Nonlinear Dynamics (2006) 44: 351–357
DOI: 10.1007/s11071-006-2019-0

Entropy-Based Independence Test

ANDREIA DIONÍSIO[1,*], RUI MENEZES[2], and DIANA A. MENDES[2]
[1]*Centre of Business Studies (CEFAGE-UE) Management Department, University of Évora, Largo dos Colegiais 2, 7000 Évora, Portugal;* [2]*ISCTE, Av. Forcas Armadas, 1649-026 Lisbo, Portugal;* *Author for correspondence (e-mail: andreia@uevora.pt; fax: +00351-266-742-494)*

(Received: 20 September 2004; accepted: 9 June 2005)

Abstract. This paper presents a new test of independence (linear and non-linear) among distributions based on the entropy of Shannon. The main advantages of the presented approach are the fact that this measure does not need to assume any type of theoretical probability distribution and has the ability to capture the linear and non-linear dependencies, without requiring the specification of any kind of dependence model.

Key words: entropy, independence test, mutual information, non-linear serial dependence, stock index markets

1. Introduction

The notion of "independence", "distance" and "divergence" between distributions has been central in statistical inference and econometrics from the earliest stages. This is also evident in the work of Kullback and Leibler, H. Jeffreys, H. Akaike, E. Shannon, Hartley, A. Renyi, E. Maasoumi, C. Granger, H. White, A. Zellner and many others (see e.g. [1]). Some authors, such as Cover et al. [2] and Maasoumi [1], moved by the "elegance" and the potential power of information theory, brought a new way of interpretation and motivation for the research in statistical inference. In addition, the axiomatic systems in information theory suggest principles of decomposition that distinguish between different information functions and "entropies", and identify desirable measures, decision criteria and indices [1].

Several measures have been used as independence tests and/or dependence measures in this field. The most known measure of dependence between random variables is the Pearson correlation coefficient. However, this is nothing but a normalized covariance and only accounts for linear (or linearly transformed) relationships (see e.g. [3, 4]). In general, this statistic may not be helpful to capture serial dependence when there are non-linearities in the data. In this context, it seems that a measure of global dependence is required, that is, a measure that captures both linear and non-linear dependencies without requiring the specification of any kind of model of dependence. Urbach [5] defends a strong relationship between entropy, dependence and predictability. This relation has been studied by several authors, namely Granger and Lin [3], Maasoumi and Racine [4], Darbellay and Wuertz [6]. On the basis of the arguments mentioned earlier, we aim to evaluate in this paper the efficiency of a new entropy-based independence test without requiring the specification of mean-variance models and theoretical distribution probabilities. Thus, in the next section we discuss the subject of information and predictability in the context of entropy, and we then illustrate our test using evidence based on empirical financial data.

2. Information and Predictability

A measure that takes the value 0 when there is total independence and 1 when there is total dependence is one of the most practical ways to evaluate (in)dependence between two vectors of random variables $\mathbf{X}$, $\mathbf{Y}$. Let $p_{\mathbf{X},\mathbf{Y}}(A \times B)$ be the joint probability distribution of $(\mathbf{X}, \mathbf{Y})$ and $p_{\mathbf{X}}(A)$, $p_{\mathbf{Y}}(B)$ the underlying marginal probability distributions, where A is a subset of the observation space of $\mathbf{X}$ and B is a subset of the observation space of $\mathbf{Y}$, such that we can evaluate the following expression:

$$\ln \frac{p_{\mathbf{X},\mathbf{Y}}(A \times B)}{p_{\mathbf{X}}(A) p_{\mathbf{Y}}(B)}. \tag{1}$$

If the two events are independent, then $p_{\mathbf{X},\mathbf{Y}}(A \times B) = p_{\mathbf{X}}(A) p_{\mathbf{Y}}(B)$, and so Equation (1) will be equal to zero.

Granger et al. [4] consider that a good measure of dependence should satisfy the following six "ideal" properties:

1. It must be well defined, both for continuous and discrete variables;
2. It must be normalized to zero if $\mathbf{X}$ and $\mathbf{Y}$ are independent, and, in general, lie between -1 and $+1$;
3. The modulus of the measure should be equal to 1 if there is an exact non-linear relationship between the variables;
4. It must be similar or simply related to the linear correlation coefficient in the case of a bivariate normal distribution;
5. It must be a metric in the sense that it is a true measure of "distance" and not just a measure of "divergence";
6. It must be an invariant measure under continuous and strictly increasing transformations.

Now, consider two vectors of random variables $(\mathbf{X}, \mathbf{Y})$. Let $p_{\mathbf{X}}$, $p_{\mathbf{Y}}$ and $p_{\mathbf{X},\mathbf{Y}}$ be the probability density function (pdf) of $\mathbf{X}$, $\mathbf{Y}$ and the joint probability distribution of $(\mathbf{X}, \mathbf{Y})$, respectively. Denote by $H(\mathbf{X})$, $H(\mathbf{X}, \mathbf{Y})$ and $H(\mathbf{Y}|\mathbf{X})$ the entropy of $\mathbf{X}$, the joint entropy of the two arguments $(\mathbf{X}, \mathbf{Y})$ and the conditional entropy of $\mathbf{Y}$ given $\mathbf{X}$. Then, mutual information can be given defined by the following expression:,

$$\begin{aligned} I(\mathbf{X}, \mathbf{Y}) &= H(\mathbf{X}, \mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \\ &= H(\mathbf{X}) + H(\mathbf{Y}) - H(\mathbf{X}, \mathbf{Y}) \\ &= \iint p_{\mathbf{X},\mathbf{Y}}(x, y) \ln \frac{p_{\mathbf{X},\mathbf{Y}}(x, y)}{p_{\mathbf{X}}(x) p_{\mathbf{Y}}(y)} \, dx \, dy. \end{aligned} \tag{2}$$

Since $H(\mathbf{X}) \geq H(\mathbf{Y}|\mathbf{X})$, we have $I(\mathbf{X}, \mathbf{Y}) \geq 0$, assuming the equality *iff* $\mathbf{X}$ and $\mathbf{Y}$ are statistically independent. Thus, the mutual information between the vectors of random variables $\mathbf{X}$ and $\mathbf{Y}$ can be considered as a measure of dependence between these variables or, even better, the statistical correlation between $\mathbf{X}$ and $\mathbf{Y}$.

The statistic defined in Equation (2) satisfies some of the desirable properties of a good measure of dependence (see e.g. [7]). In Equation (2), we have $0 \leq I(\mathbf{X}, \mathbf{Y}) \leq \infty$, which renders difficult the comparisons between different samples. In this context, Granger and Lin [3], and Darbellay [8], among others, use a standardised measure for the mutual information, the global correlation coefficient, defined by $\lambda(\mathbf{X}, \mathbf{Y}) = \sqrt{1 - e^{-2I(\mathbf{X},\mathbf{Y})}}$. This measure varies between 0 and 1 being thus directly comparable with the linear correlation coefficient r, based on the relationship between the measures of information theory and analysis of variance. The function λ captures the overall dependence, both linear and non-linear, between $\mathbf{X}$ and $\mathbf{Y}$.

According to the properties displayed by mutual information, and because independence is one of the most valuable concepts in econometrics, we can construct a test of independence based on the following hypothesis:

H_0: $p_{\mathbf{X},\mathbf{Y}}(x, y) = p_{\mathbf{X}}(x)p_{\mathbf{Y}}(y)$;
H_1: $p_{\mathbf{X},\mathbf{Y}}(x, y) \neq p_x(x)p_{\mathbf{Y}}(y)$.

If H_0, then $I(\mathbf{X}, \mathbf{Y}) = 0$ and we conclude that there is independence between the variables. If H_1, then $I(\mathbf{X}, \mathbf{Y}) > 0$ and we reject the null hypothesis of independence. The earlier hypothesis can be reformulated as follows:

$$H_0\colon I\,(\mathbf{X}, \mathbf{Y}) = 0; \quad H_1\colon I(\mathbf{X}, \mathbf{Y}) > 0.$$

In order to test adequately for the independence between variables (or vectors of variables) we need to calculate the corresponding critical values. In our case, we have simulated critical values for the null distribution or the percentile approach.[1]

One of the problems with calculating mutual information from empirical data lies in the fact that the underlying pdf is unknown. There are, essentially, three different methods for estimating mutual information: histogram-based estimators; kernel-based estimators; parametric methods. According to Kraskov et al. [9] and Moddemeijer [10], the most straightforward and widespread approach to estimate mutual information consists of partitioning the supports of **X** and **Y** into bins of finite size, i.e. using histogram-based estimators. The histogram-based estimators are divided in two groups: equidistant cells (see e.g. [10]) and equiprobable cells, i.e. marginal equiquantisation (see e.g. [8]). The second approach, marginal equiquantisation, has some advantages, since it allows for a better adherence to the data and maximizes mutual information [8].

The definition of mutual information is expressed in an abstract way and it is based on space partitions. To simplify, let us consider a finite dimension Euclidian space, $\mathbb{R}^d = \mathbb{R}^{d_A} \times \mathbb{R}^{d_B}$, and let $\Gamma_{\mathbf{X}} = \{A_i\}_{i=1}^{n_1}$, $\Gamma_{\mathbf{Y}} = \{B_j\}_{j=1}^{n_2}$ be two generic partitions of the subspaces $\mathbb{R}^{d_{\mathrm{X}}}$ and $\mathbb{R}^{d_{\mathrm{Y}}}$. Then the mutual information is a positive number defined as:

$$I(\mathbf{X}, \mathbf{Y}) \equiv \sup_{\{A_i\}\{B_j\}} \sum_{i,j} P_{\mathbf{X},\mathbf{Y}}(A_i \times B_j) \log \frac{P_{\mathbf{X},\mathbf{Y}}(A_i \times B_j)}{P_{\mathbf{X}}(A_i) \times P_{\mathbf{Y}}(B_j)}. \tag{3}$$

The supremum is taken over all finite partitions of $\mathbb{R}^{d_{\mathrm{X}}}$ and ${}^{d_{\mathrm{Y}}}$. The conventions $0 \ln(0/z) = 0$ for $z \geq 0$ and $z \ln(z/0) = +\infty$ were used. Darbellay [10] shows that mutual information is finite *iif* the measure $p_{\mathbf{X},\mathbf{Y}}$ is absolutely continuous with respect to the product measure $p_{\mathbf{X}}(A_i)p_{\mathbf{Y}}(B_j)$. The system $\Gamma = \Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{Y}}$ is a partition of $\mathbb{R}^d = \mathbb{R}^{d_{\mathrm{X}}} \times \mathbb{R}^{d_{\mathrm{Y}}}$ and is the product of two marginal partitions, one of $\mathbb{R}^{d_{\mathrm{X}}}$ and the other of $\mathbb{R}^{d_{\mathrm{Y}}}$. Marginal equiquantisation consists of dividing each edge of a cell into α ($\alpha = 2$, usually) intervals with approximately the same number of points. The approximativeness of the division has two causes: the number of points in a cell may not be exactly divisible by α, or some **X** may take repeating values (for more details see for example [8]).

[1] These values have been found through the simulation of critical values based on a white noise, for a number of sample sizes. Given that the distribution of mutual information is skewed, we can adopt a percentile approach to obtain critical values. Appendix lists the 90th, 95th and 99th percentiles of the empirical distribution of the mutual information for the process $y_t = \varepsilon_t$ with $\varepsilon_t \sim$ i.i.d. $N(0, 1)$, having been made 5000 simulations for each critical value. This methodology was applied as proposed by Granger et al. [7], and according to these authors, the critical values can be used as the base to test for time series serial independence.

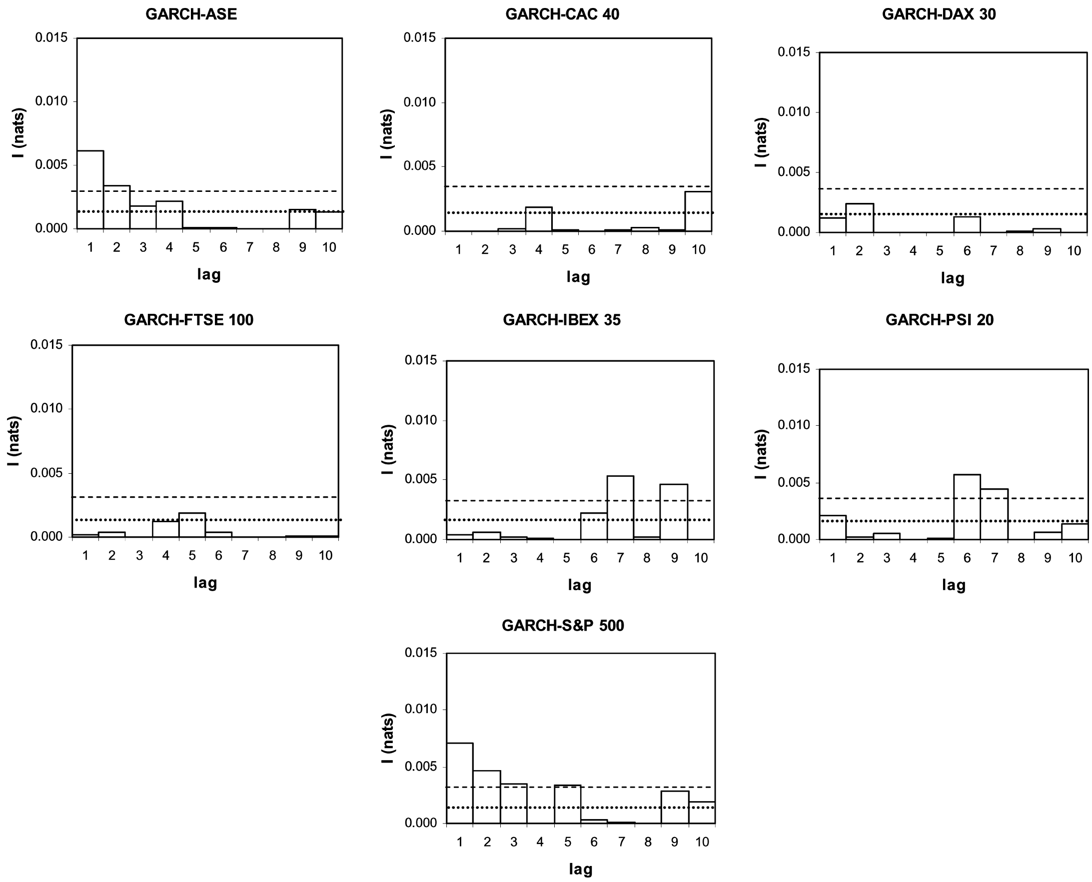

Figure 1. Mutual information for the rate of returns of the filtered *GARCH* effects indexes stocks for the lags $k = 1, \dots, 10$. The dashed line is the 1% critical value (0.0030 nats) and the point line is the 5% critical value for the mutual information (0.0015 nats) with 2 degrees of freedom.

The mutual information and the global correlation coefficient (λ) almost satisfy the property of a good measure of dependence here presented, but they are not measures of "distance", since they do not verify the triangle inequality. Kraskov et al. [11] present a modified mutual information-based measure, such that the resulting is a metric in strict sense. According to these authors, this modification presents some difficulties when we deal with continuous random variables. One solution for this problem consists of dividing mutual information by the sum or by the maximum of dimensions of the continuous variable in study [11].

3. Empirical Evidence

We now apply the concepts of mutual information and global correlation coefficient as measures of dependence in financial time series, in order to evaluate the overall performance of these measures and to extract the advantages of this approach face to the traditional linear correlation coefficient. Mutual information was estimated through marginal equiquantisation, and was applied to a number of stock market indexes.

From the data base DataStream we selected the daily closing prices of several stock market indexes: *ASE* (*Greece*), *CAC 40* (*France*), *DAX 30* (*Germany*), *FTSE 100* (*UK*), *PSI 20* (*Portugal*), *IBEX 35* (*Spain*) and *S&P 500* (*USA*), spanning the period from 4 January 1993 to 31 December 2002, which corresponds to 2596 observations per index, in order to compute the rates of return.

We filtered all the time series with ARMA(p, q) processes in order to eliminate the linear serial dependence. In order to isolate the possible sources of non-linear dependence, namely the heteroscedasticity, we filtered all the time series through GARCH(p, q) processes. The new time series did not reject the null of homocedasticity for the Engle test and the McLeod and Li test. According to the results of the *BDS test* applied to this new time series, the non-linear serial dependence disappear, since the null is not rejected in any case. We calculate the mutual information for the time series filtered from the *GARCH* effects. The results are presented in the Figure 1.

We should note the presence of values of mutual information statistically significant for some lags in all the indexes, denoting the presence of non-linear dependence for those lags.

The results obtained for the mutual information allow us to identify possible lags relatively to which it is necessary to proceed with a detailed analysis, in an attempt to identify the type of non-linearity. It is inferred to here that the sources of captured non-linearity will not be just the existence of non-linearity in the mean and heteroscedasticity. Mutual information does not provide guidance about the type of non-linearity, but it informs about which are the "most problematic" lags and on the level of existent non-linear dependence through the calculation of the global correlation coefficient (λ).

To conclude, we can say that the main advantage of the application of mutual information to financial time series is the fact that this measure captures the global serial dependence (linear and non-linear) without requiring a theoretical probability distribution or specific model of dependency. Even if this dependence is not able to refute the efficient market hypothesis, it is important for investor to know that the rate of return is not independent and identically distributed.

Appendix A

Table of critical values for testing serial independence through mutual information for $N(0, 1)$ data. Five thousand replications were computed. D.F. is the degrees of freedom for the mutual information, which correspond to the dimension (d) of the analysed vectors.

	Percentiles		
D.F	90	95	99
	$N = 100$		
2	0.0185	0.0323	0.0679
3	0.1029	0.1232	0.1933
4	0.1059	0.1260	0.1722
5	0.2290	0.2580	0.3261
6	0.6639	0.7528	0.9663
7	0.8996	0.9731	1.1586
8	1.3384	1.3839	1.5024
9	1.9030	1.9352	2.0142
10	2.5266	2.5571	2.6181

(*Continued on next page*)

	Percentiles		
D.F.	90	95	99
	$N = 500$		
2	0.0037	0.0070	0.0144
3	0.0222	0.0369	0.0501
4	0.0680	0.0788	0.1128
5	0.1756	0.2066	0.2712
6	0.3084	0.3514	0.4390
7	0.4920	0.5391	0.6339
8	0.4477	0.4843	0.5659
9	0.6661	0.6941	0.7594
10	1.0884	1.1082	1.1483
	$N = 1000$		
2	0.0019	0.0041	0.0071
3	0.0133	0.0191	0.0311
4	0.0340	0.0399	0.0568
5	0.0708	0.0865	0.1128
6	0.2119	0.2430	0.3046
7	0.3635	0.3954	0.4688
8	0.4041	0.4414	0.5252
9	0.3865	04114	0.4640
10	0.6418	0.6585	0.6942
	$N = 2000$		
2	0.0009	0.0019	0.0033
3	0.0061	0.0094	0.0147
4	0.0169	0.0203	0.0278
5	0.0701	0.0804	0.1030
6	0.1370	0.1549	0.1940
7	0.2496	0.2733	0.3224
8	0.4497	0.4864	0.5508
9	0.3036	0.3298	0.3858
10	0.3530	0.3669	0.3996
	$N = 2500$		
2	0.0008	0.0015	0.0030
3	0.0054	0.0078	0.0129
4	0.0134	0.0171	0.0251
5	0.0556	0.0648	0.0797
6	0.1203	0.1376	0.1738
7	0.2181	0.2418	0.2884
8	0.3938	0.4217	0.4719
9	0.3175	0.3409	0.4024
10	0.2931	0.3124	0.3477

References

1. Maasoumi, E., ‘A compendium to information theory in economics and econometrics’, *Econometric Reviews* **12**(2), 1993, 137–181.

2. Cover, T. and Thomas, J., *Elements of Information Theory*, Wiley, New York, 1991.
3. Granger, C. and Lin, J., 'Using the mutual information coefficient to identify lags in nonlinear models', *Journal of Time Series Analysis* **15**(4), 1994, 371–384.
4. Maasoumi, E. and Racine, J., 'Entropy and predictability of stock market returns', *Journal of Econometrics* **107**, 2002, 291–312.
5. Urbach, R., *Footprints of Chaos in the Markets – Analysing Non-linear Time Series in Financial Markets and Other Real Systems*, Prentice-Hall, London, 2000.
6. Darbellay, G. and Wuertz, D., 'The entropy as a tool for analysing statistical dependence's in financial time series', *Physica A* **287**, 2000, 429–439.
7. Granger, C., Maasoumi, E., and Racine, J., 'A dependence metric for possibly nonlinear processes', *Journal of Time Series Analysis* **25**(5), 2004, 649–669.
8. Darbellay, G., 'An adaptative histogram estimator for the mutual information', *UTIA Research Report 1889*, Academy of Sciences, Prague, 1998.
9. Kraskov, A., Stogbauer, H., and Grassberger, P., 'Estimating mutual information', preprint in http://www.arxiv:cond-mat/0305641, 2003.
10. Moddenmeijer, R., 'A statistic to estimate the variance of the histogram-based mutual information estimator on dependent pairs of observations', *Signal Processing* **75**, 1999, 51-63.
11. Kraskov, A., Stogbauer, H., Andrzejak, R., and Grassberger, P., 'Hierarchical clustering based on mutual information', preprint in http://www.arxiv:q-bio.QM/0311039, 2004.