

**Área Temática: Finanças**

**TÍTULO: Identificação de Empresas Comparáveis Através da Análise de *Clusters*: uma Aplicação no Mercado Acionário Brasileiro**

**AUTORES**

**RAFAEL CAMPOS ROLIM**

Universidade Federal de Minas Gerais  
rolim\_ufla@yahoo.com.br

**AURELIANO ANGEL BRESSAN**

Universidade Federal de Minas Gerais  
bressan@face.ufmg.br

**JOÃO GABRIEL BERNARDES DE FREITAS**

Universidade Federal de Minas Gerais  
joaogabriel197@hotmail.com

**ANDREI SALEM GONÇALVES**

Universidade Federal de Minas Gerais  
andreisalem@hotmail.com

**Resumo**

A Avaliação Relativa busca avaliar determinado ativo ou empresa, com base na precificação de ativos similares no mercado. Ou seja, busca-se possibilidades de investimento comparáveis àquela escolhida pelo investidor, a fim de se determinar uma possível sub ou sobreavaliação. Duas questões são cruciais nesse tipo de método de avaliação de empresas e continuam em aberto na literatura, sendo elas: 1) a definição do grupo de empresas comparáveis e; 2) a identificação dos múltiplos a serem utilizados. O presente buscou avaliar a viabilidade de se identificar empresas comparáveis no mercado acionário brasileiro através de análise estatística de agrupamento por semelhança de fundamentos, como possível alternativa à usual classificação por setores econômicos, a fim de se realizar uma avaliação relativa. Para tal, utilizou-se como universo de estudo todas as empresas listadas e negociadas na BM&FBOVESPA no período compreendido entre Janeiro de 2003 e Dezembro de 2008. A metodologia utilizada para identificar empresas comparáveis foi a análise de *clusters*, realizada através do *software* STATA 10 SE. Com os resultados, observou-se que o agrupamento por similaridade de fundamentos diverge do agrupamento por setores econômicos. Assim, há evidências de que os setores econômicos do mercado acionário brasileiro não representam grupos de empresas comparáveis.

**Abstract**

Relative Valuation seeks to value an asset or business, based on the pricing of similar assets in the market. In other words, it seeks investment opportunities comparable to that chosen by the investor, in order to determine a possible under or overestimation. Two issues are crucial in this type of valuation and remain without answer in the literature: 1) defining the group of comparable companies and 2) the identification of multiples to be used. This paper aims to

identify comparable companies in the Brazilian stock market through statistical analysis of grouping by similarity of fundamentals, as an alternative to the usual classification by industry, in order to perform a Relative Valuation. To this end, this paper used as universe of study all companies listed and traded on the BM&FBovespa in the period between January 2003 and December 2008. The methodology used to identify comparable companies was the cluster analysis, performed using STATA-SE 10. With the results, we found that grouping by similarity of fundamentals differs from the grouping by industry. Thus, there is evidence that the industries of the Brazilian stock market do not represent groups of comparable companies.

**Palavras-chave:** Avaliação Relativa, Empresas Comparáveis, Análise de *Clusters*.

## 1. INTRODUÇÃO

Diferentes formas de avaliação de empresas são utilizadas diariamente e tornam-se centro de discussões e debates no meio acadêmico e no mercado. Ainda que existam modelos mais aplicados em determinadas ocasiões, é certo que não há um consenso geral sobre qual o método mais adequado para se avaliar empresas.

A eficiência dos métodos de *valuation* pode variar de mercado para mercado, bem como no decorrer dos anos. Ainda que exista a tentativa de se aplicar os métodos da maneira mais objetiva possível, os aspectos subjetivos acabam sempre existindo em alguma etapa da avaliação, independente do método utilizado.

Do ponto de vista de um investidor, ao se utilizar um modelo de avaliação de empresas, procura-se identificar empresas sub ou sobreavaliadas no mercado, a fim de se obter ganhos com as mesmas. Dessa forma, investidores estão sempre comparando a viabilidade de investimento em determinada empresa, em detrimento de outra. Como então comparar de maneira mais adequada duas possibilidades de investimento? Ou melhor, como então identificar quais dessas possibilidades podem realmente ser consideradas comparáveis?

O método de avaliação de empresas denominado Avaliação Relativa busca avaliar determinado ativo ou empresa, com base na precificação de ativos similares no mercado. Ou seja, busca-se possibilidades de investimento comparáveis àquela escolhida pelo investidor, a fim de se determinar uma possível sub ou sobreavaliação. Para se analisar de maneira adequada o grupo de empresas comparáveis encontrado, o processo de avaliação relativa padroniza os preços das mesmas em torno de uma variável comum, seja ela uma medida de lucro ou uma medida patrimonial, formando então os chamados múltiplos. Duas questões são cruciais nesse tipo de método de avaliação de empresas, sendo elas: 1) a definição do grupo de empresas comparáveis e; 2) a identificação dos múltiplos a serem utilizados.

Como serão revisadas no decorrer do estudo, estas duas questões continuam em aberto na literatura e não há, ainda, consenso sobre a melhor maneira de tratá-las. Muitos estudos consideram o setor econômico como critério para identificação de empresas comparáveis, mas a eficácia dessa prática vem sendo contestada. Dessa forma, a necessidade da realização de estudos que busquem confrontar diferentes maneiras de se identificar grupos de empresas similares, bem como possíveis múltiplos a serem utilizados na avaliação relativa, levando em consideração as especificidades de diferentes mercados, torna-se evidente.

Nessa perspectiva, emerge o problema de pesquisa do presente estudo, relacionado à primeira questão do processo de avaliação relativa: *É possível identificar empresas comparáveis no mercado acionário brasileiro através de técnicas estatísticas de semelhança de fundamentos como alternativa à usual classificação por setor econômico?*

Com base no problema de pesquisa proposto, o trabalho pretende então atingir aos seguintes objetivos:

### 1.1. Objetivo Geral

- Avaliar a viabilidade de se identificar empresas comparáveis no mercado acionário brasileiro através de análise estatística de agrupamento por semelhança de fundamentos, a fim de se realizar uma avaliação relativa.

### 1.2. Objetivos Específicos

- Desenvolver revisão teórica sobre estudos que buscam identificar empresas comparáveis.
- Definir os fundamentos a serem utilizados como base para identificação de empresas semelhantes.
- Identificar a técnica estatística mais adequada para agrupamento de empresas.
- Avaliar a convergência entre o agrupamento por semelhança de fundamentos e o critério de classificação por setores econômicos.

## 2. REFERENCIAL TEÓRICO

Tendo em vista os objetivos propostos, é necessária a revisão teórica sobre os principais estudos sobre avaliação relativa e, mais especificamente, sobre aqueles que buscaram identificar empresas comparáveis, seja por técnicas estatísticas ou por agrupamento por setor econômico.

### 2.1. Avaliação Relativa

A avaliação relativa constitui-se em uma alternativa para avaliação de empresas ou ativos e sua idéia básica consiste na avaliação de um ativo com base na precificação de ativos similares no mercado. Conforme Damodaran (2007), existem três passos essenciais na avaliação relativa.

O primeiro deles consiste em identificar empresas comparáveis. Para que a avaliação relativa possa ser realizada, é necessário que o grupo de empresas a ser avaliado seja composto, efetivamente, por empresas comparáveis. Uma alternativa para esta classificação é a utilização de empresas do mesmo setor econômico como comparáveis, sob a perspectiva de que a melhor maneira de se encontrar empresas similares àquela sendo avaliada é analisando o próprio setor em que ela se insere. Ainda que tal critério de classificação permita a utilização de múltiplos específicos daquele segmento de mercado, enriquecendo a análise, ele é passível de questionamento pela possível existência de empresas com estruturas e características muito divergentes dentro de um mesmo setor.

Tendo em vista esta desvantagem, Damodaran (2006) sugere outra possibilidade de identificação de empresas comparáveis, baseando-se na proposição de que empresas comparáveis são aquelas com características de risco, crescimento nos lucros e geração de fluxo de caixa semelhantes, e não aquelas que se situam no mesmo setor econômico. Dessa forma, o autor defende que uma avaliação baseada unicamente na segmentação por setores econômicos – desconsiderando a similaridade de tais fundamentos – pode comprometer a análise. O autor justifica seu argumento exemplificando a dificuldade de se realizar tal avaliação em um setor com relativamente poucas empresas ou quando uma empresa atua em diferentes setores (DAMODARAN, 2007).

Após identificar os grupos de empresas comparáveis, a segunda etapa do processo de avaliação relativa consiste na padronização dos preços de mercado a uma variável comum. Damodaran (2006) afirma que, como as empresas a ser comparadas não são idênticas, é necessário que seus preços estejam padronizados em relação a algum fundamento como lucro, valor patrimonial ou vendas, por exemplo, obtendo-se os chamados múltiplos. Estes múltiplos abordados na avaliação relativa podem ser definidos, conforme Martelanc *et al* (2004) *apud* Pasin (2004), como indicadores padronizados dos valores das empresas e utilizados para avaliar uma empresa comparativamente a empresas similares. Kim e Ritter (1999) argumentam que a discussão sobre quais múltiplos são os melhores direcionadores de valor das empresas também pode ser considerada uma questão em aberto na literatura, com resultados que variam em diferentes mercados. Assim, pode-se considerar que não existe uma resposta clara sobre quais múltiplos são mais bem ajustados ao valor da empresa, devendo ser utilizados na avaliação.

Por fim, após se identificar empresas comparáveis e padronizar seus preços em torno de uma variável comum, Damodaran (2006) aponta como última etapa do processo de avaliação relativa o ajuste das diferenças entre os ativos. Estes ajustes podem ser realizados, conforme o autor, de três maneiras:

- 1) Ajustes subjetivos
- 2) Múltiplos modificados
- 3) Técnicas estatísticas

Explicitando, os ajustes subjetivos decorrem da utilização da experiência do analista para julgar as características da empresa que podem explicar a diferença obtida entre os múltiplos da empresa avaliada e de suas comparáveis. Já os múltiplos modificados são utilizados para controlar diferenças na variável que apresenta maior divergência entre a empresa avaliada e suas comparáveis. As técnicas estatísticas, por sua vez, podem ser utilizadas para indicar a intensidade da relação entre o múltiplo e as variáveis utilizadas, quando esta relação se torna complexa e dificulta a aplicação das duas primeiras maneiras.

Conforme desenvolvido anteriormente, o presente estudo possui como foco a primeira etapa do processo de avaliação relativa; ou seja, a identificação de empresas comparáveis. Busca-se uma possível alternativa à classificação de empresas comparáveis através do setor econômico ao qual pertencem. Dessa forma, é importante expor trabalhos que estudaram o tema, bem como as vantagens e desvantagens de cada abordagem.

## **2.2. Identificação de Empresas Comparáveis**

Uma das questões-chave na avaliação relativa para construção de modelos é a definição das empresas comparáveis e dos múltiplos a serem utilizados [Pasin (2004), Damodaran (2006)]. Nota-se na literatura uma discussão ainda em aberto sobre como identificar tais empresas comparáveis. Ainda que alguns autores citem técnicas estatísticas como possibilidade, a maioria dos estudos adota a classificação por setor econômico como critério, mesmo que seus resultados não sejam satisfatórios.

Kim e Ritter (1999) desenvolveram um estudo avaliando o uso de múltiplos em conjunto com informações contábeis em processos de avaliação de ofertas públicas iniciais. Os autores utilizaram uma amostra de cento e noventa ofertas públicas realizadas entre os anos de 1992 e 1993. Para avaliar as empresas, foram utilizados dois critérios de classificação de empresas comparáveis. Primeiramente, os IPOs a ser avaliados foram comparados com empresas inseridas no mesmo setor econômico que o seu, que haviam realizado suas ofertas públicas em, no máximo, doze meses precedentes ao IPO da empresa analisada. Além disso, os autores testaram uma forma alternativa de classificação ao critério de agrupamento por setores econômicos. Com base em técnicas estatísticas, foram classificadas como comparáveis, aquelas empresas que possuíam múltiplos similares, segundo um modelo de regressão utilizado por um grupo de pesquisa especializado em avaliar IPOs. Dentre os dois tipos de classificação adotados, a utilização de técnicas estatísticas apresentou melhores resultados, fortalecendo a necessidade de se questionar o processo de classificação por setores econômicos.

Liu, Nissim e Thomas (2002) apesar de não adotar algum critério alternativo de identificação de empresas comparáveis, utilizam e questionam a confiabilidade da adoção do critério de classificação por setores econômicos. Os autores realizaram um estudo avaliando o desempenho de um conjunto de múltiplos como direcionadores de valor em procedimentos de avaliação de empresas. O critério adotado foi o de classificação por setores econômicos e os resultados contrariaram a visão popular de que diferentes setores possuem diferentes “melhores múltiplos”, levando os autores a contestar essa usual segmentação.

Questionando também a identificação de empresas comparáveis por setor econômico, Bhojraj e Lee (2002) desenvolvem um estudo que apresenta uma abordagem alternativa de classificação de empresas. Os autores seguem a linha defendida por Damodaran (2006) de que empresas comparáveis são aquelas que apresentam características de risco, rentabilidade e crescimento similares, e não aquelas que se agrupam em determinado segmento de mercado. Utilizando os múltiplos P/VPA e EV/Vendas e identificando empresas comparáveis através da similaridade das três características apresentadas acima, os autores refutam a superioridade da classificação por setores econômicos. A escolha de empresas comparáveis foi realizada através da obtenção de um múltiplo ajustado para cada empresa e consequente identificação de empresas com múltiplos próximos ao da empresa em análise.

A discussão sobre identificação de empresas comparáveis pode se aprofundar mais ainda, como demonstrado por Bhojraj, Lee e Oler (2003). Os autores demonstram que, mesmo após adotar um critério de classificação, diferentes resultados podem ser obtidos. Explicitando, o estudo parte da classificação por setores econômicos e avalia quatro diferentes “sub-critérios” para este tipo de classificação, disponíveis no mercado. Conforme esperado, os resultados divergiram na medida em que “sub-critérios” diferentes foram utilizados. A importância do estudo se dá na demonstração de que, mesmo depois de identificado o critério a ser utilizado, deve-se atentar para possíveis variações disponíveis.

No Brasil, Pasin (2004) também desenvolve um estudo com segmentação por região e setor econômico. O autor desenvolveu um modelo de estimação relativa baseado em indicadores contábeis do valor de empresas não financeiras negociadas nas principais bolsas do mundo. Seus resultados indicaram o valor da empresa fortemente correlacionado com EBTIDA e ativo total. Apesar da classificação por país e setor econômico, o autor ressalta que a

identificação de empresas comparáveis através fundamentos contábeis similares pode ser uma alternativa viável à classificação adotada em seu estudo, corroborando então a idéia dos trabalhos de Damodaran (2006) e Kim e Ritter (1999).

Como pode ser observado, embora os setores econômicos tenham sido amplamente utilizados para se identificar empresas comparáveis, muitos estudos já apontam para a busca de novos critérios de classificação, tendo em vista a baixa eficácia dos resultados obtidos perante o agrupamento por setores.

Por fim, a discussão em aberto existente na literatura sobre questões cruciais na avaliação relativa pode ser resumida no trabalho de Bhoraj e Lee (2002):

(...) com raras exceções, a literatura de contabilidade e finanças contém poucas evidências sobre como e por que determinados múltiplos individuais, ou determinadas empresas comparáveis, devem ser selecionadas em contextos específicos.

### 3. METODOLOGIA

#### 3.1. Universo e Amostra de Estudo

O universo do presente estudo compreende todas as empresas listadas e negociadas na BM&FBOVESPA no período compreendido entre Janeiro de 2003 e Dezembro de 2008. Inicialmente, tentou-se utilizar o período de 1995 a 2008, tendo em vista que períodos anteriores ao Plano Real eram marcados por alta inflação e oscilações abruptas nas taxas nominal de juros e cambial. Contudo, o estudo se tornou inviável para os anos de 1995 a 2002, devido à reduzida quantidade de informações disponíveis sobre os fundamentos de empresas no período.

Foram utilizadas duas amostras compostas apenas por ações ordinárias das empresas listadas e ativas em bolsa nos seus respectivos anos. A primeira amostra foi composta por empresas que participaram de, no mínimo, um pregão no ano. Já na segunda amostra, foram excluídas empresas que apresentaram negociações em menos de 50% dos pregões a cada ano, obtendo-se um menor número de empresas.

Como o estudo representa a primeira etapa de um processo de avaliação relativa, também não fizeram parte da amostra final empresas que não possuíam informações disponíveis sobre os fundamentos utilizados para se identificar empresas comparáveis. As variáveis utilizadas para se determinar a similaridade entre as empresas obtidas na amostra, possibilitando a análise de agrupamento, foram: *Payout*, Beta e Taxa de Crescimento nos Lucros por Ação. Tal escolha foi baseada em Damodaran (2006, p. 65), que propôs que “*Uma empresa comparável é aquela que possui características de fluxo de caixa, potencial de crescimento e risco similares à empresa sendo avaliada*”. As *proxies* utilizadas para risco (Beta), crescimento (Taxa de Crescimento nos Lucros por Ação) e fluxo de caixa (*payout*) também foram utilizadas tendo por base o trabalho desenvolvido por Damodaran (2006).

Desta forma, a relação entre o universo de estudo e as amostras finais pode ser visualizada na tabela 1. A amostra final 1 engloba as empresas que participaram de, pelo menos, um pregão no ano. Já a amostra final 2 corresponde àquela em que foram excluídas empresas que apresentaram negociações em menos de 50% dos pregões a cada ano.

TABELA 1: Universo e Amostras de Estudo.

Ano	Universo de Estudo	Amostra Final 1	Amostra Final 2
2003	227	68	46
2004	245	74	54
2005	247	74	54
2006	250	73	53
2007	322	84	65
2008	316	51	44

Fonte: Elaborada pelos autores.

### 3.2. Coleta de Dados

Os dados relacionados às *proxies* para risco (beta), crescimento (taxa de crescimento nos lucros por ação) e fluxo de caixa (*payout*), foram coletados no *software* Economática. Entretanto, ressalta-se que as informações relacionadas ao *payout* das empresas do ano de 2008 não estavam disponíveis no *software* em questão, no momento da pesquisa. Assim, tais informações foram adquiridas através da base de dados do Instituto Nacional de Investidores (INI). Tal divergência para este ano específico não distorce a análise, uma vez que não se busca analisar o comportamento temporal do *payout* das empresas, mas sim utilizá-lo para formar grupos de empresas comparáveis, de maneira pontual no tempo.

### 3.3. Tratamento dos Dados

Para tratamento dos dados, utilizou-se metodologia estatística de análise de *clusters*, realizada através do *software* STATA 10 SE.

#### 3.3.1. Análise de Clusters

Segundo Anderberg (1973), a análise de *clusters* é um termo geral que abrange todo um segmento de estudos sobre a determinação de agrupamentos naturais em bases de dados. No contexto da avaliação relativa de empresas, a análise de *clusters* entra como uma alternativa à tradicional metodologia de agrupamento via setores econômicos. Essa metodologia de análise é bastante abrangente e, naturalmente, se divide em inúmeras formas de ser realizada.

Como ponto de partida para a análise de *clusters* é necessário definir uma medida de dissimilaridade (ou similaridade) a ser utilizada. Essa escolha é de extrema importância, pois a medida define o quão semelhantes são dois elementos de uma amostra. Existem inúmeras medidas de dissimilaridade, que podem ser encontradas em Johnson e Wichern (2001), e cada uma produz diferentes resultados. Dentre elas, a medida tradicionalmente mais utilizada é a Distância Euclidiana, a qual foi utilizada no presente estudo Ainda segundo Johnson e Wichern (2001), considerando dois elementos  $X_i$  e  $X_j$ , a distância euclidiana entre estes elementos é dada pela equação 1.

$$d(X_i, X_j) = \left[ (X_i - X_j)' \cdot (X_i - X_j) \right]^{\frac{1}{2}} \quad (1)$$

De posse da escolha da medida de dissimilaridade, outra importante decisão para realizar uma análise de *clusters* se refere à escolha da técnica de construção dos conglomerados. Essa técnica tem relevância fundamental, pois é a base conceitual do procedimento que será realizado na formação dos *clusters* e se divide basicamente em métodos hierárquicos e não hierárquicos.

A metodologia estatística de *clustering* utilizada foi a de técnicas de agrupamento hierárquicas. Essa escolha se deve principalmente devido às técnicas de agrupamento não hierárquicas necessitarem, de acordo com Mingoti (2007) e Jonhson e Wichern (2001), da prévia especificação do número  $g$  de *clusters*. Como o presente trabalho tem por objetivo determinar a quantidade de *clusters* a serem formados a cada ano, não é viável a pré-definição do número de tais *clusters*, impossibilitando a utilização de técnicas não-hierárquicas. Além disso, Jonhson e Wichern (2001) alegam também que os métodos não hierárquicos são mais designados para agrupar itens, em detrimento de variáveis, o que justifica a opção por métodos hierárquicos.

Dentre os métodos hierárquicos mais utilizados, foi escolhido o método da média das distâncias (*Average Linkage*). Mingoti (2007, p. 178) apresenta a seguinte discussão da utilização dos métodos de ligação simples (*Single Linkage*), da média das distâncias (*Average Linkage*) e de ligação completa (*Complete Linkage*):

A maioria dos métodos produz *clusters* esféricos ou elipsóides, com exceção do método de ligação simples, que tem a capacidade de gerar estruturas geométricas diferentes. No entanto, este método é incapaz de delinear grupos pouco separados(...). O método de ligação completa tende a produzir conglomerados de aproximadamente mesmo diâmetro, além de ter a tendência de isolar os valores discrepantes da amostra nos primeiros passos do agrupamento. O método da média das distâncias, por sua vez, tende a produzir conglomerados de aproximadamente mesma variância interna e, em geral, produz melhores partições que os métodos de ligação simples e completa.

Para prosseguir com o tratamento dos dados se faz necessária a realização de um procedimento estatístico de padronização das variáveis a fim de que a discrepância entre elas diminuam. Como as variáveis utilizadas no presente estudo não possuem mesma magnitude, padronizá-las se faz necessário e o método utilizado para tal foi o de padronização pela média e desvio padrão, proposto em (MINGOTI, 2007, p. 200).

A autora citada acima sugere também que outro fator que pode comprometer a análise de *clusters* é a presença de *outliers*. Dessa forma, é necessária a identificação e exclusão destes elementos discrepantes para sanar esta problemática. Utilizou-se a técnica de mineração de *outliers* baseada em estatística, cujos detalhes podem ser encontrados em Han e Kamber (2001), para se identificar e excluir os elementos discrepantes. De forma resumida, essa técnica pode ser descrita como um ajuste de distribuição aos dados e posterior exclusão de dados inverossímeis a um determinado nível de significância arbitrário (1% para a análise realizada).

Por fim, para a criação dos *clusters*, é necessário realizar testes para a detecção do número de grupos da partição natural dos dados. Diversos testes para detecção do número de *clusters* podem ser encontrados na literatura e Milligan e Cooper (1985) realizaram um estudo comparativo de 30 diferentes critérios a fim de se determinar o número  $g$  de clusters da

partição natural dos dados. Segundo o estudo, os testes que obtiveram melhores resultados foram os que utilizam as medidas pseudo-F e pseudo-T<sup>2</sup>. Assim, com base em tal justificativa, estes testes foram adotados para a identificação do número  $g$  de *clusters* a cada ano.

Além disso, existe um gráfico denominado Dendrograma que pode auxiliar na identificação do número  $g$  de *clusters* (Timm, 2002). Tal gráfico possibilita a observação do processo e auxilia na decisão do passo em que o algoritmo deve ser parado. Na prática, o Dendrograma auxilia na observação do número de *clusters* naturais que estão sendo formados ao longo do processo.

Finalmente, ao saber o número de *clusters* a serem formados, seja através de um procedimento de teste ou de análise do Dendrograma, define-se a composição dos *clusters*, verificando-se quais elementos são agrupados em conjunto. Para o contexto específico da avaliação relativa, após definir quais elementos (empresas) são agrupados em conjunto, tem-se as empresas que podem ser consideradas comparáveis e o número de grupos de empresas comparáveis é, conseqüentemente, o número de *clusters* formados.

#### 4. RESULTADOS E DISCUSSÕES

De posse da amostra final com exclusão de *outliers*, iniciou-se a análise dos resultados. Como descrito anteriormente, o método para mensuração de dissimilaridade adotado foi o da média das distâncias (*Average Linkage*). Os testes pseudo-F e pseudo-T<sup>2</sup> foram aplicados a cada ano a fim de se determinar o número  $g$  de grupos da partição natural dos dados.

Os resultados do teste pseudo-F foram inconclusivos quanto ao número de *clusters* para todos os anos. De acordo com Mingoti (2007), essa situação indica a não existência de uma partição natural dos dados, ou seja, segundo os resultados obtidos pelo teste pseudo-F, não haveria uma natural divisão das empresas em grupos, tornando inviável a análise de *clusters* para a base de dados utilizada. Em contrapartida, os resultados do teste pseudo-T<sup>2</sup> indicaram a existência de partição natural dos dados para todos os anos analisados, resultando em um número  $g_i$  de *clusters* a cada ano  $i$ . Todas as análises a serem apresentadas se referem aos resultados obtidos pela utilização do teste pseudo-T<sup>2</sup> para a definição do número  $g$  de *clusters*; ou seja, do número de grupos de empresas comparáveis.

Os resultados da análise de *cluster* para as duas amostras podem ser observados através da tabela 2. Vale ressaltar que, para a análise dos resultados, a amostra final 1 representa as empresas que participaram de, pelo menos, um pregão no ano e a amostra final 2 corresponde àquela em que foram excluídas empresas que apresentaram negociações em menos de 50% dos pregões a cada ano.

TABELA 2: Total de empresas por *cluster* durante o período de 2003 a 2008.

Clusters	Amostra Final 1						Amostra Final 2					
	2003	2004	2005	2006	2007	2008	2003	2004	2005	2006	2007	2008
Cluster 1	3	16	8	14	76	12	3	35	45	52	61	7
Cluster 2	12	37	1	2	7	13	39	9	3	1	1	1
Cluster 3	41	5	17	2	1	2	3	2	1		2	1
Cluster 4	4	12	18	12		1	1	1	3		1	1
Cluster 5	7	2	12	6		1		4	1			1
Cluster 6	1	1	11	9		1		2	1			1
Cluster 7		1	1	15		1		1				4
Cluster 8			2	1		5						1
Cluster 9			2	2		6						3
Cluster 10			1	2		1						1
Cluster 11			1	1		1						2
Cluster 12				2		1						6
Cluster 13				1		1						1
Cluster 14				1		1						4
Cluster 15				2		1						1
Cluster 16				1		2						1
Cluster 17						1						1
Cluster 18												1
Cluster 19												1
Cluster 20												1
Cluster 21												1
Cluster 22												2
Cluster 23												1
<b>Total de Clusters</b>	6	7	11	16	3	17	4	7	6	2	4	23
<b>Total de Empresas</b>	68	74	74	73	84	51	46	54	54	53	65	44

Fonte: Elaborada pelos autores.

De maneira geral, pode-se observar que o número de empresas nas amostras cresce no decorrer dos anos, com uma queda expressiva no ano de 2008. O baixo número de empresas neste ano se justifica pela utilização de uma base de dados alternativa para obtenção do indicador *payout* das empresas. Como exposto anteriormente, o *payout* do ano de 2008 não estava disponível no *software* Economática, no período da coleta de dados para o presente estudo. Já o número de *clusters* varia no decorrer dos anos, para as duas amostras, não apresentando um padrão específico.

Para dar continuidade às análises é necessário apresentar uma comparação entre os resultados obtidos com a análise de *clusters* e a tradicional metodologia de segmentação de empresas por setor econômico. Para tal, utilizou-se a classificação setorial disponível na BM&FBovespa. Os setores existentes utilizando este critério foram:

- 1) Bens Industriais
- 2) Construção e Transporte
- 3) Consumo Cíclico
- 4) Consumo Não Cíclico
- 5) Financeiro e Outros
- 6) Materiais Básicos
- 7) Petróleo, Gás e Biocombustíveis
- 8) Tecnologia da Informação
- 9) Telecomunicações
- 10) Utilidade Pública

Informações sobre os segmentos e subsetores que compõem cada setor podem ser adquiridas no portal da BM&FBovespa. Os resultados da classificação por setor econômico podem ser visualizados na tabela 3:

TABELA 3: Total de empresas por setor econômico durante o período de 2003 a 2008.

Setor	Amostra Final 1						Amostra Final 2					
	2003	2004	2005	2006	2007	2008	2003	2004	2005	2006	2007	2008
Bens Industriais	7	7	7	7	8	2	4	3	2	3	6	2
Construção e Transporte	1	1	1	3	3	2	1	1	0	3	3	2
Consumo Cíclico	4	6	6	6	8	8	2	4	4	4	6	5
Consumo Não Cíclico	5	4	3	5	6	1	5	4	4	4	5	1
Financeiro e Outros	5	5	5	5	6	2	0	1	3	2	4	2
Materiais Básicos	15	19	17	22	21	21	9	13	14	14	16	9
Petróleo, Gás e Biocomb.	3	4	5	5	7	3	1	2	3	2	4	1
Tecnologia da Informação	0	0	0	0	1	0	0	0	0	1	1	0
Telecomunicações	14	12	14	8	8	4	12	14	12	9	8	1
Utilidade Pública	14	16	16	12	16	8	12	12	12	11	12	21
<b>Total de Empresas</b>	<b>68</b>	<b>74</b>	<b>74</b>	<b>73</b>	<b>84</b>	<b>51</b>	<b>46</b>	<b>54</b>	<b>54</b>	<b>53</b>	<b>65</b>	<b>44</b>

Fonte: Elaborada pelos autores.

Em termos gerais, observa-se que os setores que agrupam maior número de empresas no Brasil, conforme o critério adotado, são os de Materiais Básicos, Telecomunicações e Energia Elétrica. As tabelas 3 e 4 foram elaboradas apenas para apresentar a disposição das empresas entre *clusters* e setores econômicos, respectivamente, no decorrer do período analisado. Para uma análise mais detalhada sobre as semelhanças e divergências entre os dois tipos de classificação, cruzou-se os dados, buscando-se identificar a relação existente entre setores e *clusters*. Através deste cruzamento de dados, pôde-se observar se empresas inseridas em um determinado setor econômico necessariamente fizeram parte de um mesmo *cluster*. Ou seja, avaliou-se a convergência da classificação por setores econômicos com a classificação baseada na definição de empresas comparáveis proposta por Damodaran (2006).

Com os resultados obtidos, observou-se que existem divergências em utilizar-se a classificação das empresas por setores econômicos e a classificação por similaridade de fundamentos. Notou-se que, muitas vezes, empresas de um mesmo setor, agruparam-se em mais de um *cluster* e determinados *clusters* obtiveram empresas de diversos setores. Tal fato

leva crer que, no Brasil, existe a possibilidade de empresas de um mesmo setor econômico não necessariamente possuírem características de risco, geração de fluxo de caixa e taxas de crescimento semelhantes entre si. É interessante observar que a classificação por semelhança de fundamentos serviu para identificar subgrupos de empresas comparáveis dentro de um mesmo setor. Para ilustrar esta situação, utilizou-se como exemplo o ano de 2003 da amostra 1, conforme tabela abaixo:

TABELA 4: Distribuição de empresas por *clusters* e por setor econômico para o ano de 2003 da amostra final 1.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	TOTAL
<b>Bens Industriais</b>			6		1		<b>7</b>
<b>Construção e Transporte</b>				1			<b>1</b>
<b>Consumo Cíclico</b>		1	3				<b>4</b>
<b>Consumo Não Cíclico</b>			4		1		<b>5</b>
<b>Financeiro e Outros</b>			4	1			<b>5</b>
<b>Materiais Básicos</b>		3	11		1		<b>15</b>
<b>Petróleo, Gás e Biocomb.</b>			3				<b>3</b>
<b>Telecomunicações</b>	3	4	4		2	1	<b>14</b>
<b>Utilidade Pública</b>		4	6	2	2		<b>14</b>
<b>TOTAL</b>	<b>3</b>	<b>12</b>	<b>41</b>	<b>4</b>	<b>7</b>	<b>1</b>	<b>68</b>

Fonte: Elaborada pelos autores.

A divergência na classificação por setores econômicos e por similaridade de fundamentos pode ser observada claramente pelos setores de Telecomunicações e Utilidade Pública. As empresas do primeiro setor em questão se distribuíram entre cinco dos seis *clusters* gerados pela análise, enquanto aquelas do segundo se distribuem entre quatro dos seis *clusters*. Tal fato sugere a existência de subgrupos de empresas comparáveis dentro de um mesmo setor econômico. Nestes casos, a avaliação relativa poderia não gerar resultados adequados quando realizada utilizando o critério de classificação por setores econômicos, tendo em vista a necessidade de similaridade de fundamentos das empresas a serem comparadas. Esta ampla distribuição de empresas de um mesmo setor em vários *clusters* distintos também foi observada nos anos 2004, 2005 e 2008 da amostra final 1 e nos anos 2004 e 2008 da amostra final 2.

Por outro lado, é interessante observar que alguns setores possuíram quase todas as empresas agrupadas em um único *cluster*, indicando convergência do critério de classificação por setores econômicos com o critério de similaridade de fundamentos. Exemplos deste caso são os setores de Bens Industriais, Consumo Cíclico, Consumo Não Cíclico e Financeiro e Outros, para os dados da tabela 4. Nestes setores, as empresas podem ser consideradas mais similares entre si quanto aos fundamentos considerados (risco, geração de fluxo de caixa e taxa de crescimento nos lucros). Ainda assim, a análise de *clusters* foi fundamental para identificar empresas não comparáveis dentro de um mesmo setor. Explicitando, todos estes setores citados acima apresentaram apenas uma empresa em um *cluster* distinto das demais, indicando que a mesma não poderia ser utilizada para a realização de uma avaliação relativa do setor. Dessa forma, caso o analista opte por realizar uma avaliação relativa por setores econômicos, a utilização de técnicas estatísticas de agrupamento é viável para complementar

sua análise, auxiliando-o na identificação de empresas muito distintas das demais em um mesmo setor, que poderiam distorcer a análise. Observou-se a situação descrita acima nos anos 2004 e 2007 para a amostra final 1 e nos anos 2003, 2005, 2006 e 2007 para a amostra final 2.

Prosseguindo a análise, conseguiu-se também identificar que empresas de diferentes setores muitas vezes não são tão distintas em relação aos seus fundamentos. Ou seja, muitas vezes, diversos setores são agrupados no mesmo *cluster*. Adotando o ano de 2005 da amostra final 1 como exemplo, a tabela 5 evidencia esta situação.

TABELA 5: Distribuição de empresas por *clusters* e por setor econômico para o ano de 2005 da amostra final 1.

	Clusters										
	1	2	3	4	5	6	7	8	9	10	11
Bens Industriais			3	3		1					
Construção e Transporte							1				
Consumo Cíclico			3	3							
Consumo Não Cíclico	1		1		1						
Financeiro e Outros			1	2	2						
Materiais Básicos	4		5	5	1	1				1	
Petróleo, Gás e Biocomb.			3	1	1						
Telecomunicações	1	1	1	4	1	3		1	1		1
Utilidade Pública	2				6	6		1	1		
<b>Total</b>	<b>8</b>	<b>1</b>	<b>17</b>	<b>18</b>	<b>12</b>	<b>11</b>	<b>1</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>

Fonte: Elaborada pelos autores.

A existência de *clusters* formados por empresas de diversos setores pode ser visualizada através dos *clusters* 3, 4 e 5 que apresentam, respectivamente, sete, seis e seis setores distintos em sua composição. Além deles, os *clusters* 1 e 6, ainda que em menor escala, também são bons exemplos ao apresentar quatro setores distintos em suas composições. Estes cinco *clusters* abordados acima englobam cerca de 90% das empresas analisadas no ano de 2005, sendo representativos então para se afirmar que no ano de 2005, de maneira geral, a análise de *clusters* conseguiu identificar grupos de empresas comparáveis que não condizem com a lógica de agrupamento por setores econômicos.

Resultados similares foram obtidos para os anos de 2003, 2004 e 2006 na amostra final 1. Já na amostra final 2, que conta com o filtro de presença em bolsa em ao menos 50% dos pregões, os resultados foram peculiares. Os anos 2003, 2005, 2006 e 2007 apresentaram mais de 80% das empresas em um único *cluster*. A tabela 6 ilustra esta situação para o ano de 2007. Os valores absolutos para todos os anos podem ser visualizados na tabela 2, apresentada anteriormente.

TABELA 6: Distribuição de empresas por *clusters* e por setor econômico para o ano de 2007 da amostra final 2.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Bens Industriais	6			
Construção e Transporte	3			
Consumo Cíclico	6			
Consumo Não Cíclico	5			
Financeiro e Outros	4			
Materiais Básicos	16			
Petróleo, Gás e Biocomb.	4			
Tecnologia da Informação	1			
Telecomunicações	5		2	1
Utilidade Pública	11	1		
<b>Total geral</b>	<b>61</b>	<b>1</b>	<b>2</b>	<b>1</b>

Fonte: Elaborada pelos autores.

Com base nestes resultados, uma discussão sobre as características das duas amostras utilizadas pode ser formulada.

Primeiramente, vale ressaltar que, mesmo excluindo o período de 1995 a 2002 da análise, ambas as amostras finais apresentaram um reduzido número de observações, influenciando de maneira direta a análise. Este fato se deve ao ainda emergente mercado de capitais brasileiro que, além de apresentar um número relativamente pequeno de empresas, possui carência na disponibilidade de dados dessas companhias. Mais especificamente, na amostra final 1, ainda que utilizando um filtro mais brando de participação nos pregões (no mínimo um pregão ao ano), cerca de 73% das empresas da população ficaram de fora da amostra, o que reitera a idéia de um mercado de capitais relativamente pequeno.

Além disso, os fundamentos obtidos das empresas da amostra final 1 para se medir a similaridade entre elas podem ser distorcidos. O cálculo da medida de risco utilizada (beta), por exemplo, pode estar comprometido, tendo em vista a existência de poucas observações sobre os preços das ações das empresas. Contudo, ao tentar sanar esta problemática inserindo o filtro de 50% de presença em bolsa, ampliou-se o problema que já era visível na amostra final 1: um reduzido número de empresas para a análise. Enfim, a intensificação deste problema na amostra final 2 pode ter sido a causa do agrupamento de grande parcela das empresas em um único *cluster*, conforme abordado anteriormente. Uma possível justificativa para isto é que, ao acrescentar o filtro de 50% de presença em bolsa, eliminaram-se as empresas menos líquidas, levando-se a crer que as empresas mais líquidas do mercado brasileiro possuem fundamentos risco, geração de fluxo de caixa e taxa de crescimento muito similares, independente do setor econômico em que se encontram.

## 5. CONCLUSÃO

A primeira conclusão obtida com o presente estudo se refere ao fato de que os resultados adquiridos pelo agrupamento por setores econômicos divergiram daqueles obtidos pela metodologia estatística de análise de *cluster*. Tendo em vista tal fato, há evidências de que os setores econômicos do mercado acionário brasileiro não representam grupos de empresas com características similares de risco, crescimento e fluxo de caixa; ou seja, grupos de empresas comparáveis.

Conclui-se também que, em algumas situações, a análise *cluster* pode ser utilizada para se identificar empresas destoantes dentro de um determinado setor econômico. Ou seja, ainda que a classificação por setores econômicos seja válida, a análise de *clusters* pode auxiliar na identificação de empresas não comparáveis que devem ser retiradas do processo de avaliação relativa por setor.

Além disso, quando se considera apenas as empresas mais líquidas do mercado acionário brasileiro, representadas pela amostra final 2, nota-se que as mesmas não divergem, em geral, nas características de risco, geração de fluxo de caixa e taxa de crescimento. Tal fato permite que a avaliação relativa seja realizada desconsiderando-se a segmentação por setor econômico.

Por fim, pode-se concluir que o estudo possui como limitação o reduzido número de empresas nas amostras formadas, ficando então, como sugestão para estudos futuros, a ampliação desta amostra para países da América Latina, por exemplo, com a finalidade de se encontrar grupos mais bem definidos de empresas comparáveis. Assim, é possível que os testes pseudo-F e pseudo-T<sup>2</sup> convirjam em seus resultados, gerando maior robustez aos agrupamentos obtidos.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

ANDERBERG, M. R. **Cluster Analysis for Applications**. Academic Press, Inc., New York, NY, 1973.

BM&F BOVESPA. Empresas Listadas. São Paulo, 2010. Disponível em: [www.bovespa.com.br](http://www.bovespa.com.br). Acesso em: 13 abr. 2010.

BHORAJ, S., LEE, C.M. Who is my peer? A valuation-based approach to the selection of comparable firms. **Journal of Accounting Research**, vol. 40, n.2, 2002.

BHOJRAJ, S. LEE, C. M. OLER, D. K. What's my line? A comparison of industry classification schemes for capital market research. **Journal of Accounting Research**, vol. 41, n. 5, 2003.

DAMODARAN, A. **Valuation Approaches and Metrics: A Survey of the Theory and Evidence**. Working Paper, Stern School of Business, New York University, 2006.

\_\_\_\_\_. **Avaliação de Empresas**. 2ªed., São Paulo, Pearson Prentice Hall, 2007.

HAN, J. KAMBER, M. **Data Mining: concepts and techniques**. Morgan Kaufmann, 2001.

JOHNSON, R. A. WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 5ª ed., New Jersey: Prentice Hall, 2001.

LIU, J., NISSIM, D., THOMAS, J. Equity Valuation using Multiples. **Journal of Accounting Research**, vol.40, n.1, 2002.

KIM, M. RITTER, J. Valuing IPOs. **Journal of Financial Economics**, vol. 53, n. 2, 1999.

MARTELANC, R.; PASIN, R.; CAVALCANTE, F. **Avaliação de empresas: Um guia para fusões & aquisições e gestão de valor**. São Paulo: Financial Times/Prentice Hall, 2004.

MILLIGAN, W. COOPER, M.C. An examination of producers for determining the number of clusters in a data set. **Psychometrika**, 50, 2, p. 159-179, 1985.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2007.

PASIN, R. M. **Avaliação Relativa de Empresas por Meio da Regressão de Direcionadores de Valor**. 2004. Dissertação (Mestrado em Administração). Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo. São Paulo.

TIMM, N. H. **Applied Multivariate Analysis**. Pittsburgh: Springer, 2002.