



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO SÓCIO-ECONÔMICO
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA
ÁREA DE CONCENTRAÇÃO: ECONOMIA E FINANÇAS

RICARDO FARIA GIGLIO

EFICIÊNCIA RELATIVA DE MERCADO SOB A PERSPECTIVA DA
TEORIA DA INFORMAÇÃO ALGORÍTMICA

Florianópolis
2008

RICARDO GIGLIO

EFICIÊNCIA RELATIVA DE MERCADO SOB A PERSPECTIVA DA
TEORIA DA INFORMAÇÃO ALGORÍTMICA

Dissertação apresentada como requisito parcial à obtenção do título de Mestre em Ciências Econômicas do Programa de Pós-Graduação em Economia da Universidade Federal de Santa Catarina (UFSC).

Orientador: Prof. Dr. Eraldo Sergio da Silva.

Florianópolis
2008

Giglio, Ricardo Faria.

Eficiência relativa de mercado sob a perspectiva da Teoria da Informação Algorítmica.

Ricardo Faria Giglio. Florianópolis, Universidade Federal de Santa Catarina – UFSC, 2008.

Dissertação (Mestrado em Economia – UFSC – Centros Sócio-Econômico – Programa de Pós-Graduação em Economia/PPGE, 2008).

Bibliografia.

Inclui Anexos.

1. Eficiência relativa de mercado 2. Teoria da Informação 3. Medidas de complexidade

RICARDO GIGLIO

**EFICIÊNCIA RELATIVA DE MERCADO SOB A PERSPECTIVA DA TEORIA DA
INFORMAÇÃO ALGORÍTMICA**

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Economia e aprovada, em sua forma final, pelo Programa de Pós-Graduação em Economia da Universidade Federal de Santa Catarina.

Prof. Dr. Roberto Meurer
Coordenador do Programa

Prof. Dr. Eraldo Sergio da Silva
Orientador

Prof. Dr. Newton Carneiro Affonso da Costa Jr.
Examinador interno

Prof. Dr. Annibal Dias de Figueiredo Neto
Instituto de Física – Universidade de Brasília

Prof. Dr. Iram Marcelo Gleria
Instituto de Física – Universidade Federal de Alagoas

Dedico este trabalho a meus pais, à Paula e a todos que o escreveram junto comigo.

Agradecimentos

Primeiramente agradeço à minha família por todo o apoio que me deram. Meu pai Vagner, minha mãe Cleusa, minha irmã Fernanda e minha avó Olga; muito obrigado de verdade. Com certeza este trabalho não teria sido realizado sem todos os tipos de ajuda e incentivo que recebi de vocês antes, durante e depois do mestrado. Agradeço também a Toninho e Mercedes pela preocupação comigo e pelas diversas demonstrações de carinho. Agradeço separadamente à minha eterna namorada Paula que me acompanhou de perto durante essa jornada e soube ter a paciência e o companheirismo necessários neste período de dedicação.

Em segundo lugar devo agradecer a todos que me proporcionaram um excelente ambiente acadêmico e pessoal durante estes dois anos na Universidade Federal de Santa Catarina, mais precisamente no Programa de Pós-Graduação em Economia do Centro Sócio-Econômico. Agradecimentos sinceros e carinhosos à Evelise Elpo, Roberto Meurer e Silvio Cairo.

Agradecimento especial ao meu orientador Sergio da Silva, que me ensinou muito sobre o trabalho do pesquisador e que será para mim sempre uma referência de um excelente profissional acadêmico.

Agradeço a todos os meus amigos, cuja lista felizmente não caberia em uma página de agradecimentos, que me apoiaram seja pela simples presença de espírito ou por meio de profundas discussões técnicas. Em especial aos amigos do mestrado, do Tatuapé e da Getulio Vargas. Não citarei nomes, sintam-se todos agradecidos.

Em último lugar, mas não menos importante, agradeço a Joseph por estar sempre do meu lado me iluminando quando os problemas se mostram maiores do que realmente são. Muito obrigado a todos vocês.

Resumo

Grande parte da literatura acerca das Hipóteses de Eficiência de Mercado (HEM) relaciona eficiência de um mercado à flutuação aleatória do preço de seus ativos. Entretanto, podemos considerar mercados perfeitamente eficientes em termos absolutos como abstrações que não ocorrem na prática (GROSSMAN, STIGLITZ, 1980), mas que servem como útil referência para o conceito de eficiência relativa (CAMPBELL; LO; MACKINLAY, 1997).

Além disso, alguns testes empíricos sobre a validade das HEM encontram uma séria dificuldade no que tange à formulação de hipóteses conjuntas (LO; MACKINLAY, 1999). A rejeição da hipótese de eficiência não determina se o mercado é realmente ineficiente ou se o modelo construído *a priori* para descrever o comportamento normal dos preços foi mal especificado.

Este trabalho utiliza a medida de complexidade proposta por Lempel e Ziv (1976), que pode ser entendida como uma medida do grau de aleatoriedade de uma seqüência (KASPAR; SCHUSTER, 1987), para verificar a eficiência de diversos mercados financeiros em termos relativos. Para tal, analisa 36 índices de ações, ações individuais de 37 empresas selecionadas e 19 taxas de câmbio relacionadas ao dólar norte-americano.

Esta medida apresenta duas características importantes: a utilização de uma série genuinamente aleatória como referência para um mercado idealmente eficiente (problema da impossibilidade de mercados perfeitamente eficientes) e a independência de um modelo especificado *a priori* para descrever o comportamento dos preços (problema da hipótese conjunta).

A quantidade de informação redundante presente nas séries financeiras, representada pelos desvios da condição de aleatoriedade, foi calculada para estabelecer comparações de eficiência em termos relativos. Sob esta medida de eficiência, constatou-se que, por exemplo, o índice *S&P500* é 99,1% eficiente e o índice *Bovespa*, 67,8% eficiente em relação a uma série genuinamente aleatória.

Abstract

The traditional market efficiency literature relates market efficiency to random price fluctuations. Yet a perfect efficient market is an idealization unattainable in practice (GROSSMAN; STIGLITZ, 1980), but that serves as a useful benchmark for measuring relative efficiency (CAMPBELL; LO; MACKINLAY, 1997).

Besides, any empirical test of the efficient market hypothesis is a joint test of market efficiency and the equilibrium model assumed that defines normal asset returns. If the efficiency hypothesis is rejected, this could be because either the market is really inefficient or a poor model of price behavior has been specified.

This work uses the complexity measure proposed by Lempel and Ziv (1976), which can be interpreted as a measure of the degree of randomness of a string (KASPAR; SCHUSTER, 1987), to assess the efficiency of financial markets in relative terms. This has been done by considering data from 36 stockmarket indices, 37 individual company stock prices, and 19 exchange rates.

This approach presents two main advantages. Firstly, it measures the deviation from randomness as the degree of efficiency of a given market. And secondly, it does not require an equilibrium model for describing price behavior (the above joint hypotheses problem).

Efficiency in relative terms has been assessed by measuring the amount of redundant information (deviation from randomness) in financial time series. Doing so, we have found, for instance, that the *S&P500* index is 99.1% efficient whereas *Bovespa* index is only 67.8% efficient.

Sumário

1 Introdução	11
2 Hipóteses de Eficiência de Mercado (HEM)	14
2.1 Introdução: eficiência de mercado	14
2.2 Taxonomia	15
2.3 Eficiência e aleatoriedade de preços	17
2.4 Alternativa às HEM: Eficiência relativa	19
3 Método	22
3.1 Contexto: aleatoriedade e complexidade	22
3.2 Medida de complexidade de Lempel e Ziv (LZ)	29
3.2.1 Formulação matemática	29
3.2.2 Construção operacional	32
3.3 Normalizações	35
3.3.1 Medida de complexidade máxima	35
3.3.2 Medida da entropia da fonte geradora	37
3.4 Aplicação em Finanças	39
3.5 Propriedades	41
3.6 Comparação com a estatística Entropia Aproximada	45
4 Experimentos	48
4.1 Bolsas	49
4.2 Ações	52
4.3 Câmbio	55
5 Conclusão	58
Referências bibliográficas	60
Apêndice	63

Lista de figuras

Figura 1 Representação artística de uma Máquina de Turing. Disponível em: <http://pt.wikipedia.org/wiki/Imagem:Turing_Machine.png>. Acesso em: 30 jan. 2008.

Figura 2 Dois experimentos que apresentam a medida LZ para 14000 janelas deslizantes de seqüências geradas pela (a) distância entre os 15001 primeiros números primos e (b) para o logaritmo natural destas distâncias.

Figura 3 Apresentação do mapa logístico para quatro diferentes valores do parâmetro de crescimento ($r=2,5$, $r=3,23$, $r=3,45$ e $r=3,92$). Para valores de $r < 3$, existe estabilidade em algum ponto fixo. Para valores entre $3 \leq r \leq 4$, não há estabilidade e o sistema entra em crescente complexidade.

Figura 4 Apresenta cinco gráficos que relacionam a medida LZ (ordenada) para 1000 janelas deslizantes (abscissa) para quatro bolsas selecionadas, sendo (a) S&P500 (LZ médio = 1,0232, grau de eficiência = 99,10%), (b) DAX30 (LZ médio = 1,0257, grau de eficiência = 98,40%), (c) *All Ords.* Austrália (LZ médio = 1,0246, grau de eficiência = 97,80%), (d) Shanghai Composite Index (LZ médio = 1,0032, grau de eficiência = 49,50%) e (e) Karachi 100 (LZ médio = 0,9918, grau de eficiência = 23,70%).

Figura 5 Apresenta cinco gráficos que relacionam a medida LZ (ordenada) para 1000 janelas deslizantes (abscissa) para preço de ações de quatro empresas selecionadas, sendo (a) Coca-Cola (LZ médio = 1,0324, grau de eficiência = 100,00%), (b) Yahoo! (LZ médio = 1,0310, grau de eficiência = 99,65%), (c) Vale (LZ médio = 1,0193, grau de eficiência = 92,75%), (d) Toyota (LZ médio = 1,0100, grau de eficiência = 76,32%) e (e) Aracruz (LZ médio = 1,0048, grau de eficiência = 66,67%).

Figura 6 Apresenta cinco gráficos que relacionam a medida LZ (ordenada) para 1000 janelas deslizantes (abscissa) para taxas de câmbio em relação ao dólar norte-americano (US\$) de quatro moedas selecionadas, sendo (a) Libra esterlina (LZ médio = 1,0223, grau de eficiência = 99,81%), (b) Euro (LZ médio = 1,0254, grau de eficiência = 99,45%), (c) Peso mexicano (LZ médio = 1,0254, grau de eficiência = 96,58%), (d) Real (LZ médio = 1,0156, grau de eficiência = 92,60%) e (e) Rúpia indiana (LZ médio = 0,9958, grau de eficiência = 43,54%).

Figura 7 Fluxograma que representa o algoritmo utilizado para calcular a medida de complexidade como descrito por Kaspar e Schuster (1987).

Lista de gráficos

Gráfico 1 A complexidade normalizada c/b de uma seqüência aleatória se aproxima por cima de seu limite superior assintótico $\lim_{n \rightarrow \infty} c/b = 1$.

Gráfico 2 Comportamento da entropia da fonte $h(p)$ para valores de p no intervalo $0 \leq p \leq 1$. A entropia é mínima ($h=0$) quando um dos caracteres aparece com probabilidade $p=1$ e máxima ($h=1$) quando $p=0,5$.

Gráfico 3 Mostra a medida de complexidade normalizada LZ de uma seqüência de 15000 números pseudo-aleatórios gerados por computador em um experimento com 14000 janelas deslizantes.

Gráfico 4 Mostra a complexidade do mapa logístico em função de seu parâmetro de crescimento r . O gráfico mostra que a medida LZ é robusta para diferenciar seqüências de complexidades distintas.

Lista de Quadros

Quadro 1 Construção operacional da medida de complexidade da seqüência $S = 0110001$. Ilustra o procedimento utilizado para medir a complexidade de uma seqüência como descrito em Kaspar e Schuster (1987).

Lista de tabelas

Tabela 1 Mostra os resultados completos do experimento com 36 bolsas selecionadas, ordenadas pelo grau de eficiência. Apresenta a bolsa analisada, seu país, índice LZ médio do procedimento de janelas e o grau de eficiência, que corresponde à probabilidade de encontrar janelas com $LZ > 1$.

Tabela 2 Mostra os resultados completos do experimento com ações de 37 empresas selecionadas, ordenadas pelo grau de eficiência. Apresenta a ação analisada, sua respectiva bolsa, índice LZ médio do procedimento de janelas e o grau de eficiência, que corresponde à probabilidade de encontrar janelas com $LZ > 1$.

Tabela 3 Mostra os resultados completos do experimento com 19 taxas de câmbio selecionadas, todas em relação ao dólar americano, ordenadas pelo grau de eficiência. Apresenta a moeda analisada, seu respectivo país, índice LZ médio do procedimento de janelas e o grau de eficiência, que corresponde à probabilidade de encontrar janelas com $LZ > 1$.

Lista de abreviaturas

<i>S&P 500</i>	Standard & Poor's
<i>DAX 30</i>	Deutsche Aktien Xchange
<i>Nikkei 225</i>	Nihon Keizai Shimbun
<i>All Ords.</i>	Australian All Ordinaries Index
<i>ATX</i>	Austrian Traded Index
<i>Dow Jones</i>	Dow Jones Industrial Average
<i>Korea Composite</i>	Korea Composite Stock Price Index
<i>TA 100</i>	Tel-Aviv 100
<i>Hang Seng</i>	Hang Seng Index
<i>Strait Times</i>	Strait Times Index
<i>CAC 40</i>	Cotation Assistée en Continu
<i>Helsinki General</i>	Helsinki General Index
<i>Kuala Lumpur SE</i>	Kuala Lumpur Stock Exchange
<i>FTSE 100</i>	Financial Times Stock Index
<i>PX 50</i>	Prague Stock Exchange
<i>Bel 20</i>	Belgium Stock Exchange
<i>IBC</i>	Indice Bursátil Caracas
<i>Madrid General</i>	Madrid General Index
<i>Swiss Market</i>	Swiss Market Index
<i>Nasdaq</i>	Nat. Ass. of Securities Dealers Automated Quotations
<i>Amsterdam EX</i>	Amsterdam Exchange Index
<i>Bovespa</i>	Bolsa de Valores de São Paulo
<i>IPC</i>	Indice de Precios y Cotaciones
<i>Merval</i>	Mercado de Valores
<i>Jakarta Composite</i>	Jakarta Composite Index
<i>Istanbul 100</i>	Istambul Stock Exchange
<i>Moscow Times</i>	Moscow Times Index
<i>KFX</i>	Copenhagen Stock Exchange
<i>Athens Composite</i>	Athens Composite
<i>Bombay SE</i>	Bombay Stock Index
<i>Taiwan Weighted</i>	Taiwan Weighted Index
<i>Shanghai Composite</i>	Shanghai Composite Index
<i>PSE</i>	Phillipines Stock Exchange
<i>Lima General</i>	Lima General Index
<i>Karachi 100</i>	Karachi 100 Index
<i>Colombo SE</i>	Colombo Stock Exchange

1 Introdução

As Hipóteses de Eficiência de Mercado (HEM) constituem um dos esquemas teóricos mais difundidos e testados empiricamente na literatura econômica. Segundo Jensen (1978), esse interesse pelas condições de eficiência se intensificou nas décadas de 50 e 60 sob a denominação de “*random walk theory*” (teoria do passeio aleatório) no campo das finanças e de “*rational expectations theory*” (teoria das expectativas racionais) na literatura *mainstream*.

As HEM, de modo geral, relacionam o conceito de eficiência de um mercado à flutuação aleatória de seus preços, baseando-se na suposição de que, se todas as expectativas e informações disponíveis aos agentes forem integralmente incorporadas aos preços dos ativos, estes se comportariam de maneira randômica, acarretando a impossibilidade de alguém obter lucros certos.

Grande parte dos trabalhos nessa linha de pesquisa se preocupa em estabelecer testes no sentido de estudar a eficiência de um mercado específico, sob diferentes formulações das hipóteses de eficiência de mercado. Todavia, existem algumas dificuldades nos trabalhos práticos referentes a problemas de hipóteses conjuntas (FAMA, 1991) e à impossibilidade de mercados perfeitamente eficientes (GROSSMAN; STIGLITZ, 1980).

De maneira geral, as HEM podem ser entendidas como uma aplicação da clássica condição de equilíbrio “lucro econômico igual a zero” de mercados perfeitamente competitivos ao comportamento dinâmico dos preços de mercados especulativos sob condições de incerteza (JENSEN, 1978). Assim, podemos pensar as condições de eficiência como situações limites inatingíveis em mercados reais, mas que servem como útil referência para o conceito de eficiência em termos relativos (CAMPBELL; LO; MACKINLAY, 1997).

Na prática, as informações e as expectativas não são homogeneamente distribuídas entre os agentes do mercado. Existem custos para a aquisição de informações, e pode não ser marginalmente interessante para todos os agentes empenharem

recursos para obtê-las. Há ainda as informações privadas, que não podem ser adquiridas por outros agentes. De outro modo, podemos dizer que ineficiências residuais estão sempre presentes, em maior ou menor grau, em mercados reais. Dessa forma, as condições de eficiência podem ser vistas não como uma pergunta do tipo sim ou não, mas sim como desvios em relação ao mercado perfeito idealizado e sem fricções.

Existe uma linha de pesquisa em finanças usualmente relacionada aos termos “não linearidade”, “teoria do caos” e “teoria da complexidade” que trabalha com uma definição alternativa de eficiência de mercado associada à idéia de complexidade (BROCK; LIMA, 1996). Mercados financeiros podem ser vistos como sistemas complexos no sentido de que transmitem informação sobre a negociação de determinado ativo na sua série histórica de preços.

A Teoria da Informação Algorítmica, conforme Mantegna e Stanley (1999), pode auxiliar o entendimento do comportamento de séries temporais financeiras no sentido de avaliar se estas possuem propriedades estatísticas que as assemelhem a outras séries de tempo puramente aleatórias.

Nesse sentido, podemos traçar um paralelo com as hipóteses de eficiência de mercado, pois uma série é dita aleatória quando não conseguimos comprimir seu tamanho por meio de nenhum algoritmo (não existem redundâncias ou padrões). Em outros termos, dizemos que tem complexidade máxima, e a menor codificação existente para a série é ela mesma.

Sob a perspectiva de que um mercado é dito eficiente quando seus preços flutuam de maneira aleatória, intuitivamente podemos concordar que determinado mercado seja considerado ineficiente se apresenta padrões, de qualquer ordem, nos seus preços históricos, sendo que isso o afaste da condição de aleatoriedade (alguém pode se utilizar destes padrões para obter lucros certos) e vice-versa.

O presente trabalho utiliza a medida de complexidade proposta por Lempel e Ziv (1976), que é tida como uma útil medida de complexidade (KASPAR; SCHUSTER, 1987), para estabelecer comparações relativas entre diferentes mercados quanto à

questão da aleatoriedade de seus preços e, por decorrência, de seus graus de eficiência.

Assim, foram realizados experimentos com índices de bolsas, ações de companhias individuais e taxas de câmbio quanto aos seus graus relativos de eficiência. Encontrou-se que, por exemplo, sob a medida utilizada neste trabalho, o índice *S&P500* é 99,1% eficiente e o índice *Bovespa*, 67,8% eficiente.

Tal abordagem permite a comparação do grau de aleatoriedade de séries financeiras (e, por decorrência, da eficiência de mercado em termos relativos), pois considera uma série aleatória como referência para máxima aleatoriedade (problema da impossibilidade de mercados perfeitamente eficientes) e por ser uma estatística que é independente de um modelo escolhido, *a priori*, para descrever a flutuação do preço analisado (problema da hipótese conjunta).

Além desta introdução, o trabalho é dividido como segue. O capítulo 2 trata das Hipóteses de Eficiência de Mercado, descreve a taxonomia utilizada pela literatura, ilustra a relação eficiência-aleatoriedade e aborda a impossibilidade de mercados perfeitamente eficientes, o problema de hipótese conjunta e o conceito de eficiência relativa.

O capítulo 3 descreve o método utilizado, assim como o contexto no qual é inserido (Teoria da Informação). Trata também das normalizações necessárias e traz algumas ilustrações sobre as propriedades da medida de complexidade utilizada. O capítulo 4 apresenta a formulação dos experimentos, os resultados obtidos e os dados utilizados. O capítulo 5 conclui o trabalho.

2 Hipóteses de Eficiência de Mercado (HEM)

2.1 Introdução: eficiência de mercado

O conceito de eficiência, seja para qualquer uso possível em diversas áreas do conhecimento, é passível a diversas interpretações. Portanto, em Finanças, devemos nos preocupar em diferenciar o termo eficiência entre seus principais usos possíveis.

A maneira pela qual os recursos disponíveis são utilizados para facilitar as operações de mercado pode ser denominada por *eficiência operacional*. Nesse sentido, estão incluídos fatores como funcionamento dos pregões, serviços financeiros e até a própria tecnologia disponível no momento. Pode-se pensar também a *eficiência alocativa* no sentido de Pareto. Este trabalho não se preocupará com tais abordagens de eficiência, mas sim com outra caracterizada como *eficiência informacional*, que é objeto das Hipóteses de Eficiência de Mercado (HEM).

Fama (1991) aponta que existem diversas definições similares para as HEM. A exibida por Mantegna e Stanley (1999) estabelece que um mercado seja considerado eficiente se toda a informação disponível é instantaneamente processada quando chega ao mercado e imediatamente refletida em um novo valor de preços dos ativos negociados. Outras definições similares podem ser encontradas em Dimson e Mussavian (1998) e Fama (1970).

Sob o ponto de vista marginalista, a definição apresentada por Jensen (1978) procura imprimir mais realismo na questão, afirmando que os preços refletem apenas as informações cujos custos de obtenção não excedam os benefícios de sua utilização.

Dada a chegada de uma nova informação, podemos checar as condições apontadas por Fama (1991) para classificar um mercado como eficiente, sendo:

- a) não existem custos de transação na negociação do ativo;

b) todas as informações estão disponíveis a todos os agentes do mercado;

Se houverem custos e apenas para alguns agentes for marginalmente interessante realizar a negociação ou obter a informação, esses agentes obterão lucros certos e os preços não incorporarão integralmente as informações disponíveis.

c) todos os agentes têm a mesma opinião sobre os impactos desta nova informação sobre os preços futuros e atuais do ativo negociado.

Sob estas condições de eficiência de mercado, os agentes concordam que a nova informação é positiva para a valoração do ativo, assim como concordam sobre o impacto da referida informação sobre seu preço, tanto atual como futuro. Em outras palavras, as expectativas são homogêneas.

Retomando as definições de mercado eficiente, se tais condições forem atendidas, o preço do ativo em questão se deslocará para um novo patamar que reflita integralmente as expectativas dos agentes. Este ajustamento é bem exemplificado por Samuelson (1965, p. 41), ao afirmar que em mercados competitivos existe um comprador para cada vendedor, e que se alguém pudesse estar certo de que um preço subiria, ele já teria subido.

De modo intuitivo, imagina-se um mercado financeiro eficiente quando alguém não pode obter lucros certos sem contar com alguma informação extra, ou seja, contando apenas com as informações disponíveis acerca da movimentação histórica do preço em questão. Em termos da taxonomia utilizada por Fama (1970), esta situação corresponde à forma (ou hipótese) fraca de eficiência de mercado.

2.2 Taxonomia

Uma definição geral das HEM é formulada, por Jensen (1978), como segue: Um mercado é dito eficiente em respeito ao conjunto de informações θ_t se é impossível ter lucro econômico maior do que zero baseando-se no conjunto de informações θ_t .

Grande parte dos testes empíricos quanto às HEM podem ser divididos em três categorias, conforme a taxonomia utilizada por Fama (1970), que fazem diferentes suposições para o conjunto de informação θ_t presente na definição acima.

Os testes que assumem que θ_t contém a série histórica do preço em questão até o momento t são considerados testes da forma *fraca* de eficiência. Ao assumir que θ_t conta com toda a informação pública disponível, fala-se em eficiência *semi-forte*. A forma *forte* de eficiência inclui ainda as informações privadas disponíveis a pelo menos um agente do mercado no momento t .

A última destas categorias (forma forte de eficiência) pode ser avaliada mais como uma extensão lógica das duas primeiras e conta com uma quantidade muito pequena de testes empíricos. A grande limitação prática encontrada é justamente ter acesso às informações privadas a fim de realizar os testes.

As duas primeiras (fraca e semi-forte) são, com certeza, as mais testadas empiricamente. Os testes da forma semi-forte de eficiência consistem em avaliar se o preço de um mercado se ajusta rápida e integralmente quando uma nova informação se torna disponível. Diferentes possibilidades empíricas dos testes de eficiência semi-forte dizem respeito ao ajustamento (velocidade e magnitude) dos preços de um ativo frente a determinados eventos, a saber: anúncios de pagamentos de dividendos ou de lucros realizados no período, de aumento do número de quotas disponíveis (*stock split*), intenção ou realização de fusões ou aquisições, dentre outros (SEILER; ROM, 1997). Mais referências sobre testes de evento podem ser encontradas em Beechey, Gruen e Vickery (2001).

Os testes realizados sob a forma fraca de eficiência consistem em avaliar a questão de previsibilidade do preço de um ativo a partir das informações contidas em sua série histórica. Mesmo que boa parte da literatura acerca deste tipo de teste de eficiência não consiga refutar as HEM (JENSEN, 1978), alguns trabalhos, como o de Lo e Mackinlay (1988), apresentam evidências de que alguns preços não se comportam de maneira totalmente aleatória e que parte de seus resultados futuros de longo prazo podem ser previstos por suas informações históricas. Devido às

características do método utilizado, este trabalho se preocupa apenas com a forma fraca de eficiência de mercado.

2.3 Eficiência e aleatoriedade de preços

Em um mercado eficiente, o preço dos ativos deve se comportar de maneira randômica através do tempo em resposta aos componentes não antecipados das novas informações (SAMUELSON, 1965). Novas informações, por definição, são não correlacionadas com informações passadas, pois se fossem, não seriam informações realmente novas e alguém poderia ter condições de prevê-las e assim obter lucros certos.

Se um mercado é suficientemente competitivo (eficiente), o instrumental da microeconomia clássica nos diz que os participantes não podem obter lucro (econômico) positivo por meio de alguma estratégia qualquer. Embora este fato apareça como um consenso nos dias de hoje, de acordo com Dimson e Maussavian (1998), não era o caso na maior parte do século passado. Estes autores apontam os trabalhos de Bachelier (1900), de Working (1934) e de Cowles e Jones (1937) como pioneiros no sentido de que todos compartilharam o resultado de que os preços de comodites flutuam de maneira randômica.

Os mencionados autores salientam que tal resultado não obteve seu devido lugar no estudo da Economia até a década de 50, com a publicação do clássico artigo de Samuelson (1965) *“Proof that properly anticipated prices fluctuate randomly”*, que, como o próprio nome diz, caracteriza um mercado como eficiente quando a plena incorporação das informações e expectativas dos participantes do mercado nos preços faz com que estes se comportem aleatoriamente.

A partir de então, e com o apoio de computadores para estudar o comportamento de séries de preços, alguns economistas acreditavam, como apontado por Kendall (1953), que se pudesse analisar uma série de preços extraindo sua tendência de longo prazo para obter informações de curto prazo a partir dos movimentos residuais. Porém, tal suposição foi contradita por este mesmo autor ao estudar séries de preços de ações e comodites no Reino Unido e concluir que as variações

aleatórias nos preços, quando analisados em intervalos curtos, são tão grandes que “escondem” qualquer efeito sistemático que pudesse estar presente nas séries.

Este fato, de que havia correlação serial bem próxima de zero nas séries de variações de preços financeiros, é a origem de uma metodologia de análise em Finanças conhecida como Modelo de Caminho Aleatório (ou Hipótese de Caminho Aleatório). A conclusão direta que se obtém por meio desta abordagem é que os preços de mercados financeiros são imprevisíveis. Desse modo, Fama (1970) diz que Roberts (1959) e Osborne (1959) encontraram evidência empírica de que séries de tempo geradas por números aleatórios eram indistinguíveis de séries históricas de preços financeiros. Adicionalmente, este autor apresentou outras evidências em favor das hipóteses de passeio aleatório, tema que se tornou um dos mais discutidos na literatura recente de Finanças.

Segundo Campbell, Lo e Mackinlay (1997), parte significativa dos trabalhos existentes na literatura compartilha desta idéia por meio dos modelos de jogo justo (*martingale*), nos quais, em termos financeiros, o ganho esperado por um investidor para determinado ativo poderia ser representado por $E(r_{t+1}|\theta_t)=0$ em que r_t é o retorno do ativo em questão e θ_t é o conjunto de informações (preços históricos) disponíveis no momento t . Isso decorre da idéia central que fundamenta os modelos de jogo justo que é a inutilidade do conhecimento dos preços passados para realizar previsões.

Definindo a diferença entre o valor esperado do retorno de um ativo j para o período $t+1$ e o retorno realmente ocorrido como Z_t , na seguinte forma $Z_{j,t+1} = r_{j,t+1} - E(r_{j,t+1}|\theta_t)$ podemos caracterizar a hipótese de jogo justo em relação ao conjunto de informação θ_t por meio da condição $E(z_{j,t+1}|\theta_t)=0$. Podemos considerar Z_t como o excesso do valor de mercado, ou seja, o montante pelo qual o valor ocorrido é diferente do esperado que ocorresse.

Seja então $V_{t+1}(\theta_t)$ um conjunto de excessos de valor de mercado gerado por qualquer sistema (estratégia) de negociação que compõe determinado grupo de

ativos, baseado em θ_t . Portanto, valendo as hipóteses de jogo justo, espera-se que $E(V_{t+1}|\theta_t) = 0$ possa ser derivado como uma implicação testável para modelos de eficiência de mercado. Conforme enfatizam Costa Jr. e Lemos (1997), as hipóteses de eficiência de mercado demandam a inexistência de quaisquer tipos de padrões de ocorrência (sistemas de negociação) no comportamento dos mercados que resultem em retornos fora dos considerados normais.

Esta idéia de jogo justo imprime aos modelos que a adotam, ou pelo menos que adotam alguma de suas variações, a importante característica de imprevisibilidade (e, por analogia, aleatoriedade) dos preços futuros. A literatura apresenta diversas variações destes modelos de passeio aleatório no comportamento de preços (MANDELBROT, 1966).

2.4 Alternativa às HEM: Eficiência relativa

Campbell, Lo e MacKinlay (1997) afirmam que, mesmo que a metodologia empírica de testes sobre as hipóteses de eficiência de mercado esteja bem estabelecida, existem algumas dificuldades sérias em interpretar seus resultados. Primeiramente, qualquer teste de eficiência depende crucialmente do modelo de equilíbrio estimado que defina o retorno normal do mercado. Nestes casos, quando a hipótese de eficiência de um mercado é rejeitada, podemos interpretar tal resultado como real ineficiência ou ainda como erro de especificação do modelo de equilíbrio sobre o qual o teste de eficiência é realizado, resultando em um problema de hipótese conjunta.

Nesse sentido, Lo e Mackinley (1999, p. 6) ressaltam que as HEM não são hipóteses bem estabelecidas e empiricamente refutáveis. Um teste de eficiência se torna um teste de diversas hipóteses auxiliares e a rejeição de tal teste diz pouco sobre qual aspecto da hipótese conjunta é inconsistente com os dados.

Em segundo lugar, temos o fato de que eficiência perfeita é uma abstração que não ocorre na prática. Mesmo no campo teórico existe espaço para retornos fora do normal, como demonstrado por Grossman e Stiglitz (1980), devido aos custos de

aquisição e utilização de informações. Para estes autores, mesmo em um mercado suficientemente grande e líquido, os custos de transação são justificativa para pequenos retornos fora do normal, ainda que fosse difícil dizer o quão pequeno mesmo que os custos de transação pudessem ser mensuráveis.

Visto que aleatoriedade pode ser entendida como uma questão de grau, e não simplesmente 'sim ou não', a noção de eficiência relativa de mercados (como *Bovespa* em relação à *Merval*, mercado de câmbio em relação a outro determinado ativo) é considerada por Campbell, Lo e MacKinlay (1997) um conceito de maior utilidade do que a noção absoluta de eficiência compartilhada por grande parte da literatura acerca das hipóteses de eficiência de mercado.

Em uma analogia a sistemas físicos, os autores citados destacam que motores a combustão, por exemplo, são comumente avaliados quanto à sua eficiência baseados na proporção relativa de energia que é convertida em trabalho. Dizer que um motor é 30% eficiente significa que, na média, 30% da energia inserida é convertida em energia mecânica, sendo que os outros 40% são transformados em outras formas como calor, luz e som. Sendo assim, não faz sentido um teste de hipótese para verificar se determinado motor é ou não 100% eficiente, o que corresponderia apenas a uma abstração teórica.

De maneira análoga, as HEM também podem ser entendidas como um caso limite, inatingível em mercados reais, mas úteis no sentido de avaliar a eficiência em termos relativos. Devido aos custos e benefícios decorrentes de buscar informações, Grossman e Stiglitz (1980) argumentam que mercados perfeitamente eficientes são uma impossibilidade. Esta afirmativa decorre do fato de que, em um mercado perfeitamente eficiente, o lucro econômico obtido por coletar e utilizar informações deve ser igual a zero e assim não haveria muitas razões para negociação e o mercado deixaria de existir.

Em outras palavras, se os agentes realmente acreditassem que buscar informações não é uma atividade lucrativa, já que o mercado se comporta de maneira perfeitamente aleatória, não existiriam razões para participar do mercado diferentes das que fazem alguém jogar na loteria ou ir a um cassino.

Como alternativa à visão sim ou não compartilhada por parte da literatura acerca das HEM, estes autores assinalam que o conceito de grau de ineficiência (ou eficiência relativa), uma vez que todos os mercados reais seriam ineficientes, poderia determinar o quanto os agentes estão dispostos a gastar para obter informações que possibilitem a obtenção de retornos positivos (explorar as ineficiências). Quanto maior o grau de ineficiência de determinado mercado, mais vale a pena buscar e utilizar informações.

3 Método

Este capítulo trata do método utilizado para estabelecer comparações em termos relativos sobre a eficiência de mercados financeiros. Como discutido no capítulo anterior, a eficiência de um mercado é associada à aleatoriedade de seus preços. A medida de complexidade de Lempel e Ziv (*LZ*) pode ser entendida como uma medida do grau de aleatoriedade de uma seqüência de caracteres e assim ser utilizada no estudo da eficiência de mercados em termos relativos (CAMPBELL; LO; MACKINLAY, 1997).

Este capítulo é dividido como segue: a primeira seção trata da discussão sobre o conceito de aleatoriedade e possíveis métodos que permitam sua verificação. A segunda descreve a formulação matemática da medida de complexidade de Lempel e Ziv (1976), utilizada neste trabalho, assim como sua construção operacional. A terceira expõe as normalizações utilizadas e a quarta trata das especificidades da utilização deste método no estudo das Finanças. A quinta seção ilumina a medida utilizada ao testá-la contra séries notoriamente conhecidas como estáveis, periódicas e aleatórias. A sexta e última seção mostra as vantagens da medida de complexidade de Lempel e Ziv em relação a outro método de verificação de aleatoriedade chamado Entropia Aproximada (*ApEn* - *Approximate Entropy*).

3.1 Contexto: aleatoriedade e complexidade

O conceito de aleatoriedade é tratado por diferentes campos teóricos (CHAITIN, 1987). Sob os fundamentos da *Teoria das Probabilidades*, diversos testes de aleatoriedade foram propostos. Wolfram (2002) coloca que estes são baseados na formulação de testes de hipótese sobre determinadas características estatísticas (distribuição de probabilidades) de seqüências numéricas, a fim de se estabelecer se dada seqüência pode ou não ser considerada aleatória. Haahr (1999) descreve diversos destes testes e alguns selecionados seguem abaixo:

1. *Teste de freqüência de valores*: O propósito deste tipo de teste é determinar se a freqüência de cada valor na seqüência analisada é aproximadamente

igual à esperada em uma seqüência verdadeiramente aleatória ($f(\textit{cara}) = 0,5$ no caso de uma moeda justa). É fundamentado na suposição de que todos os elementos possíveis de uma seqüência deveriam aparecer com a mesma freqüência e conta quantas vezes determinado número aparece e compara este valor com a ocorrência dos outros números.

2. *Teste de freqüência em blocos*: Similar à categoria anterior, mas não testa a freqüência de determinados valores em toda a seqüência, mas sim em subseqüências de determinado tamanho M (que é um argumento do teste). Sob a mesma conceituação de aleatoriedade, espera-se que a freqüência de ocorrência de determinado valor seja a mesma para diferentes valores de M .
3. *Teste serial*: Um par de elementos deve ser igualmente provável de ocorrer em qualquer ordem. O teste verifica se dois números consecutivos não têm correlação, comparando sua ocorrência conjunta com a dos outros pares possíveis. O propósito desta categoria de testes é determinar se existe correlação serial, de alguma ordem, na seqüência analisada.
4. *Teste de intervalo*: Conta quantos números existem na seqüência analisada entre duas ocorrências de números que estão dentro de um intervalo previamente definido.

A *Teoria da Informação de Shannon* (THEIL, 1967), por sua vez, caracteriza como perfeitamente aleatório o caso extremo em que o conteúdo informacional esperado (entropia de Shannon) de uma mensagem é maximizado, ou seja, o grau de incerteza é máximo e não existe nenhuma redundância na mensagem.

Suponha que, em um momento inicial, alguém receba uma mensagem de que determinado evento E pode ocorrer com probabilidade x , com $0 \leq x \leq 1$. Suponha também que, em um momento posterior, alguém receba a informação de que o evento E realmente ocorreu. Se $x = 0,99$, não haveria muita surpresa, visto que era quase certo de que E ocorreria. Em outras palavras, a mensagem que estabeleceu que o evento E ocorreu tinha pouco conteúdo informacional (no lado oposto, caso $x = 0,01$, a mensagem teria um alto conteúdo informacional, pois era esperado que E não ocorresse).

Espera-se, então, que o *conteúdo informacional* $h(x)$ de uma mensagem seja uma função decrescente da função probabilidade x , e é aceito que assuma o logaritmo do inverso da probabilidade de x , na forma:

$$h(x) = \log \frac{1}{x} = -\log x$$

Considerando agora uma seqüência de n possíveis eventos E_1, \dots, E_n de maneira que eles formem um sistema completo no sentido de que é certo de que pelo menos um deles vai ocorrer ($\sum_{i=1}^n x_i = 1$, para $i = 1, \dots, n$). Não sabemos, *a priori*, qual deles vai ocorrer, mas podemos calcular a média ou o conteúdo informacional esperado (*entropia de Shannon*) $H(x)$ da seguinte maneira:

$$H(x) = \sum_{i=1}^n x_i h(x_i) = \sum_{i=1}^n x_i \log \frac{1}{x_i} = -\sum_{i=1}^n x_i \log x_i$$

A entropia de Shannon é minimizada quando apenas um dos eventos agrupa toda a probabilidade do conjunto (por exemplo, $x_j = 1$ e $x_i = 0$, $\forall i \neq j$) e maximizada quando os eventos são equiprováveis ($x_i = \frac{1}{n}$, $\forall i$).

A *Teoria da Informação Algorítmica* (RUDICH; WIGDERSON, 2000) é uma disciplina que se relaciona tanto com a teoria da informação convencional como com os fundamentos da Ciência da Computação. A relação desta disciplina com o conceito de aleatoriedade é baseada na definição de complexidade de um objeto proposta por Kolmogorov que é, de maneira informal, dada pelo tamanho do menor programa de computador que é capaz de reproduzir tal objeto. Nas palavras de Gregory Chaitin, a Teoria da Informação Algorítmica (CALUDE *et al*, 2008):

“Is the result of putting Shannon's information theory and Turing's computability theory into a cocktail shaker and shaking vigorously. The basic idea is to measure the complexity of an object by the size in bits of the smallest program for computing it.”

A definição de complexidade apresentada por Kolmogorov formaliza os conceitos intuitivos de complexidade e simplicidade. Se determinado objeto, por conter

diversas redundâncias ou padrões, for passível de ser reproduzido por um simples (pequeno) programa de computador, podemos dizer que este objeto apresenta baixa complexidade. Pelo outro lado e no caso extremo, um objeto é dito perfeitamente aleatório quando o tamanho do menor programa de computador capaz de reproduzi-lo é no mínimo igual ao tamanho do próprio objeto, já que não existem redundâncias e padrões que possibilitem sua compressão.

Para objetos como seqüências binárias de caracteres (por exemplo, uma seqüência constituída apenas pelos caracteres '0' e '1'), a idéia de complexidade proposta por Kolmogorov se relaciona com a menor descrição possível de determinada seqüência. Considerando as duas seqüências binárias a seguir:

A 01010101010101010101
B 11001000011000011101

Podemos descrever a seqüência *A* simplesmente como “10 repetições de '01'”. Entretanto, a seqüência *B* não tem uma descrição óbvia mais curta do que sua própria descrição caractere por caractere. Então, podemos dizer que a menor descrição de *A* é menor do que a menor descrição de *B* e, por decorrência, que *B* é mais complexo do que *A*, o que está de acordo com as noções intuitivas de complexidade e simplicidade.

Pode-se perceber que esta métrica de complexidade é sensível à linguagem (ou máquina) escolhida, visto que, no exemplo acima, a descrição da seqüência *A* foi feita em português. Dessa maneira, podemos chamar de $K_F(x)$ a complexidade da seqüência x relativa ao modelo computacional (máquina, linguagem, etc.) F . Se o programa P , rodado no modelo computacional F , reproduz a seqüência x , podemos dizer em notação que $F(P) = x$. Consideremos ainda que $\langle P \rangle$ represente o tamanho do programa P . Então, a complexidade K da seqüência x , relativa ao modelo computacional F , pode ser escrita como:

$$K_F(x) = \min_P \{ \langle P \rangle : F(P) = x \}$$

Em palavras, a complexidade de x em respeito ao modelo F é dada pelo tamanho do menor programa P que reproduz x rodando em F . Para dar um tratamento mais formal à definição de complexidade proposta por Kolmogorov e expor duas de suas propriedades (invariância e incomputabilidade), cabem algumas definições prévias (MOHR, 2008), realçadas pelo sublinhado.

Algoritmos e problemas: Um *algoritmo* pode ser definido como um processo computacional que, em um finito espaço de tempo, encontra uma solução para determinado *problema*.

Os problemas que podem ser resolvidos por algum algoritmo são estabelecidos como *problemas de decisão* (tipo sim-ou-não: 1363 é primo? R: Não), *funções* (no sentido de um resultado único: qual o menor primo maior do que 1363? R: 1367) e *relações* (admitindo mais de um resultado: quais os fatores primos de 1363? R: 29 e 47).

Tanto os problemas de decisão quanto as relações podem ser representados por funções. Um problema de decisão pode ser formulado por uma função que resulte em 0 se a resposta é “Sim” e em 1 se a resposta é “Não” e uma relação por outra função que resulte em algum valor pré-determinado do conjunto correto de resultados. Podemos então generalizar estas três instâncias ao tratar da questão da *computabilidade* de uma função.

Uma função é dita computável se existe algum algoritmo efetivo que represente tal função e que este produza o resultado correto em um finito espaço de tempo e páre, dado o argumento ou conjunto de argumentos para os quais a função é definida.

Máquina de Turing: Considere o seguinte problema de decisão¹: “Existe algum processo definitivo pelo qual todas as questões matemáticas podem ser resolvidas?”. Para dar uma resposta negativa a este problema, pode-se considerar uma máquina de computação simples e abstrata chamada de *Máquina de Turing* (*MT*).

¹ Um dos problemas de Hilbert. Disponível em: <http://en.wikipedia.org/wiki/Hilbert's_problems>. Acesso: em 30 jan. 2008.

Uma *MT* (Figura 1) é composta por uma unidade de controle, um cabeçote leitor/escritor e uma fita na qual o cabeçote é posicionado (infinita e dividida em quadrados que podem ter um caractere de determinado alfabeto). As possíveis operações de uma *MT* são: a) ler o caractere atual no cabeçote; b) (re)escrever um caractere em um quadrado da fita; c) mover o cabeçote um quadrado em qualquer direção, e d) verificar se o quadrado atual do cabeçote é o inicial da fita.

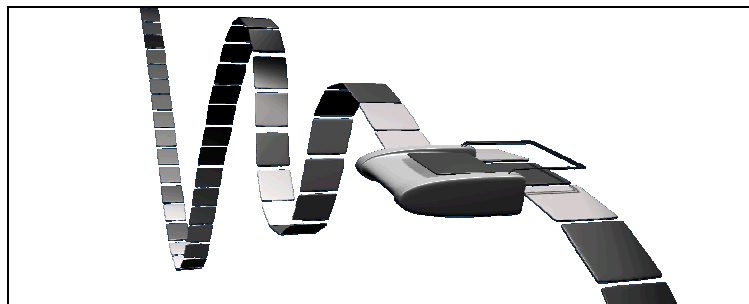


Figura 1: Representação artística de uma *MT*. Disponível em: <http://pt.wikipedia.org/wiki/Imagem:Turing_Machine.png>. Acesso em: 30 jan. 2008.

Mesmo que as *MT* não sejam usadas em computadores atuais, elas podem servir como modelos capazes de simular qualquer outro processo computacional físico possível. Considerando mais uma generalização, uma Máquina Universal de Turing *U* pode simular a computação de qualquer outra *MT* se receber a descrição desta *MT* como um número de série único e suas configurações iniciais. Então, a resposta negativa ao problema da decisão acima decorre da existência de problemas que não são computáveis por *U*. Se um destes problemas é formulado como uma função, ele é dito incomputável, e se é formulado como um problema de decisão, é dito indecidível.

A prova de que existem funções incomputáveis decorre do fato de que o conjunto de todas as *MT* possíveis é contável (isto é, tem correspondência única com os números inteiros – vide seu número de série). Entretanto, o conjunto de todas as funções que tem os inteiros como domínio e contradomínio ($f: \mathbb{Z} \rightarrow \mathbb{Z}$) é incontável, o que implica afirmar que as *MT* são capazes de computar apenas um subconjunto próprio do conjunto de todas as funções possíveis.

Incomputabilidade: Para mostrar que existem funções incomputáveis, pode-se considerar problemas que fazem referência a si mesmos, como a frase “Esta sentença é falsa” ou como o Paradoxo de Russel (seja o conjunto R que tem todos os conjuntos que não têm a si mesmos como membros – a prova de que um conjunto A não pertence a R se baseia na suposição de que A pertence a R – o que resulta em uma contradição).

A complexidade de Kolmogorov é uma função incomputável, no sentido de que não existe nenhum programa que receba uma seqüência x e produza a complexidade $K(x)$ desta seqüência. Isso significa que, na prática, a complexidade de um objeto pode apenas ser aproximada. De maneira informal, a incomputabilidade da complexidade de Kolmogorov pode ser associada a um problema de referência a si mesma, pois a complexidade (do objeto) é dada pelo tamanho do menor programa que o descreva (referência a si mesmo).

Invariância: Contudo, sabe-se que a complexidade pode ser aproximada de forma independente da escolha da máquina ou linguagem. Para demonstrar isso, alguém pode considerar os impactos da escolha da linguagem de descrição na medida de complexidade. Sendo $K_{MT}(x)$ a complexidade de x em respeito a esta determinada MT e $K_U(x)$ a complexidade de x em respeito a U , existe uma constante $C_{MT} = \langle MT \rangle$ para cada MT possível tal que

$$K_U(x) \leq K_{MT}(x) + C_{MT} \text{ (invariância) ou}$$

$$\langle K_U(x) - K_{MT}(x) \rangle \leq C_{MT} \text{ (equivalência)}$$

para $\forall x$ e $\forall MT$

Note que também aqui o símbolo $\langle P \rangle$ representa o tamanho de P . Por conseguinte, a complexidade de uma seqüência x em respeito a U (que pode simular qualquer outra MT) é no máximo igual à complexidade de x em respeito a esta certa MT acrescida de uma constante C_{MT} (que só depende de U e de MT , mas não de x) que representa o ‘custo’ em esforço computacional envolvido em simular a MT por meio de U .

3.2 Medida de complexidade de Lempel e Ziv (LZ)

3.2.1 Formulação matemática

Esta subseção é destinada a apresentar a formulação matemática da medida de complexidade proposta por Lempel e Ziv (1976), no artigo “*On the complexity of finite sequences*”. A formulação operacional da medida para fins de experimentos práticos é tratada na próxima subseção.

Considerando a incomputabilidade da definição de complexidade proposta por Kolmogorov tratada na seção anterior, Lempel e Ziv (1976) não se propuseram a desenvolver uma medida absoluta de complexidade, que eles próprios julgam ser inexistente, mas sim estudar a complexidade de uma seqüência finita sob o ponto de vista de uma máquina razoavelmente simples.

Esta máquina proposta reconstrói a seqüência contando apenas com duas operações básicas: copiar e inserir. Para tal, lê a seqüência da esquerda para a direita e adiciona, em um procedimento incremental, as novas subseqüências encontradas (que até então não tinham aparecido no processo de reconstrução) a um dicionário. De modo geral, o tamanho de tal dicionário é sugerido pelos autores como uma útil medida da complexidade da seqüência.

Nesse sentido, a medida proposta por Lempel e Ziv não resulta no tamanho $\langle P \rangle$ do menor programa de computador P capaz de reconstruir determinada seqüência S , mas sim em um número $c(S)$ que é uma medida apropriada da complexidade de Kolmogorov (KASPAR; SCHUSTER, 1987).

O restante desta seção é dividido da seguinte maneira: começa com notações prévias definidas por Lempel e Ziv (1976) para expor a medida proposta, seguidas pela exposição das operações de cópia e inserção e por detalhamentos do processo de reconstrução da seqüência. Por fim, é apresentado o modo pelo qual a complexidade da seqüência é calculada.

Notações prévias: A^* é o conjunto que contém todas as possíveis seqüências $S \in A^*$ de tamanho finito $l(S)$ formadas pelo alfabeto finito de caracteres A . A^n é o conjunto de todas as seqüências de A^* que tem tamanho n . O conjunto Λ que contém a seqüência nula ($l(S)=0$) pertence a A^* .

Uma seqüência $S \in A^n$ é especificada por $S = s_1s_2 \dots s_n$. Uma *subseqüência* de S é denotada por $S(i, j)$ se começa em i e termina em j para $i \leq j$ e por Λ para $i > j$. A *concatenação* de $Q \in A^m$ e $R \in A^n$ forma uma nova seqüência $S = QR = q_1q_2 \dots q_m r_1r_2 \dots r_n \in A^{m+n}$. S^2 representa a concatenação de S consigo mesma, $S^0 = \Lambda$ e $S^i = S^{i-1}S$ de modo geral para $i \geq 1$.

Um *prefixo* de $S \in A^*$ é definido como uma seqüência $Q = S(1, i)$, e nesse caso S é considerado uma *extensão* de Q . Um prefixo e uma extensão são definidos como *próprios* se $l(S) > l(Q)$, ou seja, se não são a mesma seqüência. Mesmo que o tamanho de S seja desconhecido, pode-se fazer referência aos seus prefixos por meio do operador π^i tal que $S\pi^i = S(1, l(S)-i)$, com $i = 0, 1, \dots$.

O *vocabulário* $v(S)$ de S é o subconjunto de A^* formado por todas as subseqüências (ou palavras) presentes em S . Exemplo: $v(010) = \{\Lambda, 0, 1, 01, 10, 010\}$. Uma palavra $Q \in v(S)$ é chamada de *auto-palavra* de S se não pertence ao vocabulário de nenhum prefixo próprio de S . O *auto-conjunto* $e(S)$ de S contém todas as auto-palavras de S . Exemplo: $e(010) = \{10, 010\}$.

Operações de cópia e de inserção: Se uma seqüência S é estendida pela concatenação com uma de suas palavras $W = S(i, j)$, a seqüência resultante $R = SW$ é obtida simplesmente por um mecanismo de cópia onde o caractere $r_{m+l(S)} = w_m$ é copiado de S_{i+m-1} , para $m = 1, 2, \dots, j-i+1$.

Uma extensão $R = SQ$ de S é *reproduzível* a partir de S se $Q \in v(R\pi) = v(SQ\pi)$, e é denotada por $S \rightarrow R$. Extensões maiores do que qualquer palavra contida no vocabulário $v(S)$ podem ser geradas por este método de reprodução se existir um

índice inicial p de S (chamado de *ponteiro* de $S \rightarrow R$) tal que $Q = R(p, l(Q) + p - 1)$.

Exemplo: $010 \rightarrow 01010101$ e $010 \xrightarrow{\sim} 01011111$ com $p = 2$, em que $\xrightarrow{\sim}$ significa a negação de \rightarrow .

Uma seqüência S é considerada *produzível* a partir de um prefixo próprio $S(1, j)$ se $S(1, j) \rightarrow S\pi$ e denotada por $S(1, j) \Rightarrow S$. O prefixo $S(1, j)$ é então chamado de *base* de S . Exemplos: $01 \Rightarrow 0100$, mas $01 \xrightarrow{\sim} 0100$ com $p = 1$. A distinção entre os conceitos de reproduzível e produzível é central na diferenciação entre as duas operações (copiar e inserir) da máquina proposta: Uma extensão é dita reproduzível a partir de dada seqüência se pode ser gerada apenas pelo procedimento de cópia proposto, enquanto o conceito de extensão produzível permite a inserção de caracteres.

Procedimento de produção: Qualquer seqüência S não nula e finita pode ser produzida a partir de algum de seus prefixos próprios Q ($Q \Rightarrow S$), mediante um *procedimento de produção* que, tem como primeiros passos $\Lambda = S(1, 0) \Rightarrow S(1, 1) = s_1 \Rightarrow S(1, 2) \cdots S(1, h_i) \Rightarrow S(1, h_{i+1})$ por meio da inserção de novos caracteres até que todos os $l(S)$ passos tenham sido dados e que a seqüência S tenha sido reconstruída.

Considera-se então o procedimento de produção de S com m passos e $S(1, h_i)$ sendo o i -ésimo estado do processo, com $i = 1, 2, \dots, m$. O resultado da repartição de S em m palavras (ou *componentes*) $H_i(S) = S(h_{i-1} + 1, h_i)$ é chamado de *história* $H(S)$ de S . Um componente $H_i(S)$ e seu respectivo passo no procedimento de produção $S(1, h_{i-1}) \Rightarrow S(1, h_i)$ são chamados *exaustivos* caso $S(1, h_{i-1}) \xrightarrow{\sim} S(1, h_i)$, ou seja, o procedimento de produção não pode prosseguir com a operação de cópia e um novo caractere deve ser inserido. A *história exaustiva* $E(S)$ de S é composta por todos os seus componentes exaustivos.

Todas as seqüências finitas não nulas têm uma história exaustiva única. Exemplo: $S = 0110001$, pode ser dividida em seus componentes exaustivos na forma $E(S) = 0 \bullet 1 \bullet 10 \bullet 001$.

Medida de complexidade proposta: Seja $c_H(S)$ o número de componentes presentes em uma história $H(S)$ de S . A complexidade $c(S)$ de S é dada por

$$c(S) = \min\{c_H(S)\}$$

Isto é, a complexidade de uma seqüência é dada pelo menor número possível de componentes pelo qual tal seqüência pode ser dividida. Um dos teoremas apresentados por Lempel e Ziv (1976), que é mostrado aqui sem prova, estabelece que

$$c(S) = c_E(S)$$

Em palavras, a complexidade de uma seqüência é dada pelo número de componentes de sua história exaustiva $c_E(S)$, ou seja, a história que tem o menor número de componentes é a história exaustiva. Exemplo: $c(0110001) = 4$, visto que $E(0110001) = 0 \bullet 1 \bullet 10 \bullet 001$ ($c_E(0110001) = 4$).

3.2.2 Construção operacional

Esta subseção exhibe a construção operacional desenvolvida por Kaspar e Schuster (1987) para a medida de complexidade proposta por Lempel e Ziv (1976), e é amplamente baseada no artigo "*Easily calculable measure for the complexity of spatiotemporal patterns*", no qual estes autores apresentam a construção da medida de complexidade, assim como também analisam algumas de suas propriedades.

Cálculo da medida de complexidade: Kaspar e Schuster (1987) desenvolveram um algoritmo para calcular a complexidade c de uma seqüência de caracteres finita, não nula e totalmente especificada S , de acordo com as formulações matemáticas propostas por Lempel e Ziv (1976).

Considere determinada seqüência $s_1 s_2 \dots s_r \dots s_n$ reconstruída pelo algoritmo proposto até o dígito recém-inserido s_r (ou seja, este é um dígito novo que não pode ser copiado, pois não constava no vocabulário $v(s_1 s_2 \dots s_{r-1})$). A questão imprimida por

Lempel e Ziv (1976) é se o restante da seqüência $(s_{r+1}s_{r+2}\cdots s_n)$ pode ser construído apenas por meio do procedimento de cópia de subseqüências previamente encontradas em $v(s_1s_2\cdots s_r)$ ou se será necessária a inserção de novos dígitos.

Para tal, verificam se a seqüência $Q = s_{r+1}$ está contida no vocabulário composto pelas subseqüências previamente encontradas $v(s_1s_2\cdots s_r)$. Em caso afirmativo, o próximo dígito é adicionado a Q tal que $Q = s_{r+1}s_{r+2}$ e verificam novamente se $Q \in v(s_1s_2\cdots s_r) = v(SQ\pi)$ ². Este procedimento pode ser estendido até o momento em que uma nova palavra é encontrada ou que Q se torne tão grande ($l(Q) > l(S)$) que não possa mais ser obtido a partir de $v(SQ\pi)$.

Na primeira das possibilidades, o algoritmo insere a nova palavra encontrada Q em S e retoma o mesmo procedimento, mas agora de um novo ponto de partida. Na última, o algoritmo insere a palavra Q e pára a computação. O número de palavras inseridas é utilizado como a medida c de complexidade da seqüência. A fim de ilustrar tal procedimento, seguem alguns exemplos:

1) $S = 00000$

No processo de reconstrução, sempre o primeiro dígito ($s_1 = 0$) deve ser inserido, então $S = 0$. Para ver se o restante de S pode ou não ser reconstruído apenas por meio da operação de cópia, devemos considerar $Q = s_2 = 0$, a concatenação $SQ = 00$, e o vocabulário previamente encontrado $v(SQ\pi) = \{0\}$ e verificar se $Q \in v(SQ\pi)$. Visto que a resposta é afirmativa, tomamos $Q = s_2s_3 = 00$, $SQ = 000$, $v(SQ\pi) = \{0, 00\}$ e verificamos novamente se $Q \in v(SQ\pi)$.

Neste exemplo de simplicidade extrema, seguiríamos este processo até a situação em que o tamanho da concatenação SQ seja igual ao da seqüência original (5) e o

² O operador especial π foi definido na seção anterior, na forma $S\pi^i = S(1, l(S) - i)$. Significa que os i últimos dígitos de uma seqüência devem ser retirados. Neste exemplo, $SQ\pi = S$.

programa insere Q em S para ter a seqüência totalmente reconstruída. Os dígitos '0' e '0000' foram inseridos, então a complexidade $c(S)$ de S é igual a 2.

2) $S = 01010$

O primeiro dígito deve ser inserido: $S = 0$. Sendo $Q = 1$, $SQ = 01$ e $v(SQ\pi) = \{0\}$, conclui-se que $Q \notin v(SQ\pi)$, então Q deve ser inserido em S . No próximo passo, percebe-se que $Q \in v(SQ\pi)$, pois $S = 01$, $Q = 0$, $SQ = 010$ e $v(SQ\pi) = \{0,1,01\}$. Dando continuidade ao processo, temos $Q = 01$, $SQ = 0101$ e $v(SQ\pi) = \{0,1,01,10,010\}$, então novamente $Q \in v(SQ\pi)$. Como o resultado da concatenação de $S = 01$ e $Q = 010$ é igual à seqüência original, insere-se Q em S e o procedimento termina com $c(S) = 3$, visto que as palavras inseridas foram $\{0,1,010\}$. Note que, nesse caso, mesmo que a última palavra da seqüência já estava no vocabulário pré-encontrado, ela também é inserida.

3) $S = 0110001$ (apresentado em forma de quadro)

S	Q	SQ	$SQ\pi$	$v(SQ\pi)$	$Q \in v(SQ\pi)$
0	1	01	0	0	Não
01	1	011	01	0,1	Sim
01	10	0110	011	0,1,01,11,011	Não
0110	0	01100	0110	0,1,01,11,10,011,110,0110	Sim
0110	00	011000	01100	0,1,01,11,10,00,011,110,100,0110,110,01100	Sim
0110	001	0110001	SQ é igual à seqüência original. Insere Q em S e pára.		

Quadro 1: Construção operacional da medida de complexidade da seqüência $S = 0110001$. Ilustra o procedimento utilizado para medir a complexidade de uma seqüência como descrito em Kaspar e Schuster (1987).

Na última linha, a concatenação sQ é igual à seqüência original. O algoritmo insere Q em s e pára. As palavras inseridas foram $\{0,1,10,001\}$, resultando em $c(S) = 4$.

O fluxograma do algoritmo descrito em Kaspar e Schuster (1987) é mostrado no anexo A, juntamente com dois códigos escritos na forma de funções da linguagem *Python* e *VBA* apresentados no anexo B.

3.3 Normalizações

3.3.1 Normalização por uma medida de complexidade máxima

Como demonstrado na última seção, a complexidade de uma seqüência aumenta com seu tamanho. Esta não é uma regra absoluta, uma vez que as seqüências '00000' e '00000000000000000000' têm a mesma complexidade igual a 2. Entretanto, pode-se esperar que, em casos não tão simples, o tamanho do dicionário, e por decorrência a complexidade, aumentem com o tamanho da seqüência.

Lempel e Ziv (1976) demonstraram que, para uma seqüência reconhecida como muito complexa (seqüências *de Brujin*, descritas no final da seção), o comportamento assintótico da complexidade c , quando o tamanho n da seqüência tende ao infinito, segue a relação abaixo:

$$\lim_{n \rightarrow \infty} c = b \equiv n / \log_a n$$

em que a é o tamanho do dicionário (número de caracteres) utilizado. No caso binário no qual os caracteres do alfabeto podem ser '0' ou '1', $a = 2$. Então, neste trabalho sempre utilizaremos a normalização

$$0 \leq c/b \leq 1$$

no lugar de apenas c . De outra forma, pode-se perceber que, apenas para seqüências aleatórias de tamanho n suficientemente grandes (infinito), a complexidade normalizada atinge seu valor assintótico

$$\lim_{n \rightarrow \infty} c/b = 1$$

Kaspar e Schuster (1987) analisaram o quão rápido a complexidade c de uma seqüência binária construída por um gerador de números pseudo-aleatórios converge para seu valor assintótico $n/\log_2 n$. O gerador produz números dentro do intervalo $x_i \in [0,1]$ e a seqüência é construída na forma $s_i = 0$ se $x_i \leq 0,5$ e $s_i = 1$ caso contrário. O Gráfico 1 abaixo mostra a complexidade normalizada c/b de uma seqüência aleatória gerada por computador em função do seu tamanho n .

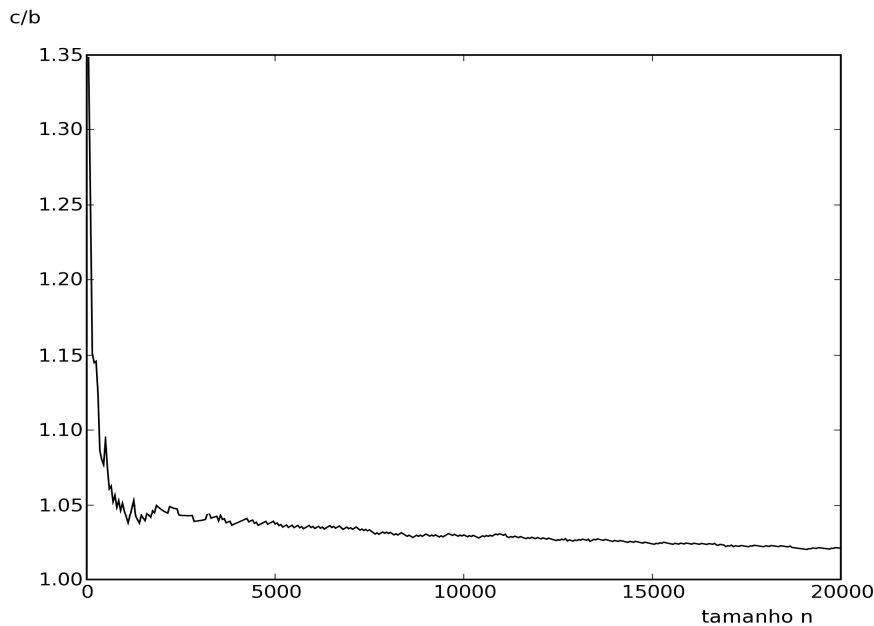


Gráfico 1: A complexidade normalizada c/b de uma seqüência aleatória se aproxima por cima de seu limite superior assintótico $\lim_{n \rightarrow \infty} c/b = 1$

Note que b descreve o comportamento assintótico de c para uma seqüência aleatória quando seu tamanho n tende ao infinito. Para fins práticos (finitos) este índice descreve um limite superior, atingido por cima, para a complexidade de uma seqüência aleatória.

Kaspar e Schuster (1987) afirmam que o algoritmo proposto alcança seu valor assintótico para seqüências de tamanho maior ou igual a 1000 com uma margem de 5%, conforme mostra o Gráfico 1 (complexidade normalizada igual a 1.05). Isso

significa que apenas quando o tamanho da seqüência é maior que 1000 pode-se falar em complexidade (normalizada) por dígito e estabelecer comparações entre seqüências de tamanhos distintos.

Seqüências de Bruijn (KJELLERSTRAND, 2008): Uma seqüência de Bruijn de tamanho a^n é a menor seqüência fechada (ou circular, no sentido de que os últimos caracteres podem ser concatenados com os primeiros) que contém todas as palavras possíveis de tamanho n construídas com os caracteres disponíveis no alfabeto A de tamanho a .

Visto que a medida de complexidade utilizada representa o menor número de palavras distintas em que se pode dividir uma seqüência, Lempel e Ziv (1976) argumentam que as seqüências de Bruijn são boas aproximações finitas de seqüências aleatórias.

Como ilustração, considere o caso binário (caracteres '0' ou '1', alfabeto de tamanho $a = 2$) e palavras de tamanho $n = 3$, temos as seguintes palavras possíveis:

000	011
001	111
010	110
101	100

A menor seqüência circular que contém todas as palavras acima é uma seqüência (circular) de Bruijn de tamanho $a^n = 2^3 = 8$, dada por 00010111, como se pode verificar.

3.3.2 Normalização pela entropia

Se em determinada seqüência binária aleatória, a probabilidade p de encontrar o caractere '1' for tal que $p \neq 0,5$, podemos esperar que a complexidade desta seja menor do que de outra na qual $p = 0,5$ (KASPAR; SCHUSTER, 1987). Por exemplo, considerando duas seqüências de mesmo tamanho, espera-se que a complexidade

de uma 90% composta pelo caractere '1' seja menor do que a da outra dividida meio a meio.

Lempel e Ziv (1976) mostram que, nestes casos em que $p \neq 0,5$, a complexidade normalizada $c/b = \frac{c}{n/\log_2 n}$ tem seu valor assintótico desviado de 1 para h

$$\lim_{n \rightarrow \infty} c/b = h$$

sendo que h é a entropia da fonte geradora da seqüência, dada por

$$h(p) = -[p \log_2(p) + (1-p) \log_2(1-p)] \leq 1, \text{ com } 0 \leq p \leq 1$$

O Gráfico 2 mostra que a entropia é máxima ($h=1$) quando $p=0,5$.

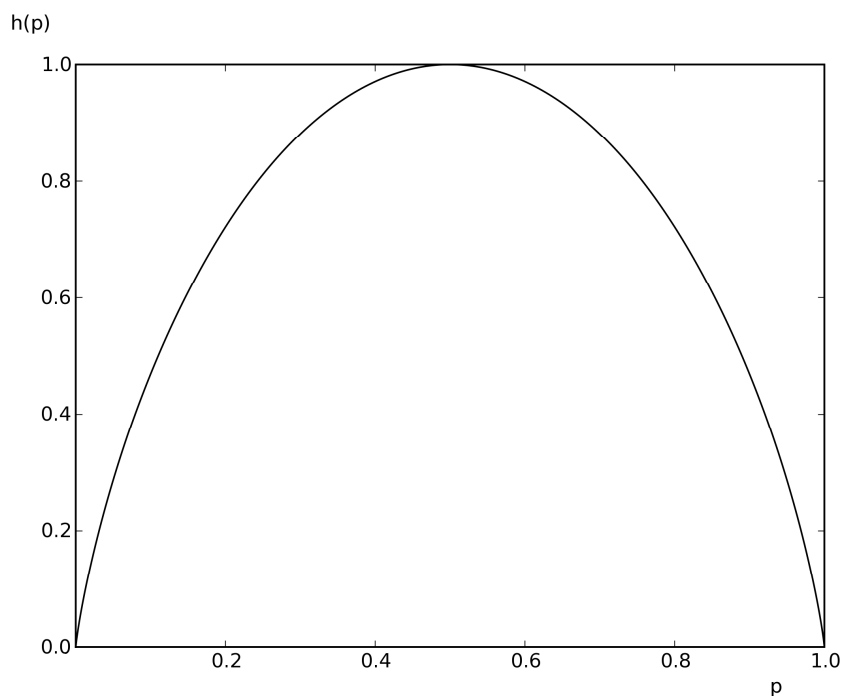


Gráfico 2: Comportamento da entropia da fonte $h(p)$ para valores de p no intervalo $0 \leq p \leq 1$. A entropia é mínima ($h=0$) quando um dos caracteres aparece com probabilidade $p=1$ e máxima ($h=1$) quando $p=0,5$.

Visto a possível redução de complexidade resultante do fato de a entropia da fonte geradora não ser máxima, Lempel e Ziv (1976) propõe a seguinte normalização adicional:

$$\lim_{n \rightarrow \infty} c / bh = 1$$

que permite estabelecer comparações entre seqüências com fontes geradoras que têm entropia distintas. No caso em que $p = 0,5$, a entropia é $h = 1$ e a normalização se torna neutra. O presente trabalho utilizou em todos os experimentos esta medida de complexidade normalizada tanto pela complexidade máxima quanto pela entropia da fonte, e sempre notada por $LZ = c / bh$.

3.4 Aplicação em Finanças

Esta subseção trata de dois pontos específicos da aplicação da medida LZ utilizados neste trabalho: a codificação das séries financeiras e o procedimento de janelas deslizantes

Codificação das séries financeiras: A codificação das séries financeiras foi feita como em Shmilovici, Alon-Brimer e Hauser (2003), ao realizar teste de compressibilidade para verificar a eficiência informacional de alguns índices. Define-se um código em que '0' representa baixa, '1' representa alta e '2' significa que a série ficou dentro de uma região de estabilidade previamente definida. Se tal região for definida como zero, o resultado será a codificação binária ('0' ou '1').

Seja v_t a série de retornos diários do índice analisado, e a região de estabilidade e s_t a seqüência codificada, o procedimento descrito acima pode ser representado por:

$$s_t = '0' \text{ se } v_t < -e$$

$$s_t = '2' \text{ se } -e \leq v_t < e$$

$$s_t = '1' \text{ se } v_t \geq e$$

Para exemplificar, considere a seguinte série de retornos apresentada pelo índice *Bovespa* na semana compreendida pelos dias 28 e 31 de agosto de 2007: 0.15%, -2.17%, 2.11%, 0.23% e 3.37%. Utilizando uma região de estabilidade $e = 0.25\%$, a seqüência codificada que representa essa semana é dada por “20121”.

Janelas deslizantes: Existe uma probabilidade positiva de que a medida de complexidade calculada seja determinada de maneira aleatória, o que justifica o tratamento por janelas deslizantes. Sendo n o número de observações consideradas e j o tamanho escolhido para a janela, a média das complexidades de todas as $n - j + 1$ janelas analisadas é calculada.

Supondo $n = 2000$ e $j = 1000$, a complexidade da janela de 1 a 1000 é calculada, depois a de 2 a 1001, assim por diante até a janela de 1001 a 2000. A complexidade do período $n = 2000$, considerando-se janelas de tamanho $j = 1000$, é dada pela média das complexidades das $n - j + 1 = 1001$ janelas deslizantes da amostra. O Gráfico 3 mostra um experimento com 15000 observações (14000 janelas de tamanho 1000) de uma série pseudo-aleatória gerada por computador.

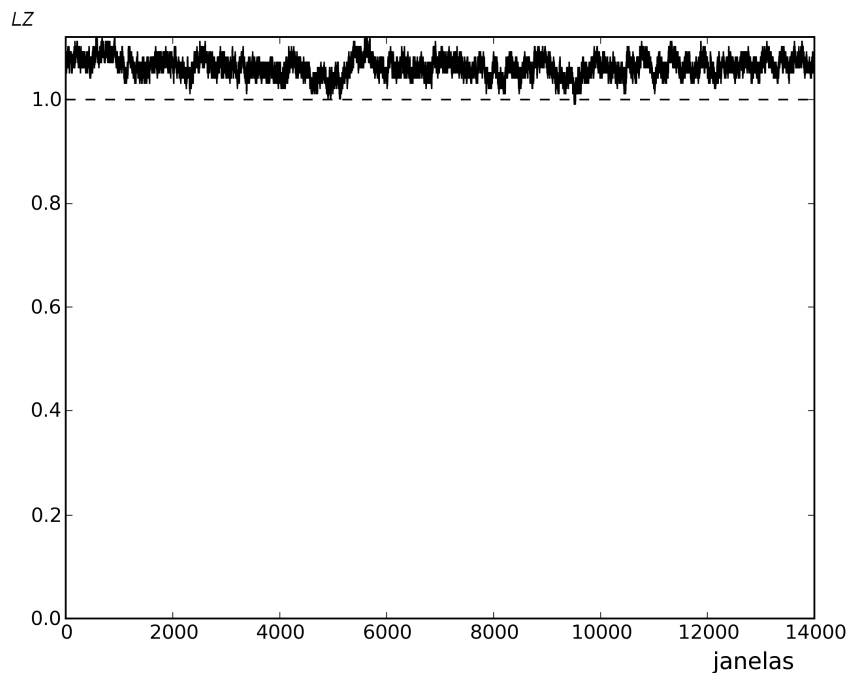


Gráfico 3: Mostra a medida de complexidade normalizada LZ de uma seqüência de 15000 números pseudo-aleatórios gerados por computador em um experimento com 14000 janelas deslizantes.

A linha tracejada indica o ponto de referência utilizado para avaliar o grau de aleatoriedade de uma seqüência. Como discutido na seção que diz respeito às normalizações, a medida LZ de uma série aleatória atinge por cima seu limite superior assintótico $\lim_{n \rightarrow \infty} LZ = 1$ quando $n \rightarrow \infty$.

Então, para seqüências finitas como as que representam séries financeiras, este trabalho considera que janelas com $LZ < 1$ constituem indícios de desvios da aleatoriedade pura. O capítulo que trata dos experimentos descreve como este ponto de referência é utilizado para estabelecer comparações relativas do grau de eficiência dos índices econômicos analisados.

3.5 Propriedades

Esta seção apresenta o comportamento da medida de complexidade normalizada LZ , descrita nas seções anteriores, frente a seqüências determinadas pelas distâncias entre números primos adjacentes, reconhecidas como aleatórias, e frente a um simples sistema dinâmico (mapa logístico) que exhibe tanto comportamento estável como caótico. A medida LZ se mostra robusta no sentido de diferenciar a complexidade destes processos distintos.

Números primos:

A Figura 2 mostra um experimento com 15001 observações (14000 janelas de tamanho 1000) para (a) distância entre os 15001 primeiros números primos e (b) idem ao anterior, porém com a distância normalizada pelo logaritmo natural das distâncias entre os primos.

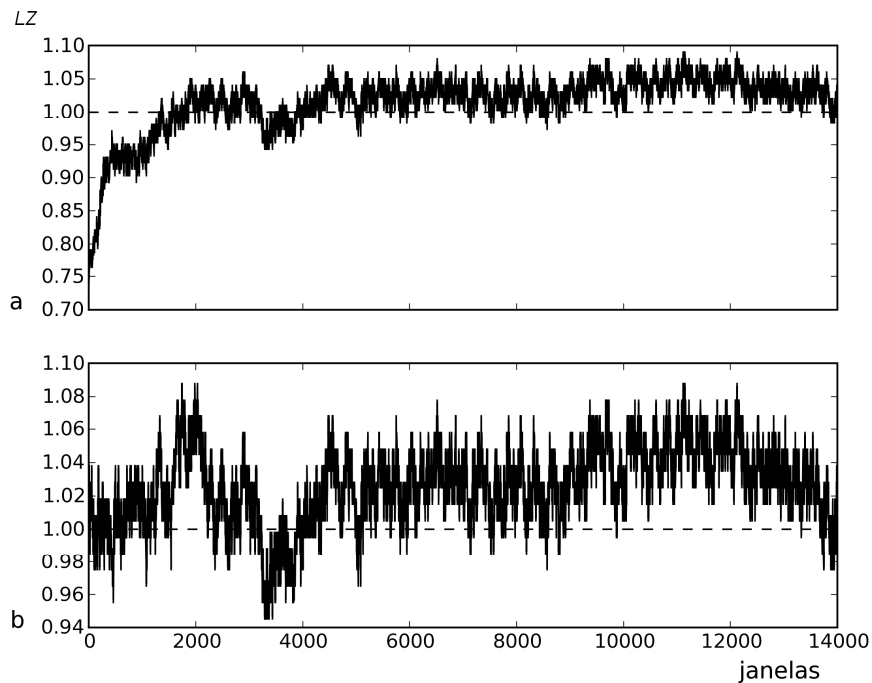


Figura 2: Dois experimentos que apresentam a medida LZ para 14000 janelas deslizantes de seqüências geradas pela (a) distância entre os 15001 primeiros números primos e (b) para o logaritmo natural destas distâncias.

A medida LZ se mostra efetiva para perceber a alta complexidade da seqüência gerada pelos números pseudo-aleatórios (exibida na seção anterior, LZ médio 1,0626), assim como a da seqüência que representa a distância entre números primos adjacentes (LZ médio 1,0143 e LZ médio 1,0255 para o logaritmo das distâncias).

Mapa logístico:

A equação logística constitui um dos modelos determinísticos mais simples que apresentam comportamento caótico (AUYANG, 1998). O mapa logístico, gerado pela equação logística, pode ser utilizado, por exemplo, para descrever a dinâmica de uma população com o passar do tempo incorporando suas taxas de nascimento e mortalidade.

O tamanho da população no período t , expressa em termos relativos à capacidade do ecossistema, pode ser descrito por $0 \leq x_t \leq 1$, sendo que, se $x_t = 0$ a população não existe ou foi extinta, e se $x_t = 1$, a população atingiu o limite do ecossistema.

Para estudarmos a dinâmica do crescimento populacional por meio do mapa logístico, assume-se que o tamanho da população em x_{t+1} é uma função do tamanho da população em x_t e da taxa de crescimento populacional r sob a seguinte função:

$$x_{t+1} \equiv f(x_t) = rx_t(1 - x_t)$$

Note que a taxa de crescimento r deve ser limitada entre $0 \leq r \leq 4$, para que o sistema tenha fronteiras e que o mapa tenha sentido em termos físicos. Dessa maneira, pode-se pensar os dois primeiros termos do lado direito da fórmula (r e x_t) como uma alimentação positiva (reprodução), ao contrário do último termo ($1 - x_t$), que seria uma alimentação negativa no tamanho da população (competição por recursos escassos).

Partindo de qualquer valor inicial para o tamanho x_0 da população, pode-se verificar o comportamento dinâmico do mapa alterando a taxa de crescimento r também chamada de parâmetro logístico. Percebe-se que, para $0 \leq r \leq 3$, o tamanho populacional é atraído para algum ponto fixo estável, $x = 0$ ou $x = 1 - 1/r$, pois um ponto fixo é definido como estável se gerar ele mesmo como próximo valor em um mapa na forma $x_{t+1} = x_t = rx_t(1 - x_t)$ que tem estes dois pontos como raízes. Por exemplo, para $r = 2,5$, o tamanho populacional é atraído para o valor $0,6$.

Estes pontos fixos como atratores imprimem um forte padrão na seqüência gerada pelo mapa, o que faz com que a medida de complexidade LZ desta região seja muito baixa. O pequeno salto de complexidade quando $r = 2$ deve-se ao fato de que, a partir deste valor, o ponto fixo é atingido após certa oscilação. Para valores de $r < 2$, o ponto fixo estável é atingido sem oscilação.

Entretanto, para valores de r compreendidos pelo intervalo $3 < r \leq 4$, o tamanho populacional não é atraído para nenhum ponto fixo estável (a não ser para tamanhos iniciais iguais a 0 ou $1 - 1/r$), mas sim entra em ciclos de períodos n (alternância entre n pontos não fixos). A Figura 3 é dividida em quatro gráficos, sendo que: (a) existe estabilidade no ponto fixo igual a 0,6 quando $r = 2,5$, (b) mostra que, para $r = 3,23$, o mapa entra em um ciclo de período 2, ou seja, alterna entre dois pontos não fixos; (c) mostra o ciclo de período 4 atingido quando $r = 3,45$ e (d) mostra o ciclo de período n onde não existem mais ciclos, e sim comportamento caótico.

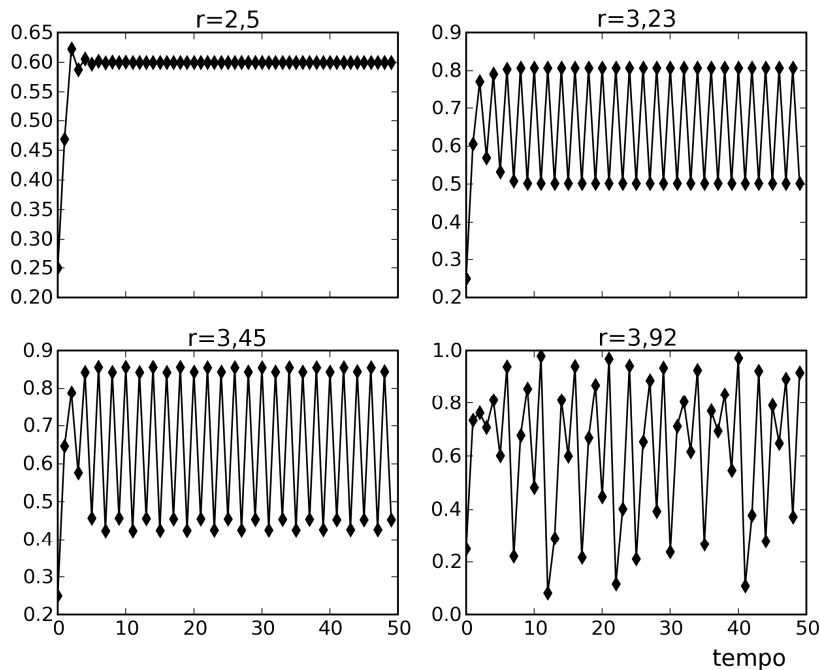


Figura 3: Apresentação do mapa logístico para quatro diferentes valores do parâmetro de crescimento ($r = 2,5, r = 3,23, r = 3,45$ e $r = 3,92$). Para valores de $r < 3$, existe estabilidade em algum ponto fixo. Para valores entre $3 \leq r \leq 4$, não há estabilidade e o sistema entra em crescente complexidade.

A medida LZ captura estes diversos momentos, vide a baixa complexidade na região de estabilidade e o aumento crescente de complexidade do mapa logístico quando este ele na fase de bifurcação (duplicação de períodos) e atinge seu máximo na fase caótica, conforme aponta o Gráfico 4 abaixo.

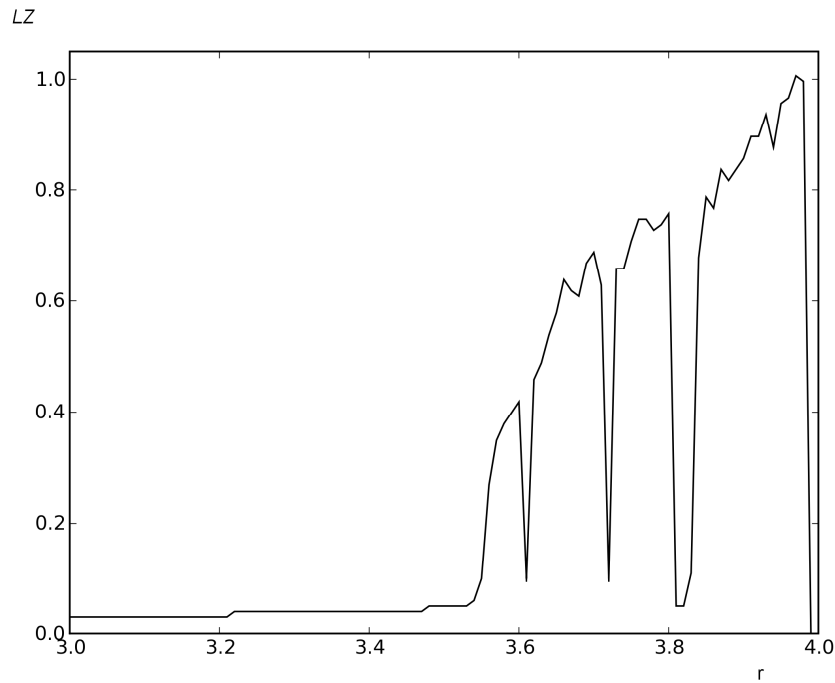


Gráfico 4: Apresenta a complexidade do mapa logístico em função de seu parâmetro de crescimento r . O gráfico mostra que a medida LZ é robusta para diferenciar seqüências de complexidades distintas.

3.6 Comparação com a Entropia Aproximada ($ApEn$)

A Entropia Aproximada é uma estatística que mede a regularidade de uma série de tempo. É exibida em Pincus (1991) e utilizada por Gabjin, Kim e Eom (2007) para estabelecer comparações de eficiência em termos relativos para algumas taxas de câmbio em uma abordagem similar à realizada neste trabalho.

Assim como a complexidade de LZ , a $ApEn$ também pode ser entendida como uma medida de aleatoriedade que independe de um modelo estabelecido, *a priori*, sobre o comportamento da série de tempo analisada. Esta subseção descreve as propriedades da $ApEn$ e apresenta uma comparação desta em relação à medida LZ .

A estatística $ApEn$ é uma medida de regularidade que quantifica a imprevisibilidade da flutuação em uma série temporal. Intuitivamente, pode-se imaginar que a presença de padrões repetitivos na flutuação de uma série de tempo faz com que esta seja mais previsível do que outra em que não existam tais padrões. A $ApEn$, de

modo geral, indica a probabilidade de que determinada seqüência de observações (um padrão) não vai ser seguida por seqüências similares.

Uma série de tempo que contém muitos padrões repetitivos tem $ApEn$ relativamente menor do que uma série mais imprevisível (mais complexa) com menos padrões repetitivos.

Seja uma seqüência S_N consistindo de N observações consecutivas S_1, S_2, \dots, S_N . Para calcular a $ApEn$, além da série de tempo, existem ainda dois parâmetros: m e r .

$$ApEn(S_N, m, r)$$

em que m especifica o tamanho do padrão e r define o critério de similaridade.

Define-se uma subseqüência (ou padrão a ser analisado) de m observações que começa em S_i pelo vetor $p_m(i)$. Duas subseqüências, $p_m(i)$ e $p_m(j)$, são consideradas similares se a diferença entre qualquer par correspondente é menor do que r .

$$|S(i+k) - S(j+k)| < r, \text{ para } \forall k \text{ em } [0, m[$$

Considerando o conjunto P_m de todas as $N - m + 1$ subseqüências de tamanho m em S_N . Seja

$$C_{im}(r) = \frac{n_m(r)}{N - m + 1}$$

em que $n_m(r)$ é o número de subseqüências em P_m que são similares (dado o critério r de similaridade) a $p_m(i)$. $C_{im}(r)$ representa a proporção de todas as subseqüências de tamanho m que são similares àquela que começa em i . Pode-se calcular $C_{im}(r)$ para todas as subseqüências em P_m , e então define-se $C_m(r)$ como a média de todos os valores $C_{im}(r)$. $C_m(r)$ representa a quantificação da ocorrência

de padrões repetitivos de tamanho m em S_N . Finalmente, define-se a entropia aproximada de S_N para subsequências de tamanho m e critério de proximidade r como:

$$ApEn(S_N, m, r) = \ln[C_m(r) / C_{m+1}(r)]$$

Em palavras, $ApEn$ é o logaritmo natural da ocorrência relativa de padrões repetitivos de tamanho m comparado à ocorrência de padrões de tamanho $m+1$. Estes dois valores ($C_m(r)$ e $C_{m+1}(r)$) serão próximos em séries claramente periódicas e se distanciarão mais quanto menos padrões ocorrerem na seqüência.

Então, se encontrarmos padrões similares em alguma série de tempo, $ApEn$ é uma estimativa da verossimilhança logarítmica de que o próximo intervalo $m+1$ difere de m (e a similaridade de subsequências não é uma mera coincidência que ocorre para diversos intervalos).

Valores baixos de $ApEn$ indicam grande probabilidade de que a quantificação de similaridade em um intervalo é próxima à quantificação da similaridade para o próximo intervalo. Se a série de tempo é altamente irregular, a ocorrência de padrões similares não é útil para prever a similaridade do próximo intervalo e então o valor de $ApEn$ é alto.

Contudo, deve-se notar que a medida $ApEn$ é estatisticamente fraca, devido à sua forte dependência dos parâmetros m e r (RICHMAN; MOORMAN, 2000). Outra questão referente à $ApEn$ é apontada por Ferrario, Signorini e Magenes (2005), que comparam o comportamento desta medida com o da medida proposta por Lempel e Ziv (LZ). Tais autores mencionam que ambas as medidas são aptas a captar a crescente complexidade do mapa logístico, porém, apontam para o fato de que a medida $ApEn$ de uma seqüência A pode ser maior do que outra B para certos valores de m e r e que esta relação pode ser invertida para outros valores de m e r , ocasionando um problema de transitividade não encontrado na medida LZ .

4 Experimentos

Todos os dados utilizados foram obtidos em duas fontes: Yahoo!Finance e EconStats. As séries históricas referentes às bolsas e taxas de câmbio estão disponíveis em www.econstats.com e as séries de preços de ações de empresas, em finance.yahoo.com. Último acesso em 30 de janeiro de 2008.

Neste trabalho, foram utilizados dados de retornos diários (retornos percentuais de um período) de diversas séries históricas financeiras. Campbell, Lo e Mackinley (1997) apontam duas vantagens em trabalhar com os retornos dos preços, em contrapartida às séries em nível. Em primeiro lugar, os retornos servem como índice livre de escala para o investidor e, em segundo, apresentam propriedades estatísticas vantajosas (como estacionaridade, por exemplo).

Para as séries financeiras de bolsas e ações analisadas, foram utilizadas as 2000 observações mais recentes anteriores a 15 de junho de 2007, totalizando aproximadamente sete anos de observações diárias. Para as taxas de câmbio, também foram utilizadas 2000 observações, porém as mais recentes anteriores a 20 de dezembro de 2007.

Estas 2000 observações foram usadas no procedimento de janelas deslizantes descrito na seção anterior para constituir 1000 janelas (2000 observações em nível resultam em 1999 observações de retornos de um período que, por sua vez, geram 1000 janelas deslizantes).

Assim como discutido anteriormente, o presente trabalho utiliza a referência $LZ < 1$ como indício de desvios da condição de aleatoriedade. Esta seção apresenta, para cada experimento, o LZ médio das 1000 janelas deslizantes e o grau de eficiência de mercado, definido como $P(LZ > 1)$, ou seja, o número de janelas dentre as 1000 que apresentaram $LZ > 1$.

O restante deste capítulo apresenta os três experimentos realizados, na ordem: (1) 36 bolas selecionadas, (2) 37 ações de empresas negociadas na Bovespa, NYSE e Nasdaq e (3) taxas de câmbio de 19 moedas selecionadas, todas em relação ao

dólar americano. As tabelas e figuras com os resultados são apresentados em páginas separadas no decorrer do texto.

4.1 Bolsas

Este experimento foi realizado com 2000 observações diárias (as mais recentes anteriores a 15 de julho de 2007) do preço de abertura de pregão de 36 bolsas de valores selecionadas. A Figura 4 mostra o experimento realizado com séries históricas de preços dos índices (a) S&P500, (b) DAX30, (c) *All Ords.* Austrália, (d) Shanghai composite e (e) Karachi 100. A linha tracejada indica o ponto de referência utilizado $LZ = 1$ como indício de desvios da condição de aleatoriedade.

Os gráficos destacam que todas as cinco séries analisadas são bastante complexas, verificação apoiada por boa parte da literatura sobre as hipóteses de eficiência de mercado. Elas se parecem mais com séries aleatórias do que com séries redundantes ou bastante previsíveis.

Entretanto, considerando o ponto de referência para desvios da condição de aleatoriedade, podemos verificar quantas das 1000 janelas calculadas apresentam $LZ > 1$ e considerar este número como um indício de aleatoriedade. Visto que este trabalho se baseia na relação entre aleatoriedade e eficiência, chamamos este número de grau de eficiência. A Tabela 1 contém os resultados completos para as 36 bolsas analisadas. Para cada bolsa a tabela apresenta o LZ médio do procedimento de janelas e o grau de eficiência, em porcentagem, que representa o número de janelas que se situam na região de aleatoriedade $LZ > 1$.

LZ

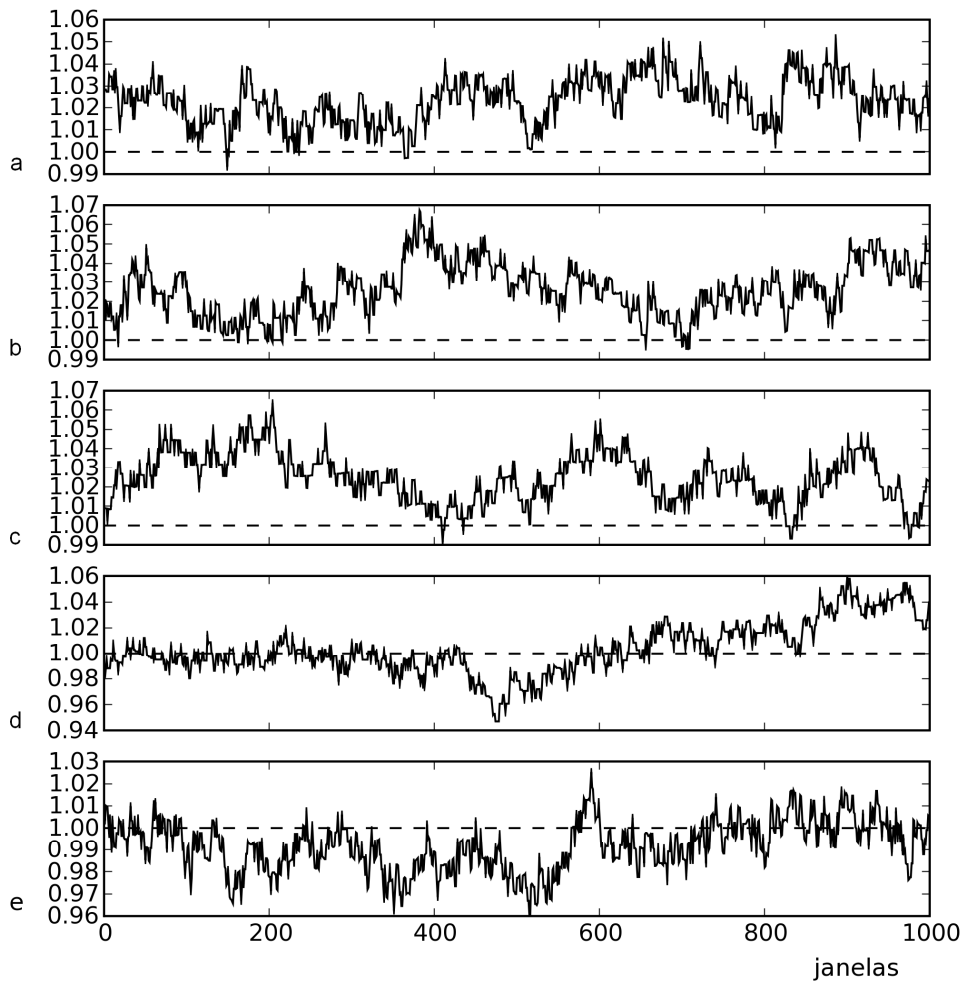


Figura 4: Apresenta cinco gráficos que relacionam a medida *LZ* (ordenada) para 1000 janelas deslizantes (abscissa) para quatro bolsas selecionadas, sendo (a) S&P500 (*LZ* médio = 1,0232, grau de eficiência = 99,10%), (b) DAX30 (*LZ* médio = 1,0257, grau de eficiência = 98,40%), (c) *All Ords.* Austrália (*LZ* médio = 1,0246, grau de eficiência = 97,80%), (d) Shanghai Composite Index (*LZ* médio = 1,0032, grau de eficiência = 49,50%) e (e) Karachi 100 (*LZ* médio = 0,9918, grau de eficiência = 23,70%).

Bolsa	País	LZ médio	Grau de eficiência, %
S&P 500	Estados Unidos	1.0232	99.10
DAX 30	Alemanha	1.0257	98.40
Nikkei 225	Japão	1.0432	98.20
All Ords.	Austrália	1.0246	97.80
ATX	Áustria	1.0173	97.40
DJIA	Estados Unidos	1.0201	95.40
Korea Composite	Coréia do Sul	1.0163	94.90
TA 100	Israel	1.0187	92.90
Hang Seng	Hong Kong	1.0151	91.50
Straits Times	Singapura	1.0153	90.30
CAC 40	França	1.0138	88.40
Helsinki General	Finlândia	1.0149	88.40
Kuala Lumpur SE	Malásia	1.0158	88.00
FTSE 100	Reino Unido	1.0106	86.60
PX 50	República Checa	1.0139	81.00
Bel 20	Bélgica	1.0118	80.40
IBC	Venezuela	1.0110	79.90
Madrid General	Espanha	1.0201	79.30
Swiss Market	Suíça	1.0101	78.40
Nasdaq	Estados Unidos	1.0080	75.40
Amsterdam EX	Holanda	1.0100	74.40
Bovespa	Brasil	1.0127	67.80
IPC	México	1.0060	64.00
Merval	Argentina	1.0050	62.90
Jakarta Composite	Indonésia	1.0054	62.10
Istanbul 100	Turquia	1.0085	61.30
Moscow Times	Rússia	1.0050	59.20
Copenhagen	Dinamarca	1.0025	58.70
Athens Composite	Grécia	1.0048	56.90
Bombay SE	Índia	1.0010	53.30
Taiwan Weighted	Taiwan	1.0006	50.30
Shanghai Composite	China	1.0032	49.50
PSE	Filipinas	0.9987	43.10
Lima General	Peru	0.9903	37.90
Karachi 100	Paquistão	0.9918	23.70
Colombo SE	Sri Lanka	0.9795	10.50

Tabela 1: Mostra os resultados completos do experimento com 36 bolsas selecionadas, ordenadas pelo grau de eficiência. Apresenta a bolsa analisada, seu país, índice *LZ* médio do procedimento de janelas e o grau de eficiência, que corresponde à probabilidade de encontrar janelas com $LZ > 1$.

4.2 Ações

Assim como o anterior, este experimento foi realizado com 2000 observações diárias (as mais recentes anteriores a 15 de julho de 2007) do preço de abertura de pregão de ações de 37 empresas selecionadas. Como ilustração, a Figura 5 apresenta o experimento realizado com os preços de (a) Coca-Cola, (b) Yahoo!, (c) Vale, (d) Toyota e (e) Aracruz. A linha tracejada indica o ponto de referência utilizado $LZ = 1$ como indício de desvios da condição de aleatoriedade.

Os resultados completos para as ações das 37 empresas analisadas encontram-se na Tabela 2 a seguir.

LZ

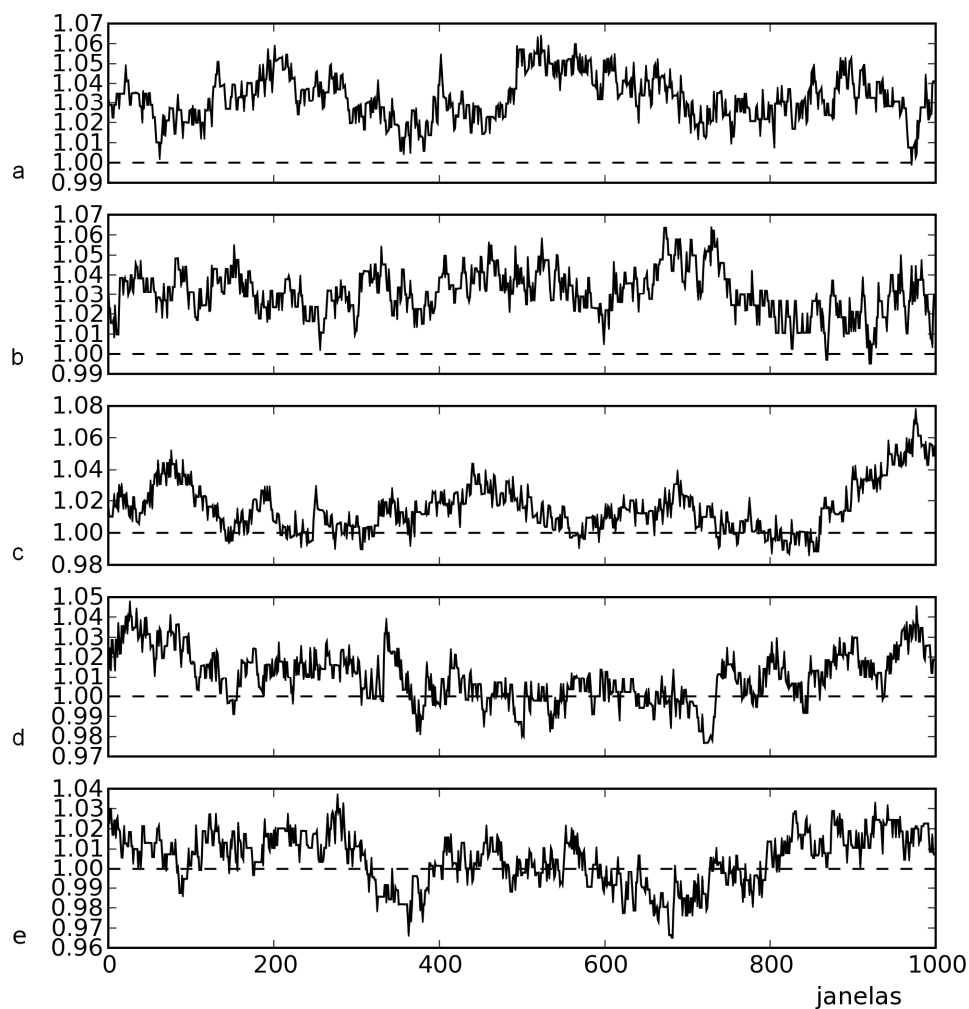


Figura 5: Apresenta cinco gráficos que relacionam a medida *LZ* (ordenada) para 1000 janelas deslizantes (abscissa) para preço de ações de quatro empresas selecionadas, sendo (a) Coca-Cola (*LZ* médio = 1,0324, grau de eficiência = 100,00%), (b) Yahoo! (*LZ* médio = 1,0310, grau de eficiência = 99,65%), (c) Vale (*LZ* médio = 1,0193, grau de eficiência = 92,75%), (d) Toyota (*LZ* médio = 1,0100, grau de eficiência = 76,32%) e (e) Aracruz (*LZ* médio = 1,0048, grau de eficiência = 66,67%).

Empresa	Bolsa	LZ médio	Grau de eficiência, %
Amazon	NYSE	1.0416	100.00
Coca-Cola	NYSE	1.0324	100.00
P&G	NYSE	1.0264	99.97
Intel	Nasdaq	1.0292	99.92
eBay	Nasdaq	1.0377	99.80
General Electric	NYSE	1.0274	99.66
Yahoo	Nasdaq	1.0310	99.65
Texaco	NYSE	1.0264	99.46
Cisco	Nasdaq	1.0357	99.44
Petrobras	Bovespa	1.0284	99.43
Pfizer	NYSE	1.0327	99.39
HP	NYSE	1.0298	99.38
Microsoft	Nasdaq	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	1.0178	97.58
Merrill Lynch	NYSE	1.0279	97.33
JP Morgan	NYSE	1.0281	96.70
Oracle	Nasdaq	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	86.60
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

Tabela 2: Mostra os resultados completos do experimento com ações de 37 empresas selecionadas, ordenadas pelo grau de eficiência. Apresenta a ação analisada, sua respectiva bolsa, índice *LZ* médio do procedimento de janelas e o grau de eficiência, que corresponde à probabilidade de encontrar janelas com $LZ > 1$.

4.3 Câmbio

Da mesma forma que os experimentos anteriores, este foi realizado com 2000 observações diárias, todavia as mais recentes anteriores a 20 de dezembro de 2007, do preço de abertura de pregão de 19 moedas selecionadas, todas em relação ao dólar norte-americano (US\$). Como ilustração, a Figura 6 na seqüência apresenta o experimento realizado com as taxas de câmbio em relação ao US\$ da (a) Libra esterlina (b) Euro, (c) Peso mexicano, (d) Real e (e) Rúpia indiana. Assim como nas figuras anteriores, a linha tracejada indica o ponto de referência utilizado $LZ = 1$ como indício de desvios da condição de aleatoriedade.

As razões que geram as ineficiências em mercados específicos não são o escopo deste trabalho, mas quanto ao mercado de câmbio, uma constatação se torna pertinente. O controle da taxa de câmbio, exercido pelas respectivas autoridades monetárias na forma de câmbio fixo ou administrado, pode imprimir forte padrão nas seqüências acarretando baixa complexidade e, por decorrência, baixa eficiência (em um mercado em que o preço é exogenamente determinado, não pode ser considerado eficiente).

Os resultados completos para as 19 taxas de câmbio analisadas são apresentados na Tabela 3.

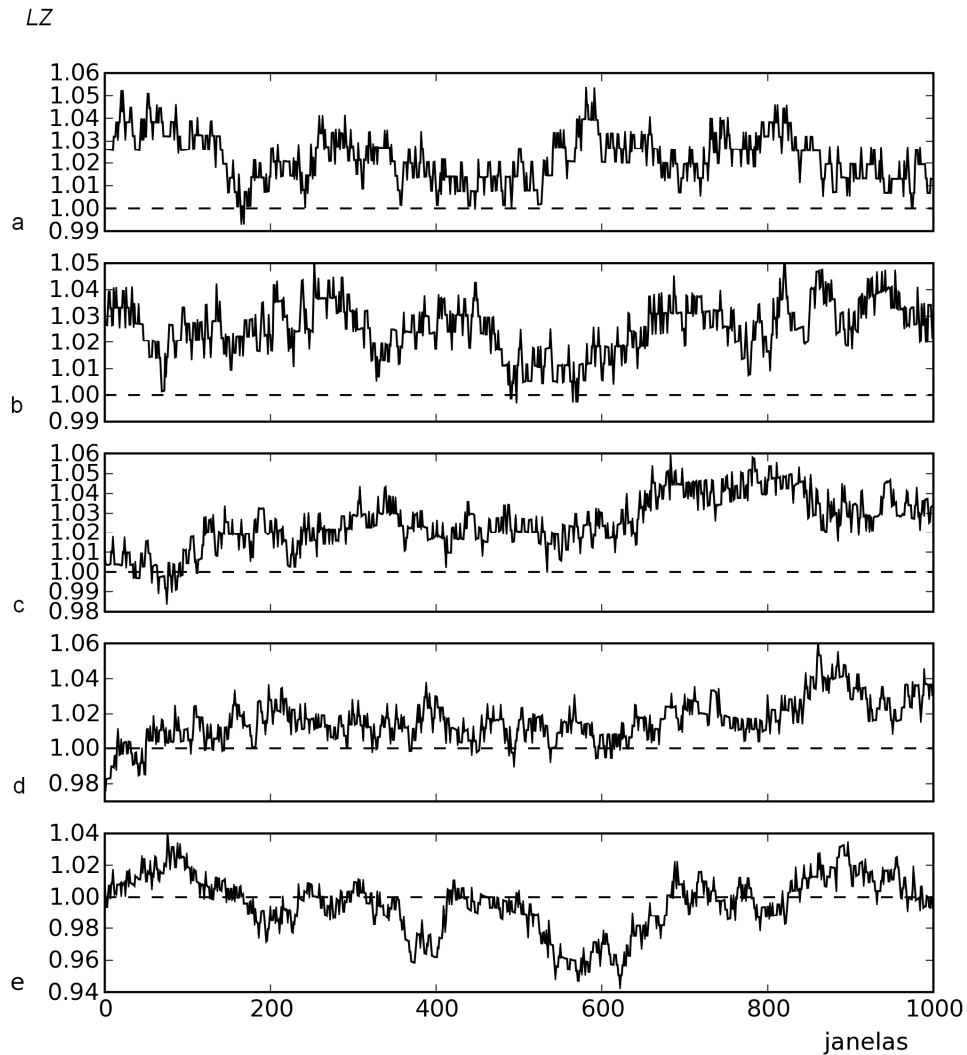


Figura 6: Apresenta cinco gráficos que relacionam a medida *LZ* (ordenada) para 1000 janelas deslizantes (abscissa) para taxas de câmbio em relação ao dólar norte americano (US\$) de quatro moedas selecionadas, sendo (a) Libra esterlina (*LZ* médio = 1,0223, grau de eficiência = 99,81%), (b) Euro (*LZ* médio = 1,0254, grau de eficiência = 99,45%), (c) Peso mexicano (*LZ* médio = 1,0254, grau de eficiência = 96,58%), (d) Real (*LZ* médio = 1,0156, grau de eficiência = 92,60%) e (e) Rúpia indiana (*LZ* médio = 0,9958, grau de eficiência = 43,54%).

Moeda	País	<i>LZ</i> médio	Grau de eficiência, %
Libra Esterlina	Inglaterra	1.0223	99.81%
Coroa sueca	Suécia	1.0236	99.71%
Coroa norueguesa	Noruega	1.0314	99.60%
Euro	Zona do Euro	1.0253	99.45%
Dólar neozelandês	Nova Zelândia	1.0248	99.20%
Franco suíço	Suíça	1.0169	99.12%
Króna	Islândia	1.0184	97.48%
Peso mexicano	México	1.0254	96.58%
Real	Brasil	1.0156	92.60%
Dólar canadense	Canadá	1.0219	90.07%
Rand	África do Sul	1.0177	86.41%
Yene	Japão	1.0153	85.51%
Dólar	Singapura	1.0074	66.48%
Dólar australiano	Austrália	1.004	63.41%
Rúpia indiana	Índia	0.9957	43.54%
Peso colombiano	Colômbia	0.9913	21.98%
Novo Dólar	Taiwan	0.9794	21.17%
Yuan	China	0.9265	17.94%
Rúpia	Sri Lanka	0.9687	11.84%

Tabela 3: Mostra os resultados completos do experimento com 19 taxas de câmbio selecionadas, todas em relação ao dólar americano, ordenadas pelo grau de eficiência. Apresenta a moeda analisada, seu respectivo país, índice *LZ* médio do procedimento de janelas e o grau de eficiência, que corresponde à probabilidade de encontrar janelas com $LZ > 1$.

5 Conclusão

Este trabalho apresenta um modo de verificar a eficiência de mercados financeiros em termos relativos por meio da medida de complexidade proposta por Lempel e Ziv (1976), que é uma útil medida da idéia de complexidade presente na Teoria da Informação Algorítmica (KASPAR; SCHUSTER, 1987).

A complexidade de um objeto, segundo a Teoria da Informação Algorítmica, é definida como igual ao tamanho do menor programa de computador que pode computá-lo. Nesse sentido, uma seqüência é considerada aleatória quando o tamanho do menor programa de computador que pode reconstruí-la é igual ao tamanho da própria seqüência.

Em outras palavras, uma seqüência aleatória não tem informações redundantes que possibilitem sua compressão e representação por um pequeno (menor do que ela própria) programa de computador.

Existe literatura em Finanças que relaciona a eficiência de mercados financeiros com a aleatoriedade de seus preços, por meio das Hipóteses de Eficiência de Mercado.

Um mercado é dito eficiente quando toda a informação disponível é instantaneamente processada quando chega ao mercado e é imediatamente refletida em um novo valor de preços dos ativos negociados. Como as novas informações não podem ser previstas, o preço deve flutuar de maneira aleatória (SAMUELSON, 1965).

Campbell, Lo e MacKinlay (1997) argumentam que as hipóteses de eficiência de mercado podem ser vistas como situações idealizadas não presentes em mercados reais, mas que podem ser úteis para avaliar a eficiência em termos relativos.

Sistemas físicos geralmente são avaliados quanto à eficiência em termos relativos. Um motor é dito 30% eficiente se converte esta taxa da energia presente no combustível para energia mecânica. O presente trabalho utiliza uma avaliação de

eficiência similar, baseada na quantidade relativa de informações redundantes presente em uma série financeira. O preço em um mercado idealmente eficiente não conteria informações redundantes e o mercado seria dito 100% eficiente.

Em mercados reais, os preços refletem apenas as informações cujos custos de aquisição não excedam os benefícios de utilizá-las. As expectativas dos agentes do mercado não necessariamente são homogêneas. Existem ainda diferentes custos de transação e informações privadas como possíveis geradores de fricção. Grossman e Stiglitz (1980) dizem que mercados perfeitamente eficientes são uma impossibilidade.

Neste sentido, as ineficiências residuais estarão sempre presentes nos mercados reais, e não podemos esperar que estes sejam 100% eficientes em termos absolutos. Considerando a referência de seqüências conhecidas como aleatórias, podemos verificar a quantidade relativa de informações redundantes presentes em séries históricas de preços de diversos mercados financeiros.

Este trabalho utilizou dados de 36 bolsas mundiais, preços de ações de 37 empresas e 19 moedas relacionadas ao dólar norte-americano. Percebeu-se, por exemplo, que o índice *Dow Jones* é 95,4% e que o índice *Merval* é 62,9 % eficiente. Isto significa que a série histórica do índice norte-americano apresenta menos informações redundantes do que o argentino.

Referências bibliográficas

AUYANG, S. *Foundations of Complex-system Theory in Economics, Evolutionary Biology and Statistical Physics*. Cambridge, England, Cambridge University Press, 1998.

BACHELIER, L. **Théorie de la spéculation**. Annales scientifiques de l'É.N.S. 3^a série, tome 17, p. 21-86, 1900. Disponível em:
<http://www.numdam.org/item?id=ASENS_1900_3_17__21_0>. Acesso em 30 jan. 2008.

BEECHEY, M.; GRUEN, D.; VICKERY, J. “The efficient market hypothesis: a survey”. *RBA Research Discussion Papers 2000-01*, Reserve Bank of Australia, 2000.

BROCK, W.; LIMA, P. “Nonlinear time series, complexity theory and finance”. In: MADALLA, G.; RAO, C. (Eds), *Handbook of Statistics: Statistical methods in finance*, North Holland, Amsterdam, v. 14, p. 317-361, 1996.

CALUDE, C. *et al.* (Orgs.) **Algorithmic information theory**. Centre for Discrete Mathematics and Theoretical Computer Science. Disponível em:
<<http://www.cs.auckland.ac.nz/CDMTCS/docs/ait.html>>. Acesso em: 30 jan. 2008.

CAMPBELL, J.; LO, A.; MACKINLAY, C. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, 1997.

CHAITIN, G. *Information randomness and incompleteness: Papers on Algorithmic Information Theory*. World Scientific, Series in Computer Science, v. 8, 1987.

COOPER, S. *Computability Theory*. Chapman & Hall/CRC Press, Boca Raton, FL, 2004.

COSTA JR, N da.; LEMOS, M. “O efeito de sobre-reação no curto prazo no mercado de capitais brasileiro”. *Revista Brasileira de Economia*, v. 51, n. 3, p. 323-324, 1997.

COWLES, A.; JONES, H. “Some a posteriori probabilities in stock market action”. *Econometrica*, v. 5, p. 280-294, 1937.

DIMSON, E.; MUSSAVIAN, M. “A brief history of market efficiency”. *European Financial Management*, v. 4, n. 1, p. 91-103, 1998.

FAMA, E. F. “Efficient capital markets: a review of theory and empirical work.” *Journal of Finance*, v. 25, n. 2, p. 383-417, 1970.

FAMA, E. F. “Efficient capital markets II”. *Journal of Finance*, v. 46, n. 5, p. 1575-1617, 1991.

FERRARIO, M.; SIGNORINI, M.; MAGENES, G. “Complexity analysis of the fetal heart rate for the identification of pathology in fetuses”. *Computers in Cardiology*, v. 32, p. 989-992, 2005.

GABJIN, O., KIM, S., EOM, C. “Market efficiency in Foreign Exchange Markets”, *Physica A: Statistical Mechanics and its Applications*, v. 382, n. 1, p. 209-212, 2007.

GROSSMAN, S.; STIGLITZ, J. "On the impossibility of informationally efficient markets". *The American Economic Review*, v. 70, n. 3, p. 393-408, 1980.

HAAHR, M. **Introduction to randomness and random numbers**. random.org, 1999. Disponível em <<http://random.org/essay/html>>, Acesso em 30 jan. 2008.

JENSEN, M. C. "Some anomalous evidence regarding market efficiency". *Journal of Financial Economics*, v. 6, n. 2/3, p. 96-101, 1978.

KASPAR, F.; SCHUSTER, H. "Easily calculable measure for the complexity of spatiotemporal patterns". *The American Physical Society, Physical Review A*, v. 36, n. 2, p. 842-848, 1987.

KENDALL, M. "The analysis of economic time series, part I: prices". *Journal of the Royal Statistical Society*, v. 96, p. 11-25, 1953.

KJELLERSTRAND, H. **de Bruijn Sequences**. Disponível em: <<http://www.hakank.org/comb/debruijn.cgi>>, Acesso em: 30 jan. 2008.

LEMPEL, A.; ZIV, J. "On the complexity of finite sequences". *IEEE Transactions on Information Theory*, v. 22, n. 1, p. 75-81, 1976.

LO, A.; MACKINLAY, C. "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, 1988.

LO, A.; MACKINLAY, C. *Non Random Walk Wall Street*. Princeton University Press, 1999.

MANDELBROT, B. "Forecasts of future prices, unbiased markets, and martingale models". *Journal of Business*, v. 39, n. 1, p. 242-255, 1966.

MANTEGNA, R.; STANLEY, H. *An Introduction to Econophysics: correlation and complexity in finance*. Cambridge University Press, 1999.

MOHR, J. **Turing Machines and Computability**. Disponível em: <http://www.augustana.ab.ca/~mohrj/courses/1998.fall/csc110/lecture_notes/turing_machines.html>. Acesso em: 30 jan. 2008.

OSBORNE, M. "Brownian motion in the stockmarket". *Operations research*, v. 7, p. 145-173, 1959.

PINCUS, S. "Approximate Entropy as a Measure of System Complexity", *Proceedings of the National Academy of Sciences*, v. 88, n. 6, p. 2297-2301, 1991.

RICHMAN, J. MOORMAN, R. "Physiological time-series analysis using approximate entropy and sample entropy", *Am J Physio Heart Circ Physio*, v. 278, p. 2039-2049, 2000.

ROBERTS, H. V. "Stock market pattern and financial analysis: methodological suggestions". *Journal of Finance*, v. 14, n. 1, p. 1-10, Mar. 1959.

RUDICH, S.; WIGDERSON, A. (Eds.) *Computational Complexity Theory*. IAS/Park-City Mathematics Series, v. 10, American Mathematical Society, 2000.

SAMUELSON, P. “Proof that properly anticipated prices fluctuate randomly”. *Industrial Management Review*, v. 5, n. 1, p. 41-49, 1965.

SEILER, M.; ROM, W. “A Historical Analysis of Market Efficiency: Do Historical Returns Follow a Random Walk?” *Journal of Financial and Strategic Decisions*, v. 10, n. 2, p. 49-57, 1997.

SHMILOVICI, A.; ALON-BRIMER, Y.; HAUSER, S. “Using a stochastic complexity measure to check the efficient market hypothesis”. *Computational Economics*, v. 22, p. 273-284, 2003.

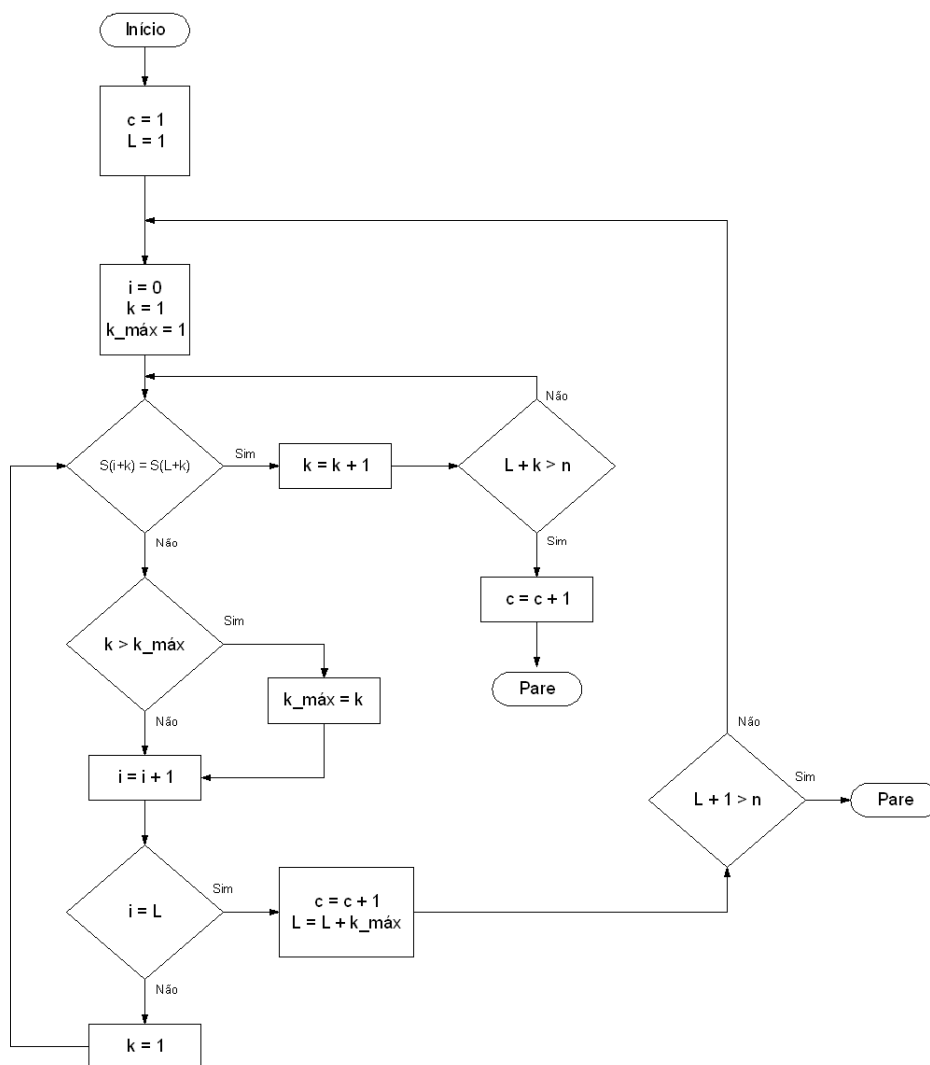
THEIL, H. *Economics and Information Theory*. Amsterdam, North Holland, 1967.

WOLFRAM, S. **A new kind of science – Online**. Wolfram Media, 2002. Disponível em <<http://www.wolframscience.com/nksonline/toc.html>>. Acesso em: 30 jan. 2008.

WORKING, H. “A random difference series for use in the analysis of time series”. *Journal of the American Statistical Association*, v. 29, p. 11-24, 1934.

Apêndice

Apêndice A – Figura 7: Fluxograma que representa o algoritmo utilizado para calcular a medida de complexidade, como descrito por Kaspar e Schuster (1987).



Apêndice B – As duas funções abaixo tomam uma seqüência de caracteres ‘s’ como argumento e retornam um número inteiro ‘c’ que representa a complexidade de ‘s’, como descrito em Kaspar e Schuster (1987). Esquerda: linguagem Python. Direita: linguagem VBA (*Visual Basic for Applications*).

Python

```
def complexidade(s):
    i = 0
    k = 1
    l = 1
    k_max = 1
    n = len(s) - 1
    c = 1
    while True:
        if s[i + k - 1] == s[l + k - 1]:
            k = k + 1
            if l + k >= n:
                c = c + 1
                break
        else:
            if k > k_max:
                k_max = k
            l = i + 1
            if i == l:
                c = c + 1
                l = l + k_max
                if l + 1 > n:
                    break
            else:
                i = 0
                k = 1
                k_max = 1
        else:
            k = 1
    return c
```

VBA

```
Function complexidade(s As String)
    i = 0
    k = 1
    l = 1
    k_max = 1
    n = Len(s)
    c = 1
    Do While 1 = 1
        If Mid(s, i + k, 1) = Mid(s, l + k, 1) Then
            k = k + 1
            If l + k > n Then
                c = c + 1
                Exit Do
            End If
        Else
            If k > k_max Then k_max = k
            i = i + 1
            If i = l Then
                c = c + 1
                l = l + k_max
                If l > n Then
                    Exit Do
                Else
                    i = 0
                    k = 1
                    k_max = 1
                End If
            Else
                k = 1
            End If
        End If
    Loop
    complexidade = c
End Function
```