

Prêmio por informação: uma investigação empírica da microestrutura do mercado acionário do Brasil

Fábio Massáud Caetano

Professor - Departamento de Economia – Universidade Federal de Pelotas (UFPel)
Endereço: Rua Gomes Carneiro, 1 – Pelotas/RS – Rio Grande do Sul – Brasil
CEP: 96010-610 – E-mail: fabio.massaud.caetano@ufpel.edu.br

Nelson Seixas dos Santos

Professor - Universidade Federal do Rio Grande do Sul (UFRGS)
Endereço: Avenida João Pessoa, 52 – Centro – Porto Alegre/RS – Brasil
CEP: 90040-000 – E-mail: nelson.seixas@ufrgs.br

Gilberto de Oliveira Kloeckner

Professor – Faculdade São Francisco de Assis
Endereço: Rua Florêncio Ygartua, 271 – Sala 204 – Porto Alegre/RS – Brasil
CEP: 90430-010 – E-mail: gilberto@saofranciscodeassis.edu.br

Recebido em 09 de setembro de 2013. Aceito em 09 de março de 2015.

Resumo

Este artigo estima a probabilidade de informação privilegiada (PIN) para ações do Índice Brasil durante o período de 2006 até 2011. O PIN é uma *proxy* para informação privada e é incorporado ao método de Fama French (1993) para separar os *portfolios* e explicar seus retornos. A combinação do PIN com as variáveis valor de mercado e índice *book-to-market* possibilita, para alguns *portfolios*, aumento do retorno e diminuição do risco. A significância das variáveis na explicação dos retornos é testada utilizando modelos de *Markov-switching*. Os resultados demonstram que o PIN é um fator importante na explicação dos retornos dos *portfolios*.

Palavras-Chave

Probabilidade de negociação informada. Modelos de mudança de regime markoviano. Seleção de *portfolios*. Algoritmo do *tick-test*.

Abstract

This article makes use of tick-by-tick data to estimate the probability of informed trading (PIN) to IBRX stocks. We add PIN in Fama French (1993) method to separate portfolios and explain their returns. Our evidence points out we may find some portfolios, which have statistically higher return and lower risk in a Markov-switching regression. The results demonstrate PIN is an important factor in explaining portfolio returns.

Keywords

Probability of informed negotiation. Markov-switching models. Portfolios selection. Tick-test algorithm.

JEL Classification

G11. C58.

1. Introdução

Diversos trabalhos, na área de microestrutura do mercado, abordam o comportamento dos *traders* informados e não informados. O estudo seminal de Bagehot (1971), bem como as elaborações de Kyle (1985) e Easley e O'Hara (1987), apontam que o *trader* informado obtém vantagem em relação aos *traders* não informados. Easley *et al.* (1996) desenvolvem um modelo empiricamente testável em que se constrói uma medida de probabilidade de negociação informada do mercado, denominada PIN, que tem possibilitado melhor avaliação empírica da teoria de microestrutura de mercado e recentemente tem sido utilizado na literatura de apreçamento de ativos.

Easley, Hvidkjaer, O'Hara (2002) investigam como a negociação baseada em informação afeta os retornos dos ativos ao incorporar o PIN ao modelo de Fama e French (1992). Easley e O'Hara (2004) argumentam que ações com maior informação apresentam maior retorno esperado. Evidência contrária a Easley e O'Hara (2004) é encontrada em Duarte e Young (2009), ao investigarem o problema de assimetria de informação e o retorno *cross-section* das ações da *New York Stock Exchange* (NYSE) durante o período de 1983 a 2004, e, utilizando o modelo de Fama-Macbeth (1973), concluem que a negociação baseada na informação não afeta os retornos.

Já Copeland, Wong e Zeng (2009), empregando método de Fama e French (1992), mostram que a probabilidade de negociação informada desempenha um papel importante na explicação dos retornos mensais das ações negociadas em Shanghai durante o período de 2001 a 2006.

Easley, Hvidkjaer, O'Hara (2010), utilizando ações preferenciais da NYSE e *American Stock Exchange* (AMEX) durante os anos de 1983 a 2001, demonstram que *portfolios* que utilizam o PIN geram retornos anormais significativos que não são explicados pelos fatores de Fama e French (1993). Considerando os trabalhos de Easley *et al.* (1996) e Easley, Hvidkjaer, O'Hara (2010), este artigo estima o PIN para ações que participam do IBRX durante o período de 2006 até 2011, e utiliza esta variável para selecionar *portfolios* baseados no método de Fama e French (1993), incorporando ao método um prêmio por informação. Dessa maneira, pretende-se identificar se, no Brasil, a composição de *portfolios* que levam em consideração a

variável PIN apresenta maiores retornos. A estimação dos fatores de Fama e French (1993) e do prêmio por informação é feita empregando modelos de *markov-switching*. Dessa forma, é possível verificar se os parâmetros estimados são significativos em regimes de alta e de baixa volatilidade.

No Brasil, a pesquisa em microestrutura do mercado é recente. Silva (2009), Barbedo, Silva e Leal (2009) consideram sua aplicação na obtenção do PIN como um dos primeiros trabalhos realmente inseridos na área de microestrutura no Brasil.

Este artigo contribui para aproximar a literatura de microestrutura de mercado com a de apreçamento ao incorporar o prêmio por informação ao método de Fama e French (1993). Além disso, desenvolve uma aplicação empírica no mercado brasileiro, visando a identificar se o PIN é importante na explicação dos retornos dos *portfolios*.

Para isso, este artigo está estruturado em seis partes. A primeira delas é propriamente esta introdução. A segunda apresenta o referencial teórico abordando o trabalho de Easley *et al.* (1996), o algoritmo do *tick-test*, e o modelo de Fama e French (1993). A terceira, os dados e seus tratamentos. A quarta parte explica a metodologia econométrica, enquanto a quinta apresenta os resultados. Por fim, a sexta seção exhibe as considerações finais.

2. Referencial teórico

2.1. Modelo de Easley, Kiefer, O'Hara e Paperman (1996)

Easley, Kiefer, O'Hara, e Paperman (1996) desenvolvem um modelo de negociação sequencial no qual é possível estimar a probabilidade de negociação informada em um mercado. Admitindo um *market maker* competitivo que negocia ações com *traders* informados e não informados, as negociações ocorrem durante os dias de $i = 1, \dots, I$ e, durante a cada dia, o tempo é contínuo.

Entre os dias de negociação, ocorre um evento informacional sobre o valor de um ativo com probabilidade α , e $(1-\alpha)$ é a probabilidade do evento informacional não ocorrer. Admitindo que ocorra o even-

to informacional, existe uma probabilidade δ de que o evento traga boas notícias, e de que o valor do ativo na data i seja \overline{V}_i . Existe uma probabilidade $1 - \delta$ de que o evento traga más notícias e, conseqüentemente, de que o valor do ativo na data i seja \underline{V}_i .

Durante o dia de negociação, os *traders* informados e não informados chegam ao mercado de acordo com processos independentes de Poisson. O *market maker* determina os preços de compra e de venda a cada tempo t , com $t \in [0, T]$ durante o dia de negociação. As ordens dos *traders* informados chegam a uma taxa μ nos dias em que ocorre um evento informacional. Os *traders* informados compram se têm boas notícias e vendem se recebem más notícias. As ordens dos *traders* não informados chegam a uma taxa ε . A Figura 1 demonstra esse processo de negociação.

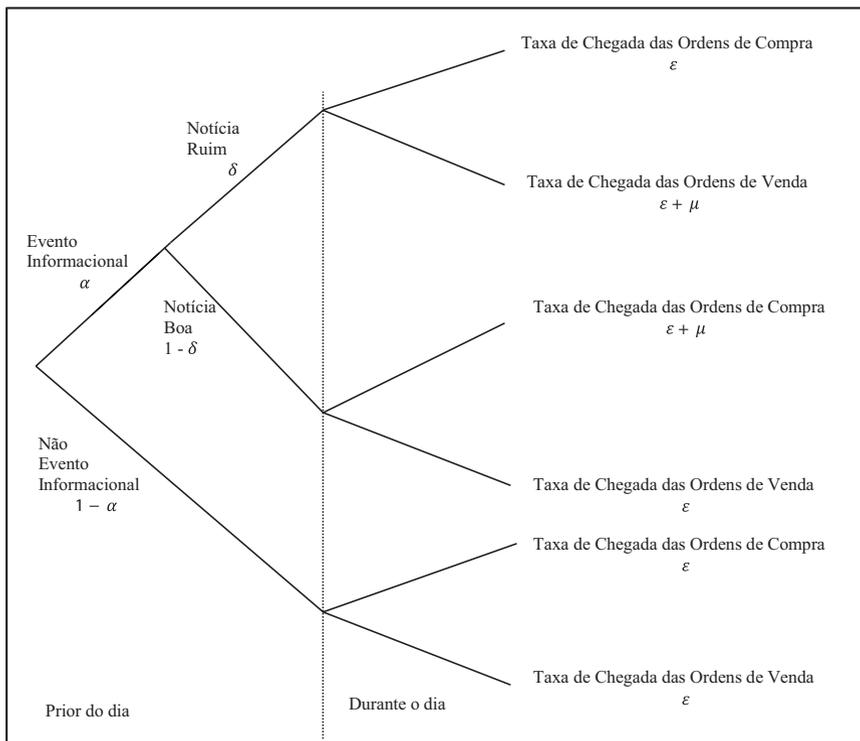


Figura 1 – Processo de negociação.

Fonte: Easley, Kiefer, O'Hara e Paperman (1996)

Pela estrutura do modelo, dado que um evento informacional ocorre, é esperado um número maior de ordens de compra em dias de boas notícias e um número maior de ordens de vendas em dias de más notícias. Em dias sem evento informacional, é esperado um número menor de operações devido à ausência de *traders* informados. Os parâmetros de $(\alpha, \delta, \varepsilon, \mu)$ não são diretamente observáveis e são estimados por uma função de verossimilhança.

Como as taxas de chegada das ordens seguem a distribuição de Poisson, o modelo supõe que as ordens de compra e venda seguem três processos independentes, refletindo os eventos de boa notícia, má notícia ou a não ocorrência de evento informacional. Considerando um dia de má notícia, as ordens de venda chegam a uma taxa $(\varepsilon + \mu)$, e as de compra, a uma taxa ε . A probabilidade, de uma sequência que contém B ordens de compra e S ordens de venda, no final do tempo de negociação T , em um dia de más notícias, é dada por:

$$e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-(\mu + \varepsilon)T} \frac{[(\mu + \varepsilon)T]^S}{S!} \quad (1)$$

No caso de o dia ser de boas notícias, a probabilidade é representada por:

$$e^{-(\mu + \varepsilon)T} \frac{[(\mu + \varepsilon)T]^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \quad (2)$$

Enquanto que, em um dia sem evento informacional, a probabilidade é:

$$e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \quad (3)$$

Assim, a função de verossimilhança de observar B e S em um dia qualquer é a média ponderada, pelas probabilidades dos eventos, das Equações (1), (2) e (3) e é dada por:

$$\mathcal{L}((B_t, S_t) | \theta) = (1 - \alpha) e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} + \alpha \delta e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-(\mu + \varepsilon)T} \frac{[(\mu + \varepsilon)T]^S}{S!} + \alpha(1 - \delta) e^{-(\mu + \varepsilon)T} \frac{[(\mu + \varepsilon)T]^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \quad (4)$$

Na Equação, (B, S) representam o número total de compras e vendas para cada dia, e $\theta=(\mu, \varepsilon, \alpha, \delta)$ é o vetor de parâmetros. Para um determinado dia, os parâmetros α e δ assumem valores 0 ou 1, refletindo a ocorrência ou não do evento. Considerando os dias como independentes, a função de verossimilhança para os I dias é

$$\mathcal{L}((B, S)|\theta) = \prod_{i=1}^I \left\{ \begin{aligned} &(1 - \alpha) \left[e^{-\varepsilon T \frac{B_i}{B_i!}} e^{-\varepsilon T \frac{S_i}{S_i!}} \right] + \\ &\alpha \delta \left[e^{-\varepsilon T \frac{B_i}{B_i!}} e^{-(\mu + \varepsilon) T \frac{[(\mu + \varepsilon) T] S_i}{S_i!}} \right] + \\ &\alpha(1 - \delta) \left[e^{-(\mu + \varepsilon) T \frac{[(\mu + \varepsilon) T] B_i}{B_i!}} e^{-\varepsilon T \frac{S_i}{S_i!}} \right] \end{aligned} \right\} \quad (5)$$

Estimando o vetor de parâmetros $\theta=(\mu, \varepsilon, \alpha, \delta)$, é possível utilizá-lo para calcular o PIN, que é dado por:

$$PIN = \frac{\alpha \mu}{\alpha \mu + 2\varepsilon} \quad (6)$$

A probabilidade de negociação informada é calculada pelos parâmetros estimados e mede a fração de ordens que chegam de *traders* informados, $\alpha\mu$, em relação ao total de ordens informadas e não informadas, $\alpha\mu + 2\varepsilon$.

Easley, Kiefer e O’Hara (1997) e Easley, Hvidkjaer e O’Hara (2002) apresentam uma transformação logarítmica da Equação (5), classificando a taxa de chegada das ordens dos *traders* não informados, ε , em taxa de chegadas das ordens de compra, ε_b , e venda, ε_s , dos *traders* não informados.

$$\mathcal{L}((B_t, S_t)_{t=1}^T|\theta) = \sum_{t=1}^T [-\varepsilon_b - \varepsilon_s + M_t(\ln x_b + \ln x_s) + B_t \ln(\mu + \varepsilon_b) + S_t \ln(\mu + \varepsilon_s)] + \sum_{t=1}^T \ln[\alpha(1 - \delta) e^{-\mu x_s^{s_t - M_t}} x_b^{-M_t} + \alpha \delta e^{-\mu x_b^{B_t - M_t}} x_s^{-M_t} + (1 - \alpha) x_s^{s_t - M_t} x_b^{B_t - M_t}] \quad (7)$$

$$\text{Sendo: } M_t = \min(B_t, S_t) + \max\left(\frac{B_t - S_t}{2}\right);$$

$$x_S = \varepsilon_S / (\mu + \varepsilon_S)$$

$$x_b = \varepsilon_b / (\mu + \varepsilon_b)$$

Estimados os parâmetros da Função (7), o cálculo do PIN é dado por:

$$PIN = \frac{\alpha \mu}{\alpha \mu + \varepsilon_S + \varepsilon_b} \quad (8)$$

A vantagem de empregar a Função (7) segundo Duarte e Young (2009) e Yan (2009), é que a transformação logarítmica facilita o processo de otimização numérica, diminuindo os problemas de *overflow* e *underflow*.¹

2.1.1. Algoritmo do tick-test

Para o desenvolvimento do PIN, é necessário classificar as ordens executadas em ordens de compra ou ordens de venda. Lee-Ready (1991) apresenta, além do seu algoritmo, o algoritmo do *tick-test*.

No método do *tick-test*, a classificação é feita considerando o preço da última ordem executada. Portanto, a primeira ordem do dia não pode ser classificada. Até haver uma mudança no preço das ordens subsequentes, não é possível classificá-las. Se o preço da ordem seguinte for superior, então a ordem é classificada como *uptick* e, portanto, é considerada uma ordem de compra; se o preço for inferior, a ordem é classificada com *downtick* e é considerada como venda. Dado um *uptick*, se a próxima ordem for executada ao mesmo preço, então a ordem é classificada como *zerouptick* e é considerada como ordem de compra. Da mesma maneira, dado um *downtick*, se a próxima ordem for executada ao mesmo preço, a ordem é classificada como *zerodowntick* e, portanto, é considerada como ordem venda.

¹ O sistema de pontos flutuantes utilizado no cálculo numérico computacional pode originar problemas de *overflow* e *underflow*. Normalmente, problemas de *underflow* estão associados a uma divisão por zero, enquanto problemas de *overflow* a um número infinito. Em ambos os casos, o computador não consegue armazenar os números gerados no cálculo numérico.

2.2. Modelo de Fama e French (1993)

O modelo de Fama e French (1993) consiste de uma equação de regressão linear múltipla para os retornos das ações, em que as variáveis explicativas são o prêmio de risco de mercado, o valor de mercado (VM) da empresa e a razão entre o valor patrimonial (VP) e o valor de mercado (índice VP/VM). Justifica-se pela teoria de apreçamento por arbitragem (APT) de Ross (1976 e 1978) e nos achados empíricos de Basu (1977), Banz (1981) e Rosember, Reid e Lantein (1985). Com efeito, a escolha das variáveis explicativas se reduz à adição à equação do modelo de apreçamento de ativos de capital (CAPM) de Sharpe (1964), Lintner (1965) e Mossin (1966) das variáveis valor de mercado e do índice VP/VM em vista dos testes realizados em Fama e French (1992), que evidenciam a existência de um prêmio das empresas com menor valor de mercado (*Small*) sobre as empresas com maiores valores de mercado (*Big*), assim como das empresas com alto índice VP/VM (*High*) sobre as empresas com baixo índice VP/VM (*Low*).

Com essa nova especificação, Fama e French adaptaram o método de teste por construção de carteiras de Black, Jensen e Scholes (1972), usando como critério a diferença entre o valor de mercado entre a maior e a menor empresa na amostra – variável explicativa *Small Minus Big* (SMB) - e na diferença entre o maior e o menor índice VP/VM – variável *High Minus Low* (HML). Efetuaram a estimação e os testes do modelo com uma amostra dos retornos mensais de todas as empresas listadas na NYSE, AMEX e NASDAQ entre os anos de 1963 e 1991. Tal procedimento resultou, como esperado, em superioridade em termos de poder explicativo R^2 em relação ao CAPM, com os coeficientes dos fatores SMB e HML estatisticamente significativos na grande maioria das carteiras. Formalmente, o modelo testado é apresentado na Equação (9) abaixo:

$$R_{i,t} - R_{f,t} = a_i + b_i(R_{m,t} - R_{f,t}) + s_i[SMB_t] + h_i[HML_t] + \varepsilon_{i,t} \quad (9)$$

Onde:

$R_{i,t}$ = retorno da carteira i no mês t;

$R_{m,t}$ = retorno da carteira de mercado no mês t;

$R_{f,t}$ = retorno do ativo livre de risco no mês t;

SMB_t = prêmio pelo fator tamanho no mês t ;

HML_t = prêmio pelo fator B/M no mês t ;

$\varepsilon_{i,t}$ = resíduo do modelo referente à carteira i no mês t .

3. Dados

O primeiro passo foi identificar todas as ações que participaram do IBRX-100² durante o período janeiro de 2006 até abril de 2011. De posse dessas informações e do histórico de negócios intradiários da BMF&BOVESPA, aplicou-se o algoritmo do *tick-test*, descrito na seção anterior, para classificar as ordens executadas como ordens de compra (*buy*) ou ordens de venda (*sell*). Obtendo-se o total diário de *buy* e *sell* para cada ativo, foi possível estimar o PIN para cada trimestre.

As informações sobre o valor de mercado das empresas, a relação entre o valor do patrimônio líquido e valor de mercado (*book-to-market*), evolução do Índice Bovespa (IBOVESPA), o preço diário das ações e as taxas diárias dos Certificados de Depósito Interbancário (CDI) foram obtidas junto à BMF&BOVESPA, Comissão de Valores Mobiliários (CVM), Economatica e Cetip. Por meio dessas informações, tornou-se possível compor as carteiras de forma similar àquela feita por Fama e French (1993) e Easley, Hvidkjaer e O'Hara (2010). Para isso, dois conjuntos de *portfolios* foram criados: (X) e (Y). O conjunto de *portfolios* (X) leva em consideração as variáveis valor de mercado e *book-to-market*, enquanto que o conjunto de *portfolios* (Y) incorpora a variável PIN. É importante destacar que as ações que participaram no conjunto (Y) foram as mesmas do conjunto (X). Tanto no conjunto (X) quanto no (Y) ocorreu um ajuste trimestral dos *portfolios*, ou seja, a composição das carteiras no trimestre t foi feita com as informações do trimestre anterior $(t-1)$ ³.

² O IBRX-100 foi preferido ao IBOVESPA por apresentar um número maior de ações para compor os *portfolios*.

³ É importante destacar que as informações de Patrimônio Líquido necessárias para o índice *book-to-market* e para composição das carteiras no tempo t são consideradas conhecidas no final do trimestre $t-1$. Isso é válido tanto no conjunto (X) como no conjunto (Y).

Os conjuntos de *portfolios* (X) e (Y) foram criados pela interseção das variáveis. Presume-se que cada variável possa ter valores alto ou baixo, e sua separação é feita pela mediana. Assim, para o conjunto (X), criaram-se quatro *portfolios*. São eles:

- a) BH: Alto valor de mercado e alto índice *book-to-market*;
- b) BL: Alto valor de mercado e baixo índice *book-to-market*;
- c) SH: Baixo valor de mercado e alto índice *book-to market*; e
- d) SL: Baixo valor de mercado e baixo índice *book-to-market*.

No conjunto (Y), por considerar a variável PIN, formaram-se as seguintes carteiras:

- a) BHG: Alto VM, Alto B/M e Alto PIN.
- b) BHP: Alto VM, Alto B/M e Baixo PIN.
- c) BLG: Alto VM, Baixo B/M, Alto PIN.
- d) BLP: Alto VM, Baixo B/M, Baixo PIN.
- e) SHG: Baixo VM, Alto B/M, Alto PIN.
- f) SHP: Baixo VM, Alto B/M, Baixo PIN.
- g) SLG: Baixo VM, Baixo B/M, e Alto PIN.
- h) SLP: Baixo VM, Baixo B/M e Baixo PIN.

Após a composição, foi calculado diariamente o retorno das carteiras, o prêmio pelo fator de risco de mercado, o prêmio pelo fator de tamanho, o prêmio pelo fator *book-to-market* e o prêmio pelo fator de informação (apenas para o conjunto Y).

O retorno de cada carteira é dado por:

$$R_i = \sum_{j=1}^m \frac{VM_j}{VM_p} R_{j,i} \quad (10)$$

Sendo: R_i o retorno do *portfolio* no dia i ;

VM_j o valor de mercado da ação j ;

VM_p o valor de mercado do *portfolio*;

$R_{j,i}$ retorno da ação j no dia i .

O prêmio pelo fator de tamanho [SMB] foi calculado diariamente da seguinte forma:

$$[SMB_i] = \overline{R_{S,i}} - \overline{R_{B,i}} \quad (11)$$

Sendo:

SMB_i o prêmio pelo fator de tamanho no dia i ;

$\overline{R_{S,i}}$ o retorno médio das carteiras com baixo valor de mercado no dia i ; e

$\overline{R_{B,i}}$ o retorno médio das carteiras com alto valor de mercado no dia i .

O prêmio por fator de risco *book-to-market* é obtido por:

$$[HML_i] = \overline{R_{H,i}} - \overline{R_{L,i}} \quad (12)$$

Sendo:

HML_i o prêmio pelo fator *book-to-market* no dia i ;

$\overline{R_{H,i}}$ o retorno médio das carteiras com alto *book-to-market* no dia i ; e

$\overline{R_{L,i}}$ o retorno médio das carteiras com baixo *book-to-market* no dia i .

O prêmio por fator de informação é calculado para cada dia como sendo:

$$[GMP_i] = \overline{R_{G,i}} - \overline{R_{P,i}} \quad (13)$$

Na fórmula:

GMP_i é o prêmio por fator de informação no dia i ;

$\overline{R_{G,i}}$ o retorno médio das carteiras com alto *PIN* no dia i ; e

$\overline{R_{P,i}}$ o retorno médio das carteiras com baixo *PIN* no dia i .

Tendo calculado os prêmios, foi possível estimar as seguintes equações para cada *portfolio* dos conjuntos (X) e (Y) respectivamente:

$$R_i(t) - RF(t) = a(s_t) + b(s_t)[RM(t) - RF(t)] + s(s_t)[SMB(t)] + h(s_t)[HML(t)] + e(t) \quad (14)$$

$$R_i(t) - RF(t) = a(s_t) + b(s_t)[RM(t) - RF(t)] + s(s_t)[SMB(t)] + h(s_t)[HML(t)] + p(s_t)[GMP(t)] + e(t) \quad (15)$$

Sendo:

$RF(t)$ a renda fixa, no caso o CDI diário; e

$RM(t)$ o retorno do mercado, no caso, o IBOVESPA.

A estimação foi feita utilizando modelos de *markov-switching* para os quatro *portfolios* do conjunto (X) e para os oito *portfolios* do conjunto (Y). Um teste empírico do método tradicional de Fama-French (1993), utilizando modelos de *markov-switching*, encontra-se em Coggi e Manescu (2004).

A estimação em dois regimes distintos (alta e baixa volatilidade) permite observar se os parâmetros do modelo são significativos quando ocorrem mudanças de regime, principalmente o prêmio por informação (GMP). A significância dos parâmetros da variável prêmio por informação, em regimes de baixa e alta volatilidade, corroboram os modelos teóricos de microestrutura do mercado, que apontam que o *trader* informado obtém vantagem em relação ao *trader* não informado, além de possibilitar testes empíricos sobre a relevância dos fatores utilizados.

4. Modelos de Markov-Switching

Os modelos com mudanças de regime são propostos por Hamilton (1989) ao analisar a taxa de crescimento do *Gross National Product* (GNP) nos Estados Unidos. Os modelos de *markov-switching vector autoregression* (MS-VAR) podem ser considerados como uma generalização de um modelo VAR de ordem p .

Supondo que exista N regimes possíveis e que s_t representa um regime que prevalece no tempo t , tal que $s_t \in \{1, 2, \dots, N\}$, a probabilidade condicionada ao vetor de séries de tempo observado, y_t , é dada por:

$$P(y_t | \psi_{t-1}, s_t) = \begin{cases} f(y_t | \psi_{t-1}, \theta_1) & \text{se } s_t = 1 \\ \vdots \\ f(y_t | \psi_{t-1}, \theta_N) & \text{se } s_t = N \end{cases} \quad (16)$$

Na equação, θ_n é o vetor de parâmetros do VAR no regime $n = 1, 2, \dots, N$, e ψ_{t-1} as observações da série de tempo $\{y_{t-j}\}_{j=1}^{\infty}$. Considerando (16) um modelo MS-VAR ajustado à média é expresso por:

$$y_t = \mu(s_t) + \sum_{j=1}^p A_j(s_t) [y_{t-j} - \mu(s_{t-j})] + u_t \quad (17)$$

Sendo: $u_t \sim IND(0, \Sigma(s_t))$; e $\mu(s_t) = (I_k - \sum_{j=1}^p A_j(s_t))^{-1} v(s_t)$ o vetor de médias

Essa é uma especificação geral do modelo. Os modelos MS-VAR, de acordo com Krolzig (1997), podem ter variações. Essas variações surgem pela mudança na média condicional do processo (M), no intercepto (I), nos parâmetros autorregressivos (A), na variância (H) ou pela combinação desses parâmetros. Dessa forma, tem-se:

- a) M = *Markov switching Mean*, média dependente da variável não observável s_t ;
- b) I = *Markov switching Intercept*, intercepto dependente da variável não observável s_t ;
- c) A = *Markov switching Autoregressive*, parâmetros autorregressivos, dependentes da variável não observável s_t ;
- d) H = *Markov switching Heteroskedasticity*, matriz de variância e covariância dependente da variável não observável s_t ;

Considerando que os parâmetros da Equação (17) dependem do regime s_t , que, por sua vez, presume-se ser estocástico e não observado, faz-se necessário um processo gerador dos regimes. Considera-se que, nos modelos MS-VAR, regimes são gerados por uma cadeia de Markov ergódica, irreduzível, com tempo e estados discretos e probabilidade de transição dada por:

$$p_{ij} = Pr(s_{t+1} = j | s_t = i), \sum_{j=1}^N p_{ij} = 1 \quad \forall i, j \in \{1, \dots, N\} \quad (18)$$

Representando as probabilidades de transição na forma matricial, tem-se a matriz de probabilidade de transição dada por:

$$P = \begin{bmatrix} p_{11} & p_{21} & \dots & p_{N1} \\ p_{12} & p_{22} & \dots & p_{N2} \\ \dots & \dots & \dots & \dots \\ p_{1N} & p_{2N} & \dots & p_{NN} \end{bmatrix} \quad (19)$$

Para o processo de estimação de modelos MS-VAR, é considerada a matriz de séries de tempo $\psi_{t-1} = (y'_{t-1}, y'_{t-2}, \dots, y'_0, \dots, y'_{1-p})'$, sendo empregadas variáveis endógenas. Definindo ξ_t como um determinado regime, dado que u_t em (17) é distribuído normalmente e que $s_t = j$ na data t , então a densidade condicional de y_t é expressa por:

$$f(y_t | \xi_t = l_j, \psi_t; \lambda) = \ln(2\pi)^{-\frac{1}{2}} \ln|\Sigma|^{-\frac{1}{2}} \exp\{(y_t - \bar{y}_{jt})\} \quad (20)$$

Sendo l_j a j -ésima coluna da matriz identidade I_N , $\bar{y}_{jt} = E[y_t | \xi_t, \psi_{t-1}]$ é a esperança condicionada de y_t , dado que o processo está em j e que λ é um vetor que contém os parâmetros da autorregressão, θ , e as probabilidades de transição que determinam a cadeia de Markov dos estados não observados. O vetor ξ_t recebe a informação a respeito dos estados da cadeia de Markov, assumindo valores *booleanos* a partir de uma função indicadora, sendo expresso por:

$$\xi_t = \begin{bmatrix} I(s_t = 1) \\ \dots \\ I(s_t = N) \end{bmatrix} \quad I(s_t = m) = \begin{cases} 1, & \text{se } s_t = N \\ 0, & \text{caso contrário} \end{cases} \quad (21)$$

As densidades condicionadas, para os N possíveis regimes, são definidas por:

$$\eta_t = \begin{bmatrix} f(y_t | \xi_t = l_j, \psi_{t-1}) \\ \dots \\ f(y_t | \xi_t = l_n, \psi_{t-1}) \end{bmatrix} \quad (22)$$

A probabilidade de densidade marginal y_t é obtida em duas etapas. Primeiramente, a densidade conjunta de y_t e ξ_t é escrita como produto das densidades marginal e condicional. No segundo passo, y_t é integrada em relação a todos os regimes. O resultado de y_t é interpretado como uma média ponderada das densidades condicionais, na qual os fatores de peso são as probabilidades dos regimes. Portanto, é necessário fazer alguma inferência sobre o regime não observado, e isso ocorre pelo filtro e suavizador BLHK (Baum-Lindgren-Hamilton-Kim), que possibilita fazer inferência sobre o estado do processo por meio das probabilidades filtradas e suavizadas, sendo:

$\hat{\xi}_{t|\tau}, \tau \leq t$ probabilidade predita do regime;

$\hat{\xi}_{t|\tau}, \tau = t$ probabilidade filtrada do regime; e

$\hat{\xi}_{t|\tau}, t \leq \tau \leq T$ probabilidade suavizada do regime.

A inferência ótima e a previsão para cada data t na amostra pode ser encontrada pela iteração do seguinte par de equações:

$$\hat{\xi}_t = \frac{(\hat{\xi}_{t|t-1} \otimes \eta_t)}{1'(\hat{\xi}_{t|t-1} \otimes \eta_t)} \quad (23)$$

$$\hat{\xi}_t = P^* \hat{\xi}_{t|t} \quad (24)$$

Nessa equação, η_t é o vetor com as densidades condicionais, P é a matriz de probabilidade de transição e 1 , o vetor de dimensão $(N \times 1)$, enquanto que o símbolo \otimes denota a multiplicação de elemento por elemento. Então, dado um valor inicial $\hat{\xi}_{1|0}$ e um valor para o vetor de parâmetros da população, λ , iterege-se as Equações (23) e (24) para $t = 1, 2, \dots, T$, calculando-se os valores $\hat{\xi}_{t|t}$ e $\hat{\xi}_{t+1|t}$ para cada data t da amostra.

5. Resultados

Com base nos dados divulgados pela BMF&Bovespa, foi possível identificar 203 ações que, em algum momento, participaram do IBRX-100 durante o período de janeiro de 2006 até abril de 2011. Utilizando o histórico de registro de negócios da BMF&BOVESPA, observou-se que, dessas 203 ações, apenas 73 apresentaram negócios durante todos os 1317 dias de negociação. Dessas 73 ações, só foi possível coletar os dados das variáveis valor de mercado e *book-to-market* para 68. Dessas 68 ações, foi possível estimar o PIN trimestral para apenas 55 ações.

Ao selecionar as ações que sempre estiveram listadas no IBRX-100 e que tiveram negócios durante todos os dias (73 ações) assume-se um viés que apenas as ações que apresentaram um maior Índice de Negociabilidade⁴ participam do trabalho. Durante o processo de estimação dos parâmetros que permitem o cálculo do PIN teve-se uma perda de 13 ações por causa de problemas de otimização. Assim o viés identificado nas empresas sobreviventes está associado ao Índice de Negociabilidade, quanto a um possível viés gerado pela estimação do PIN isto será considerado em pesquisas futuras.

5.1. Estimativa do PIN

A estimativa do PIN foi feita utilizando o histórico de negócios da BMF&BOVESPA. Esses dados foram disponibilizados em arquivos de formato texto. Os arquivos apresentam as seguintes informações: data da sessão, papel, número de negócio, preço do negócio, quantidade, hora com precisão tandem,⁵ indicador de anulação,⁶ data da oferta de compra, sequência da oferta de compra, data da oferta de venda e sequência da oferta de venda. Considerando os 74 meses e

⁴ O IBRX-100 considera em sua metodologia o Índice de Negociabilidade (IN) dado por: $IN = \sqrt{(n_i/N \times (v_i/V))}$, sendo IN = índice de negociabilidade; n_i = número de negócios com a ação "i" no mercado à vista (lote-padrão); N = número total de negócios no mercado à vista da BOVESPA (lote-padrão); v_i = volume financeiro gerado pelos negócios com a ação "i" no mercado à vista (lote-padrão); e V = volume financeiro total do mercado à vista da BOVESPA (lote-padrão).

⁵ Na precisão tandem o tempo é expresso por: HH:MM:SS.SSSSSS, onde HH representa a hora, MM o minuto, SS o segundo e SSSSSS o milionésimo de segundo. Sendo assim, o registro de cada negócio ocorrido leva em consideração até o milionésimo de segundo.

⁶ O indicador de anulação informa se um registro de negócio foi anulado.

que, em média, cada arquivo ocupa 540 *megabytes* de informação, tem-se, no total, 40 *gigabytes*. Em média, cada arquivo representa uma matriz de 5.000.000 linhas por 12 colunas.

Utilizando linguagem de programação em *R*, foi extraído⁷ para cada ação, durante todos os dias, as informações necessárias para aplicar o algoritmo do *tick-test*. Empregando o *tick-test*, identificaram-se as ordens que não eram classificadas, e as demais foram classificadas como *uptick*, *zero-uptick*, *down-tick* e *zero-downtick*. As *up-tick*, *zero-uptick* são consideradas ordens de compra, e as *down-tick* e *zero-downtick*, ordens de venda. Com o número de *buy* e *sell* diário é estimado, por máxima verossimilhança, os seguintes parâmetros: α , probabilidade de um evento informacional ocorrer; δ , probabilidade de um evento informacional ser de más notícias; ε_s , taxa de chegada das ordens de venda para os *traders* não informados; ε_b , taxa de chegada das ordens de compra para os *traders* não informados; e μ , taxa de chegada das ordens dos *traders* informados. O PIN foi calculado trimestralmente para cada ação, utilizando os parâmetros estimados. Foram 21 estimações por ação e 1.155 estimações no total. Na Tabela 1, por exemplo, são apresentados os parâmetros estimados, bem como o PIN para cada ação durante o primeiro trimestre de 2006. Os valores obtidos ficam compreendidos entre 0 e 1, pois o PIN é uma medida de probabilidade. Um PIN com valor 0 indica ausência de negociação com informação privilegiada, enquanto que um PIN com valor 1 indica que todas as negociações ocorreram com informação privilegiada.

⁷ Rotinas em *R* foram desenvolvidas para obtenção do PIN, composição das carteiras, cálculo dos retornos e dos prêmios por informação.

Tabela 1 – Parâmetros estimados e PIN para o primeiro trimestre de 2006.

Ação	α	δ	ε_b	ε_s	μ	PIN	Ação	α	δ	ε_b	ε_s	μ	PIN
AMBV4	0,4343	0,4664	163,10	174,79	130,83	0,1439	ITSA4	0,4169	0,3749	203,73	206,61	157,94	0,1383
BBAS3	0,2903	0,5493	196,23	196,00	212,11	0,1357	KLBN4	0,1936	0,0834	129,82	135,42	195,63	0,1250
BBDC4	0,3821	0,4589	706,17	695,03	468,90	0,1134	LAME4	0,4157	0,2035	86,99	92,74	72,84	0,1442
BRAP4	0,2675	0,5815	215,11	199,71	168,53	0,0980	LREN3	0,4182	0,6686	50,21	52,93	61,69	0,2001
BRKM5	0,5076	0,7835	536,09	468,93	399,06	0,1677	NATU3	0,4194	0,9231	122,52	84,91	108,40	0,1798
BRTO4	0,4549	0,6229	293,44	274,90	218,63	0,1489	OHLB3	0,4029	0,7606	51,17	40,61	68,50	0,2312
CCRO3	0,3710	0,8696	273,00	201,31	356,74	0,2182	PETR3	0,3205	0,6229	313,29	310,12	189,57	0,0888
CGAS5	0,4572	0,5339	78,68	75,36	72,09	0,1762	PETR4	0,2665	0,8938	2160,10	2006,13	1296,78	0,0766
CMIG4	0,3340	0,1732	291,99	337,25	260,84	0,1216	POMO4	0,3381	0,6183	19,71	20,58	41,27	0,2573
CPFE3	0,4041	0,3634	83,97	88,66	87,16	0,1695	PSSA3	0,3857	0,2088	22,09	31,68	60,39	0,3023
CNFB4	0,4965	0,6770	48,90	67,44	94,68	0,2878	RAPT4	0,2578	0,4377	43,99	42,95	73,78	0,1795
CPLE6	0,3034	0,8519	304,26	273,52	232,47	0,1088	RENT3	0,4412	0,5846	33,66	32,69	42,05	0,2185
CRUZ3	0,3876	0,2499	114,12	120,21	108,58	0,1523	RSID3	0,0968	0,1667	102,88	127,95	584,35	0,1968
CSNA3	0,3536	0,3098	470,56	492,06	396,01	0,1270	SBSP3	0,4164	0,9217	166,43	111,13	142,65	0,1763
CYRE3	0,4926	0,3783	60,61	66,61	59,39	0,1870	SUZB5	0,6064	0,1400	35,78	31,19	43,50	0,2826
ELET5	0,3115	0,5561	257,47	232,54	203,93	0,1148	TAMM4	0,1935	0,5002	91,46	87,63	208,64	0,1840
ELET6	0,2929	0,6086	376,90	339,03	258,92	0,0958	TBLE3	0,5420	0,0613	72,47	110,85	124,56	0,2691
EMBR3	0,2929	0,6086	376,90	339,03	258,92	0,0958	TCSL3	0,2010	0,3317	93,46	90,68	126,23	0,1211
EMBR3	0,1774	0,1818	93,94	120,97	163,24	0,1188	TCSL4	0,1596	0,6128	331,22	296,85	444,31	0,1014
ENBR3	0,3062	0,1053	57,69	78,69	107,45	0,1944	TMAR5	0,4177	0,4951	124,59	132,48	117,83	0,1607
FFTL4	0,4514	0,5031	19,00	27,54	26,89	0,2069	TNLP3	0,4442	0,6665	126,65	117,97	72,50	0,1163
GETI3	0,3703	0,3051	8,85	11,18	25,22	0,3180	TNLP4	0,4266	0,4411	518,50	546,28	232,77	0,0853
GETI4	0,6910	0,7371	22,37	24,20	24,01	0,2627	TRPL4	0,4285	0,4536	103,45	113,34	107,67	0,1755
GGBR3	0,3542	0,4544	46,08	43,29	49,23	0,1632	UGPA4	0,3249	0,5535	39,23	30,49	71,89	0,2509
GGBR4	0,3155	0,0585	430,04	489,03	344,36	0,1057	USIM5	0,4847	0,3701	565,26	586,27	373,21	0,1358
GOA4	0,3696	0,3704	202,48	197,89	142,51	0,1163	VALE3	0,4427	0,9272	303,28	232,16	184,66	0,1325
GOLL4	0,4947	0,7985	129,19	90,64	154,46	0,2579	VALE5	0,2415	0,3837	1148,14	1167,37	929,97	0,0884
IDNT3	0,1136	0,5679	46,00	47,31	183,32	0,1825							

Fonte: elaboração própria

Observam-se, na Tabela 1, as estimativas dos parâmetros (α , δ , ε_b , ε_s , μ) para todas as 55 ações no primeiro trimestre. O cálculo do PIN segue a Equação (8) e é apresentado na sétima coluna. Por exemplo, para a ação AMBV4, percebe-se que, no trimestre, a probabilidade de ocorrer um evento informacional é de aproximadamente 43,43%.

Dada a ocorrência do evento informacional, existe uma probabilidade de 46,64% de que o evento traga más notícias. A taxa média de chegada de ordens de compra (ε_b), dos *traders* não informados é de 163,10, e a taxa média de chegada das ordens de venda (ε_s), para os *traders* não informados é de 174,79. Já a taxa média de chegada de ordens de *traders* informados é de 130,82. Com base nessas informações e na Equação (8), obtém-se, para AMBV4, no primeiro trimestre de 2006, um PIN de 14,39%. Nesse trimestre, o maior PIN obtido foi da ação GETI3, ficando em 31,79%; já o menor PIN

foi o da ação PETR4, ficando em 7,66%. No Gráfico 1, é apresentado o comportamento das estatísticas descritivas do PIN para cada trimestre e todas as ações.

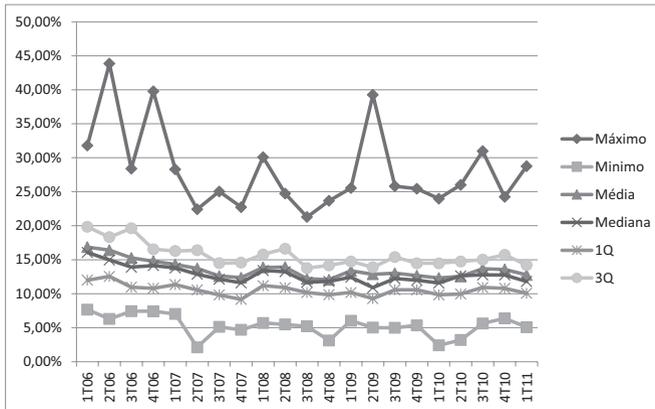


Gráfico 1 – Estatísticas descritivas do PIN por trimestre para todas as ações

Fonte: elaboração própria.

No Gráfico 1, é observado a evolução das estatísticas descritivas (média, mediana, máximo, mínimo, primeiro quartil (1Q) e terceiro quartil (3Q)) do PIN por trimestre. O PIN médio durante o período ficou entre 16,86% (primeiro trimestre de 2011) e 12,06% (último trimestre de 2008). Percebe-se também que a mediana e o primeiro e terceiro quartil tiveram a tendência de acompanhar o PIN médio. O PIN mínimo ficou entre 2,12% (segundo trimestre de 2007) e 6,29% (primeiro trimestre de 2006). As maiores oscilações aconteceram com o PIN máximo que, no período, oscilou entre 43,84% (segundo trimestre de 2006) e 21,28% (terceiro trimestre de 2008).

5.2. Seleção de portfólios

Os conjuntos de portfólios (X) e (Y) foram criados conforme descrito na seção 3. É importante destacar que as variáveis *valor de mercado* e *PIN* são conhecidas no final de cada trimestre para compor os *portfólios* no trimestre seguinte. O mesmo não acontece com o *book-to-market* que, embora seja do final de cada trimestre, não é imediatamente disponibilizado nas infor-

mações trimestrais de resultados (ITR). Neste ensaio, parte-se do pressuposto de que o *book-to-market* também seja conhecido no início do trimestre subsequente quando ocorre a nova composição das carteiras. Esse procedimento impacta igualmente os conjuntos (X) e (Y), de modo que é esperado que os demais resultados obtidos possam ser explicados pela variável PIN.

Para o conjunto (X), são criados 4 *portfolios*, considerando as variáveis *valor de mercado* e índice *book-to-market*. No conjunto (Y), formam-se 8 *portfolios* ao incorporar a variável PIN na seleção. Admite-se que cada variável possa assumir valores altos ou baixos, sendo a separação feita pela mediana, considerando que as ações que compõem os conjuntos (X) e (Y) são as mesmas.

Após a composição dos *portfolios* do conjunto (X) e (Y), foi obtido o retorno das carteiras e os prêmios pelos fatores de risco. As estatísticas descritivas dos retornos e do desvio-padrão são apresentadas na Tabela 2.

Tabela 2 – Média dos retornos e desvio padrão entre jan-2006 até mar-2011.

	X			Y	
	Média	DP		Média	DP
BH	0,1040%	2,0775%	BHG	0,0826%	2,2540%
BL	0,0822%	2,2664%	BHP	0,1097%	2,0681%
SH	0,0924%	1,8110%	BLG	0,0566%	2,1486%
SL	0,0753%	1,7577%	BLP	0,0922%	2,2874%
			SHG	0,1054%	1,8273%
			SHP	0,0911%	2,0002%
			SLG	0,1034%	1,7211%
			SLP	0,0340%	2,1055%
IBOVESPA	0,06963%	2,08807%			

Fonte: elaboração própria.

Na Tabela 2, é possível comparar as médias diárias dos retornos e o desvio padrão dos *portfolios* do conjunto (X) e (Y). No conjunto (X), a carteira BH é a que apresenta maior retorno médio diário, de 0,1040%, seguida da SH, com 0,0924%. As carteiras com alto índice *book-to-market* (BH e SH) apresentaram maiores retornos, enquanto que as carteiras com menor valor de mercado (SH e SL) apresentaram um desvio-padrão menor. Quando comparados os *portfolios* (SH, SL, BH, BL) com a média de retornos diários e com o desvio-padrão do IBOVESPA, constata-se que SH apresenta maior retorno médio diário e menor desvio-padrão. No conjunto (Y), os *portfolios* SLP e BLG apresentaram retornos médios diários inferiores ao IBOVESPA

e desvio-padrão maior. Os *portfolios* que apresentaram maior retorno médio diário foram BHP (0,1097%), SHG (0,1054%) e SLG (0,1034%). É importante destacar que os *portfolios* SLG e SHG ainda aprearam o menor desvio-padrão no conjunto (Y).

5.3. Estimação do modelo

Os resultados das estimações dos modelos MS para os *portfolios* do conjunto (X) são apresentados na Tabela 3. O modelo é estimado admitindo a existência de dois regimes, um de baixa e outro de alta volatilidade.

Tabela 3 - Estimativas dos parâmetros para o conjunto X

$R_i(t) - RF(t) = \alpha(s_i) + b(s_i)[RM(t) - RF(t)] + s(s_i)[SMB(t)] + h(s_i)[HML(t)] + e(t)$									
		Regime 1				Regime 2			
	C	RM-RF	SMB	HML	C	RM-RF	SMB	HML	
BH	0,0001 1,0805	0,8814 90,1689	-0,2827 -12,8129	0,6982 37,1827	0,0001 1,0805	0,8221 31,3564	-0,3755 -5,8085	0,7784 12,043	
BL	0,0002 2,1128	0,9192 110,208	-0,4444 -31,0805	-0,5233 -32,5896	0,0001 0,1113	0,8817 30,7883	-0,4072 -5,613	-0,6894 -8,9282	
SH	0,00025 2,1383	0,93699 119,32	0,57469 40,76	0,47882 27,822	-0,00017 -0,15564	0,77786 14,945	0,44647 4,4357	0,18235 1,5356	
SL	0,0001 0,8484	0,8811 90,1306	0,7164 32,3973	-0,3018 -16,0333	0,0006 0,798	0,8219 31,2684	0,6265 9,7416	-0,2251 -3,4903	

Fonte: elaboração própria

Notas: A primeira linha com dados numéricos de cada *portfolio* é o valor dos parâmetros estimados para cada variável e a segunda linha o valor da estatística t ;

O Teste LR teve a hipótese nula de linearidade do modelo rejeitada para todos os *portfolios*. Admite-se um modelo não linear com mudança de regime.

Na Tabela 3, são apresentadas as estimativas dos parâmetros e as estatísticas t para ambos os regimes do conjunto (X). No regime 1, caracterizado por menor volatilidade dos retornos, é observado que os prêmios, SMB e foram significativos ao nível de significância de 1%⁸, acontecendo o mesmo no regime 2, com exceção do prêmio HML no *portfolio* SH. O prêmio ($RM - RF$) apresenta uma relação

⁸ Admitindo um teste bilateral, o valor crítico da estatística t é 2,576.

direta na explicação dos retornos diários das carteiras em ambos os regimes. O prêmio por tamanho de mercado (*SMB*) apresenta uma relação direta para as carteiras com baixo valor de mercado e uma relação inversa para as carteiras com alto valor de mercado. Ao observar o prêmio *HML*, percebe-se uma relação direta com o retorno diário das carteiras com alto índice *book-to-market* e uma relação inversa com as carteiras de baixo índice.

Na Tabela 4, são apresentados os resultados dos modelos MS para o conjunto de *portfolios* (*Y*).

Tabela 4 – Resultados das estimações para o conjunto Y.

$$R_i(t) - RF(t) = a(s_i) + b(s_i)[RM(t) - RF(t)] + s_i(s_i)[SMB(t)] + h(s_i)[HML(t)] + p(st)[PINF(t)] + e(t)$$

Carteira	Regime 1					Regime2				
	Const	RM-RF	SMB	HML	GMP	Const	RM-RF	SMB	HML	GMP
BHG	-0,0004	0,8347	-0,6179	0,8115	0,2793	-0,0004	0,9615	-0,0795	0,6624	0,5691
	-2,031	20,9316	-5,1866	7,9463	1,9101	-2,031	53,2569	-1,7803	22,9843	12,3679
BHP	-0,0001	0,7502	-0,4066	0,8188	-0,5606	-0,0001	0,8665	-0,2435	0,492	-0,1869
	-0,5311	31,9795	-7,3756	13,6814	-6,8381	-0,5311	58,7879	-8,4727	15,9505	-4,1327
BLP	0,0001	0,5635	-0,854	-1,0441	-0,8607	0,0001	0,9766	-0,3721	-0,4202	-0,2949
	0,6395	8,1569	-5,6184	-6,3925	-4,1286	0,6395	77,7542	-15,4253	-19,2792	-9,4568
BLG	-0,0003	0,8392	-0,4551	-0,4312	0,6868	-0,0003	0,8983	-0,2082	-0,7309	0,7487
	-1,3925	49,0545	-10,1775	-11,5454	13,6301	-1,3925	45,2577	-4,3984	-15,1090	12,2706
SHG	-9,56E-05	0,71855	0,56698	0,19667	0,42056	9,56E-05	1,0197	0,69975	0,48942	0,88301
	-0,57611	25,341	9,7782	3,3512	6,2028	-0,57611	64,131	20,427	14,922	15,825
SHP	-0,0001	0,8376	0,5742	0,5815	-0,3311	-0,0001	0,9556	0,6213	0,4765	-0,3074
	-0,4654	33,0148	15,1857	19,3793	-7,3265	-0,4654	36,4401	9,9664	8,1694	-3,508
SLG	0,0002	0,7944	0,5421	-0,235	0,5845	0,0002	0,8382	0,6394	-0,2548	0,6759
	1,15	51,4577	18,0935	-8,904	16,2643	1,15	37,9555	11,2577	-4,5975	8,4992
SLP	-0,0007	0,7902	0,6284	-0,3494	-0,6987	-0,0007	0,9841	0,9987	-0,4447	-0,3515
	-2,9475	18,3416	3,4718	-3,0838	-3,5618	-2,9475	38,3082	13,5586	-9,5574	-5,523

Fonte: elaboração própria.

Nota: A primeira linha com dados numéricos de cada *portfolio* é o valor dos parâmetros estimados para cada variável e a segunda linha o valor da estatística *t*.

O Teste LR teve a hipótese nula de linearidade do modelo rejeitada para todos os *portfolios*. Admite-se um modelo não linear com mudança de regime.

Na Tabela 4, é possível visualizar os parâmetros estimados e o valor das estatísticas t em ambos os regimes. O regime 1 é caracterizado por uma menor volatilidade dos retornos, enquanto que o regime 2 tem por característica uma maior volatilidade. O prêmio por risco de mercado ($RM - RF$) é significativo em ambos os regimes ao nível de significância de 1% e apresenta uma relação direta com os retornos dos *portfolios*.

O prêmio por tamanho de mercado (SMB) é significativo em todas as carteiras no regime 1 e apresenta, na explicação dos retornos, uma relação inversa com os *portfolios* com alto valor de mercado (B) e uma relação direta com os de baixo valor de mercado (S). O mesmo comportamento é observado no regime 2. No entanto, SMB, na carteira BHG, só é significativo a 5% admitindo-se um teste unilateral.

O prêmio HML é significativo a 1% em ambos os regimes em todas as carteiras, e apresenta uma relação direta com os retornos das carteiras com alto índice *book-to-market* e uma relação inversa com as carteiras de baixo índice.

O prêmio por informação (GMP) é significativo a 1% em quase todas as carteiras nos dois regimes, exceto no regime 1 na carteira BHG. Nas carteiras com baixo PIN, tem-se uma relação inversa do fator em relação aos retornos diários, e uma relação direta com alto PIN.

6. Comentários finais

Neste artigo, foi verificado se a probabilidade de negociação informada (PIN), desenvolvida por Easley, Kiefer, O'Hara, e Paperman (1996), contribui para explicar os retornos de carteiras de investimentos compostas nos moldes de Fama e French (1993). O estudo foi desenvolvido utilizando ações que participaram do IBRX entre janeiro de 2006 até março de 2011.

Para obtenção do PIN, fez-se necessário classificar as ordens executadas diariamente, e esse procedimento foi realizado aplicando-se o algoritmo *tick-test* aos dados de negócios. O PIN mediano no período oscilou entre 11% e 14%. No entanto, houve variação significativa do PIN estimado dentro de cada trimestre. O ponto fundamental é

se essa *proxy* de informação é significativa para explicar os retornos de *portfolios*. Com base em Fama e French (1993), foram construídos dois conjuntos de carteiras para o mesmo grupo de ações. A diferença na composição dos *portfolios* desses grupos é que, para a seleção dos mesmos no primeiro grupo, foram consideradas apenas as variáveis de Fama e French (1993), enquanto que, no segundo, a separação foi feita considerando a inserção da variável PIN. Todas as quatro carteiras do primeiro conjunto apresentaram um desempenho superior à carteira de mercado. No segundo grupo, seis das oito carteiras apresentaram um retorno médio superior ao do mercado. Contudo, o importante é comparar as carteiras que utilizam a variável PIN e as que não utilizam. A utilização do PIN para carteiras com baixo valor de mercado apresenta um comportamento interessante, pois possibilita uma melhor composição das mesmas, melhorando a relação risco/retorno. A carteira SH, quando emprega o PIN, compõe SHG e SHP. SHG teve um aumento do retorno em relação à SH e praticamente manteve o desvio-padrão, enquanto que SHP teve uma pequena queda na média dos retornos diários com aumento do risco quando comparada com SH. No entanto, é em SL que se percebe o maior efeito do PIN ao compor SLG e SLP. Ações com pequeno valor de mercado, baixo índice *book-to-market*, mas com alto PIN, compõem uma das melhores carteiras, pois apresentam um dos melhores retornos e o menor desvio-padrão. Já a SLP apresentou o menor retorno médio diário, ficando abaixo da média do IBOVESPA e apresentando um desvio-padrão elevado.

A estimação do método de Fama e French (1993), utilizando modelos de *markov-switching* demonstrou que os prêmios pelos fatores foram significativos em ambientes de alta e baixa volatilidade, acontecendo o mesmo para o prêmio por informação, que apresentou uma relação direta para carteiras com alto PIN e uma relação inversa para carteiras de baixo PIN.

Referências

- BAGEHOT, W. The only game in town. *Financial Analysts Journal*, v. 27, p. 12–14, 1971.
- BANZ, R. The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics*, v. 9, n. 1, p. 3-18, 1981.
- BARBEDO, C.; SILVA, E.; LEAL, R. Probabilidade de informação privilegiada no mercado de ações, liquidez intra-diária e níveis de governança corporativa. *Revista Brasileira de Economia*, v. 63, n.1, p. 51-62, jan., 2009.
- BASU, S. Investment Performance on Common Stocks in Relation to Their Price/Earnings Ratio: A Test of the Efficient Market Hypothesis. *The Journal of Finance*, v. 32, n. 3, p. 663-682, 1977.
- BLACK, F.; JENSEN, M. C.; SCHOLES, M. The Capital Asset Pricing Model: Some Empirical Testes. In M.C. Jensen (org.). *Studies in the Theory of Capital Markets*. New York: Praeger, 1972.
- COGGI, P.; MANESCU, B. A multifactor model of stock returns with endogenous regime switching. Jan, 2004. Disponível em: < http://papers.ssrn.com/sol3/papers.cfm?abstract_id=526064> . Acesso em 05 ago. 2011.
- COPELAND, L.; WONG, W.; ZENG, Y. Information-based trade in Shanghai stock market. *Global Finance Journal*, v. 20, n. 2, p. 180-190, jun., 2009.
- DUARTE, J.; YOUNG L. Why is PIN priced? *Journal of Financial Economics*, v. 91, n. 2, p. 119-138, fev., 2009.
- EASLEY D.; KIEFER, N.; O'HARA, M.; PAPERMAN, J. Liquidity, information, and infrequently traded stocks, *Journal of Finance*, v. 51, n. 4, p. 1405–1436, set., 1996.
- EASLEY, D.; HVIDKJAER, S.; O'HARA, M. Factoring information into returns, *Journal of Financial Quantitative Analysis*, v.45, n. 2, p. 293-309, fev., 2010.
- EASLEY, D.; HVIDKJAER, S.; O'HARA, M. Is information risk a determinant of asset returns? *Journal of Finance*, v. 57, n. 5, p. 2185–2222, out., 2002.
- EASLEY, D.; O'HARA, M. Information and the cost of capital. *Journal of Finance*, v. 59, n. 4, p. 1553–1583, ago., 2004.
- EASLEY, D.; O'HARA, M. Price, trade size, and information in securities markets. *Journal of Financial Economics*, v. 19, n. 1, p. 69–90, set., 1987.
- EASLEY, D.; KIEFER, N.; O'HARA, M. One day in the life of a very common stock, *Review of Financial Studies*, v. 10, n. 3, p. 805-835, 1997.
- FAMA, E.; MACBETH, J. Risk, return and equilibrium: empirical tests. *Journal of Political Economy*, n. 81, n. 3, p. 607–636, jun., 1973.
- FAMA, F.; FRENCH, K. Common risk factors in the returns on stock and bonds. *Journal of Financial Economics*, v. 33, n. 1, p. 3-56, fev., 1993.
- FAMA, F.; FRENCH, K. The Cross-Section of Expected Stock Returns. *Journal of Finance*, v. 47, n. 2, p. 427–465, jun., 1992.
- HAMILTON, J. A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, v. 57, n. 2, p. 357-384, mar., 1989.
- KROLZIG, H-M, *Markov-Switching vector autoregression: modeling, statistical inference and application to business cycle analysis (lectures notes in economics and mathematical systems)*. New York: Springer, 1997, 357p.
- KYLE, A. S. Continuous auctions and insider trading, *Econometrica*, v. 53, n. 6, p. 1315–1335, nov., 1985.
- LEE, C.; READY, J. Inferring trade direction from intraday data. *Journal of Finance*, v. 46, n. 2, p. 733–746, jun. 1991.

- LINTNER, J. The Valuation of Risk Assets and the Selection of Risk Investments in Stocks Portfolios and Capital Budgets. *The Review of Economics and Statistics*, v. 47, p. 13-37, feb., 1965.
- MOSSIN, J. Equilibrium in a Capital Asset Market. *Econometrica*, v. 34, p. 768-783, oct., 1966.
- ROSENBERG, B.; REID, K.; LANSTEIN, R. Persuasive Evidence of Market Inefficiency. *Journal of Portfolio Management*, v. 11, n. 2, p. 9-17, 1985.
- ROSS, S.A. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13, pp. 341-360, 1976.
- ROSS, Stephen A. A simple approach to the valuation of risky streams. *Journal of Business*, p. 453-475, 1978.
- SHARPE, W. F. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance*, v. 19, p. 425-442, sept., 1964.
- SILVA, E. Dois ensaios sobre microestrutura de Mercado e probabilidade de informação privilegiada no mercado de ações brasileiro. 2009. Tese (Doutorado em Administração) – Programa de Pós-Graduação em Administração, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.