

### 3

## **Critérios de Formação de Carteiras de Ativos Através de *Hierarchical Clusters***

### 3.1

#### **Introdução**

Este capítulo tem por objetivo principal a introdução de uma ferramenta muito conhecida em trabalhos de estudos organizacionais, e ainda não utilizada de maneira efetiva na área de finanças, conhecida como análise de *cluster*. É feita aqui uma introdução deste instrumento, aplicando os dados de retornos de ações da Bolsa de Valores de São Paulo (Bovespa) no Modelo de Multifatores de Fama e French (1996), que parte do princípio de que o prêmio pelo risco do mercado é uma variável insuficiente para explicar o comportamento dos ativos ao longo do tempo.

A principal motivação deste capítulo foi a insatisfação com a maneira tradicional pelas quais os acadêmicos e os profissionais da área de finanças organizam as carteiras de investimentos para testar os modelos de risco. Normalmente são divididas através de medidas de posição, que podem ordenar adequadamente de acordo com determinadas características, mas que também podem incorrer em alguns erros, como, por exemplo, o de termos distribuições bem assimétricas, com muitas observações na fronteira entre os quantis, distorcendo o resultado nas regressões dos modelos que posteriormente utilizam estas carteiras.

Aqui foi feita uma comparação entre a forma comumente utilizada para formação de carteiras em trabalhos acadêmicos<sup>3</sup> e um instrumento da estatística multivariada e de mineração de dados<sup>4</sup>, conhecida como análise de *clusters*. Esta ferramenta tem por objetivo a formação de grupos com características comuns,

---

<sup>3</sup> Ver Grinblatt e Moskowitz (2002) e Fama e French (1996).

<sup>4</sup> Ver Han e Kamber (2001).

sendo muito utilizada em trabalhos de análise organizacional de característica quantitativa e de descoberta de conhecimento em grandes bases de dados.

O pressuposto aqui assumido é o de que modelos baseados em grupos mais homogêneos tendem a apresentar melhor adequação do que os apenas separados através de medidas de posição, isto é, que ordenam as observações através de seus valores. Deseja-se, portanto, verificar se alguns instrumentos da estatística multivariada e mineração de dados podem contribuir de maneira efetiva no conhecimento intrínseco dos modelos financeiros e na melhoria de sua performance.

Este capítulo está dividido em quatro partes principais, além da introdução. Primeiro, é apresentada a justificativa teórica para esta proposição. Em seguida, a metodologia utilizada é descrita, bem como o Modelo de Fama e French (1996), que servirá de parâmetro para a análise comparativa. Sua escolha se deve ao fato de que é um modelo muito conhecido, e com características de utilização que servem ao caso brasileiro, como visto por Lucena e Figueiredo (2004). Na terceira parte, são discutidos os resultados encontrados nas regressões isoladas e nas regressões em painel do Modelo de Fama e French (1996) para as duas metodologias distintas. Por fim, na quarta parte, as conclusões são apresentadas.

## 3.2

### Justificativa Teórica

O objetivo deste capítulo é discutir o melhor caminho para a formação de carteiras e para a formulação de modelos na área financeira, particularmente os de formação de ativos de capital baseados em pressupostos de anomalias de mercado, como o de multifatores de Fama e French (1996), que será utilizado aqui para teste, além de propor uma metodologia pouco utilizada em trabalhos na área financeira, apesar de ser bastante popular na área de organizações. A comparação empírica entre as metodologias também é de suma importância para o desenvolvimento do restante da tese.

É comum a utilização da separação por quantis<sup>5</sup>, a partir de determinadas variáveis, como o tamanho da empresa, ou ainda outro indicador. Eles separam

---

<sup>5</sup> Ver Grinblatt e Moskowitz (2002) e Fama e French (1996).

justamente os extremos, e pode-se formar as carteiras a partir deles, porém devem ser levadas em consideração as características que fazem com que eles estejam localizados em determinado quantil. Lembrando que este apenas separa de acordo com sua posição ordinal, não levando em consideração a distância proporcional que o separa de seu vizinho mais próximo.

A idéia aqui é introduzir alguns conceitos de estatística multivariada, que podem ser muito úteis em trabalhos na área de finanças e que ainda são pouco utilizados. Partindo do princípio de que os quantis separam apenas de maneira ordinal, e que não leva em consideração o valor da variável em si, seria melhor encontrar uma maneira de dividir os grupos através de suas reais características. O que se quer dizer aqui é que, apesar de a ordenação ser um critério válido, ela não necessariamente agrupa na maneira adequada.

No modelo de multifatores de Fama e French (1996), que será utilizado aqui, as carteiras são formadas a partir de duas variáveis contábeis, que são o tamanho da empresa e o seu *book-to-market*. No trabalho original, os autores compõem 25 carteiras a partir do cruzamento dos quintis das duas variáveis. Porém, o que deve ser observado é que não existe certa linearidade nestas informações, isto é, o que leva um ativo a compor uma carteira pode ser uma diferença muito pequena, que não necessariamente justifique sua indicação para aquele grupo. Em outras palavras, se os grupos estiverem sendo separados por tamanho, duas empresas praticamente com o mesmo valor podem estar em grupos separados, quando possuem a mesma característica que se pretende analisar. A solução para isto seria agrupá-los de maneira adequada, e a proposição mais correta seria a separação dos grupos através de uma técnica conhecida como análise de *clusters*.

Desta forma, é esperado encontrar grupos mais homogêneos, evitando assim com que empresas que possivelmente estivessem na fronteira que diferencia os quantis sejam separadas apenas por estarem classificadas por ordenação.

### 3.3

#### Metodologia

Pretende-se aqui realizar uma comparação entre a formação de carteiras a através da divisão por quantis e a análise de *cluster*. O primeiro agrupa os ativos

através de ordenamento, e o segundo de maneira teoricamente mais homogênea, levando em consideração suas reais características. O objetivo aqui é testar se a análise de *clusters* pode apresentar melhores resultados nas regressões posteriores que utilizam carteiras de ativos, como é o caso do Modelo de Multifatores de Fama e French (1996).

A idéia principal do capítulo é comparar a formação de carteiras através dos quantis e da análise de *clusters*, utilizando o Modelo Multifatores de Fama e French (1996). Para isto, tal como no trabalho original destes autores, aqui é utilizada a base de dados mais ampla possível.

Para a organização dos dados, foi utilizada a planilha eletrônica do *Microsoft Excel* 2003, na sua versão 2003; para a análise de *clusters* o *SPSS 12.0*; e para as regressões do Modelo de Fama e French (1996), o pacote estatístico *eviews*, na versão 3.1.

### 3.3.1

#### Dados

Os dados são referentes a preços mensais de fechamento dos ativos na Bovespa, de 213 ações, além do *book-to-market* médio e do tamanho médio de cada uma delas, entre 1994 e 2004. Estes dados foram coletados a partir do banco de dados da Economática<sup>6</sup>. As 213 ações foram mais bem detalhadas no primeiro capítulo. Deve-se salientar, porém, que 2 delas não possuíam dados suficientes de tamanho<sup>7</sup> e *book-to-market*, e foram retiradas da amostra. Os dados de preços são referentes ao período de julho de 1994 a agosto de 2004, o *book-to-market* e o valor da empresa tem base anual. Os dados foram também deflacionados com base em agosto de 2004 pelo IPCA/IBGE acumulado, com o objetivo de evitar distorções no resultado final. Foi utilizado aqui o *book-to-market* médio e o tamanho médio, visto que nem todas as informações estavam disponíveis para todos os anos. Estas 213 ações foram selecionadas levando-se em consideração pelo menos uma negociação nos últimos 15 dias do mês<sup>8</sup>. Desta forma, evita-se

---

<sup>6</sup> [www.economatica.com.br](http://www.economatica.com.br), disponível no IAG/PUC-Rio.

<sup>7</sup> Tamanho aqui é visto como o valor de mercado da empresa, fornecido pela Economática.

<sup>8</sup> Foi estabelecida esta flexibilidade para que pudesse ao menos ter uma negociação no mês, com um período espaçado de tempo, evitando que houvesse apenas uma negociação no último dia de um mês, e outra no dia seguinte.

que títulos sem negociação por vários meses estivessem compondo a amostra desta tese. Posteriormente, 8 destas 213 ações foram retiradas da amostra por apresentarem *book-to-market* negativo, o que poderia distorcer o resultado final. Nossa amostra final é então composta por 205 ações. Os dividendos foram ajustados pela própria base de dados.

Para o cálculo do retorno mensal das ações, foi utilizada a seguinte fórmula:

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

onde  $P_t$  é o preço da ação no período  $t$  e  $P_{t-1}$  é o preço da ação no mês anterior, lembrando que o preço de fechamento da ação foi o último preço disponível nos últimos 15 dias do mês, como descrito anteriormente.

### 3.3.2

#### Formação de Carteiras através da Análise de *Clusters*

Para a formação das carteiras por quintis, com base nas 205 ações selecionadas, as ações foram ordenadas pelo seu tamanho e pelo seu *book-to-market*, e em seguida determinado seu posicionamento. Desta forma, pode-se formar 25 carteiras, se forem consideradas conjuntamente as duas variáveis. É importante destacar que não necessariamente as carteiras possuem o mesmo número de ativos, pois não estão distribuídas de maneira homogênea entre as duas variáveis.

Para a formação de *clusters*, é necessário um instrumento bem mais refinado e trabalhoso. Foi realizado aqui um procedimento conhecido como *hierarchical clusters*, que determina os grupos através das distâncias entre os valores de cada observação. Desta maneira, estes são formados a partir de características mais homogêneas. O principal problema deste procedimento é a existência de *outliers*, que podem desconfigurar completamente os grupos. O ideal neste caso seria a retirada destes *outliers* e a sua inclusão posterior de acordo com suas características. Isso se faz necessário para que não existam *clusters* com apenas uma empresa e *clusters* muito grandes. Da mesma maneira, pode-se ter carteiras com diferentes números de ativos, ou ainda ter carteiras que não possuam

ativos, pois o cruzamento dos *clusters* referentes a tamanho e *book-to-market* não obedece a uma ordem simétrica.

A análise de *clusters* de maneira hierárquica, como visto, tem o objetivo de formar grupos mais homogêneos, e estes são formados através da combinação individual dos ativos, um a um, em diferentes *clusters*, até a formação do número proposto de grupos, que neste caso específico foram cinco. Este número é justamente escolhido para comparação com o trabalho original de Fama e French (1996), que também forma 5 grupos, mas através de quintis. O processo hierárquico de formação de *clusters* é feito através de um dendograma, que, na verdade, é uma árvore que vai combinando os ativos através de suas diferenças. Este processo é facilitado pelo uso do *software SPSS*, pois, apesar de ser um procedimento matemático simples, é muito trabalhoso, porque a amostra é composta por 205 ativos. A formação destes grupos é feita, então, através de uma espécie de árvore de agrupamentos, determinada pela distância entre os pontos.

### 3.3.2.1

#### ***Hierarchichal Clusters***

O procedimento utilizado nesta tese é conhecido como análise hierárquica de clusters, sendo a técnica mais comumente usada. A maneira de organização dos dados se dá através de uma matriz de similaridades, com a detecção dos dados de maior similaridade, no nosso caso através de números mais próximos, tanto de valor como de *book-to-market*.

Parte-se de uma matriz inicial, que no nosso caso é representada pelas medidas de *book-to-market* médio e tamanho médio das empresas da amostra, conforme apresentado na seção de dados. Temos então uma matriz  $[n \times p]$ , sendo  $n$  o número de empresas, aqui tendo  $n = 205$ , e  $p = 1$ . É utilizada então a distância euclidiana, com o propósito de realizar os agrupamentos.

O número de *clusters* a serem utilizados não segue um padrão fixo, pois depende muito do objetivo do pesquisador. Não há um critério estatístico padrão para inferência, como testes ou coisas semelhantes. Um procedimento muito conhecido é a regra de parada, que examina alguma medida de similaridade ou distância entre agrupamentos em cada passo sucessivo, com a solução de *clusters* sendo definida quando a medidade de similaridade excede a um valor

especificado, isto é, quando a distância entre dois pontos ultrapassar um valor pré-definido pelo pesquisador. O procedimento utilizado nesta tese representa uma restrição prática, que é o de forçar um número pré-determinado de *clusters* como igual a 5 (cinco), mas que é bastante útil no processo de comparação que pretende-se realizar. A escolha de 5 (cinco) grupos se deu pelo fato de Fama e French (1996) também utilizarem os quintis para no seu trabalho original.

A medida de distância euclidiana, que é parte componente do *software* SPSS, utilizado neste capítulo, expressa o grau de similaridade que pretende-se obter. A distância  $D$  entre dois pontos, cuja localização é especificada através de coordenadas cartesianas é dada por:

$$D_{1,2} = \sqrt{(BM_1 - BM_2)^2 + (T_1 - T_2)^2}$$

sendo  $BM$  o book-to-market da empresa e  $T$  o tamanho. Nesta tese não serão utilizados *clusters* conjuntos, então a equação sofre uma transformação, sendo apenas a distância entre dois pontos na mesma reta:

$$DBM_{1,2} = BM_1 - BM_2$$

sendo  $DBM$  a distância entre duas empresas na variável *book-to-market*. O mesmo procedimento se aplica para a variável tamanho.

### 3.3.3

#### Modelo de Multifatores de Fama e French

As anomalias de mercado, que na verdade são as variáveis adicionais ao Modelo CAPM, foram popularmente disseminadas a partir do trabalho de DeBondt e Thaler (1985) e também de Fama e French (1996). Este último partiu do pressuposto de que o tamanho da empresa e o índice *book-to-market* também são fatores determinantes na taxa de retorno exigida das empresas, e devem ser analisados. Outras variáveis vêm sendo acrescentadas em outros trabalhos científicos, mas estes são os mais comuns, e servem aos propósitos principais desta tese, que é a comparação entre as diversas metodologias de formação de carteiras. A versão original do modelo de Fama e French (1996) é apresentada abaixo:

$$E(R_t) - R_f = \beta_1(E(R_M) - R_f) + \beta_2SMB_t + \beta_3HML_t$$

onde  $R_i$  é o retorno do ativo;  $R_f$ , o retorno do ativo livre de risco;  $R_M$ , o retorno de mercado;  $SMB$ , o retorno da carteira formada por ativos de empresas pequenas menos o retorno das grandes (*small minus big*);  $HML$ , o retorno da carteira de empresas com alto índice *book-to-market* menos o retorno das de baixo índice (*high minus low*). Em toda a tese, é utilizado o Ibovespa como ativo médio de mercado, e o retorno do CDI bancário como o retorno de ativo livre de risco.

Com este modelo, será feita uma regressão em painel com as possíveis 25 carteiras que serão formadas, que são dadas por  $n \times m$ , onde  $n$  e  $m$  são as quantidades de quintis ou *clusters* formados pelo tamanho e pelo *book-to-market*. Como visto anteriormente, pode-se ter menos do que 25 carteiras formadas, visto que não é obrigatória a existência de cruzamentos entre os grupos. A idéia aqui é testar se serão encontrados resultados mais satisfatórios em termos de explicabilidade do modelo com a formação de carteiras através de *clusters* do que com os quantis, pelo fato de ter uma maior homogeneidade nos mesmos.

### 3.3.4

#### Regressão de Dados em Painel

Em muitos casos, há dois tipos de informação disponível para modelagem através de regressão, as informações *cross-section* e as de ordem cronológica. Neste trabalho, optou-se por uma técnica conhecida como regressão de dados em painel, que pode ser de duas formas: de efeitos aleatórios ou de efeitos fixos. Neste caso trata-se da primeira delas, já que não são considerados efeitos fixos no modelo. Como a idéia é comparar as duas metodologias para formação de carteiras, e as regressões individuais por carteiras pouco revelam, a opção pela regressão em painel se mostra a mais adequada.

A regressão em painel pode apresentar alguns problemas, principalmente quando são estimadas por mínimos quadrados ordinários (MQO), porém os erros, em sua maioria, são nos erros padrões. Segundo Johnston e DiNardo (2001), o método MQO produz estimativas consistentes de  $\beta$ , apenas os erros-padrão pouco dizem.

A regressão de Fama e French (1996) em painel se daria então da seguinte forma.

$$(R_i - R_f)_t = \beta_1(R_m - R_f)_t + \beta_2(SMB)_t + \beta_3(HML)_t$$

Como pode ser visto na equação acima, os ativos são os dados em painel, isto é, retornos de diferentes carteiras ao longo do tempo. Neste caso, os retornos de mercado, *SMB* e *HML* variam apenas ao longo do tempo, já que são variáveis de mercado e independem dos ativos para seu valor. Considera-se aqui  $t \leq 25$ , que seria o número máximo de carteiras. O *software eviews* possui um comando para cálculos desta natureza.

### 3.4

#### Resultados Encontrados

Primeiramente foram divididos os ativos por quintis, nas duas variáveis, que são tamanho e *book-to-market*. Desta forma, foram encontradas 25 carteiras, formadas quantitativamente de maneira heterogênea, como visto na tabela abaixo, no número de ativos de cada uma.

Quando analisada isoladamente cada carteira e sua regressão, pode-se verificar que os prêmios pelo risco medido pelo Modelo CAPM e seus parâmetros são significativos em todos os casos, o que de certa forma já era esperado. Nas outras variáveis, em alguns casos os parâmetros não se mostraram significativos, o que não invalida o modelo, visto que se pretende realizar uma regressão em painel para encontrar uma fórmula que represente o conjunto das carteiras, a qual, apresentado a seguir, mostrou-se significativa.

Inicialmente serão rodadas as regressões isoladamente por carteiras, e posteriormente, a regressão em painel, com todas as informações.

Tabela 4 – Dados da Regressão Individual de Cada Carteira formada por Quintis

<i>Book-to-market</i>										
	$\beta_1$					<i>p-value</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	Alto
<b>Pequena</b>	1,15	0,76	0,77	0,70	0,87	0	0,01	0	0	0
<b>2</b>	1,02	0,76	0,61	0,73	0,71	0	0	0	0	0
<b>3</b>	0,86	0,82	0,72	0,90	0,91	0	0	0	0	0
<b>4</b>	0,80	0,86	0,80	0,91	0,70	0	0	0	0	0
<b>Grande</b>	0,75	0,85	0,89	0,84	1,08	0	0	0	0	0
<i>Book-to-market</i>										
	$\beta_2$					<i>p-value</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	Alto
<b>Pequena</b>	0,85	2,06	1,05	0,79	1,08	0	0	0	0	0
<b>2</b>	1,39	0,56	0,56	0,69	0,19	0	0	0	0	0,17
<b>3</b>	0,67	0,29	0,49	0,47	0,03	0	0	0	0	0,91
<b>4</b>	0,61	0,49	0,35	0,40	-0,08	0	0	0,02	0,02	0,74
<b>Grande</b>	0,19	-0,05	-0,13	0,03	-0,25	0,12	0,53	0,31	0,81	0,13
<i>Book-to-market</i>										
	$\beta_3$					<i>p-value</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	Alto
<b>Pequena</b>	-0,77	-1,23	-0,42	-0,21	0,21	0	0,01	0,01	0,07	0,03
<b>2</b>	-1,09	-0,15	-0,16	-0,12	0,39	0	0,26	0,18	0,41	0
<b>3</b>	-0,60	-0,09	0,01	0,22	1,26	0	0,48	0,90	0,10	0
<b>4</b>	-0,58	-0,20	-0,13	-0,23	0,38	0	0,18	0,33	0,13	0,07
<b>Grande</b>	-0,47	-0,04	0,12	-0,03	0,40	0	0,63	0,33	0,78	0,01
<i>Book-to-market</i>										
	$R^2$					<i>Número de ativos</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	Alto
<b>Pequena</b>	0,45	0,21	0,29	0,38	0,63	3	3	6	10	19
<b>2</b>	0,19	0,37	0,29	0,32	0,48	3	10	9	12	7
<b>3</b>	0,45	0,46	0,41	0,48	0,38	5	11	13	4	8
<b>4</b>	0,43	0,41	0,45	0,43	0,25	15	6	9	6	5
<b>Grande</b>	0,58	0,77	0,64	0,56	0,64	15	11	4	9	2

Como dito, a maioria dos parâmetros se mostrou significativo, conforme testes de primeira ordem na tabela acima (testes *t*), e este resultado se confirmou na regressão em painel, também utilizando o *software eviews*. A equação de regressão em painel, então, foi:

$$R_i - R_f = 0,830(E(R_M) - R_f) + 0,480SMB - 0,122HML \quad R^2=0,31$$

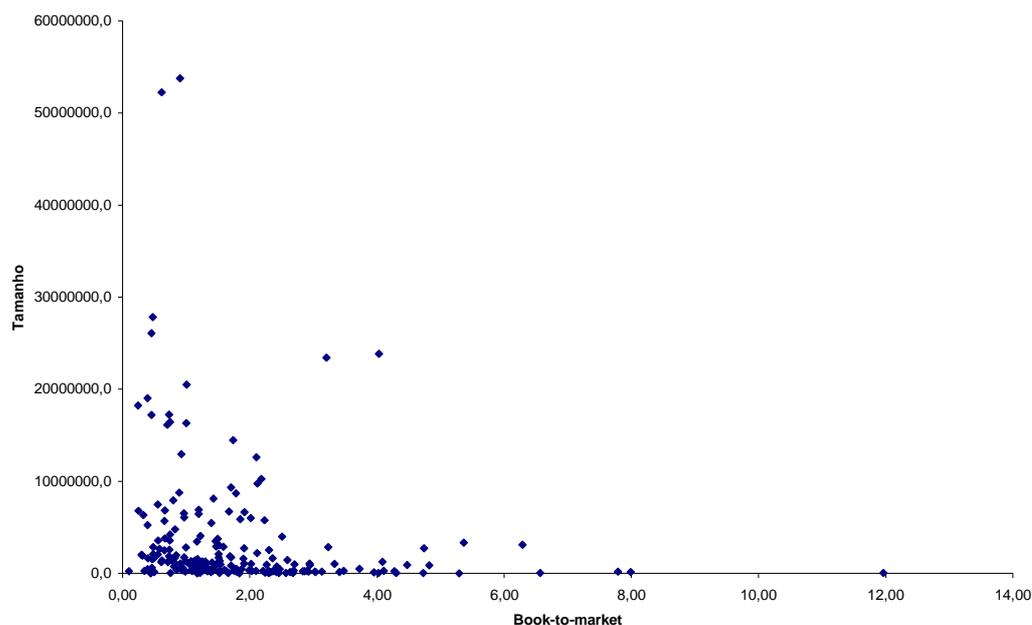
(37,16)
(12,49)
(-3,54)

Nos resultados da equação acima, os valores entre parênteses abaixo dos parâmetros são o *t*'s encontrados para os testes de primeira ordem. Pode-se afirmar que todos são significantes, e que as duas anomalias consideradas pelos autores, tamanho e *book-to-market*, devem ser levadas em consideração para o

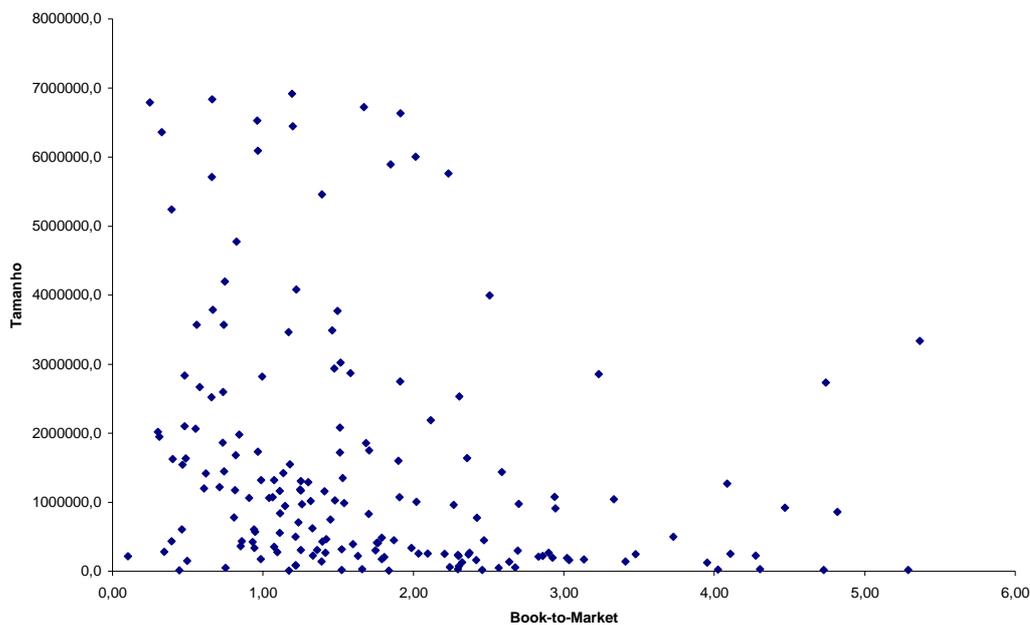
mercado de capitais brasileiro. O coeficiente de determinação, importante para a comparação com a formação através da análise de *clusters*, se mostrou razoável, visto que se trabalha aqui com séries financeiras, no valor de 0,31. O teste *F* de significância conjunta dos parâmetros apresentou o *p-value* de 0,000.

Quando as carteiras são formadas através de análise de *clusters*, espera-se encontrar grupos mais homogêneos em relação às duas anomalias aqui consideradas, que são o tamanho e o *book-to-market*. O principal problema diz respeito aos *outliers*, que devem ser tratados caso a caso, para que não produzam distorções no resultado final. O procedimento mais correto seria o de retirá-los e posteriormente recolocá-los em um grupo similar no qual a observação possa pertencer. No gráfico abaixo, percebe-se que não há condições de determinar os *clusters* com alguns resultados muito dispersos.

**Figura 1 – Distribuição dos Dados das 205 Empresas Estudadas (*book-to-market* e tamanho) com *Outliers***



**Figura 2 – Distribuição dos Dados das 205 Empresas Estudadas (*book-to-market* e tamanho) sem *Outliers***



Quando os *outliers* são retirados, pode-se perceber que alguns grupos com características comuns se formam. Vale ressaltar que este gráfico está refletindo a análise conjunta que será apresentada mais abaixo, na regressão da última equação. Se fossem formados grupos (*clusters*) a partir das informações que formaram a Figura 1, haveria um *cluster* com muitas empresas, na verdade quase a totalidade, e cada *outlier* representaria um *clusters* específico. Na Figura 1, pode-se verificar que uma empresa tem um *book-to-market* bem superior às demais (próximo de 12). Esta empresa é um exemplo típico de um *outlier* que poderia ser agrupada em um grupo chamado “empresas com *book-to-market* alto”.

Os resultados individuais pouco mostram em análise comparativa, mas pode ser verificado que a distribuição de ativos nas carteiras se mostrou muito mais assimétrica do que na separação por quintis.

**Tabela 5 – Dados da Regressão Individual de cada carteira formada por Análise de Clusters**

Tamanho	<i>Book-to-market</i>									
	$\beta_1$					<i>p-value</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	alto
Pequena	0,83	1,09	0,89	0,43	0,80	0	0	0	0	0
2	0,65	0,81	0,93	0,87	0,90	0	0	0	0	0
3	0,83	0,68	0,87	0,92	Não	0	0	0	0	Não
4	1,03	Não	0,89	Não	Não	0	Não	0	Não	Não
Grande	0,84	1,04	0,74	Não	1,03	0	0	0	Não	0
<i>Book-to-market</i>										
	$\beta_2$					<i>p-value</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	Alto
	Pequena	0,61	0,65	0,66	-0,07	0,40	0	0,03	0	0,81
2	0,80	0,95	1,11	1,19	0,75	0	0	0	0	0
3	0,31	0,53	0,54	0,27	Não	0,04	0,04	0	0,33	Não
4	0,35	Não	0,31	Não	Não	0,01	Não	0,01	Não	Não
Grande	0,03	-0,42	-0,06	Não	-0,34	0,85	0,02	0,52	Não	0,03
<i>Book-to-market</i>										
	$\beta_3$					<i>p-value</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	Alto
	Pequena	-0,06	0,18	-0,15	0,20	-0,36	0,29	0,23	0,02	0,13
2	-0,15	-0,03	-0,20	1,19	0,05	0,03	0,64	0	0	0,62
3	-0,07	-0,22	-0,16	0,35	Não	0,31	0,08	0,03	0,01	Não
4	-0,02	Não	-0,16	Não	Não	0,79	Não	0,01	Não	Não
Grande	0,26	0,27	-0,09	Não	0,25	0,26	0	0,07	Não	0
<i>Book-to-market</i>										
	$R^2$					<i>Número de Ativos</i>				
	Baixo	2	3	4	Alto	Baixo	2	3	4	Alto
	Pequena	0,59	0,27	0,59	0,17	0,23	19	5	32	1
2	0,38	0,57	0,65	0,67	0,36	35	13	22	6	6
3	0,48	0,16	0,52	0,29	Não	8	1	11	3	Não
4	0,63	Não	0,63	Não	Não	6	Não	9	Não	Não
Grande	0,54	0,65	0,67	Não	0,66	7	1	17	Não	1

A regressão em painel do Modelo de Fama e French (1996) e seus resultados são mostrados abaixo.

$$E(R_i) - R_f = 0,873(E(R_M) - R_f) + 0,448 SMB - 0,044 HML + \varepsilon \quad R^2=0,36$$

(34,81)                      (10,15)                      (2,02)

Como pode ser visto acima nos resultados da aplicação do Modelo de Fama e French (1996), quando rodado por carteiras formadas através de análise de *clusters*, os resultados são melhores, como observado no coeficiente de determinação, que neste caso foi de 0,36. O teste *F* de significância conjunta dos parâmetros também apresentou o *p-value* de 0,000, como no caso anterior. Além

disso, nos testes de primeira ordem, todos os parâmetros se mostraram significativos. O importante foi a verificação da melhoria dos resultados da regressão.

O resultado dos coeficientes mostra coerência em relação ao Modelo tradicional de ativos de capital (CAPM), com o resultado do parâmetro de prêmio pelo risco sendo não tão distante de 1(un). O resultado de  $\beta_1$  foi de 0,83 e 0,873, respectivamente. O resultado dos demais parâmetros também se mostrou convergente com o Modelo de Fama e French, sendo positiva a diferença entre os quintis da carteira de tamanho e negativo de *book-to-market*. Neste último caso, houve uma razoável diferença na carteira formada pela análise de *clusters*, indo de -0,122 para -0,044. É como se nesta nova metodologia este parâmetro perdesse força, mesmo mostrando significância estatística.

### 3.5

#### Conclusões

Os trabalhos na área de finanças, em particular os que derivam de modelagem de risco, como os de Sharpe (1964), Fama e French (1996), e Grinblatt e Moskowitz (2004), podem dizer muito à respeito das relações que o comportamento dos títulos disponíveis no mercado têm com as diversas variáveis, como o comportamento da média do mercado, ou mesmo com o que é chamado de anomalia, assunto muito discutido a partir de alguns trabalhos de Fama e French (1991, 1996).

A literatura relevante nesta área, em particular os trabalhos com testes empíricos, mostra que é comum a utilização de formação de carteiras de ativos para que os testes se realizem. E os *portfólios* são formados normalmente através de medidas de posição de ordenação, como os quintis, quartis ou percentis. O que pretendia-se aqui era apresentar uma forma alternativa de formação de carteiras, conhecida como análise de *clusters*, que agrupa as variáveis através de uma metodologia mais refinada do que simplesmente a ordenação.

O objetivo da introdução da análise de *clusters* em trabalhos de finanças pode ser de grande valia, visto que algumas observações podem estar na fronteira entre determinados grupos de posição, e isso pode gerar resultados não consistentes nas regressões posteriores.

Ao utilizar a análise de *clusters*, e aplicar à regressão de multifatores de Fama e French (1996), descobriu-se que, no caso brasileiro, podem ser encontrados resultados mais satisfatórios, que vão de encontro à hipótese inicial deste capítulo, que é a de que algumas técnicas de estatística multivariada, neste caso a análise de *cluster*, podem ser muito úteis a testes empíricos da teoria financeira.

Os resultados encontrados corroboram a premissa inicial, e sugere-se a sua utilização na formação de carteiras para a construção ou aplicação de modelos de risco, pois os grupos formados tendem a terem características mais homogêneas do que simplesmente a ordenação através de medidas de posição.

A principal limitação deste capítulo é a utilização de indicadores médios de tamanho e *book-to-market* em detrimento de valores anuais nos quais a carteira pudesse ser rebalanceada, pois os ativos poderiam migrar de uma carteira para outra, isto é, ela poderia aumentar de tamanho, passando para outro *cluster*. Essa opção se deu pelo fato de que grande parte das séries estaria incompleta, e, neste caso, muitas empresas teriam que ser retiradas da amostra. Isso comprometeria um dos principais objetivos deste capítulo, que eram selecionar uma amostra ampla para a tese e que o modelo a ser proposto posteriormente fosse formado a partir de uma boa representação do mercado brasileiro de ações. Neste aspecto, foi realmente uma opção de pesquisa, mas que de certa forma delimita as conclusões posteriores. Deve ser citada aqui uma restrição que foi outra decisão tomada durante a pesquisa: o fato de forçar o número de *clusters* em 5 (cinco), com o intuito de fazer a comparação com os quintis do trabalho original de Fama e French (1996). Apesar de não ser a técnica mais correta, é uma restrição que facilitaria a comparação de resultados, o que já pode ser considerada como válida. Outra limitação, comum a trabalhos acadêmicos no Brasil, é o início de informações captadas a partir de julho de 1994, excluindo-se o alto período inflacionário da década de 70, 80 e início dos anos 90. É também uma opção que limita a pesquisa, mas que oferece mais vantagens que desvantagens ao processo de análise de resultados.

Fica como sugestão de pesquisas posteriores a aplicação desta metodologia a diferentes modelos da área de finanças que utilizam a formação de carteiras, como trabalhos com o *Value at Risk* e outros modelos de estudos de anomalias, como as de calendário (efeito segunda-feira, efeito janeiro, etc.).