ECONOMICS BULLETIN

The relative efficiency of stockmarkets

Ricardo Giglio
Department of Economics, Federal University of Santa Catarina, Brazil

Raul Matsushita
Department of Statistics, University of Brasilia, Brazil

Sergio Da Silva
Department of Economics, Federal University of Santa Catarina, Brazil

Abstract

Financial economists usually assess market efficiency in absolute terms. This is a shortcoming. One way of dealing with the relative efficiency of markets is to resort to the efficiency interpretation provided by algorithmic complexity theory. This paper employs such an approach in order to rank 36 stock exchanges and 37 individual company stocks in terms of their relative efficiency.

SDS acknowledges financial support from the Brazilian agencies CNPq and CAPES-Procad, and RG acknowledges support from CAPES-Procad.
Citation: Giglio, Ricardo, Raul Matsushita, and Sergio Da Silva, (2008) "The relative efficiency of stockmarkets." *Economics Bulletin*, Vol. 7, No. 6 pp. 1-12
Submitted: January 22, 2008. **Accepted:** April 21, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume7/EB-08G10001A.pdf

1. Introduction

If price changes fully incorporate the expectations and information of all market participants such changes are unpredictable, and the market is said to be informationally efficient. Stockmarkets are complex systems in that they convey information about a given stock in its price time series. In an efficient market populated by rational agents if the price is properly anticipated then it must fluctuate randomly. The stochastic process in question is a martingale that is, roughly, a probabilistic model of a fair game, one in which gains and losses cancel each other. This is conventional wisdom in financial economics.

After presenting an overview of market efficiency in their classic financial econometrics textbook, Campbell *et al.* (1997) observed that (p. 24) the notion of *relative* efficiency, i.e. the efficiency of one market measured against another may be a more useful concept than the all-or-nothing (absolute) view taken by much of the traditional market efficiency literature. They made an analogy with physical systems that are usually given an efficiency rating based on the relative proportion of energy converted to work. Rating a piston engine 60% efficient means that on average 60% of the energy contained in the engine's fuel is used to turn the crankshaft, with the remaining 40% lost to other forms of work such as heat, light, or noise. It makes no sense to test statistically whether the engine is perfectly efficient. Similarly, market efficiency is an idealization that is unattainable, but that serves as a useful benchmark for measuring relative efficiency.

Indeed, one must regard the market efficient hypothesis as a limiting case. In practice, prices reflect only the information for which the acquisition costs cannot outweigh the benefits. There are also transaction costs. And information may not be widespread, i.e. there can be inside traders. Following the arrival of new information, market participants may diverge from each other in how they think it will impact prices; in other words, expectations are heterogeneous. Residual inefficiencies are always present in actual markets. These inefficiencies can introduce artificial patterns and thus redundant information in real-world financial price series. Thus it is inappropriate to assess whether a given actual market is efficient or not. This is not a yes-no question; rather, efficiency should measure to what extent one market departs from the idealized efficient market. Relative efficiency is what really matters.

Algorithmic complexity theory makes a connection between the efficient market hypothesis and the unpredictable character of stock returns because a time series that has a dense amount of nonredundant information (such as that of the idealized efficient market) exhibits statistical features that are almost indistinguishable from those observed in a time series that is random (Mantegna and Stanley 2000). As a result, measurements of the deviation from randomness provide a tool to assess the degree of efficiency of a given market. Because algorithmic complexity theory cannot discriminate between trading on noise and trading on information, it detects no difference between a time series conveying a large amount of nonredundant information and a pure random process.

This paper adopts such an approach. Doing so, we will be able to rank both stock exchanges and individual company stocks in terms of their relative efficiency. We will find, for instance, that the S&P 500 is 99.1% efficient whereas the Colombo Stock Exchange of Sri Lanka is only 10.5% efficient. This means that prices in the American stockmarket incorporate much more nonredundant information than its Sri Lankan counterpart does.

The absolute efficiency of stockmarkets has been investigated in a huge number of papers (for a survey see Beechey *et al.* 2000), but we could track only two previous attempts similar to ours to deal with their relative efficiency. Shmilovici *et al.* (2003) provide a test for the efficient market hypothesis (and not exactly for the relative efficiency of stockmarkets) that is based on the insight that compression of the time series of an efficient market is not possible since there are no patterns. In that case, "stochastic complexity" is highest. The stochastic complexity of a time series is a measure of the number of binary digits needed to represent and reproduce the information in the time series. The authors use the Rissanen context tree algorithm to track patterns and then compress the series of both 13 stock exchange indices and the stock prices of the companies listed on the Tel-Aviv 25. (Shmilovici *et al.* claim that the approach in Chen and Tan (1999) is one particular case of theirs.) Section 2 will show our distinct perspective applied to a larger database as well as a simpler methodology that is based straightforwardly on the Lempel-Ziv ("deterministic") complexity index.

The rest of this paper is organized as follows. Section 2 discusses algorithmic complexity theory in more detail. Section 3 presents data and performs analysis. And Section 4 concludes.

2. Lempel-Ziv algorithmic complexity

Shannon entropy of information theory implies that a genuinely random series is the polar case where its expected information content is maximized, in which case there is maximum uncertainty and no redundancy in the series. The algorithmic (Kolmogorov) complexity of a string is given by the length of the shortest computer program that can produce the string. The shortest algorithm cannot be computable, however. Yet there are several ways to circumvent this problem. Lempel and Ziv (1976) suggest a useful measure that does not rely on the shortest algorithm. (Rissanen context tree algorithm of stochastic complexity is another alternative.) Kaspar and Schuster (1987) provide an easily calculable measure of the Lempel-Ziv index which runs as follows.

A program either inserts a new digit into the binary string $S = s_1, \ldots, s_n$ or copies the new digit to S. The program then reconstructs the entire string up to the digit $s_r < s_n$ that has been newly inserted. Digit s_r does not come from the substring $s_1, \ldots, s_{r-1}$; otherwise, s_r could simply be copied from $s_1, \ldots, s_{r-1}$. To learn whether the rest of S can be reconstructed by either simply copying or inserting new digits we take s_{r+1}, and then check whether this digit belongs to one of the substrings of S, in which case it can be obtained by simply copying it from S. If s_{r+1} can indeed be copied the routine goes on until a new digit (which once again needs to be inserted) appears. The number of newly inserted digits plus one (if the last copy step is not followed by inserting a digit) gives the complexity measure c of the string S.

As an illustration, consider the following three strings of 10 binary digits each.

A 0000000000
B 0101010101
C 0110001001

At first sight one might correctly guess that A is less random so that A is less complex than B, which in turn is less complex than C. The complexity index c agrees with such an intuition. For the string A one has only to insert the first zero and then rebuild the entire string by copying this digit; thus $c = 2$, where c is the number of steps necessary to create a string. For the string B one has to additionally insert digit 1 and then copy the substring 01 to reconstruct the entire string; thus $c = 3$. For the string C one has to further insert 10 and 001, and then copy 001; thus $c = 5$.

The complexity of a string grows with its length. The genuinely random string asymptotically approaches its maximum complexity r as its length n grows following the rule $\lim_{n \to \infty} c = r = \frac{n}{\log_2 n}$ (Kaspar and Schuster 1987). One may thus compute a positive finite normalized complexity index $LZ = \frac{c}{r}$ to get the complexity of a string relative to that of a genuinely random one. As the string approaches infinite $LZ \to 1$; however, very complex series in practical finite experiments usually have an *LZ* a little bit above one. The index also makes it possible to compare strings of distinct lengths as long as their lengths ≥ 1,000. Figure 1 shows a computer-generated pseudo-random string reaching the bulk of its convergence as it nears 1,000; from this threshold on there is slow asymptotical convergence toward an *LZ* index of one.

Here we consider sliding time windows, calculate the index for every window, and then get the average. For instance, for a time series of 2,000 datapoints and a chosen time window of 1,000 observations we first compute the *LZ* index of the window from 1 to 1,000, then the index of the window from 2 to 1,001, and so on, up to the index of the window from 1,001 to 2000. Then we take the average of the indices.

As an illustration, Figure 2 shows three time series of 15,000 observations each, and the computed *LZ* indices of 14,000 sliding time windows of length 1,000. Figure 2a displays the index evolution of the series of computer-generated pseudo-random numbers (average *LZ* index = 1.062622). Figure 2b shows the index evolution of the series of the distances ("returns") between the first 15,001 adjacent prime numbers (average *LZ* index = 1.014342). And Figure 2c shows the index evolution of the series of natural logs of the distances between the first 15,001 adjacent primes (average *LZ* index = 1.025574). The distances between adjacent primes are believed to be genuinely random, and this agrees with our computed indices in Figure 2. Figure 3 shows the evolution of the *LZ* index for different parameter values of the logistic equation (1,000 iterations with the starting value set at 0.25). The solution to this equation is a series that depends on the value of its growth parameter. The series gets stable for low values of the parameter, which means low complexity. As the parameter grows the series behaves periodically, and then goes chaotic as the parameter approaches 4. This increased complexity agrees with the *LZ* index evolution in Figure 3.

3. Data and analysis

We collected seven years of daily data from July 2000 to July 2007 (2,000 observations) from 36 stock exchange indices (Table 1) as well as 37 stock prices of companies listed on the NYSE, Nasdaq, and Bovespa (Table 2). The source was Yahoo Finance and EconStats.

Analysis was performed with simple returns of the raw series. The return series were coded as ternary strings as follows (Shmilovici *et al.* 2003). Assuming a stability basin b for a return observation ρ_t, a datapoint d_t of the ternary string was coded as $d_t = 0$ if $\rho_t \le -b$, $d_t = 1$ if $\rho_t \ge +b$, and $d_t = 2$ if $-b < \rho_t < +b$. The series would have become binary if we had shrunk the stability basin to the attractor zero, i.e. $b = 0$; yet we assumed $b = 0.0025$ following Shmilovici *et al.*. (We checked for the effects of changing b only to realize that the rankings did not alter too much; yet future research may wish to consider a more sophisticated analysis in the choice of b.) As an illustration, we compare five daily percentage returns of the S&P 500 with 0.25%.

From 18 to 22 June 2007 the returns were, respectively, 0.652%, –0.1226%, 0.1737%, –1.381%, and 0.6407%. Thus the trading week was coded as 12201.

Figure 3 shows the evolution of the index using 1,000 sliding windows for (a) the computer-generated pseudo-random series (average *LZ* = 1.0180), (b) returns of the Dow Jones (average *LZ* = 1.0201), (c) returns of the Shanghai Composite (average *LZ* index = 1.0032), and (d) returns of the Karachi 100 (average *LZ* index = 0.9918). Table 1 shows the average *LZ* index for the other stock exchanges. As can be seen, all the series seem to be very complex. They look more like the genuinely random series than the totally redundant, perfectly predictable series. (Check Figure 3 again to see that a periodic series has an *LZ* well below one.) Inspired by the experiment in Figure 1 we decided to consider $LZ = 1$ as a threshold in order to compare the relative efficiency of the series. We counted the number of occurrences where the *LZ* index was caught above one, and then considered that as a measure of relative efficiency. For the pseudo-random series the $LZ = 1$ threshold was surpassed 98.8% of the times; thus we say that it is 98.8% efficient.

The Dow Jones, Shanghai Composite, and Karachi 100 were found to be, respectively, 95.4%, 49.5%, and 23.7% efficient. Note that the Dow Jones series nears the pseudo-random series. Table 1 shows the measures for the other stock exchanges. As can be seen, the S&P 500 even beat the pseudo-random series. Thus it is safe to conclude that this American stockmarket is almost efficient. By contrast, the Colombo Stock Exchange was found to be only 10.5% efficient, which means that stock prices in that market convey some redundant information.

The procedure above was repeated for selected company stock prices (Table 2). Figure 4 shows the evolution of the *LZ* index using 1,000 sliding windows for (a) Coca-Cola (100% efficient), (b) Yahoo (99.65% efficient), (c) Vale (92.75% efficient), and (d) Aracruz (66.67% efficient).

4. Conclusion

By considering data from 36 stockmarket indices and 37 individual company stock prices, this paper puts forward one way to assess the relative efficiency of stockmarkets. This is made possible thanks to the efficiency interpretation provided by algorithmic complexity theory. The latter makes a connection between the efficient market hypothesis and the unpredictable character of stock returns. The idealized efficient market generates a time series that has a dense amount of nonredundant information, and thus presents statistical features similar to a genuinely random time series.

Physical systems are usually given an efficiency rating based on the relative proportion of energy converted to work. We suggest a similar efficiency rating based on the relative amount of nonredundant information conveyed by financial prices. The price of the idealized efficient market conveys information that is fully nonredundant; this market is then said to be 100% efficient.

Yet prices in real-world markets reflect only the information for which the acquisition costs cannot outweigh the benefits. Also, there are transaction costs, inside trading, and heterogeneous expectations. Since such residual inefficiencies are always present in actual markets one should not expect them to be efficient in absolute terms. Yet considering the random efficient market as a benchmark one can, for instance, say that the S&P 500 is 99.1% efficient whereas the Colombo Stock Exchange is only 10.5% efficient. This means that prices in the American stockmarket incorporate much more nonredundant information than its Sri Lankan counterpart does.

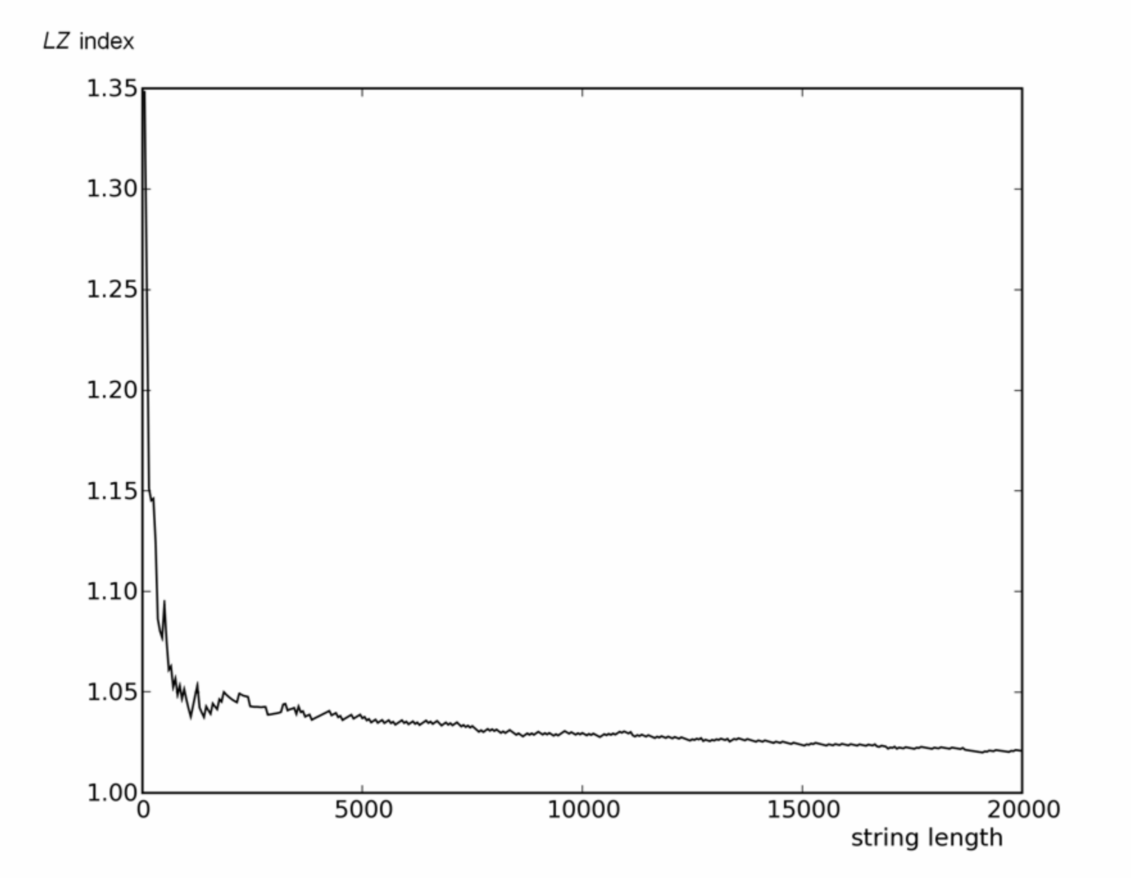

Figure 1. As its length increases, a typical, computer-generated pseudo-random string seems to asymptotically converge to an *LZ* index of one.

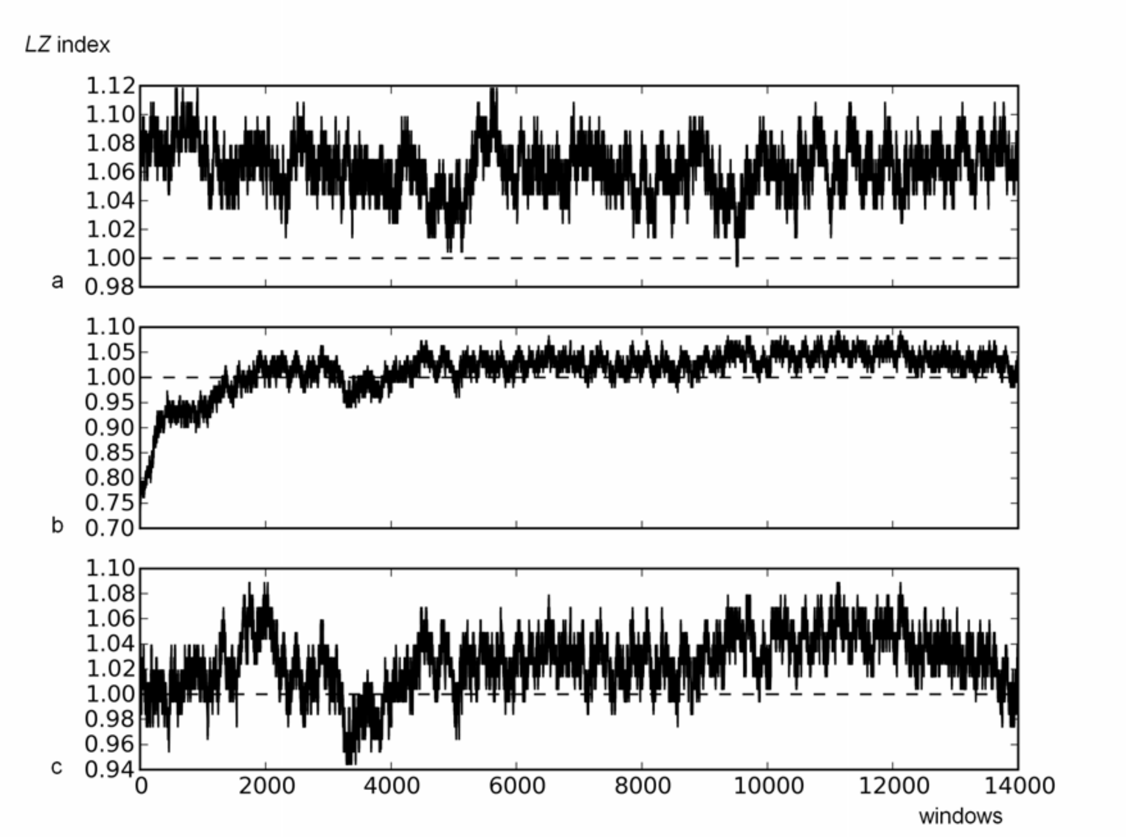

Figure 2. *LZ* index evolution of (a) a series of computer-generated pseudo-random numbers (average *LZ* index = 1.062622), (b) a series of the distances between the first 15,001 adjacent prime numbers (average *LZ* index = 1.014342), and (c) a series of natural logs of the distances between the first 15,001 adjacent primes (average *LZ* index = 1.025574).

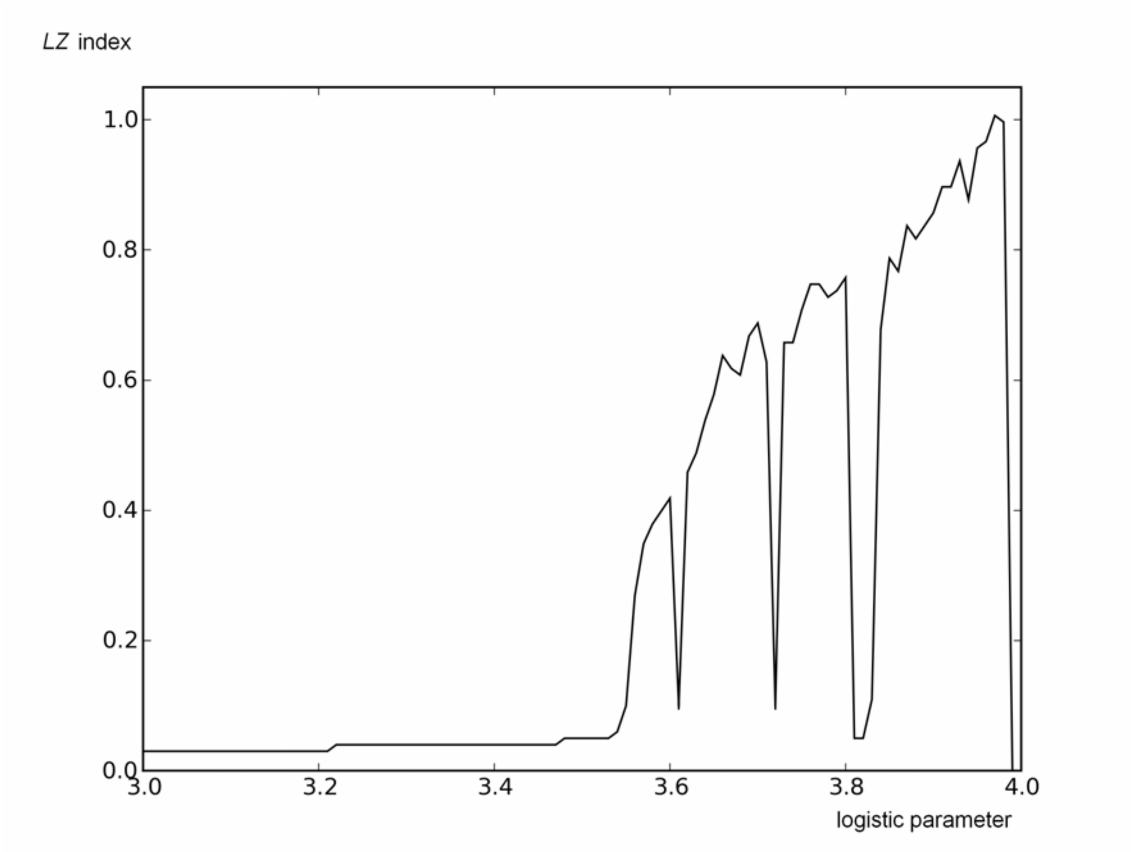

Figure 3. *LZ* index for increased values of the logistic growth parameter (1,000 iterations with the starting value set at 0.25). The series gets stable and then periodic for low values of the parameter (*LZ* complexity index well below one), and then goes chaotic as the parameter approaches 4.

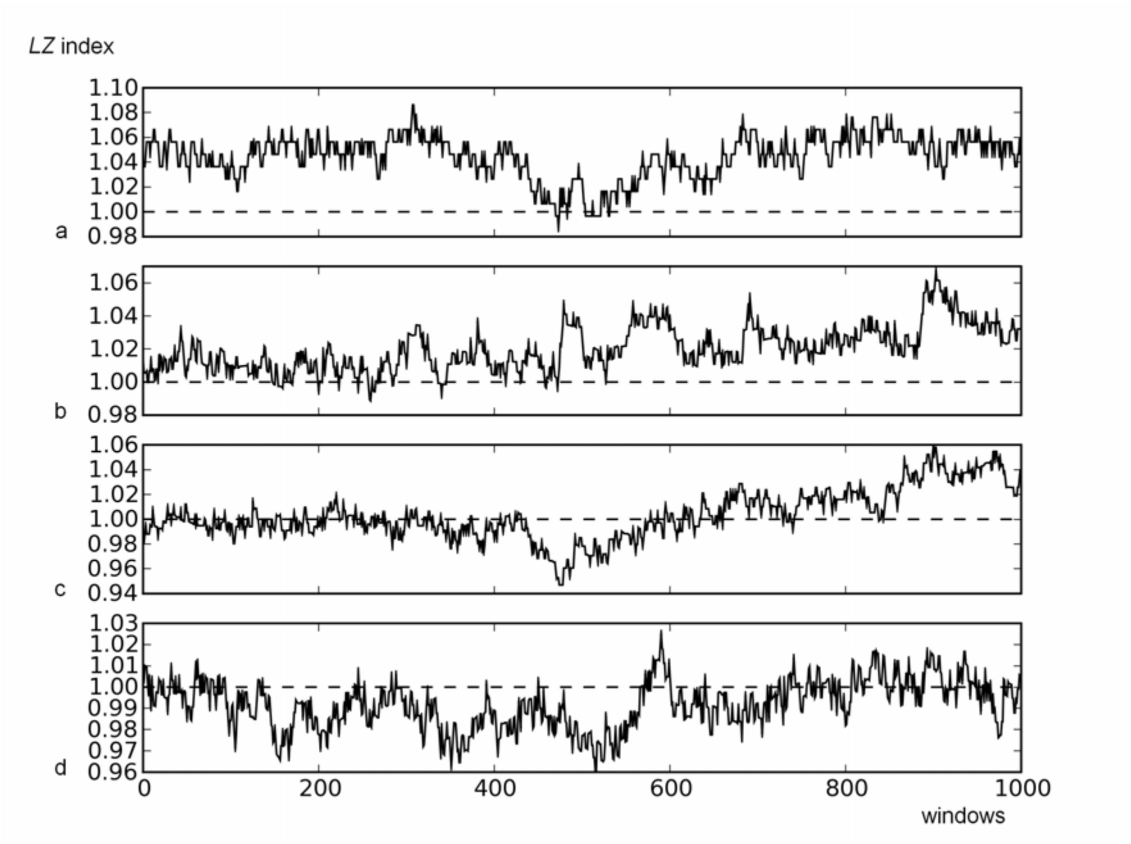

Figure 4. *LZ* index evolution over 1,000 sliding windows for (a) a computer-generated pseudo-random series (average *LZ* = 1.0180), (b) returns of the Dow Jones (average *LZ* = 1.0201), (c) returns of the Shanghai Composite (average *LZ* index = 1.0032), and (d) returns of the Karachi 100 (average *LZ* index = 0.9918).

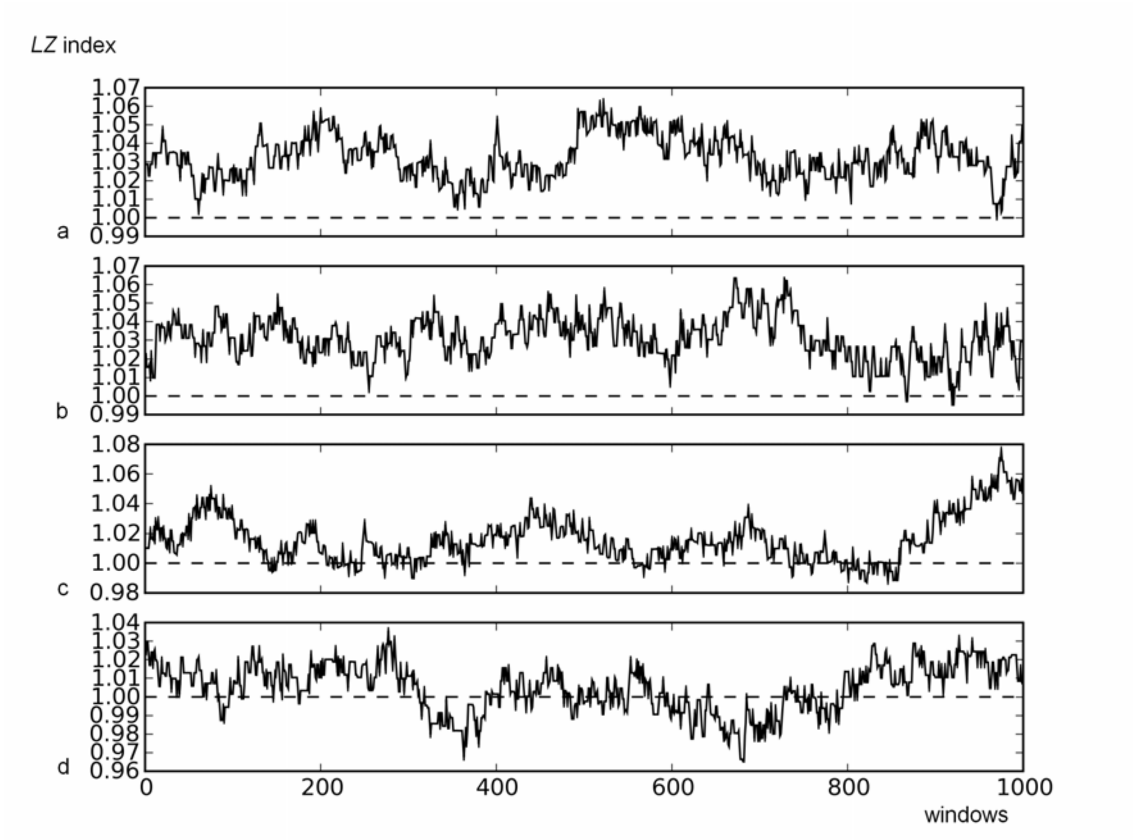

Figure 5. Evolution of the *LZ* index using 1,000 sliding windows for (a) Coca-Cola (100% efficient), (b) Yahoo (99.65% efficient), (c) Vale (92.75% efficient), and (d) Aracruz (66.67% efficient).

Table 1. The relative efficiency of selected stockmarket indices

Stock Exchange	Country	Average *LZ* Index	Degree of Efficiency*, %
S&P 500	USA	1.0232	99.1
DAX 30	GER	1.0257	98.4
Nikkei 225	JPN	1.0432	98.2
All Ordinaries	AUS	1.0246	97.8
ATX	AUT	1.0173	97.4
Dow Jones	USA	1.0201	95.4
Korea Composite	KOR	1.0163	94.9
Tel-Aviv 100	ISR	1.0187	92.9
Hang Seng	HKG	1.0151	91.5
Straits Times	SIN	1.0153	90.3
CAC 40	FRA	1.0138	88.4
Helsinki General	FIN	1.0149	88.4
Kuala Lumpur SE	MAS	1.0158	88
FTSE 100	UK	1.0106	86.6
Prague X	CZE	1.0139	81
Bel 20	BEL	1.0118	80.4
IBC	VEN	1.0110	79.9
Madrid General	ESP	1.0201	79.3
Swiss Market	SUI	1.0101	78.4
Nasdaq Composite	USA	1.0080	75.4
Amsterdam EX	NED	1.0100	74.4
Bovespa	BRA	1.0127	67.8
IPC	MEX	1.0060	64
Merval	ARG	1.0050	62.9
Jakarta Composite	IDN	1.0054	62.1
Istanbul 100	TUR	1.0085	61.3
Moscow Times	RUS	1.0050	59.2
Copenhagen	DEN	1.0025	58.7
Athex Composite	GRE	1.0048	56.9
Bombay SE	IND	1.0010	53.3
Taiwan Weighted	TPE	1.0006	50.3
Shanghai Composite	CHN	1.0032	49.5
Philippines	PHI	0.9987	43.1
Lima General	PER	0.9903	37.9
Karachi 100	PAK	0.9918	23.7
Colombo SE	SRI	0.9795	10.5

* Hits above the threshold $LZ = 1$

Table 2. The relative efficiency of selected company stocks

Company	Stock Exchange	Average *LZ* Index	Degree of Efficiency*, %
Amazon	NYSE	1.0416	100
Coca-Cola	NYSE	1.0324	100
P&G	NYSE	1.0264	99.97
Intel	Nasdaq Composite	1.0292	99.92
eBay	Nasdaq Composite	1.0377	99.8
General Electric	NYSE	1.0274	99.66
Yahoo	Nasdaq Composite	1.0310	99.65
Texaco	NYSE	1.0264	99.46
Cisco	Nasdaq Composite	1.0357	99.44
Petrobras	Bovespa	1.0284	99.43
Pfizer	NYSE	1.0327	99.39
HP	NYSE	1.0298	99.38
Microsoft	Nasdaq Composite	1.0286	99.25
Goldman Sachs	NYSE	1.0311	98.78
J&J	NYSE	1.0275	98.73
Unilever	NYSE	1.0297	98.44
Nissan	Nasdaq Composite	1.0178	97.58
Merrill Lynch	NYSE	1.0279	97.33
JP Morgan	NYSE	1.0281	96.7
Oracle	Nasdaq Composite	1.0206	94.93
Citigroup	NYSE	1.0314	94.59
Vale	Bovespa	1.0193	92.75
Embraer	Bovespa	1.0258	91.59
Itau	Bovespa	1.0183	86.74
FedEx	NYSE	1.0186	86.6
Bradesco	Bovespa	1.0172	85.88
Exxon	NYSE	1.0161	85.56
Ford	NYSE	1.0152	84.26
Marcopolo	Bovespa	1.0072	77.36
Americanas	Bovespa	1.0136	76.68
Ipiranga	Bovespa	1.0111	76.55
Toyota	NYSE	1.0100	76.32
Wal-Mart	NYSE	1.0074	71.42
Ambev	Bovespa	1.0108	70.27
Aracruz	Bovespa	1.0048	66.67
Duratex	Bovespa	1.0048	65.17
Celesc	Bovespa	1.0005	50.03

* Hits above the threshold $LZ = 1$

References

Beechey, M., D. Gruen, and J. Vickery (2000) "The Efficient Market Hypothesis: A Survey" Reserve Bank of Australia research discussion paper 2000-01.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997) *The Econometrics of Financial Markets*, Princeton University Press: Princeton.

Chen, S. H., and C. W. Tan (1999) "Estimating the complexity function of financial time series: an estimation based on predictive stochastic complexity" *Journal of Management and Economics* **3**, No. 3.

Kaspar, F., and H. G. Schuster (1987) "Easily calculable measure for the complexity of spatiotemporal patterns" *Physical Review A* **36**, 842–848.

Lempel, A., and J. Ziv (1976) "On the complexity of finite sequences" *IEEE Transactions on Information Theory* **22**, 75–81.

Mantegna, R. N., and E. Stanley (2000) *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press: Cambridge.

Shmilovici, A., Y. Alon-Brimer, and S. Hauser (2003) "Using a stochastic complexity measure to check the efficient market hypothesis" *Computational Economics* **22**, 273–284.