



Munich Personal RePEc Archive

## **Criteria of portfolio formation of stocks through hierarchical clusters**

Lucena, Pierre and Figueiredo, Antonio Carlos and  
Lachtermacher, Gerson

PUC-Rio, Universidade Federal de Pernambuco

20 February 2008

Online at <https://mpa.ub.uni-muenchen.de/38105/>  
MPRA Paper No. 38105, posted 15 Apr 2012 14:39 UTC



## CRITÉRIOS DE FORMAÇÃO DE CARTEIRAS DE ATIVOS POR MEIO DE *HIERARCHICAL CLUSTERS*

### CRITERIA OF PORTFOLIO FORMATION OF STOCKS THROUGH HIERARCHICAL CLUSTERS

#### **PIERRE LUCENA**

*Doutor em Administração/Finanças pelo Programa de Pós-Graduação em Administração de Empresas da Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ).  
Professor do Centro de Ciências Sociais Aplicadas da Universidade Federal de Pernambuco (UFPE).  
Rua Professor Moraes Rego, 1235, CDU – Recife – PE – Brasil – CEP 50670-901  
E-mail: pierrelucena@uol.com.br*

#### **ANTONIO CARLOS FIGUEIREDO PINTO**

*Doutor em Administração/Finanças pelo Programa de Pós-Graduação em Administração de Empresas da Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ).  
Professor do Instituto de Administração e Gerência da Pontifícia Universidade Católica do Rio de Janeiro.  
Rua Marquês de São Vicente, 225, Gávea – Rio de Janeiro – Brasil – CEP 22453-900  
E-mail: figueiredo@iag.puc-rio.br*

#### **GERSON LACHTERMACHER**

*Ph.D. em Management Sciences pelo Department of Management Sciences da Faculty of Engineering da University of Waterloo.  
Professor da Escola de Administração e Economia da Fundação Getúlio Vargas (FGV) e da Faculdade de Ciências Econômicas da Universidade do Estado do Rio de Janeiro (UERJ).  
Avenida São Francisco Xavier, 524, 8º andar, bloco B,  
Maracanã – Rio de Janeiro – RJ – Brasil – CEP 20550-013  
E-mail: rglachter@fgv.br*

## RESUMO

Este artigo tem como objetivo principal apresentar e testar uma ferramenta de estatística multivariada em modelos financeiros. Essa metodologia, conhecida como análise de *clusters*, separa as observações em grupos com suas determinadas características, em contraste com a metodologia tradicional, que é somente a ordem com os quantis. Foi aplicada essa ferramenta em 213 ações negociadas na Bolsa de São Paulo (Bovespa), separando os grupos por tamanho e *book-to-market*. Depois, as novas carteiras foram aplicadas no modelo de Fama e French (1996), comparando os resultados numa formação de carteira para quantil e análise de *cluster*. Foram encontrados melhores resultados na segunda metodologia. Os autores concluem que a análise de *cluster* pode ser mais adequada porque tende a formar grupos mais homogêneos, sendo sua aplicação útil para a formação de carteiras e para a teoria financeira.

## PALAVRAS-CHAVE

Quantis; Análise de *cluster*; *Data Mining*; Anomalias; Modelo de Fama e French.

## ABSTRACT

This paper has as main objective to present and to test a tool of multivariate statistics in financial models. This methodology, known as clusters analysis, separates the observations in groups through its determined characteristic, in contrast of the traditional methodology, which is only the order through quantiles. This tool was applied in 213 shares negotiated in the São Paulo Stock Exchange (Bovespa), separating to the groups for size and book-to-market. Later, the new portfolios were applied in the Fama and French Model (1996), comparing the results in a portfolio formation for quintiles and for cluster analysis. Better results were found in the second methodology. The authors conclude that the cluster analysis can be more adequate, because tends to form more homogeneous groups, being useful its application for portfolio formation, and for financial theory.

## KEYWORDS

Quantiles; Cluster analysis; Data Mining; Anomalies; Fama and French model.

## 1 INTRODUÇÃO

Este artigo 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 desse 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 trabalho foi a insatisfação com a maneira tradicional pela qual os acadêmicos e os profissionais da área de finanças organizam as carteiras de investimentos para testar os modelos de risco. Normalmente, elas são divididas por meio de medidas de posição que podem ordenar as carteiras adequadamente de acordo com determinadas características. No entanto, esse procedimento pode gerar alguns erros, sobretudo em relação às distribuições bem assimétricas, com muitas observações na fronteira entre os quantis, distorcendo o resultado nas regressões dos modelos que posteriormente utilizam essas carteiras.

Aqui foi feita uma comparação entre a forma comumente utilizada para formação de carteiras em trabalhos acadêmicos (GRINBLATT; MOSKOWITZ, 2004; FAMA; FRENCH, 1996) e um instrumento da estatística multivariada e de mineração de dados (HAN; KAMBER, 2001), conhecida como análise de *clusters*. Essa ferramenta tem por objetivo a formação de grupos com características comuns, 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 podem apresentar melhor adequação do que os apenas separados por meio de medidas de posição, isto é, que ordenam as observações por meio de seus valores. Deseja-se, portanto, verificar se alguns instrumentos da estatística multivariada e mineração de dados podem contribuir de maneira efetiva para o conhecimento intrínseco dos modelos financeiros e para a melhoria de sua *performance*.

O trabalho está dividido em quatro partes principais, além da introdução. Primeiro, é apresentada a justificativa teórica para essa 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, apresentam-se as conclusões.

## 2 REFERENCIAL TEÓRICO

O objetivo deste artigo 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 trabalho.

É comum a utilização da separação por quantis (GRINBLATT; MOSKOWITZ, 2004; FAMA; FRENCH, 1996) com base em determinadas variáveis, como o tamanho da empresa, ou outro indicador. Os quantis separam justamente os extremos, e podem-se formar as carteiras a partir deles. No entanto, devem ser consideradas as características que fazem 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 ideia 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 levam em consideração o valor da variável em si, seria melhor encontrar uma maneira de dividir os grupos por meio 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. Em alguns casos, a separação pode estar adequada e não apresentar mudança significativa, mas isso nem sempre ocorre.

No modelo de multifatores de Fama e French (1996), que será utilizado aqui, as carteiras são formadas a partir de duas variáveis contáveis: tamanho da

empresa e 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 nessas 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 isso seria agrupá-los de maneira adequada, e a proposição mais correta seria a separação dos grupos por meio de uma técnica conhecida como análise de *clusters*.

Dessa forma, espera-se encontrar grupos mais homogêneos, evitando assim que empresas que possivelmente estejam na fronteira que diferencia os quantis sejam separadas apenas por estarem classificadas por ordenação.

### 3 PROCEDIMENTO METODOLÓGICO

Pretende-se aqui realizar uma comparação entre a formação de carteiras por meio da divisão por quantis e a análise de *cluster*. O primeiro agrupa os ativos por meio de ordenamento, e o segundo faz isso de maneira teoricamente mais homogênea, levando em consideração suas reais características. O objetivo 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 ideia principal deste trabalho é comparar a formação de carteiras por meio dos quantis e da análise de *clusters*, utilizando o modelo multifatores de Fama e French (1996). Para isto, tal como no trabalho original desses autores, utiliza-se aqui a base de dados mais ampla possível.

Para a organização dos dados, utilizou-se a planilha eletrônica do Microsoft Excel, 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.

#### 3.1 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. Esses dados foram coletados a partir do banco de dados da Economatica. Os dados de preços são referentes ao período de julho de 1994 a agosto de 2004, e o *book-to-market* e o valor da empresa têm 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. Utilizaram-se aqui o *book-to-market* médio e o tamanho médio, visto que nem todas as informações estavam disponíveis para todos os anos. Essas 213 ações foram selecionadas levando-se em consideração pelo menos uma negociação nos últimos 15 dias do mês<sup>1</sup>. Dessa forma, evitou-se que títulos sem negociação por vários meses fizessem parte desta amostra. Posteriormente, 8 dessas 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.

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. Os dados de preços foram ajustados para dividendos e proventos pelo banco de dados da Economatica.

### 3.2 FORMAÇÃO DE CARTEIRAS POR MEIO DA ANÁLISE DE CLUSTERS

Para a formação das carteiras por quintis, com base nas 205 ações selecionadas, as ações foram ordenadas por seu tamanho e *book-to-market*, e, em seguida, determinou-se seu posicionamento. Dessa forma, podem-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 por meio das distâncias entre os valores de cada observação. Dessa maneira, estes são formados a partir de características mais homogêneas. O principal problema desse procedimento é a existência de *outliers*, que podem desconfigurar completamente os grupos. O ideal nesse caso seria a retirada desses *outliers* e a sua inclusão posterior de acordo com suas características. Isso se faz necessário para que não existam *clusters* com apenas uma empre-

<sup>1</sup> Foi estabelecida essa flexibilidade para que pudesse haver, ao menos, 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.

sa e *clusters* muito grandes. Da mesma maneira, podem-se obter carteiras com diferentes números de ativos ou ainda carteiras que não possuam ativos, pois o cruzamento dos *clusters* referentes a tamanho e *book-to-market* não obedece a uma ordem simétrica.

Os *clusters* são formados pela combinação individual dos ativos, um a um, em diferentes grupos, até a formação do número proposto, que neste caso específico foram cinco, escolhido para comparação com o trabalho original de Fama e French (1996), que também forma cinco grupos, mas por meio de quintis. O processo hierárquico de formação de *clusters* é feito por meio de um dendograma, que, na verdade, é uma árvore que vai combinando os ativos por meio de suas diferenças. Esse processo é facilitado pelo uso do *software* SPSS. Apesar de ser um procedimento matemático simples, é muito trabalhoso, porque a amostra é composta por 205 ativos. A formação desses grupos é feita por meio de uma espécie de árvore de agrupamentos, determinada pela distância entre os pontos.

### 3.2.1 Hierarchical clusters

O procedimento utilizado neste trabalho é conhecido como análise hierárquica de *clusters*, sendo a técnica mais comumente usada. A maneira de organização dos dados se dá por meio de uma matriz de similaridades, com a detecção dos dados de maior similaridade, no nosso caso por meio 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 medida de similaridade excede a um valor especificado, isto é, quando a distância entre dois pontos ultrapassar um valor predefinido pelo pesquisador.

O procedimento utilizado neste trabalho representa uma restrição prática, que é o de forçar um número predeterminado de *clusters* como igual a cinco, mas que é bastante útil no processo de comparação que se pretende realizar. A escolha de cinco grupos se deu pelo fato de Fama e French (1996) também uti-



lizarem os quintis no seu trabalho original. Outros critérios poderiam ser utilizados para a composição de grupos, mas este pareceu mais adequado para efeito comparativo com o trabalho original de Fama e French (1996).

A medida de distância euclidiana, que é parte componente do *software* SPSS, utilizado aqui, expressa o grau de similaridade que se pretende obter. A distância  $D$  entre dois pontos, cuja localização é especificada por meio 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. Como neste artigo não serão utilizados *clusters* conjuntos, 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 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 do presente trabalho, que é a comparação entre as diversas metodologias de formação de carteiras. A versão original do modelo de Fama e French (1996) é apresentada a seguir:

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

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*). Utilizam-se o Ibovespa como ativo médio de mercado e o retorno do CDI bancário como o retorno de ativo livre de risco.

Com esse 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 ideia aqui é testar se serão encontrados resultados mais satisfatórios em termos de explicabilidade do modelo com a formação de carteiras por meio de *clusters* do que com os quantis, pelo fato de haver uma maior homogeneidade nos *clusters*.

### 3.4 REGRESSÃO DE DADOS EM PAINEL

Em muitos casos, há dois tipos de informação disponível para modelagem por meio 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 ideia é comparar as duas metodologias para formação de carteiras, como as regressões individuais por carteiras pouco revelam, a opção pela regressão em painel se mostra a mais adequada.

As regressões em painel podem apresentar alguns problemas, principalmente quando são estimadas por mínimos quadrados ordinários (MQO), porém os erros, em sua maioria, são nos erros padrão. 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 nessa equação, os ativos são os dados em painel, isto é, retornos de diferentes carteiras ao longo do tempo. Nesse 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 dessa natureza.

## 4 RESULTADOS ENCONTRADOS

Primeiramente, foram divididos os ativos por quintis, nas duas variáveis, que são tamanho e *book-to-market*. Dessa forma, encontraram-se 25 carteiras,

formadas quantitativamente de maneira heterogênea, como visto na Tabela 1, 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, apresentada a seguir, mostrou-se significativa.

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

**TABELA 1**

**DADOS DA REGRESSÃO INDIVIDUAL DE  
 CADA CARTEIRA FORMADA POR QUINTIS**

		BOOK-TO-MARKET										
		$\beta_1$					<i>p-value</i>					
		BAIXO	2	3	4	ALTO	BAIXO	2	3	4	ALTO	
TAMANHO	PEQUENA	1,15	0,76	0,77	0,70	0,87	0	0,01	0	0	0	
	2	1,02	0,76	0,61	0,73	0,71	0	0	0	0	0	
	3	0,86	0,82	0,72	0,90	0,91	0	0	0	0	0	
	4	0,80	0,86	0,80	0,91	0,70	0	0	0	0	0	
	GRANDE	0,75	0,85	0,89	0,84	1,08	0	0	0	0	0	
			BOOK-TO-MARKET									
			$\beta_2$					<i>p-value</i>				
			BAIXO	2	3	4	ALTO	BAIXO	2	3	4	ALTO
		PEQUENA	0,85	2,06	1,05	0,79	1,08	0	0	0	0	0
		2	1,39	0,56	0,56	0,69	0,19	0	0	0	0	0,17
	3	0,67	0,29	0,49	0,47	0,03	0	0	0	0	0,91	
	4	0,61	0,49	0,35	0,40	-0,08	0	0	0,02	0,02	0,74	
	GRANDE	0,19	-0,05	-0,13	0,03	-0,25	0,12	0,53	0,31	0,81	0,13	

(continua)

**TABELA 1 (CONTINUAÇÃO)**

**DADOS DA REGRESSÃO INDIVIDUAL DE  
 CADA CARTEIRA FORMADA POR QUINTIS**

		BOOK-TO-MARKET										
		$\beta_3$					p-value					
		BAIXO	2	3	4	ALTO	BAIXO	2	3	4	ALTO	
TAMANHO	PEQUENA	-0,77	-1,23	-0,42	-0,21	0,21	0	0,01	0,01	0,07	0,03	
	2	-1,09	-0,15	-0,16	-0,12	0,39	0	0,26	0,18	0,41	0	
	3	-0,60	-0,09	0,01	0,22	1,26	0	0,48	0,90	0,10	0	
	4	-0,58	-0,20	-0,13	-0,23	0,38	0	0,18	0,33	0,13	0,07	
	GRANDE	-0,47	-0,04	0,12	-0,03	0,40	0	0,63	0,33	0,78	0,01	
			BOOK-TO-MARKET									
			R <sup>2</sup>					NÚMERO DE ATIVOS				
			BAIXO	2	3	4	ALTO	BAIXO	2	3	4	ALTO
		PEQUENA	0,45	0,21	0,29	0,38	0,63	3	3	6	10	19
		2	0,19	0,37	0,29	0,32	0,48	3	10	9	12	7
	3	0,45	0,46	0,41	0,48	0,38	5	11	13	4	8	
	4	0,43	0,41	0,45	0,43	0,25	15	6	9	6	5	
	GRANDE	0,58	0,77	0,64	0,56	0,64	15	11	4	9	2	

Fonte: Elaborada pelos autores.

Como dito, a maioria dos parâmetros se mostrou significativa, conforme testes de primeira ordem na Tabela 1 (testes *t*), e esse resultado se confirmou na regressão em painel, também utilizando o *software eviews*. A equação de regressão em painel, então, foi:

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

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

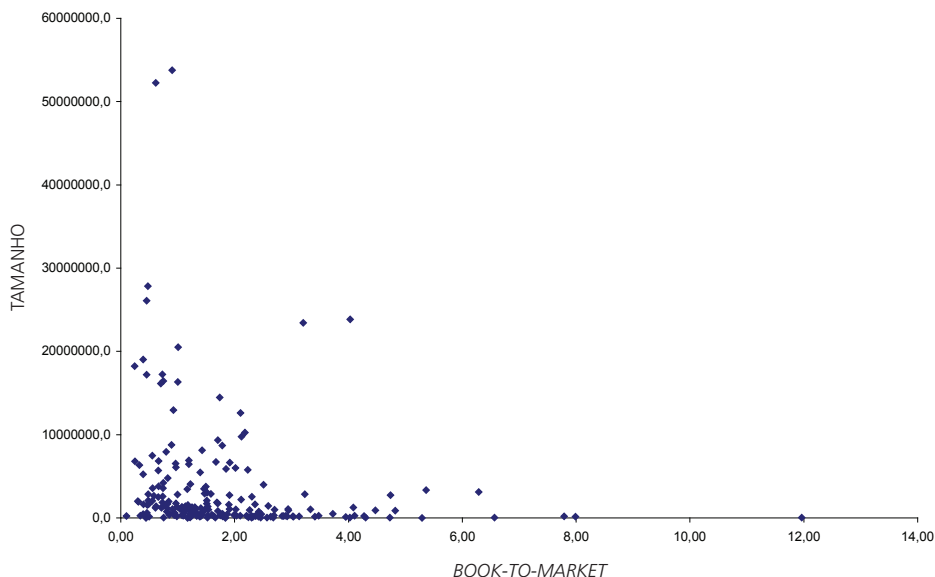
Nos resultados dessa equação, os valores entre parênteses abaixo dos parâmetros são o *t* encontrados para os testes de primeira ordem. Pode-se afirmar que todos são significantes, e que as duas anomalias consideradas pelos autores,

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 por meio 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 por meio 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 retirá-los e posteriormente recolocá-los em um grupo similar, no qual a observação possa pertencer. No Gráfico 1, percebe-se que não há condições de determinar os *clusters* com alguns resultados muito dispersos.

GRÁFICO 1

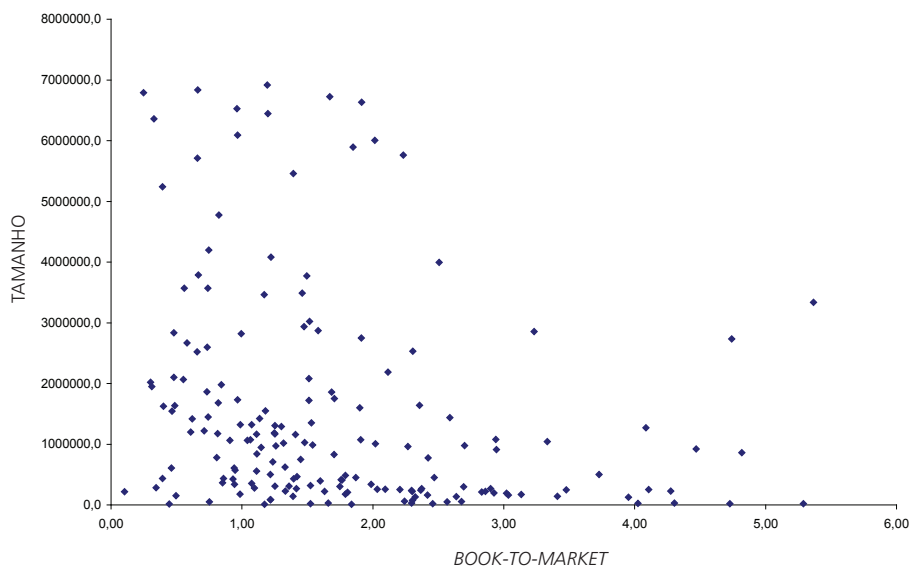
DISTRIBUIÇÃO DOS DADOS DAS 205 EMPRESAS ESTUDADAS  
(BOOK-TO-MARKET E TAMANHO) COM OUTLIERS



Fonte: Elaborado pelos autores.

GRÁFICO 2

DISTRIBUIÇÃO DOS DADOS DAS 205 EMPRESAS ESTUDADAS  
(BOOK-TO-MARKET E TAMANHO) SEM OUTLIERS



Fonte: Elaborado pelos autores.

Quando os *outliers* são retirados (Gráfico 2), pode-se perceber que alguns grupos com características comuns se formam. Vale ressaltar que esse gráfico está refletindo a análise conjunta que será apresentada mais adiante, na regressão da última equação. Se fossem formados grupos (*clusters*) a partir das informações que formaram o Gráfico 1, haveria um *cluster* com muitas empresas, na verdade quase a totalidade, e cada *outlier* representaria um *cluster* específico. No Gráfico 1, pode-se verificar que uma empresa tem um *book-to-market* bem superior às demais (próximo de 12). Essa 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 2**

**DADOS DA REGRESSÃO INDIVIDUAL DE CADA CARTEIRA FORMADA POR ANÁLISE DE CLUSTERS**

BOOK-TO-MARKET										
TAMANHO	$\beta_1$					p-value				
	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
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
BOOK-TO-MARKET										
TAMANHO	$\beta_2$					p-value				
	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
BOOK-TO-MARKET										
TAMANHO	$\beta_3$					p-value				
	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

(continua)

**TABELA 2 (CONTINUAÇÃO)**

**DADOS DA REGRESSÃO INDIVIDUAL DE CADA  
 CARTEIRA FORMADA POR ANÁLISE DE CLUSTERS**

		BOOK-TO-MARKET									
		R <sup>2</sup>					NÚMERO DE ATIVOS				
		BAIXO	2	3	4	ALTO	BAIXO	2	3	4	ALTO
TAMANHO	PEQUENA	0,59	0,27	0,59	0,17	0,23	19	5	32	1	2
	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

Fonte: Elaborada pelos autores.

A regressão em painel do modelo de Fama e French (1996) e seus resultados são mostrados a seguir.

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

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

Como pode ser visto nos resultados da aplicação do modelo de Fama e French (1996), quando rodado por carteiras formadas por meio de análise de *clusters*, os resultados são melhores, como observado no coeficiente de determinação, que nesse 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 um. 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 (1996), 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 nessa nova metodologia esse parâmetro perdesse força, mesmo mostrando significância estatística.



## 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 a respeito das relações que o comportamento dos títulos disponíveis no mercado tem com as diversas variáveis, como o comportamento da média do mercado, ou mesmo com o que é chamado de anomalia, assunto muito discutido com base em alguns trabalhos de Fama (1991) e Fama e French (1996).

A literatura relevante nessa 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 por meio de medidas de posição de ordenação, como os quintis, quartis ou percentis. O que se pretendia aqui era apresentar uma forma alternativa de formação de carteiras, conhecida como análise de *clusters*, que agrupa as variáveis por meio 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.

Pela análise de *clusters* e regressão de multifatores de Fama e French (1996), descobriu-se que, no caso brasileiro, podem ser encontrados resultados mais satisfatórios que vão ao encontro da hipótese inicial deste trabalho, de que algumas técnicas de estatística multivariada, nesse 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 ter características mais homogêneas do que simplesmente a ordenação por meio de medidas de posição.

Apesar de não necessariamente apresentar resultados melhores que os quintis, essa metodologia pode ser útil em trabalhos em que o simples ordenamento não forme grupos satisfatoriamente homogêneos.

A principal limitação deste estudo 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, nesse caso, muitas empresas teriam que ser retiradas da amostra. Isso comprometeria um dos principais objetivos deste trabalho, que era selecionar uma amostra

ampla e que o modelo a ser proposto posteriormente fosse formado a partir de uma boa representação do mercado brasileiro de ações. Nesse 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 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 considerado válido. 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 das décadas de 1970, 1980 e início dos anos 1990. 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 dessa 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.).

## REFERÊNCIAS

- ANDREWS, D. W. K. Tests for parameter instability and structural change with unknown change point. *Econometrica*, v. 61, n. 4, p. 821-856, 1993.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedacity. *Journal of Econometrics*, v. 31, p. 307-372, 1986.
- BONOMO, M.; AGNOL, I. D. Retornos anormais e estratégias contrárias. *Revista Brasileira de Finanças*, Rio de Janeiro, v. 1, n. 2, p. 165-215, dez. 2003.
- BOX, G. E. P.; JENKINS, G. M. *Time series analysis: forecasting and control*. San Francisco, California: Holden-Day, 1976.
- BRITO, N. R. O. Eficiência informacional fraca de mercados de capitais sob condições de inflação. *Revista Brasileira de Mercado de Capitais*, Rio de Janeiro, v. 4, n. 10, p. 63-85, jan./abr. 1978.
- \_\_\_\_\_. Eficiência informacional fraca no mercado à vista da BVRJ no período 1980/1984. *Relatório Técnico n° 78*, Rio de Janeiro, Coppead, 1985.
- CASSETTARI, A. Uma forma alternativa para alocação ótima de capital em carteiras de risco. *Revista de Administração da USP*, São Paulo, v. 36, n. 3, p. 70-85, abr./ago. 2001.
- CERETTA, P. S. Comportamento das variações de preços nos mercados da América Latina. In: ENCONTRO NACIONAL DA ANPAD, 25., 2001, Campinas. *Anais...* Campinas: Anpad, 2001.
- CHAN, K. C.; CHEN, N. Structural and return characteristics of small and large firms. *Journal of Finance*, v. 46, n. 4, p. 1467-1484, Sept. 1985.
- CONTADOR, C. Uma análise espectral dos movimentos da Bolsa de Valores do Rio de Janeiro. *Revista Brasileira de Mercado de Capitais*, Rio de Janeiro, Ibmecc, v. 1, n. 1, jan./abr. 1975.

- COSTA JR., N. C. A. Sobre-reação a longo prazo no mercado de capitais brasileiro. In: COSTA JR., N. C. A. da N.; LEAL, R. P. C.; LEMGRUBER, E. F. (Org.). *Mercado de capitais: análise empírica no Brasil*. São Paulo: Atlas, 2000. p. 174-183.
- COSTA JR., N. C. A.; NEVES, M. B. E. Variáveis fundamentalistas e retornos das ações. In: COSTA JR., N. C. A. da N.; LEAL, R. P. C.; LEMGRUBER, E. F. (Org.). *Mercado de capitais: análise empírica no Brasil*. São Paulo: Atlas, 2000. p. 99-110.
- DEBONDT, W. F. M.; THALER, R. Does the stock market overreact? *Journal of Finance*, v. 40, n. 3, p. 793-805, July 1985.
- ECONOMÁTICA: banco de dados. Disponível em: <<http://www.economatica.com.br>>. Acesso: entre set. 2004 e fev. 2005.
- ENGLE, R. F. *ARCH: selected readings*. Oxford: Oxford University Press, 1995.
- ESTRADA, J. Systematic risk in emerging markets: the D-CAPM. *Emerging Markets Review*, v. 3, n. 4, p. 365-379, 2002.
- FAMA, E. F. Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, v. 25, p. 383-417, 1970.
- \_\_\_\_\_. Efficient capital markets II. *Journal of Finance*, v. 46, p. 1575-1617, 1991.
- FAMA, E. F.; FRENCH, E. Multifactor explanations of asset pricing anomalies. *Journal of Finance*, v. 51, n. 1, p. 55-84, March 1996.
- GREENE, W. *Econometric analysis*. 4. ed. New York: Prentice-Hall, 2000.
- GRINBLATT, M.; MOSKOWITZ, T. J. Predicting stock price movements from past returns: the role of consistency and tax-loss selling. *Journal of Financial Economics*, v. 71, p. 541-579, 2004.
- HAIR, J. F. et al. *Multivariate data analysis*. 4. ed. New York: Prentice Hall, 1995.
- HAN, J.; KAMBER, M. *Data mining: concepts and techniques*. San Francisco: Morgan Kaufmann, 2001.
- JEGADEESH, N.; TITMAN, S. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, v. 48, p. 65-91, 1993.
- \_\_\_\_\_. Profitability of momentum strategies: an evolution of alternative explanations. *Journal of Finance*, v. 61, p. 699-720, 2001.
- JOHNSTON, J.; DINARDO, J. *Métodos econométricos*. 4. ed. São Paulo: McGraw Hill, 2001.
- LEAL, R.; AMARAL, A. S. Um momento para o “insider trading”: o período anterior ao anúncio de uma emissão pública de ações. *Revista Brasileira de Mercado de Capitais*, Rio de Janeiro, Ibmecc, v. 15, n. 41, p. 21-26, 1990.
- LEMGRUBER, E. F.; BECKER, J. L.; CHAVES, T. B. S. O efeito fim de semana no comportamento dos retornos diários de índices de ações. In: ENCONTRO NACIONAL DA ANPAD, 12., 1998, Belo Horizonte. *Anais...* Belo Horizonte: Anpad, 1988. p. 873-878.
- LO, A. W.; MaCKINLAY, A. C. *A non-random walk down wall street*. USA: Princeton University Press, 1999.
- LUCENA, P.; FIGUEIREDO, A. C. Pressupostos de eficiência de mercado: um estudo empírico na Bovespa. *Revista Eletrônica de Gestão Organizacional*, v. 2, n. 3, set./dez. 2004. Disponível em: <<http://www.gestaoorg.dca.ufpe>>. Acesso em: 15 mar. 2005.
- LUCENA, P.; MOTTA, A. C. da. Aplicação de um novo modelo de análise de risco na Bovespa: o D-CAPM. *Revista Eletrônica de Administração*, v. 10, n. 5, 2004. Disponível em: <<http://read.adm.ufrgs.br>>. Acesso em: 10 abr. 2005.

- LUCENA, P.; ANDRES, T.; NESS, W. L. Não-normalidade, autocorrelação dos retornos e existência de caudas densas: um estudo empírico na Bovespa. In: TOMEI, P. A.; WETZEL, U. (Org.). *Estudos em negócios II*. Rio de Janeiro: Mauad, 2003. p. 17-32.
- MALAGA, F. K.; SECURATO, J. R. Aplicação do modelo de três fatores de Fama e French no mercado acionário brasileiro: um estudo empírico do período 1995-2003. ENCONTRO NACIONAL DA ANPAD, 28., 2004, Curitiba. *Anais...* Curitiba: Anpad, 2004. p. 1-16.
- MARKOWITZ, H. Portfolio selection. *Journal of Finance*, v. 8, p. 77-91, 1952.
- MENEZES, J. C. F. *Mercado acionário brasileiro: a evolução recente e a sua eficiência informacional*. 1981. Dissertação (Mestrado)–Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.
- NAKAMURA, W. T.; CAMARGO JR., A. S. Análise da validade dos modelos CAPM e APT no mercado brasileiro de ações. *Working Paper Investsul*, 2003. Disponível em: <[http://www.investsul.com.br/textos\\_academicos](http://www.investsul.com.br/textos_academicos)>. Acesso em: 15 mar. 2005.
- NEVES, M. B. E. das; LEAL, R. P. C. Existe relação entre o crescimento do PIB brasileiro e os efeitos tamanho, valor e momento? In: ENCONTRO NACIONAL DA ANPAD, 17., 2003, Atibaia. *Anais...* Atibaia: Anpad, 2003. p. 1-16.
- PROCIANOY, J. L.; ANTUNES, M. A. Os efeitos das decisões de investimento das firmas sobre os preços de suas ações no mercado de capitais. In: ENCONTRO BRASILEIRO DE FINANÇAS, 1., 2001, São Paulo. *Anais...* São Paulo, 2001. Disponível em: <<http://www.sbfin.org.br>>. Acesso em: 15 mar. 2005.
- ROSS, S. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, v. 13, p. 341-360, 1976.
- SHARPE, W. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, v. 19, p. 425-442, 1964.
- STOCK, J.; WATSON, M. W. *Econometria*. São Paulo: Adson Wesley, 2004.
- TORRES, R.; BONOMO, M.; FERNANDES, C. A aleatoriedade do passeio na Bovespa: testando a eficiência do mercado acionário brasileiro. In: BONOMO, M. (Org.). *Finanças aplicadas ao Brasil*. Rio de Janeiro: FGV, 2002. p. 193-233.
- WHITE, H. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, v. 48, p. 817-838, 1980.