

A Dependence Metric for Possibly Nonlinear Processes*

C.W. Granger

Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508

E. Maasoumi[†]

Department of Economics
Southern Methodist University
Dallas, TX 75275-0496

J. Racine

Department of Economics & Center for Policy Research
Syracuse University
Syracuse, NY 13244-1020

December 6, 2003

Abstract

A transformed metric entropy measure of dependence is studied which satisfies many desirable properties, including being a proper measure of *distance*. It is capable of good performance in identifying dependence even in possibly nonlinear time series, and is applicable for both continuous and discrete variables. A nonparametric kernel density implementation is considered here for many stylized models including linear and nonlinear MA, AR, GARCH, integrated series and chaotic dynamics. A related permutation test of independence is proposed and compared with several alternatives.

*We would like to thank, without implicating, several referees, the Editor, Timo Terasvirta and seminar participants in the UK, Canada, Australia, Sweden, and the US.

[†]E. Maasoumi is the corresponding author. His contact information is Department of Economics, Southern Methodist University, Dallas, TX 75275-0496, Email: maasoumi@mail.smu.edu, Tel: (214) 768-4298.

1 Introduction

The need to detect and properly measure *association* and *dependence* is an essential task in economic model building and forecasting. Numerous diagnostics, such as the Durbin Watson, Lagrange Multiplier, and runs tests, are used to examine model *residuals* for departure from “independence”, *iid*, reversibility, martingale difference and other properties.

Most commonly used measures of dependence and test statistics are convenient functions of “correlation” which is motivated by *linear* relations involving *continuous* variables and/or Gaussian processes. These measures tend to fail when variables are discrete, or in detection, as when they face nonlinear, or non-Gaussian processes. The currently dominant measures tend to be functions of only one or two moments of the underlying processes. While this has the advantage of simplicity, it can mislead when distinctions between the tail areas and higher order moments are germane. Examples in both economic and finance processes abound; See Hsieh (1989) on foreign exchange rates, Pagan & Schwert (1990), Hong & White (2000), Chen & Kuan (2002), and Maasoumi & Racine (2002) and their references. In macroeconomics, the regime change models of Hamilton (1993), and threshold models of the US macro series in, for example, Perron (1989), successfully compete with linear (possibly integrated) time series hypotheses. Of course, nonlinearity in the cost and production functions is widely acknowledged.

Furthermore, the residuals of empirical economic models may have nonlinearities, heterogeneity, and serial dependence for a multitude of reasons, including unknown forms of misspecification. Recent examples of unexplained nonlinear dependence in the *residuals* of (GARCH and other) models for exchange rates and S&P returns are, Hsieh (1989), Hong & White (2000), Qi (1999) and her references, and Maasoumi & Racine (2002). It is clearly desirable for measures of association and dependence to be robust toward possible (but unknown) nonlinearities and non-Gaussian processes.

The *concept* of statistical independence is well defined in terms of the joint distribution of variables. But for reasons of simplicity and convenience, the traditional criteria tend to measure different implications of independence. Examples of criteria that at least incorporate the divergence of the joint distributions from the product of their marginals include Chan & Tran (1992), who use an L_1 norm, Ahmad & Li (1997) among others, who use the von-Mises (L_2) norm, and

Skaug & Tjøstheim (1996) who use the Hellinger and several other measures¹. Examples of other “well-informed” measures include the moment generating and characteristic functions, as well as many *entropy* functionals developed in information theory. Entropies are defined over the space of *distributions* which form the bases of independence/dependence concepts in both continuous and discrete cases. Entropy is also “dimension-less” as it applies seamlessly to univariate and multivariate contexts.

For these reasons, Shannon’s *mutual information* function has been increasingly utilized in the literature; see Joe (1989), Robinson (1991), Skaug & Tjøstheim (1996), and Granger & Lin (1994). Shannon’s relative entropy and almost all other entropies fail to be “metric”, however, as they violate either symmetry, or the triangularity rule, or both. This means that they are measures of *divergence*, not *distance*. A metric measure would have the additional advantage of allowing multiple comparisons of departures/distances. It would be helpful to motivate and support any proposed measure by examining whether it satisfies certain useful and clearly stated properties such as those discussed in Granger & Lin (1994), in the excellent treatment by Skaug & Tjøstheim (1996), and in this paper. It is also desirable to provide the framework for assessing statistical significance of any proposed measure.

The robustness of *nonparametric* implementation of entropy indices is one of the main reasons for the recent surge in their popularity. The interested reader is directed to Tjøstheim’s (1996) survey on the subject. Also see Delgado (1996), Chan & Tran (1992), Granger & Lin (1994), Hsieh (1989), Skaug & Tjøstheim (1993), Aparicio & Escribano (1999), Hong (1998), and Aparicio (1998), who all report superior performance for nonparametric entropy measures of dependence over the traditional measures.

Motivated by these arguments, we consider a robust nonparametric implementation of a *metric* entropy measure of association and dependence that satisfies several desirable properties. These properties embody the advantages of the proposed method and motivate its use in preference to existing indices. We also provide a framework for assessing the statistical significance of deviations from independence. We will focus on a popular application, measuring serial dependence in a time

¹The BDS test may also be interpreted as an unusual measure of divergence between the moments of the joint and the marginal distributions; see Brock, Deckert, Scheinkman & LeBaron (1996).

series. Of course, one could apply the metric in many settings in which one wishes to measure fit and dependence, for example detecting nonlinear dependence between a variable being predicted and the predictions obtained from various models, thereby serving as a general measure of (nonlinear) goodness-of-fit.

Much like the BDS and other tests in this arena, we do not test for nonlinearity *per se*, rather, we aim to measure and test for the degree of departure from independence. Those interested in applications of entropy to model adequacy tests and related tests for nonlinearity are referred to Hong & White (2000) and the references therein. Our method complements the prevalent approach in empirical econometrics and finance of parametric model search based on tests of specification for the conditional means and/or variances. Hong (1999) and Hong & Lee (2003) follow a similar philosophy in developing comparable tests for serial dependence based on generalized spectral distributions. A robust nonparametric approach as proposed by us provides a sound preliminary basis for an otherwise *a priori* exclusion/inclusion of model classes and processes. It may also help to reduce “data mining” consequences of numerous parametric tests, as well as the dangers of tests of hypotheses within misspecified models.

Additionally, diagnostic testing may be done using our technique to examine the *residuals* of fitted models for “generic” serial dependence, or for association with possibly excluded factors.

In Section 2 we outline desirable properties of any measure of dependence and briefly review the metric entropy measure ‘ S_ρ ’ which is a normalized form of the Bhattacharya-Matusita-Hellinger measure, and in Section 3 we outline a kernel-based implementation of S_ρ which we denote ‘ \hat{S}_ρ ’. Section 4 considers using the metric as the basis for a permutation test for serial independence. Section 5 considers simulations designed to gauge the finite-sample performance of the estimator for a number of popular dependent processes, and presents an application to chaotic series. We offer power comparisons with other tests which complements the earlier work of Granger & Lin (1994) and Skaug & Tjøstheim (1993), Skaug & Tjøstheim (1996). Section 6 concludes.

2 An Ideal Measure of Dependence: S_ρ

Several axioms may embody what we consider as desirable in any measure/index and motivate its use. A measure of functional dependence for a pair of random variables X and Y may be required to satisfy the following six “ideal” 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 (or a maximum) if there is a *measurable* exact (nonlinear) relationship, $Y = m(X)$ 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 $\psi(\cdot)$. This is useful since X and Y are independent if and only if $\psi(X)$ and $\psi(Y)$ are independent. Invariance is important since otherwise clever or inadvertent transformations would produce different levels of dependence.

We consider a normalization of the Bhattacharya-Matusita-Hellinger measure of dependence given by

$$\begin{aligned} S_\rho &= \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(f_1^{1/2} - f_2^{1/2} \right)^2 dx dy \\ &= \frac{1}{2} \int \int \left[1 - \frac{f_2^{1/2}}{f_1^{1/2}} \right]^2 dF_1(x, y), \end{aligned}$$

where $f_1 = f(x, y)$ is the joint density and $f_2 = g(x) \cdot h(y)$ is the product of the marginal densities of the random variables X and Y . The second expression is in a moment form which is often replaced with a sample average, especially for theoretical developments; see the section below on relation to copulas. Importantly, $H_0 : \text{independence } (f_1 = f_2) \iff S_\rho = 0$, otherwise (under H_1)

$S_\rho > 0$. Power and *consistency* of the tests based on consistent estimates of S_ρ arise from this last property.

An impressive body of literature demonstrates the desirable axiomatic properties of entropy measures which is instructive in anticipating the properties of related indices. This literature would also make it clear that choices of indices are not as arbitrary as may seem. To see the relation of our normalized measure to entropy *divergence* measures, consider the k -class entropy family of Havrda & Charvat (1967):

$$\begin{aligned} H_k(f) &= (k-1)^{-1}(1 - Ef^{k-1}), \dots k \neq 1 \\ &= -E \log f, \quad (\text{Shannon's entropy}) \quad \text{for } k = 1, \end{aligned}$$

where E denotes expectation with respect to a distribution f . For any two density functions f_1 and f_2 , the asymmetric (with respect to f_2) k -class entropy *divergence measure* is:

$$I_k(f_2, f_1) = \frac{1}{k-1} \left[\int (f_1^k / f_2^k) dF_2 - 1 \right], \quad k \neq 1,$$

such that $\lim_{k \rightarrow 1} I_k(\cdot) = I_1(\cdot)$, the Shannon relative entropy (divergence) measure. Once the divergence in both directions of f_1 and f_2 are averaged, a symmetric measure is obtained which, for $k = 1$, is well known as the Kullback-Leibler measure. Consider the symmetric k -class measure at $k = \frac{1}{2}$ as follows:

$$\begin{aligned} I_{1/2} &= I_{1/2}(f_2, f_1) + I_{1/2}(f_1, f_2) = 2M(f_1, f_2) = 4B(f_1, f_2) \\ \text{where } M(\cdot) &= \int (f_1^{1/2} - f_2^{1/2})^2 dx \text{ is known as the Matusita or Hellinger distance, and,} \\ B(\cdot) &= 1 - \rho^* \text{ is known as the Bhattacharya distance with} \\ 0 \leq \rho^* &= \int (f_1 f_2)^{1/2} \leq 1 \text{ being a measure of "affinity" between the two densities.} \end{aligned}$$

Note that $S_\rho = B(\cdot) = \frac{1}{2}M(\cdot) = \frac{1}{4}I_{1/2}$. $B(\cdot)$ 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 inde-

pendence) but may not be appropriate when these distances are compared across models, sample periods, or agents. Such comparisons are often made, albeit implicitly, in routine inferences. See Hirschberg, Maasoumi & Slottje (2001) for an example in cluster analysis where “distances” are meaningful.

The measure S_ρ satisfies properties 1-3. Property 5 is not difficult to establish, and property 6 was established by Skaug & Tjøstheim (1996) for the Hellinger measure. As for property 4, we note that when $f_1(x, y) = N(0, 0, 1, 1, \rho)$ and $g(x) = N(0, 1) = h(y)$,

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

2.1 Relation to Copula

Consider the monotonic ‘probability integral transformations’ $U = G(X)$ and $V = H(Y)$ which are standard uniformly distributed variables, and with X and Y as in our moment form definition of S_ρ above. The copula, $C(u, v) = F^*(x, y)$, is a joint distribution function, serves as a measure of dependence, and is unique for continuous variables. If it is twice differentiable (and the marginals are once differentiable), we have:

$$f^*(x, y) = \frac{\partial^2 C(u, v)}{\partial u \partial v} g(x) h(y). \quad (1)$$

Since our measure is invariant under such monotonic transformations as $G(\cdot)$ and $H(\cdot)$, and since $\partial^2 C(u, v) / \partial u \partial v = c(u, v)$ is a density with uniform marginals, we may verify the following relations:

$$\begin{aligned} S_\rho &= \frac{1}{2} \int \int \left[1 - \frac{g^{1/2}(x) h^{1/2}(y)}{f^{*1/2}(x, y)} \right]^2 dF(x, y) \\ &= \int \int \left[1 - c^{1/2}(u, v) \right] du dv, \end{aligned}$$

The study of copulas for convenient characterizations of dependence has received increasing attention in finance and other fields; See Nelsen (1999) or Genest & MacKay (1986).

3 A Kernel Density Implementation

We consider using S_ρ to measure the degree of dependence present in time-series data. We employ kernel estimators of the densities involved originally proposed by Parzen (1962). A second-order Gaussian kernel is used throughout, and bandwidths are obtained via likelihood cross-validation (Silverman (1986, page 52)) which produces “optimal” density estimators according to the Kullback-Leibler criterion.

Replacing the unknown densities in S_ρ with kernel estimators yields

$$\hat{S}_\rho = \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\sqrt{\hat{f}_1(a, b)} - \sqrt{\hat{g}(a)} \sqrt{\hat{h}(b)} \right)^2 da db,$$

and multivariate numerical quadrature was used for computing the double integral (Lau (1995, pg 303)). For discrete processes, one simply replaces the relevant integrals in the equation above with the summation operator and replaces $\hat{f}(\cdot)$, $\hat{g}(\cdot)$, and $\hat{h}(\cdot)$ with the kernel probability estimators in Li & Racine (2003). Proofs of the asymptotic properties of this implementation are available from the authors upon request.

A few words regarding the application of S_ρ to potentially nonstationary processes are in order. Application of this method to potentially nonstationary processes raises the question of interpretation of nonparametric density estimators. Phillips & Park (forthcoming) have examined the properties of kernel estimators for nonstationary density estimation demonstrating that the kernel estimator remains meaningful as a type of density estimate even in the nonstationary case. The estimate tells us how dense the process is about a particular (spatial) point and *in this sense* can be interpreted as a form of ‘density’ estimator²; See also Karlsen & Tjøstheim (2001). In nonstationary settings our measure is therefore to be interpreted as assessing the affinity of the processes in terms of their ‘denseness’ rather than their densities.

²Strictly speaking, $\hat{f}(x)$ estimates the local time spent in the immediate vicinity of a point that is determined by the value of x .

4 A Permutation Test for Serial Independence

In this section we outline the use of S_ρ for testing serial independence against alternatives of dependence which can be of a general and nonlinear nature. A limiting normal distribution for the statistic \hat{S}_ρ was established in Granger, Maasoumi, and Racine (2002) who also examined its asymptotic power properties. But there are good reasons to expect that this approximation would be poor, while in addition the outcomes of asymptotic-based kernel tests tend to be quite sensitive to the choice of bandwidth. Skaug & Tjøstheim (1993), Skaug & Tjøstheim (1996) have also studied this issue in the context of testing for serial independence. They note that the asymptotic variance is very poorly estimated in the case of the Hellinger and mutual information measures, which combined with various competing methods of bandwidth selection, renders asymptotic inferences quite unreliable. These same reasons suggest that bootstrapping “asymptotically pivotal” statistics may not perform well in this context.

A permutation test approach can be used to test the simple null of serial independence and would be expected to be robust to the underlying distribution; see Efron & Tibshirani (1993). By applying a random shuffle (permutation) to the dataset at hand one can generate replications which are serially independent having marginal distributions identical to the original data. Randomly reordering the data leaves the marginal distributions intact while generating an independent bivariate distribution. This reshuffle can be used to recompute the statistic using data generated under the null, and this can be repeated a large number of times to generate the empirical null distribution of the statistic which can be used to compute finite-sample critical values. Proofs of consistency of this test are available from the authors upon request. Extensive simulations, also available from the authors upon request, reveal that the permutation-based test has valid size and possesses power in the direction of numerous interesting alternative directions. A comparison of permutation tests and bootstrap tests in this setting can be found in Skaug & Tjøstheim (1996).

5 Applications

5.1 Finite-Sample Behavior

We present some evidence on the finite-sample performance of this kernel-based implementation, and we use many non-linear models and simulations including those of Granger & Lin (1994) as our benchmark. The traditional indices are known to fail in certain instances, sometimes very badly indeed, whereas the new measure is seen to be successful in detecting dependence, and often, revealing the correct dynamic structure.

Granger & Lin (1994) considered the following data generating processes (DGPs) with $\epsilon_t \sim iid, N(0, 1)$.

$$\text{Model 1 } y_t = \epsilon_t + 0.8\epsilon_{t-1}^2$$

$$\text{Model 2 } y_t = \epsilon_t + 0.8\epsilon_{t-2}^2$$

$$\text{Model 3 } y_t = \epsilon_t + 0.8\epsilon_{t-3}^2$$

$$\text{Model 4 } y_t = \epsilon_t + 0.8\epsilon_{t-1}^2 + 0.8\epsilon_{t-2}^2 + 0.8\epsilon_{t-3}^2$$

$$\text{Model 5 } y_t = |y_{t-1}|^{0.8} + \epsilon_t$$

$$\text{Model 6 } y_t = \text{sign}(y_{t-1}) + \epsilon_t$$

$$\text{Model 7 } y_t = 0.8y_{t-1} + \epsilon_t$$

$$\text{Model 8 } y_t = y_{t-1} + \epsilon_t$$

$$\text{Model 9 } y_t = 0.6\epsilon_{t-1}y_{t-2} + \epsilon_t$$

$$\text{Model 10 } y_t = 4y_{t-1}(1 - y_{t-1}) \quad \text{for } t > 1, 0 < y_1 < 1.$$

For the simulations that follow we augment Granger & Lin's (1994) DGPs with the following.

$$\text{Model 0 } y_t = \epsilon_t$$

$$\text{Model 11 } y_t = \sqrt{h_t}\epsilon_t, h_t = 0.01 + 0.94h_{t-1} + 0.05y_{t-1}^2.$$

Models 1-4 are nonlinear MA processes of order 1, 2, 3, and 3 respectively. We expect a good measure to exhibit the theoretical properties of these MA processes which require zero “dependence”

at lags beyond their nominal orders. Models 5-7 are AR(1) autoregressions with various decaying memory properties. Model 8 is a simple I(1) process with persistent memory, and Model 9 is bilinear with white noise characteristics. Model 10 is the logistic function generating chaotic dynamics. Granger & Lin (1994) found the usual correlation function measures to be inadequate in recognizing nonlinear relationships. They found that the relative entropy did very well, and Kendall's τ did well for MA processes but not for the AR models or models 9-10. A directly relevant measure, a portmanteau version of the Hellinger index over a number of lags, was shown by Skaug & Tjøstheim (1996) to do very well indeed for ARCH (1), GARCH (1), nonlinear MA, an Extended nonlinear MA, and Threshold Autoregressive of order 1 (TAR(1)). They showed that the correlation function measures, such as Ljung & Box (1978), can fail to detect the dependence and/or the order of the lags. This failure has been observed by others in a variety of settings.

The AR(1) Model 6 has been further studied in Granger & Terasvirta (1999). The process is Markovian and stationary. Its theoretical autocorrelations should decline exponentially, as would also be expected by a linear stationary AR(1) process. Granger & Terasvirta (1999) observed that the usual autocorrelation measures can point to a fractionally integrated process, indicating long memory where a short memory process is appropriate. The important nonlinear/switching regime behavior of this process is lost to correlation measures.

Model 11 is a GARCH (1,1) process commonly employed to represent the errors in financial applications. Its coefficients were taken from an empirical model of S&P weekly returns obtained by Hong & White (2000), in conjunction with an AR(3) model for the mean.

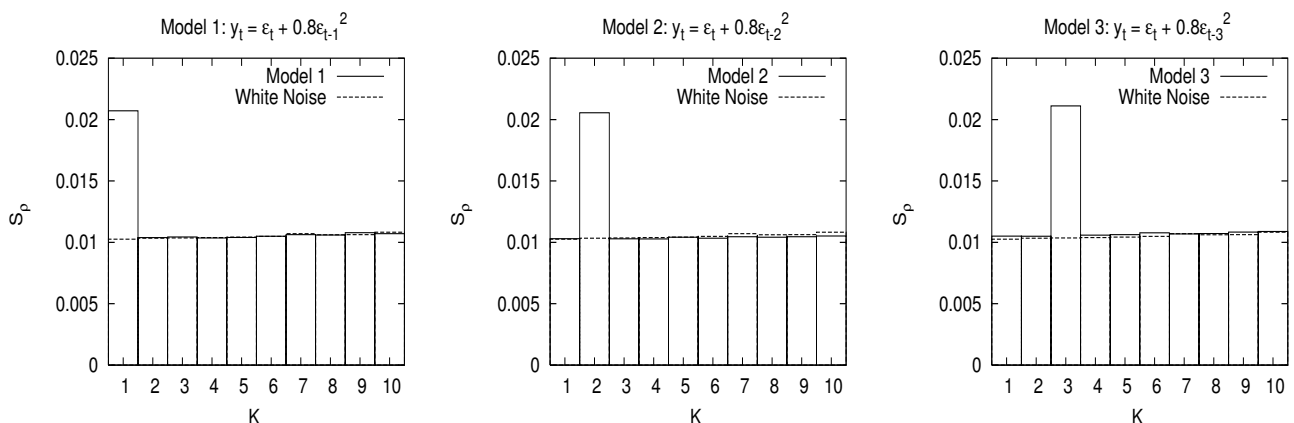
We add Model 0, $y_t = \epsilon_t$, a simple *iid* white noise Gaussian process, as a benchmark. As well, the sampling distribution of Model 0 provides one method of obtaining critical values for the proposed test of independence outlined in Section 4.

We shall use these models to evaluate the performance of our dependence metric in finite samples. A minimum of 1,000 Monte Carlo replications from each model are computed, and $K = 10$ lags are considered. Code was written in the C programming language. Random number generation employed the portable random number routines `ran1` and `gasdev` which use three linear congruential generators and the Box-Muller method found in Press, Flannery, Teukolsky & Vetterling

(1990, pages 210, 216). Likelihood cross-validation was used *for each replication* for selection of the bandwidths.

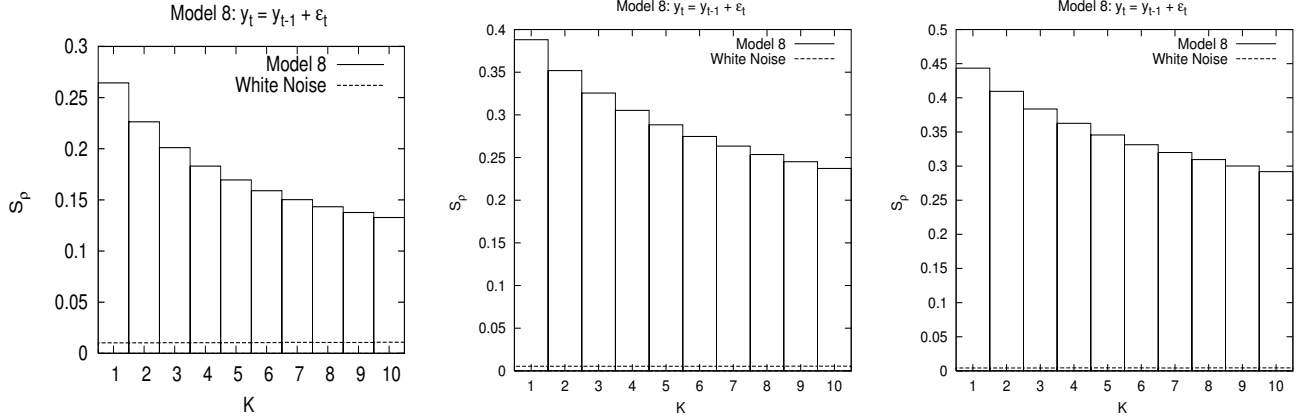
We have undertaken an extensive range of experiments for models 0 through 11, and have considered sample sizes of $n = 50, 75, 100, 150, 200, 300, 400$, and 500. We present results of the *average* value of the \hat{S}_ρ statistic for *each lag* for a given model, with the average computed over the total number of Monte Carlo replications. This is therefore analogous to a sample autocorrelation function for linear time-series models. Following Granger & Lin (1994) we tabulate the mean and standard deviation for each lag and model. In addition, the distribution of the statistic is skewed and bounded below by zero, therefore the median and interquartile ranges are also tabulated. We also consider the empirical distribution of the statistic for the *iid* white noise process which will be useful for determining significant deviations of S_ρ from zero, the theoretical value of S_ρ for an *iid* white noise process. For this last process we tabulate the 90th, 95th, and 99th percentiles from the empirical distribution of S_ρ based upon the Monte Carlo replications.

By way of example, consider the simulation results for $n = 100$ for models 1-3, the nonlinear MA processes of order 1, 2, 3 presented in the following three figures. Each figure plots the average value of \hat{S}_ρ over 1,000 Monte Carlo replications for each DGP for lags 1 through 10 along with the average value of \hat{S}_ρ for a white noise process.



The statistic clearly detects dependence in each *nonlinear* process and it does so often at the correct lag. In addition, the statistic does not behave differently from that for an *iid* process for the remaining lags.

In order to examine the behavior of \hat{S}_ρ as n increases, consider the results for Model 8 for $n = 100, 300$, and 500 (the leftmost figure is for $n = 100$, middle for $n = 300$, and the rightmost for $n = 500$).



We also summarize the central tendency and dispersion of \hat{S}_ρ for Model 8 for $n = 100, 300$, and 500 in the following tables. The columns represent the lag, mean, median, standard deviation, and interquartile range in that order.

Model 8 $y_t = y_{t-1} + \epsilon_t$					Model 8 $y_t = y_{t-1} + \epsilon_t$					Model 8 $y_t = y_{t-1} + \epsilon_t$				
Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{iqr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{iqr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{iqr}}$
1	0.264	0.261	0.098	0.147	1	0.388	0.393	0.091	0.121	1	0.443	0.443	0.084	0.122
2	0.226	0.219	0.099	0.150	2	0.352	0.354	0.094	0.129	2	0.409	0.406	0.088	0.131
3	0.201	0.192	0.100	0.154	3	0.326	0.326	0.098	0.138	3	0.384	0.379	0.093	0.144
4	0.183	0.171	0.100	0.153	4	0.305	0.303	0.100	0.144	4	0.363	0.358	0.096	0.148
5	0.170	0.155	0.100	0.155	5	0.288	0.287	0.103	0.149	5	0.346	0.340	0.099	0.150
6	0.159	0.142	0.100	0.154	6	0.275	0.274	0.104	0.153	6	0.331	0.325	0.100	0.152
7	0.150	0.131	0.100	0.153	7	0.263	0.262	0.106	0.152	7	0.320	0.312	0.102	0.155
8	0.143	0.121	0.099	0.153	8	0.253	0.251	0.107	0.154	8	0.309	0.301	0.104	0.160
9	0.138	0.115	0.098	0.152	9	0.245	0.241	0.108	0.158	9	0.300	0.293	0.106	0.163
10	0.133	0.109	0.098	0.147	10	0.237	0.231	0.109	0.158	10	0.292	0.282	0.107	0.168

Some important improvements occur as n increases. First, as n increases the values for \hat{S}_ρ for the *iid* series decrease and approach zero as expected. Second, for processes where there indeed exists dependence at the K th lag, \hat{S}_ρ increases as n increases. These two features are particularly useful for detecting exactly which lags are significant as n grows large.

Since they are completely representative, and for brevity, we present the remaining simulation

results (only for $n = 100$) in Appendix A. Other results are available from the authors upon request.

A larger set of tests, including the portmanteau version of the \hat{S}_ρ test (the sum of the computed statistic at a finite number of lags), were studied by Skaug & Tjøstheim (1996). These included the portmanteau versions of the Shannon relative entropy, a test due to the authors (absolute difference of the two densities), the ACF test, the BDS, and the van der Waerden test. Their power results at the 95% level, and for $n = 250$, may be summarized as follows:

The \hat{S}_ρ test and the BDS test were often comparable except that results for BDS at higher lags than 3 were not made available because of computational issues. BDS was well beaten by the \hat{S}_ρ test for the TAR model. Indeed the \hat{S}_ρ test and the relative entropy (the best here) beat other tests rather handily for this model, especially at higher lags. The van der Waerden test was often marginally better for the ARCH, GARCH, Nonlinear MA and Extended Nonlinear MA models considered by the authors. The ACF and, surprisingly, the Spearman’s rank test did quite badly in almost all of the nonlinear cases. Power of all tests declines with the lag order, and is generally best at the “correct lag”. The picture that emerges is that the \hat{S}_ρ metric, as well as the relative entropy, is always close to the best tests (when not the best), and maintains very high power for processes that can be missed by tests that are powerful in other situations. The power property of the \hat{S}_ρ test thus appears quite robust to nonlinear model classes, an important characteristic in a world of specification uncertainty. These results also confirm serious failings of the traditional measures. The good power performance of the \hat{S}_ρ measure in detecting memory structure/lags is also notable and gives rise to an expectation that it may form a suitable basis for constructive specification searches.

5.2 Power Comparisons of the Permutation Test with Existing Tests (BDS and Ljung-Box)

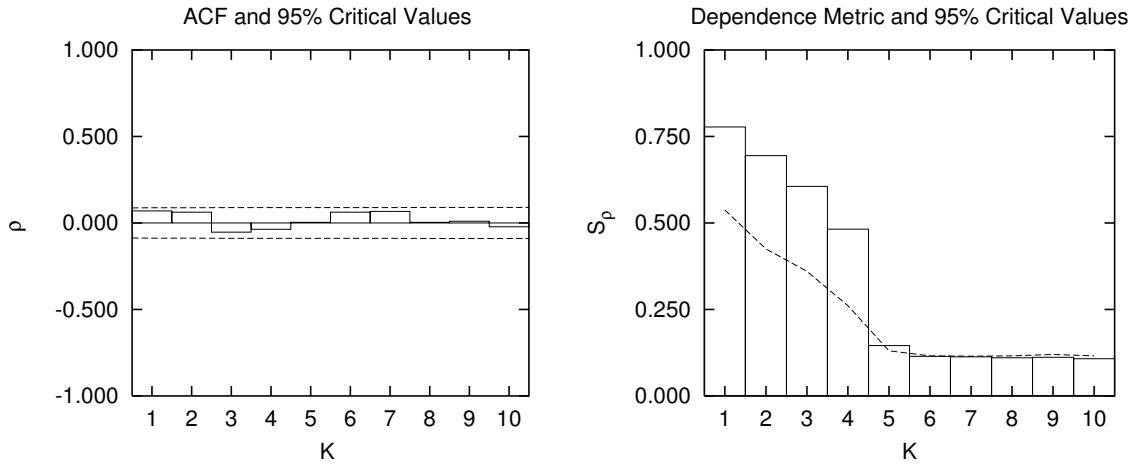
We now compare the finite-sample power of our proposed test with two widely-used tests for serial independence, the BDS and Ljung-Box tests. The Ljung-Box test is correlation-based, while the BDS test has its origins in the deterministic chaos literature; see Brock et al. (1996). We shall

consider testing $H_0 : x_t$ is serially independent, $t = 1, 2, \dots, T$, and by way of example, we consider the logistic map (Model 10), a deterministic chaotic process. Given the origins of the BDS test, it seems fitting to proceed with this model.

A few words on the size of the BDS test are in order. The BDS statistic has a limiting $N(0, 1)$ null distribution. However, we too have observed that, even in large samples, the asymptotic critical values provide a test that has invalid size (see Maasoumi & Racine (2002) for examples). It has been suggested that one *not* use the asymptotic-based test for samples less than $n = 500$. We find the preferred approach to be a permutation-based version of the test which indeed has correct size. We therefore use the more favorable (to BDS) permutation-based version of the test³. In addition to the test's size issues, the BDS requires the user to set the embedding dimension (m) and the dimension distance (ϵ), the choice of which can rather dramatically affect the test's power.

First, we consider one draw from a logistic map for a sample of size $n = 500$. We construct the Ljung-Box, BDS-permutation, and S_ρ -permutation tests for nominal size $\alpha = 0.05$. We graph the ACF and S_ρ for lags $k = 1, 2, \dots, 10$ along with their pointwise critical values (dotted line) under H_0 in Figure 1. Table 1 presents the BDS test statistic for recommended ranges of m and ϵ highlighting values which are significant at the $\alpha = 0.05$ level.

Figure 1: The ACF and S_ρ test statistics for one draw from the logistic map ($n = 500$).



³This will also permit direct power comparisons between our proposed permutation-based test and the size-adjusted i.e. permutation-based BDS test.

Table 1: The BDS test statistic for one draw from the logistic map ($n = 500$) for a recommended range of values of m and ϵ . Entries marked with an asterisk are significant at the $\alpha = 0.05$ level using permutation-based (and asymptotic-based) critical values.

m	ϵ	
	0.36 ($\hat{\sigma}_x$)	0.53 ($1.5\hat{\sigma}_x$)
2	175.03*	8.23*
3	149.44*	2.17
4	143.36*	1.98
5	134.22*	0.89
6	129.01*	0.01
7	132.96*	-0.48
8	138.91*	-0.45

Based upon Figure 1 we observe that, for this sample of $n = 500$ drawn from the logistic map, the ACF fails to detect this (nonlinear) alternative and would lead one to conclude falsely that this series is serially independent⁴. Examining Table 1, we see that the BDS test gives conflicting results depending on one's choice of m and ϵ . One often encounters advice on setting ϵ equal to 1 or 1.5 times the series' standard error, and setting m in the range $2, \dots, 8$, but we can observe that the outcome of the test hinges crucially on this choice: one can fail to reject or reject for a range of m depending on one's choice of ϵ . The S_ρ test, though, shows no such weakness.

A more serious comparison of power can only be obtained via Monte Carlo simulation. We therefore consider a modest simulation again using this DGP in order to more convincingly compare the power of the three tests. We draw 1,000 replications from Model 10, construct each test as described above, and tabulate empirical rejection frequencies over the 1,000 Monte Carlo replications. As we implement the permutation-based version of the BDS test, it is valid to apply the test for small samples, thus we conduct a power comparison of the test for a sample size of $n = 50$. Results are summarized in tables 2 and 3, and as the Ljung-Box test has power that does not differ significantly from size, we omit these results for space considerations. We also include the permutation-based portmanteau version of the \hat{S}_ρ test mentioned earlier, denoted $\sum_{i=1}^k \hat{S}_{\rho,i}$.

We observe that, for samples that are extremely small by kernel estimator standards, our proposed test has high power rejecting the null of serial independence at lag $k = 1$ with power close

⁴This has been noted by Granger & Lin (1994, page 379).

Table 2: Empirical rejection frequency for the BDS test, $\alpha = 0.05$, $n = 50$.

m	$0.5\sigma_x$	$1.0\sigma_x$	$1.5\sigma_x$	$2.0\sigma_x$
2	0.798	0.540	0.222	0.746
3	0.766	0.464	0.170	0.642
4	0.750	0.340	0.182	0.552
5	0.776	0.296	0.172	0.498
6	0.776	0.284	0.154	0.470
7	0.788	0.222	0.088	0.378
8	0.750	0.198	0.056	0.330

Table 3: Empirical rejection frequency for the S_ρ test, $\alpha = 0.05$, $n = 50$.

Metric	$K = 1$	$K = 2$	$K = 3$	$K = 4$	$K = 5$
\hat{S}_ρ	0.920	0.790	0.444	0.204	0.086
$\sum_{i=1}^k \hat{S}_{\rho,i}$	0.920	0.936	0.876	0.722	0.476

to unity. The BDS test, however, has power ranging from 0.058 for $m = 8$ and $\epsilon = 1.5\hat{\sigma}_x$ to 0.798 for $m = 2$ and $\epsilon = 0.5\hat{\sigma}_x$.

6 Conclusions

We believe that the proposed metric S_ρ shows strong promise as a general statistic which can be used to detect generic association and serial dependence present in a time-series. A new white-noise test is proposed, and applications to nonlinear time-series demonstrate the value added by the proposed approach relative to traditional measures. Further work on the constructive specification search applications of this measure is in progress, as is its utility in more general testing for causality and exogeneity.

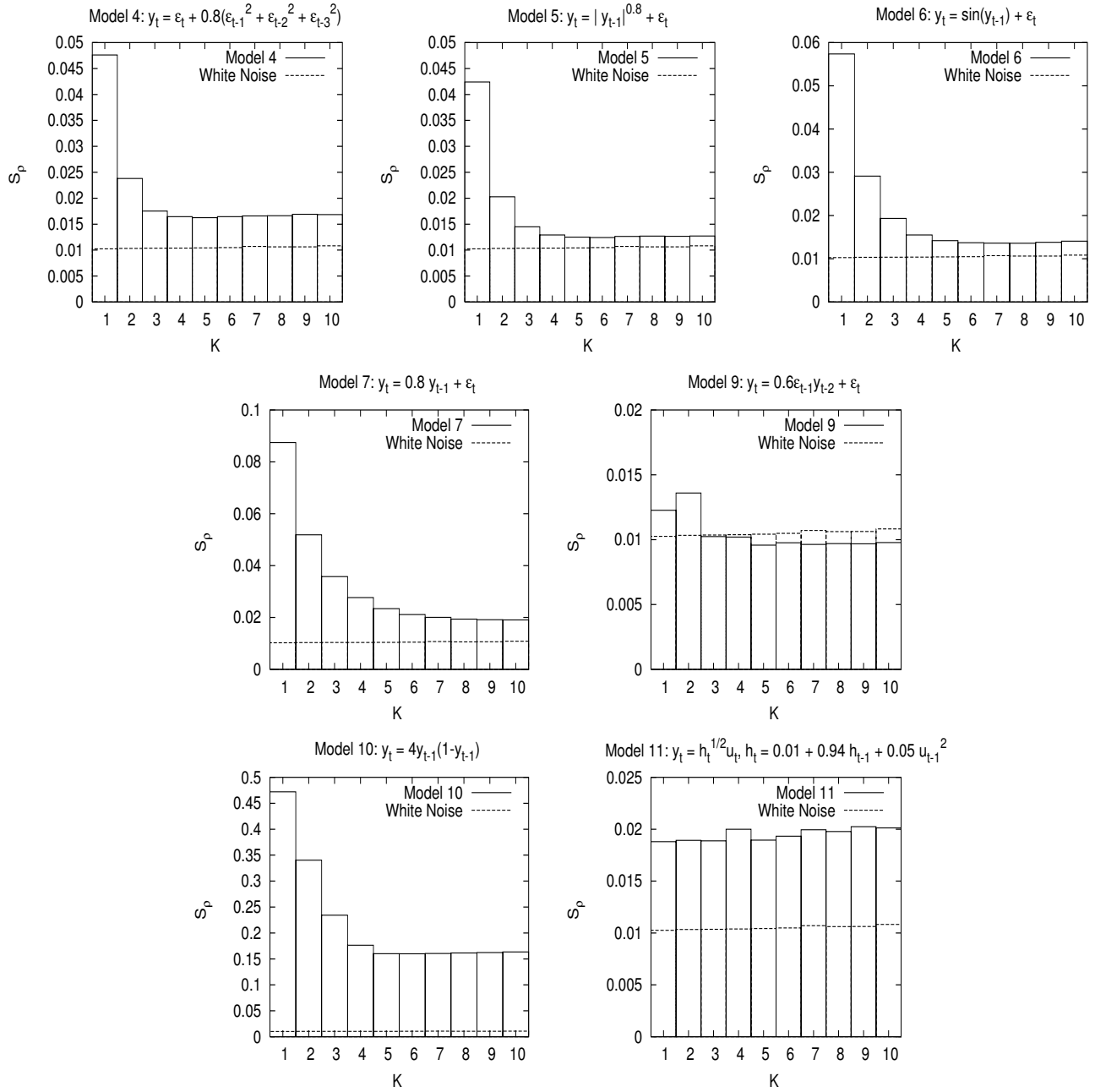
References

- Ahmad, I. & Li, Q. (1997), ‘Testing independence by nonparametric kernel method’, *Statistics and Probability Letters* **34**, 201–210.
- Aparicio, F. (1998), ‘Testing independence by nonparametric kernel method’, *Statistics and Probability Letters* **34**, 201–210.
- Aparicio, F. & Escribano, A. (1999), ‘Information-theoretic analysis of serial dependence and cointegration’, *Studies in Nonlinear Dynamics and Econometrics* **3**(3), 119–140.

- Brock, W., Deckert, W., Scheinkman, J. & LeBaron, B. (1996), ‘A test for independence based on the correlation dimension’, *Econometric Reviews* **15**(3), 197–236.
- Chan, H. & Tran, T. L. (1992), ‘Nonparametric tests for serial dependence’, *Journal of Time Series Analysis* **13**(1), 169–175.
- Chen, Y. & Kuan, C. (2002), ‘Time irreversibility and EGARCH effects in US stock index returns’, *Journal of Applied Econometrics* **17**, 565–578.
- Delgado, M. (1996), ‘Testing serial independence using the sample distribution function’, *The Journal of Time Series Analysis* **17**(3), 271–285.
- Efron, B. & Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York, London.
- Genest, C. & MacKay, R. J. (1986), ‘The joy of copulas: Bivariate distribution with uniform marginals’, *The American Statistician* **40**, 280–283.
- Granger, C. & Lin, J. L. (1994), ‘Using the mutual information coefficient to identify lags in nonlinear models’, *The Journal of Time Series Analysis* **15**(4), 371–384.
- Granger, C. W. & Terasvirta, T. (1999), ‘A simple nonlinear time series model with misleading linear properties’, *Economics Letters* .
- Hamilton, J. D. (1993), Estimation, inference, and forecasting of time series subject to changes in regime, in G. Madalla, C. Rao & H. Vinod, eds, ‘Handbook of Statistics, Volume 11’, North Holland.
- Havrda, J. & Charvat, F. (1967), ‘Quantification method of classification processes: concept of structural α -entropy’, *Kybernetika Císlo I. Ročník* **3**, 30–34.
- Hirschberg, D., Maasoumi, E. & Slottje, D. (2001), ‘Clusters of attributes and well-being in the US’, *Journal of Applied Econometrics* **61**(3), 445–460.
- Hong, Y. (1998), ‘Testing for pairwise serial independence via the empirical distribution function’, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **60**(2), 429–453.
- Hong, Y. (1999), Testing serial independence via the empirical characteristic function, Technical report, Department of Economics, Cornell University, Ithaca, NY USA.
- Hong, Y. & Lee, T. W. (2003), ‘Diagnostic checking for the adequacy of nonlinear time series models’, *Econometric Theory* **19**(6), 1065–1121.
- Hong, Y. & White, H. (2000), ‘Asymptotic distribution theory for nonparametric entropy measures of serial dependence’, *Mimeo, Department of Economics, Cornell University, and UCSD* .
- Hsieh, D. (1989), ‘Testing for nonlinear dependence in foreign exchange rates: 1974-1983’, *Journal of Business* **62**, 339–68.
- Joe, H. (1989), ‘Relative entropy measures of multivariate dependence’, *Journal of the American Statistical Association* **84**, 157–164.

- Karlsen, H. A. & Tjøstheim, D. (2001), ‘Nonparametric estimation in null recurrent time series’, *The Annals of Statistics* **29**(2), 372–416.
- Lau, H. T. (1995), *A Numerical Library in C for Scientists and Engineers*, CRC Press, Tokyo.
- Li, Q. & Racine, J. (2003), ‘Nonparametric estimation of distributions with categorical and continuous data’, *Journal of Multivariate Analysis* **86**(2), 266–292.
- Ljung, G. & Box, G. (1978), ‘On a measure of lack of fit in time series models’, *Biometrika* **65**, 297–303.
- Maasoumi, E. & Racine, J. (2002), ‘Entropy and predictability of stock market returns’, *Journal of Econometrics* **107**, 291–312.
- Nelsen, R. (1999), *An Introduction to Copulas*, Springer-Verlag.
- Pagan, A. & Schwert, G. (1990), ‘Alternative models for conditional stock volatility’, *Journal of Econometrics* **45**, 267–90.
- Parzen, E. (1962), ‘On estimation of a probability density function and mode’, *The Annals of Mathematical Statistics* **33**, 1065–1076.
- Perron, P. (1989), ‘The great crash, the oil price shock, and the unit root hypothesis’, *Econometrica* **57**, 1361–14001.
- Phillips, P. C. B. & Park, J. Y. (forthcoming), ‘Nonstationary density estimation and kernel autoregression’, *Econometric Theory*.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1990), *Numerical recipes in C*, Cambridge University Press, New York.
- Qi, M. (1999), ‘Nonlinear predictability of stock returns using financial and economic variables’, *Journal of Business and Economic Statistics* **17**(4), 419–429.
- Robinson, P. M. (1991), ‘Consistent nonparametric entropy-based testing’, *Review of Economic Studies* **58**, 437–453.
- 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.
- Skaug, H. & Tjøstheim, D. (1996), Testing for serial independence using measures of distance between densities, in P. Robinson & M. Rosenblatt, eds, ‘Athens Conference on Applied Probability and Time Series’, Springer Lecture Notes in Statistics, Springer.
- Tjøstheim, D. (1996), ‘Measures and tests of independence: A survey’, *Statistics* **28**, 249–284.

A \hat{S}_ρ versus K : $n = 100$



B Mean, Median, Standard Deviation, and Interquartile Range:

$n = 100$

Model 0 $y_t = \epsilon_t$					Model 1 $y_t = \epsilon_t + 0.8\epsilon_{t-1}^2$					Model 2 $y_t = \epsilon_t + 0.8\epsilon_{t-2}^2$				
Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$
1	0.010	0.007	0.014	0.005	1	0.021	0.017	0.016	0.012	1	0.010	0.008	0.010	0.005
2	0.010	0.007	0.014	0.005	2	0.010	0.008	0.011	0.005	2	0.021	0.016	0.014	0.012
3	0.010	0.007	0.013	0.005	3	0.010	0.008	0.011	0.006	3	0.010	0.008	0.009	0.005
4	0.010	0.007	0.013	0.005	4	0.010	0.008	0.011	0.005	4	0.010	0.008	0.009	0.006
5	0.010	0.007	0.014	0.005	5	0.010	0.008	0.012	0.005	5	0.010	0.008	0.010	0.006
6	0.010	0.007	0.014	0.005	6	0.010	0.008	0.011	0.006	6	0.010	0.008	0.010	0.005
7	0.011	0.007	0.014	0.005	7	0.011	0.008	0.012	0.006	7	0.010	0.008	0.009	0.006
8	0.011	0.007	0.014	0.005	8	0.011	0.008	0.012	0.006	8	0.010	0.008	0.009	0.005
9	0.011	0.007	0.014	0.005	9	0.011	0.008	0.012	0.006	9	0.010	0.008	0.010	0.006
10	0.011	0.007	0.014	0.005	10	0.011	0.008	0.012	0.006	10	0.011	0.008	0.010	0.006

Model 3 $y_t = \epsilon_t + 0.8\epsilon_{t-3}^2$					Model 4 $y_t = \epsilon_t + 0.8(\epsilon_{t-1}^2 + \epsilon_{t-2}^2 + \epsilon_{t-3}^2)$					Model 5 $y_t = y_{t-1} ^{0.8} + \epsilon_t$				
Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$
1	0.011	0.008	0.011	0.005	1	0.048	0.043	0.024	0.024	1	0.042	0.037	0.024	0.023
2	0.010	0.008	0.011	0.005	2	0.024	0.020	0.017	0.014	2	0.020	0.016	0.017	0.014
3	0.021	0.017	0.016	0.012	3	0.018	0.013	0.015	0.011	3	0.014	0.010	0.015	0.009
4	0.011	0.008	0.011	0.006	4	0.016	0.012	0.015	0.010	4	0.013	0.009	0.015	0.007
5	0.011	0.008	0.011	0.006	5	0.016	0.013	0.014	0.010	5	0.013	0.008	0.015	0.007
6	0.011	0.008	0.011	0.006	6	0.016	0.013	0.015	0.010	6	0.012	0.008	0.015	0.007
7	0.011	0.008	0.011	0.005	7	0.017	0.013	0.015	0.010	7	0.013	0.008	0.015	0.007
8	0.011	0.008	0.011	0.006	8	0.017	0.013	0.015	0.010	8	0.013	0.008	0.015	0.007
9	0.011	0.008	0.011	0.006	9	0.017	0.013	0.015	0.010	9	0.013	0.008	0.015	0.007
10	0.011	0.008	0.011	0.006	10	0.017	0.013	0.015	0.010	10	0.013	0.008	0.015	0.007

Model 6 $y_t = \text{sign}(y_{t-1}) + \epsilon_t$					Model 7 $y_t = 0.8y_{t-1} + \epsilon_t$				
Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$
1	0.057	0.051	0.028	0.028	1	0.087	0.080	0.037	0.044
2	0.029	0.024	0.021	0.019	2	0.052	0.046	0.029	0.034
3	0.019	0.015	0.018	0.013	3	0.036	0.030	0.025	0.027
4	0.016	0.011	0.016	0.010	4	0.028	0.022	0.022	0.022
5	0.014	0.010	0.016	0.009	5	0.023	0.017	0.020	0.018
6	0.014	0.009	0.016	0.008	6	0.021	0.015	0.019	0.016
7	0.014	0.009	0.016	0.008	7	0.020	0.014	0.019	0.014
8	0.014	0.009	0.016	0.008	8	0.019	0.014	0.018	0.013
9	0.014	0.009	0.016	0.008	9	0.019	0.014	0.018	0.013
10	0.014	0.009	0.017	0.008	10	0.019	0.014	0.018	0.013

Model 9 $y_t = 0.6\epsilon_{t-1}y_{t-2} + \epsilon_t$					Model 10 $y_t = 4y_{t-1}(1 - y_{t-1})$					Model 11 $y_t = h_t^{1/2}\epsilon_t$				
Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$	Lag	\hat{S}_ρ	$\hat{S}_{\rho_{med}}$	$\hat{\sigma}_{\hat{S}_\rho}$	$\hat{S}_{\rho_{igr}}$
1	0.012	0.009	0.012	0.007	1	0.472	0.469	0.066	0.084	1	0.019	0.012	0.020	0.017
2	0.014	0.010	0.012	0.008	2	0.340	0.333	0.117	0.113	2	0.019	0.012	0.022	0.017
3	0.010	0.007	0.010	0.005	3	0.234	0.227	0.087	0.121	3	0.019	0.011	0.021	0.016
4	0.010	0.007	0.010	0.005	4	0.176	0.164	0.089	0.118	4	0.020	0.012	0.023	0.017
5	0.010	0.007	0.010	0.005	5	0.160	0.148	0.076	0.099	5	0.019	0.012	0.020	0.015
6	0.010	0.007	0.010	0.005	6	0.160	0.149	0.072	0.093	6	0.019	0.012	0.021	0.018
7	0.010	0.007	0.010	0.005	7	0.161	0.150	0.071	0.094	7	0.020	0.012	0.022	0.017
8	0.010	0.007	0.010	0.005	8	0.161	0.151	0.071	0.094	8	0.020	0.012	0.022	0.017
9	0.010	0.007	0.010	0.005	9	0.162	0.152	0.072	0.096	9	0.020	0.013	0.022	0.017
10	0.010	0.007	0.010	0.005	10	0.163	0.152	0.076	0.095	10	0.020	0.012	0.022	0.017