Ranking the stocks listed on Bovespa according to their relative efficiency

Giglio, Ricardo and Da Silva, Sergio

Federal University of Santa Catarina

2009

Ranking the Stocks Listed on Bovespa According to their Relative Efficiency

Ricardo Giglio

Graduate Program in Economics
Federal University of Santa Catarina
88049970 Florianopolis SC, Brazil
rgiglio@gmail.com

Sergio Da Silva

Graduate Program in Economics
Federal University of Santa Catarina
88049970 Florianopolis SC, Brazil
professorsergiodasilva@gmail.com

Abstract

A methodology based on the algorithmic complexity theory has been applied to assess the relative efficiency of the stocks listed on Bovespa. We provide eight alternative listings of the top ten stocks according to their efficiency rates.

Keywords: Algorithmic complexity theory, Econophysics, Financial efficiency

1 Introduction

When physicists talk about the efficiency of a system, they imply its relative efficiency. An efficiency rate refers, for example, to the relative proportion of energy converted to work. If a piston engine is rated as 30% efficient, this implies that, on average, 30% of the engine's fuel is consumed for useful work, with the remaining 70% lost to heat, light, or noise. In contrast, economists commonly think of the efficiency of markets—that is, the capacity of the market prices to convey nonredundant information in absolute terms [1]. However, efficiency rates based on the relative amount of nonredundant information conveyed by financial prices can be calculated with the help of the algorithmic complexity theory [3]. The price of an idealized, absolutely efficient market conveys completely nonredundant information; in such a case, this market is said to be 100% efficient. The idealized efficient market generates a time series with a dense amount of nonredundant information. Essentially, the algorithmic complexity theory predicts that such a series shows statistical features that are almost indistinguishable from those observed in genuinely random time series [6]. Accordingly, measuring the deviation from randomness yields the relative efficiency of a market. The efficiency rates of stock exchanges, foreign-exchange markets [3], and the markets for the stocks of selected individual companies listed on the New York Stock Exchange, the Nasdaq Stock Exchange, and the Sao Paulo Stock Exchange (Bovespa) [2] have been already calculated using that approach. In this study, we apply the same methodology for evaluating the majority of the stocks listed on Bovespa. We also extend the methodology to consider eight alternative methods of finding the efficiency rates.

The rest of this article is organized as follows: Section 2 explains the measure of algorithmic complexity, Section 3 presents data and elaborates further on the methodology applied, Section 4 conducts the analysis, and Section 5 concludes the study.

2 Measuring algorithmic complexity

In Shannon's entropy of information theory, the expected information content of a series is maximized if the series is genuinely random. In this case, there is maximum uncertainty, and no redundancy, in the series. The algorithmic complexity of a string is the length of the shortest computer program that can reproduce the string. However, the shortest algorithm cannot be computed. Nevertheless, there are several methods to circumvent this problem. Lempel and Ziv [5] suggest a useful measure that does not rely on the shortest algorithm; furthermore, Kaspar and Schuster [4] 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 $S = s_1, \dots, s_n$ 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 originate in the substring $s_1, \dots, s_{r-1}$; otherwise, s_r could simply be copied from $s_1, \dots, s_{r-1}$. To learn whether the rest of S can be reconstructed by either simply copying or inserting new digits, s_{r+1} is initially chosen and subsequently checked as to whether it belongs to one of the substrings of S; in such a case, it can be obtained by simply copying it from S. If s_{r+1} can indeed be copied, the routine continues until a new digit (which once again needs to be inserted) appears. The number of newly inserted digits plus one (if the last copying step is not followed by insertion of a digit) yields 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. In 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. In the string B, one has to additionally insert digit 1 and then copy the substring 01 to reconstruct the entire string; thus, $c = 3$. In 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 $\lim_{n\to\infty} c = r = \frac{n}{\log_2 n}$ [4]. One may thus compute a positive finite normalized complexity index $LZ = \frac{c}{r}$ to obtain the complexity of a string relative to that of a genuinely random one. Under the broad definition of complexity proposed by Lempel and Ziv [5], almost all sequences of sufficiently large length are found to be complex. To obtain a useful measure of complexity, they then consider a De Bruijn sequence, which is commonly viewed as a good finite approximation of a complex sequence [5]. After proving that the De Bruijn sequence is indeed complex according to their definition and that its complexity index cannot be less than one, they decided to fix it as a benchmark against which other sequences could be compared. Thus, a finite sequence with a complexity index greater than one is guaranteed to be more complex than (or at least as complex as) a De Bruijn sequence of the same size. Note that the *LZ* index is not an absolute measure of the complexity (which is perhaps nonexistent), nor is the index ranged between zero and one. A previous work [3] by the authors of this study provides more details on the *LZ* index.

3 Data and methodology

The data regarding the stocks of companies listed on the Bovespa were collected from the website of the stock exchange. For this study, 55 companies (Table 1) that had data for the entire period ranging from August 2000 to September 2008 were considered. This nearly corresponds to eight years of daily data, that is, 2,000 observations of opening prices. This analysis was carried out using the returns of such series.

To calculate the *LZ* index using the method described in Section 2, one needs first to express the original series of returns as binary, ternary, quaternary data, and so on. Unfortunately, the calculation of the *LZ* index is sensitive to whether the coding is binary, ternary, or quaternary, in addition to the stability basin chosen [3]. To remedy this problem, in this study, eight alternative methods of coding the data have been adopted—four ternary and four quaternary.

The return series were coded as ternary strings as follows [3]. Assuming a stability basin b for a return observation ρ_t, a data point 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 $-b < \rho_t < +b$. It was arbitrarily considered that b = 0.25% and b = 0.50%. Although these values have been found to be appropriate using numerical experiments, future research is still needed to rigorously justify the choice of *b*. Furthermore, it should be noted that one can get a binary series by shrinking the stability basin to $b = 0$. As an illustration, the five daily percentage returns of the week from 18 to 22 June 2007 of the Standard & Poor's 500 index can be compared using the value of $b = 0.25\%$ to produce the following percentage returns: 0.652, –0.1226, 0.1737, –1.381, and 0.6407. Thus, the trading week should be coded as 12201.

The return series were coded as quaternary strings adopting the following criterion: $d_t = 0$ if $\rho_t < -b$; $d_t = 1$ if $0 > \rho_t > -b$; $d_t = 2$ if $+b > \rho_t > 0$; and $d_t = 3$ if $\rho_t > +b$. Each series was coded considering the center of the stability basin either at zero or at the historical mean of the series (Table 2).

To find the *LZ* index of the time series, sliding time windows were considered; the index for every window was calculated and then the average was obtained. For example, in the time series of 2,000 data points and a chosen time window of 1,000 observations, the *LZ* index of the window from 1 to 1,000 was first computed; then the index of the window from 2 to 1,001 was derived, and so on, up to the index of the window from 1,001 to 2,000. Then the average of the indices was calculated. To obtain the efficiency rate, *LZ* = 1 was considered as the threshold, with the number of occurrences where the *LZ* index had reached values above one being considered as the measure of relative efficiency.

4 Analysis

Table 3 shows the top ten stocks in terms of their efficiency rate according to the ternary methods, and Table 4 presents the top ten based on the quaternary methods. Table 3 shows that Gerdau and Unipar appeared among the top ten in all the four ternary methods. Petrobras, Itau, and Cteep appeared in three out of four ternary methods. Vale, Itausa, Duratex, Ultrapar, Randon, Copel, and Souza Cruz appeared among the top ten in two out of four ternary methods. Table 3 also suggests that the choice of the stability basis *b* weighs more than the core of the stability basin chosen, whether zero or the historical mean. For example, Table 3 shows that, with the exception of Duratex, the choice of *b* did not alter the rankings of the stocks that appeared at least twice.

Table 4 shows that Bradesco and Gerdau appeared among the top ten in all the quaternary methods. Unipar ranked first in three out of four quaternary methods. Itau, Confab, Souza Cruz, and Duratex also appeared in three out of four quaternary methods. Itausa, Petrobras, Banco do Brasil, Siderurgica Nacional, and Embratel appeared in two out of four quaternary methods. Overall, the efficiency rates were more concentrated under the ternary methods than under the quaternary ones. For example, Duratex dropped from its 99.6% efficiency rate in the third ternary method to 60.4% in the seventh quaternary method. The reductions in the efficiency rates under the quaternary methods are because the maximum complexity of four digit strings is greater than that of three digit strings.

Table 5 shows the number of appearances of 26 stocks among the top ten considering all the eight methods applied herein. Not surprisingly, those appearing more number of times are the stocks with greater weight in the Bovespa index (Table 6).

Figures 1 and 2 show the *LZ* complexity index of selected stocks derived over 1,000 sliding windows using two of the ternary methods. Figure 1 shows the index using method 1 for (a) Bradesco (average *LZ* = 1.0357; efficiency rate = 99.8%), (b) Itau (average *LZ* = 1.0323; efficiency rate = 99.5%, (c) Celesc (average *LZ* = 1.004; efficiency rate = 56.1%), and (d) Aracruz (average *LZ* = 0.9981; efficiency rate = 44.8%).

Figure 2 shows the index using method 2 for (a) Gerdau (average *LZ* = 1.0420; efficiency rate = 100.0%), (b) Unipar (average *LZ* = 1.0339; efficiency rate = 99.9%), (c) Bombril (average *LZ* = 1.0006; efficiency rate = 50.2%), and (d) Acesita (average *LZ* = 1.0004; efficiency rate = 44.0%).

5 Concluding remarks

Financial efficiency is commonly treated as absolute efficiency. The current authors have previously devised a methodology based on the algorithmic complexity theory, which facilitates calculation of the relative efficiency of financial markets. In this study, we have further applied this methodology to the markets of the stocks listed on the Bovespa. Thereafter, the stocks have been ranked according to their efficiency rate.

Though the relative efficiency has an edge over absolute efficiency in theoretical terms, problems do persist with the measure based on algorithmic complexity; this fact should be addressed by future research. Rankings are dependent both on the stability-basin chosen and on the method by which the data are coded. This article has tried to remedy such a deficiency by applying eight alternative methods—four ternary and four quaternary, in addition to using two distinct stability basins. Finally, the results have succeeded in indicating economic wisdom. The stocks that have been rated as the most efficient ones are those of Gerdau, Unipar, Itau, Petrobras, Bradesco, Confab, Souza Cruz, and Duratex. Not surprisingly, these are exactly the ones with greater weights in the Bovespa index.

Table 1. Trade and company names of the 55 companies listed on Bovespa, which have been considered in this work

Stock	Company
Acesita	Acesita
Ambev	Companhia de Bebidas das Américas
Americanas	Lojas Americanas
Aracruz	Aracruz Celulose
Banco do Brasil ON	Banco do Brasil
Banco do Brasil PN	Banco do Brasil
Bombril	Bombril
Bradesco	Banco Bradesco
Celesc	Centrais Elétricas de Santa Catarina
Cemig ON	Companhia Energética de Minas Gerais
Cemig PN	Companhia Energética de Minas Gerais
Cesp	Companhia Energética de São Paulo
Coelce	Companhia Energética do Ceará
Confab	Confab Industrial
Copel ON	Companhia Paranaense de Energia
Copel PN	Companhia Paranaense de Energia
Coteminas	Companhia Tecidos Norte de Minas
Cteep	Companhia de Transmissão de Energia Elétrica Paulista
Duratex	Duratex
Eletrobras ON	Centrais Elétricas Brasileiras
Eletrobras PN	Centrais Elétricas Brasileiras
Emae	Empresa Metropolitana de Águas e Energia
Embraer	Empresa Brasileira de Aeronáutica
Embratel ON	Embratel Participações
Embratel PN	Embratel Participações
Forja Taurus	Forjas Taurus
Fosfertil	Fertilizantes Fosfatados
Gerdau	Gerdau
Itau	Banco Itaú
Itausa	Investimentos Itaú
Magnesita	Magnesita Refratários
Marcopolo	Marcopolo
Pao de Acucar	Companhia Brasileira de Distribuição
Paranapanema	Paranapanema
Petrobras ON	Petróleo Brasileiro
Petrobras PN	Petróleo Brasileiro
Randon	Randon Implementos e Partipações
Sabesp	Companhia de Saneamento Básico do Estado de São Paulo
Siderurgica Nacional	Companhia Siderúrgica Nacional
Souza Cruz	Souza Cruz
Telemar ON	Tele Norte Leste Participações
Telemar PN	Tele Norte Leste Participações
Telemig ON	Telemig Celular Participações
Telemig PN	Telemig Celular Participações
Telenorte	Tele Norte Celular Participações
Telesp ON	Telecomunicações de São Paulo
Telesp PN	Telecomunicações de São Paulo
Tim ON	TIM Participações
Tim PN	TIM Participações
Ultrapar	Ultrapar Participações
Unipar	União de Indústrias Petroquímicas
Usiminas	Usinas Siderúrgicas de Minas Gerais
Vale ON	Companhia Vale do Rio Doce
Vale PN	Companhia Vale do Rio Doce
Vcp	Votorantim Celulose e Papel

Note:
ON common stock
PN preferred stock

Table 2. The eight methods considered in this study

Method	Coding	Stability basin	Core of the stability basin
1	ternary	0.0025	0
2	ternary	0.0025	historical mean
3	ternary	0.005	0
4	ternary	0.005	historical mean
5	quaternary	0.0025	0
6	quaternary	0.0025	historical mean
7	quaternary	0.005	0
8	quaternary	0.005	historical mean

Table 3. Top ten companies in terms of greatest efficiency rates: ternary methods

Method 1		Method 2		Method 3		Method 4	
Stock	Efficiency rate, %	Stock	Efficiency rate, %	Stock	Efficiency rate, %	Stock	Efficiency rate, %
Bradesco	99.80	Gerdau	100.00	Unipar	99.60	Gerdau	99.60
Itausa	99.50	Unipar	99.90	Duratex	99.60	Confab	99.20
Telesp ON	98.70	Cteep	99.00	Gerdau	99.20	Unipar	98.90
Gerdau	98.20	Banco do Brasil ON	97.50	Cteep	97.80	Copel PN	98.80
Unipar	98.20	Ultrapar	96.40	Randon	97.30	Souza Cruz	96.40
Ultrapar	98.00	Petrobras ON	95.20	Copel PN	96.50	Pao de Acucar	96.00
Cteep	97.30	Forja Taurus	95.00	Souza Cruz	96.50	Itau	95.10
Itau	97.20	Itausa	94.80	Confab	96.30	Randon	94.10
Petrobras ON	97.10	Sabesp	94.70	Vale PN	95.70	Marcopolo	94.00
Duratex	97.10	Itau	94.20	Petrobras ON	95.50	Vale PN	93.70

Note:
ON common stock
PN preferred stock

Table 4. Top ten stocks in terms of greatest efficiency rates: quaternary methods

Method 5		Method 6		Method 7		Method 8	
Stock	Efficiency rate, %	Stock	Efficiency rate, %	Stock	Efficiency rate, %	Stock	Efficiency rate, %
Unipar	98.20	Unipar	93.40	Unipar	73.30	Duratex	56.10
Itausa	88.80	Gerdau	78.00	Duratex	60.40	Souza Cruz	56.00
Usiminas	81.60	Itausa	68.90	Petrobras PN	54.20	Bradesco	52.80
Itau	78.00	Souza Cruz	67.20	Cteep	49.30	Confab	52.80
Bradesco	77.80	Embratel PN	63.20	Itau	49.20	Banco do Brasil ON	52.50
Tim ON	72.60	Petrobras PN	63.00	Siderurgica Nacional	46.40	Embratel PN	48.60
Confab	69.10	Itau	61.20	Confab	46.40	Gerdau	45.30
Banco do Brasil ON	66.70	Duratex	60.90	Souza Cruz	45.40	Embraer	40.70
Gerdau	62.50	Siderurgica Nacional	58.30	Gerdau	43.20	Randon	37.10
Paranapanema	59.60	Bradesco	55.70	Bradesco	38.20	Pao de Acucar	33.60

Note:
ON common stock
PN preferred stock

Table 5. Most ranked stocks among the top ten under all the eight methods applied

Company	Number of appearances in the top ten
Gerdau	8
Unipar	7
Itau	6
Petrobras, Bradesco, Confab, Souza Cruz, Duratex	5
Itausa, Cteep	4
Banco do Brasil, Randon	3
Vale, Pao de Acucar, Siderurgica Nacional, Ultrapar, Embratel, Copel	2
Embraer, Usiminas, Telesp, Sabesp, Marcopolo, Tim, Paranapanema	1

Table 6. Weights of the stocks in the Bovespa index as of October 2008, %

Stock	Weight
Petrobras	15.4
Vale	12.7
Bradesco	3.6
Siderurgica Nacional	3.3
Itau	3.2
Usiminas	3.0
Gerdau	2.9
Banco do Brasil	2.4
Unibanco	2.4
Itausa	2.3

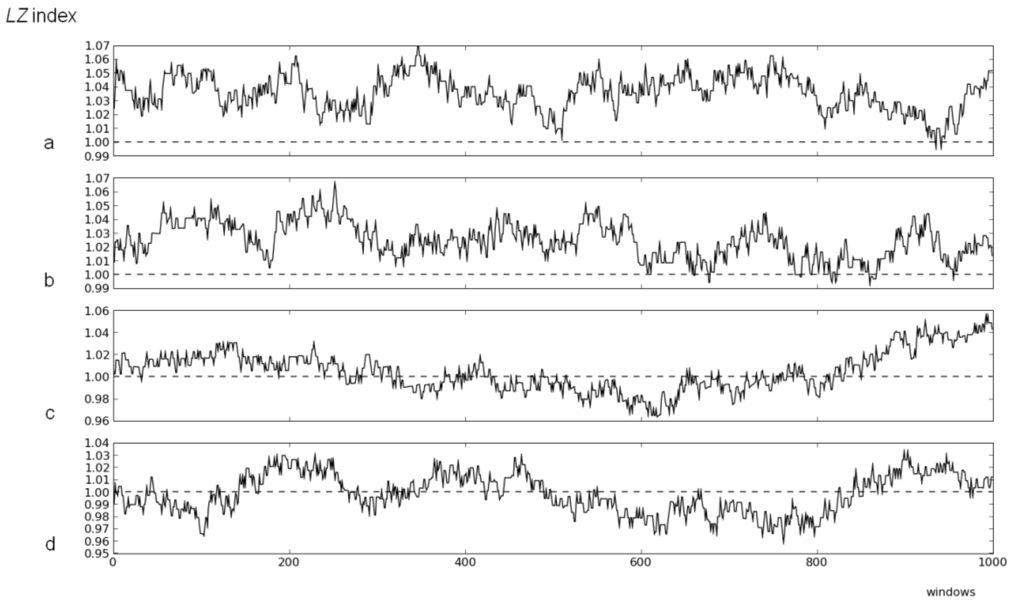

Figure 1. *LZ* complexity index over 1,000 sliding windows for (a) Bradesco (average *LZ* = 1.0357; efficiency rate = 99.8%), (b) Itau (average *LZ* = 1.0323; efficiency rate = 99.5%, (c) Celesc (average *LZ* = 1.004; efficiency rate = 56.1%), and (d) Aracruz (average *LZ* = 0.9981; efficiency rate = 44.8%). These results were obtained using method 1.

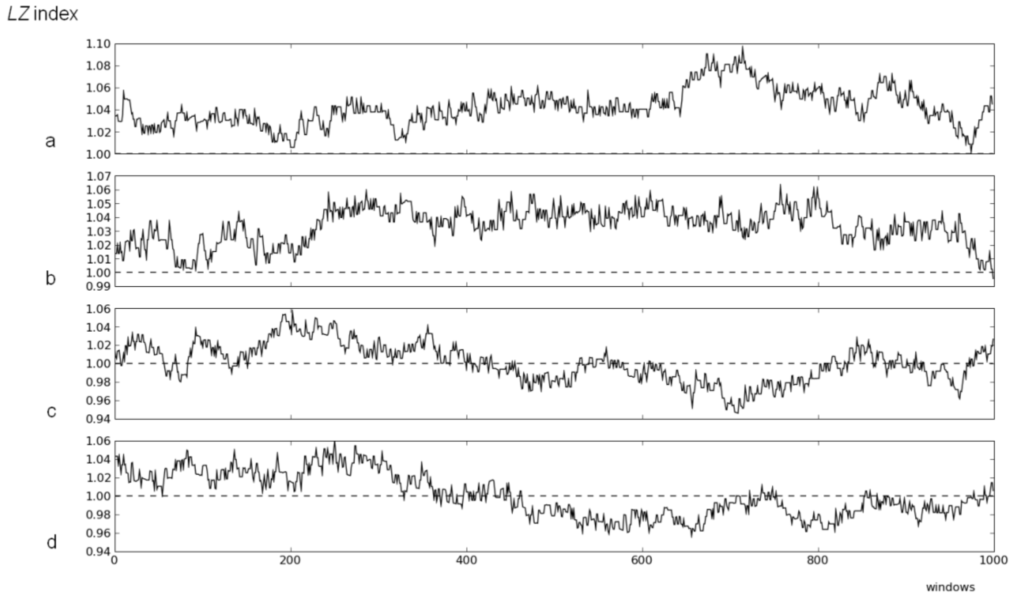

Figure 2. *LZ* complexity index over 1,000 sliding windows for (a) Gerdau (average *LZ* = 1.0420; efficiency rate = 100.0%), (b) Unipar (average *LZ* = 1.0339; efficiency rate = 99.9%), (c) Bombril (average *LZ* = 1.0006; efficiency rate = 50.2%), and (d) Acesita (average *LZ* = 1.0004; efficiency rate = 44.0%). These results were obtained using method 2.

References

[1] J. Y. Campbell, A. W. Lo, A. C. MacKinley, The Econometrics of Financial Markets, Princeton University Press, Princeton, 1997.

[2] R. Giglio, R. Matsushita, S. Da Silva, The relative efficiency of stockmarkets, Economics Bulletin, 7(2008a), 1−12.

[3] R. Giglio, R. Matsushita, A. Figueiredo, I. Gleria, S. Da Silva, Algorithmic complexity theory and the relative efficiency of financial markets, Europhysics Letters, 84(2008b), 48005.

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

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

[6] R. N. Mantegna, H. E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, 2000.