Entropy-Based Dynamic Specification Searches*

Jeff Racine
Department of Economics
Syracuse University
Syracuse, NY 13244-1020, USA
jracine@maxwell.syr.edu

Esfandiar Maasoumi
Department of Economics
Southern Methodist University
Dallas, TX 75275-0496, USA
maasoumi@mail.smu.edu

August, 2003

Abstract

We consider a metric entropy capable of providing guidance in the refinement of parametric specification in time-series models. In particular, we demonstrate how an entropy-based test of nonlinear dependence can be a valuable prescriptive aid in assessing goodness of fit as well as in the model specification process. Illustrative applications to financial models of six major US stock market indices are undertaken. We contrast the proposed measure with competing diagnostic techniques which are meant to test the white noise and/or reversibility properties of model residuals. We examine the BDS, correlation-type, characteristic function-based and information-theoretic approaches.

*The authors would like to thank Dee Dechert for his useful comments and suggestions. Racine would like to acknowledge support from the Center for Policy Research at Syracuse University. The usual caveat applies.

1 Introduction

The most common diagnostic tools used to guide the process of time-series model specification are 'correlation-based' statistics such as the ACF and Ljung-Box (Ljung & Box (1978)) Q statistics, though of course information-theoretic statistics such as AIC (Akaike (1981)) have also proved to be very popular. Alternative statistics used for testing serial independence and other dynamic assumptions have also been used to detect nonlinearity and to assess correct specification. For instance the chaos-based BDS statistic (Brock, Dechert & Scheinkman (1987), Brock, Dechert, Scheinkman & LeBaron (1996)) is often employed as a model diagnostic, and more recently the characteristic function-based "time-reversibility" statistic (Chen & Kuan (2002)) have been proposed and applied for the same purpose. The generally unsatisfactory performance of several approaches to testing for nonlinearity is described in Barnett, Gallant, Hinich, Jungeilges, Kaplan & Jensen (1997).

It is axiomatic that detecting departures from serial independence, particularly in the higher order moments of a distribution, is a necessary component for specification searches in constructing modern time-series models. Numerous diagnostic statistics have been proposed which form the basis for tests that are applied to a model's residuals to provide guidance as to whether or not a specification is 'adequate', i.e. data consistent. Tests which indicate failure of a model but which do not offer guidance in a particular direction are of course less useful in a model specification search framework. Unfortunately, the majority of existing tests lie in this camp for one of two reasons; either i) they lack power in the direction of interesting alternatives, or ii) though having good power, rejection of the null is uninformative from a model specification standpoint.

The BDS test for serial independence is a very popular diagnostic tool, and is routinely applied to residuals from econometric models. See, for example, Cromwell, Labys & Terraza (1994, pg 32–36) who outline its use as a diagnostic tool for linear time series modeling. The popularity of the test derives mainly from an expectation of power against linear, nonlinear, and chaotic (deterministic) alternatives. The test has been incorporated into programs such as Eviews and Dataplore, while modules and libraries have been written for Matlab, Gauss, and S-plus and R making it widely available to the practitioner, so much so that journal referees frequently request its inclusion among the comparison group when proposing improved tests for the detection of nonlinear dependence.

Information theoretic tests are increasingly found to be superior in the same settings. See Hong & White (2000), Granger, Maasoumi & Racine (2002), Skaug & Tjøstheim (1993), among others. BDS and other correlation integrals too may be viewed as an approximation of various mutual information measures. Indeed, this relation may be used to obtain alternative nonparametric estimates for entropy measures, as proposed by Diks & Manzan (2002). But BDS may be seen to be a poor measure of complete statistical "independence". In related work, BDS is also found to be ill-suited to testing the related 'reversibility' property, for example in US stock indices; see Chen & Kuan (2002). Motivated by this failure, the authors proposed an interesting alternative statistic to test for reversibility that appeared to have good performance when used to identify suitable models and lags.

In this paper, we examine an alternative metric entropy statistic which can serve as a diagnostic tool for guiding model specification, and we examine its performance relative to both the "Robust BDS", and the test of Chen & Kuan (2002). We adopt and extend the results from Chen & Kuan (2002) who consider a bootstrap implementation of BDS that is robust to the true (non-Gaussian) underlying distributions. This is an improvement over the vast majority of BDS implementations in the current literature. We also highlight the versatility of the statistic by considering its use as a robust measure of goodness of fit in potentially nonlinear models. The rest of the paper proceeds as follows. Section 2 presents a detailed overview of the diagnostic tests considered herein. Section 3 presents results extending the application performed by Chen & Kuan (2002) on financial models of six major US stock indices, while Section 4 presents some concluding remarks.

2 Description of Test Statistics

We briefly describe the tests which are compared in the current paper, the BDS test (Brock et al. (1987), Brock et al. (1996)), the 'time reversibility' (TR) test of Chen & Kuan (2002), and the entropy-based test S_ρ (Granger et al. (2002), Maasoumi & Racine (2002)). We refer interested readers to the original papers for detailed descriptions of size and power performances of the respective tests.[1]

2.1 The BDS Test

The BDS test statistic is based on the correlation integral of a time series $\{Y_t\}_{t=1}^T$. The generalized K-order correlation integral is given by:

$$C_K(Y,\epsilon) = \left[\int \left(\int I_{(\|y-y'\|\leq\epsilon)} f_Y(y')dy'\right)^{K-1} f_Y(y)dy\right]^{\frac{1}{K-1}},$$

where $I_{()}$ denotes the indicator function, and $\|Y\| = \sup_{i=1,\ldots,\dim Y} |y_i|$, the sup norm. The distance parameter ϵ is like a bandwidth and behaves accordingly. When the elements of Y are i.i.d., the correlation integral factorizes. The BDS test statistic is based on C_K, $K=2$. This gives the expected probability of $\epsilon - neighborhoods$.

For small ϵ and dimensionality parameter m, the inner integral (probability) in $C_K(\cdot)$ behaves as $\epsilon^m f_Y(y)$ over the ϵ-neighborhood. This allows us to see an approximate relationship between the correlation integral and various entropies.

Before proceeding, we note that, in applied settings, the user is required to set the embedding dimension (m) and the size of the dimensional distance (ϵ). One often encounters advice to avoid using the test on samples of size 500 or smaller, while one also encounters advice on setting ϵ in the range $0.5\sigma_y$ to $2.0\sigma_y$ of a time series $\{Y_t\}_{t=1}^T$ along with advice on setting m in the range $2-8$.

[1]For size and power performance of the "Robust BDS" (permutation-based BDS) test, the interested reader is referred to Belaire-Franch & Contreras (2002).

The test's finite-sample distribution has been found to be poorly approximated by its limiting $N(0,1)$ distribution. In particular, the asymptotic-based test has been found to suffer from substantial size distortions, often rejecting the null 100% of the time *when the null is in fact true*. Recently, tables providing quantiles of the finite-sample distribution have been constructed in certain cases which attempt to correct for finite-sample size distortions arising from the use of the asymptotic distribution (see Kanzler (1999) who assumed true Gaussian error distributions), though the asymptotic version of the test is that found in virtually all applied settings. A number of authors have noted that its finite-sample distribution is sensitive to the embedding dimension, dimension distance, and sample size, thus tabulated values are not likely to be useful in applied settings. However, a simple permutation-based resampled version of the BDS statistic does yield a correctly sized test (Belaire-Franch & Contreras (2002), Diks & Manzan (2002)), hence we elect to use this "Robust BDS" approach implemented by Chen & Kuan (2002) for what follows.

2.2 Time Reversibility Tests

Recently, Chen & Kuan (2002) have suggested using a modified version of the 'time reversibility' test of Chen, Chou & Kuan (2000) as a diagnostic test for time series models. This is a characteristic function-based test, and its authors recommend it in part as it requires no moment conditions hence is of wider applicability than existing TR tests. A stationary process is said to be 'time reversible' if its distributions are invariant to the reversal of time indices; independent processes are time reversible. If time reversibility does not hold (i.e. the series is 'time irreversible'), then there is asymmetric dependence among members of the series in the sense that the effect of, say, Y_s on Y_t is different from that of Y_t on Y_s; the threshold autoregressive (TAR) model is one example of a time irreversible series.

When a series is time reversible then the distribution of $Y_t - Y_{t-k}$ is symmetric ($k = 1, 2 \dots$), while failure of this symmetry condition indicates asymmetric dependence. A distribution is symmetric if and only if the imaginary part of its characteristic function is zero (i.e. $h(\omega) = E[\sin(\omega(Y_t - Y_{t-k}))] = 0 \quad \forall \quad \omega \in \mathbb{R}^+$). This can be seen to be equivalent to testing

$$E[\psi_g(Y_t - Y_{t-k})] = E\left[\int_{\mathbb{R}^+} \sin(\omega(Y_t - Y_{t-k}))g(\omega)\, d\omega\right] = 0, \tag{1}$$

where $g(\cdot)$ is a weighting function, and Chen et al. (2000) therefore propose a test based on the sample analog of (1) given by

$$C_{g,k} = \sqrt{T_k}\left(\frac{\bar{\psi}_{g,k}}{\bar{\sigma}_{g,k}}\right),$$

where $T_k = T - k$, $\bar{\psi}_{g,k} = \sum_{t=k+1}^{T} \psi_g(Y_t - Y_{t-k})/T_k$, and $\bar{\sigma}^2_{g,k}$ is a consistent estimator of the asymptotic variance of $\sqrt{T_k}\bar{\psi}_{g,k}$. Chen et al. (2000) choose $g(\cdot)$ to be the exponential distribution function with parameter $\beta > 0$ yielding a test that is straightforward to compute, and under H_0, it can be shown that $C_{g,k}$ has a limiting $N(0,1)$ distribution. However, this version of the test suffers from substantial size distortions. A modified version of this test appropriate for testing asymmetry of residuals arising from a time series model which is called the TR test (Chen & Kuan (2002)) is given by

$$\widehat{C}_{g,k} = \sqrt{T_k}\left(\frac{\hat{\psi}_{g,k}}{\hat{\nu}_{g,k}}\right),$$

where $\hat{\psi}_{g,k} = \sum_{t=k+1}^{T} \psi_g(\hat{\epsilon}_t - \hat{\epsilon}_{t-k})/T_k$ (for our purposes $\hat{\epsilon}_t$ represents standardized residuals from a time series model) and where $\hat{\nu}^2_{g,k}$ is a consistent estimator of the asymptotic variance of $\sqrt{T_k}\hat{\psi}_{g,k}$ which is obtained by bootstrapping (Chen & Kuan (2002, pg 568)).

2.3 Entropy-Based Tests

Entropy-based measures of divergence have proved to be powerful tools for a variety of tasks. They may form the basis for measures of nonlinear dependence, as in Granger et al. (2002) who consider the case of the K-Class entropy, as in:

$$H_K(Y) = \frac{1}{K-1}\left[1 - \int f_Y^{K-1} f_Y(y) dy\right], \quad K \neq 1$$

(equal to Shannon's entropy for $K = 1$),

$$\simeq \frac{1}{K-1}\left\{1 - [C_K(Y;\epsilon)/\epsilon^m]^{K-1}\right\},$$

or, for Renyi's entropy:

$$H_q(Y) = \frac{1}{q-1} \log \int [f_Y(y)]^{q-1} f_Y(y) dy, q \neq 1,$$

$$\simeq -\log C_q(Y;\epsilon) + m \log \epsilon.$$

A nonparametric estimate of $C_q(\cdot)$ may then be used to obtain a similar estimate of $H_q(\cdot)$, as cogently argued by Diks & Manzan (2002).

BDS has an asymptotic standard Gaussian distribution under the null of i.i.d. variables. *Mutual Information* measures test the significance of different metrics of divergence between the joint distribution and the product of the marginals. For instance:

$$I(X,Y) = \int\int \ln(f_{X,Y}(x,y)/f_X(x)f_Y(y)) \times f_{X,Y}(x,y)\, dx\, dy$$

is the simple Shannon mutual information. And, generally

$$Iq(X,Y) = Hq(X) + Hq(Y) - Hq(X,Y),$$

for any q, relates mutual information to the respective entropies. The conditional mutual information, given a third variable Z, is:

$$I(X,Y|Z) = \int\int\int \ln(f_{X|y,z}(x|y,z)/f_{X|z}(x|z)) f_{X,Y,Z}(x,y,z)\, dx\, dy\, dz.$$

More generally,

$$\widehat{I}q(X,Y|Z) = \ln \widehat{C}q(X,Y,Z;\epsilon) - \ln \widehat{C}q(X,Z;\epsilon) - \ln \widehat{C}q(Y,Z;\epsilon) + \ln \widehat{C}q(Z;\epsilon)$$

reveals the relation between conditional mutual information and the correlation integral as the basis for its nonparametric estimation. The unconditional relation is obtained by removing Z. Extensive results on the connection between correlation integrals and information theory are given in Prichard & Theiler (1995).

The choice $q = 2$ is by far the most popular in chaos analysis as it allows for efficient estimation algorithms. But the conditional mutual information $Iq(X, Y|Z)$ is not positive definite for $q \neq 1$. It is thus possible to have variables X and Y which are conditionally dependent given Z, but for which $I_2(X, Y|Z)$ is zero or negative. As noted by Diks & Manzan (2002), if $I_2(X, Y|Z)$ is zero the test based on it does not have unit power asymptotically against conditional dependence. This situation, while exceptional, is quite undesirable. Since $I_2(X, Y|Z)$ is usually either positive or negative a one-sided test rejecting for $I_2(X, Y|Z)$ large, is not optimal. Diks & Manzan (2002) argue that, in practice I_2 behaves much like I_1 in that we usually observe larger power for one-sided tests (rejecting for large I_2) than for two-sided tests. This led them to propose $q = 2$, together with a one-sided implementation of the test.

Following the arguments in Granger et al. (2002) and Maasoumi & Racine (2002) in favor of metric entropies, we choose $K = 1/2$ in the K-class entropy defined above. That is, we propose a test statistic that satisfies the following properties:

1. It is well defined for both continuous and discrete variables.

2. It is normalized to zero if X and Y are independent, and lies between 0 and 1.

3. The modulus of the measure is equal to unity if there is a *measurable* exact (nonlinear) relationship, $Y = g(X)$ say, between the random variables.

4. It is equal to or has a simple relationship with the (linear) correlation coefficient in the case of a bivariate normal distribution.

5. It is metric, that is, it is a true measure of "distance" and not just of divergence.

6. The measure is invariant under continuous and strictly increasing transformations $h(\cdot)$. This is useful since X and Y are independent if and only if $h(X)$ and $h(Y)$ are independent. Invariance is important since otherwise clever or inadvertent transformations would produce different levels of dependence.

This is a normalization of the Bhattacharya-Matusita-Hellinger measure of dependence given by

$$\begin{aligned} S_\rho &= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(f^{\frac{1}{2}} - f_1^{\frac{1}{2}} f_2^{\frac{1}{2}}\right)^2 dx\, dy \\ &= \frac{1}{4} I_{\frac{1}{2}}, \text{ see below.} \end{aligned} \tag{2}$$

where $f = f(x, y)$ is the joint density and $f_1 = f(x)$ and $f_2 = f(y)$ are the marginal densities of the random variables X and Y. If X and Y are independent, this metric will yield the value zero, and is otherwise positive and less than one. For two density functions f_1 and $f_{2,}$ the asymmetric (with respect to f_1) K-class entropy *divergence measure* is:

$$I_k(f_2, f_1) = \frac{1}{k-1}\left[\int 1 - (f_2^k / f_1^k) dF_1\right], \quad k \neq 1,$$

such that $\lim_{k\to 1} I_k(\cdot) = I_1(\cdot)$, the Shannon cross entropy (divergence) measure. When one distribution is the joint, and the other is the product of the marginals, this measure is called the "mutual information" outlined earlier. Once the divergence in both directions (of f_1 *and* f_2) are added, a symmetric measure is obtained which, for $K = 1$, is well known as the Kullback-Leibler measure. We consider the symmetric K-class measure at $K = 1/2$ as follows:

$$\begin{aligned} I_{\frac{1}{2}} &= I_{\frac{1}{2}}(f_2, f_1) + I_{\frac{1}{2}}(f_1, f_2) \\ &= 2M(f_1, f_2) \\ &= 4B(f_1, f_2), \end{aligned}$$

where $M(\cdot) = \int (f_1^{\frac{1}{2}} - f_2^{\frac{1}{2}})^2 \, dx$ is known as the Matusita *distance*, and,

$$B(\cdot) = 1 - \rho^*$$

is known as the Bhattacharya *distance* with

$$0 \leq \rho^* = \int (f_1 f_2)^{\frac{1}{2}} \leq 1$$

being a measure of "affinity" between the two densities.

$B(\cdot)$ and $M(\cdot)$ are rather unique among measures of divergence since they satisfy the triangular inequality and are, therefore, proper measures of *distance*. Other *divergence* measures are capable of characterizing desired null hypotheses (such as independence) but may not be appropriate when these distances are compared across models, sample periods, or agents. These comparisons are often made, and more often implicit in inferences. See also the discussion by Diks & Manzan (2002) summarized above.

Note that when $f(x, y) = N(0, 0, 1, 1, \rho)$ and $f(x) = N(0, 1) = f(y)$,

$$\begin{aligned} \rho^* &= \frac{(1-\rho^2)^{\frac{5}{4}}}{(1-\frac{\rho^2}{2})^{\frac{3}{2}}} \\ &= 1 \text{ if } \rho = 0 \\ &= 0 \text{ if } \rho = 1. \end{aligned}$$

S_ρ is used to measure and test the degree of dependence present in time-series data. We employ kernel estimators of $f(y, y_{-j})$, $f(y)$, and $f(y_{-j})$, $j = 1, 2, \ldots, K$ originally proposed by Parzen (1962) (see Appendix A for implementation details). For testing the null of serial independence, the null distribution of the kernel-based implementation of S_ρ is obtained via a resampling approach identical to that used for the "Robust BDS" test described above (see Granger et al. (2002) for details).

R Code for computing the entropy metric and for computing the Robust BDS test (Ihaka & Gentleman (1996)[2]) is available from the authors upon

[2] See `http://www.r-project.org`

request.

In the next section we compare the relative performance of the TR, "Robust BDS", and the S_ρ entropy-based tests on the basis of their diagnostic ability in an empirical setting. All tests are correctly-sized due mainly to their reliance on resampling procedures, therefore relative performance boils down to a comparison of power. We note that it is known that BDS is a test of the hypothesis $E[f(X,Y)] - E[f(X)] \times E[f(Y)] = 0$, whereas mutual information tests are concerned with the expected divergence between $f(x,y)$ and $f(x) \times f(y)$ (relative to $f(x,y)$). The latter are one-to-one representations of independence and imply the null of concern to BDS, but not vice versa.

3 Stock Return Dynamic Model Specification

3.1 Data Sources

We use the data series found in Chen & Kuan (2002), who apply their TR characteristic function-based test to residuals from a variety of models of daily returns of six major US stock indices. The indices are the Dow Jones Industrial Averages (DJIA), New York Stock Exchange Composite (NYSE), Standard and Poor's 500 (S&P500), National Association of Securities Dealers Automated Quotations Composite (NASDAQ), Russel 2000 (RS2000), and Pacific Exchange Technology (PE-TECH). Each series contains $T = 2,527$ observations running from January 1 1991 through December 31 2000, and we let $Y_t = 100 \times (\log P_t - \log P_{t-1})$ denote the daily return of the index P_t.

3.2 Assessing Dependence in a Raw Data Series

We begin by considering whether or not there exists potential nonlinear dependence in the raw series themselves. We therefore compute our metric entropy $\hat{S}_\rho$ using $k = 1, 2, \ldots, 5$ lags for each raw series. Bandwidths were selected via least-squares cross-validation, and the Gaussian kernel was used throughout. We then construct P-values for the hypothesis that each series is a serially independent white-noise process. Results are summarized in Table 1.

Table 1: P-values from the entropy-based test for serial independence at various lags on the raw data series.

Series	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
DJIA	0.00	0.00	0.00	0.00	0.00
NASDAQ	0.00	0.00	0.00	0.00	0.00
NYSE	0.00	0.00	0.00	0.00	0.00
PETECH	0.00	0.00	0.00	0.00	0.00
RS2000	0.00	0.00	0.00	0.00	0.00
S&P500	0.01	0.00	0.00	0.00	0.00

As can be seen from Table 1, there is extremely strong and robust evidence in favor of dependence being present in all of these series. Given this expected finding, we proceed along the lines of Chen & Kuan (2002) who assess the suitability of two classes of popular time-series models which have been used to model such processes.

3.3 Models

As in Chen & Kuan (2002), we consider two models, the GARCH(p,q) (Bollerslev (1986)) and EGARCH(p,q) (Nelson (1991)) specifications for a time series $Y_t|\Psi_{t-1} = \epsilon_t \sim N(0,h_t)$ which we now briefly outline. The GARCH(p,q) model may be expressed as

$$\text{GARCH}(p,q): \quad h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j}, \tag{3}$$

where $p \geq 0$, $q > 0$, $\omega > 0$, $\alpha_i \geq 0$, and $\gamma_j \geq 0$, while the EGARCH(p,q) model may be written as

$$\begin{gathered}\text{EGARCH}(p,q): \quad \ln(h_t) = \omega + \sum_{i=1}^{q} \alpha_i g(z_{t-i}) + \sum_{j=1}^{p} \gamma_j \ln(h_{t-j}), \\ \text{where } g(z_t) = \theta z_t + \gamma[|z_t| - E|z_t|] \text{ and } z_t = \epsilon_t/\sqrt{h_t}.\end{gathered} \tag{4}$$

3.4 Goodness of Fit

We first consider the question of whether these nonlinear models differ in terms of their goodness of fit for a given series as these models are often considered equal on the basis of "goodness of fit" criteria. However, goodness of fit criteria such as R^2 are correlation-based. Out of concern that the 'equality' of models might be an artifact of using correlation-based measures of fit, we therefore compute the entropy measure with $f = f(\hat{Y}_t, Y_t)$ as the joint density of the predicted and actual excess returns, and $f_1 = f(\hat{Y}_t)$ and $f_2 = f(Y_t)$ as the respective marginal densities. If predicted and actual returns are independent, this metric will yield the value zero, and will increase as the model's predictive ability improves.

In order to determine whether or not two model's measures of goodness of fit differ statistically, we require the sampling distribution of the goodness of fit measure itself. To obtain the percentiles for our goodness of fit statistic we employed the stationary bootstrap of Politis & Romano (1994) and 99 bootstrap replications. Results are summarized in Table 2.

Given results summarized in Table 2, it is evident that there does not appear to be any significant difference between models in terms of their fidelity to the data for a given series. This common finding leads naturally to residual-based specification testing to which we now proceed.

3.5 Model Specification

Next, we focus on using S_ρ as a constructive model diagnostic tool. Under the null of correct model specification, the model residuals would be indistinguishable from white noise at all lags. Table 3 reports the associated P-values from the entropy-based test for serial independence at various lags on time-series models' residuals using 99 permutation replications, results for the Ljung-Box Q test, those for the modified BDS test, and those for Chen & Kuan's (2002) TR test.

Table 3 reveals that the correlation-based Q statistic almost uniformly fails to have power in the direction of misspecification for both the GARCH and EGARCH specifications. The Robust BDS test performs relatively bet-

Table 2: $\hat{S}_\rho$ measures of goodness of fit along with their resampled 90% interval estimates.

Series	$\hat{S}_\rho$	$[pct_5, pct_{95}]$
DJIA [EGARCH11]	0.07	[0.06, 0.15]
DJIA [EGARCH1k]	0.06	[0.05, 0.16]
DJIA [GARCH11]	0.07	[0.06, 0.13]
DJIA [GARCH1k]	0.07	[0.06, 0.14]
NASDAQ [EGARCH11]	0.12	[0.09, 0.25]
NASDAQ [EGARCH1k]	0.12	[0.09, 0.29]
NASDAQ [GARCH11]	0.12	[0.10, 0.23]
NASDAQ [GARCH1k]	0.12	[0.09, 0.21]
NYSE [EGARCH11]	0.06	[0.05, 0.09]
NYSE [EGARCH1k]	0.05	[0.05, 0.09]
NYSE [GARCH11]	0.06	[0.05, 0.09]
NYSE [GARCH1k]	0.06	[0.05, 0.10]
PETECH [EGARCH11]	0.16	[0.14, 0.19]
PETECH [EGARCH1k]	0.16	[0.13, 0.19]
PETECH [GARCH11]	0.16	[0.14, 0.20]
PETECH [GARCH1k]	0.16	[0.14, 0.21]
RS2000 [EGARCH11]	0.08	[0.05, 0.14]
RS2000 [EGARCH1k]	0.07	[0.05, 0.14]
RS2000 [GARCH11]	0.08	[0.06, 0.14]
RS2000 [GARCH1k]	0.08	[0.06, 0.15]
S&P500 [EGARCH11]	0.08	[0.06, 0.10]
S&P500 [EGARCH1k]	0.07	[0.06, 0.11]
S&P500 [GARCH11]	0.08	[0.06, 0.10]
S&P500 [GARCH1k]	0.07	[0.06, 0.10]

ter, though relative to the TR test which does quite well, the BDS also fails to have power in a number of instances. The TR test performs quite well, though we note that it too lacks power for several cases of EGARCH1k. In particular, Chen & Kuan (2002), on the basis of their reversibility tests, conclude that expanded EGARCH specifications are correct, further noting that "the [proposed] test detects volatility asymmetry that cannot be detected by the BDS test [...] providing more information regarding how a model should be refined" (Chen & Kuan (2002, pg 577)). Table 3 reveals that the correlation-based Q test, the chaos-based BDS test, and characteristic function-based TR test uniformly fail to reject the EGARCH1k specification across series. In contrast to this, however, the entropy-based test detects misspecification across EGARCH1k models for every series at lags 1 and/or 2 at the 5% level except the NASDAQ (though the appropriateness of this model is rejected at lag 2 at the 10% level.). We conclude that the failure to detect this misspecification by competing tests merely reflects their lack of power relative to the entropy-based test in finite-sample settings. Relative to its peers, the entropy-based test has two features to recommend its use as a diagnostic tool:

1. It has higher power than competing correlation-based and characteristic-function-based tests in finite-sample settings

2. It clearly indicates where models fail (e.g. at lags 1 and 2) hence provides prescriptive advice for model refinement.

4 Conclusion

We consider the application of a metric entropy for detecting departures from serial independence to aid in the construction of parametric time-series models. Applications indicate the the approach not only offers *constructive* prescriptions for model specification, but does not suffer from drawbacks found in numerous tools employed in the field.

Table 3: Test results from various tests for serial independence at various lags on time-series models' residuals. For the $\hat{S}_\rho$ and Q-tests we present P-values for the null of correct specification, while for the BDS and TR tests we present the actual statistics and flag those values which are significant with an asterisk. Therefore, for comparison purposes, all entries which are significant at the $\alpha = 0.05$ level are marked with an asterisk (i.e. P-values and actual statistics). Throughout we use k to denote lag, and n to denote embedding dimension.

	S_ρ Entropy Test					Q Test	BDS Test				TR Test				
Series	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=10$	$n=1$	$n=2$	$n=3$	$n=4$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
GARCH11															
DJIA	0.00*	0.00*	0.00*	0.41	0.59	0.17	-0.55	-0.07	-0.21	-0.12	-3.19*	-3.28*	-0.99	-1.16	-1.26
NASDAQ	0.00*	0.00*	0.49	0.82	0.30	0.98	0.68	1.90	1.88	2.05*	-5.52*	-4.15*	-2.11*	0.36	-2.02*
NYSE	0.00*	0.00*	0.00*	0.52	0.01*	0.06	-1.29	-0.49	-0.69	-0.42	-2.61*	-4.65*	-0.72	-1.77	-1.81
PETECH	0.00*	0.00*	0.05*	0.88	0.46	0.44	-0.62	0.19	0.25	0.42	-4.45*	-3.59*	-2.40*	-0.76	-1.25
RS2000	0.00*	0.01*	0.13	0.51	0.58	0.66	2.55*	3.68*	3.69*	3.89*	-3.23*	-3.26*	-2.56*	0.32	-0.83
S&P500	0.00*	0.00*	0.07	0.34	0.01*	0.28	-1.69	-0.81	-0.79	-0.30	-3.38*	-4.65*	-0.89	-1.91	-2.59*
GARCH1k															
DJIA	0.00*	0.00*	0.00*	0.23	0.47	0.14	-1.20	-1.49	-1.61	-1.32	-2.93*	-2.80*	-0.72	-1.03	-1.29
NASDAQ	0.00*	0.00*	0.19	0.72	0.54	0.96	-0.09	0.25	0.20	0.18	-5.38*	-3.81*	-2.23*	0.45	-1.86*
NYSE	0.00*	0.00*	0.01*	0.31	0.03*	0.05*	-1.41	-1.55	-1.72	-1.32	-2.75*	-4.43*	-0.46	-1.38	-1.65
PETECH	0.00*	0.00*	0.05*	0.94	0.31	0.41	-0.85	-0.65	-0.45	-0.04	-4.41*	-3.19*	-2.22*	-0.73	-1.30
RS2000	0.01*	0.01*	0.34	0.10	0.20	0.51	1.77	2.03*	2.09*	2.01*	-3.27*	-3.04*	-2.62*	0.13	-0.37
S&P500	0.00*	0.00*	0.20	0.33	0.00*	0.25	-1.67	-1.75	-1.77	-1.14	-3.63*	-4.06*	-0.72	-1.74	-2.47*
EGARCH11															
DJIA	0.00*	0.25	0.60	0.43	0.85	0.18	-1.04	-0.74	-0.87	-0.79	-1.96*	-1.98*	-0.14	-0.41	-0.54
NASDAQ	0.00*	0.00*	0.65	0.60	0.53	0.99	0.42	1.56	1.55	1.73	-3.70*	-2.72*	-1.03	1.29	-1.27
NYSE	0.00*	0.06	0.12	0.75	0.08	0.12	-1.83	-1.21	-1.41	-1.08	-1.11	-3.27*	-0.31	-0.77	-0.81
PETECH	0.00*	0.00*	0.33	0.32	0.68	0.59	-1.33	-0.66	-0.50	-0.31	-3.09*	-2.06*	-1.28	0.21	-0.26
RS2000	0.01*	0.02*	0.48	0.26	0.84	0.69	2.79*	3.88*	3.89*	4.26*	-1.56	-1.89	-1.96*	1.10	-0.14
S&P500	0.00*	0.00*	0.67	0.61	0.06	0.21	-2.57*	-1.83	-1.90	-1.39	-1.98*	-3.01*	0.27	-0.92	-1.65
EGARCH1k															
DJIA	0.00*	0.29	0.10	0.22	0.73	0.13	-0.83	-0.66	-0.72	-0.56	-0.10	0.40	-0.52	-0.92	-0.83
NASDAQ	0.16	0.06	0.44	0.79	0.58	0.94	-1.00	-0.50	-0.67	-0.57	-0.69	0.06	-1.11	-0.29	-1.07
NYSE	0.01*	0.10	0.04*	0.42	0.05*	0.07	-0.95	-0.74	-0.96	-0.61	0.75	-0.13	-0.22	-1.31	-1.02
PETECH	0.13	0.00*	0.34	0.48	0.19	0.59	-1.18	-0.65	-0.64	-0.48	-1.11	0.42	-0.63	-0.62	-1.09
RS2000	0.02*	0.11	0.57	0.03*	0.66	0.48	1.48	1.47	1.32	1.73	0.79	-0.13	-1.90	-0.95	-1.31
S&P500	0.00*	0.10	0.50	0.57	0.13	0.36	-1.17	-0.96	-1.14	-0.84	-0.51	-0.25	0.65	-2.01*	-0.98

References

Akaike, H. (1981), 'Likelihood of a model and information criteria', *Journal of Econometrics* **16**, 3–14.

Barnett, B., Gallant, A. R., Hinich, M. J., Jungeilges, J. A., Kaplan, D. T. & Jensen, M. J. (1997), 'A single-blind controlled competition among tests for nonlinearity and chaos', *Journal of Econometrics* **82**, 157–192.

Belaire-Franch, J. & Contreras, D. (2002), 'How to compute the BDS test: a software comparison', *Journal of Applied Econometrics* **17**, 691–699.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–27.

Brock, W. A., Dechert, W. D. & Scheinkman, J. A. (1987), 'A test for independence based on the correlation dimension', *University of Wisconsin-Madison Social Systems Reerach Institute Working Paper* (8702).

Brock, W. A., Dechert, W. D., Scheinkman, J. A. & LeBaron, B. (1996), 'A test for independence based on the correlation dimension', *Econometric Reviews* **15**(3).

Chen, Y., Chou, R. & Kuan, C. (2000), 'Testing time reversibility without moment restrictions', *Journal of Econometrics* **95**, 199–218.

Chen, Y. & Kuan, C. (2002), 'Time irreversibility and EGARCH effects in US stock index returns', *Journal of Applied Econometrics* **17**, 565–578.

Cromwell, J. B., Labys, W. C. & Terraza, M. (1994), *Univariate Tests for Time Series Models*, Sage.

Diks, C. & Manzan, S. (2002), 'Tests for serial independence and linearity based on correlation integrals', *Studies in Nonlinear Dynamics and Econometrics* **6**.

Granger, C., Maasoumi, E. & Racine, J. S. (2002), 'A dependence metric for possibly nonlinear time series', *Manuscript* .

Hong, Y. & White, H. (2000), 'Asymptotic distribution theory for nonparametric entropy measures of serial dependence', *Mimeo, Department of Economics, Cornell University, and UCSD* .

Ihaka, R. & Gentleman, R. (1996), 'R: A language for data analysis and graphics', *Journal of Computational and Graphical Statistics* **5**(3), 299–314.

Kanzler, L. (1999), 'Very fast and correctly sized estimation of the BDS statistic', *Christ Church and Department of Economics, University of Oxford* .

Lau, H. T. (1995), *A Numerical Library in C for Scientists and Engineers*, CRC Press, Tokyo.

Ljung, G. & Box, G. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**, 297–303.

Maasoumi, E. & Racine, J. S. (2002), 'Entropy and predictability of stock market returns', *Journal of Econometrics* **107**(2), 291–312.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica* **59**(2), 347–370.

Parzen, E. (1962), 'On estimation of a probability density function and mode', *The Annals of Mathematical Statistics* **33**, 1065–1076.

Politis, D. N. & Romano, J. P. (1994), 'The stationary bootstrap', *Journal of the American Statistical Association* **89**, 1303–1313.

Prichard, D. & Theiler, J. (1995), 'Generalized redundancies for time series analysis', *Physica D* **84**, 476–493.

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley.

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

Skaug, H. & Tjøstheim, D. (1993), Nonparametric tests of serial independence, *in* S. Rao, ed., 'Developments in Time Series Analysis', Chapman and Hall, pp. 207–229.

A A Kernel Implementation of S_ρ

The kernel estimator of the bivariate density of the random variables A and B evaluated at the point (a, b) based upon a sample of observations of size n is given by

$$\hat{f}(a,b) = \frac{1}{nh_a h_b} \sum_{j=1}^{n} K\left(\frac{a - a_j}{h_a}, \frac{b - b_j}{h_b}\right),$$

while that for the marginal densities evaluated at the points a and b are given by

$$\hat{f}(a) = \frac{1}{nh_a} \sum_{j=1}^{n} K\left(\frac{a - a_j}{h_a}\right), \quad \hat{f}(b) = \frac{1}{nh_b} \sum_{j=1}^{n} K\left(\frac{b - b_j}{h_b}\right),$$

where $K(\cdot)$ is a pth order univariate kernel function and where h_a and h_b are bandwidths. For excellent treatments of kernel density estimation see Silverman (1986) and Scott (1992).

Choice of the kernel imparts properties such as continuity and bias on the resultant estimator. Common choices are bounded kernels such as the Epanechnikov kernel and unbounded kernels such as the Gaussian kernel. It is known that kernel choice is not important, rather, it is choice of the bandwidth that drives the behavior of the resulting estimator. We therefore employ the widely used second-order Gaussian kernel (the product Gaussian kernel is employed for the bivariate estimator), while least-squares cross-validation (Silverman (1986, page 48)) is used for bandwidth selection as this method is known to work well for 'fat-tailed' distributions which often arise in time-series settings.

Replacing the unknown densities in (2) with kernel estimators yields

$$\begin{aligned}\widehat{S}_\rho &= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\sqrt{\hat{f}_1(a,b)} - \sqrt{\widehat{g}(a)}\sqrt{\widehat{h}(b)}\right)^2 da\,db \\ &= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\sqrt{\frac{1}{nh_ah_b}\sum_{j=1}^{n}K\left(\frac{a-a_j}{h_a},\frac{b-b_j}{h_b}\right)}\right. \\ &\qquad \left. - \sqrt{\frac{1}{nh_a}\sum_{j=1}^{n}K\left(\frac{a-a_j}{h_a}\right)}\sqrt{\frac{1}{nh_b}\sum_{j=1}^{n}K\left(\frac{b-b_j}{h_b}\right)}\right)^2 da\,db\end{aligned} \tag{5}$$

Evaluation of this integral is not straightforward. We consider multivariate numerical quadrature for its evaluation using the `tricub()` algorithm of Lau (1995, pg 303) written in the C programming language. For our implementation, data was first re-scaled to lie in the range $\mathrm{S} = [-0.5, 0.5]$. Employing Lau's (1995, pg 303) notation, the vertices used for his tricub() algorithm were $(x_i, y_i) = (2.0, 1.5)$, $(x_j, y_j) = (-2.0, 1.5)$, $(x_k, y_k) = (0.0, -2.5)$, and the relative and absolute errors used were both 1.0e-05[3]. Note that the data for the bivariate density therefore lie in the square (S_2) with vertices $(-0.5, 0.5)$, $(0.5, 0.5)$, $(-0.5, -0.5)$, $(0.5, -0.5)$ which lies exactly at the center and strictly in the interior of the triangle defined by the (x, y) vertices listed above. This algorithm was bench-marked by integrating a bivariate kernel estimator which integrated to 1.000 (that is, we obtained $\int\int \hat{f}(a, b)\, da\, db = 1$ for arbitrary data).

[3]If fewer than 1,000 evaluations of the integrand were computed, the absolute and relative errors were reduced by 1.0e-01 and the integral was re-evaluated.