Identificando sinais esparsos para o Ibovespa

Evandro Amorim[†, a]
Ricardo Buscariolli[‡, b]
Marcelo Fernandes[*, a]

[a]Escola de Economia de São Paulo, Fundação Getulio Vargas, Brasil
[b]Universidade Federal do ABC, Brasil

Resumo Neste trabalho, prevemos o Ibovespa um passo à frente usando os retornos passados de todas as suas ações constituintes. Consideramos retornos nas frequências de 1, 5 e 10 minutos em janelas móveis de 30 observações. Adotamos uma regularização via LASSO para reduzir a dimensionalidade do problema e identificar sinais esparsos. A nossa estratégia de previsão traz ganhos estatisticamente significativos de poder preditivo em relação a modelos autorregressivos, embora os ganhos líquidos sejam limitados pelos custos de transação.

Palavras-chave: Esparsidade; LASSO; Previsibilidade; Regularização.
Códigos JEL: C55, C58, G12, G14.

1. Introdução

Neste trabalho, calculamos a previsão dos retornos intradiários do IBOVESPA a partir dos retornos passados de suas ações constituintes, considerando uma janela móvel de 30 observações apenas. Para tanto, precisamos reduzir a dimensionalidade do problema de previsão, já que o número de regressores (a saber, três defasagens dos retornos de 60 ações) supera em muito o número de observações temporais. Assim como em Chinco et al. (2019), empregamos o procedimento de LASSO para identificar quais previsores entre os diversos candidatos são realmente relevantes a cada instante de tempo, sem depender de qualquer intuição arbitrária (Gabaix, 2014). Essa abordagem impõe esparsidade no sentido que apenas alguns poucos retornos individuais têm capacidade genuína de prever o Ibovespa a cada instante do tempo.

Como citar: Amorim, E., Buscariolli, R. & Fernandes, M. (2026). Identificando sinais esparsos para o Ibovespa. *Revista Brasileira de Finanças*, *24*, e202601. https://doi.org/10.12660/rbfin.v24n1.2026.94372
Submetido em 22 de setembro de 2025. Publicado *on-line* em 22 de janeiro de 2026.
Editor responsável: Bruno Giovannetti.
[†]evandro.jivando@gmail.com
[‡]ricardo.buscariolli@ufabc.edu.br
[*]marcelo.fernandes@fgv.br

2026 – Os Autores, sob uma licença Creative Commons Attribution 4.0 International.

O algoritmo é de tal sorte que atribui-se valor zero para os coeficientes de todos os retornos que não contribuem para a previsão, permitindo assim solucionar o problema de dimensionalidade (Tibshirani, 1996). Além de esparsos, procuramos sinais de curta direção, já que contemplamos janelas móveis de apenas 30 observações intradiárias. No mais, para investigar a diferença na capacidade de previsão de cada ação em diferentes frequências, rodamos o modelo de previsão para o retorno do Ibovespa em intervalos de tempo de 1, 5 e 10 minutos. Nossos resultados mostram que a utilização do LASSO aumenta a aderência fora da amostra em comparação com modelos puramente autorregressivos, seja qual for a frequência.

Nosso trabalho insere-se na literatura de previsibilidade (McLean e Pontiff, 2016; Harvey et al., 2015; Linnainmaa e Roberts, 2018), especialmente ao que concerne o emprego de técnicas de regularização (DeMiguel et al., 2009). As evidências mostram que pequenos incrementos em previsibilidade podem gerar grandes benefícios para os investidores (Campbell e Thompson, 2008). Feng et al. (2020) empregam o LASSO para selecionar fatores de risco que contribuem para melhorar a previsão do prêmio de risco entre 150 candidatos. Por sua vez, Freyberger et al. (2020) usam o LASSO para selecionar características que afetam os retornos esperados de cada ação. Kozak et al. (2020) mostram que previsões regularizadas por LASSO (penalização em ℓ_1) ou por ridge (penalização em ℓ_2) melhoram o índice de Sharpe em dados mensais.

No entanto, nossa abordagem está mais próxima àquela de Chinco et al. (2019), que usam LASSO para prever o retorno de ações individuais com base nos retornos passados de todas as ações da Bolsa de Valores de Nova Iorque. Em trabalho independente e concomitante, Alexandre et al. (2024) praticamente replicam o exercício empírico de Chinco et al. (2019) para ações individuais da B3, porém usando diferentes métodos de regularização. Nosso trabalho difere em duas dimensões, ao menos. Primeiro, buscamos prever apenas o retorno do Ibovespa a partir dos retornos passados de suas principais ações constituintes. O foco no índice de mercado, em vez de ações individuais, tem implicações mais diretas para a gestão de carteiras de renda variável e estratégias de hedge com contratos futuros e ETFs. No mais, ao restringir a atenção às ações constituintes, atenuamos problemas de liquidez que possam distorcer a análise, visando identificar sinais mais robustos. Segundo, não estamos tão interessados na comparação meramente estatística entre os diferentes métodos. Em particular, avaliamos a relevância econômica das previsões através de estratégias de negociação. A análise demonstra que, embora os ganhos líquidos sejam limitados pelos custos de transação, há evidências de sinais esparsos de curta duração na B3.

O resto do trabalho procede da seguinte forma. Na seção 2, apresentamos os dados e a estratégia empírica. Na seção 3, discutimos os resultados empíricos sobre a previsão do retorno do Ibovespa. A seção 4 conclui.

2. Estratégia Empírica

A ideia é prever o retorno do Ibovespa no próximo instante de tempo com base nos retornos passados das principais ações que compõem o índice. Em particular, contemplamos uma uma especificação com até três defasagens:

$$r_{IB,t} = \beta_0 + \sum_{n=1}^{N}\sum_{j=1}^{3}\beta_{n,j}r_{n,t-j} + \varepsilon_t, \tag{1}$$

em que $r_{IB,t}$ é o retorno do Ibovespa no instante $t \in \{1,\ldots,T\}$, $r_{n,t}$ é o retorno da ação $n \in \{1,\ldots,N\}$ da carteira teórica do Ibovespa e ε_t é um ruído branco que captura a incerteza inerente à previsão.[1]

A regressão em (1) tem um número bastante alto de parâmetros $(3N+1)$, especialmente considerando uma janela móvel de $T = 30$ observações apenas. Para tornar factível a estimação, usamos o procedimento de LASSO para impor esparsidade, atribuindo valor zero para os coeficientes dos retornos passados de menor contribuição para a previsão do Ibovespa. Mais precisamente, minimizamos a perda quadrática sob uma penalização da norma ℓ_1 do vetor de parâmetros:

$$\arg\min_{\beta}\left\{\sum_{t=1}^{T}\left(r_{IB,t} - \beta_0 - \sum_{n=1}^{N}\sum_{j=1}^{3}\beta_{n,j}r_{n,t-j}\right)^2 + \lambda\sum_{n=1}^{N}\sum_{j=1}^{3}|\beta_{n,j}|\right\}, \tag{2}$$

em que λ é o parâmetro de regularização que define a severidade da penalização.

Para selecionar o grau de regularização, utilizamos o método de validação cruzada com 10 partições dos dados, assim como em Chinco et al. (2019). De acordo com Bergmeir et al. (2018), o método de validação cruzada funciona bem no contexto de séries temporais caso o processo gerador dos dados seja puramente autorregressivo e os resíduos não apresentem correlação serial. Em nosso estudo, selecionamos o valor de λ dentre 100 valores que minimizam

[1] Em princípio, uma alternativa seria controlar pelos retornos passados do próprio Ibovespa, além dos retornos individuais das ações constituintes. Porém, o retorno do índice Ibovespa é uma combinação linear dos retornos das ações constituintes, trazendo problemas de colinearidade que costumam afetar gravemente o procedimento de LASSO ao colocar em xeque a condição de irrepresentabilidade.

o erro quadrático médio fora da amostra entre partições. Além do modelo baseado em LASSO, calculamos previsões a partir de modelos autorregressivos com uma estrutura fixa de defasagens — AR(1) a AR(6) — e com estrutura determinada, em cada janela de estimação, pelo critério de informação de Akaike (AIC) com, no máximo, 10 defasagens.

Para medir o erro de previsão, calculamos primeiro o quanto nosso modelo de previsão explica da variação total dos retornos através da seguinte regressão:

$$r_{IB,t} = a_L + b_L \left(\frac{\hat{r}^{(L)}_{IB,t} - \bar{r}^{(L)}_{IB}}{s^{(L)}} \right) + e^{(L)}_t \tag{3}$$

em que $\hat{r}^{(L)}_{IB,t}$ denota a previsão do retorno do Ibovespa segundo o modelo estimado por LASSO, $\bar{r}^{(L)}_{IB}$ a média dessas previsões e $s^{(L)}$ seu desvio padrão. No mais, assim como em Chinco et al. (2019), comparamos ainda a contribuição marginal do modelo LASSO usando uma regressão que acrescenta as previsões de modelos autorregressivos:

$$r_{t+1} = a_c + b_L \left(\frac{\hat{r}^{(L)}_{IB,t} - \bar{r}^{(L)}_{IB}}{s^{(L)}} \right) + b_{\text{AR}} \left(\frac{\hat{r}^{(\text{AR})}_{IB,t} - \bar{r}^{(\text{AR})}_{IB}}{s^{(\text{AR})}} \right) + e^{(c)}_t. \tag{4}$$

Comparando os coeficientes ajustados de determinação R^2 de cada regressão, podemos verificar facilmente o quanto a identificação de sinais esparsos nos retornos passados das ações constituintes ajuda na previsão dos retornos do Ibovespa.

2.1 Dados

Coletamos dados intradiários de negociação das ações listadas na B3 entre maio de 2017 e dezembro de 2018 diretamente do servidor FTP da B3. Em seguida, calculamos os retornos nas frequências de 1, 5 e 10 minutos a partir dos preços de fechamento destes intervalos. Consideramos apenas os negócios realizados entre os leilões de abertura e fechamento, excluindo qualquer observação no período de pós-negociação. A amostra contempla apenas as ações pertencentes à carteira teórica do Ibovespa, cuja composição está disponível no UP2DATA. Conforme a metodologia do Ibovespa, a cada quatro meses, há uma revisão da composição da carteira com base no índice de negociabilidade de cada ação. Durante o período amostral, a composição do Ibovespa muda quatro vezes, sempre compreendendo entre 58 e 65 ações, das

Tabela 1
Estatísticas descritivas dos retornos percentuais do Ibovespa

	agregação temporal		
	1 minuto	5 minutos	10 minutos
média	0,00	0,00	0,00
mínimo	−4,58	−7,42	−8,00
primeiro quartil	−0,02	−0,06	−0,08
mediana	0,00	0,00	0,00
terceiro quartil	0,02	0,06	0,08
máximo	2,57	2,93	2,80
desvio padrão	62,27	30,44	23,27
número de observações	173.440	34.875	14.897

quais entre 7 e 9 possuem mais de 50% do peso conjuntamente.[2] Naturalmente, no momento em que há um rebalanceamento do índice, deixamos de prever as trinta primeiras observações dos retornos, qualquer que seja a frequência.

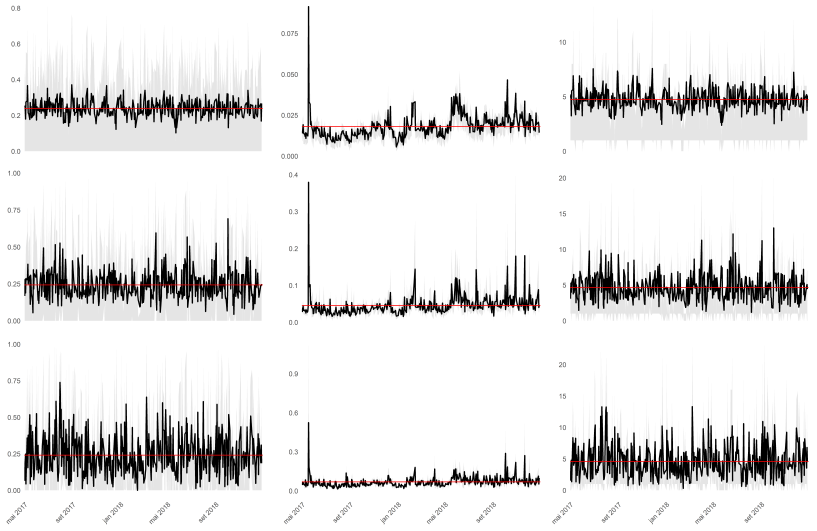
A Tabela 1 mostra que os retornos do Ibovespa são tipicamente muito próximos de zero, não importando a frequência. Já a variação dos retornos aumenta consideravelmente com a frequência: a volatilidade anualizada dos retornos de 1 minuto é de aproximadamente 62% e cai para 23,27% nos retornos de 10 minutos, enquanto que a amplitude (ou seja, a diferença entre os valores máximo e mínimo) vai de 7,15% para 10,80% por conta principalmente da cauda negativa. O excesso de volatilidade nas frequências mais alatas indica a provável presença de ruídos de microestrutura.

3. Resultados e discussão

A primeira coluna da Figura 1 estampa as médias diárias dos coeficientes ajustados de determinação $\bar{R}^2$ dos modelos estimados por LASSO para os retornos de 1, 5 e 10 minutos do Ibovespa, além dos respectivos intervalos interquartis (área hachurada). O valor médio do $\bar{R}^2$ no período amostral (linha vermelha) é muito próximo a 24% para todas as frequências, apesar da variabilidade do ajuste diminuir com a frequência. De fato, o desvio padrão do $\bar{R}^2$ aumenta de 4,44% no intervalo de um minuto (frequência mais alta) para 13,78% no intervalo de 10 minutos (frequência mais baixa). O primeiro quartil do $\bar{R}^2$ é muito próximo de zero em todos os dias da amostra. Isso acontece porque, em aproximadamente 28% das vezes, o algoritmo LASSO exclui todos

[2] No caso de uma oferta pública de aquisição (OPA), o peso do papel na carteira teórica da Ibovespa é distribuído pelas demais ações constituintes de acordo com o peso delas. Isso ocorre na nossa amostra com as ações da BVMF3, SMLE3, SUZB5 e VALE5.

Figura 1
Ajuste dentro da amostra das previsões por LASSO nas diferentes frequências



A primeira coluna refere-se aos valores do $\bar{R}^2$ do modelo de previsão estimado por LASSO para os retornos de 1, 5 e 10 minutos, respectivamente. A segunda coluna reporta os valores do parâmetro de regularização λ estimado por validação cruzada, enquanto que a terceira coluna estampa o número de previsores selecionados pelo algoritmo LASSO. Em todos os gráficos, as linhas preta e vermelha correspondem respectivamente às médias em cada dia e no período inteiro, enquanto que a área hachurada representa o intervalo interquartil em cada dia da amostra.

os regressores do modelo, estimando apenas o intercepto. Já o terceiro quartil indica que o modelo de previsão por LASSO tem dias de alta aderência. De fato, $\bar{R}^2 \geqslant 70\%$ em aproximadamente 15% dos dias em todas as frequências.

A segunda coluna da Figura 1 reporta os valores médios do parâmetro de penalização λ para cada dia (linha preta) e na amostra completa (linha vermelha), além dos seus intervalos interquartis diários (área hachurada). O procedimento de validação cruzada seleciona valores de λ que crescem, em média, conforme aumentamos o intervalo de amostragem: a saber, 0,018, 0,046 e 0,068 para 1, 5 e 10 minutos, respectivamente. Apesar da evolução diária do lambda médio ser bastante semelhante entre as diferentes frequências, a dispersão diária claramente aumenta com a frequência. Já a terceira coluna da Figura 1 mostra quantos previsores sobrevivem a penalização LASSO, em média, por dia. O número médio de coeficientes diferentes de zero varia entre 4,57 e 4,73 decaindo levemente com o intervalo de amostragem. Não há muitos casos em que o LASSO mantém muitos previsores: por exemplo,

selecionamos modelos com 15 previsores ou mais em aproximadamente 11% dos dias.

Como o número de previsores escolhidos pelo LASSO é baixo, podemos dizer que os sinais são esparsos. Resta saber se têm realmente vida curta, como em Chinco et al. (2019). Para tanto, verificamos a persistência de cada previsor ao longo do tempo calculando o número de vezes consecutivas que a estimativa de seu coeficiente é diferente de zero. Na média, utilizamos cada previsor menos de 2 vezes consecutivas em qualquer frequência. De fato, como a mediana é igual a um para todas as frequências, sabemos que não usamos o mesmo preditor consecutivamente em mais da metade das vezes. E, quando permanece no modelo de previsão, não persiste por muito tempo, já que o terceiro quartil é igual a dois.

A Tabela 2 apresenta as estimativas dos coeficientes das regressões (3) e (4) usando as amostras completas dos retornos de 1, 5 e 10 minutos. As estimativas dos efeitos parciais das previsões por LASSO são estatisticamente significantes ao nível de 1% em praticamente todos os casos. A única exceção corresponde à especificação que acrescenta as previsões de um modelo AR(1) para os retornos de 10 minutos. Os modelos autorregressivos também contribuem bastante para a previsão dos retornos do Ibovespa, já que a vasta maioria das estimativas dos respectivos coeficientes de regressão são significativas ao nível de 1%. Há três exceções apenas: AR(3) para retornos de 1 minuto, que é significativo apenas ao nível de 10%; AR(2) para os retornos de 5 minutos; e AR(3) para retornos de 10 minutos.[3]

A variação do ajuste fora da amostra quando adicionamos as previsões por LASSO mostra que sua contribuição marginal é claramente substancial, mais que dobrando o valor de $\bar{R}^2$ na vasta maioria dos casos. A pior melhora, mesmo que substancial, é sempre para o modelo AR(1), variando entre 43% e 78% dependendo da frequência dos dados. A inclusão das previsões por LASSO é particularmente benéfica para os modelos AR de ordem mais alta por terem uma estrutura desnecessariamente rica de defasagens. O caso mais extremo refere-se às previsões do modelo AR(6) para os retornos de 5 minutos. Por conta do provável excesso de parâmetros, seu desempenho é pífio, apresentando inclusive $\bar{R}^2$ negativo fora da amostra e, portanto, distoando muito dos demais modelos AR. Quando adicionamos as previsões por LASSO, o poder de ajuste fora da amostra passa a ser não apenas positivo (daí a variação percentual infinita), mas também comparável aos demais modelos.

[3] Na comparação dentro da amostra, testes de Chow mostram que a melhoria no ajuste é estatisticamente significativa ao nível de 1% em todos os casos, exceto para o modelo AR(1) usando retornos de 10 minutos. Esses resultados estão naturalmente disponíveis mediante solicitação aos autores.

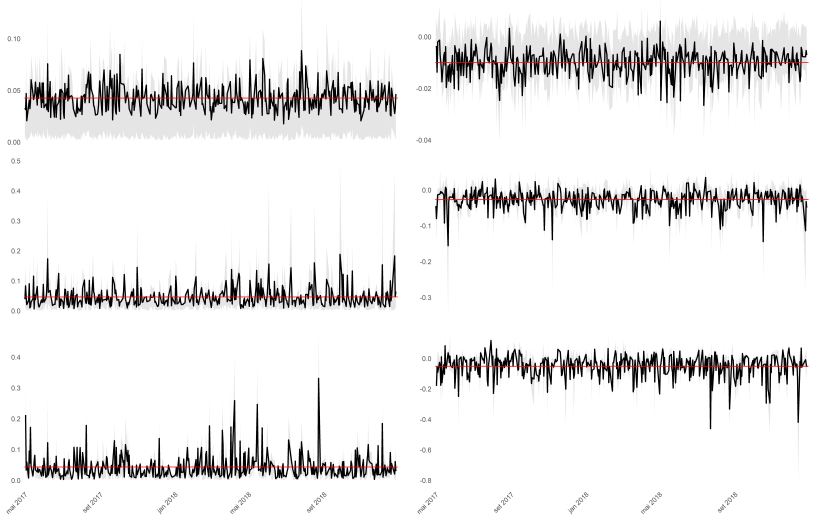
Tabela 2
Ajuste das previsões para cada frequência

		estrutura de defasagem do modelo AR						
	LASSO	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	k=AIC
retornos de 1 minuto								
constante	–0,014***	–0,014***	–0,014***	–0,014***	–0,014***	–0,014***	–0,014***	–0,014***
	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)
LASSO	–0,070***	–0,086***	–0,082***	–0,065***	–0,088***	–0,086***	–0,092***	–0,090***
	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)
AR		–0,110***	–0,092***	–0,025***	–0,172***	–0,174***	–0,230***	–0,200***
		(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)	(0,014)
$\bar{R}^2$	0,013	0,051	0,038	0,015	0,100	0,100	0,170	0,130
$\Delta\bar{R}^2$		43%	111%	175%	238%	329%	539%	116%
retornos de 5 minutos								
constante	–0,065***	–0,065***	–0,065***	–0,065***	–0,065***	–0,065***	–0,065***	–0,065***
	(0,075)	(0,075)	(0,075)	(0,075)	(0,075)	(0,075)	(0,075)	(0,075)
LASSO	–0,350***	–0,410***	–0,340***	–0,250***	–0,250***	–0,290***	–0,280***	–0,260***
	(0,070)	(0,070)	(0,075)	(0,076)	(0,076)	(0,076)	(0,076)	(0,076)
AR		–1,050***	–0,100***	–0,480***	–0,490***	–0,340***	–0,410***	–0,450***
		(0,070)	(0,075)	(0,076)	(0,076)	(0,076)	(0,076)	(0,076)
$\bar{R}^2$	0,060	0,619	0,062	0,171	0,175	0,114	0,140	0,159
$\Delta\bar{R}^2$		78%	158%	419%	767%	1,200%	∞	191%
retornos de 10 minutos								
constante	–0,110***	–0,110***	–0,110***	–0,110***	–0,110***	–0,110***	–0,110***	–0,110***
	(0,150)	(0,150)	(0,150)	(0,150)	(0,150)	(0,150)	(0,150)	(0,150)
LASSO	–0,730***	–0,230***	–0,530***	–0,640***	–0,540***	–0,520***	–0,560***	–0,420***
	(0,150)	(0,170)	(0,160)	(0,160)	(0,150)	(0,150)	(0,150)	(0,016)
AR		–0,950***	–0,430***	0,220***	–0,630***	–0,700***	–0,570***	–0,750***
		(0,170)	(0,160)	(0,160)	(0,150)	(0,150)	(0,150)	(0,160)
$\bar{R}^2$	0,120	0,280	0,160	0,130	0,210	0,230	0,190	0,230
$\Delta\bar{R}^2$		47%	139%	134%	192%	575%	521%	129%

A coluna LASSO exibe as estimativas dos coeficientes da regressão (3) e seus respectivos erros padrão, multiplicados por 100. As demais colunas reportam os coeficientes da regressão (4) que adicionam as previsões de modelos autorregressivos, com número fixo de defasagens ($k = 1,\ldots,6$) ou determinado pelo critério de informação de Akaike (k =AIC). No mais, reportamos o poder de ajuste de cada regressão dentro da amostra, medido pelo coeficiente ajustado de determinação $\bar{R}^2$ (em %), assim como o quanto a adição das previsões por LASSO melhora o ajuste do modelo AR fora da amostra ($\Delta\bar{R}^2$). Denotamos por *, ** e *** significância estatística aos níveis de 10%, 5% e 1%, respectivamente.

A Figura 2 exibe os valores médios diários de $\bar{R}^2$ das previsões por LASSO fora da amostra, assim como as estimativas dos coeficientes de regressão em (3) usando uma janela móvel de 30 observações, para cada frequência. Não há grandes diferenças entre o poder de ajuste médio entre as frequências, variando entre 4,25% e 4,61%. Apesar do intervalo interquartil reduzir significativamente quando agregamos os retornos para intervalos de 5 ou 10 minutos, a proporção de previsões com $\bar{R}^2 < 1\%$ ou com $\bar{R}^2 > 10\%$ é semelhante: a saber, por volta de 36,5% e 12,5%, respectivamente. Já as estimativas médias dos coeficientes de regressão em (3) crescem em magnitude com a agregação temporal: -0.01, -0.03 e -0.05, respectivamente. Em linha com os resultados para a amostra toda, as estimativas são majoritariamente negativas, com valores positivos apenas em cerca de um terço das vezes.

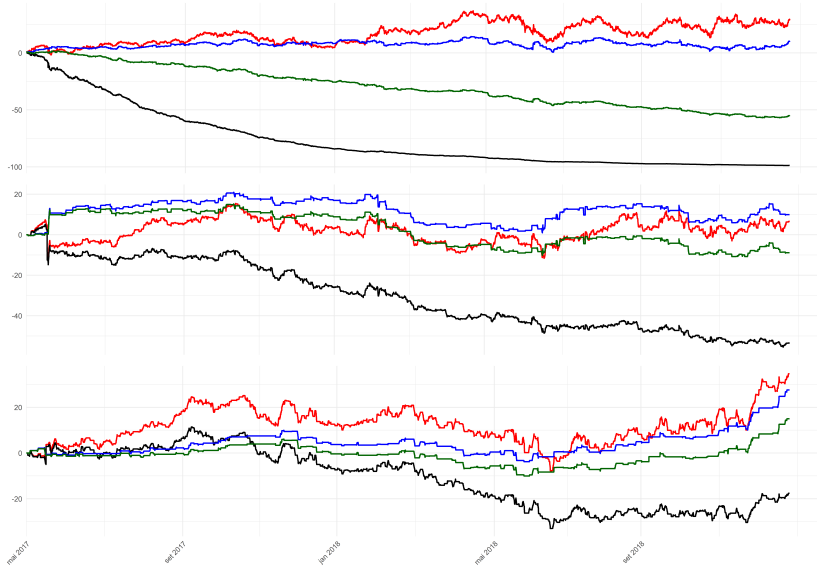
Figura 2
Ajuste fora da amostra das previsões por LASSO nas diferentes frequências



A primeira coluna refere-se aos valores do $\bar{R}^2$ fora da amostra das previsões por LASSO para os retornos de 1, 5 e 10 minutos, respectivamente. A segunda coluna estampa as estimativas do coeficiente de regressão em (3) usando uma janela móvel de 30 observações. Em todos os gráficos, as linhas preta e vermelha correspondem respectivamente às médias em cada dia e no período inteiro, enquanto que a área hachurada representa o intervalo interquartil em cada dia da amostra.

Finalmente, avaliamos a relevância econômica das previsões a partir de uma estratégia de negociação com base em nossas previsões. Para evitar viés de antecipação, estimamos a regressão (3) sem normalização em uma janela móvel de 30 observações. Utilizamos inicialmente uma estratégia simples

Figura 3
Evolução do capital investido na estratégia baseada nas previsões por LASSO



As linhas mostram o desempenho das estratégias ao longo do tempo, os retornos são acumulados entre maio de 2017 e dezembro de 2018 usando retornos de 1, 5 e 10 minutos, respectivamente. A linha vermelha representa os retornos acumulados da estratégia simples sem custo de transação. A linha preta ilustra os retornos acumulados da estratégia simples com custo de transação. A linha azul corresponde aos retornos acumulados da estratégia que controla pela previsibilidade ($\bar{R}^2 > 5\%$) na ausência de custos de transação. A linha verde ilustra os retornos acumulados dessa estratégia quando controlamos pelos custos de transação.

composta por dois critérios: (1) compra o índice quando a previsão é positiva e vende caso a previsão seja negativa e (2) tanto a compra quanto a venda só são efetuadas se b_L da equação (3) for maior que zero.

Talvez por conta de sua simplicidade, essa estratégia não resulta em ganhos líquidos consideráveis. A Figura 3 ilustra os retornos acumulados obtidos entre maio de 2017 e dezembro de 2018. As linhas vermelhas ilustram os ganhos da estratégia simples sem custo de transação: a saber, 29,73%, 6,64% e 34,67% para retornos de 1, 5 e 10 minutos, respectivamente. Considerando o custo mínimo de 0,011% para negociação de ETF na B3 (*day trade*), o resultado líquido passa a ser negativo para todas as frequências: mais especificamente, $-98{,}73\%$, $-53{,}43\%$ e $-17{,}61\%$ para retornos de 1, 5 e 10 minutos, respectivamente. Isso não surpreende já que sinais esparsos de curta duração implicam necessariamente alto *turnover*.

Uma laternativa é condicionar a estratégia a momentos em que o poder de

ajuste dentro da amostra seja suficientemente bom: por exemplo, $\bar{R}^2 > 5\%$ para entrar na posição. As linhas azuis na Figura 3 mostram que essa estratégia tem desempenho aceitável, retornando 10,44%, 9,89% e 27,57% nas frequências de 1, 5 e 10 minutos, respectivamente. Apesar do resultado negativo para retornos de 1 e 5 minutos quando consideramos custos de transação ($-4{,}95\%$ e $-8{,}91\%$, respectivamente), observamos um ganho líquido considerável de 14,99% para retornos de 10 minutos.

4. Conclusão

O procedimento LASSO é capaz de identificar a presença de sinais esparsos de vida curta entre as ações constituintes do Ibovespa. Além da significância estatística, demonstramos que nossos resultados são economicamente significativos, ao menos para a frequência de 10 minutos, a partir de uma estratégia de negociação direcional.

Nosso estudo possui naturalmente diversas limitações. Primeiro, o período amostral é relativamente curto, podendo refletir regimes específicos de volatilidade e liquidez. Segundo, adotamos uma regra de negociação propositadamente parcimoniosa. Em particular, não otimizamos o limiar de previsibilidade em termos de $\bar{R}^2$, nem execução ótima.

Essas limitações abrem frentes naturais para pesquisa futura.

Agradecimentos Fernandes agradece o apoio financeiro da FAPESP (2023/01728-0) e do CNPq (302153/2022-5).

Conflito de interesses Os autores declaram não possuir qualquer conflito de interesses.

Inteligência Artificial Os autores declaram que nenhuma ferramenta de Inteligência Artificial foi utilizada no desenvolvimento desta pesquisa, seja na elaboração do manuscrito, na redação do texto ou na implementação e execução dos códigos em R.

Disponibilidade de dados Todos os dados utilizados neste estudo estão disponíveis mediante solicitação ao autor correspondente.

Referências

Alexandre, H., Arismendi-Zambrano, J. e De Genaro, A. (2024). Intraday returns forecasting using machine learning: Evidence from the brazilian stock market. University of São Paulo, University College Dublin School of Business, and São Paulo School of Business Adminstration, FGV.

Bergmeir, C., Hyndman, R. J. e Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction, *Computational Statistics and Data Analysis*, *120*, 70–83.

Campbell, J. Y. e Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies*, *21*(4), 1509–1531.

Chinco, A., Clark-Joseph, A. D. e Ye, M. (2019). Sparse signals in the cross-section of returns, *Journal of Finance*, *74*(1), 449–492.

DeMiguel, V., Garlappi, L., Nogales, F. J. e Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science*, *55*(5), 798–812.

Feng, G., Giglio, S. e Xiu, D. (2020). Taming the factor zoo: A test of new factors, *Journal of Finance*, *75*(3), 1327–1370.

Freyberger, J., Neuhierl, A. e Weber, M. (2020). Dissecting characteristics nonparametrically, *Review of Financial Studies*, *33*(5), 2326–2377.

Gabaix, X. (2014). A sparsity-based model of bounded rationality, *Quarterly Journal of Economics*, *129*(4), 1661–1710.

Harvey, C. R., Liu, Y. e Zhu, H. (2015). . . . and the Cross-Section of Expected Returns, *Review of Financial Studies*, *29*(1), 5–68.

Kozak, S., Nagel, S. e Santosh, S. (2020). Shrinking the cross-section, *Journal of Financial Economics*, *135*(2), 271–292.

Linnainmaa, J. T. e Roberts, M. R. (2018). The history of the cross-section of stock returns, *Review of Financial Studies*, *31*(7), 2606–2649.

McLean, R. D. e Pontiff, J. (2016). Does academic research destroy stock return predictability?, *Journal of Finance*, *71*(1), 5–32.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B*, *58*(1), 267–288.