A Dependence Metric for Nonlinear Time Series*

C.W. Granger
Department of Economics
UC San Diego
La Jolla, CA 92093-0508

E. Maasoumi†
Department of Economics
Southern Methodist University
Dallas, TX 75275-0496

J. Racine
Department of Economics
University of South Florida
Tampa, FL 33620-5500

January 20, 2000

Abstract

A transformed metric entropy measure of dependence is studied which satisfies several desirable properties and is capable of impressive performance in identifying nonlinear dependence in time series. The measure is applicable for both continuous and discrete variables. A nonparametric kernel density implementation is considered here for ten models including MA, AR, integrated series and chaotic dynamics.

*We thank participants at our seminars in Canada, USA, Sweden and elsewhere. Timo Terasvirta provided especially valuable comments and references.

†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) 768-4298/740-1802.

1 Introduction

The most widely used measures of dependence are geared towards linear relations and continuous variables. Correlation-based indices are common and are defined over the realizations of the random variables. Entropy-based indices are defined over the actual distributions which are the bases of independence/dependence concepts, and more fully represent their underlying variables, both continuous and discrete. For example, the relative entropy measure based on Shannon's entropy function has been increasingly utilized in the literature (see Joe (1989), Robinson (1991), Skaug & Tjøstheim (1996), and Granger & Lin (1994)). But relative entropy, and almost all other entropies, fail to be "metric" as they violate the triangularity rule. We wish to consider a metric entropy measure which also satisfies several other desirable properties stated in Granger & Lin (1994).

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 -1 and +1.

3. The modulus of the measure should equal unity if there is an exact (nonlinear) relationship, $X = g(Y)$ say, between the 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 $h(.)$. This is useful since X and Y are independent if and only if $h(X)$ and $h(Y)$ are independent.

Granger & Maasoumi (1993) considered a normalization of the Matusida-Bhattacharya-Hellinger measure of dependence given by

$$\begin{aligned} S_\rho &= 1 - \rho^* \\ &= \frac{1}{4} I_{\frac{1}{2}} \\ &= \frac{1}{2} M(f, f_1, f_2) \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(f^{\frac{1}{2}} - f_1^{\frac{1}{2}} f_2^{\frac{1}{2}} \right)^2 \, da \, db \end{aligned} \tag{1}$$

where $f = f(a, b)$ is the joint density and $f_1 = f(a)$ and $f_2 = f(b)$ are the marginal densities of the random variables A and B, and where $0 \leq \rho^* \leq \int\int (f \times f_1 f_2)^{\frac{1}{2}} \, da \, db \leq 1$. The measure is one half of the Hellinger (last equality) and the Matusida $M(.)$ measures which are non-negative.

S_ρ satisfies properties 1-3. Property 5 is established in the literature (see Maasoumi (1993)), and property 6 was established by Skaug & Tjøstheim (1996) for the Hellinger measure. As for property 4, we note that when $f(x, y) = N(0, 0, 1, 1, \rho)$ and $f(x) = N(0, 1) = f(y)$,

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

We consider using S_ρ to measure the degree of dependence present in time-series data. To this end we employ a Nadaraya-Watson (Nadaraya (1965), Watson (1964)) kernel estimator of $f(y, y_{-j})$, $f(y)$, and $f(y_{-j})$, $j = 1, 2, \ldots, K$. We investigate the performance of this kernel-based implementation, and we use the many non-linear models and simulations of Granger & Lin (1994) as our benchmark. The traditional, correlation-based measures fail, sometimes very badly indeed, whereas our measure is very successful in detecting dependence, and often, revealing the dynamic structure.

2 Kernel Estimators of Densities

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

$$\hat{f}(a,b) = \frac{1}{nh_a h_b} \sum_{j=1}^{n} K\left(\frac{a-a_j}{h_a}, \frac{b-b_j}{h_b}\right) \tag{3}$$

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

$$\hat{f}(a) = \frac{1}{nh_a} \sum_{j=1}^{n} K\left(\frac{a-a_j}{h_a}\right), \quad \hat{f}(b) = \frac{1}{nh_b} \sum_{j=1}^{n} K\left(\frac{b-b_j}{h_b}\right) \tag{4}$$

where $K()$ is a pth order univariate kernel function and where h_a and h_b are bandwidths.

2.1 Kernel Choice

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. We employ the widely used second-order Gaussian kernel, and the product Gaussian kernel is employed for the bivariate estimator.

2.2 Bandwidth Selection

In applied settings the bandwidth is typically chosen via data-driven approaches such as plug-in or cross-validatory methods. The bandwidth choice drives the behavior of the resultant estimator far more than the kernel choice.

For univariate and bivariate density estimators based upon a second order kernel, the optimal bandwidths[1] are given by $h_i = c_i \sigma_i n^{-1/5}$ and $h_i = c_i \sigma_i n^{-1/6}$, $i = a, b$, respectively. The constants c_i are unknown and are functions of the density being estimated, and data-driven methods can be employed to determine their likely values. For what follows, these constants are obtained via likelihood cross-validation (Silverman (1986, page 52)). Given that the same vector of data is used to construct our marginal and joint densities, common values of both c and σ are appropriate throughout and are used to obtain the bandwidths for both the marginal and joint densities when suitably normalized. The common values employed are those for the density of the entire sample of data, $\{y_1, \ldots, y_n\}$, and σ is replaced with its sample analog.

3 Kernel-Based Evaluation of $\widehat{S}_\rho$

Replacing the unknown densities in $S\rho$ with kernel estimators yields

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

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 $[-0.5, 0.5]$. 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[2]. Note that

[1] Those which minimize integrated mean square error of the estimator.

[2] 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.

the data for the bivariate density therefore lie in the square 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).

4 Simulation

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

$$
\begin{aligned}
&\text{Model 1 } y_t = \epsilon_t + 0.8\epsilon_{t-1}^2 \\
&\text{Model 2 } y_t = \epsilon_t + 0.8\epsilon_{t-2}^2 \\
&\text{Model 3 } y_t = \epsilon_t + 0.8\epsilon_{t-3}^2 \\
&\text{Model 4 } y_t = \epsilon_t + 0.8\epsilon_{t-1}^2 + 0.8\epsilon_{t-2}^2 + 0.8\epsilon_{t-3}^2 \\
&\text{Model 5 } y_t = |y_{t-1}|^{0.8} + \epsilon_t \\
&\text{Model 6 } y_t = \text{sign}(y_{t-1}) + \epsilon_t \\
&\text{Model 7 } y_t = 0.8y_{t-1} + \epsilon_t \\
&\text{Model 8 } y_t = y_{t-1} + \epsilon_t \\
&\text{Model 9 } y_t = 0.6\epsilon_{t-1}y_{t-2} + \epsilon_t \\
&\text{Model 10 } y_t = 4y_{t-1}(1 - y_{t-1}) \quad \text{for } t > 1, 0 < y_1 < 1
\end{aligned}
$$

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. A portmanteau version of the Hellinger measure 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 can be very misleading.

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 linear measures.

In addition, we add Model 0, $y_t = \epsilon_t$, a simple white noise Gaussian process, as a benchmark. We use these models to evaluate the performance of our dependence metric in finite samples. We consider sample sizes ranging from $n = 50$ through $n = 500$. 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, Flanery, Teukolsky & Vetterling (1990, pages 210, 216).

Simulation results are summarized in appendices B through F. In these appendices we graph the average value of the 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 a white noise process which will be useful for determining significant deviations of S_ρ from zero, the theoretical value of S_ρ for a 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.

5 Hypothesis Testing

We consider using S_ρ for testing serial independence of a time series against alternatives of dependence which can be of a general and nonlinear nature. We require the distribution of this statistic under the null in order to proceed with inference, and consider three common approaches towards obtaining critical values for our test statistic under this null.

5.1 Asymptotic Approximations

One approach is to approximate the null distribution of our statistic based upon asymptotic theory. Skaug & Tjøstheim (1996) obtained this distribution using Robinson's trimming weights (see Robinson (1991)). When these weights are equal to unity, as in our computations, no asymptotic distribution has been established. The trimming technique essentially takes out outliers, however, which is quite sensible in practice. So a normal distribution theory is essentially available for very large samples. But there are good reasons to expect that this would be met with limited success. One quite serious problem with this approach in a kernel context is that the asymptotic-based null distribution would not depend on the bandwidth, while the value of the test statistic depends directly on the bandwidth. This is due in part to the fact that the bandwidth is a quantity which vanishes asymptotically. This is a serious drawback in practice, since the outcome of such asymptotic-based tests tends to be quite sensitive to the choice of bandwidth. This has been noted by a number of authors including Robinson (1991) and Rilstone (1991). For his kernel-based statistic, Robinson (1991) noted that "substantial variability in the [test statistic] across bandwidths was recorded", which would be quite disturbing in applied situations. 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 relative entropy measures, which combined with the problem of window width, renders asymptotic inferences quite unreliable. These same reasons argue against the use of "pivotal statistics" in resampling alternatives.

5.2 Simulated Critical Values

Another approach would be to simulate critical values based upon a white-noise data series for a number of sample sizes. Given that the distribution of our statistic is skewed, one would adopt a percentile approach to obtaining critical values using, for example, the 95th percentile for a given sample size as a critical value.

Appendix A lists the 90th, 95th, and 99th percentiles of the empirical distribution of S_ρ for the process $y_t = \epsilon_t$ with $\epsilon_t \sim i.i.d., N(0,1)$ for a range of sample sizes[3]. These values can be used as the basis for tests of serial independence of a time-series. The alternatives against which this test has power contain general nonlinear dependent processes. The metric S_ρ combined with the percentiles tabulated in Appendix A indicate that the proposed test for serial independence has power against general nonlinear alternatives and that the power increases quite quickly with the sample size as can be seen by examining appendices B through F. Another approach would be to generate the white noise data from a known distribution.

5.3 Resampled Critical Values

Note that, taking the approach of Section 5.2 above, one is testing for serially independent normally distributed data, that is, the null is compound. However, if one uses a resampling approach, then we may test the simple null of serial independence regardless of the underlying distribution. That is, a resampling approach would be expected to be robust to the underlying distribution given that standard regularity conditions on the underlying distribution required for consistency of the kernel estimator are met.

[3]A minimum of 2,000 Monte Carlo replications were computed.

Thus, a third approach would be to resample the test statistic under the null of serial independence. One can generate replications which are serially independent having marginal distributions identical to the original data simply by applying a random shuffle to the dataset at hand. Randomly reordering the data leaves the marginal distributions intact while generating an independent bivariate distribution. This reshuffle is 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 distribution of the statistic under the null. One could then use the empirical distribution of this resampled statistic to compute finite-sample critical values. The null distribution will be that for a given bandwidth and will therefore adapt to bandwidth choice as in Racine (1997).

Extensive simulations of this approach are not undertaken at this time. However, preliminary results suggest that this approach gives values which are comparable to that from Section 5.2 above for the experimental setup contained therein.

5.4 Application - Detecting Dependence for Chaotic Time-Series

We consider testing for serial independence of a sample drawn from the DGP given by Model 10, a chaotic time-series generated by the logistic map. We consider the application of traditional time-series methods versus the proposed metric.

The following graphs plot the autocorrelation function (ACF) ($\hat{\rho}$) and the proposed metric ($S\hat{\rho}$) for an arbitrary sample of size $n = 500$ drawn from this DGP along with critical values under the null of white noise for each[4]. The critical values are given by the dotted lines in each graph. $K = 10$ lags were computed for each.

The logistic map, it turns out, has ACF and partial autocorrelation (PACF) functions which behave like white noise (Granger & Lin (1994,

[4]The critical values for the proposed metric are based upon the percentile approach from Section 5.3. A 5% level of significance was used for both the ACF and the proposed metric.

page 379)). The ACF incorrectly leads us to conclude that there is no dependence in the series, while the proposed metric correctly leads us to conclude strong dependence in the series. The proposed metric has good power in this direction for all sample sizes considered as can be observed upon examining appendices B through F.

This simple example is additional evidence which highlights the limitations of traditional, linear time series measures for detecting serial dependence as well as the value added by using the proposed metric for such purposes.

Our results for the remaining models are in agreement with the findings of Granger & Lin (1994). They confirm serious failings of the correlation-based measures, and a patently strong performance of the S_ρ measure in detecting many types of non-linear dependence. Its good performance for detecting memory structure/lags is also notable and gives rise to an expectation that it may be a suitable basis for constructive specification tests.

6 Conclusion

We believe that the proposed metric S_ρ shows strong promise as a general statistic which can be used to detect general nonlinear dependence present in a time-series. New white-noise tests are proposed, and applications to nonlinear time-series demonstrate the value added by the proposed approach relative to standard time-series 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

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. & Maasoumi, E. (1993), 'A metric entropy measure', *unpublished notes, Department of Economics, UCSD and SMU* .

Granger, C. W. & Terasvirta, T. (1999), 'A simple nonlinear time series model with misleading linear properties', *Economics Letters* .

Joe, H. (1989), 'Relative entropy measures of multivariate dependence', *Journal of the American Statistical Association* **84**, 157–164.

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

Maasoumi, E. (1993), 'A compendium to information theory in economics and econometrics', *Econometric Reviews* **12**, 1–53.

Nadaraya, E. A. (1965), 'On nonparametric estimates of density functions and regression curves', *Theory of Applied Probability* **10**, 186–190.

Press, W. H., Flanery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1990), *Numerical recipes in C*, Cambridge University Press, New York.

Racine, J. (1997), 'Consistent significance testing for nonparametric regression', *Journal of Business and Economic Statistics* **15**(3), 369–379.

Rilstone, P. (1991), 'Nonparametric hypothesis testing with parametric rates of convergence', *International Economic Review* **32**, 209–227.

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.

Watson, G. S. (1964), 'Smooth regression analysis', *Sanikhya* **26:15**, 175–184.