

ANÁLISE DISCRIMINANTE E REGRESSÃO LOGÍSTICA – RECONHECIMENTO DE PADRÕES PARA A SELEÇÃO DE PORTFÓLIOS NO MERCADO ACIONÁRIO BRASILEIRO

Alysson Ramos Artuso¹, Anselmo Chaves Neto¹

Resumo: *Nesse artigo propõe-se a aplicação das técnicas multivariadas de análise discriminante e regressão logística para o reconhecimento de padrões no mercado acionário brasileiro. Com base no método de múltiplos, foram selecionados 22 indicadores econômico-financeiros levantados para todas as empresas não-financeiras do período de 1999 a 2009. 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 acima de 80%. Além disso, foram identificadas as variáveis preço de mercado, preço por valor contábil e taxa de dividendos como os atributos que melhor discriminam os grupos. Um estudo de carteira demonstrou a aplicabilidade dos modelos, inclusive contornando o problema de diversificação, com retornos ajustados ao risco significativos especialmente na medida do alfa de Jensen. Os resultados apontam para ineficiências no mercado acionário brasileiro, com preferência pelo investimento em ações de valor em detrimento ao investimento ações de crescimento.*

Palavras-chave: Análise Discriminante, Regressão Logística, Seleção de Portfólio, Análise Multivariada, Hipótese de Mercado Eficiente.

Abstract: *In this article, we suggested the application of the multivariate techniques discriminant analysis and logistic regression for the pattern recognize in Brazilian stock market. Based on the multiple model, 22 economic-financial indicators were selected for all non-financial companies in the period between 1999 and 2009. The classification rules constructed were able to discriminate the groups with higher or lower stock returns than the market with an efficiency greater than 80%. In addition, we identified the variables market price, price for book value and dividend yield as the attributes that best discriminate the groups. A portfolio approach demonstrated the applicability of models, including solving the diversification problem, with significant abnormal returns especially for Jensen's alpha. The results point to inefficiencies in the Brazilian stock market, with a preference for investing in value stocks over growth stocks.*

Keywords: Discriminant Analysis, Logistic Regression, Portfolio Selection, Multivariate Analysis, Efficient Market Hypothesis.

Código JEL: G11 - Portfolio Choice; Investment Decisions

Introdução

A decisão de investir em ações deve ser precedida de uma análise das expectativas dos rendimentos a serem obtidos ao longo do prazo de permanência com a posição acionária. Dessa forma, a principal tarefa do investidor está na avaliação do retorno esperado em conjunto com o risco assumido devido ao capital aplicado.

¹ Departamento de Estatística – UFPR – alysson.artuso@gmail.com , anselmo @est.ufpr.br

São diversos os fatores que podem influenciar nos preços das ações, tanto internos, como os relacionados ao desempenho da empresa emissora, quanto externos, relacionados à economia do país e à expectativa dos agentes econômicos.

Uma das formas de se analisar o investimento em ações e se construir portfólios é o método de múltiplos da escola fundamentalista, que se baseia em indicadores de mercado e em indicadores econômico-financeiros, os quais visam identificar ações subavaliadas em relação a seus pares.

Estudos empíricos sobre esses indicadores são relativamente comuns em países com economia consolidada e mercado de capitais popularizado. Porém, sua utilização a outras épocas e a outros mercados não é simples, necessitando testes mais específicos.

No Brasil, a comunidade científica tem voltado sua atenção para esse tema, o que já gerou alguma base científica sobre o assunto. Porém, é pouco difundida a utilização da Análise Multivariada nesses estudos, pois, tipicamente, as variáveis são avaliadas de maneira isolada. Exceções são os trabalhos de TAVARES (2010) e ARTUSO e CHAVES NETO (2010), que utilizaram técnicas de reconhecimento de padrões para analisar conjuntamente diversas variáveis fundamentalistas.

Nessa lacuna, sugere-se a aplicação do reconhecimento de padrões multivariados de resposta dicotômica – a análise discriminante via função de Fisher e a regressão logística. Essas técnicas são capazes de contribuir tanto para o desenvolvimento de estratégias para seleção de ativos em um portfólio, como para a compreensão do funcionamento do mercado de capitais nacional, pois favorece a identificação de variáveis relevantes. Em geral, com melhores resultados na classificação do que modelos unidimensionais, os métodos multivariados são capazes de evidenciar relações complexas entre a estrutura de dados, além de apontar para as variáveis mais relevantes em termos de discriminação das observações.

Para cumprir o objetivo de construir modelos para a construção de carteiras e contribuir para a compreensão do mercado acionário brasileiro, foram analisadas as empresas não-financeiras da Bolsa de Valores de São Paulo no período de 1999 a 2009. A partir de 22 indicadores 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 com um nível de significância de 5%. Para comparar métodos de reconhecimento de padrões, foram utilizadas duas ferramentas distintas, a análise discriminante e a regressão logística.

De modo geral, os modelos construídos tiveram um índice de acerto superior a 80%, produzindo retornos acima dos de mercado de maneira consistente ao longo dos anos, com vantagem para o modelo baseado na Análise Discriminante. Os modelos também apontaram as variáveis Preço de Mercado, Taxa de Dividendos e Preço por Valor Contábil como as de maior relevância no cenário nacional. Tais resultados trazem indícios contrários à Hipótese do Mercado Eficiente para o contexto brasileiro.

O conteúdo do presente artigo está organizado da seguinte maneira. A próxima seção apresenta o método de múltiplos, com as variáveis fundamentalistas utilizadas nesse estudo, a seção seguinte trata das medidas de risco habituais das finanças e da Hipótese do Mercado Eficiente, a quarta seção fornece um resumo da técnica de análise discriminante, a quinta aborda a regressão logística, seguida de uma seção sobre pesquisas similares. Na sétima seção são detalhados os dados utilizados e explicitada a metodologia de pesquisa, a oitava seção traz os resultados e discussões da aplicação das duas técnicas de análise multivariada aplicadas para a construção de portfólios, fechando com as considerações finais, as referências e um anexo, no qual são mostradas as definições das variáveis fundamentalistas utilizadas.

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 (ASSAF NETO, 2006; COPLAND; KOLLER; MURRIN, 2002; DAMODARAN, 2002, 2006; GRAHAM; DODD, 1951, GRAHAM, 2007; MELLAGI FILHO; ISHIKAWA, 2003; PÓVOA, 2007), selecionou-se 22 indicadores amplamente aceitos que serão analisados neste trabalho. São eles: Preço de Mercado (PM), Taxa de Dividendos (DY), Preço/Lucro (P/L), Preço/Vendas (P/V), Preço/Valor contábil (P/VC), Preço/Valor contábil tangível (P/VCT), Preço/Capital de Giro Líquido (P/CGL), Retorno sobre o Patrimônio Líquido (ROE), Retorno sobre Ativos (ROA), Retorno sobre o Capital (ROC), Margem líquida (ML), Média do Crescimento dos Lucros por Ação (MCL), Liquidez Corrente (LC), Liquidez Seca (LS), Liquidez Geral (LG), Liquidez Imediata (LI), Grau de Endividamento (GE), Participação de Capitais de Terceiros Sobre Recursos Totais (PCTRT), Garantia do Capital Próprio ao Capital de Terceiros (GCPCT), Composição do Endividamento (CP), Liquidez Geral Modificada (LGM) e Grau de Endividamento Modificado (GEM). Embora esses indicadores sejam amplamente conhecidos, podem existir nomenclaturas ou definições variadas, motivo pelo qual eles são especificados no Anexo.

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, 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.

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 neperiano dos preços das ações:

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

onde 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 normalidade dos dados e de homocedasticidade 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.

Assim, o alfa de Jensen é calculado por

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

onde α_i é a medida de performance de Jensen; $E(r_i)$ é o retorno médio do portfólio; 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.

E o índice de Sharpe é obtido por

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

onde IS_i é o índice de Sharpe da carteira de investimento i ; $E(r_i)$ e r_f foram vistas anteriormente e σ_i é o desvio padrão de r_i . Com os retornos considerados sendo sempre os retornos logarítmicos.

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.

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_t , 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 (1970) 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 (COSTA JR; COSTA, 2000). Tais indícios de ineficiência são ainda mais fortes nos mercados em desenvolvimento como o brasileiro (COSTA JR, LEAL e LEMGRUBER, 2000; LIMA, 2003; FORTI, PEIXOTO e SANTIAGO, 2009).

Análise discriminante – Método de Fisher

As técnicas estatísticas multivariadas que compõem a Análise Discriminante se referem à solução de problemas de classificação de novos objetos em conjuntos previamente definidos, e têm o objetivo específico de utilizar as diversas variáveis na composição de funções discriminantes que classificarão novos objetos no conjunto (grupo) mais adequado. Esta técnica é frequentemente utilizada no contexto de reconhecimento de padrões com o objetivo de construir regras de reconhecimento e classificação.

As discriminações podem ser feitas através de processos supervisionados, que são utilizados quando se conhece os padrões, ou através de processos não-supervisionados, usados quando não se tem um padrão reconhecido. Os métodos Função Discriminante Linear de Fisher (FDLF) e Modelo de Regressão Logístico Múltiplo (MRLM), FDLF são métodos supervisionados, enquanto a Análise de Agrupamentos (*Clusters*) é um método não-supervisionado.

Na Análise Discriminante, as categorias definidas (duas no caso deste trabalho) são denominadas por π_1 e π_2 . Os objetos são separados ou classificados usando a base de suas medidas associadas a p variáveis aleatórias. Para avaliar se os grupos são significativamente diferentes, utiliza-se o teste T^2 de Hotteling, equivalente ao teste t para médias amostrais univariadas. A estatística do teste é

$$T^2 = [(\bar{x}_1 - \bar{x}_2) - \delta_0] \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_p \right]^{-1} [(\bar{x}_1 - \bar{x}_2) - \delta_0]$$

onde S_p é a matriz de covariância amostral calculada de forma conjunta e que estima o parâmetro Σ , considerado comum. A forma de S_p é:

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

e o produto da estatística T^2 pela razão de graus de liberdade tem uma distribuição F com p e $n_1 + n_2 - p - 1$ graus de liberdade, ou seja:

$$T^2 \frac{(n_1 + n_2 - p - 1)}{(n_1 + n_2 - 2)p} \sim F_{p, n_1 + n_2 - p - 1}(1 - \alpha)$$

O método de Fisher consiste em transformar as observações multivariadas $\underline{\mathbf{X}}$ nas observações univariadas Y tal que os Y 's nas populações π_1 e π_2 sejam separadas tanto quanto possível.

Isso é feito através de combinações lineares dos $\underline{\mathbf{X}}$ criando os Y 's, tendo em vista que combinações lineares são funções de $\underline{\mathbf{X}}$ e por outro lado são de fácil cálculo. Portanto, seja μ_{1Y} a média dos Y 's obtidos dos $\underline{\mathbf{X}}$'s pertencentes à população p_1 e μ_{2Y} a média dos Y 's obtidos dos $\underline{\mathbf{X}}$'s pertencentes a p_2 , então se seleciona a combinação linear que maximiza a distância quadrática entre μ_{1Y} e μ_{2Y} relativamente à variabilidade dos Y 's (GUIMARÃES, 2000). Portanto, seja $\mu_1 = E(\underline{\mathbf{X}}|p_1)$: valor esperado de uma observação multivariada de p_1 ; $\mu_2 = E(\underline{\mathbf{X}}|p_2)$: valor esperado de uma observação multivariada de p_2 e supondo a matriz de covariância $\Sigma = E[(\underline{\mathbf{X}} - \mu_i)(\underline{\mathbf{X}} - \mu_i)']$, $i = 1, 2$ como sendo a mesma para ambas as populações, então considerando a combinação linear $Y = \underline{\mathbf{C}}' \underline{\mathbf{X}}$ tem-se:

$$\mu_{1Y} = E(Y|p_1) = E(\underline{\mathbf{C}}' \underline{\mathbf{X}}|p_1) = \underline{\mathbf{C}}' E(\underline{\mathbf{X}}|p_1) = \underline{\mathbf{C}}' \underline{\mu}_1,$$

da mesma forma

$$\mu_{2Y} = E(Y|p_2) = E(\underline{\mathbf{C}}' \underline{\mathbf{X}}|p_2) = \underline{\mathbf{C}}' E(\underline{\mathbf{X}}|p_2) = \underline{\mathbf{C}}' \underline{\mu}_2$$

e

$$\sigma_Y^2 = V(Y) = V(\underline{\mathbf{C}}' \underline{\mathbf{X}}) = \underline{\mathbf{C}}' V(\underline{\mathbf{X}}) \underline{\mathbf{C}} = \underline{\mathbf{C}}' \Sigma \underline{\mathbf{C}},$$

que é a mesma para as duas populações.

Segundo esse método, a melhor combinação linear vem da razão entre o quadrado da distância entre as médias e a variância de Y , ou seja,

$$\frac{(\mu_{1Y} - \mu_{2Y})^2}{\sigma_Y^2} = \frac{(\underline{\mathbf{C}}' \underline{\mu}_1 - \underline{\mathbf{C}}' \underline{\mu}_2)^2}{\underline{\mathbf{C}}' \Sigma \underline{\mathbf{C}}} = \frac{(\underline{\mathbf{C}}' \underline{\delta})^2}{\underline{\mathbf{C}}' \Sigma \underline{\mathbf{C}}},$$

onde: $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ e $V(\underline{\mathbf{X}}) = \Sigma$

A razão anterior é maximizada por considerar que a matriz de covariância é positiva definida, e então a demonstração é feita aplicando o lema de maximização baseado nas desigualdades de Cauchy-Schwarz, resultando em $\underline{\mathbf{C}} = K \Sigma^{-1} \underline{\delta} = K \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$ para qualquer $K \neq 0$ (JOHNSON; WICHERN, 1998).

Escolhendo-se $K = 1$ tem-se

$$\underline{C} = \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \text{ e}$$

$$Y = \underline{C}' \underline{X} = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X}$$

que é conhecida como Função Discriminante Linear de Fisher (FDLF).

A FDLF transforma as populações multivariadas p_1 e p_2 em populações univariadas tais que as médias das populações univariadas são separadas tanto quanto possível relativamente à variância populacional considerada comum.

Toma-se

$$Y_0 = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X}_0$$

como valor da FDLF para uma nova observação \underline{X}_0 e consideramos o ponto médio m entre as duas populações univariadas como

$$m = \frac{1}{2}(\mu_{1Y} + \mu_{2Y}) = \frac{1}{2}(\underline{C}' \underline{\mu}_1 + \underline{C}' \underline{\mu}_2) = \frac{1}{2} \underline{C}' (\underline{\mu}_1 + \underline{\mu}_2)$$

$$m = \frac{1}{2}[(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2)]$$

Pode-se mostrar, portanto, que $E(Y_0/p_1) - m \geq 0$ e $E(Y_0/p_2) - m < 0$. Ou seja, se $\underline{X}_0 \in p_1$, é esperado que $Y_0 \geq m$. Por outro lado se $\underline{X}_0 \in p_2$, é esperado que $Y_0 < m$.

Como se trabalha com amostras, a FDLF será estimada por:

$$\hat{Y} = \hat{\underline{C}}' \underline{X} = (\underline{X}_1 - \underline{X}_2)' S_p^{-1} \underline{X},$$

sendo o ponto médio m estimado por

$$\hat{m} = \frac{1}{2}(\underline{X}_1 - \underline{X}_2)' S_p^{-1} (\underline{X}_1 + \underline{X}_2) = \frac{\bar{Y}_1 + \bar{Y}_2}{2},$$

$$\text{com } S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

Uma vantagem desse método é não necessitar da suposição de que as populações sejam distribuídas normalmente, entretanto, é assumido que as matrizes de covariância populacional Σ seja a mesma para os dois grupos.

Uma maneira de construir modelos cuja variabilidade seja mais bem explicada, ao mesmo tempo em que pode maximizar a diferença entre os grupos, é o método *stepwise* (por etapas). Esse método sequência busca estimar a função discriminante selecionando variáveis para serem acrescentadas (*forward stepwise*) ou eliminadas (*backward stepwise*) do modelo até que alguma medida de critério geral seja alcançada. Este tem por objetivo selecionar variáveis que maximiza a previsão com o menor número de variáveis empregadas. Com isso, pode-se identificar as variáveis estatisticamente mais significantes para compor a função discriminante, uma vez somente variáveis que aumentem significativamente a capacidade discriminativa do modelo são adicionadas no método *forward*. De maneira análoga, uma variável presente na função discriminante é removida se a capacidade do modelo não for significativamente reduzida pela remoção dessa variável no método *backward*.

Uma das estatísticas mais utilizadas para denotar a significância estatística do poder discriminatório de uma função discriminante é o lambda de Wilks. Ele é um dos critérios baseados na Análise de Variância Multivariada (MANOVA) para avaliar diferenças multivariadas entre grupos (HAIR, ANDERSON, TATHAM; 1987). Sua formulação é:

$$\Lambda = \frac{\det({}_p A_p)}{\det({}_p A_p + {}_p B_p)} = \frac{1}{\det(I + A^{-1}B)}$$

onde ${}_pA_p$ é a matriz de dispersão multivariada dentro do grupo (no caso geral, com distribuição $W_p(\Sigma, m)$; $m > p$) e ${}_pB_p$ a matriz de dispersão multivariada entre grupos (no caso geral, com distribuição $W_p(\Sigma, n)$).

Sendo λ um autovalor da matriz $\mathbf{A}^{-1}\mathbf{B}$, então $1 + \lambda$ é um valor próprio $\mathbf{I} + \mathbf{A}^{-1}\mathbf{B}$:

$$(\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})\mathbf{x} = \mathbf{x} + \mathbf{A}^{-1}\mathbf{B}\mathbf{x} = \mathbf{x} + \lambda\mathbf{x} = (1 + \lambda)\mathbf{x}.$$

Logo, o Lambda de Wilks pode ser escrito como:

$$\Lambda = \frac{1}{\det(\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})} = \frac{1}{\prod_{i=1}^p (1 + \lambda_i)} \sim \Lambda(p, m, n)$$

onde os λ_i são os autovalores de $\mathbf{A}^{-1}\mathbf{B}$; p é a dimensão; m são os graus de liberdade dos resíduos e n os graus de liberdade da hipótese para o contexto apresentado.

A distribuição exata do Lambda de Wilks não é conhecida, entretanto, se tem aproximações à distribuição (MARDIA, KENT, BIBBY; 1979). Para $n = 2$ e $p > 2$:

$$\frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \cdot \frac{as - \frac{pn}{2} + 1}{pn} \sim F\left(pn, as - \frac{pn}{2} + 1\right)$$

$$\text{com } a = m - \frac{p - n + 1}{2} \text{ e } s = \sqrt{\frac{p^2 n - 4}{p^2 + n^2 - 5}}.$$

Regressão logística

Os métodos de regressão têm como objetivo principal descrever a relação entre uma variável resposta (dependente) e uma ou mais variáveis explicativas (independentes). O que distingue o modelo de regressão logística do modelo de regressão linear é que a variável resposta no primeiro caso é binária ou dicotômica (GUIMARÃES, 2000).

O modelo de regressão linear é do tipo

$$E(Y|x) = \beta_0 + \beta_1 X$$

sendo que essa expressão implica que é possível para a esperança tomar qualquer valor entre $-\infty$ e ∞ .

Considerando dados dicotômicos a média condicional varia de 0 a 1, ou seja, $0 \leq E(Y|x) \leq 1$, a distribuição acumulada de uma variável aleatória pode ser usada para proporcionar um modelo para $E(Y|x)$ no caso que Y é dicotômica, é o caso da Regressão Logística.

Usando a notação $\pi(x) = E(Y|x)$, o modelo de regressão logística aqui adotado é

$$E(Y|x) = \pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

com a transformação logit

$$\ln \pi(x) = \ln \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

pode ser colocada na forma

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x$$

A importância dessa transformação é que $g(x)$ tem diversas propriedades desejáveis do modelo de regressão linear, pois a logit $g(x)$ é linear nos seus parâmetros. Porém existem diferenças em relação ao modelo de regressão linear, como a média condicional da equação de regressão que deve variar entre 0 e 1 e a distribuição dos

erros que possui distribuição binomial com média zero e variância igual a $\pi(x)[1-\pi(x)]$.

Assim, a Regressão Logística é uma técnica de análise que relaciona variáveis independentes, as quais podem ser métricas ou não métricas, a uma variável dependente binária. FÁVERO *et al* (2009, p. 13) afirmam que a previsão resultante é um valor entre zero e um ou, de outra forma, a probabilidade de obtenção do valor um, sendo, portanto, uma técnica “[...] útil para aferição da probabilidade de ocorrência de um evento e para identificação das características dos elementos pertencentes a cada categoria estabelecida pela dicotomia da variável dependente”.

Esse desenvolvimento pode ser estendido para diversas variáveis independentes, de forma que o modelo de regressão logística múltipla a ser utilizado é expresso por:

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

onde: $g(x) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$ (HOSMER; LEMESHOW, 1989)

Da mesma forma que na Análise Discriminante via Função Linear Discriminante de Fisher, também podem utilizados métodos *stepwise*, tendo como estatística avaliadora o Lambda de Wilks, que podem reduzir o número de variáveis incorporadas ao modelo.

Diferentemente do que ocorre com a Regressão Linear, a estimação exata dos parâmetros na Regressão Logística não possui uma forma analítica fechada, sendo que as soluções dos sistemas não-lineares de equações para estimação dos β 's requerem a utilização de métodos iterativos na estimação desses parâmetros.

Entretanto, considerando que o problema de estimação dos parâmetros pode ser entendido como um problema de otimização, pode-se considerar a utilização de algum método de minimização para esse fim.

A estimação dos parâmetros para ajuste do modelo de regressão logística é comumente efetuado pelo Método Quase-Newton, pelo Método da Função Discriminante ou pelo Modelo de Localização de Variáveis Mistas. A opção adotada, por eficiência na estimação e facilidades computacionais, foi a do Método Quase-Newton. O *software* utilizado para tal foi o STATISTICA 8.0.

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 (1992) identificou que tamanho e *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 (1992) pode-se definir o investimento em valor como aquele no qual se procura empresas grandes e consolidadas, cujas ações estejam subvalorizadas 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 (2000) 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 (2000) 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 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 no trabalho de COSTA JR e NEVES (2000) 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.

BRUNI (1998), analisando o comportamento de 265 ações preferenciais e 65 ações ordinárias de empresas brasileiras negociadas na Bovespa, no período compreendido entre julho de 1986 e julho de 1997, encontrou as seguintes associações entre o retorno apresentado pelas ações e variáveis fundamentalistas.

Ações ordinárias e preferenciais com o múltiplo P/VC baixo apresentaram retornos significativamente mais elevados para o período de 1988 a 1995. Em relação ao múltiplo P/L os resultados observados não foram considerados significativos para nenhum momento do período estudado, o mesmo ocorrendo para a relação preço por fluxo de caixa (P/FC).

Com base na menor relação P/VC o portfólio constituído pelas ações ordinárias apresentou um retorno anual médio de 88%, frente a um retorno de 50% das ações preferenciais com baixa relação P/VC e 41% do índice de mercado Ibovespa.

Não foi encontrada relação significativa entre risco e retorno das ações, a não ser para o período de 1988 a 1994 para as ações preferenciais. Nesse período foi encontrada uma associação positiva entre o retorno e o beta.

Uma análise empírica da Taxa de Dividendos no mercado brasileiro foi realizada BUENO (2002) com dados de 1994 até 1999 que não observou evidências de uma associação entre a taxa de dividendos e o retorno das ações. Ao realizar um método de carteiras para comparação das rentabilidades, o autor observou que o portfólio composto por empresas que não distribuem dividendos obteve a maior rentabilidade, porém nenhum teste estatístico foi conclusivo sobre a significância dessa e das outras observações.

SILVA (2003) analisou as variáveis Preço de Mercado, Preço por Lucro, Preço por Valor Contábil e o Retorno sobre o Patrimônio Líquido sobre o retorno das ações no mercado acionário brasileiro no período de 1993 a 2003. A pesquisa apresentou diferenças estatisticamente significativas apenas nos índices P/L e P/VC, com as carteiras compostas por baixos múltiplos apresentando retornos estatisticamente superiores, 117,4% e 107,5% ao ano, respectivamente, contra 50,4% do Ibovespa.

Analisando a correlação entre variáveis fundamentalistas, o risco beta e o retorno das ações, NAGANO, MERLO e SILVA (2003) concluíram que existe uma

correlação positiva entre o retorno e o beta e liquidez das ações, e negativa entre o retorno e o valor de mercado, preço por valor patrimonial e Preço por Lucro. A amostra do estudo contou com as 55 empresas componentes do índice Ibovespa do período de 1995 a 2000. A correlação mais significativa foi a do índice P/L, seguida por valor de mercado. Não foram identificadas correlações significativas entre o retorno das ações e os índices Preço/Fluxo de Caixa, Preço/Vendas, Taxa de Dividendos e os níveis de endividamento das companhias.

ROSTAGNO, SOARES e SOARES (2008) analisaram 37 variáveis fundamentalistas para o período de 1995 a 2002 e identificaram a relevância do endividamento, tamanho, Retorno sobre Ativos e patrimônio líquido, P/L e DY, entre outras, como variáveis capazes de separar as carteiras entre vencedores e perdedoras no período.

SANTOS (2010) 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 (2010) 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 (2010) fizeram uma avaliação dos múltiplos introduzida por Benjamin Graham, considerado um dos criadores da análise fundamentalista, mostrando que a filtragem passiva, que pode ser considerada uma forma rudimentar de reconhecimento de padrões, poder ser adaptada ao mercado brasileiro e gerar retornos ajustados superiores ao Ibovespa. 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 alcançou rentabilidades significativamente superiores ao mercado, com alfa de Jensen de 26,26% ao ano.

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*, entendidos como as observações que ultrapassaram três desvios padrões nas duas primeiras componentes principais, e das empresas que não continham todos os dados disponíveis, houve uma média de 200 observações a cada ano.

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.

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. Caso contrário, os ativos foram alocados no grupo Fracasso.

As funções discriminantes foram construídas pelo método *forward stepwise*, no qual apenas algumas variáveis são incluídas no modelo – o que favorece a identificação das variáveis, a separação dos grupos e, conseqüentemente, a eficiência das funções discriminantes. O teste T^2 de Hotelling foi significativo para todos os anos, conforme mostra a Tabela 1.

TABELA 1: TESTES T^2 DE HOTELLING PARA MODELO *FORWARD STEPWISE*

Ano	Quantidade de variáveis incluídas	Lambda de Wilks	<i>p-value</i>
1999	14	0,590628	0,000000
2000	6	0,781776	0,000001
2001	9	0,862613	0,001066
2002	3	0,965280	0,049084
2003	7	0,823822	0,000059
2004	4	0,872876	0,000034
2005	5	0,887949	0,000247
2006	7	0,817520	0,000019
2007	4	0,881922	0,000284
2008	6	0,905841	0,002686
2009	3	0,943123	0,009031

Os modelos de classificação foram avaliados segundo a abordagem de Lachenbruch. Essa abordagem consiste num método de validação cruzada, na qual uma observação é excluída, constrói-se o modelo de classificação sem considerá-la e então se classifica a observação retirada anteriormente, processo repetido para todas as observações.

Na seqüência, as regras de classificação foram utilizadas para alocar ativos no ano seguinte a sua construção. Ou seja, foram aplicadas para dados de fora do conjunto de treinamento e avaliados segundo um estudo de carteira, primeiro com as alocações originais e, num segundo momento, somente com as 12 alocações de maior score no grupo Sucesso.

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 (DAMODARAN, 2006; COSTA JR; NEVES, 2000). 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.

Resultados e discussão

As técnicas de reconhecimento de padrões utilizadas, Análise Discriminante via função linear de Fischer (FDLF) e Regressão Logística (MRLM) foram utilizadas, primeiro, para identificar as variáveis mais relevantes para diferenciar ativos com rentabilidade superior ao mercado dos que não possuem tal rentabilidade. Para esse objetivo, contou-se com o método *forward stepwise*. A Tabela 2 traz variáveis incluídas a cada ano nos modelos.

TABELA 2: VARIÁVEIS PRESENTES NAS FUNÇÕES DISCRIMINANTES

Ano	FDLF	MRLM
1999	PM, P/L, P/V, P/VC, P/CGL, ROE, ROA, ROC, MCL, LI, GE, PCTRT, LGM, GEM	PM, LI, GE, GEM
2000	PM, DY, P/VC, P/CGL, GCPCT, CP	PM, P/VC
2001	PM, DY, P/V, P/CGL, LC, LG, LI, GCPCT, LGM	DY, P/CGL, LI
2002	DY, ML, LGM	LG, GCPCT
2003	PM, DY, P/L, P/VC, ROA, PCTRT, CP	PM,
2004	PM, DY, LC, LI	PM, DY
2005	P/VCT, P/CGL, MCL, LC, CP	P/VC, CP
2006	PM, DY, P/VC, P/CGL, ROC, MCL, CP, GEM	PM, P/VC
2007	PM, DY, GE,	PM, DY, ROC, LC
2008	DY, P/VC, LC, GE, GCPCT, GEM	P/VC,
2009	DY, P/VC, ROC	DY, PCTRT

As funções para o reconhecimento de padrões são construídas a partir dos grupos 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, três variáveis se mostraram recorrentes ao longo dos anos: Preço de Mercado (PM), Taxa de dividendos (DY) e Preço por Valor Contábil (P/VC).

O Preço de Mercado é uma variável dada como significativa em diversos trabalhos nacionais e internacionais de diferentes épocas, como em COSTA JR e NEVES (2000), SILVA (2003) e NAGANO, MERLO e SILVA (2003). Nos estudos levantados parece não haver uma direção muito bem definida sobre qual direção é mais vantajosa para o Preço de Mercado, alguns apontando para vantagens de pequenas, outros para grandes empresas quando se trata de investimento em valor.

Os estudos sobre a distribuição de dividendos (BUENO, 2002; ROSTAGNO, SOARES e SOARES, 2008) também não são conclusivos, mas quando apontam relevância para essa variável sempre relacionam altas taxas de distribuição de dividendos com maiores retornos dos ativos.

Uma possível explicação, assim como para o PM, também passa pelos períodos analisados. 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 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 fato costumeiro no mercado acionário que, em períodos de alta volatilidade, os investidores migrem de ativos mais arriscados para os entendidos como de menor risco. Mantendo-se no campo acionário, ativos de grandes companhias e que pagam altos dividendos são entendidos como tal, 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.

Com essa premissa e dados os estudos passados e a janela temporal dessa pesquisa, não se pode descartar a conjuntura econômica do período como fator condicional para a importância ou não, e em qual direção, das variáveis PM e DY. Diante disso, com base nos modelos construídos, esperasse que os investimentos em ativos de grandes empresas e com alta distribuição de dividendos continuem se destacando se os momentos de instabilidade continuarem frequentes nas próximas décadas.

Era de se esperar que o índice P/VC 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, como BRUNI (1998); RAMOS, PIKANÇO; COSTA JR (2000); COSTA JR; NEVES (2000) e SILVA (2003). Dessa forma, diferentes técnicas se reforçam ao apontar esse indicador como fundamental na análise de ativos, independente do mercado ou do momento econômico vivido.

Em menor grau que as três variáveis discutidas até então, mas também frequente nas funções discriminantes construídas, está a razão Preço por Capital de Giro Líquido (P/CGL). Assim, como o P/VC, ela funciona como um filtro de preço, indicando se a empresa está barata ou cara em relação ao capital de giro descontada a dívida financeira da empresa. Esse também é o critério mais restritivo dos filtros de Graham, que já foi testado com sucesso por ARTUSO e CHAVES NETO (2010).

Pelos resultados e pelos ativos selecionados, constata-se que os modelos de reconhecimento de padrões indicam ativos de alto PM, alto DY e baixo P/VC como os de rentabilidade acima da média de mercado. Essas são justamente as características que definem as ações de valor e esse resultado é um grande indício do destaque do investimento em valor face ao investimento em crescimento para o mercado acionário brasileiro, conforme observado por SANTOS (2010). O que também se mostra coerente com os estudos de baixos múltiplos de preço conduzidos por RAMOS, PIKANÇO e COSTA JR (2000), NAGANO, MERLO e SILVA (2003) e ROSTAGNO, SOARES e SOARES (2008).

Para avaliar os modelos construídos de reconhecimento de padrões, conferindo peso aos apontamentos feitos anteriormente, tem-se a Abordagem Lachenbruch. A Tabela 3 traz o índice médio de acerto das regras de classificação. Os percentuais

mostrados correspondem a quantas ações do grupo Sucesso (ou Fracasso) foram efetivamente classificadas como de Sucesso (ou Fracasso) pelo modelo. O que resta para 100% foi classificada de maneira incorreta pelo modelo.

TABELA 3: MÉDIA DO ÍNDICE DE ACERTO DAS FUNÇÕES DISCRIMINANTES PELA ABORDAGEM LACHENBRUCH

Grupo	FDLF <i>Stepwise</i>			MRLM <i>Stepwise</i>		
	Percentual correto	Class como Sucesso	Class como Fracasso	Percentual correto	Class como Sucesso	Class como Fracasso
Sucesso	15,54%	5,18	29,82	14,37%	4,55	30,55
Fracasso	97,62%	3,45	141,27	98,60%	1,91	142,73
Total	81,96%	8,64	171,09	82,30%	6,45	173,27

Os modelos tiveram avaliação bastante próxima, com um índice de acerto total de 81,96% para a FDLF e 82,30% para o MRLM. Eles se mostraram particularmente eficientes para discriminar as empresas sem rentabilidade acima do mercado (grupo Fracasso), acertando 97,62% e 98,60% das classificações, respectivamente. Também foram estáveis as medidas ano a ano, com um acerto mínimo de 68,1% e máximo de 91,7% de acerto total para a FDLF e 70,1% e 91,8% para o MRLM.

A maior concentração de erro dos modelos se deu ao classificar uma empresa Sucesso no grupo Fracasso. Na FDLF, 15,54% das empresas eram alocadas corretamente. No MRLM, o índice de acerto bastante próximo: 14,37%. A princípio, tal procedimento não é tão prejudicial para o investidor como seria classificar um alto índice de empresas do grupo Fracasso como Sucesso, apenas se mostra conservador ao alocar poucos ativos para o grupo sucesso.

Esse grau de acerto na avaliação é cerca de 15 pontos percentuais superior aos alcançados por TAVARES (2010) em estudo similar já citado. Tal diferença se deve, possivelmente, às diferentes variáveis de entrada do modelo (60% delas são diferentes) e à maneira de separação dos grupos: aqui com base em três medidas de rentabilidade em comparação com o Ibovespa, no estudo comentado uma divisão 50%-50% com base somente no retorno dos ativos.

Uma maneira de validar o modelo, ao mesmo tempo em que se propõe sua utilização como estratégia de investimento, é testar as regras de classificação para a montagem de portfólios nos anos posteriores à construção do modelo. Ao se fazer um estudo de carteira nesses portfólios tem-se as rentabilidades indicadas nas Tabelas 5 e 6. A média dos retornos foi calculada pelo logaritmo neperiano da razão entre o preço carteira no final do período pelo preço da carteira no início do período. O Índice de Sharpe foi calculado considerando o retorno médio da carteira, seu desvio padrão e a taxa livre de risco (tomada como a Taxa Selic). O Alfa de Jensen foi calculado considerando o retorno médio da carteira, o retorno médio do mercado (tomado como o Índice Bovespa), o coeficiente beta de ajuste da carteira em relação ao mercado e a taxa livre de risco (tomada como a Taxa Selic).

A média dos retornos das estratégias foi superior ao mercado em todos os períodos analisados. Na carteira de um ano, o retorno médio máximo foi alcançado pelo Modelo de Regressão Logística Múltipla, 25,18%. Contudo, nenhum dos valores se mostrou significativo.

TABELA 5: MÉDIA DO RETORNO, ÍNDICE DE SHARPE E ALFA DE JENSEN PARA DIFERENTES PERÍODOS – FDLF *STEPWISE*

	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	18,49%	16,05%	0,0415	0,0140	-6,62%
2 anos	49,12%	36,24%	0,0427	0,0155	**36,44%
3 anos	79,32%	56,88%	0,0401	0,0136	**76,86%
5 anos	120,94%	101,78%	*0,0360	0,0149	**76,55%

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

** valores estatisticamente positivos/superiores ao Ibovespa no Wilcoxon-Mann-Whitney

TABELA 6: MÉDIA DO RETORNO, ÍNDICE DE SHARPE E ALFA DE JENSEN PARA DIFERENTES PERÍODOS – MRLM *STEPWISE*

	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	25,18%	16,05%	0,0509	0,0140	13,14%
2 anos	48,97%	36,24%	0,0343	0,0155	17,58%
3 anos	73,59%	56,88%	0,0281	0,0136	23,13%
5 anos	114,25%	101,78%	0,0267	0,0149	*23,98%

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

** valores estatisticamente positivos/superiores ao Ibovespa no Wilcoxon-Mann-Whitney

O índice de Sharpe dos modelos também foi maior que o índice de Sharpe do Ibovespa em todos os períodos analisados, novamente com destaque para a estratégia MRML e sem significância estatística.

O alfa de Jensen se mostrou em todos os períodos, exceto para a carteira de um ano da FDLF. Entretanto, esse medidor, que não pode ser assumido como gaussiano nesse caso, foi significativamente positivo segundo o teste não paramétrico de Wilcoxon-Mann-Whitney para as demais janelas temporais desse mesmo modelo. No MRLM, o alfa de Jensen só foi significativo para a carteira de 5 anos.

Um fator que contribuiu para a dispersão dos retornos e, conseqüentemente, para menor sucesso dos testes de hipótese, foi a alta variabilidade da quantidade de ativos nas carteiras montadas nas duas estratégias. Mesmo com os bons resultados do modelo, em alguns anos muitos papéis foram selecionados, desfazendo o caráter conservador que os modelos possuíam quando avaliados sobre a base de dados em que foram construídos. Houve ano (2003) com somente um ativo selecionado, enquanto houve ano (2008) com 100 ações em carteira. Isso acarreta uma maior volatilidade dos retornos e baixos valores nas duas medidas ajustadas ao risco. Na média, a FDLF selecionou 23,8 ativos por ano e o MRLM selecionou 34,1.

Estudos do mercado brasileiro (CERETTA e COSTA JR, 2000; ASSAF NETO, 2006; OLIVEIRA e PAULA, 2008) têm indicado a quantidade ótima de ativos para o investidor médio entre 10 e 12. Assim, é possível utilizar os escores das funções discriminantes para se selecionar, por exemplo, apenas os 12 ativos de maior escore do grupo Sucesso, criando sempre portfólios com um número equilibrado de ações.

Procedendo dessa maneira, têm-se um novo estudo de carteira, cujas medidas de rentabilidade estão sintetizadas nas tabelas 7 e 8.

A inclusão das 12 primeiras ações melhorou a rentabilidade do modelo em praticamente todas as medidas, em comparação com o modelo anterior e, novamente, supera as marcas de mercado, mas nem sempre de forma significativa. A carteira de um ano teve desempenho bastante parecido nas duas estratégias com 12 ativos, 23,89% de retorno e 12,01% de alfa de Jensen para a FDLF e 23,62% de retorno e 12,16% para o alfa de Jensen do MRLM.

TABELA 7: MÉDIA DO RETORNO, ÍNDICE DE SHARPE E ALFA DE JENSEN PARA DIFERENTES PERÍODOS – 12 ATIVOS FDLF *STEPWISE*

	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	23,89%	16,05%	*0,0590	0,0140	*12,01%
2 anos	45,36%	36,24%	0,0422	0,0155	13,41%
3 anos	72,41%	56,88%	0,0382	0,0136	*22,31%
5 anos	122,68%	101,78%	0,0354	0,0149	*35,14%

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

** valores estatisticamente positivos/superiores ao Ibovespa no Wilcoxon-Mann-Whitney

TABELA 8: MÉDIA DO RETORNO, ÍNDICE DE SHARPE E ALFA DE JENSEN PARA DIFERENTES PERÍODOS – 12 ATIVOS MRLM *STEPWISE*

	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	23,62%	16,05%	0,0472	0,0140	*12,16%
2 anos	41,11%	36,24%	0,0342	0,0155	*11,17%
3 anos	64,87%	56,88%	0,0277	0,0136	15,20%
5 anos	110,29%	101,78%	0,0245	0,0149	*24,74%

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

** valores estatisticamente positivos/superiores ao Ibovespa no Wilcoxon-Mann-Whitney

O Índice de Sharpe foi significativo na carteira de um ano da FDLF, já o alfa de Jensen foi significativamente superior a zero em três dos quatros períodos para cada um dos modelos.

Esses resultados significativos nas medidas ajustadas ao risco, em especial no alfa de Jensen, dão indícios da existência de informações nos balanços contábeis das empresas suficientes para separá-las classificá-las de acordo com sua rentabilidade e, assim, gerar estratégias capazes de identificar ações como potencialmente bons investimentos.

Além disso, apesar de pequenas diferenças na escolha de variáveis entre os modelos, a tônica para separar ações vencedoras das demais se dá pela escolha de investimentos em valor. Tanto que as ações selecionadas para os modelos são bastante estáveis, sendo recorrentes as presenças de empresas como Petrobrás (PETR4), Vale (VALE5), Telesp/Telefônica (TLPP4), Ambev (AMBV3), Itausa (ITSA3), AES Tietê (GETI4), Eletrobrás (ELET3), Companhia Siderúrgica Nacional (CSN3), Embraer (3), Telemar (TNLP4) e Telemar Norte Leste (TMAR5) entre as escolhidas nas 10 carteiras montadas entre 2000 e 2009. O quadro 9 traz os principais ativos selecionados modelos pelas técnicas de reconhecimento de padrões.

QUADRO 9 – PRINCIPAIS AÇÕES SELECIONADAS PELOS MODELOS FDLF E MRLM 12 STEPWISE

ATIVOS	FLFD	MRLM	Total
PETR4	7	6	13
VALE5	6	7	13
TLPP4	6	7	13
AMBV4	4	4	8
ITSA3	2	5	7
GETI4	3	2	5
ELET3	2	3	5
CSNA3	2	3	5
EMBR3	1	3	4
TMAR5	1	3	4
TNLP4	1	3	4

Desse modo, mostra-se que há espaço para a aplicação de técnicas multivariadas de reconhecimento de padrões para esclarecer a dinâmica de funcionamento do mercado acionário brasileiro e, também, como estratégia para a construção de portfólios com rentabilidade acima do mercado. Sem elas, de modo apenas univariado, não seria possível comparar da maneira como foi feito a influência de diversas variáveis e determinar as de maior relevância para a construção dos modelos. Além disso, o uso dos escores das funções discriminantes são uma opção para a montagem de portfólios com o grau de diversificação que for desejado, contornando o problema de excesso ou escassez de ativos.

Entretanto, não se pode afirmar categoricamente que a Hipótese de Mercado Eficiente não é válida. Isso porque, ao investigá-la, é preciso tomar como opção um ou mais modelos de mensuração de risco, que, como dito, também acabam sendo testados. Por isso, em geral, não é possível fazer afirmações conclusivas sobre se o mercado é ineficiente ou se o modelo de mensuração de risco que é inadequado. 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 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 sua inaplicabilidade.

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.

Para refinar os modelos é possível utilizar outras técnicas de seleção de variáveis, como *backward stepwise*, ou modificar a forma de definição dos grupos Sucesso e Fracasso, que foi restritiva ao se considerar as três medidas simultaneamente. Mesmo porque, na composição de um portfólio, modifica-se o risco individual de cada ativo, com algumas combinações de ativos podendo ser mais favorecidas que outras.

Como alternativa às técnicas de reconhecimento de padrões empregadas, há o uso de redes neurais artificiais, utilizada com sucesso em diversos campos similares, máquinas de vetor de suporte (SVM) e árvores de decisão baseadas na entropia da informação.

Na mesma linha, sugere-se o uso da teoria dos conjuntos aproximativos como ferramenta de *data mining*, na qual se pode utilizar os conceitos de núcleo e redução para identificar as variáveis mais significativas na separação de grupos e, a partir delas, adotar regras de decisão. Uma das vantagens dos conjuntos aproximativos é que eles não necessitam de nenhuma informação preliminar ou adicional sobre os dados. Além dessa característica, pode-se fazer um tratamento quantitativo da incerteza, se utilizar métricas estatísticas para avaliar a importância das regras e combiná-los com outras arquiteturas, como a redes neurais.

A análise realizada se baseou no estudo de carteiras ano a ano, ou período ao período, em separado. Com isso, não se levou em consideração uma possível correlação entre as informações de um ano para outro dentro de uma mesma empresa. Essa estrutura temporal também pode ter rico potencial discriminatório, sendo um interessante caminho de pesquisa usar da teoria de modelagem de dados longitudinais para a construção de apenas um modelo simultaneamente para todos os anos.

Por fim, a divisão de ações entre rentáveis ou não foi bastante conservadora, considerando as rentáveis somente as que apresentavam rentabilidade acima do mercado

para as três medidas. Embora seja uma medida bastante sensata para o investidor, se poderia trabalhar com mais informações se essa classificação fosse feita em mais de 2 categorias. Uma alternativa seria criar 4 categorias, em que a primeira delas seria o grupo de ações que não apresentam rentabilidade acima do mercado para nenhuma das três medidas, a segunda, para apenas duas medidas, e assim por diante. Neste caso, um Modelo Logístico Multinomial e as próprias técnicas de Análise Discriminante poderiam ser usados no processo de classificação.

A exclusão das empresas que não continham toda a base de dados completas também é uma decisão conservadora que não precisa ser seguida em trabalhos futuros, uma vez que os métodos de classificação são capazes de lidar com esse tipo de situação e classificar os dados utilizando toda a informação disponível. O mesmo pode ser dito da retirada das empresas consideradas *outliers*, que não necessariamente precisariam ser excluídas. Uma possibilidade é proceder o estudo com essas observações atípicas e verificar se elas causam problemas de mensuração. Além disso, existem métodos robustos à presença de *outliers* que podem ser utilizadas, como o Modelos de Misturas.

Considerações finais

Nesse trabalho apresentou-se uma forma de se utilizar técnicas estatísticas multivariadas para reconhecer e classificar empresas que estejam possivelmente subavaliadas pelo mercado. Tal processo se baseia no uso de modelos de reconhecimento de padrões, nos conceitos teóricos da escola fundamentalista e, particularmente, no método de múltiplos, com a identificação de variáveis explicativas entre índices bursáteis e econômico-financeiros, todos publicamente disponíveis. Com isso, abre-se a oportunidade de se testar também a Hipótese de Mercado Eficiente em sua forma semiforte no mercado acionário brasileiro.

As técnicas de reconhecimento de padrões empregadas, Função Discriminante Linear de Fisher e Modelo de Regressão Logística Múltipla, foram aplicadas pelo método *forward stepwise*. Os resultados da avaliação de Lachenbruch foram satisfatórios, com um índice de acerto na classificação de cerca de 82% nos dois casos. Entretanto, as funções se mostraram bastante conservadores para designar uma ação como de sucesso, estando aí a maior concentração de erros das funções discriminantes, o que não é de todo prejudicial, mas pode acarretar problemas de diversificação.

Pelos resultados desses modelos, que foram muitos semelhantes, extrai-se que basicamente três variáveis são as capazes de diferenciar significativamente as ações com retorno acima do mercado das demais: o Preço de Mercado, a Taxa de Dividendos e o Preço por Valor Contábil. Também com alguma relevância, há a presença de outro múltiplo, a variável Preço por Capital de Giro Líquido. Diante dessas variáveis e da classificação das ações nos grupos Sucesso e Fracasso, observou-se que a designação de sucesso é frequentemente atribuída às empresas de valor, em contraposição às empresas de crescimento.

Um estudo de carteira entre os anos de 2000 e 2009 forneceu solidez para esse argumento, uma vez que os retornos obtidos foram superiores ao mercado, entretanto nem sempre houve significância estatística nesses resultados. Um dos problemas encontrados nesses portfólios foi a grande variabilidade do número de ativos, podendo incluir apenas uma ou mais de 100 ações em carteira. Para solucionar essa questão, usou-se os escores obtidos a partir das funções discriminantes de modo a incluir somente as 12 ações melhor ranqueadas a cada ano em cada um dos dois modelos. Com isso, os resultados foram um pouco superiores e elimina-se o problema de excesso ou escassez de ativos em carteira.

Com as carteiras formadas por 12 ações, para o período de um ano, a média do retorno logarítmico da FDLF foi de 23,89%, do MRLM de 23,62% enquanto do

Ibovespa foi de 16,05%, porém o resultado ainda não foi significativo. Por outro lado, o alfa de Jensen das duas estratégias e o índice de Sharpe da FDLF tiveram valores estatisticamente superiores ao do mercado. No caso do alfa de Jensen, significativo em mais períodos, houve um retorno anormal de 12,01% nas carteiras refeitas anualmente quando utilizada a função de Fisher e de 12,16% com o uso de regressão múltipla.

Os valores próximos de avaliação e de rentabilidade mostram certa equivalência entre os dois modelos de reconhecimento de padrões. As medidas de retorno ajustadas ao risco, em especial o alfa de Jensen, estatisticamente significativas e superiores ao mercado apontam, num primeiro momento, para ineficiências no mercado acionário brasileiro, uma vez que as estratégias propostas permitem identificar, de maneira sistemática, ativos subavaliados pelo mercado.

Para embasar esses resultados e se aprofundar na discussão da Hipótese de Mercado Eficiente, sugere-se, em trabalhos futuros, a utilização de formas de mensuração de risco diversas das utilizadas neste trabalho e do emprego de outras técnicas de reconhecimento de padrões, como redes neurais, a partir da análise fundamentalista.

Referências

- [1] 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, 2010.
- [2] ASSAF NETO, A. **Mercado Financeiro**. 7 ed. São Paulo: Atlas, 2006.
- [3] BRUNI, A. L. **Risco, retorno e equilíbrio: uma análise do modelo de precificação de ativos financeiros na avaliação de ações negociadas na Bovespa**. São Paulo, 1998. 173 p. Dissertação (Mestrado) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.
- [4] BUENO, A. F. Os Dividendos Como Estratégia de Investimentos em Ações. **Revista Contabilidade & Finanças**, v. 28, n.1, p. 39-55, abr./2002.
- [5] CERETTA, P. S.; COSTA JR, N. C. A. Quantas ações tornam um portfólio diversificado no mercado de capitais brasileiro? In: 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. p. 19-33.
- [6] 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.
- [7] 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.
- [8] 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, mar./2000.
- [9] DAMODARAN, A. **Investment Valuation**. 2 ed. New York: John Wiley and Sons, 2002.

- [10] _____. **Filosofias de Investimento**. Rio de Janeiro: Qualitymark, 2006.
- [11] FAMA, E. F. Efficient capital markets: a review of theory and empirical work. **Journal of Finance**, v. 25, n.2, p. 383-417, mai./1970.
- [12] FAVERO, L. P.; BELFIORE, P.; SILVA, F. L.; CHAN, B. **Análise de dados: Modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.
- [13] 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**, São Caetano do Sul, v.25, n.75, set./dez. 2009.
- [14] GRAHAM, B. **O Investidor Inteligente**. Rio de Janeiro: Nova Fronteira, 2007.
- [15] GRAHAM, B; DODD, D. L. **Security Analysis**. 3. ed. New York: McGraw-Hill, 1951.
- [16] GUIMARÃES, I. A. **Construção e avaliação de uma regra de reconhecimento e classificação de clientes de instituição financeira com base na Análise Multivariada**. Curitiba, 2000, 115 f. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná.
- [17] HAIR, J.F.; ANDERSON, R.E.; TATHAM R.L. **Multivariate Data Analysis**. New York: Macmillan, 1987.
- [18] HOSMER, D. W; LEMESHOW, S. **Applied Logistic Regression**. Amherst: John Wiley & Sons, 1989.
- [19] JOHNSON, R. A; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 4. ed. New Jersey: Prentice Hall, 1998.
- [20] LIMA, L. A. O. Auge e Declínio da Hipótese dos Mercados Eficientes. **Revista de Economia Política**, São Paulo, v. 23, n. 4 (92), out./dez. 2003.
- [21] MARDIA, K. V; KENT, J. T; BIBBY, J. M. **Multivariate Analysis**. Londres: Academic Press, 1979.
- [22] MELLAGI FILHO, A.; ISHIKAWA, S. **Mercado Financeiro e de Capitais**. 2. ed. São Paulo: Atlas, 2003.
- [23] NAGANO, M. S; MERLO, E. M; SILVA, M. C. As Variáveis Fundamentalistas e seus Impactos na taxa de Retorno de Ações no Brasil. **Revista da FAE**, Curitiba, v. 6, n. 2, p. 13-28, 2003.
- [24] OLIVEIRA, F.; PAULA, E. Determinando o Grau Ótimo de Diversificação para Investidores Usuários de *Home Brokers*. **Revista Brasileira de Finanças**, Rio de Janeiro, v. 6, n.3, p. 437-461, set./dez. 2008.
- [25] PÓVOA, A. **Valuation – Como Precificar Ações**. 2. ed. São Paulo: Globo, 2007.
- [26] 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.

- [27] ROSTAGNO, L. M.; SOARES, K; T; C; SOARES, R. O. O Perfil Fundamentalista das Carteiras Vencedoras e Perdedoras na Bovespa no Período de 1995 a 2002. **Brazilian Business Review**, vol. 5, n. 3, p.271-288 , dez./2008.
- [28] SANTOS, L. R. **Aplicação de estratégias de *value investing* no mercado acionário brasileiro**. Rio de Janeiro, 2010. 85 f. Dissertação (Mestrado Profissionalizante em Administração) – Programa de Pós Graduação em Administração e Economia, Faculdade de Economia e Finanças Ibmec.
- [29] SHARPE, W. F. Asset allocation: management style and performance measurement. **The Journal of Portfolio Management**, v. 18, n.2, p. 7-19, jul./1992.
- [30] SILVA, L. A. F. **A verificação das relações entre estratégias de investimento e as hipóteses de eficiência de mercado** – um estudo na bolsa de valores de São Paulo. São Paulo, 2003. 110 f. Tese (Doutorado em Engenharia de Produção) – Escola Politécnica, Universidade de São Paulo.
- [31] TAVARES, A. L. **A eficiência da análise financeira fundamentalista na previsão de variações no valor da empresa**. Natal, 2010, 184 f. Tese (Doutorado em Ciências Contábeis), Programa de Doutorado Multinstitucional e Inter-regional de Pós Graduação em Ciências Contábeis, Universidade Federal de Brasília, Universidade Federal da Paraíba e Universidade Federal do Rio Grande do Norte.

Anexo – Definição dos indicadores utilizados

1. Preço de Mercado (PM) – multiplicação da quantidade de ações pela cotação do ativo em bolsa, resultando no valor de mercado da empresa.
2. Taxa de Dividendos (DY) – indica a remuneração obtida em forma de proventos sobre o capital investido do acionista.

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

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.

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

7. Preço/Capital de Giro Líquido (P/CGL) – divisão entre o valor de mercado da empresa e seu capital de giro líquido, entendido como ativo circulante menos dívida total.

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

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

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

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 dos lucros por ação da empresa nos últimos 5 anos.
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 total e o passivo total.
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.