Entropia de Shannon e Rényi Aplicadas no Reconhecimento de Padrões Para a Seleção de Ativos no Mercado Acionário Brasileiro

Shannon and Rényi Entropies Applied in the Pattern Recognition for the Selection of Assets in the Brazilian Stock Market

Alysson Ramos Artuso
Editora Positivo, Curitiba, Pr
alysson.artuso@gmail.com

Resumo: O mercado de capitais possui destacada importância no processo de desenvolvimento econômico. Uma de suas discussões são as estratégias de seleção de ativos face a Hipótese do Mercado Eficiente. Nesse artigo, propõe-se a aplicação da técnica de árvores de decisão baseadas no cálculo de entropia de Shannon e Rényi. Com base no método de múltiplos da análise fundamentalista de ações, foram selecionados 22 indicadores econômico-financeiros cobrindo índices de mercado, rentabilidade, liquidez e estrutura de capital. Foram avaliadas todas as empresas não-financeiras do período de 1999 a 2009 em função de seu retorno logarítmico, índice de Sharpe e alfa de Jensen. As regras de classificação construídas foram capazes de discriminar corretamente os grupos de ações com maior ou menor rentabilidade que o mercado com uma eficiência superior a 85%. Além disso, foram identificadas as variáveis preço de mercado, preço por valor contábil tangível, *dividend yield* e média de crescimento dos lucros como os atributos que melhor discriminam os grupos. Tanto as variáveis identificadas quanto os ativos selecionados indicam como mais adequadas para o mercado acionário brasileiro do período o investimento em valor em detrimento ao investimento em crescimento. Um estudo de carteira demonstrou a aplicabilidade dos modelos, com retornos anormais significativos especialmente para o uso da entropia de Shannon. Entretanto, houve problemas de diversificação das carteiras, ora com poucos ou nenhum ativo, ora com um número excessivo. Por fim, o modelo

Recebido em 12/03/2014 - Aceito em 24/09/2014.

apresentado pode ser expandido para outras técnicas de reconhecimento de padrões, como redes neurais artificiais e *support vector machine* (SVM).

Palavras-chave: árvore de decisão; entropia; hipótese de mercado eficiente; seleção de ativos; teoria da informação.

Abstract: The capital market has outstanding importance in the process of economic development. One of their discussions is the asset selection strategies against the Efficient Market Hypothesis. In this article, we propose the application of the technique of Decision Trees, based on the estimation of Shannon and Rényi entropies. It was selected 22 economic-financial indicators, covering market, profitability, liquidity and capital structure index, based on the multiple model from fundamental analysis of stocks. It was evaluated all non-financial companies for the period between 1999 and 2009 due to its logarithmic return, Sharpe ratio and Jensen's alpha. The rules of classification built were able to correctly discriminate groups of stocks with higher or lower profitability than the market with greater than 85% efficiency. In addition, it was identified the variables market price, price for tangible book value, dividend yield and average earnings growth as the attributes that best discriminate the groups. Both the identified variables and the selected assets indicate that the most appropriate is the value investment rather than the growth investment for the Brazilian stock market. One study has demonstrated the applicability of the portfolio models, with significant abnormal returns especially for the use of Shannon entropy. However, there were problems of portfolio diversification, sometimes with little or no assets, sometimes with an excessive number. In sum, the model presented can be extended to other pattern recognition techniques such as artificial neural networks and support vector machine (SVM).

Key words: asset selection; decision tree; efficient market hypothesis; entropy; information theory.

1 Introdução

Identificar e compreender o que influencia o desempenho das ações tem se tornado um desafio crescente para investidores e pesquisadores. Apesar de existirem pesquisas buscando variáveis explicativas dos retornos no mercado acionário, há espaço para aplicação de técnicas matemáticas-estatísticas refinadas, em especial, vinculadas ao reconhecimento de padrões.

Nessa lacuna, propõe-se a aplicação da técnica de árvores de decisão baseadas no cálculo de entropia, com a inovação da proposição de uso da família de entropias de Rényi e que pode ser estendida para outras técnicas de reconhecimento de padrões, como *support vector machine* (SVM) e redes neurais artificiais.

Com isso, busca-se identificar variáveis relevantes na predição do retorno de ativos acionários, compreender melhor a dinâmica de funcionamento do mercado acionário brasileiro e aplicar os modelos desenvolvidos para a seleção de ativos para um portfólio, bem como analisá-los em termos de rentabilidade.

Para cumprir tais objetivos, foram analisadas as empresas não-financeiras da Bolsa de Valores de São Paulo no período de 1999 a 2009. A partir de 22 indicadores de mercado e econômico-financeiros, foram construídas regras de classificação avaliadas pelo método de Lachenbruch e utilizadas para um estudo de carteira posterior. As carteiras foram avaliadas por três medidas de rentabilidade – o retorno logarítmico, o índice de Sharpe e o alfa de Jensen – com a execução de testes estatísticos paramétricos e não-paramétricos ao nível de significância de 5%.

2 Método de múltiplos

Os modelos de avaliação objetivam, em sua essência, avaliar uma empresa e projetar o comportamento futuro de seus ativos financeiros. Uma forma de se fazer isso, ligada à análise fundamentalista, é o método de múltiplos.

O método de múltiplos procura avaliar os reflexos do desempenho da empresa por meio de indicadores econômico-financeiros em comparação com o valor de mercado de suas ações para apontar se elas estão sub ou sobre-avaliadas em comparação a seus pares. Para tal, são necessários indicadores balizados em termos de lucro, valor

contábil ou receitas geradas.

Baseado na literatura da área [1-7], selecionou-se 22 indicadores amplamente aceitos que serão analisados neste trabalho, calculados por meio da plataforma Economatica. No entanto, vale ressaltar que não há concordância entre todos os indicadores na literatura da área, por isso a necessidade de explicitar a maneira de cálculo:

1. Preço de Mercado (PM) – multiplicação da quantidade de ações pela cotação do ativo em bolsa, resultando no valor de mercado da empresa.

2. *Dividend Yield* (DY) – indica a remuneração obtida em forma de proventos sobre o capital investido do acionista.

$$DY = \frac{\text{Dividendos Totais}}{\text{Preço de Mercado}} = \frac{\text{Dividendos por Ação}}{\text{Preço por Ação}}$$

3. Preço/Lucro (P/L) – relação entre o preço de mercado da ação e o lucro por ação do período.

$$P/L = \frac{\text{Preço de Mercado}}{\text{Lucro Líquido}} = \frac{\text{Preço por Ação}}{\text{Lucro por Ação}}$$

4. Preço/Vendas (P/V) – quociente entre o preço de mercado da ação e a receita líquida

5. Preço/Valor contábil (P/VC) – divisão entre o preço de mercado do ativo e o seu patrimônio líquido.

6. Preço/Valor contábil tangível (P/VCT) – similar ao P/VC, porém sem contabilizar no patrimônio líquido os ativos intangíveis. Antes de 2005, os ativos intangíveis informados pelas empresas brasileiras eram praticamente desprezíveis, o que diferencia ainda menos esse indicador do P/VC.

$$P/VCT = \frac{\text{Preço de Mercado}}{\text{Patrimônio Líquido - Ativos Intangóveis}}$$

7. Preço/Capital de Giro Líquido (P/CGL) – divisão entre o valor de mercado da empresa e seu capital de giro líquido, aqui entendido como ativo circulante menos dívida total, embora existam diferentes formas dele ser definido [4].

$$\text{P/CGL} = \frac{\text{Preço de Mercado}}{\text{Ativo Circulante - Dívida Total}}$$

8. Retorno sobre o Patrimônio Líquido (ROE) – razão do lucro líquido pelo patrimônio líquido da empresa.

9. Retorno sobre Ativos (ROA) – quociente entre o lucro líquido e o ativo total.

10. Retorno sobre o Capital (ROC) – similar ao ROE, considera a dívida financeira juntamente com o patrimônio líquido na divisão.

$$\text{ROC} = \frac{\text{Lucro Líquido}}{\text{Dívida Finamceira} + \text{Patrimônio Líquido}}$$

11. Margem líquida (ML) – quociente entre o lucro líquido e a receita líquida da empresa.

12. Média do Crescimento dos Lucros por Ação (MCL) – média aritmética do crescimento logarítmico dos lucros por ação da empresa nos últimos 5 anos [4].

13. Liquidez Corrente (LC) – quociente entre o ativo circulante e o passivo circulante, indicando a capacidade de pagamento em curto prazo da empresa.

14. Liquidez Seca (LS) – similar à LC, mas excluindo os estoques do ativo circulante.

15. Liquidez Geral (LG) – razão entre o ativo circulante e de longo prazo pelo passivo circulante e de longo prazo.

16. Liquidez Imediata (LI) – mostra quanto se dispõe imediatamente para liquidar as dívidas de curto prazo (contabilizando o dinheiro em caixa e disponível nas contas bancárias e nas aplicações de curtíssimo prazo), calculada por essa disponibilidade dividida pelo passivo circulante.

$$LI = \frac{\text{Caixa} + \text{Bancos} + \text{Aplicações de Curtíssimo Prazo}}{\text{Passivo Circulante}}$$

17. Grau de Endividamento (GE) – razão entre o passivo total e o patrimônio líquido, refletindo a dependência a terceiros.

18. Participação de Capitais de Terceiros Sobre Recursos Totais (PCTRT) – similar ao GE, divide o passivo total pelo passivo total mais o patrimônio líquido.

$$PCTRT = \frac{\text{Passivo Circulante} + \text{Exigível de Longo Prazo}}{\text{Passivo Circulante} + \text{Exigível de Longo Prazo} + \text{Valor Contábil Tangível}}$$

19. Garantia do Capital Próprio ao Capital de Terceiros (GCPCT) – é a relação de cada unidade monetária de capital próprio disponível para garantir uma unidade monetária de capital de terceiros.

$$GCPCT = \frac{\text{Patrimônio Líquido}}{\text{Passivo Circulante} + \text{Exigível de Longo Prazo}}$$

20. Composição do Endividamento (CP) – mostra a participação do endividamento de curto prazo ao dividir o passivo circulante pelo passivo total.

$$CP = \frac{\text{Passivo Circulante}}{\text{Passivo Circulante} + \text{Exigível de Longo Prazo}}$$

21. Liquidez Geral Modificada (LGM) – razão entre o capital de giro líquido e o realizável a longo prazo com o passivo circulante e o exigível a longo prazo.

$$LGM = \frac{\text{Capital de Giro Líquido}}{\text{Passivo Circulante} + \text{Passivo Exigível a Longo Prazo}}$$

22. Grau de Endividamento Modificado (GEM) – razão entre o passivo total e o patrimônio líquido tangível.

3 Risco e hipótese de mercado eficiente

Em finanças, o risco refere-se à probabilidade de se obter um retorno diferente do esperado, seja maior ou menor. Por isso a comparação de retornos, isoladamente, não é suficiente para avaliar uma estratégia de investimento, afinal um retorno maior pode ser proveniente de uma exposição maior ao risco.

Duas das maneiras mais comuns de se ajustar retornos ao risco são o uso do modelo CAPM [8], que possibilita o cálculo do alfa de Jensen, e o uso do desvio padrão como medida de risco, que dá origem ao índice de Sharpe [9].

Para seus cálculos, e mesmo para o uso de outros conceitos estatísticos, é frequente assumir que a série histórica de um ativo é contínua e se calcular o retorno a partir do logaritmo natural dos preços das ações:

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right)$$

em que r_t é o retorno da ação num momento t; P_t é o preço da ação num momento t; e P_{t-1} é o preço da ação num momento anterior a t.

O objetivo dessa transformação de cálculo para os retornos é satisfazer as hipóteses de gaussianeidade dos dados e de homogeneidade da variância dos dados necessária para posteriores testes estatísticos paramétricos. Essa transformação normalizante é conhecida na literatura de Finanças e a premissa de que os retornos logarítmicos possuem distribuição normal e os preços possuem distribuição lognormal ao final de qualquer intervalo finito de tempo é frequentemente assumida [10].

Assim, o alfa de Jensen é calculado por

$$\alpha_i = E(r_i) - [r_f + \beta_i(r_m - r_f)]$$

em que α_i é a medida de performance de Jensen; $E(r_i)$ é o retorno médio do investimento; r_f é o retorno livre de risco; β_i é o coeficiente beta estimado para o investimento i; e r_m é o retorno médio do mercado [11].

E o índice de Sharpe é obtido por

$$IS_i = \frac{E(r_i) - r_f}{\sigma_i}$$

em que IS_i é o índice de Sharpe da carteira de investimento i; $E(r_i)$ é o retorno médio da carteira de investimento; r_f é o retorno livre de risco e σ_i é o desvio padrão dos retornos da carteira de investimento. Com os retornos considerados sendo sempre os retornos logarítmicos [8].

A Hipótese de Mercado Eficiente (HME) está associada à ideia de que as séries de variações de preços dos ativos negociados no mercado de capitais comportam-se de maneira aleatória, não sendo possível discernir qualquer tendência nessas séries que permita a um investidor obter um retorno, ajustado para o risco, melhor que o do mercado.

Segundo Fama [12], um mercado é considerado eficiente se a posse de um conjunto de informações I_t, sobre o mesmo não altera o retorno esperado ao investir no mercado. Ou seja, $E(R_{i,t+1}|I_t) = E(R_{i,t+1})$, onde $E(R_{i,t+1}|I_t)$ é o retorno esperado do ativo i no período $t+1$, condicionado ao conjunto de informações I, disponível no período t, e $E(R_{i,t+1})$ é o retorno esperado não condicionado desse ativo. Em outras palavras, o preço dos ativos em qualquer momento é uma estimativa não viesada de toda a informação disponível.

Fama [13] caracterizou o conjunto I_t de três formas diferentes:

a) Quando I_t é composto por todas as informações de preços passados, há o mercado eficiente em sua forma fraca;

b) Quando I_t é composto por todas as informações públicas (receitas, balanços etc), há o mercado eficiente em sua forma semiforte;

c) Quando I_t é composto por todas as informações públicas e privadas (informações privilegiadas), há o mercado eficiente em sua forma forte.

Durante as décadas de 1960 e 1970, grande parte das publicações na área das finanças tentou comprovar a hipótese de eficiência informacional do mercado, com quase todas favorecendo a HME. Porém, nas décadas seguintes essa situação se inverteu devido, principalmente à evolução da informática, ao uso de bancos de dados maiores

e mais confiáveis e às técnicas estatísticas cada vez mais sofisticadas [14]. Tais indícios de ineficiência são ainda mais fortes nos mercados em desenvolvimento como o brasileiro [15-17].

4 Teoria da informação, entropia e informação mútua

Shannon [18] foi um pioneiro ao considerar a comunicação como um problema matemático rigorosamente embasado na estatística, criando um ramo da teoria da probabilidade e da estatística chamado Teoria da Informação. Apesar de ser originalmente desenvolvida para informações perdidas na compressão e transmissão de mensagens com ruídos em um canal de comunicação, sua aplicabilidade se expandiu para outros domínios da engenharia, informática, estatística e economia. Um de seus conceitos centrais é a entropia *(H)*, medida construída para quantificar a incerteza de uma transmissão sujeita a eventos de probabilidade $p(x_i)$ [19]:

$$H(X) = -\sum_{i=1}^{n} p(x_i) log(p(x_i))$$

Na área de reconhecimento de padrões, o interesse se volta para a Teoria da Informação pela sua capacidade de identificação de variáveis relevantes e utilização em métodos classificatórios. Isso se dá, a princípio, por meio de dois conceitos nomeados por Shannon: entropia e informação mútua. Nesse contexto, a entropia funciona como uma medida de incerteza de variáveis aleatórias isoladas ou combinadas. A informação mútua refere-se à dependência estocástica entre variáveis aleatórias, quantificando a informação comum entre elas [20].

A entropia é um termo que originalmente se refere a um conceito físico termodinâmico. Num primeiro momento, remete aos trabalhos do físico alemão Rudolf Clausius na segunda metade do século XIX, sendo definida como uma função de estado relacionada com a passagem de calor e a temperatura. Em 1877, com o desenvolvimento da Mecânica Estatística, uma definição alternativa de entropia foi estabelecida com base na probabilidade de arranjos de átomos ou moléculas para formarem microestados, sendo usualmente definida por $S = -Nk_B \sum_j p_j log(p_j)$, cuja semelhança com a medida de Shannon fez com que esta fosse também chamada de entro-

pia.

Chama-se a atenção para o fato de que a entropia *H(X)* não é função da variável aleatória *X*, mas sim da distribuição de probabilidade dessa variável. Em outras palavras, não dependente dos valores que *X* assume, mas das suas probabilidades.

Assim, sejam *X* e *Y* dois eventos quaisquer e $p(x_i, y_j)$ a probabilidade conjunta de ocorrência do primeiro e do segundo evento, então a entropia conjunta *H(X,Y)* é dada por:

$$H(X,Y) = H(Y,X) = -\sum_i \sum_j p(x_i, y_j) log\, p(x_i, y_j)$$

e a entropia condicional *H (Y | X)* é definida como:

$$H(X|Y) = -\sum_i p(i) \sum_j p(j|i) log\, p(j|i) = -\sum_j \sum_i p(j|i) log\, p(j|i)$$

de modo que *H (Y | X) = H (X, Y) - H (X)*.

Como *H (X, Y) ≤ H (X) + H (Y)*, tem-se *H (Y | X) ≤ H (Y)*, com igualdade apenas se *X* e *Y* forem independentes. O que se justifica pelo fato da entropia diminuir com o conhecimento de *X*, pois diminui a incerteza que existe relativamente a *Y*, a menos que as variáveis *X* e *Y* sejam independentes. Nesse caso, qualquer informação sobre *X* não diminui a entropia de *Y*.

Caso *X* possa diminuir o grau de incerteza sobre *Y*, pode-se considerar tal diminuição como um ganho de informação, representado pela informação mútua *I(Y,X)*:

$$I(Y,X) = H(Y) - H(Y|X)$$

de onde tem-se que I (Y, X) = H (Y) – H (Y | X) = H (X) – H (X | Y) = I (X, Y) e

$$I(X,Y) = \sum_i \sum_j p(x_i, y_j) log \left(\frac{p(x_i, y_j)}{p(x_i).p(y_j)} \right)$$

Com a aplicação do método de histograma, utilizando o histograma de frequências relativas [21] com a discretização das variáveis contínuas, a entropia e a Informa-

ção Mútua de Shannon podem ser estimadas a partir dos dados amostrais por:

$$\hat{H}(X) = -\sum_{i=1}^{N} \hat{f}_x(x_i) log \hat{f}_x(x_i)$$

$$\hat{I}(X,Y) = -\sum_{i=1}^{N}\sum_{j=1}^{N} \hat{f}_{xy}(x_i,y_i) log \frac{\hat{f}_{xy}(x_i,y_i)}{\hat{f}_x(x_i)\hat{f}_y(y_j)}$$

Uma extensão da entropia de Shannon foi alcançada por Rényi, que definiu uma família de entropias com base no parâmetro $\alpha \geq 0$ e $\alpha \neq 1$, da qual Shannon é um caso particular quando $\alpha \to 1$:

$$H_\alpha = \frac{1}{1-\alpha} log \left(\sum_{i=1}^{n} p^\alpha(x_i) \right)$$

Utilizando-se $\alpha = 2$, tem-se a entropia quadrática de Rényi:

$$H_\alpha = log \left(\sum_{i=1}^{n} p_i^2 \right)$$

Essa escolha faz com que a estimação da entropia de Rényi seja facilitada, podendo-se utilizar uma janela de Parzen que atribui um *kernel* sobre os dados da amostra e os soma com uma normalização adequada. Uma possibilidade é [19]:

$$\hat{f}_x(x) = \frac{1}{N\sigma} \sum_{i=1}^{N} \varkappa \left(\frac{x - x_i}{\sigma} \right)$$

e a entropia quadrática, utilizando um kernel gaussiano, pode ser estimada por [19]:

$$\hat{H}_2(X) = -log \left(\frac{1}{N^2} \sum_{i=1}^{N}\sum_{j=1}^{N} G_{\sigma\sqrt{2}}(x_j - x_i) \right)$$

O parâmetro σ deve ser selecionado pelo usuário, normalmente com base no método de *cross validation* ou pela regra de Silverman para N observações e d dimensões [22]:

$$\sigma_{opt} = \sigma_X (4N^{-1}(2d+1)^{-1})^{\frac{1}{d+4}}$$

A princípio a informação mútua de Rényi não pode ser expressa em termos da entropia, como foi feito pela eq. 2.92 para a entropia de Shannon. No entanto, se for usada a divergência de Cauchy-Schwarz para se definir a informação mútua, é possível estabelecer uma relação de forma que [20]:

$$I_{CS}(X,Y) = H_2(f_{XY} \times f_X f_Y) - \frac{1}{2}H_2(f_{XY}) - \frac{1}{2}H_2(f_X f_Y)$$

que pode ser estimada, de maneira análoga ao feito para o H_2(X), por [23]:

$$\hat{I}_{CS}(X,Y) = \frac{\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G_{\sqrt{2\sigma}}(x(i)-x(j)) \times G_{\sqrt{2\sigma}}(y(i)-y(j))\right)\left(\hat{V}_x \hat{V}_y\right)}{\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\left(\sum_{j=1}^{N} G_{\sqrt{2\sigma}}(x(i)-x(j))\right)\left(\frac{1}{N}\left(\sum_{j=1}^{N} G_{\sqrt{2\sigma}}(y(i)-y(j))\right)\right)\right)^2}$$

com $\hat{V}_k = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G_{\sqrt{2\sigma}}(k(i)-k(j))$,para $k = x, y$.

5 Árvores de decisão

Árvores de decisão são modelos de mineração de dados cuja estrutura apresenta-se no formato de uma árvore. Cada nó interno da árvore indica um teste sobre um atributo, cada ramo representa um resultado do teste, e os nós terminais (folhas) correspondem a classes ou distribuições de classes. A profundidade de uma árvore é definida pela maior distância entre uma folha e a raiz (primeiro nó). Com isso, tem-se uma técnica que constrói regras de classificação passíveis de avaliação, interpretação e posterior aplicação.

Algumas das vantagens apresentadas pelas árvores de decisão são sua flexibilidade, pois não assumem uma distribuição única dos dados, sendo métodos não-paramétricos; robustez, uma vez a seleção interna de características produz árvores que tendem a ser bastante robustas mesmo com a adição de variáveis irrelevantes; interpretabilidade, já que todas as decisões são baseadas nos valores (conhecidos) dos atributos usados para descrever o problema; e velocidade, pois a maioria dos algoritmos constrói rapidamente as árvores de decisão [24].

Em geral, o procedimento de uma árvore de decisão consiste em apresentar um

conjunto de dados ao nó raiz da árvore e avaliá-lo segundo um teste lógico. Dependendo do resultado, a árvore ramifica-se para um dos nós descendentes e este procedimento é repetido até que uma folha conceda a classificação dos dados. A figura 1 exemplifica a estrutura de uma árvore de decisão.

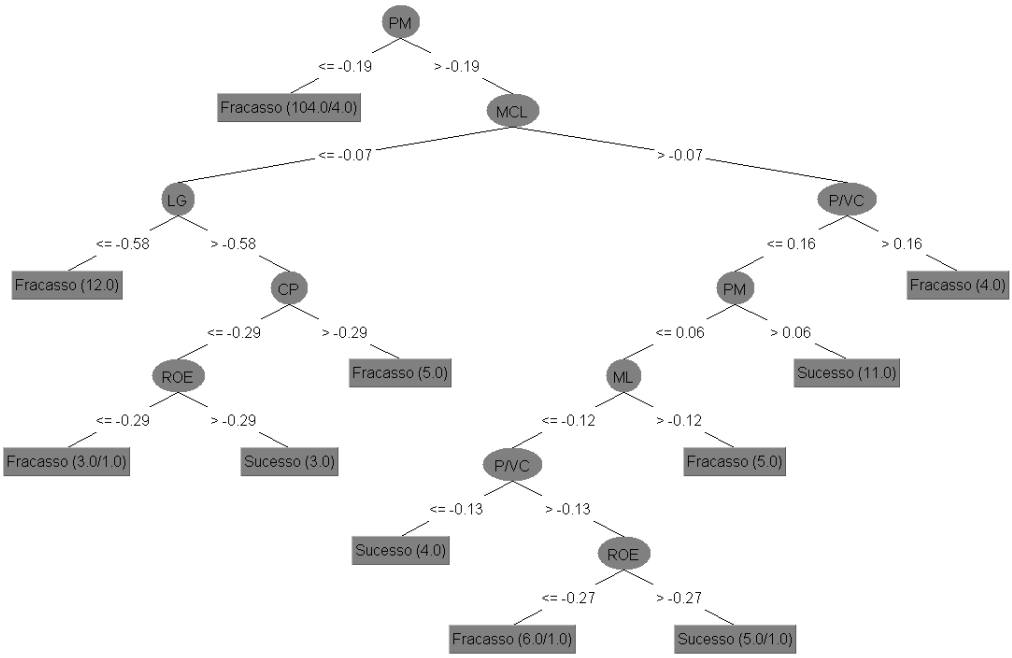

Figura 1. Exemplo de árvore de decisão.

As árvores de decisão são construídas usando um algoritmo de partição recursiva. Uma das possibilidades é este algoritmo construir uma árvore por divisões recursivas binárias que começa no nó raiz e desce até os nós folhas. Nesse caso, têm-se dois fatores principais no algoritmo de partição: a forma de selecionar uma divisão para cada nó intermediário (crescimento) e uma regra para determinar quando um nó é terminal (poda).

Algoritmos de indução de árvores de decisão foram desenvolvidos e refinados ao longo de vários anos por John Ross Quinlan. Sua contribuição inicial foi o algoritmo ID3. Várias melhorias foram realizadas nesse algoritmo, culminando no surgimento do algoritmo C4.5 [25], muito utilizado em aplicações práticas e pesquisas acadêmi-

cas.

Em ambos, a escolha do atributo que geram as ramificações é feita a partir de uma medida conhecida como ganho de informação, que nada mais é do que a Informação Mútua de Shannon entre a variável avaliada na iteração i e o grupo desfecho. O atributo que proporcionar o maior ganho de informação (princípio da minimização da entropia) é selecionado como atributo teste do nó corrente.

Para aprimorar a seleção pelo ganho de informação, [25] propôs o uso conjunto da Razão de Ganho (*Gain Ratio*), que consiste no ganho de informação relativo como critério de avaliação:

$$Razão\ de\ Ganho = \frac{Ganho\ de\ Informação}{entropia(nó)}$$

Pela equação anterior, é possível perceber que a razão não é definida quando o denominador é igual a zero. Além disso, a razão de ganho favorece atributos cujo denominador – a entropia – possui valor pequeno. Em [26], é sugerido que a avaliação pela razão de ganho seja realizada em duas etapas. Na primeira, é calculado o ganho de informação para todos os atributos. Após isso, consideram-se apenas aqueles que obtiveram um ganho de informação acima da média, e então se escolhe aquele que apresenta a melhor razão de ganho. Com esse procedimento, Quinlan mostrou que a Razão de Ganho supera o ganho de informação tanto em termos de acurácia, quanto em termos de complexidade das árvores de decisão geradas, sendo esse o método implementado no algoritmo C4.5 [25].

Como o algoritmo divide recursivamente o conjunto de dados de treinamento original, as divisões estão sendo avaliadas com amostras cada vez menores. Isto significa que as estimativas de erro têm menos confiabilidade à medida que a árvore cresce. Com intuito de minimizar este problema e evitar o *overfitting* dos dados de treinamento, usam-se métodos conhecidos como podagem.

Uma forma de podagem, após ter a árvore completa, é para cada nó interno da árvore, o algoritmo calcular a taxa de erro caso a sub-árvore abaixo desse nó seja podada. Em seguida, é calculada a taxa de erro caso não haja a poda. Se a diferença entre essas duas taxas de erro for menor que um valor pré-estabelecido, a árvore é podada. Caso

contrário, não ocorre a poda. Esse processo se repete progressivamente, gerando um conjunto de árvores podadas. Por fim, para cada uma delas é calculada a acurácia na classificação de um conjunto de dados. No algoritmo C4.5 é feita uma busca na árvore, de baixo para cima, transformando em nós folha os nós intermediários sempre que o valor do erro do nó for inferior à soma dos erros de seus descendentes. O erro é definido como a razão k de classificações erradas pelas n observações do conjunto de treinamento. Essa técnica é conhecida por error *based pruning*.

Para avaliar funções de classificação, como as árvores de decisão, pode ser utilizada a abordagem de Lachenbruch. Ela consiste num método de *cross validation* na qual uma observação é excluída, constrói-se o modelo de classificação para as demais e classifica-se a observação retirada anteriormente, repetindo o método para as n observações.

6 Pesquisas similares

Com o avanço das pesquisas na área, vários estilos de investimentos foram identificados e estudados, sobretudo a partir de 1970 nos Estados Unidos. Desses trabalhos, a Análise Fundamentalista se dividiu em recomendações de, basicamente, duas correntes: o investimento em valor (*value*) e o investimento em crescimento (*growth*).

De forma resumida, o investimento em valor se concentra em empresas grandes e consolidadas no mercado, com taxas de crescimento estável, maior distribuição de dividendos e baixos múltiplos P/L e P/VC. A oposição a esses conceitos são as ações de crescimento (*growth stocks*) com altas taxas de crescimento, distribuição pequena ou inexistente de dividendos e elevados múltiplos P/L e P/VC, que refletem uma grande expectativa sobre os lucros futuros.

Sharpe [27] identificou que tamanho, *value* e *growth* eram características capazes de explicar o retorno obtido por portfólios de ação. Analisando o desempenho de fundos de investimento nos Estados Unidos, no período de 1985 até 1989, o autor verificou que, praticamente, duas variáveis eram capazes de diferenciar os retornos dos fundos: *value/growth* e *small/large*.

A partir do estudo de Sharpe [27] pode-se definir o investimento em valor como aquele no qual se procura empresas grandes e consolidadas, cujas ações estejam sub-

valorizadas em relação aos ativos que já possui. O investimento em crescimento também consiste na seleção de ações subvalorizadas, mas do ponto de vista do potencial de crescimento de seus lucros.

No Brasil, Ramos, Picanço e Costa Jr [28] compararam, para o período de 1988 até 1994, o retorno e risco de ações de valor e de crescimento concluindo que as ações denominadas value, devido ao baixo índice P/VC, apresentaram rentabilidade média superior ao portfólio de ações composto pelas ações denominadas *growth*. Constataram também que o coeficiente beta do modelo CAPM do portfólio *value* é, em média, um pouco menor do que o coeficiente beta do portfólio *growth*.

Costa Jr e Neves [29] realizaram um estudo no mercado brasileiro cujo objetivo principal foi verificar a influência das variáveis fundamentalistas nas rentabilidades médias das ações. O período de análise compreende de 1987 a 1996 e as variáveis explicativas utilizadas nesse estudo foram: Preço/Lucro, Preço de Mercado, Preço/Valor Contábil e o beta do modelo CAPM.

Os resultados, obtidos no estudo de Costa Jr e Neves [29], mostraram um relacionamento negativo entre a rentabilidade média das carteiras e as variáveis P/L, P/VC e PM. Contudo, embora as variáveis fundamentalistas analisadas em [29] tenham influência nas explicações das variações das rentabilidades médias das ações, o beta é fortemente representativo, sendo a variável que mais se destacou nessa explicação. Assim, baseados nos testes realizados, os autores afirmaram que o CAPM está mal especificado, devido à possibilidade de inclusão de outros fatores no comportamento dos retornos dos ativos, além do beta.

Santos [30], também, analisou a rentabilidade das ações de valor em comparação com as ações de crescimento no mercado brasileiro de 1989 a 2008. Como critério para a separação dos grupos, ele utilizou os quartis extremos dos múltiplos P/L e P/VC, que, quando baixos, caracterizariam empresas de valor. Foram analisados diversos contextos (mercados de alta/baixa, pré/pós plano Real, contração/expansão econômica) e em todos eles a carteira composta pelos ativos com os dois múltiplos baixos teve retorno significativamente superior (ao nível de 10%) à carteira de crescimento.

Na proposição de modelos de reconhecimento de padrões, Tavares [31] analisou a

aplicação da Análise Discriminante por meio de 23 indicadores no mercado acionário brasileiro entre 2005 e 2007. Dentre esses indicadores 14 são diferentes dos propostos na presente pesquisa: prazo de pagamentos a fornecedores, prazo de recebimentos, prazo de estocagem, giro do ativo, giro do patrimônio líquido, dívida financeira por ativo total, dívida financeira por patrimônio líquido, lucro operacional por dívida financeira, margem bruta, margem operacional, grau de alavancagem operacional, EBITDA por ação, lucro por ação e patrimônio líquido por ação. Os outros 9 são iguais: LI, LC, LS, PCTRT, GE, CP, ML, ROA e ROE.

As empresas foram separadas em grupo de Vencedoras e Perdedoras de acordo com o retorno de suas ações, com 50% em cada grupo e foram utilizadas funções discriminantes baseadas na regra do qui-quadrado mínimo, na função linear discriminante de Fisher e no modelo *logit* (Regressão Logística). O resultado aponta para a aplicabilidade da Análise Discriminante para a seleção de ativos, já que o sucesso de alocação dos ativos foi em torno de 60% e 70% nos três modelos, seja com o uso de todas as variáveis ou de somente as de maior poder discriminante, com os melhores resultados sendo dados pelo modelo *logit*. As variáveis que se mostraram significativas foram poucas, o que concede grande poder de síntese aos modelos, porém instáveis a cada ano, a saber, margem bruta (2005); prazo de recebimento, dívida financeira por patrimônio líquido e CP (2006); grau de alavancagem operacional (2007). Não foi realizado um estudo de carteira, nem aplicada as funções discriminantes de um ano para o período subsequente de forma a analisar a regra de alocação de maneira como seria possível de se proceder na prática.

Artuso e Chaves Neto [32] fizeram uma avaliação dos múltiplos introduzida por Benjamin Graham, considerado um dos criadores da análise fundamentalista, mostrando que a filtragem passiva poder ser adaptada ao mercado brasileiro e gerar retornos ajustados superiores ao Ibovespa. A filtragem passiva é uma forma rudimentar de reconhecimento de padrões, também baseada em regras *if-then* como as árvores de decisão. No estudo foram consideradas nove variáveis, com somente uma não presente nesta pesquisa, o número de anos com lucros em declínio de uma empresa. Todas as demais variáveis estão entre as 22 deste trabalho: P/L, DY, P/VCT, P/CGL, LGM, GEM, MCL E LC. Nos estudos de carteira realizados, a estratégia proposta alcan-

çou rentabilidades significativamente superiores ao mercado, com alfa de Jensen de 26,26% ao ano.

O uso da entropia no mercado acionário brasileiro, de maneira próxima ao proposto neste trabalho, foi observado em Rocha, Hein e Kroenke [33], no qual foi utilizada entropia de Shannon para avaliar os indicadores econômico-financeiros de empresas participantes dos níveis de governança corporativa da Bovespa entre os períodos de 2005 a 2009 do setor de materiais básicos. Entre os 14 indicadores de liquidez, endividamento, atividade e rentabilidade levantados, o que apresentou maior quantidade de informação foi o Retorno sobre Patrimônio Líquido (ROE) do ano de 2008. Devido aos objetivos da pesquisa, não houve a inclusão de múltiplos de mercado para análise.

7 Material e métodos

A amostragem dos dados deste estudo consiste em todas as empresas não-financeiras listadas na Bovespa negociadas no ano entre 1999 e 2009. Os dados foram levantados da plataforma Economatica. Após a retirada dos *outliers* (feita através do método das duas primeiras componentes principais padronizadas, quando os escores apresentados fugirem ao intervalo de mais ou menos dois desvios padrões) e das empresas que não continham todos os dados disponíveis, houve uma média de 200 ativos a cada ano. Foram entre 3 a 5 empresas caracterizadas como *outliers* anualmente, não caracterizando um grupo significativo a ponto de merecer uma descrição pormenorizada. Contudo, esse descarte é útil tanto para maior eficiência dos modelos de classificação, quanto para descartar dados que foram digitados de forma incorreta.

Para os 200 ativos *inliers*, foram considerados os balanços anuais e as cotações diárias de fechamento para os cálculos relativos a preços. A taxa Selic foi utilizada como taxa livre de risco e o alfa de Jensen foi estimado por meio do modelo CAPM, conforme equação citada na Seção 3 – Risco e Hipótese de Mercado Eficiente. Na mesma seção encontra-se a equação utilizada para o Índice de Sharpe, na qual a volatilidade α_i é o desvio padrão dos retornos do ativo ou da carteira de investimento.

Como os balanços das empresas são divulgados até março do ano posterior ao qual se refere, foi tomado o último dia útil do mês de março como data para o levantamento

de dados e cotações. Assim, o modelo de 2009, por exemplo, foi construído a partir dos dados disponíveis em 31/03/2010 e seus ativos foram classificados de acordo com a rentabilidade até 31/03/2011.

O arquivo de dados com as informações contábeis dos ativos foi organizado em planilhas conforme mostra a tabela 1, que representa um trecho ilustrativo dos dados para o ano de 2001. Como se observa no exemplo, as variáveis levantadas, detalhadas na Seção 2 – Método de Múltiplos, foram padronizadas.

Tabela 1. Exemplo da estrutura de dados.

	PM	DY	P/L	P/V	P/VC	P/VCT	P/CGL	ROE	ROA	ROC	ML
FJTA4	-0,28	0,88	-0,39	-0,18	-0,28	-0,28	-0,36	-0,13	0,39	0,08	-0,14
MAGS5	-0,26	0,44	-0,27	-0,17	-0,25	-0,25	-0,35	-0,18	-0,01	-0,17	-0,19
CNFB4	-0,25	-0,52	2,65	-0,17	-0,21	-0,21	-0,35	-0,10	0,29	0,13	-0,17
POMO4	-0,25	-0,13	0,32	-0,18	-0,11	-0,11	-0,31	-0,11	-0,21	-0,26	-0,23
LAME4	-0,24	0,07	-0,38	-0,19	-0,08	-0,08	-0,37	-0,06	-0,31	-0,26	-0,23
ETER3	-0,24	4,45	-0,37	-0,08	-0,09	-0,09	-0,26	-0,13	0,50	0,08	-0,12
RPSA4	-0,22	0,28	-0,25	-0,15	-0,25	-0,25	-0,28	-0,16	0,09	-0,17	-0,14
UNIP6	-0,22	0,66	0,01	-0,16	-0,21	-0,21	-0,29	-0,13	0,29	0,03	-0,16
DURA4	-0,17	-0,13	-0,38	-0,11	-0,18	-0,18	-0,28	-0,21	-0,31	-0,41	-0,20
TRPL4	-0,16	-0,23	-0,39	-0,02	-0,30	-0,30	-0,29	-0,24	-0,51	-0,51	-0,11
FJTA4	-0,07	1,43	0,78	1,27	-0,68	-0,24	-0,61	0,37	0,70	1,21	-0,12
MAGS5	-0,08	1,27	0,77	1,29	1,16	-0,24	-0,66	0,48	1,07	1,95	-0,13
CNFB4	-0,08	0,27	0,02	0,32	0,32	-0,20	-0,24	-0,12	0,79	0,61	-0,09
POMO4	-0,08	-0,05	0,08	0,03	-0,12	-0,11	0,14	-0,31	0,79	-0,07	0,01
LAME4	-0,09	0,22	0,26	0,00	0,51	0,00	0,30	-0,37	0,01	0,45	0,13
ETER3	-0,08	1,79	1,58	1,32	2,47	-0,24	-0,72	0,63	0,05	1,84	-0,13
RPSA4	-0,09	1,16	1,18	-0,15	1,75	-0,22	-0,47	0,13	-1,10	0,06	-0,11
UNIP6	-0,10	0,44	0,47	0,06	0,88	-0,23	-0,52	0,19	-0,04	0,41	-0,12
DURA4	-0,07	1,12	0,95	0,07	1,43	-0,23	-0,49	0,16	-0,59	0,48	-0,11
TRPL4	-0,07	0,87	1,18	0,01	1,29	-0,25	-0,91	1,70	-1,05	0,52	-0,14

Foram incluídas no grupo Sucesso as ações que apresentaram, simultaneamente, rentabilidade acima do mercado nas três medidas usadas: retorno logarítmico, índice de Sharpe e alfa de Jensen. Portanto, foi calculado o retorno logarítmico, o índice de Sharpe e o alfa de Jensen de cada um dos 200 ativos e encontradas as respectivas médias – entendidas como a média do mercado para esses três índices de rentabilidade. Para uma ação ser classificada como Sucesso, ela precisava ter as três medidas – retorno logarítmico, o índice de Sharpe e o alfa de Jensen – com valores superiores à média dos 200 ativos. Caso contrário, o papel foi alocado no grupo Fracasso.

Na construção das árvores de decisão, foi utilizado o algoritmo C4.5 [25] implementado no pacote WEKA 3.4.11. O algoritmo original (J58) faz uso da entropia

de Shannon na elaboração da árvore e um algoritmo para uso da entropia de Rényi foi desenvolvido por Lima, Assis e Souza [34] e gentilmente cedido para aplicação na presente pesquisa. O modelo foi compilado com $\alpha = 2$ para a entropia de Rényi (quadrática), $\alpha = 1$ (entropia de Shannon), com pós-podagem pelo método *error based pruning* e utilizando a regra de Silvermann para a estimação da janela de Parzen no cálculo da informação mútua.

Para se testar as regras de classificação fora dos dados em que foram construídas, foi utilizado um estudo de carteira com os ativos selecionados distribuídos com o mesmo peso dentro da carteira e o tempo de manutenção dos portfólios foi de 1, 2, 3 e 5 anos com os pesos dos ativos permanecendo o inicial de modo a analisar a rentabilidade do investimento em diferentes janelas temporais, uma medida típica de estudos semelhantes [15, 30, 31]. A seguir os três índices de rentabilidade – já citados tiveram suas significâncias testadas ao nível de 5%. Quando as premissas foram satisfeitas, foi utilizado o teste t, caso contrário, a comparação das médias foi feita pelo teste de Wilcoxon-Mann-Whitney.

A abordagem de carteira pode trazer indícios acerca de eficiência do mercado, caso encontre resultados significativamente superiores. Todavia, muitos são os cuidados com essas conclusões. O primeiro é que um teste de eficiência de mercado tanto testa a eficiência de mercado como testa a eficácia do modelo utilizado para calcular os retornos esperados. Ou seja, quando surge uma evidência de retornos excedentes em um teste de eficiência de mercado, isto tanto pode ser um indicativo de que os mercados são ineficientes quanto pode indicar que o modelo utilizado para calcular os retornos esperados está equivocado, ou ambos. Embora essa situação possa parecer um dilema inescapável, se as conclusões de um estudo forem confirmadas por vários modelos e para vários períodos, é muito mais provável que os resultados estejam sendo gerados devido a verdadeiras ineficiências do mercado do que a má especificação do modelo.

Os próprios modelos utilizados para calcular os retornos esperados possuem vieses dos quais o pesquisador precisa estar consciente. O uso do retorno logarítmico tem um viés em direção a estratégias de alto risco, o índice de Sharpe possui tendência de penalizar carteiras que não sejam bem diversificadas, o modelo CAPM tende

a subestimar o risco de ações menos negociadas e assim por diante [3, 29]. Por isso, não deve ser usado apenas uma medida isolada de rentabilidade para se testar uma estratégia e, como consequência, a eficiência dos mercados. De certa maneira, as três medidas aqui utilizadas são complementares, visto que possuem diferentes vieses, justificando suas escolhas.

8 Resultados e discussão

As árvores de decisão são um método de reconhecimento de padrões que permite identificar variáveis explicativas e utilizá-las para a classificação de ativos. No caso, elas foram avaliadas em termos da entropia de Rényi ou de Shannon para a construção dos modelos de classificação. A tabela 2, contém as variáveis incluídas nos modelos a cada ano. A seguir, também, é mostrado um exemplo de árvore de decisão utilizando a entropia de Shannon (o exemplo é do ano de 2007).

Os anos de 2005, 2006 e 2009 para a entropia de Rényi e 2008 para a entropia de Shannon não tiveram variáveis incluídas na construção de suas regras de classificação, tendo todos os ativos sendo classificados como Fracasso. Como o grupo Sucesso foi definido pelos ativos que possuíam as três medidas de rentabilidade acima da média, conforme explicitado na Seção 7 – Material e métodos, e não por uma divisão igualitária 50%-50%, a quantidade de ativos no grupo Sucesso era sempre menor que no grupo Fracasso. Nos anos citados, o algoritmo de construção das árvores não foi capaz de encontrar variáveis que discriminassem os grupos com um erro menor do que classificar todos como de rentabilidade abaixo do mercado.

As árvores de decisão são construídas a partir dos grupos de sucesso e fracasso. Como o mercado acionário é dinâmico, ou seja, como o conjunto de empresas que apresentam alta rentabilidade não é muito estável ao longo do tempo, poderia se esperar uma alternância nas variáveis que diferenciam esses grupos nas regras de seleção. Contudo, quatro variáveis se mostraram recorrentes ao longo dos anos: Preço de Mercado (PM), Média de Crescimento dos Lucros (MCL), *Dividend Yield* (DY) e Preço por Valor Contábil (P/VC).

As variáveis P/VC e P/VCT estão intimamente ligadas, afinal a diferença é o cômputo, ou não, do Ativo Intangível. Entretanto, nos balanços das empresas brasileiras

Tabela 2. Variáveis presentes nas árvores de decisão.

Ano	Rényi	Shannon
1999	PM, LG, LC	PM
2000	PM, MCL, P/VCT, LG, CP, ROE, ML, GE	PM, MCL, P/VCT, LG, CP, ROE, ML
2001	PM, ML, ROE, GEM, MCL	PM, ML, ROE, GEM, MCL
2002	GEM, PM, CP, LG, MCL	PM, GEM, CP, P/L, DY
2003	PM, P/V, LG	PM, ML, P/V, ROC, MCL, PCTRT, P/CGL, ML
2004	PM, GEM, DY	DY, P/CGL, MCL, PCTRT, P/VCT
2005	-	PM, CP, ROC, P/VC, MCL
2006	-	PM, P/VCT, CP, P/L, MCL, PCTRT, DY, LC
2007	PM, LC, LS, ROC, MCL, P/VCT, P/CGL, LGM, P/L, DY, P/V	PM, LC, ROC, P/VCT, LGM, P/L, DY, P/V
2008	P/VCT, ROC, DY, P/V, ROE, GEM	-
2009	-	PM, P/VC, ROC, P/V, P/L, PCTRT

os lançamentos de Ativo Intangível são, geralmente, pequenos em comparação com o Ativo Total, o que causa uma correlação bastante grande entre as duas variáveis, em especial no período anterior a 2005. Ainda assim, é esclarecedor perceber que as soluções das árvores de decisão são mais conservadoras ao valorizar somente o patrimônio líquido tangível das empresas, deixando de lado os valores intangíveis, de mensuração mais complexa e passíveis de causarem distorções no múltiplo P/VC.

No trabalho de Rocha, Hien e Kroenke [33], que também utilizaram a entropia para detectar a relevância de índices econômico-financeiros, não foram usados indicadores de mercado, como o PM, P/VCT e DY. O destaque do estudo são os indicadores de rentabilidade, em especial o ROE, como importante fator para a análise de empresas.

Em conjunto, as quatro medidas de destaque priorizam grandes empresas, negociadas com um baixo múltiplo P/VCT, de crescimento de destaque e boa distribuição de lucros.

Uma possível explicação para esse fato passa pelo período analisado. Entre 1999 e 2011, o Brasil e o mundo passaram por diversos momentos de inquietação dos mercados financeiros, que provocaram rápidos movimentos e grandes quedas nas bolsas

```
PM <= -0.1921: Fracasso (107.0/9.0)
PM > -0.1921
|   LC <= 0.5116
|   |   ROC <= 0.2078
|   |   |   P/VCT <= 0.0471
|   |   |   |   LGM <= -0.8082: Fracasso (4.0)
|   |   |   |   LGM > -0.8082
|   |   |   |   |   PM <= 0.1198
|   |   |   |   |   |   P/L <= -0.2549: Sucesso (4.0)
|   |   |   |   |   |   P/L > -0.2549
|   |   |   |   |   |   |   DY <= 0.3924
|   |   |   |   |   |   |   |   P/V <= -0.1059: Fracasso (3.0/1.0)
|   |   |   |   |   |   |   |   P/V > -0.1059: Sucesso (3.0)
|   |   |   |   |   |   |   DY > 0.3924: Fracasso (5.0)
|   |   |   |   |   PM > 0.1198: Sucesso (10.0)
|   |   |   P/VCT > 0.0471
|   |   |   |   P/CGL <= 1.6154: Fracasso (13.0)
|   |   |   |   P/CGL > 1.6154: Sucesso (3.0/1.0)
|   |   ROC > 0.2078: Sucesso (6.0)
|   LC > 0.5116: Fracasso (15.0)

Número de folhas:     11
Tamanho da árvore:    21
```

Figura 2. Árvore para detalhar a estrutura de dados.

de valores. Como citado, houve o estouro da "Bolha da Nasdaq" em 2000, a crise Argentina e o 11 de setembro, em 2001, o risco político da eleição do presidente Lula em 2002 e a crise do subprime em 2008, além da crise da dívida pública da Zona do Euro em 2010 e 2011. Lembrando que as regras de classificação são construídas a partir da variação da série temporal das cotações, portanto, mesmo eventos ocorridos no início de 2011 podem ter impactado sobre a construção dos modelos discriminantes.

É costumeiro no mercado acionário que, em períodos de alta volatilidade, os investidores migrem de ativos mais arriscados para os entendidos como de menor risco. No campo acionário, ativos de grandes companhias e que pagam altos dividendos são entendidos como de baixo risco, afinal, do ponto de vista desses investidores, o tamanho e a distribuição dos lucros da empresa minimizam o risco da perda de capital ao se aplicar naquele papel.

Era de se esperar que o índice P/VC, ou o P/VCT, dada sua proximidade, fosse um dos mais relevantes na construção dos modelos de classificação dos papéis. Afinal, ele é tido como a principal variável a influenciar os retornos de ações em dezenas de

trabalhos científicos, [28-30] para ficar somente entre os já citados.

Pela abordagem de Lachenbruch, também as árvores de decisão podem ser avaliadas, com a tabela 3 trazendo o índice médio de acerto dos modelos.

Tabela 3. Média do índice de acerto das árvores de decisão.

	Rényi			Shannon		
Grupo	Percentual correto	Class como Sucesso	Class como Fracasso	Percentual correto	Class como Sucesso	Class como Fracasso
Sucesso	31,99%	8,55	26,45	49,22%	17,00	18,00
Fracasso	99,21%	1,27	145,00	96,49%	4,91	141,36
Total	85,27%	9,82	171,45	87,75%	21,91	159,36

Na avaliação geral, as árvores de decisão com base na entropia quadrática de Rényi tiveram um índice de acerto de 85,27%, enquanto com a entropia de Shannon 87,75% dos ativos foram classificados corretamente. Esses resultados são superiores ao melhor resultado encontrado por Tavares [31] em estudo similar (71,2%), provavelmente em razão da escolha de variáveis e do método baseado na entropia.

Ponto que chama a atenção é a baixa classificação de empresas do grupo Sucesso como Fracasso com a utilização da entropia de Rényi. Porém, a quantidade de empresas classificadas como Sucesso é pequena, evidenciando certo grau de conservadorismo nas regras de decisão. Nesse quesito, o destaque é do modelo com a entropia de Shannon, que classifica corretamente 49,22% das empresas de maior rentabilidade, o que dá uma média de 17 empresas no grupo Sucesso por ano.

Uma forma de se utilizar na prática essas regras de classificação é aplicá-las no ano seguinte aos dados de sua construção. Os portfólios formados dessa maneira são avaliados por um estudo de carteira, em busca de retornos que superem o mercado. As tabelas 4 e 5 expõem esses resultados e indicam os valores estatisticamente significativos.

Os retornos logarítmicos baseados na estratégia de árvores de decisão com entropia de Rényi foram muito próximos aos do mercado. O modelo foi potencialmente prejudicado pelos anos em que todos os ativos foram classificados como Fracasso e, portanto, não foram formadas carteiras. Ainda assim, em todos os anos em que houve a elaboração portfólios, o modelo obteve valores de rentabilidade acima do mercado.

Tabela 4. Média do retorno, índice de Sharpe e alfa de Jensen para diferentes períodos – Rényi.

	Média dos Retornos da Carteira (a.p)	**Média dos Retornos do Ibovespa (a.p)**	**Índice de Sharpe da carteira**	**Índice de Sharpe do Ibovespa**	**Média do Alfa de Jensen (a.p.)**
1 ano	17,81%	16,05%	0,0328	0,0140	*7,80%
2 anos	35,61%	36,24%	0,0327	0,0155	14,05%
3 anos	46,91%	56,88%	0,0245	0,0136	14,88%
5 anos	92,90%	101,78%	0,0272	0,0149	*31,47%

* Valores estatisticamente positivos/superiores ao Ibovespa no teste t de Student.

Tabela 5. Média do retorno, índice de Sharpe e alfa de Jensen para diferentes períodos – Shannon.

	Média dos Retornos da Carteira (a.p)	**Média dos Retornos do Ibovespa (a.p)**	**Índice de Sharpe da carteira**	**Índice de Sharpe do Ibovespa**	**Média do Alfa de Jensen (a.p.)**
1 ano	22,67%	16,05%	0,0504	0,0140	*9,15%
2 anos	48,06%	36,24%	**0,0473	0,0155	*17,81%
3 anos	65,66%	56,88%	0,0355	0,0136	*19,85%
5 anos	124,12%	101,78%	*0,0369	0,0149	*41,42%

* Valores estatisticamente positivos/superiores ao Ibovespa no teste t de Student.
** Valores estatisticamente positivos/superiores ao Ibovespa no Wilcoxon-Mann-Whitney.

Por outro lado, a despeito do retorno próximo, o alfa de Jensen foi significativo para as carteiras refeitas anualmente e a cada cinco anos.

Resultados melhores são notados com a estratégia de árvores de decisão baseadas na entropia de Shannon. O retorno da carteira anual foi de 22,67%, mas sem magnitude ou variabilidade suficiente para ser estatisticamente superior ao retorno de 16,05% do Ibovespa, com o mesmo ocorrendo para os demais períodos. Já o índice de Sharpe foi significativo nos portfólios de dois e cinco anos. O alfa de Jensen, por sua vez, indicou retornos anormais significativos para todos os períodos.

A baixa significância do índice de Sharpe pode ser justificada pelo fato das carteiras não serem sempre bem diversificadas, fato melhor discutido posteriormente. Esse medidor penaliza portfólios com poucos ativos, o que pode ter contribuído para a não significância estatística do indicador em alguns períodos.

Ainda assim, com as medida significativas nos índices corrigidos pelo risco, no-

vamente se tem indícios da capacidade de modelos detectarem empresas subavaliadas e utilizarem essa identificação para obter lucro. Contudo, as estratégias enfrentaram o problema de diversificação, o que pode ter prejudicado seus resultados. Nos dois modelos, houve anos em que nenhum ativo foi selecionado, enquanto em outros se ultrapassou 40 ações em carteira. Como indica a tabela 6, a média de ativos usando-se a entropia de Rényi foi de 12,65 ativos, com a entropia de Shannon foi de 26,5.

Tabela 6. Quantidade de ativos em carteira.

	Rényi	**Shannon**
Mínimo	0	0
Máximo	49	77
Média	12,5	26,5

Essa grande variação da quantidade de ativos ocasiona um problema de diversificação na construção da carteira, ora por ser excessivamente diversificada, podendo elevar demasiadamente os custos de formação do portfólio, ora por conter poucos (ou mesmo nenhum) ativo, sobrelevando o risco. Esse problema pode ser contornado com o cálculo de escores fatoriais, advindos a técnica estatística multivariada de análise fatorial, tomando como base as variáveis incluídas no modelo ou pelo emprego de lógica *fuzzy*. A operacionalização dessas propostas é deixada como sugestões de melhorias do modelo para trabalhos futuros.

O conjunto de anos sem variáveis na árvore de decisão (que se refletiu nas carteiras sem ativos), o índice de acerto maior dos outros modelos e os retornos bons, mas não significativos, do estudo de carteira para os dados de fora da amostra de construção, podem indicar algum superajustamento (*overfitting*) na construção das árvores. Nesse caso, técnicas distintas de crescimento, como a utilização de critérios diferentes do ganho de informação e da razão de ganho, e podagem podem ser empregadas para se tentar refinar o modelo, sendo mais uma proposição para outras pesquisas.

Voltando à discussão de investimento em valor, levantada pelas variáveis presentes nas árvores de decisão, cabe identificar quais são as ações que compuseram as carteiras. O previsto a partir das variáveis se confirma nos portfólios: empresas consideradas de valor foram sistematicamente selecionadas.

Tabela 7. Principais ações selecionadas pelos modelos de árvores de decisão.

	Rényi	**Shannon**	**Total**
PETR4	3	5	8
ITSA3	2	5	7
ELET3	3	4	7
BRTP3	2	5	7
ACES4	3	4	7
BRKM5	3	3	6
DURA4	3	3	6
EEEL3	2	4	6
TCSL4	2	4	6
FFTL4	2	4	6
TMCP4	2	4	6
TMGC7	2	4	6
TLPP4	2	3	5
AMBV4	2	3	5
EMBR3	1	4	5

Pela tabela 7, vê-se a presença de empresas como Petrobrás (PETR4), Telesp/Telefônica (TLPP4), Ambev (AMBV3), Itausa (ITSA3), Eletrobrás (ELET3) e Embraer (EMBR3), Brasil Telecom (BRTP3), Braskem (BRKM5), Duratex (DURA4), Tim (TCSL4), Telemig (TMCP4) e Telemig Celular (TMGC7). Todas grandes empresas de seus setores, substanciando a tese de investimento em valor.

Outras melhorias ao modelo podem ser buscadas variando-se o algoritmo de indução da árvore, o alfa adotado para a entropia de Rényi e a forma de estimação da janela de Parzen. Além de aplicar os conceitos da teoria da informação na construção de árvores de decisão, eles também podem ser estendidos para outras técnicas de reconhecimento de padrões, como redes neurais artificiais e *support vector machine* (SVM).

Mais uma possibilidade é construir regras não com base somente nas informações contábeis do último ano, mas tomando os últimos dois, três ou cinco anos como dados amostrais. Todavia, essa proposição tem que ser analisada à luz dos momentos econômicos vividos, já que bruscas modificações podem deturpar os resultados das técnicas de reconhecimento de padrões.

O conjunto de resultados, com a identificação de variáveis que discriminam os

grupos de ativos de sucesso dos de fracasso e o estudo de carteira que mostra a existência de retornos ajustados ao risco significativo, em especial com o uso da entropia de Shannon, mostra que há fragilidades na Hipótese de Mercado Eficiente.

Entretanto, não se pode afirmar categoricamente que a Hipótese de Mercado Eficiente não é válida, dadas as questões expostas no tópico Materiais e métodos. Dessa forma, uma sugestão para trabalhos futuros é a utilização de outras medidas de avaliação além do alfa de Jensen e do índice de Sharpe, como o M ao quadrado, índice de Treynor, modelos APM ou outros, de forma a se colher mais indícios sobre o sucesso das estratégias. Afinal, quanto mais indicadores de risco apontarem para a existência de estratégias de seleção que superam significativamente o mercado, mais consistentes são as evidências de ineficiência.

Também, apesar da extensão temporal, não se pode descartar a possibilidade das estratégias terem se beneficiado de ocorrências presentes somente nos anos estudados (de 1999 a 2009). Apesar de esse período ter compreendido momentos de crise e euforia do mercado financeiro, o modelo pode não ser tão satisfatório em outros períodos e em outras realidades econômicas. A manutenção desse estudo nos anos posteriores, sua aplicação em outros intervalos temporais, como janelas trimestrais, e em outros países é uma forma de elucidar a estabilidade desses resultados.

9 Considerações finais

Nesse trabalho, apresentou-se uma forma de se utilizar a teoria da informação e as árvores de decisão para identificar variáveis explicativas e classificar empresas que estejam subavaliadas pelo mercado. Além do uso corrente da entropia de Shannon, foi proposta a utilização da entropia quadrática de Rényi. Vale ressaltar que tal emprego pode ser expandido para outras técnicas de reconhecimento de padrões, como redes neurais e SVM.

O levantamento inicial das variáveis do modelo se baseou nos conceitos teóricos da escola fundamentalista por meio do método de múltiplos. O uso de árvores de decisão e das entropias foi capaz de reduzir a quantidade de variáveis e discriminar as empresas nos grupos Sucesso e Fracasso. Com isso, abre-se a oportunidade de se testar também a Hipótese de Mercado Eficiente em sua forma semiforte no mercado

acionário brasileiro.

Tendo como base para a seleção de variáveis as entropias de Rényi e Shannon, as árvores de decisão destacaram quatro variáveis como relevantes para discriminar os ativos de rentabilidade superior a do mercado: Preço de Mercado, Média de Crescimento dos Lucros, Preço por Valor Contábil Tangível e *Dividend Yield*. Essas variáveis fazem parte das usualmente utilizadas para segregar as ações de valor das de crescimento, mostrando que essa talvez seja a principal diferença nos retornos esperados para o mercado acionário brasileiro.

Na avaliação dos modelos pela abordagem de Lachenbruch, percebeu-se resultados satisfatórios, com acertos totais de 85,27% para a entropia quadrática de Rényi e 87,75% para a de Shannon. A árvore de decisão baseada na entropia de Shannon também trouxe a capacidade de classificar com maior índice de acerto do grupo Sucesso, obtendo o maior índice entre todos os modelos citados: 49,22%.

O estudo de carteira obteve retornos logarítmicos próximos ao do mercado, com o modelo utilizando entropia de Shannon o superando, mas não de maneira significativa. Contudo, quando o retorno é ajustado ao risco por meio do alfa de Jensen, foram encontrados estatisticamente significativos em todas as janelas de tempo para a entropia de Shannon (9,15% na carteira de um ano) e nas carteiras de um e cinco anos para a de Rényi (7,80% na carteira de um ano). O índice de Sharpe das estratégias só foi significativamente maior que o do mercado para o caso da entropia de Shannon nas carteiras de dois e cinco anos. Entretanto, colabora com esse fato o viés do índice de Sharpe penalizar carteiras pouco diversificadas, o que aconteceu com certa frequência no uso de árvores de decisão.

Com momentos sem formação de portfólio ou com poucos ativos em carteiras, em especial com uso da entropia de Rényi, depara-se com o problema de diversificação. Nas árvores de decisão não há uma medida de escore que permita contornar o problema, mas tal escore pode ser estimado com o uso da técnica multivariada de análise fatorial, por exemplo. Ou outras estratégias podem ser empregadas, como o uso de lógica *fuzzy*, para se obter refinamentos do modelo.

Referências

[1] ASSAF NETO, A. Mercado Financeiro. 7 ed. São Paulo: Atlas, 2006.

[2] DAMODARAN, A. Investiment Valuation. 2 ed. New York: John Wiley and Sons, 2002.

[3] DAMODARAN, A. Filosofias de Investimento. Rio de Janeiro: Qualitymark, 2006.

[4] GRAHAM, B; DODD, D. L. Security Analisys. 3. ed. New York: McGraw-Hill, 1951.

[5] GRAHAM, B. O Investidor Inteligente. Rio de Janeiro: Nova Fronteira, 2007.

[6] MELLAGI FILHO, A.; ISHIKAWA, S. Mercado Financeiro e de Capitais. 2. ed. São Paulo: Atlas, 2003.

[7] PÓVOA, A. Valuation – Como Precificar Ações. 2. ed. São Paulo: Globo, 2007.

[8] SHARPE, W. F. Capital Asset Prices: A Theory of Capital Asset Pricing. *J Financ*, v. 19, n.3, p. 425-442, 1964.

[9] SHARPE, W. F. ALEXANDER, G. J.; BAILEY, J. V. Investments. 5 ed. New Jersey : Prentice Hall, 1995.

[10] MORETIN, P. A.; TOLOI, C. M. C. Análise de Séries Temporais. São Paulo: Edgard Blücher, 2006.

[11] JENSEN, M.C. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *J Bus*, vol. 42, n.2, p.167-247, 1969.

[12] FAMA. E. F. Efficient Capital Markets II. *J Financ*, v. 46, n.5, p. 1575-1617, 1991.

[13] FAMA. E. F. Efficient capital markets: a review of theory and empirical work. *J Financ*, v. 25, n.2, p. 383-417, 1970.

[14] COSTA JR, N. C. A; COSTA, N. C. A. Teoria do Caos e Mercado Financeiro. In: COSTA JR, N. C. A; LEAL, R. P. C; LEMGRUBER, E. F. (Orgs) Mercado de Capitais: análise empírica no Brasil. São Paulo: Atlas, 2000. p. 168-173.

[15] COSTA JR, N. C. A.; LEAL, R. P. C.; LEMGRUBER, E. F. Mercado de Capitais – Análise Empírica no Brasil. Rio de Janeiro: Coppead/UFRJ, 2000.

[16] LIMA, L. A. O. Auge e Declínio da Hipótese dos Mercados Eficientes. *Revista de Economia Política*, São Paulo, v. 23, n. 4 (92), 2003.

[17] FORTI, C. A. B.; PEIXOTO, F. M.; SANTIAGO, W. P. Hipótese da Eficiência de Mercado: Um Estudo Exploratório no Mercado de Capitais Brasileiro. *Gestão & Regionalidade*, v.25, n.75, 2009.

[18] SHANNON, C. E. A Mathematical Theory of Communication. *Bell Syst Tech J*, n. 27, p. 379-423 e p. 623-656, 1948.

[19] ARTUSO, A. R. Entropias de Shannon e Rényi aplicadas ao Reconhecimento de Padrões. *Revista CIATEC-UPF*, v. 3, p. 56-72, 2011.

[20] GONÇALVES, L. B. Entropia de Rényi e informação mútua de Cauchy-Schwarz aplicadas ao algoritmo de seleção de variáveis MIFS-u: um estudo comparativo. Dissertação de Mestrado – Pontifícia Universidade Católica do Rio de Janeiro, 2008.

[21] SCOTT, D. W. Multivariate Density Estimation. New York: John Wiley & Sons, 1992.

[22] JENSSEN, R.; PRINCIPE, J. C.; ERDOGMUS, D.; ELTOFT, T. The Cauchy-Schwarz Divergence and Parzen Windowing: Connections to Graph Theory and Mercer Kernels. *J Frankl Inst*, v. 343, n 6, p. 614–629, 2006.

[23] PRINCIPE, J. C. Information theoretic learning: Rényi's entropy and kernel perspectives. New York: Springer Verlag, 2010.

[24] GAMA, J. M. P. Combining classification algorithms. 195p. Tese de Doutorado, Universidade do Porto, 1999, 195p.

[25] QUINLAN, J. R. C4.5: Programs for Machine Learning. San Diego (EUA): Morgan Kaufmann, 1993.

[26] QUINLAN, J. R. Decision trees and multivalued attributes. *Mach Intell*, n. 11, p. 305-318, 1988.

[27] SHARPE, W. F. Asset allocation: management style and performance measurement. *J Porfolio Manage*, v. 18, n.2, p. 7-19, 1992.

[28] RAMOS, P. B; PICANÇO, M. B; COSTA JR., N. C. Retornos e Riscos das Value e Growth Stocks no Mercado Brasileiro. In: COSTA JR, N. C. A; LEAL, R. P. C; LEMGRUBER, E. F. (Orgs) Mercado de Capitais: análise empírica no Brasil. São Paulo: Atlas, 2000. p. 124-138.

[29] COSTA JR., N. C. A; NEVES, M. B. E. das. As variáveis fundamentalistas retornos das ações no Brasil. *Revista Brasileira de Economia*, v. 54, n.1, p. 123-137, 2000.

[30] SANTOS, L. R. Aplicação de estratégias de value investing no mercado acionário brasileiro. Dissertação de Mestrado, Faculdade de Economia e Finanças Ibmec, Rio de Janeiro, 2010.

[31] TAVARES, A. L. A eficiência da análise financeira fundamentalista na previsão de variações no valor da empresa. Tese de Doutorado, Universidade Federal do Rio Grande do Norte, Natal, 2010.

[32] ARTUSO, A. R.; CHAVES NETO, A. O uso de quartis para a aplicação dos filtros de Graham na Bovespa (1998-2009). *Revista Contabilidade & Finanças*, v. 21, n. 1, 2010.

[33] ROCHA, I.; HEIN, N.; KROENKE, A. Grau de entropia da informação em indicadores econômico-financeiros das empresas do setor econômico materiais básicos participantes dos níveis de governança corporativa da BM&FBovespa no período de 2005 a 2009. In: XLIII Simpósio Brasileiro de Pesquisa Operacional, 2011. Anais..., Ubatuba: SBPO, 2011.

[34] LIMA, C. F. L.; ASSIS, F. M.; SOUZA, C. P. Árvores de Decisão baseadas nas entropias de Shannon, Rényi e Tsallis para Sistemas Tolerantes a Intrusão. In: La Novena Conferencia Iberoamericana en Sistemas, Cibernética e Informática CISCI 2010, 2010, Orlando. Anais... Orlando: CISCI, 2010.