

DETECÇÃO DE PADRÕES EM RETORNOS DE AÇÕES UTILIZANDO A TEORIA DA INFORMAÇÃO ALGORÍTMICA ¹

DETECTING PATTERNS IN STOCK RETURNS USING ALGORITHMIC INFORMATION THEORY

Adhemar Ranciaro Neto²

Iram Marcelo Gléria³

Luiz Carlos Marques dos Anjos⁴

Daniel José Cardoso da Silva⁵

Resumo: O objetivo deste trabalho foi desenvolver uma regra temporal para a negociação de um ativo em bolsa de valores com base num indicador de grau de aleatoriedade de uma sequência de caracteres. Foram aplicadas a medida de complexidade normalizada de Lempel e Ziv e uma medida de eficiência relativa de mercado, baseadas na teoria da informação algorítmica, em dados de alta frequência de retornos para diferentes períodos de negociação. Os resultados foram obtidos para um ativo, no qual foi possível determinar o conjunto de períodos em que pudesse haver maior chance de detectar previsibilidade nos retornos. Conseguiu-se descrever, em termos probabilísticos, em quais intervalos podem ocorrer padrões para um determinado ativo. A medida define um padrão, mas aquela sofre com flutuações estatísticas podendo, em algumas observações, mostrar-se errônea apesar de haver uma chance de 1% para sua ocorrência.

Palavras-chave: Eficiência de mercado. Teoria da informação algorítmica. Medidas de complexidade.

Abstract: The purpose of this study was to develop a rule for the time of trading an stock exchange asset based on an indicator of degree of randomness of a sequence of characters. It was applied the Lempel and Ziv standardized complexity measure and a measure of relative efficiency of the market, based on algorithmic information theory, to high frequency data of returns for different periods of negotiation. The results were obtained for an asset, in which it was possible to determine the set of periods in which there might be more likely to detect predictability in returns. We were able to describe, in probabilistic terms, in which intervals may occur standards for a particular asset. The measure defines a standard, but that suffers from statistical fluctuations can be, in some observations, wrong despite a 1% chance of its occurrence.

Key-words: Market efficiency. Algorithmic information theory. Complexity measures.

¹ Artigo apresentado no VI Seminário UFPE de Ciências Contábeis – outubro/2012

² Aluno de Pós Graduação em Física da Matéria Condensada na Ufal – adhrraneto@gmail.com

³ Doutor em Física pela UNB.

⁴ Doutorando em Ciências Contábeis pelo Programa Multi-Institucional e Inter-Regional UNB / UFPB / UFRN – lcانjos@feac.ufal.br

⁵ Mestre em Ciências Contábeis pela UFPE – danieljcsilva@gmail.com

1 Introdução

Nos mercados financeiros existe a ideia, por parte de alguns teóricos, de que não é possível prever o comportamento de preços futuros dos instrumentos negociados, dado que toda a informação nova que chega ao mercado é rapidamente transmitida aos agentes e é refletida integralmente nos preços (hipótese dos mercados eficientes). Como as novas informações são imprevisíveis, os preços flutuam de forma aleatória ao longo do tempo. Com isso, a possibilidade de lucro seria nula, o que desestimularia os agentes a participarem do mercado. Todavia, este fato não é verificado no mundo real.

Empiricamente, observou-se que os preços não seguem passeios aleatórios isto é, há um grau de previsibilidade sobre seus valores, o que permite especulações sobre como encontrar tais padrões. Por exemplo, estudiosos da análise técnica (popularmente conhecida como “grafista”) possuem métodos definidos para se observar padrões no comportamento dos preços, analisando formas e outras características nas séries temporais de preços.

Este trabalho teve por objetivo apresentar uma regra para auxiliar a busca por intervalos de tempo em que haja chance de se encontrar padrões nos retornos de ativos negociados naqueles mercados a partir do estudo do valor de uma estatística obtida com base na teoria da informação algorítmica, que mede o grau de desvio da aleatoriedade de uma sequência de caracteres.

Dentre os motivos que fundamentam este trabalho, pode-se expressar que este apresenta uma ferramenta de auxílio à tomada de decisão do negociador, tendo maior chance de alocar seus recursos de forma a se obter um resultado adequado às suas necessidades e que serve como mais um teste empírico a ser utilizado no debate sobre a eficiência de um mercado.

A organização deste texto foi realizada da seguinte forma: na seção 2 foi apresentado o arcabouço teórico referente à discussão sobre a hipótese dos mercados eficientes, a introdução a uma ideia alternativa, denominada eficiência relativa de mercado e a uma medida deste conceito, baseada na teoria da informação algorítmica, na seção 3 foi apresentado o método utilizado para a determinação da regra temporal de observação que aumenta a chance de previsibilidade de retornos, na seção 4 foram apresentados os resultados e a discussão e na seção 5 foi apresentada a conclusão.

2 Referencial Teórico

Na literatura, o estudo da evolução dos preços dos ativos nos mercados financeiros teve como uma de suas bases teóricas, o exposto em Bachelier (1900) que, ao analisar preços de instrumentos financeiros negociados em bolsa, verificou que seu valor futuro não podia ser previsto, mas que haveria possibilidade de se conhecer sua distribuição de probabilidade. Tal autor definiu a distribuição futura de preços como sendo a gaussiana (normal). Após mais de seis décadas, Samuelson (1965) observou, por meio de modelos espaciais de distribuição de preços que, se os preços incorporassem as informações e as expectativas de todos os participantes do mercado, aqueles flutuam aleatoriamente. Fama (1970), interessado pela análise empírica da evolução dos preços nos mercados, desenvolveu uma série de testes econométricos de precificação de ativos, observou regularidades e anomalias nos mercados de ações, de títulos, de câmbio e de commodities e realizou estudos de evento nestes mercados.

Com tais contribuições, o autor formalizou a hipótese dos mercados eficientes (apresentada por Samuelson (1965) em sua análise teórica de preços), classificando um mercado como tal, se os preços conseguissem refletir integralmente toda a informação disponível.

Logo, se toda a informação disponível é refletida nos preços, então não é possível realizar previsões e, portanto é impossível “vencer o mercado”, ou seja, o lucro esperado ao se portar um ativo seria nulo em um mercado eficiente. Lucas (1978) afirmou que, em mercados onde os agentes possuíssem expectativas racionais, a eficiência informacional proposta por Fama (1970) e Samuelson (1965) seria verificada.

Dentre outros defensores da existência de mercados eficientes, Jensen (1978) afirmava que, em tais mercados, os preços refletem apenas as informações cujos custos de obtenção não excedam os benefícios de sua utilização. O mesmo autor verificou, por meio de testes de evento, a consistência da hipótese dos mercados eficientes em diversos casos.

Grossman e Stiglitz (1980) apontaram que, em um mercado eficiente, não haveria estímulos para que houvesse negociação, dado que o lucro esperado é nulo. Eles propuseram que os mercados não deveriam ser avaliados com eficiência absoluta, mas sim em relação a um padrão ideal (eficiência relativa). Quanto menos eficiente um mercado, maiores as possibilidades de ocorrência de resultado esperado diferente de zero, havendo, assim, estímulo para a negociação.

Outros autores também apontaram suas opiniões em favor da não eficiência de mercado. Malkiel (2003), por exemplo, ilustrou a incapacidade de um grande estrategista do mercado financeiro gerar resultados positivos de maneira persistente em um mercado eficiente, mostrando a equivalência de suas escolhas com as realizadas por chimpanzé vendendo atirando dardos sobre os ativos que quisesse possuir em sua carteira. Algumas estratégias oriundas da análise técnica, evidenciando ganhos positivos enfatizam o posicionamento acima (BROCK, LAKONISHOK & LeBARON, 1992; MURPHY, 2009).

Campbell, Lo e MacKinlay (1997) não discutiram a eficiência de mercado em si, mas argumentaram que a hipótese, conforme a definição Fama não era possível de ser testada. Os mesmos autores, baseados no trabalho de Grossmann e Stiglitz (1980) ensinaram que a eficiência de um mercado deveria ser aferida em termos relativos, ou seja, com base na comparação daquele com um mercado que fosse considerado eficiente, de modo análogo ao que é feito com o estudo de rendimento de uma máquina térmica, sendo este determinado em termos relativos. Por exemplo, um motor a combustão possui rendimento de 50%, ou seja, tal máquina realiza metade do trabalho de um motor ideal.

Observando a proposição de aleatoriedade de preços proposta inicialmente por Samuelson (1965) e sua relação com a eficiência de mercado, Fama (1991) propôs que a hipótese dos mercados eficientes poderia ser observada em mercados em que não houvesse fricções, como por exemplo, custos de transação, que as informações estariam disponíveis a todos os agentes do mercado e que os agentes deveriam ter a mesma opinião sobre os impactos da informação sobre os preços futuros (expectativas homogêneas). Lo e MacKinlay (1988) detectaram a partir de um teste de razão de variância, que os preços nos mercados reais não flutuavam aleatoriamente, explicando que o comportamento não randômico dos preços não significava ineficiência dos

mercados, mas sim, efeitos relacionados às fricções existentes, o que foram chamados de efeitos da microestrutura de mercado (CAMPBELL, LO & MACKINLAY, 1997).

Inspirados na ideia de eficiência relativa proposta por Grossman e Stiglitz (1980), e no estudo de aleatoriedade das sequências de caracteres, Mantegna e Stanley (1999) propuseram um estudo de previsibilidade com base no uso da teoria da informação algorítmica proposta de forma independente por Kolmogorov (1965), por Solomonoff (1964) e por Chaitin (1966). Para tal teoria, uma sequência de símbolos era considerada aleatória quando o menor programa de computador utilizado para reproduzir a sequência tivesse um tamanho em bits (medida de unidade de informação) igual ou superior ao da mesma. Ou seja, uma sequência aleatória somente conseguiria ser reproduzida se for copiada integralmente. Logo, quanto menor fosse o tamanho do programa em relação ao tamanho da cadeia de caracteres, maior seria o desvio do objeto relativo à aleatoriedade.

Um dos instrumentos utilizados para medir o nível de aleatoriedade de um objeto baseado na teoria da informação algorítmica é a medida de Lempel e Ziv (1976) de complexidade. Tal medida foi empregada por Giglio, Matsushita e Silva (2008), utilizando a adaptação proposta por Kaspar e Schuster (1987) acoplando a ideia de região de estabilidade proposta por Shmilovici, Alon-Brimer e Hauser (2003) para determinar o grau de eficiência relativa em 36 índices de bolsas de valores e em 36 tipos diferentes de ações negociadas em diversos locais do mundo utilizando séries diárias de preços.

Ranciaro Neto (2010) estudou a medida em dados de alta frequência testando sua sensibilidade em relação a mudanças em parâmetros, dentre eles a região de estabilidade, identificando convergência da medida para a unidade em sequências de tamanho muito grande e diminuição de seu valor ao se aumentar o tamanho da região de estabilidade, na maioria das séries de tempo analisadas. Das 15 ações analisadas, apenas uma não apresentava comportamento monótono da medida de Lempel e Ziv (1976) para variações daquela grandeza.

Taufemback, Giglio e Silva (2011) verificaram a perda de eficiência relativa de mercado em ações após a crise de 2008 (falência do banco Lehman Brothers) com base nas variações da medida de Lempel e Ziv (1976). Outras medidas de previsibilidade dos mercados financeiros e de sequências numéricas foram apresentadas em Campbell, et al.(1997) e em Rukhin (2003),

Neste trabalho foi construída uma regra para a observação de padrões nos retornos de uma série temporal de preços de alta frequência de um ativo negociado em bolsa de valores utilizando a aplicação da medida de Lempel e Ziv (1976) encontrada em Giglio et al (2008) para um mesmo ativo, porém em intervalos de tempo distintos, desta forma, observando o comportamento da estatística ao longo das diferentes subséries de tempo analisadas para que se possa verificar, com elevada significância, sua previsibilidade em função do nível de desvio desejado de aleatoriedade.

3 Metodologia

Nesta seção, apresenta-se a medida do desvio de aleatoriedade de uma sequência numérica proposto por Lempel e Ziv (1976), com extensão realizada por Kaspar e Schuster (1987) e com aplicação em mercado financeiro desenvolvida por Giglio, Matsushita e Silva (2008) e por

Ranciaro Neto (2010). Em seguida, foi mostrada a aplicação do método nos dados financeiros das ações preferenciais da empresa Eletropaulo S/A negociadas na BM&F-BOVESPA. Na seção 3.3 foi definida uma regra para detecção de padrões de preços em intervalos de tempo diferentes.

3.1 A medida de Lempel e Ziv

Lempel e Ziv (1976) propuseram uma aferição da aleatoriedade de uma sequência finita de caracteres (tamanho finito e número finito de tipos de símbolos) baseada em uma máquina idealizada (máquina de Turing) que realiza duas operações: leitura e impressão. Para explicar o procedimento, os referidos autores propuseram algumas definições que foram detalhadas abaixo.

O número de termos em uma sequência é denominado comprimento desta. Por exemplo: a sequência $S = 0010$ tem comprimento $l(S) = 4$.

Uma sequência sem caracteres (vazia) possui comprimento zero e é denotada por $S = \Lambda$.

Uma subsequência de S é toda sequência gerada a partir dos termos de S , que obedeça a ordem estabelecida em S . A sequência vazia é uma subsequência de qualquer sequência.

O vocabulário de uma sequência de caracteres ($v(S)$) é o conjunto de todas as palavras (subseqüências) de S . Por exemplo: $v(0010) = \{\Lambda, 0, 1, 00, 01, 10, 001, 010, 0010\}$

Um prefixo próprio de S é uma subsequência que tenha o mesmo início de S , mas com comprimento inferior a S . Exemplo: $S = 0010$. Os prefixos próprios de S são: $T = 0$ $Q = 00$, $R = 001$. Para representar tais prefixos, definiu-se um operador (π^i) que retira da sequência original os i últimos termos. Exemplo: $S = 0010$. Logo, $S\pi^2 = 00$ e $S\pi = 001$.

Uma palavra $Y \in v(S)$ é chamada autopalavra de S quando ela não pertencer a nenhum prefixo próprio de S . O autovocabulário $e(S)$ é o conjunto das autopalavras de S . Por exemplo, para $S = 0010$ um conjunto possível de autopalavras de S é: $e(S) = \{10, 010, 0010\}$

Sejam duas seqüências S e Q . A extensão $R = SQ$ é dada pela concatenação de S e de Q , isto é, o termo imediatamente à direita do último termo de S é o primeiro termo de Q e R termina com o último termo de Q .

Há dois processos de construção de uma seqüência: por reprodução (cópia dos termos posteriores a partir de termos anteriores) e por produção (termos novos não existentes nos prefixos da seqüência). Uma extensão $R = SQ$ é reproduzível a partir de S ($S \rightarrow R$) se $Q \in v(SQ\pi)$. Uma seqüência S é produtível a partir de seu prefixo $S(1, j)$, se $S(1, j) \rightarrow S\pi$ e $j < l(S)$. A notação utilizada é $S(1, j) \Rightarrow S$ e $S(1, j)$ é uma base de S . O processo de produção de S é dado pela repetição dos passos de produção $S(1, h_i) \Rightarrow S(1, h_{i+1})$ consecutivamente iniciando com $i = 0$ e acabando em $i = l(S) - 1$. O resultado da produção $S(1, h_k)$ do k -ésimo passo é chamado k -ésimo estado do processo.

A história de produção de uma seqüência S , $H(S)$ é dada por: $H(S) = S(1, h_1)S(h_1 + 1, h_2)S(h_2 + 1, h_3) \dots S(h_{m-1} + 1, h_m)$, com $i=1,2,\dots,m$. Os componentes

de $H(S)$ são as palavras $H_i(S) = S(h_{m-1} + 1, h_m)$ e $h_0 = 0$. O componente $H_i(S)$ e sua produção correspondente $S(1, h_{i-1}) \Rightarrow S(1, h_i)$ são chamados de exaustivos se $S(1, h_{i-1}) \rightarrow S(1, h_i)$.

A história $H(S)$ é chamada de exaustiva se todos os seus componentes forem exaustivos, com exceção do último, que pode ou não ser exaustivo. Denota-se história exaustiva por $E(S)$. Por exemplo: a sequência $S = 0001101001000101$ pode ser decomposta da seguinte forma: $0 \cdot 001 \cdot 10 \cdot 100 \cdot 1000 \cdot 10$.

Os termos à esquerda de cada ponto são exaustivos. O último termo não é exaustivo (não possui ponto à direita).

A medida de complexidade de uma sequência finita proposta por Lempel e Ziv (1976) $c(S)$ é calculada da seguinte forma: $c(S) = \min \{c_H(S)\}$, onde $c_H(S)$ é o número de componentes de uma história de S . A minimização é dada sobre todas as histórias de S . A história que produz o menor valor de complexidade, de acordo com os autores, é a história exaustiva de S . Logo, $c(S) = c_E(S)$

Dado que, para todo número positivo ε , $\lim_{n \rightarrow \infty} P\left(c(S) < \frac{n(1-\varepsilon)}{\log(n)} \mid l(S) = n\right) = 0$ onde n é o número de elementos da sequência analisada, é possível que sequências não aleatórias possam assumir complexidades elevadas pelo fato de possuírem um número elevado de termos. Para corrigir o problema, foi necessário considerar a sequência como se tivesse sido emanada de uma fonte discreta e ergódica e relacionar sua complexidade com a entropia da fonte. Com isso obteve-se o limite superior $c(s) = \frac{hn}{\log(n)}$, onde h é a entropia normalizada da fonte.

De acordo com Shannon (1948), a entropia da fonte é dada por: $h = -k \sum_{i=1}^a p_i \log(p_i)$, onde k é apenas um fator de escala, p_i é a proporção do caractere i na sequência e o logaritmo é na base a (a é o tamanho do alfabeto).

O valor $\frac{c(S) \log(n)}{hn}$ é chamado de complexidade de Lempel e Ziv normalizada pela entropia (neste trabalho tal valor foi denominado LZ). Logo, quando o valor de LZ assume o valor de 1, a sequência analisada pode ser considerada aleatória. Tal fato foi comprovado pelos autores ao testarem a medida em uma família de sequências de 2 caracteres (binárias) construídas por De Bruijn (1975), consideradas como boas aproximações finitas de sequências complexas (LEMPER & ZIV 1976). À medida que o valor de LZ se afasta da unidade para baixo, a sequência começa a possuir padrões, o que torna previsível, assintoticamente, a determinação de seus termos posteriores.

3.2 Procedimento computacional proposto por Kaspar e Schuster

Kaspar e Schuster (1987) propuseram um algoritmo eficiente, baseado em cópia e inserção, para o cálculo de LZ. O procedimento foi proposto a seguir.

Seja uma sequência $s_1 s_2 s_3 \dots s_n$ que foi reconstruída a partir de um programa até o dígito s_r e que este dígito foi inserido (ou seja, não foi copiado a partir de $s_1 s_2 s_3 \dots s_{r-1}$). Seja $S = s_1 s_2 s_3 \dots s_r$; onde o \cdot indica que s_r foi inserido. Para determinar se o resto da sequência foi inserido ou copiado, procede-se da seguinte forma: toma-se $Q = s_{r+1}$ e verifica se $Q \in v(SQ\pi)$.

Em caso afirmativo, Q é obtida por meio de cópia de uma palavra de S . Então, concatena-se o próximo elemento a Q , obtendo $Q = s_{r+1}s_{r+2}$ e verifica se $Q \in v(SQ\pi)$ até o momento em que Q não possa mais ser obtida por meio de cópia de um elemento de $v(SQ\pi)$ sendo, então, inserida em S e $Q = \Lambda$ voltando ao início do procedimento. O número $c(S)$ de inserções em S (passos de produção) somado com 1 (um), caso a última cópia não seja seguida por uma inserção, é a complexidade da sequência $s_1s_2s_3 \dots s_n$.

Segue o exemplo abaixo para a compreensão do procedimento para o cálculo de c a partir da sequência 0010.

Passo 1: O primeiro dígito deve sempre ser inserido $\rightarrow 0 \cdot$

Passo 2: $S = 0, Q = 0, SQ = 00, SQ\pi = 0, Q \in v(SQ\pi) \rightarrow 0 \cdot 0$

Passo 3: $S = 0, Q = 01, SQ = 001, SQ\pi = 00, Q \notin v(SQ\pi) \rightarrow 0 \cdot 01 \cdot$

Passo 4: $S = 001, Q = 0, SQ = 0010, SQ\pi = 001, Q \in v(SQ\pi) \rightarrow 0 \cdot 01 \cdot 0$ (última cópia não seguida de inserção)

Então, $c(S) = 3$ para 0010.

O algoritmo foi descrito em linguagem de fluxograma na Figura 1 a seguir.

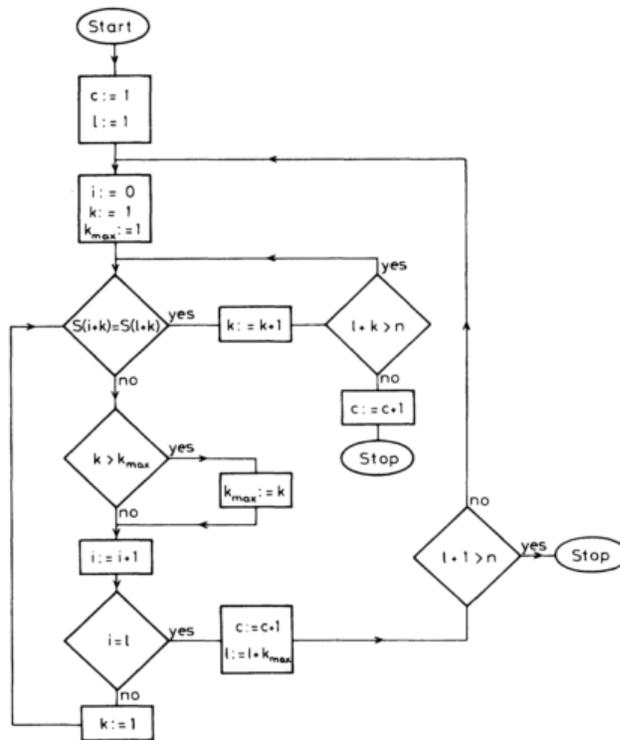


Figura 1: Fluxograma do algoritmo proposto por Kaspar e Schuster.

3.3 Aplicação da medida **LZ** aos dados para a definição da regra de observação de previsibilidade.

A medida **LZ** foi aplicada em uma série temporal de preços de ações em alta frequência (*tick by tick*) obtida junto à BM&F BOVESPA. A categoria de preço analisada foi a que estava denominada por preço de negócio, que é o preço com que a transação foi efetivada.

O procedimento foi realizado para um só ativo, com o objetivo de demonstrar sua aplicação para a análise de sua sequência temporal de retornos, podendo ser aplicado da mesma forma para as demais ações. A sequência de preços foi convertida em retornos dados pela expressão:

$$r(t) = \frac{P(t) - P(t-1)}{P(t-1)}, \text{ onde } P(t) \text{ é o preço do ativo no instante } t.$$

Em seguida, os $r(t)$ foram codificados em caracteres pertencentes ao alfabeto $\{0,1,2\}$ com base no critério proposto por Shmilovici, *et al.* (2003) que utiliza regiões de estabilidade. Seja x o valor a ser convertido, ρ a região de estabilidade pré determinada e s o valor codificado a partir de x . O critério de codificação é:

$$x > \rho \Rightarrow s = 1$$

$$-\rho < x \leq \rho \Rightarrow s = 2$$

$$x \leq \rho \Rightarrow s = 0$$

Após a codificação da série de retornos, todos os caracteres foram concatenados na ordem temporal formando, assim, uma sequência. A partir da sequência formada, partiu-se para o cálculo da complexidade normalizada pela entropia de Shannon (**LZ**).

Baseado no trabalho de Giglio *et al.* (2008) foi possível perceber que não era necessário utilizar todos os elementos da sequência de caracteres para calcular o **LZ**, pois há probabilidade não nula de se obter valores altos de **LZ** para uma sequência não complexa devido à flutuação aleatória da medida (LEMPER & ZIV, 1976). Aquele autor fez menção ao uso do método das janelas deslizantes. O método consiste em utilizar subsequências de comprimento fixo chamadas de janelas, que fazem parte da sequência original, para o cálculo do **LZ**.

Para a obtenção desta medida, deve-se seguir o procedimento abaixo:

- a) Define-se o comprimento da janela (j) e o comprimento do salto (s), que consiste determinar em que posição iniciará a nova janela. Exemplo: Para uma janela de tamanho igual a 4 e salto de tamanho igual a 3, a sequência 0123456789 terá as seguintes subsequências: 0123, 3456, 6789;
- b) Extraí-se o **LZ** de cada janela; e
- c) Calcula-se a média dos **LZ** de todas as janelas. Esta estatística é representativa do **LZ** da sequência toda.

Para o cálculo da entropia de Shannon (h), foi feita a suposição, neste trabalho, de que cada janela teria sido gerada a partir de uma fonte ergódica. Com isso, as probabilidades de ocorrência dos caracteres "0", "1" e "2" têm como estimativas as frequências relativas em que aparecem em cada janela.

Após o cálculo de LZ para a série temporal original, foram construídas outras séries temporais cujos dados possuíam intervalos a cada i ticks, com i variando de 1 a 200 negócios, mantendo-se sempre a primeira negociação de cada dia, para que se pudesse captar a chegada de informação durante o período de fechamento da bolsa. Para todas as séries foram calculados os valores de LZ e estes foram comparados estatisticamente para a definição dos intervalos de tempo em que houvesse maior capacidade preditiva dos preços. Também foi calculado o índice de eficiência relativa proposto por Giglio *et. al.* (2008), denotado por IG , que consiste no número de janelas em que o LZ assume valor maior ou igual a 1, em relação ao número total de janelas analisadas.

Depois de realizados os cálculos, foi definida como regra de observação de previsibilidade de preços o conjunto de intervalos de tempo de negociação que satisfaz $IG < c$, onde c é dado pelo maior grau de eficiência relativa (aleatoriedade) admitido pelo agente para simplificar as regiões de observação dos retornos do ativo em estudo. A partir desta regra, o agente pode utilizar menos dados para poder verificar padrões em preços observando o comportamento dos retornos ao longo de um dos intervalos de tempo que pertence ao conjunto mencionado.

4 Análise e Interpretação dos Dados

O estudo realizado foi sobre as ações preferenciais da empresa Eletropaulo S.A. (Código: ELPL6) negociadas na bolsa de valores BM&F-BOVESPA. A base de dados utilizada foi uma série de preços de alta frequência. O período utilizado para análise foi de 1º de janeiro de 2007 a 31 de dezembro de 2008, O número de negociações realizadas (registros de preços) foi de 451.312.

Foram calculados os valores de LZ médio, seu desvio padrão e a medida IG (número de janelas com LZ médio maior que 1 em relação ao número total de janelas) para diferentes intervalos de negociação (de 1 a 200) utilizando janelas de tamanho 1000 caracteres, saltos de tamanho 1 caractere e regiões de estabilidade de tamanhos 0,05% e 0,25%.

As figuras 2 e 3 mostram a evolução do LZ médio e do IG com a distância entre as negociações. Foi possível observar que, à medida que se aumenta a distância entre negociações consecutivas, o LZ médio aumenta de valor, indicando a presença de aleatoriedade para intervalos de negociação muito maiores que 1 tick. A diminuição do desvio padrão do LZ médio apresenta uniformidade para $\rho = 0,25\%$ e apresenta dispersão a partir do intervalo 100 para o caso $\rho = 0,05\%$.

Já para o IG , este apresentou aumento de valor com relação à evolução dos intervalos entre negociações e foi observado que seu desvio padrão possui uma diminuição dispersa após um curto aumento de valor. Tal aumento pára em 9 negócios para $\rho = 0,05\%$ e em 26 negócios para $\rho = 0,25\%$.

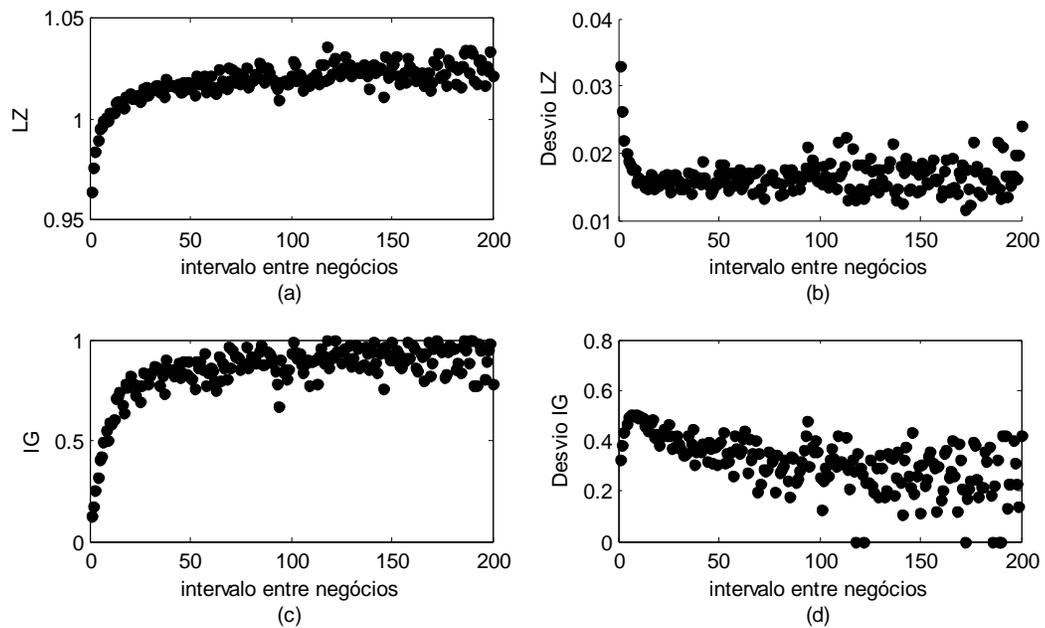


Figura 2: Diagramas das estatísticas LZ médio (a) e IG (c) e de seus respectivos desvios padrão (b) e (d). A região de estabilidade utilizada foi de 0,05%.

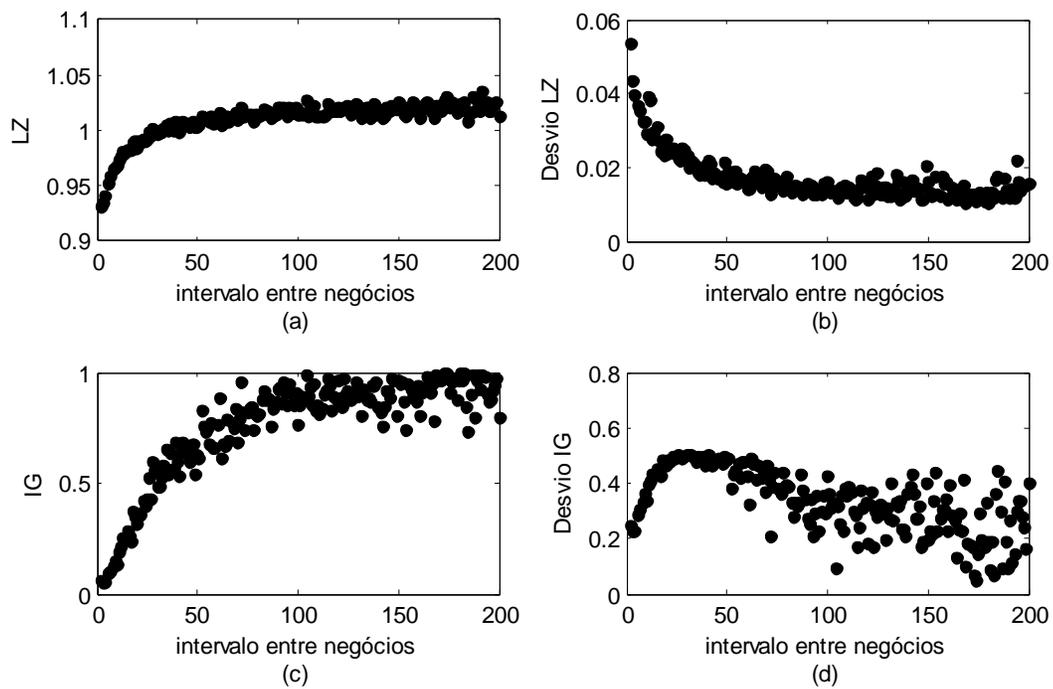


Figura 3: Diagramas das estatísticas LZ médio (a) e IG (c) e de seus respectivos desvios padrão (b) e (d). A região de estabilidade utilizada foi de 0,25%.

Na etapa seguinte, foi calculado o conjunto de intervalos de tempo (em números de negócios) em que os padrões entre retornos pudessem desviar da aleatoriedade em pelo menos $(1-c)\%$ das janelas, ou seja, $IG < c$.

Como o IG é uma estatística obtida a partir de uma amostra, foi utilizado um teste de hipóteses para proporções, onde a hipótese a ser testada foi $H_0: IG \geq c$ contra $H_a: IG < c$. Para um nível de significância de 1%, foram determinados os intervalos de tempo adequados para a observação dos retornos em diferentes valores de c e de região de estabilidade. A Tabela 1 mostra os resultados obtidos.

Tabela 1: Intervalos de tempo que obedecem à regra de observação $IG < c$.

Região de estabilidade 0,05%		Região de estabilidade 0,25%	
c	Intervalos de tempo (em números de negócios) onde $IG < c$	C	Intervalos de tempo (em números de negócios) onde $IG < c$
10%	NÃO HÁ	10%	1 a 6.
15%	1.	15%	1 a 10.
20%	1, 2.	20%	1 a 11.
25%	1, 2.	25%	1 a 12.
30%	1, 2, 3.	30%	1 a 17.
35%	1, 2, 3, 4.	35%	1 a 17.
40%	1, 2, 3, 4.	40%	1 a 22.
45%	1 a 6.	45%	1 a 26.
50%	1 a 7.	50%	1 a 27, 30 e 31.
55%	1 a 7.	55%	1 a 27, 30 a 32.
60%	1 a 11.	60%	1 a 38.
65%	1 a 12.	65%	1 a 42, 45 a 51,
70%	1 a 12, 16 e 17.	70%	1 a 51, 58 e 59, 62 a 64.
75%	1 a 17,	75%	1 a 51, 58 e 59, 62 a 64
80%	1 a 26, 93 e 94, 191 e 192	80%	1 a 68, 70 e 71.
85%	1 a 31, 35 e 36, 45 e 46, 50 a 52, 58 e 59, 62 e 63, 93 e 94, 191 e 192.	85%	1 a 81, 87 e 88, 141 e 142.
90%	1 a 68, 75 a 77, 86 e 87, 91 a 99, 105 a 107, 134 a 136, 154 a 155, 162 a 163, 169 a 170, 176 a 177, 191 e 192	90%	1 a 89, 100 e 101, 109 a 112, 121 e 122, 134 a 136, 140 a 143, 153 e 154, 184 e 185, 195 e 196.
95%	Todos os intervalos analisados, exceto 122 a 124, 126 e 127, 132 e 133, 140 e 141, 171 a 175, 178 a 181, 185 a 190, 193 a 195, 198 e 199.	95%	1 a 136, 140 a 146, 153 a 163, 184 a 186, 194 a 197.

Para que possa ser compreendida a regra de observação, foi tomado como exemplo um elemento do conjunto de intervalos de tempo que satisfaz $IG < 10\%$ para a região de estabilidade de $0,25\%$ (intervalo igual a 5 *ticks*). Ou seja, ao se observar as variações de preços a cada 5 negócios (levando em consideração, também, o preço obtido na primeira negociação de cada dia), há uma chance alta de que sejam encontrados padrões na série temporal, dado que menos de 10% das janelas apresentam indícios de aleatoriedade neste caso.

O tamanho de janela utilizado, bem como o tamanho de salto não foram alterados na análise dos cenários, pois a ideia foi a de comparar padrões de retornos em sequências produzidas por uma mesma fonte e, de acordo com Ranciaro Neto (2010), não há alterações significativas no padrão de comparação do LZ médio de duas sequências ao se alterar o comprimento daquelas grandezas. Naquele trabalho foram utilizadas janelas de tamanhos 5000 a 30000 e saltos variando de 1000 a 10000. Giglio, Matsushita e Silva (2008) utilizou janelas de tamanho 1000 e saltos de comprimento igual a 1. Mas, ao se alterar a região de estabilidade, há evidência de alterações da medida LZ médio e IG em magnitudes variadas dependendo do tipo de configuração dos caracteres na série temporal analisada (RANCIARO NETO, 2010).

Foram adotadas sequências de tamanho 1000, pois de acordo com Kaspar e Schuster (1987), a probabilidade de convergência da estatística LZ é inferior a 5% . Ao adotar valores muito grandes de janela, poder-se-ia incorrer em medidas de LZ iguais ou superiores à unidade sem haver aleatoriedade na sequência.

A interpretação utilizada para as regiões de estabilidade é a de que elas seriam um tipo de grau de tolerância das variações dos retornos de um ativo. Por exemplo, em uma região de estabilidade de $0,05\%$, são desconsiderados os movimentos nos retornos que variem de $0,05\%$ para baixo a $0,05\%$ para cima. Valores que recaem sobre tal intervalo ficam codificados sob um mesmo tipo de caractere. Ao se aumentar a região de estabilidade, as mudanças são desconsideradas em um intervalo maior de valores, diminuindo, conseqüentemente, o valor de LZ e aumentando a ocorrência de padrões na sequência. Para o exemplo apresentado anteriormente, caso fosse utilizada a região de estabilidade $0,05\%$, não seria possível observar, para intervalos de 5 negócios, padrões nos retornos em que menos de 10% das janelas possuam aleatoriedade.

Observando os resultados da Tabela 1, verificou-se que o número de intervalos de tempo que $IG < c$, aumentava para crescimento no valor de c . Houve concentração do número de intervalos de menor tamanho e, à medida que o tamanho do intervalo aumentava, um número menor e mais disperso destes satisfazia a regra de observação. Tal fato pode ser explicado pelos efeitos das fricções do mercado, que se tornam mais aparentes em séries de preços de alta frequência (CAMPBELL, *et al.* 1997).

A Tabela 2 apresenta os intervalos em que o LZ médio é estatisticamente menor que 1 ao nível de significância de 1% para os casos de região de estabilidade iguais a $0,05\%$ e a $0,25\%$.

Tabela 2: Intervalos de tempo (em número de negócios) em que o LZ médio é inferior a 1.

Região de estabilidade 0,05%	Região de estabilidade 0,25%
1 a 9	1 a 32

Tal resultado corrobora a presença de padrões na série de retornos para intervalos de tempo pequenos, dado que o valor da estatística ao longo das janelas tende a se afastar de 1, na média.

5 CONCLUSÕES

A partir do uso da medida LZ e IG, não foi possível descrever o tipo de padrão existente entre retornos em um dado intervalo de tempo entre eles. O que se tornou viável, foi descrever, em termos probabilísticos, em quais intervalos podem ocorrer padrões para um determinado ativo. A medida define um padrão, mas aquela sofre com flutuações estatísticas podendo, em algumas observações, mostrar-se errônea apesar de haver uma chance pequena para sua ocorrência (1%).

Porém, o custo de se analisar um volume grande de informação em busca de padrões fica reduzido no curto prazo, e, conseqüentemente, os efeitos provocados pelas flutuações devido à transmissão de novas informações sobre os preços ficam reduzidos, facilitando tomada de decisão no sentido de se aumentar a precisão e de se diminuir o tempo, pois a aplicação de modelos que utilizam muitos dados elevam o custo computacional de sua implementação.

Deve-se atentar para o fato que, para um prazo mais longo, poderá haver alterações na evolução das séries temporais, invalidando a análise em questão como, por exemplo, a ideia da imprevisibilidade dos efeitos de alterações na política econômica (LUCAS, 1976) ou, de acordo com Mises (2010), a evolução para o estágio avançado de um ciclo econômico gerado pela expansão exagerada da oferta monetária, onde os agentes não conseguem mais projetar seus resultados devido ao excesso de oferta induzido em alguns setores e à inflação de preços, o que agravaria a incerteza sobre os preços futuros.

Sugere-se aplicar as medidas com outras informações, como bid-ask spread, para auxiliar na determinação de padrões de liquidez de um ativo, o que em conjunto com a análise de preços poderia gerar mais informação para o negociante definir suas estratégias dentro de seu perfil de investimento.

Os intervalos de tempo foram considerados constantes (em números de negócios) na análise. Pode-se desenvolver outra medida com tempo físico (horas, minutos e segundos), com regras de mudança temporal ou em períodos definidos do dia a fim de se captar regularidades ou anomalias relativas aos preços dos ativos analisados.

Referências

- Bachelier, L. (1900). Théorie de la speculation. **Annales scientifiques de l'É.N.S.** 3^e série, v. 17, p. 21-86. Recuperado em 15 de abril de 2012 em http://archive.numdam.org/ARCHIVE/ASENS/ASENS_1900_3_17_/ASENS_1900_3_17_21_0/ASENS_1900_3_17_21_0.pdf.
- Campbell, J. Y., Lo, A. W.; & Mackinlay, A. C. (1997). **The Econometrics of Financial Markets**. Princeton University Press, Princeton. New Jersey.
- Chaitin, G. J. (1966). On the length of programs for computing finite binary sequences. **Journal of the ACM**, v. 13, n. 4, p. 547-569.
- De Bruijn, N. G. (1975). **Acknowledgement of Priority to C. Flye Sainte-Marie on the counting of circular arrangements of 2n zeros and ones that show each n-letter word exactly once**, T.H.-Report 75-WSK-06, Technological University Eindhoven, 13 p.
- Fama, E. F. (1970). Efficient Capital Markets: A review of Theory and Empirical Work. **Journal of Finance**. v. 25, n. 2, p. 383-417.
- Fama, E. F. (1991). Efficient Capital Markets II. **Journal of Finance**. v. 46, n. 5, p. 1575-1617.
- Giglio, R., Matsushita, R., & Silva, S. (2008). The relative efficiency of stockmarkets **Economics Bulletin**. v. 7, n. 6, p. 1-12.
- Grossman, S., & Stiglitz, J. (1980). On the impossibility of informationally efficient markets. **The American Economic Review**, v. 70, n. 3, p. 393-408.
- Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. **Journal of Financial Economics**, v.6, n. 2/3, p. 96-101.
- Kaspar, F., & Schuster, H. (1987). Easily calculable measure for the complexity of spatiotemporal patterns. **Physical Review A**, v/ 36, n. 2, p. 842-848.
- Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. **Problems of Information Transmission**. v. 1, n. 1, p. 3-11.
- Lempel, A., & Ziv, J. (1976). On the complexity of finite sequences. **IEEE Transactions on Information Theory**, v. 22, n. 1, p. 75-81.
- Lo, A., & Mackinlay, C. (1988). Stock market prices do not follow random walks: evidence from a simple specification test. **Review of Financial Studies**. v. 1, n. 1, p. 41-66.
- Lucas, R. E. (1976). Econometric Policy Evaluation: A Critique. In K. Brunner and A.H. Meltzer (eds.). *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, v. 1, **Journal of Monetary Economics**, supplementary issue, p. 19-46.
- Lucas, R. E. (1978) Asset prices in an exchange economy. **Econometrica**. v.46, p. 1429-1446.
- Malkiel, B. G. (2003). The efficient market hypothesis and its critics. **CEPS Working Paper n. 91.**, 47 p.

Mantegna, R. N., & Stanley, H. E. (1999). **An introduction to Econophysics: correlation and complexity in finance**. Cambridge University Press.

Mises, L. (2010). **Ação Humana: Um tratado de Economia**. 3 ed. São Paulo: Instituto Ludwig von Mises Brasil. Texto original impresso em 1966.

Ranciaro Neto, A. (2010). **Teoria da informação algorítmica, eficiência relativa de mercado e perda de memória em séries de retornos de alta frequência em ativos negociados na BM&F-BOVESPA**. Dissertação de Mestrado, Economia. Universidade Federal de Alagoas.

Rukhin, A. L. (2000). Testing randomness: a suite of statistical procedures. **Theory of Probability and its Applications**, v. 45, n. 1, p. 111-132.

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. **Industrial Management Review**. v. 5, n. 1, p. 41-49.

Shannon, C. E. (1948). A Mathematical Theory of Communication. **Bell System Technical Journal**, v. 27, p. 379-423.

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

Solomonoff, R. J. (1964). A formal theory of inductive inference, Part I and Part II. **Inform. and Control A**, n. 7, p. 1-22 e 224-254.

Taufemback, C., Giglio, R., & Silva, S. (2011). Algorithmic Complexity Theory Detects Decreases in the Relative Efficiency of Stock Markets in the Aftermath of the 2008 Financial Crisis **Economics Bulletin** v. 31, n.2, p. 1631-1647.

<p>Adhemar Ranciaro Neto é Mestre em Ciências Econômicas pela Universidade Federal de Alagoas (Ufal), Aluno de Pós Graduação em Física da Matéria Condensada pela Ufal, e professor nesta mesma instituição - adhranneto@gmail.com Av. Lourival Melo Mota, s/n, bl. 16, 1º Andar Tabuleiro dos Martins, Maceió – AL 57072-970</p>	<p>Iram Marelo Gléria é Doutor em Física pela UNB, e professor na Ufal - iramgleria@hotmail.com Av. Lourival Melo Mota, s/n, bl. 12, 1º Andar Tabuleiro dos Martins, Maceió – AL 57072-970</p>
<p>Luiz Carlos Marques dos Anjos é mestre em Ciências Contábeis pela UFPE e doutorando em Ciências Contábeis pelo Programa Multi-Institucional e Inter-Regional UNB / UFPB / UFRN, e professor na Ufal. lcjanjos@feac.ufal.br Av. Lourival Melo Mota, s/n, bl. 16, 1º Andar Tabuleiro dos Martins, Maceió – AL 57072-970</p>	<p>Daniel José Cardosos da Silva é mestre em Ciências Contábeis pela UFPE e professor na Ufal. lcjanjos@feac.ufal.br Av. Lourival Melo Mota, s/n, bl. 16, 1º Andar Tabuleiro dos Martins, Maceió – AL 57072-970</p>