

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

VOLATILIDADE IDIOSINCRÁTICA E RETORNO DE AÇÕES NO MERCADO
BRASILEIRO

FERNANDO MATTOSO DE AZEVEDO

NO. DE MATRÍCULA: 810561

ORIENTADOR: RUY RIBEIRO

NOVEMBRO DE 2017

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

VOLATILIDADE IDIOSINCRÁTICA E RETORNO DE AÇÕES NO MERCADO
BRASILEIRO

FERNANDO MATTOSO DE AZEVEDO

NO. DE MATRÍCULA: 810561

ORIENTADOR: RUY RIBEIRO

NOVEMBRO DE 2017

Declaro que o presente trabalho é de minha autoria e que não recorri para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor.

As opiniões expressas neste trabalho são de responsabilidade única e exclusiva do autor.

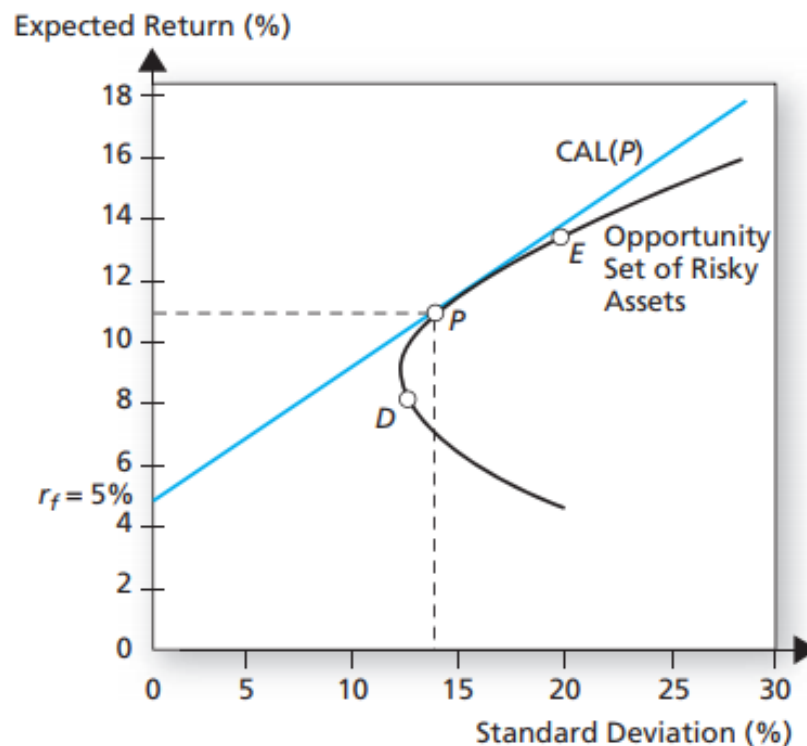
SUMÁRIO

1 INTRODUÇÃO	5
2 REVISÃO DA LITERATURA.....	8
3 DADOS	10
4 MÉTODO.....	13
5 RESULTADOS	16
6 CONCLUSÃO.....	18
7 REFERÊNCIAS BIBLIOGRÁFICAS	19
8 ANEXO: SCRIPT	21

1 INTRODUÇÃO

A Teoria Moderna de Portfólios, iniciada por Markowitz (1952), parte da ideia de que agentes racionais vão buscar otimizar suas carteiras de investimentos, de modo a conseguir o maior retorno possível com o menor risco possível. Entre portfólios com o mesmo risco (definido como desvio-padrão do retorno), não faria sentido um investidor escolher o portfólio que não fosse aquele de maior retorno. Desta maneira, as escolhas ótimas se encontram todas na chamada fronteira eficiente., conforme a Figura 1.

Figura 1:



Fonte: Bodie, Kane and Marcus, 2014.

A partir desta teoria, surge o modelo CAPM, desenvolvido por Sharpe (1964) e Lintner (1965), que relaciona o retorno esperado de ações individuais ou o retorno de

portfólios com o retorno esperado do portfólio de mercado. A fórmula do modelo CAPM é a seguinte:

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

Onde $E(r_i)$ é a taxa de retorno esperada do ativo ou portfólio i , R_f é a taxa de juros livre de risco, $E(r_m)$ é a taxa de retorno esperada do portfólio de mercado e β_i é a sensibilidade do valor esperado de retorno do ativo ou portfólio i como relação a variações no excesso de retorno esperado do portfólio de mercado.

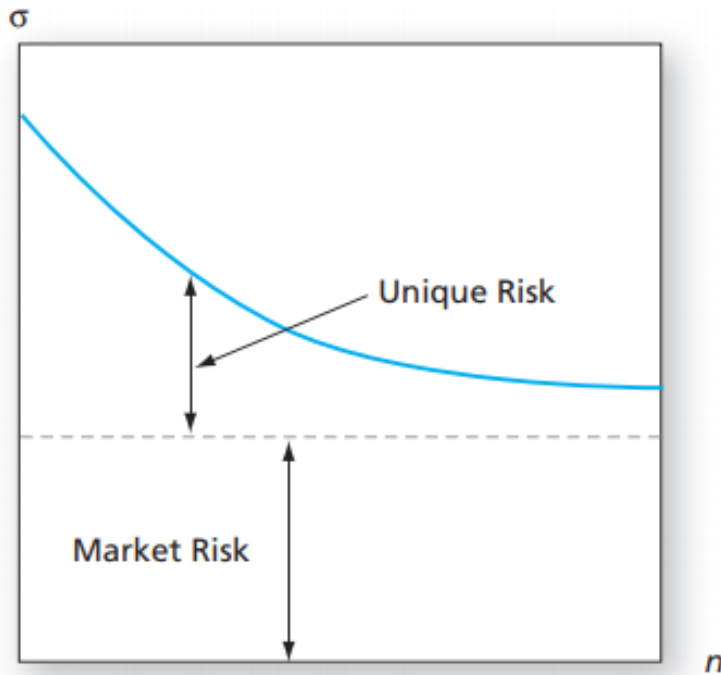
Outro modelo de precificação de ativos que tem bastante relevância e melhor respaldo empírico em relação ao CAPM é o modelo de Fama e French (1996), que é um modelo com três fatores. Além do risco de mercado, os autores incluem um fator relacionado ao tamanho da empresa (seu valor de mercado) e à razão *book-to-market*, que é a razão entre o seu valor contábil e seu valor de mercado (valor patrimonial por ação / preço). A fórmula do modelo Fama-French é a seguinte:

$$R_{it} = \alpha_i + \beta_i R_{mt} + h_i HML_t + s_i SMB_t + \varepsilon_{it}$$

Onde R_{it} é o excesso de retorno do ativo i no período t , R_{mt} é o excesso de retorno do portfólio de mercado, HML_t é o fator de valor no ano t , SMB_t é o fator de tamanho no ano t , α_i é o intercepto que indicaria a habilidade de um gestor em conseguir um retorno maior do que o que a teoria prediria, ε_{it} o resíduo da regressão e β_i , h_i e s_i as sensibilidades do excesso de retorno em relação a cada um dos fatores de risco.

Em ambos os modelos, há a ideia de um risco idiossincrático, que tem a ver com características próprias de cada firma e um risco sistêmico, um risco de mercado, ao qual todas as firmas estão expostas. Na Teoria Moderna dos Portfólios, quanto maior o número de ativos numa carteira, menor o risco total. Com uma diversificação perfeita, o risco total seria igual ao risco sistêmico, conforme se pode observar na Figura 2, onde n é o número de ativos na carteira e σ o desvio padrão do retorno do portfólio.

Figura 2:



Fonte: Bodie, Kane and Marcus, 2014.

No entanto, estudos mais recentes como o de Ang et al (2006) mostram que portfólios com volatilidade idiossincrática (definida como desvio-padrão dos resíduos do modelo de 3 fatores de Fama e French) diferentes apresentam retornos diferentes. Estes autores interpretam a volatilidade idiossincrática como um fator de risco a ser acrescentado ao modelo de Fama e French (1996). O presente trabalho utiliza a metodologia de Ang et al (2006) e estuda a correlação entre volatilidade idiossincrática e o retorno de ações no mercado brasileiro, partindo de um modelo estendido de Fama e French (1996) com um fator a mais, a diferença de retorno de ações com alta volatilidade idiossincrática e o retorno de ações de baixa volatilidade idiossincrática.

Em todo o trabalho, foi utilizado o *software* R para fazer a limpeza e a análise dos dados. O *script* anexo contém o passo a passo da pesquisa, o que permite replicar a análise.

2 REVISÃO DA LITERATURA

Segundo o modelo CAPM de Sharpe (1964) e Lintner (1965), o risco idiossincrático deveria tender a zero conforme se diversificam os portfólios, sobrando apenas o risco sistemático na precificação.

Fama e French (1992) perceberam que valor de mercado e razão *book-to-market* são fatores a se levar em conta na hora de se precificar ativos. Estes fatores são considerados riscos específicos das firmas que não são capturados pelo fator de mercado do CAPM.

Em Fama e French (1996), o valor esperado do excesso de retorno de um portfólio i em relação à taxa livre de risco, $[E(R_i) - R_f]$ é explicado pela sensibilidade do retorno a três fatores: (i) o excesso de retorno de um portfólio de mercado, (ii) a diferença de retorno entre um portfólio composto de ações pequenas e um portfólio composto de ações grandes (SMB, *small minus big*) e (iii) a diferença de retorno entre um portfólio composto de ações com alta razão *book-to-market* e um portfólio composto de ações com baixa razão *book-to-market* (HML, *high minus low*):

$$E(R_i) - R_f = b_i[E(R_M) - R_f] + s_iE(\text{SMB}) + h_iE(\text{HML})$$

Onde $[E(R_i) - R_f]$, $E(\text{SMB})$ e $E(\text{HML})$ são retornos esperados e as sensibilidades b_i , s_i e h_i são as inclinações da seguinte regressão:

$$R_i - R_f = \alpha_i + b_i(R_M - R_f) + s_i\text{SMB} + h_i\text{HML} + \varepsilon_i.$$

Ang et al (2006) definem a volatilidade idiossincrática como a volatilidade dos resíduos no modelo de Fama e French (1996) descrito acima.

Merton (1987) sugere que investidores podem ser recompensados com retornos mais altos por terem portfólios subdiversificados e com maior volatilidade idiossincrática.

Indivíduos podem não manter portfólios diversificados por diversas razões, entre elas custos de transação e de informação.

Ang et al (2006) testam a volatilidade idiossincrática como um fator esquecido no modelo de Fama e French (1996) e chegam à conclusão de que portfólios com alta (baixa) volatilidade idiossincrática apresentam retorno esperado menor (maior), resultado que é contraditório com a teoria de Merton e com as premissas da Teoria do Portfólio, de que o risco idiossincrático pode ser eliminado através da diversificação.

Desta forma, um modelo à maneira de Fama e French (1996), mas com um fator a mais que capture o risco da volatilidade idiossincrática, explicaria melhor o retorno das ações.

3 DADOS

Para construir os fatores e calcular os betas do modelo Fama-French(1996), foram utilizados dados do Economática, disponível no IAG PUC-Rio.

Os dados obtidos foram de 381 empresas ativas do mercado acionário brasileiro, sendo 597 ações de classes diferentes destas empresas. Para montar os portfólios de valor e calcular os fatores HML, foram utilizados dados de valor patrimonial por ação (VPA) e preço da ação. O quociente VPA/P nos dá a razão *book-to-market*.

Os dados de VPA estão em reais e ajustados para proventos e os dados de preço são dados de fechamento diário, em reais, não ajustado para proventos.

Para o fator SML, de tamanho, foram utilizados dados de valor de mercado das empresas, que são dados diários, em milhões de reais. O valor de mercado é definido como o número de ações em circulação multiplicado pelo preço da ação.

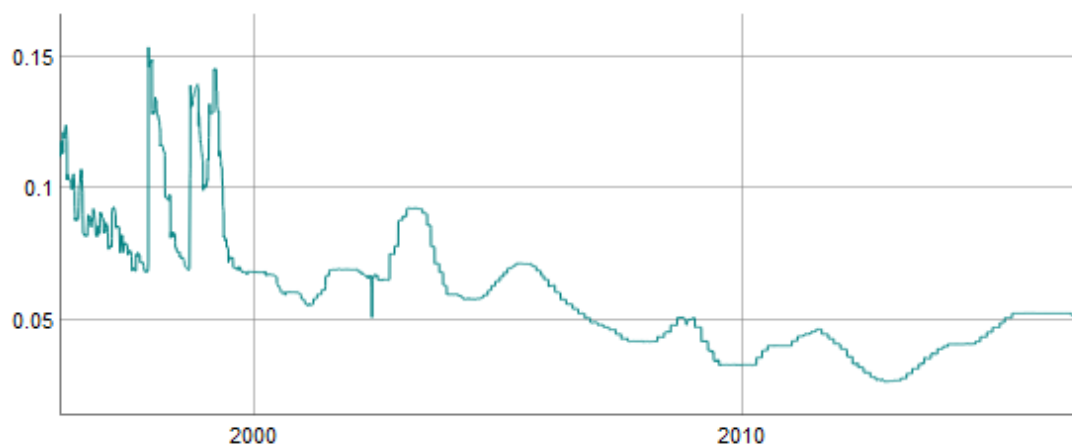
Também foram utilizados dados de retorno diários das ações e dados de retorno diário do IBOV, índice utilizado como referência para um portfólio de mercado.

Os dados de retorno das ações são dados diários de fechamento, com ajuste de proventos e os dados de retorno do IBOV são dados de retorno diário, também com ajuste de proventos.

Os dados das taxas livres de risco foram obtidos a partir da série da Cetip, com dados diários de 1/1/1996 até 31/12/2016. Para obter a série, foi utilizado um pacote da linguagem R de programação chamado BETS (*Brazilian Economic Time Series*), que importa séries produzidas por instituições como FGV/IBRE, IBGE e BACEN.

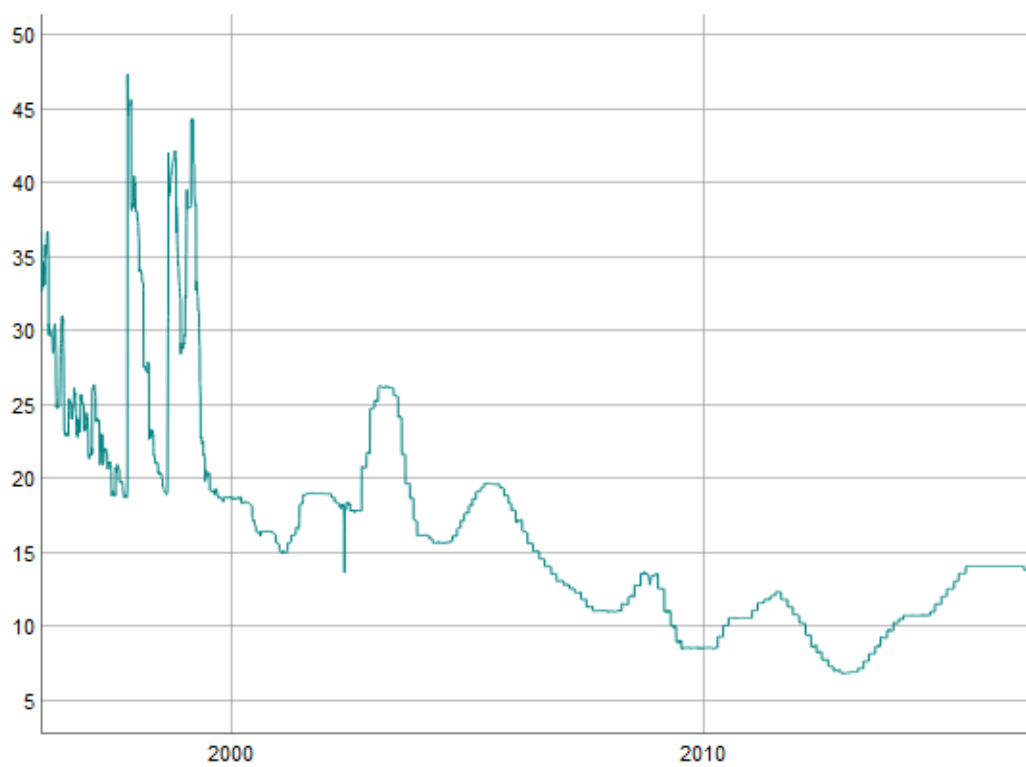
A Figura 3 apresenta a evolução da taxa CDI no período 1996-2016, em % ao dia. O gráfico foi gerado dentro do R, enquanto a Figura 4 apresenta a evolução do CDI em % ao ano.

Figura 3:



Fonte: Elaboração própria

Figura 4:



Fonte: Elaboração própria

O período de estudo é de 21 anos, de 1/1/1996 a 31/12/2016. De cada variável, foram obtidos dados referentes a esse período. Em alguns casos foi preciso fazer um tratamento, levando em conta o fato de só haver dados para dias úteis.

Ao montar os dados em painel foram filtradas as datas comuns a todos os *data frames*, que eram: retorno diário das ações, cdi diário, vpa diário, preço diário, valor de mercado diário e retorno do IBOV diário.

Com estes dados foi possível montar os fatores SMB e HML diários e, com uma regressão com dados em painel utilizando o modelo Fama e French (1993) de três fatores, obter a volatilidade idiossincrática, definida como em Ang et al (2009) como o desvio padrão dos resíduos (ε_i) desta regressão.

$$R_i - R_f = \alpha_i + b_i(R_M - R_f) + s_i\text{SMB} + h_i\text{HML} + \varepsilon_i.$$

4 MÉTODO

A metodologia utilizada se baseia em Liu e Iorio (2012), Ang et al (2009), Fama e French (1993) e Drew, Naughton e Veeraraghavan (2006).

Após tratar os dados, cria-se a razão *book-to-market*, que é definida como Valor Patrimonial por Ação / Preço (VPA/P). Com dados diários de VPA/P e Valor de Mercado (ME), podemos ranquear as empresas segundo estes dois critérios para construir 6 portfólios, conforme Fama e French (1996).

Para construir os portfólios utilizamos dados de valor de mercado do mês de junho de cada ano t e dados de VPA e preço do mês de dezembro de cada ano t .

Quadro 1:

Carteira	VM	VPA/P
1	S	V
2		N
3		G
4	B	V
5		N
6		G

Fonte: Elaboração própria

As ações são divididas entre abaixo da mediana do valor de mercado e acima e divididas em tercis de acordo com a razão *book-to-market*.

Desta maneira, temos 6 combinações, que são as carteiras SV, SN, SG, BV, BN e BG, S significando *small*, baixo valor de mercado, B significando *big*, alto valor de mercado, V significando *value*, alta razão *book-to-market*, N significando *neutral*, razão *book-to-market* intermediária e G significando *growth*, baixa razão *book-to-market*.

Para construir os fatores SMB (*small minus big*) e HML (*high minus low*), calculamos a diferença de retorno diária das estratégias da seguinte maneira:

SMB é a média de retorno dos três portfólios pequenos menos a média de retorno dos três portfólios grandes.

$$\text{SMB} = 1/3 (\text{Small Value} + \text{Small Neutral} + \text{Small Growth}) \\ - 1/3 (\text{Big Value} + \text{Big Neutral} + \text{Big Growth}).$$

HML é a média de retorno dos dois portfólios de valor menos a média de retorno dos dois portfólios de crescimento.

$$\text{HML} = 1/2 (\text{Small Value} + \text{Big Value}) \\ - 1/2 (\text{Small Growth} + \text{Big Growth}).$$

HML é a média de retorno dos dois portfólios de valor menos a média de retorno dos dois portfólios de crescimento.

Após calcularmos os fatores SMB e HML diários transformamos todos os dados diários em dados mensais. São eles retorno das ações, retorno do IBOV, CDI, SMB e HML.

Com isso, temos todos os elementos para rodar uma regressão do modelo e estimar os betas mensais, utilizando as ações presentes em pelo menos um dos seis portfólios.

A partir da regressão do modelo Fama e French (1996) de três fatores, separamos os resíduos e calculamos seu desvio padrão para cada mês.

O desvio padrão dos resíduos é a volatilidade idiossincrática, como definida por Ang et al (2009).

Montamos então dois portfólios, um com alta volatilidade idiossincrática (HI) e outro com baixa volatilidade idiossincrática (LI). A partir destes portfólios, criamos o fator HIMLI (*high idiosyncratic minus low idiosyncratic*).

Este fator é testado em uma regressão num modelo de dois fatores, dois modelos de três fatores e um modelo de quatro fatores.

O modelo básico de dois fatores é o seguinte:

$$R_{it} = \alpha_i + \beta_i R_{mt} + i_t HIMLI_t + \varepsilon_{it}$$

Depois, o fator HIMLI é testado na presença de cada um dos outros fatores, SMB e HML:

$$R_{it} = \alpha_i + \beta_i R_{mt} + h_i HML_t + i_t HIMLI_t + \varepsilon_{it}$$

$$R_{it} = \alpha_i + \beta_i R_{mt} + s_i SMB_t + i_t HIMLI_t + \varepsilon_{it}$$

Por último, fazemos a regressão do modelo com os quatro fatores:

$$R_{it} = \alpha_i + \beta_i R_{mt} + h_i HML_t + s_i SMB_t + i_t HIMLI_t + \varepsilon_{it}$$

Onde R_{it} é o excesso do portfólio i , R_{mt} é o excesso de retorno do portfólio de mercado (IBOV), SMB e HML são os fatores que incorporam o risco de tamanho e razão *book-to-market* e HIMLI é o fator que incorpora o risco de volatilidade idiossincrática.

5 RESULTADOS

A metodologia foi seguida, criando-se os portfólios e os fatores SMB e HML, calculados como a diferença de retorno entre as estratégias.

Após várias etapas de tratamento, que podem ser verificadas no script anexo, foram montadas matrizes com dados mensais de retorno, cdi, retorno do portfólio de mercado, SMB e HML.

Estas matrizes servem de input para rodar uma regressão do modelo Fama e French (1996).

A partir dos resíduos da regressão, é possível criar portfólios de acordo com a volatilidade idiossincrática (desvio padrão dos resíduos) e calcular o fator HIMLI a partir da diferença de retorno dos dois portfólios.

Com o fator HIMLI, podemos testar diferentes versões de modelos. Apenas com o risco de mercado e o fator HIMLI, o modelo de três fatores com HIMLI substituindo cada um dos fatores SMB e HML e, por último, um modelo com quatro fatores.

Tabela 1: Estatísticas descritivas dos retornos diários dos portfólios value-weighted

	BG	BN	BV	SG
Min.	:-9.0161	Min. :-17.55589	Min. :-15.29533	Min. :-13.42
92				
1st Qu.:	-0.5726	1st Qu.: -0.79864	1st Qu.: -1.09628	1st Qu.: -0.49
16				
Median :	0.1041	Median : 0.09997	Median : 0.08167	Median : 0.02
76				
Mean :	0.1479	Mean : 0.10955	Mean : 0.13196	Mean : 0.22
16				
3rd Qu.:	0.8406	3rd Qu.: 1.03480	3rd Qu.: 1.26129	3rd Qu.: 0.71
72				
Max. :	16.3451	Max. : 17.90815	Max. :169.76633	Max. : 51.22
49				
NA's :	78	NA's :77	NA's :77	NA's :778
	SN	SV		
Min. :	-11.1655	Min. :-14.67927		
1st Qu.:	-0.3348	1st Qu.: -0.34724		
Median :	0.0025	Median : 0.00458		
Mean :	0.1756	Mean : 0.15120		
3rd Qu.:	0.4836	3rd Qu.: 0.49212		
Max. :	39.6945	Max. : 29.50125		
NA's :	555	NA's :239		

Fonte: Elaboração própria

Tabela 2: Estatísticas descritivas da série de CDI diária

date	taxa cdi diária
Min. :1996-01-02	Min. :0.02626
1st Qu.:2001-04-05	1st Qu.:0.04183
Median :2006-07-04	Median :0.05568
Mean :2006-07-04	Mean :0.05962
3rd Qu.:2011-10-03	3rd Qu.:0.06925
Max. :2016-12-30	Max. :0.15400

Fonte: Elaboração própria

Tabela 3: Estatísticas descritivas dos retornos diários da carteira de mercado (IBOV)

ibov_datas	ibov_retorno
Min. :1996-01-02	Min. :0.0400
1st Qu.:2001-04-05	1st Qu.:0.2500
Median :2006-07-04	Median :0.4500
Mean :2006-07-04	Mean :0.3937
3rd Qu.:2011-10-03	3rd Qu.:0.5700
Max. :2016-12-30	Max. :0.6000
	NA's :76

Fonte: Elaboração própria

6 CONCLUSÃO

É difícil tirar conclusões a partir do modelo Fama e French (1996) para o mercado brasileiro, assim como quando incluímos o fator de volatilidade idiossincrática.

Por muito tempo tivemos juros altíssimos, o que torna o prêmio de risco negativo em muitos momentos. Por isso a prática de se utilizar os betas do mercado americano e acrescentar um risco Brasil.

No entanto, dado que trabalhos empíricos demonstram que a volatilidade idiossincrática pode ser incluída como um quarto fator no modelo Fama e French (1996) no mercado americano, podemos nos utilizar deste fato para estimarmos retornos de ativos também no Brasil.

7 REFERÊNCIAS BIBLIOGRÁFICAS

ANG, A.; HODRICK, R. J.; XING, J.; ZHANG, X. The cross section of volatility and expected returns. **The Journal of Finance**, v. 61, n.1, p.259-299, 2006.

ANG, A.; HODRICK, R. J.; XING, J.; ZHANG, X. High idiosyncratic volatility and low returns: International and further US evidence. **Journal of Financial Economics**, v. 91, n.1, p.1-23, 2009.

BODIE, Z.; KANE, A.; MARCUS, A. Investments. 10th Edition. New York: McGraw-Hill Education, 2014.

DREW, M. E., NAUGHTON, T., VEERARAGHAVAN, M. Is idiosyncratic volatility priced? Evidence from the Shanghai Stock Exchange. **International Review of Financial Analysis**, v.13, n.3, p.349-366, 2004.

FAMA, E.F; FRENCH, K.R. Common risk factors in the returns on stocks and bonds. **Journal of Financial Economics**, v.33, n.1, p.3-56, 1993.

FAMA, E.F.; FRENCH, K.R. Multifactor explanations of asset pricing anomalies. **The Journal of Finance**, v.51, n.1, p.55-84, 1996.

LINTNER, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. **The Review of Economics and Statistics**, v.43, p.13-37, 1965.

LIU, B.; IORIO, A. Idiosyncratic Volatility, Stock Returns and Economy Conditions: The Role of Idiosyncratic Volatility in the Australian Stock Market. **Global Finance Conference**, 2012.

MARKOWITZ, H. M. Portfolio Selection. **The Journal of Finance**, v.7, n.1, p.77-91, 1952.

MERTON, R. C. A simple model of capital market equilibrium with incomplete information. **Journal of Finance**, v.42, p. 483-510, 1987.

SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. **The Journal of Finance**, v.19, n.3, p.425-442, 1964.

8 ANEXO: SCRIPT

Script R

```

# Modelo Fama-French com dados do Economática

##### Pacotes #####

# Instalando

install.packages("data.table")
install.packages("xts")
install.packages("devtools")
install_github("pedrocostaferreira/BETS")
install.packages("dygraphs")
install.packages("matrixStats")
install.packages("janitor")
install.packages("plm")

# Carregando

library(data.table)
library(xts)
library(devtools)
library(BETS)
library(dygraphs)
library(matrixStats)
library(janitor)
library(plm)

##### Importação dos dados #####

# Diretório com bases de dados do Economática (input)

setwd('C:/Users/Fernando/Desktop/Economia PUC/Monografia/Ruy/Dados Econom
atica/R/input')

# Importando os dados de empresas ativas

mercado <- fread('ME_ativas.csv', skip = 1);
vpa <- fread('vpa_ativas.csv', skip = 1);
retorno <- fread('retorno_ativas_ajust.csv', skip = 1);
precos <- fread('quotes_najust_ativas.csv', skip = 1);

# Os warnings são por conta de valores NA que estão como "-" e são conver
tidos
# para character

```

```

# Importando dados diários do CDI (taxa de juros livre de risco)

# Procurar séries disponíveis que contenham a palavra CDI
BETS.search(description = "CDI")
# CDI anual e diário
cdi_anual <- BETS.get("4389", from = "1996-01-01", to="2016-12-30")
str(cdi_anual)

# Fonte: BCB Demab; % ao ano; CDI em termos anuais (base 252);
# periodicidade diária

cdi_xts <- xts(cdi_anual[,-1], order.by=cdi_anual[,1])
str(cdi_xts)

dygraph(cdi_xts)

cdi_diario <- BETS.get("12", from = "1996-01-01", to="2016-12-31")
cdi_diario_xts <- xts(cdi_diario[,-1], order.by=cdi_diario[,1])

dygraph(cdi_diario_xts)

# Fonte: Cetip; % ao ano; CDI em termos diários; periodicidade diária

# Importando dados de retorno do IBOV (retorno diário)

ibov_retorno <- fread("retorno_ibov_ajust.csv")

##### Tratamento dos dados #####

# Convertendo coluna 1 para formato de data
mercado$V1 <- as.Date(mercado$V1, "%m/%d/%Y")
vpa$V1 <- as.Date(vpa$V1, "%m/%d/%Y")
retorno$V1 <- as.Date(retorno$V1, "%m/%d/%Y")
precos$V1 <- as.Date(precos$V1, "%m/%d/%Y")
ibov_retorno$Data <- as.Date(ibov_retorno$Data, "%m/%d/%Y")
str(ibov_retorno)

datas <- mercado$V1

# Dados vão de 1/1/1996 até 30/12/2016

# Convertendo outras colunas para numeric

mercado <- data.frame(lapply(mercado[,-1], as.numeric), stringsAsFactors=
FALSE)
mercado <- cbind(datas, mercado)
vpa <- data.frame(lapply(vpa[,-1], as.numeric), stringsAsFactors=FALSE)
vpa <- cbind(datas, vpa)
retorno <- data.frame(lapply(retorno[,-1], as.numeric), stringsAsFactors=

```

```

FALSE)
retorno <- cbind(datas, retorno)
precos <- data.frame(lapply(precos[, -1], as.numeric), stringsAsFactors=FALSE)
precos <- cbind(datas, precos)
ibov_datas <- ibov_retorno$Data
ibov_retorno <- ibov_retorno[, -1]
ibov_retorno <- as.numeric(unlist(ibov_retorno))
str(ibov_retorno)
ibov_retorno <- data.frame(ibov_datas, ibov_retorno)
str(ibov_retorno)

# Warnings são por conta dos valores NAs

str(mercado)
str(vpa)
str(retorno)
str(precos)

# Verificando se colunas de vpa e preço são as mesmas
# Precisamos que sejam as mesmas para calcular a razão Book-to-market

colnames(mercado)==colnames(vpa)

# São iguais até a coluna 110. Investigando, chega-se à conclusão de
# que é uma coluna que está contida em apenas um dos dois data.frames (vpa)

head(mercado[109])
head(vpa[109])

# BPAC5 depois vem BPAC11 que só está no data frame vpa
head(vpa[110])

# Removendo a coluna 110
vpa <- vpa[, -110]

colnames(mercado)==colnames(vpa)

# Ao retirar a coluna 110 temos dois dfs do mesmo tamanho, mas ainda há colunas
# diferentes. Investigando, percebemos que são colunas com a ordem trocada.
# Trocando ordem para termos empresas na mesma ordem nos dois dataframes.

vpa <- vpa[, c(1:474, 476, 475, 477:598)]

colnames(mercado)==colnames(vpa)

vpa <- vpa[, c(1:508, 510, 509, 511:598)]

```



```

colnames(mercado)==colnames(vpa)

# Agora temos todas as colunas iguais e podemos calcular a razão B-t-M

##### Trabalhando com os dados para criar os portfólios #####

# Filtrando uma observação de junho de cada ano para montar os portfólios de
# acordo com tamanho (size) - fator SMB

startDate <- as.Date("1996-06-01")
ano_t <- seq(startDate, by="1 year", length.out=21)
size <- mercado[match(ano_t, mercado[,1]),]

ano_t <- as.Date(c("1996-06-03", "1997-06-03", "1998-06-01", "1999-06-01"
, "2000-06-01", "2001-06-01",
"2002-06-03", "2003-06-03", "2004-06-01", "2005-06-01"
, "2006-06-01", "2007-06-01",
"2008-06-03", "2009-06-01", "2010-06-01", "2011-06-01"
, "2012-06-01", "2013-06-03",
"2014-06-03", "2015-06-01", "2016-06-01"))

# Aos dias que caíram em finais de semana foram adicionados 2 dias
size <- mercado[match(ano_t, mercado[,1]),]

# Agora temos um valor de mercado para cada empresa para cada ano

# Fazendo o mesmo para BTM (VPA/P)

# Primeiro trabalhando com VPA

startDate <- as.Date("1996-12-31")
ano_t2 <- seq(startDate, by="1 year", length.out=21)

ano_t2 <- as.Date(c("1996-12-31", "1997-12-31", "1998-12-31", "1999-12-31"
, "2000-12-29", "2001-12-31",
"2002-12-31", "2003-12-31", "2004-12-31", "2005-12-30"
, "2006-12-29", "2007-12-31",
"2008-12-31", "2009-12-31", "2010-12-31", "2011-12-30"
, "2012-12-31", "2013-12-31",
"2014-12-31", "2015-12-31", "2016-12-30"))

vpa_dez <- vpa[match(ano_t2, vpa[,1]),]

# Agora trabalhando com preço

startDate <- as.Date("1996-12-30")
ano_t3 <- seq(startDate, by="1 year", length.out=21)

```

```

ano_t3 <- as.Date(c("1996-12-30", "1997-12-30", "1998-12-30", "1999-12-30",
", "2000-12-28", "2001-12-28",
", "2002-12-30", "2003-12-30", "2004-12-30", "2005-12-29",
", "2006-12-28", "2007-12-28",
", "2008-12-30", "2009-12-30", "2010-12-30", "2011-12-29",
", "2012-12-28", "2013-12-30",
", "2014-12-30", "2015-12-30", "2016-12-29"))

precos_dez <- precos[match(ano_t3, precos[,1]),]

# Dividindo VPA pelo preço (Book-to-market ratio)

anos <- seq(from = 1996, by = 1, length = 21)

BTM_anos <- cbind(anos, (vpa_dez[,-1]/precos_dez[,-1]))
BTM <- (vpa_dez[,-1]/precos_dez[,-1])

# Criando vetores com as medianas de cada ano de Size e BTM

m1 <- as.matrix.data.frame(BTM)
BTM_medianas <- rowMedians(m1, na.rm=T)
BTM_medianas

m2 <- as.matrix.data.frame(size[,-1])
size_medianas <- rowMedians(m2, na.rm=T)
size_medianas

# Separando quem está abaixo ou acima da mediana para Size
# Separando quem está em cada tercil para BTM

# Funções que retornam valor acima, abaixo ou no meio de certo valor

size_abaixo <- function (aux, valor) {

  return(which(aux<= valor))

}

size_acima <- function (aux, valor) {

  return(which(aux> valor))

}

size_meio <- function (aux, valor_inf, valor_sup) {

  return(which(aux> valor_inf & aux<= valor_sup))

```

```

}

# Criando as listas com valores acima ou abaixo da mediana e em cada terc
il

lista_abaixo <- NULL
lista_acima <- NULL
lista_t1 <- NULL
lista_t2 <- NULL
lista_t3 <- NULL

for (i in 1:nrow(m2)){

  mediana <- median(m2[i,], na.rm=T)
  tercil <- quantile(m1[i,], na.rm=T, prob = c(1/3,2/3))

  lista_abaixo[[i]] <- colnames(m2)[size_abaixo(m2[i,],mediana)]
  lista_acima[[i]] <- colnames(m2)[size_acima(m2[i,], mediana)]

  lista_t1[[i]] <- colnames(m1)[size_abaixo(m1[i,], tercil[1])]
  lista_t2[[i]] <- colnames(m1)[size_meio(m1[i,], tercil[1],tercil[2])]
  lista_t3[[i]] <- colnames(m1)[size_acima(m1[i,], tercil[2])]

}

lista_abaixo
lista_acima
lista_t1
lista_t2
lista_t3

##### Criando os 6 portfólios #####

# Os nomes se referem ao valor de mercado da ação (small - S e big - B)
# E à sua razão Book-to-market (value - V, neutral - N e growth - G)
# Os nomes abaixo são as combinações que correspondem aos 6 portfólios

SV <- NULL
SN <- NULL
SG <- NULL
BV <- NULL
BN <- NULL
BG <- NULL

for (i in 1:length(lista_acima)) {

  BG[[i]] <- intersect(lista_acima[[i]], lista_t1[[i]])
  BN[[i]] <- intersect(lista_acima[[i]], lista_t2[[i]])
  BV[[i]] <- intersect(lista_acima[[i]], lista_t3[[i]])

```

```

SG[[i]] <- intersect(lista_abaixo[[i]], lista_t1[[i]])
SN[[i]] <- intersect(lista_abaixo[[i]], lista_t2[[i]])
SV[[i]] <- intersect(lista_abaixo[[i]], lista_t3[[i]])
}

##### Construindo as estratégias diárias #####

# Datas em comum de todos
a <- intersect(retorno$datas, cdi_diario$date)
b <- intersect(a, vpa$datas)
C <- intersect(b, precos$datas)
d <- intersect(C, mercado$datas)
datas_comum <- intersect(d, ibov_retorno$ibov_datas)

# Convertendo datas do formato excel para formato de data do R
retorno$datas[1] - excel_numeric_to_date(9496)
datas_comum <- datas_comum + 25569

datas_comum <- excel_numeric_to_date(datas_comum)

# Filtrando os dados só para as datas em comum
retorno <- retorno[retorno$datas %in% datas_comum,]
cdi_diario <- cdi_diario[cdi_diario$date %in% datas_comum,]
vpa <- vpa[vpa$datas %in% datas_comum,]
precos <- precos[precos$datas %in% datas_comum,]
ibov_retorno <- ibov_retorno[ibov_retorno$ibov_datas %in% datas_comum,]
mercado <- mercado[mercado$datas %in% datas_comum,]

# Criando as estratégias diárias

# Os portfólios são value-weight

ANOS <- 1996:2016
estrategia <- NULL
for (i in 1:length(BG)){

  # Portfólio BG

  aux.retorno <- retorno[substr(retorno$datas,1,4)==ANOS[i], colnames(retorno) %in% BG[[i]]]
  aux.mercado <- mercado[substr(mercado$datas,1,4)==ANOS[i], colnames(mercado) %in% BG[[i]]]

  for (j in 1:nrow(aux.retorno)){

    r1 <- as.numeric(aux.retorno[j,])*as.numeric(aux.mercado[j,])
    if(all(is.na(r1))){

```

```

    estrategia$BG <- c(estrategia$BG, NA)
    next()

  }
  estrategia$BG <- c(estrategia$BG, sum(na.omit(r1))/sum(na.omit(as.numeric(aux.mercado[j,]))))

}

# Portfólio BN

aux.retorno <- retorno[substr(retorno$datas,1,4)==ANOS[i], colnames(retorno) %in% BN[[i]]]
aux.mercado <- mercado[substr(mercado$datas,1,4)==ANOS[i], colnames(mercado) %in% BN[[i]]]

for (j in 1:nrow(aux.retorno)){

  r1 <- as.numeric(aux.retorno[j,])*as.numeric(aux.mercado[j,])
  if(all(is.na(r1))){
    estrategia$BN <- c(estrategia$BN, NA)
    next()

  }
  estrategia$BN <- c(estrategia$BN, sum(na.omit(r1))/sum(na.omit(as.numeric(aux.mercado[j,]))))

}

# Portfólio BV

aux.retorno <- retorno[substr(retorno$datas,1,4)==ANOS[i], colnames(retorno) %in% BV[[i]]]
aux.mercado <- mercado[substr(mercado$datas,1,4)==ANOS[i], colnames(mercado) %in% BV[[i]]]

for (j in 1:nrow(aux.retorno)){

  r1 <- as.numeric(aux.retorno[j,])*as.numeric(aux.mercado[j,])
  if(all(is.na(r1))){
    estrategia$BV <- c(estrategia$BV, NA)
    next()

  }
  estrategia$BV <- c(estrategia$BV, sum(na.omit(r1))/sum(na.omit(as.numeric(aux.mercado[j,]))))

}

```

```

}

# Portfólio SG
aux.retorno <- retorno[substr(retorno$datas,1,4)==ANOS[i], colnames(retorno) %in% SG[[i]]]
aux.mercado <- mercado[substr(mercado$datas,1,4)==ANOS[i], colnames(mercado) %in% SG[[i]]]

for (j in 1:nrow(aux.retorno)){

  r1 <- as.numeric(aux.retorno[j,])*as.numeric(aux.mercado[j,])
  if(all(is.na(r1))){
    estrategia$SG <- c(estrategia$SG, NA)
    next()
  }
  estrategia$SG <- c(estrategia$SG, sum(na.omit(r1))/sum(na.omit(as.numeric(aux.mercado[j,]))))

}

# Portfólio SN

aux.retorno <- retorno[substr(retorno$datas,1,4)==ANOS[i], colnames(retorno) %in% SN[[i]]]
aux.mercado <- mercado[substr(mercado$datas,1,4)==ANOS[i], colnames(mercado) %in% SN[[i]]]

for (j in 1:nrow(aux.retorno)){

  r1 <- as.numeric(aux.retorno[j,])*as.numeric(aux.mercado[j,])
  if(all(is.na(r1))){
    estrategia$SN <- c(estrategia$SN, NA)
    next()
  }
  estrategia$SN <- c(estrategia$SN, sum(na.omit(r1))/sum(na.omit(as.numeric(aux.mercado[j,]))))

}

# Portfólio SV
aux.retorno <- retorno[substr(retorno$datas,1,4)==ANOS[i], colnames(retorno) %in% SV[[i]]]

```

```

  aux.mercado <- mercado[substr(mercado$datas,1,4)==ANOS[i], colnames(mercado) %in% SV[[i]]]

  for (j in 1:nrow(aux.retorno)){

    r1 <- as.numeric(aux.retorno[j,])*as.numeric(aux.mercado[j,])
    if(all(is.na(r1))){
      estrategia$SV <- c(estrategia$SV, NA)
      next()
    }
    estrategia$SV <- c(estrategia$SV, sum(na.omit(r1))/sum(na.omit(as.numeric(aux.mercado[j,]))))

  }
}

str(estrategia)
estrategia <- as.data.frame(estrategia)

# Estratégia Small Minus Big (SMB) diária
SMB <- (1/3)*(apply(estrategia[, 4:6], 1, sum) - apply(estrategia[,1:3], 1, sum))

# Estratégia High Minus Low (HML) diária
HML <- (1/2)*(apply(estrategia[,c(3,6)], 1, sum) - apply(estrategia[,c(1,4)], 1, sum))

SMB
str(SMB)
HML
str(HML)

##### Portfólio que vamos usar

# Utilizando ativos comuns às estratégias
portfolio <- NULL
for (i in 1:length(BG)){

  portfolio[[i]] <- sort(unique(c(BG[[i]], BN[[i]], BV[[i]], SG[[i]], SN[[i]], SV[[i]])))

}

```

```
##### Valores mensais #####

# Vetor com os meses
MESES <- c('01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12')

# Transformando tudo em mensal

retorno$ano <- as.character(substr(retorno$datas, 1, 4))
retorno$mes <- as.character(substr(retorno$datas, 6, 7))
cdi_diario$ano <- retorno$ano
cdi_diario$mes <- retorno$mes
ibov_retorno$ano <- retorno$ano
ibov_retorno$mes <- retorno$mes
# HML.bkp <- HML
HML <- data.frame(HML)
HML$ano <- retorno$ano
HML$mes <- retorno$mes
SMB <- data.frame(SMB)
SMB$ano <- retorno$ano
SMB$mes <- retorno$mes
is(HML)
is(SMB)
is(ibov_retorno)

for (i in length(HML))HML <- c(HML, HML[[i]])
# SMB.bkp <- SMB
for (i in length(SMB))SMB <- c(SMB, SMB[[i]])
HML <- as.data.frame(HML)
SMB <- as.data.frame(SMB)

retorno_mensal <- matrix(ncol = ncol(retorno)-3, nrow = 12*21)
cdi_mensal <- matrix(ncol = ncol(retorno)-3, nrow = 12*21)
ibov_mensal <- matrix(ncol = ncol(retorno)-3, nrow = 12*21)
hml_mensal <- matrix(ncol = ncol(retorno)-3, nrow = 12*21)
smb_mensal <- matrix(ncol = ncol(retorno)-3, nrow = 12*21)

opcoes <- expand.grid(MESES, ANOS)
for (i in 1:nrow(opcoes)) {

  aux <- retorno[retorno$ano == as.numeric(opcoes[i,2]) & retorno$mes ==
as.numeric(opcoes[i,1]),
                colnames(retorno) %in% portfolio[[which(ANOS == opcoes[i
,2])]]]
  aux <- (aux+100)/100
  aux.cdi <- cdi_diario[cdi_diario$ano == as.numeric(opcoes[i,2]) & cdi_d
iario$mes == as.numeric(opcoes[i,1]), ]
  aux.cdi <- (aux.cdi+100)/100
  aux.ibov <- ibov_retorno[ibov_retorno$ano == as.numeric(opcoes[i,2]) &
ibov_retorno$mes == as.numeric(opcoes[i,1]), ]
  aux.ibov <- (aux.ibov+100)/100
}
```



```

    aux.hml <- HML[HML$ano == as.numeric(opcoes[i,2]) & HML$mes == as.numeric(opcoes[i,1]),
                colnames(HML) %in% portfolio[[which(ANOS == opcoes[i,2])]]]
    aux.hml <- (aux.hml+100)/100
    aux.smb <- SMB[SMB$ano == as.numeric(opcoes[i,2]) & SMB$mes == as.numeric(opcoes[i,1]),
                colnames(SMB) %in% portfolio[[which(ANOS == opcoes[i,2])]]]
    aux.smb <- (aux.smb+100)/100
    # apply(aux, 2, prod)^(1/nrow(aux))

    for (j in 1:ncol(aux)){
        if(all(is.na(aux[,j]))){
            retorno_mensal[i,j] <- NA
            next()
        }

        retorno_mensal[i,j] <- (prod(na.omit(aux[,j]))^(1/length(na.omit(aux[,j])))-1)*100

        if(all(is.na(aux.cdi[,j]))){
            retorno_mensal[i,j] <- NA
            next()
        }

        cdi_mensal[i,j] <- (prod(na.omit(aux.cdi[,j]))^(1/length(na.omit(aux.cdi[,j])))-1)*100

        if(all(is.na(aux.ibov[,j]))){
            ibov_mensal[i,j] <- NA
            next()
        }

        ibov_mensal[i,j] <- (prod(na.omit(aux.ibov[,j]))^(1/length(na.omit(aux.ibov[,j])))-1)*100

        if(all(is.na(aux.hml[,j]))){
            hml_mensal[i,j] <- NA
            next()
        }
    }

```

```
hml_mensal[i,j] <- (prod(na.omit(aux.hml[,j]))^(1/length(na.omit(aux.
hml[,j]))) - 1) * 100

  if(all(is.na(aux.smb[,j]))){

    smb_mensal[i,j] <- NA
    next()

  }
  smb_mensal[i,j] <- (prod(na.omit(aux.smb[,j]))^(1/length(na.omit(aux.
smb[,j]))) - 1) * 100

}

}

##### Regressão Fama-French #####

ret.xcess <- retorno_mensal - cdi_mensal
merc.xcess <- ibov_mensal - cdi_mensal

ffregression <- lm(ret.xcess ~ merc.xcess + smb_mensal + hml_mensal, na.a
ction = na.exclude)

print(summary(ffregression))
```