing procedures for the null hypothesis of white noise rather than for the martingale difference hypothesis as the titles of their papers suggest. Deo acknowledges this fact in p.292. Testing the martingale hypothesis is a very challenging task; for a review of statistical procedures to test the martingale difference hypothesis, see section 2 in Dominguez and Lobato (2003). Durlauf considered a variety of tests for testing for white noise for the case when the asymptotic covariance matrix of the sample autocorrelations is the identity, and Deo extended Durlauf's approach to the case where the previous matrix is diagonal but not the identity. We consider both the Cramer von Mises and the Kolmogorov-Smirnov statistics and following Deo's notation, they will be denoted by CVM, KS, CVM_c and KS_c; the subindex c denotes the correction proposed by Deo.

Under the null two white noise processess are considered, one that is an MDS example and another one which is non-MDS. Both models are white noise, but not independent and they are covariance stationary with finite eighth moment.. The null hypothesis is tested at nominal levels 0.05 and 0.10. The estimates (empirical rejection probabilities) are calculated using 5,000 replications for sample sizes n = 100 and 500.

The MDS example is the ARCH (1) model, $y_t = z_t \cdot \sigma_t$ where z_t is a sequence of IID N(0,1) and $\sigma_t^2 = 1+0.3y_{t-1}^2$. For this process, the asymptotic covariance matrix of the sample autocorrelations is diagonal but not the identity, so Durlauf's procedures (CVM and KS) do not asymptotically control properly the type I error, contrary to Deo's tests (CVM_c and KS_c). The results are reported in Table I. In this example the sample autocorrelations are asymptotically independent but heteroskedastic, hence the tests CVM and KS are not able to control properly asymptotically the type I error. Table I indicates precisely the magnitude of these distortions. Notice that even for $n = 100$, these distortions are considerable. On the contrary, the tests CVM_c, KS_c, M_n and T_n control properly the type I error, especially for $n = 500$. Notice that in this particular MDS example the asymptotic covariance matrix is diagonal, but as commented in the introduction, for many MDS models this matrix is non-diagonal (cf. assumption (vii) in Deo). Notice that KS_c, M_n and T_n are somewhat conservative for n=100.

The non-MDS example is the bilinear model: $y_t = z_t + 0.5z_{t-1}y_{t-2}$ where $\{z_t\}$ is a sequence of iid N(0, 1) random variables. For this process, the sample autocorrelations are not asymptotically independent, hence the tests CVM, KS, CVM_c and KS_c are not able to control properly asymptotically the type I error. Table II indicates precisely the magnitude of these distortions. These distortions are considerable for CVM and KS. The only tests that are able to control the type I error are M_n and T_n that are a bit conservative for $n = 100$, as above.

Notice that for example I, CVM_c controls the type I error better than KS_c; this is the typical case when comparing the Cramer von Mises versus the Kolmogorov-Smirnov test. In spite of that, the test T_n, which is in the Kolmogorov-Smirnov spirit, controls better the size that M_n, which is a Cramer von Mises-type statistic.

Next we reports the empirical powers of the T_n and M_n tests in a small Monte Carlo study where the time series are generated by...MA(1) and AR(1): T_n es mucho menos potente que M_n, pero con el modelo $y_t = a\cos(0.3\pi t) + e_t$donde e_t es iidN(0,1) y a es una realización de una N(0,σ^2); de tal forma que para una replica a es fijo, ahora T_n es comparable o mejor que M_n,

6 References

An-Chen-Hannan, (1983), "The maximum of the periodogram", J. of Multivariate Analysis, 13, 383-400.

Bartlett, M.S. (1955), An Introduction to Stochastic Processes with Special Reference to Methods and Applications, Cambridge University Press, London.

Bera, A.K. and M.L. Higgins (1993) "ARCH models: properties, estimation and testing", Journal of Economic Surveys, 7, 305-362.

Bera, A.K. and M.L. Higgins (1997) "ARCH and bilinearity as competing models for nonlinear dependence", Journal of Business and Economic Statistics, 15, 43-51.

Brillinger, D.R. (1981), Time Series: Data Analysis and Theory, Holden Day, San Francisco.

Bollerslev, T. (1988), "On the correlation structure for the generalized autorregressive conditional heteroskedastic process,", J. of Time Setries Analysis, 9, 121-131.

Box, G.E.P. and D. A. Pierce (1970) Distribution of residual autocorrelations in autoregressive integrated moving average time series models. Journal of the American Statistical Association 65, 1509-1526.

Chen, H. and J.P. Romano (1999), "Bootstrap-assisted goodness-of-fit- tests in the frequency domain", Journal of Time Series Analysis, 20, 619-654.

Chen, Z.-G. and E.J. Hannan (1980), "The distribution of periodogram ordinates", Journal of Time Series Analysis, 1, 73-82.

Dahlhaus, R. (1985), "On the asymptotic distribution of Bartlett's U_p-statistic", Journal of Time Series Analysis, 6, 213-227.

Dahlahus, R. and D. Janas (1996), "A frequency domain bootstrap for ratio statistics in time series analysis", The Annals of Statistics, 24, 1934-1965.

Davis, R.A. and T. Mikosch (1999), The maximum of the periodogram of a non-Gaussian sequence", The Annals of Probability, 27, 522-536.

Deo, R.S. (2000), "Spectral tests of the martingale hypothesis under conditional heteroscedasticity", *Journal of Econometrics*, **99**, 291-315.

Deo, R.S. and W. Chen (2000), "On the integral of the squared periodogram", *Stochastic Processes and their Applications*, 85,159-176.

Dominguez, M. and I.N. Lobato (2003) "Testing the martingale difference hypothesis", Econometric Reviews, 22, 351-377.

Durbin, J. and G.S. Watson, (1950), "Testing for Serial Correlation in Least Squares Regression: I", Biometrika, 37, 409-428.

Durlauf, S. N. (1991), "Spectral based testing of the martingale hypothesis", *Journal of Econometrics*, **50**, 355-376.

Fay, G. and P. Soulier (2001), The peridogram of an i.i.d. sequence", Stochastic Processes and their Applications, 92, 315-343.

Fisher, R.A. (1929), "Tests of significance in harmonic analysis", Proceedings of the Royal Statistical Society, Series A 125, 54-59. (CV: este articulo tampoco lo he visto, pero no creoq ue sea muy importante)

Francq. C., R. Roy & J.-M. Zakoïan (2003) Goodness-of-fit tests for ARMA models with uncorrelated errors", Rapport de recherche CRM-2925, Université de Montréal.

Franke , J. and W. Hardle (1992), "On bootstrapping kernel spectral estimates", The Annals of Statistics, 20, 121-145.

Freedman, D. and DS. Lane (1980), "The empirical distribution of the Fourier coefficients", The Annals of Statistics, 8, 1244-1251.

Grenander, U. and M. Rosenblatt (1957), Statistical Analysis of Stationary Time Series, Chelsea Publishing Company, New York.

Gotze, F. and C. Hipp (1983), "Asymptotic expansions for sums of weakly dependent random vectors", Z. Wahrsch. verw. Gebiete, 64, 211-239.

He, C. and T. Teräsvirta (1999) Properties of moments of a family of GARCH processes. Journal of Econometrics, 92, 173-192.

Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form", Econometrica, 64, 837-864.

Horowitz, J., I.N. Lobato, J.Nankervis and N.E. Savin, "Bootstrapping the Box-Pierce Q test: a robust test of uncorrelatedness", Journal of Econometrics, forthcoming.

Kokoszka, P. and T. Mikosck (2000), "The periodogram at the Fourier frequencies", Stochastic Processes and their Applications, 86, 49-79.

Lobato, I.N., Nankervis J.C.,and Savin N. E. (2002), "Testing for Zero Autocorrelation in the Presence of Statistical Dependence," *Econometric Theory,* **18**, 730-743.

Milhøj, A. (1981), "A test of fit in time series models", Biometrika, 68, 177-187.

Pham, D.T. (1985), "The mixing property of bilinear and generalized random coefficient autoregressive models", Stochastic Processes and Applications, 23, 291-300.

Phillips, P.C.B. and V. Solo (1992) Asymptotics for linear processes. Annals of Statistics 20, 971-1001.

Terdik, G. (1999) Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis. New York: Springer Verlag

Yule, G.U. (1926), "Why do we sometimes get nonsense correlations between time-series? A study in sampling and the nature of time-series. Journal of the Royal Society", vol. 89, pp.2-9, 30-41.

Weiss, A. (1986), "ARCH and bilinear time series models: comparison and combination", Journal of Business and Economic Statistics, 4, 59-70.

7 Tables

Table 1. Percentage of rejections at nominal 10% and 5% levels. The DGP is the ARCH (1) model $y_t = z_t \cdot \sigma_t$ where z_t is a sequence of IID N(0,1) and $\sigma_t^2 = 1 + 0.3y_{t-1}^2$. The sample sizes are 100 and 500. The number of replications is 5000.

n	100		500	
%	10	5	10	5
CvM	19.1	11.9	21.7	13.3
KS	12.0	7.34	17.2	10.7
CvM_c	11.1	5.12	10.0	4.60
KS_c	7.50	3.80	8.10	4.12
M_n	6.92	4.26	8.84	5.24
T_n	8.32	3.60	9.46	4.62

Table 2. Percentage of rejections at nominal 10% and 5% levels. The DGP is a bilinear model, $y_t = z_t + 0.5z_{t-1}y_{t-2}$, where $\{z_t\}$ is IID N(0,1). The sample sizes are 100 and 500. The number of replications is 5000.

n	100		500	
%	10	5	10	5
CvM	20.5	12.8	24.8	16.4
KS	14.0	8.54	20.9	13.6
CvM_c	14.5	7.98	14.8	8.10
KS_c	10.2	5.56	12.8	6.90
M_n	8.16	5.14	9.58	5.30
T_n	8.66	4.10	9.80	4.64