

IND2604 - Projeto Final

Investigando o impacto de notícias no mercado de ações brasileiro: uma análise dos *google trends*

Matheus Nogueira (2412716)

PUC-Rio

matnogueira@gmail.com

RESUMO

O objetivo deste trabalho é estudar se notícias ou o interesse popular em assuntos relacionados à economia e finanças podem impactar o mercado financeiro brasileiro. Para isso, foram escolhidos os *google trends* como medidas quantitativas de interesse popular, uma vez que eles representam o volume de buscas de determinados termos na plataforma de pesquisa *Google*. A partir dos *trends* coletados, foi implementado um modelo linear para avaliar se esse interesse, aproximado pelos *trends*, impacta, ou ajuda a explicar, o mercado de capitais brasileiro, representado pelo índice Ibovespa. É possível concluir, com base nessa metodologia, que os *google trends* possuem, de fato, capacidade de explicar o índice Ibovespa e, consequentemente, que notícias ou interesse popular impactam o mercado de ações do país.

PALAVRAS CHAVE. Google Trends. Séries Financeiras. Regressão Linear.

1. Introdução

É de conhecimento geral que notícias podem impactar o preço de ativos nas bolsas de valores de diferentes países e que, somado a isso, a percepção do público sobre um determinado ativo ou cenário macroeconômico é uma variável importante para entender, ou até prever, o desempenho do mercado de capitais.

Dentro desse cenário, diversos trabalhos já exploraram os *Google Trends*, que serão abreviados como *gt*, como medida quantitativa da percepção e interesse da população sobre um determinado tema para avaliar o impacto de notícias nos preços/retornos de ativos financeiros listados em bolsa de valores.

Xu e Berkely [2014], por exemplo, criaram um índice de *gt* junto de informações disponíveis no *Yahoo Finance* para classificar notícias como "boas", "ambíguas" e "ruins", com o intuito de modelar a persistência do impacto dessas notícias na variação do preço semanal de ações. Preis et al. [2013], por outro lado, implementaram uma estratégia de compra e venda de um ativo financeiro da bolsa americana baseado na variação do *trends* de diferentes termos relacionados à finanças e economia, obtendo resultados financeiros significativos. Wilcoxson et al. [2020] demonstraram que modelos simplificados baseados em *gt*, utilizando técnicas de redução de dimensionalidade como AIC, BIC, lasso e ridge, superam previsões feitas por modelos tradicionais e modelos de passeio aleatório para taxas de câmbio, melhorando a acurácia das previsões.

Muitos trabalhos utilizam o *gt* para o mercado de criptomoedas. Aslanidis et al. [2022], por exemplo, construíram um índice de *gt* para criptomoedas por meio de entropia de transferência, concluindo que existe um fluxo bidirecional de informação entre os *gt* e esse mercado. O trabalho de Kristoufek [2013], por sua vez, investiga a relação entre o preço do Bitcoin, *gt* e buscas na *Wikipedia*, encontrando uma forte correlação bidirecional e assimétrica entre o interesse online e o preço do criptoativo. Kaneko [2021] também investigou a correlação entre *gt* e flutuações de preços de criptoativos, concluindo que, na média, os *trends* possuem impacto negativo, porém estatisticamente não significantes no preços de *altcoins*.

Como exemplo de que o interesse em modelar o impacto de notícias nos preços de ativos financeiros é anterior ao *gt*, Aggarwal e Schirm [1998] documentaram a existência de correlação assimétrica entre preços de câmbio e *equity* e a escala das notícias sobre balanço comercial de países e empresas.

Uma vez estabelecida a relevância do estudo do impacto de notícias e do interesse popular em temas correlatos à finanças e economia, quantificados por meio dos *google trends*, o presente trabalho propõe-se a investigar a existência de uma correlação entre esses dados e as variações diárias do valor de fechamento do índice Ibovespa. Para isso, este artigo encontra-se dividido em 3 principais seções. Primeiramente, a seção 2 apresenta todos os dados utilizados (seção 2.1), bem como descreve sobre a análise exploratória realizada (2.2) e o modelo escolhido para estudar a relação entre *gt* e o mercado brasileiro (2.3). Em seguida, a seção 3 exibe e comenta os resultados obtidos a partir da metodologia proposta. Por fim, a seção 4 traz as conclusões deste projeto.

2. Metodologia

2.1. Dados utilizados

Os dados a serem utilizados neste projeto final consistem em dois tipos:

- Séries financeiras históricas
- Séries históricas do Google Trends

As séries financeiras consistem, primeiramente, da série histórica de fechamento ajustado diário do índice IBOVESPA, reconhecido pelo *ticker* "BVSP". Essa série foi retirada do site [Yahoo Finance](#).

Além dessa série, também serão utilizadas séries de variáveis macroeconômicas, a saber, SELIC e CDI diários, obtidos por meio de um pacote em *Python* [python-bcb](#) que disponibiliza APIs para o site do [Banco Central do Brasil](#).

Os principais dados deste projeto serão os *Google Trends* históricos diários de 10 termos relacionados com economia e investimentos. Essa métrica representa o volume de buscas de determinados termos na plataforma de pesquisa *Google*. Os 10 termos são: 'Ibovespa', 'Bolsa de Valores', 'Ações', 'Dividendos', 'Renda Fixa', 'Inflação', 'CDI', 'Dolar', 'Bitcoin' e 'Renda Variável'. Todos os *trends* foram obtidos a partir de um pacote *Python* [pytrends](#) que implementa APIs de acesso ao site [Google Trends](#).

Como estratégia para selecionar os *keywords*, primeiramente foi utilizada o próprio *google trends*, que possui uma funcionalidade de indicar pesquisas relacionadas a um *keyword* fornecido. No entanto, observei que essa estratégia gerava uma lista de termos muito similares. Por exemplo, ao fornecer o termo "Ibovespa" para essa funcionalidade, as 10 principais pesquisas relacionadas fornecidas foram "Ibovespa hoje", "dolar", "dolar hoje", "cotação ibovespa", "dólar", "ibovespa dólar hoje", "ibovespa ações", "dólar hoje", "petr4" e "bitcoin". Note que muitos termos são, praticamente, repetições do termo original. Por isso, para tentar gerar uma maior abrangência dos termos utilizados, a escolha foi baseada tanto nos *keywords* utilizados em alguns dos artigos revisados na seção 1, como os trabalhos de Wilcoxson et al. [2020], Aslanidis et al. [2022] e Preis et al. [2013], quanto que arbitrariamente escolhidos com o intuito de representar diferentes componentes do mercado nacional, de tal modo que fosse possível capturar de forma generalizada o interesse popular em relação à economia e finanças.

É importante ressaltar alguns fatos sobre a natureza do *Google Trends*. Primeiramente, os valores de *trends* não são absolutos, mas sim relativos ao período de busca selecionado e sempre contidos no intervalo de 0 a 100. Isso quer dizer que, para um mesmo termo em uma mesma data, o valor do *trend* pode ser diferente dependendo do período de busca selecionado. Além disso, não é possível especificar, a priori, a granularidade dos *trends*, isto é, a plataforma não permite que o usuário escolha por *trends* diários, semanais ou mensais. Essa granularidade é definida a partir do período de busca, por exemplo: buscas de intervalo de 1 mês retornam *trends* diários e buscas de intervalo de 1 ano geram *trends* semanais. Essas duas características dos *google trends* são um desafio tanto para a obtenção dos dados quanto da interpretação de seus resultados. Por fim, a plataforma do *Google Trends* permite especificar a região de busca dos termos. Isso foi utilizado para garantir que as buscas dos 10 termos listados acima fosse restrita ao Brasil, visto que este projeto se limita a estudar o mercado de capitais brasileiro.

Todos os dados históricos foram obtidos para o intervalo de 01-01-2010 até 01-01-2024, ignorando feriados e finais de semana, isto é, contabilizando apenas os dias úteis.

Uma estratégia comum ao modelar dados financeiros como preço de ativos é utilizar, ao invés do preço em si, a variação do preço de um dia para o outro. Isto é, se y_t é a série temporal do fechamento diário do Ibovespa no tempo t , é recomendado utilizar uma série $r_t = \Delta y_t = y_t - y_{t-1}$. Neste trabalho, no entanto, será utilizada apenas a série de fechamento diário, com o objetivo de verificar, estatisticamente, se o valor do *google trends*, usado como *proxy* do interesse popular sobre o assunto, possui poder explicativo sobre essa variável.

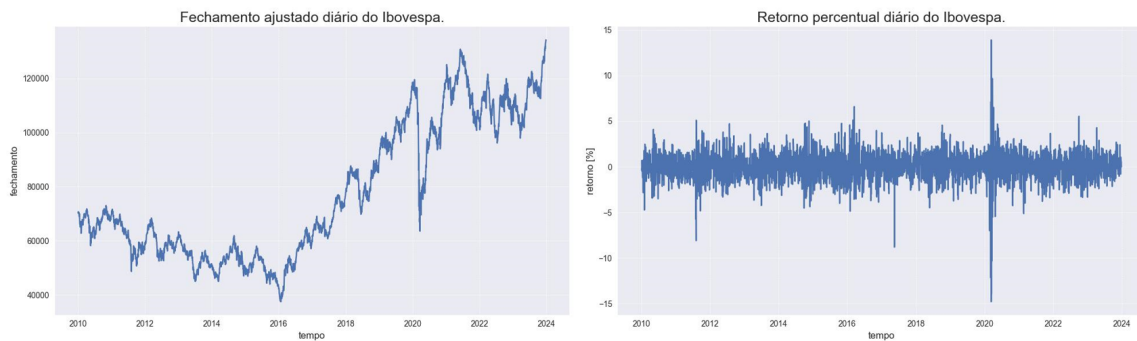


Figura 1: Gráfico dos preços de fechamento diário e do retorno percentual diário do Ibovespa.

2.2. Análise Exploratória

A primeira etapa da metodologia do trabalho é explorar, visual e numericamente, os dados coletados. A Figura 1 exibe os gráficos temporais do fechamento ajustado diário e do retorno diário do Ibovespa. Pode-se perceber que, visualmente, o mercado brasileiro segue uma tendência de alta desde 2016, exceto o grande vale no início de 2020 causado pela pandemia de COVID-19. Note, inclusive, os picos observados na série de retorno do índice nessa época, muito superiores, em módulo, a todas as variações observadas desde 2010.

Como o interesse é modelar a influência das variáveis explicativas no fechamento diário do Ibovespa, é natural analisar os gráficos de dispersão de cada regressor com a variável resposta. A Figura 2 exibe esses gráficos para o primeiro grupo de variáveis explicativas: variação percentual da SELIC, variação percentual do CDI e o valor defasado de 1 dia do próprio índice Ibovespa. Os títulos de cada figura indicam o coeficiente de correlação linear de pearson entre as variáveis. Dado que o CDI e a SELIC são duas variáveis de comportamento extremamente similar, é de se esperar que a correlação entre ambas as variáveis e o índice sejam similares. Também é esperada a correlação negativa, uma vez que SELIC e CDI representam taxas básicas de juros e, usualmente, em situação de aumento de juros, os investidores tendem a deixar de priorizar a renda variável, o que pode gerar variações negativas no índice Ibovespa.

A Figura 2 também exibe um outro resultado conhecido: o valor defasado de 1 dia do índice apresenta correlação quase perfeita com o seu valor corrente. Não obstante, é comum modelar o preço de ativos financeiros, p_t , a partir de um passeio aleatório, isto é:

$$p_t = p_{t-1} + \epsilon_t \quad \epsilon_t \sim N(0, \sigma)$$

Esse modelo advém da hipótese de eficiência do mercado, que diz que o preço já comporta toda a informação relevante sobre o ativo de tal modo que é impossível realizar previsões sobre o futuro de seu comportamento (Malkiel [1989]).

É essencial esclarecer que, para garantir a coerência temporal entre cada regressor e a variável resposta, todos os regressores foram adicionados com um dia de defasagem, de tal modo que estamos sempre modelando o valor da variável explicativa de ontem para explicar o valor da variável dependente de hoje.

O próximo grupo de variáveis a serem exploradas são os *google trends*. A Figura 3 exibe os gráficos de dispersão de cada um dos 10 termos escolhidos, defasados de 1 dia, com o valor do Ibovespa. Novamente, o título das figuras exibe a correlação de pearson. Note que, a priori, existe uma forte correlação positiva entre o valor do *trend* "Ibovespa" do dia anterior com o

Gráficos de dispersão de $\Delta BVSP$ X variáveis econômicas

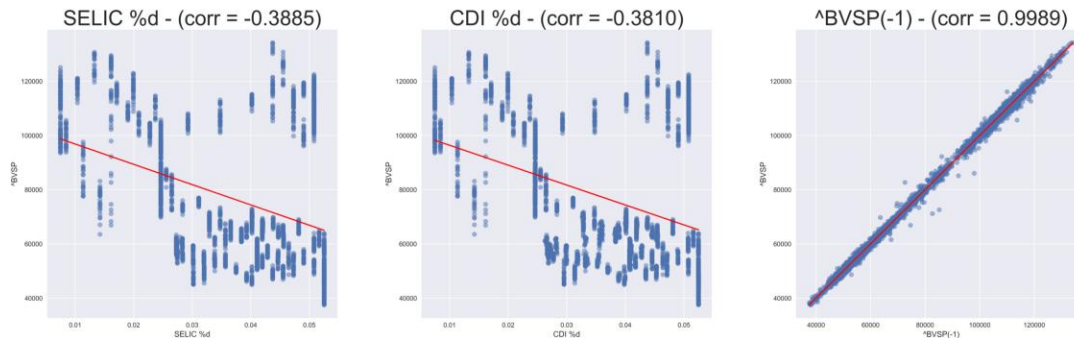


Figura 2: Gráfico de dispersão entre o fechamento diário do Ibovespa e as variáveis econômicas: variação do CDI, variação da SELIC e fechamento defasado em 1 dia do Ibovespa.

valor do índice Ibovespa no dia corrente. Dentre todos os termos escolhidos, este é o que apresenta maior correlação, 0.81. Isso pode indicar que, quanto maior o interesse no termo "Ibovespa", mais pessoas estão positivamente interessadas no mercado brasileiro e, consequentemente, maior o valor do índice do mercado. Além desse, os termos "Dividendos", "Bolsa de Valores", "Bitcoin" e "Inflação" são os que apresentam coeficiente superior, em módulo, a 0.3. Interessante notar a correlação negativa entre o interesse em inflação e o índice. Poderia-se supor, por exemplo, que o maior interesse no termo "inflação" ocorra em cenários de altas desse indicador e que, nesses cenários, é compreensível uma piora do mercado brasileiro, dada a diminuição do poder de compra da população e, consequentemente, a diminuição dos investimentos na bolsa. No entanto, esta é apenas uma suposição cuja comprovação foge do escopo deste trabalho.

Uma vez realiza a breve análise exploratória da relação entre cada variável explicativa e o valor de fechamento do índice Ibovespa, o próximo passo é apresentar a modelagem utilizada para estudar se notícias e interesse popular, aproximados pelos *google trends*, são capazes de explicar o valor de fechamento do índice e, consequentemente, o andamento do mercado de ações brasileiro.

2.3. Modelagem

O modelo de regressão linear múltipla foi escolhido para este trabalho por dois motivos. Primeiramente, por ser uma técnica consolidada, de fácil implementação e interpretação, trazendo excelentes funcionalidades de inferência sobre os seus parâmetros estimados e, com isso, trazendo *insights* úteis sobre a situação modelada. Somado a isso, as Figuras 2 e 3 revelaram que existem relações lineares entre as variáveis explicativas e a variável dependente.

Sendo assim, o modelo implementado pode ser escrito como:

$$y_t = \beta_0 + \beta_1 \cdot x_{t-1}^{(S)} + \beta_2 \cdot x_{t-1}^{(C)} + \beta_3 \cdot y_{t-1} + \sum_{i=4}^{14} \beta_i \cdot x_{i,t-1}^{(g)} + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

onde $\beta_i \forall i \in \{0, \dots, 14\}$ são os coeficiente fixos e desconhecidos a serem estimados, y_t o valor corrente e y_{t-1} o valor defasado do fechamento diário do Ibovespa, $x_{t-1}^{(S)}$ é o valor defasado da variação da SELIC, $x_{t-1}^{(C)}$ o valor defasado da variação do CDI, $x_{i,t-1}^{(g)} \forall i \in \{4, \dots, 14\}$ os valo-

Gráficos de dispersão de $\Delta BVSP$ X Google Trends

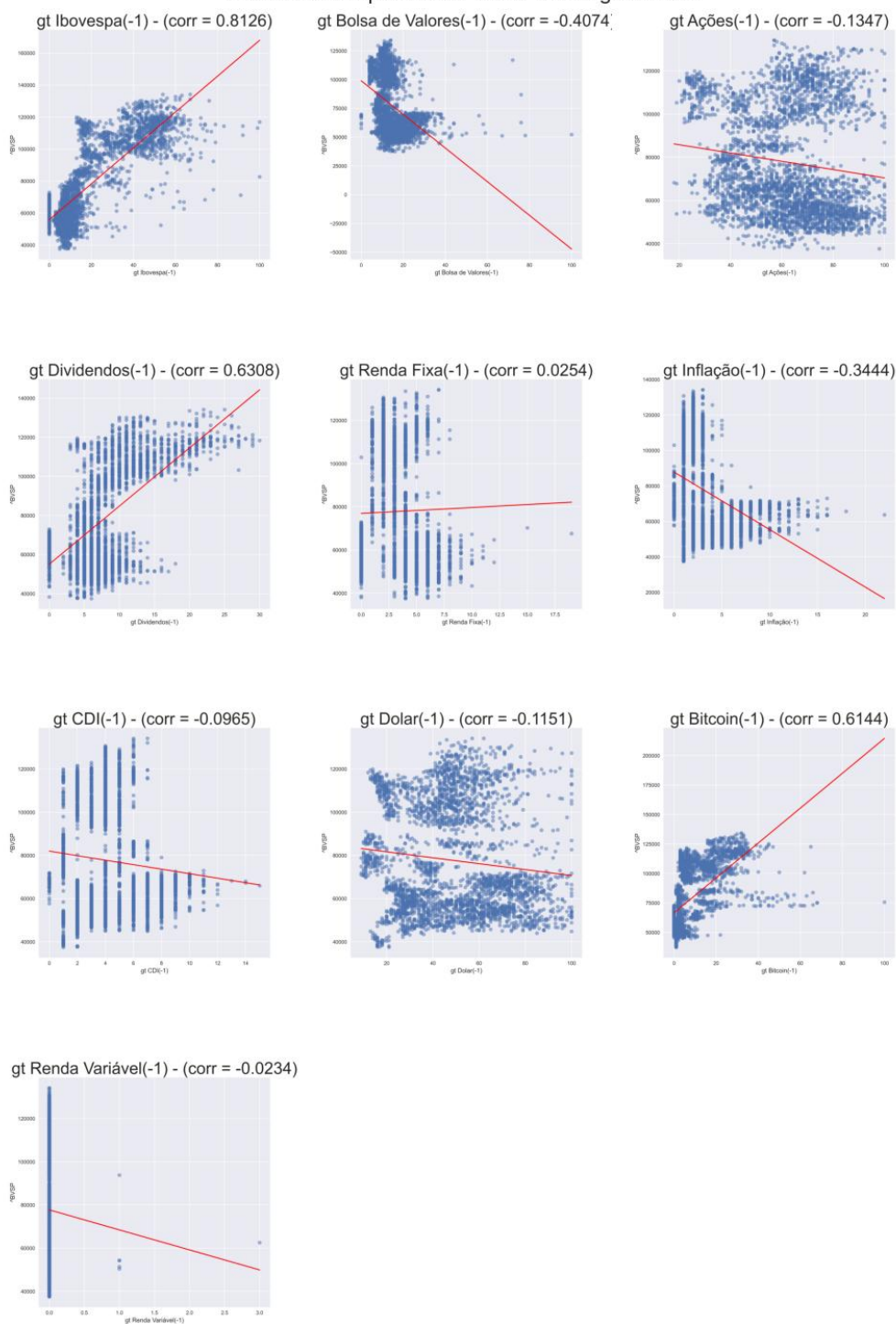


Figura 3: Gráfico de dispersão entre o fechamento diário do Ibovespa e os *google trends*.

res defasados dos 10 *google trends* e ϵ_t o termo de erro aleatório que, por hipótese, seguem uma distribuição normal padrão e IID.

Uma vez montada a regressão, temos um modelo com 14 coeficientes a serem estimados e 2468 observações.

Para avaliar se cada *google trend* possui poder explicativo sobre o fechamento diário do Ibovespa, serão realizadas 2 principais inferências. Primeiramente, testes t para averiguar se cada coeficiente estimado é estatisticamente diferente de zero. Além disso, também será realizado um teste F para estudar se, conjuntamente, todos os coeficientes associados a *google trends* são diferentes de zero, dados os 3 regressores econômicos. Esse segundo teste é de suma importância pois pode averiguar se, conjuntamente, os *google trends* são capazes de adicionar poder explicativo sobre o índice Ibovespa dada a informação, por exemplo, do próprio índice defasado. Somado a isso, também serão realizados os diagnósticos típicos de resíduos, a saber testes de normalidade, ausência de autocorrelação e homocedasticidade, uma vez que um modelo bem especificado deve exibir resíduos sem autocorrelação significativa, normais e homocedásticos (Hyndman e Athanasopoulos [2021]). Apenas com a garantia da boa estimação do modelo é que seremos capazes de realizar inferência sobre os parâmetros estimados.

Para testar normalidade, será utilizado o teste de Jarque-Bera, cuja hipótese nula é de normalidade dos resíduos. Para testar autocorrelação, utilizarei o teste de Ljung-Box, cuja hipótese nula é de resíduos descorrelatados. Por fim, para testar homocedasticidade, será utilizado o teste de Breusch-Pagan, cuja hipótese nula é de homocedasticidade.

Para compor esses diagnósticos, também serão exibidos o gráfico, qqplot e gráfico de autocorrelação dos resíduos padronizados.

Importante ressaltar que todos os diagnósticos serão realizados sobre os resíduos padronizados. Isto é, se $\hat{\epsilon}_t$ são os resíduos do modelo $f(\cdot)$, os resíduos padronizados são:

$$\hat{\epsilon}_t^* = \frac{\hat{\epsilon}_t}{\sqrt{SE}}$$

onde SE é o erro padrão do modelo.

3. Resultados

Os primeiros resultados a serem comentados estão na Tabela 1. Nela estão todos os valores estimados dos coeficientes da regressão, bem como suas estatísticas t , p-valores e intervalo de confiança. Ao final dela, também estão descritos os p-valores dos testes de hipótese dos diagnósticos dos resíduos e do teste F para todos os coeficientes do modelo.

Ao avaliar os p-valores dos testes t para os coeficientes associados aos *google trends*, bem como seus intervalos de confiança, podemos perceber que apenas 2 seriam estatisticamente significantes a nível de 5%, o "Ibovespa" e "Ações". O termo "Dividendos" também apresenta um p-valor baixo o suficiente para concluirmos a favor de sua significância. Analisando esses resultados, alguns coeficientes parecem ser estatisticamente indistinguíveis de zero, como o "Bolsa de Valores", "Renda Fixa", "Dolar" e "Renda Variável". Todos esses coeficientes possuem p-valor superior a 50%. Importante notar que as 3 variáveis econômicas escolhidas possuem p-valor abaixo de 3%, baixo o suficiente para indicar que elas são significantes para o modelo. Isso é um resultado importante pois essas variáveis foram escolhidas, justamente, para garantir um efeito *ceteri paribus* ao realizar a inferência dos coeficientes dos *google trends*.

Também analisando a Tabela 1, notamos que o p-valor do teste F para todos os regressores (ANOVA para regressão) é zero, o que indica que, conjuntamente, as variáveis escolhidas são superiores ao modelo restrito apenas com o intercepto.

	coef	std err	t	p-value	LB	UB
const	837.403	207.292	4.040	0.000	430.976	1243.830
SELIC %d	-353173.793	156498.244	-2.257	0.024	-660012.239	-46335.348
CDI %d	351063.637	155795.322	2.253	0.024	45603.377	656523.896
^BVSP(-1)	0.990	0.002	507.200	0.000	0.986	0.994
gt Ibovespa(-1)	8.304	2.545	3.263	0.001	3.314	13.294
gt Bolsa de Valores(-1)	0.244	4.321	0.056	0.955	-8.229	8.716
gt Ações(-1)	-4.547	1.642	-2.769	0.006	-7.767	-1.327
gt Dividendos(-1)	13.484	7.160	1.883	0.060	-0.555	27.523
gt Renda Fixa(-1)	-1.453	10.132	-0.143	0.886	-21.319	18.412
gt Inflação(-1)	-12.279	11.726	-1.047	0.295	-35.270	10.711
gt CDI(-1)	18.690	14.080	1.327	0.184	-8.915	46.296
gt Dolar(-1)	0.970	1.596	0.608	0.543	-2.159	4.099
gt Bitcoin(-1)	2.603	2.733	0.953	0.341	-2.755	7.960
gt Renda Variável(-1)	188.605	316.411	0.596	0.551	-431.765	808.976
Teste F	-	-	-	0.000	-	-
Jarque-Bera	-	-	-	0.0000	-	-
Ljung-Box	-	-	-	0.0000	-	-
Breusch-Pagan	-	-	-	0.0000	-	-

Tabela 1: Parâmetros estimados e testes de hipótese para o modelo de regressão linear múltipla.

Por outro lado, os p-valores dos testes de hipótese para diagnóstico dos resíduos indicam que os resíduos do modelo são não normais, heterocedásticos e correlatados. Embora esses resultados sejam indesejados, há a necessidade de cautela ao concluir, apenas com base nesses testes, que o modelo não está bem especificado, visto que o tamanho da amostra pode facilitar a rejeição da hipótese nula desses testes. Para complementar esses diagnósticos, veja as Figuras 4, 5 e 6, que exibem, respectivamente, o gráfico, qqplot e função de autocorrelação dos resíduos padronizados.

Podemos notar, pela Figura 4, que existem alguns *outliers* significativos, isto é, observações que o modelo não conseguiu capturar. Esses *outliers* estão relacionados com o período da pandemia de COVID-19, sendo, portanto, justificáveis. É possível, entretanto, que a simples presença destes outliers estejam impactando fortemente os p-valores dos testes de hipótese da Tabela 1. A Figura 5, por exemplo, mostra que esses *outliers* podem ser a causa da não normalidade dos resíduos padronizados. Além disso, a Figura 6 nos revela que, embora o teste de Ljung-Box tenha rejeitado a hipótese de ausência de autocorrelação, não existe nenhuma correlação de curto prazo relevante. Mesmo o *lag* de 7 dias, cuja FAC extrapola o intervalo de confiança, apresenta autocorrelação de 0.1, valor pequeno o suficiente para classificá-lo como irrelevante. Sendo assim, considero seguro concluir que o modelo foi bem especificado e, com isso, considero confiáveis as inferências realizadas sobre os parâmetros estimados.

O último resultado de interesse é o teste F para avaliar a hipótese de que todos os coeficientes associados aos *google trends* são diferentes de zero. Isto é:

$$\begin{cases} H_0 : & \beta_4 = \beta_5 = \dots = \beta_{14} = 0 \\ H_1 : & c.c. \end{cases}$$

Este teste é importante para verificar se, conjuntamente, a adição dos *google trends* au-

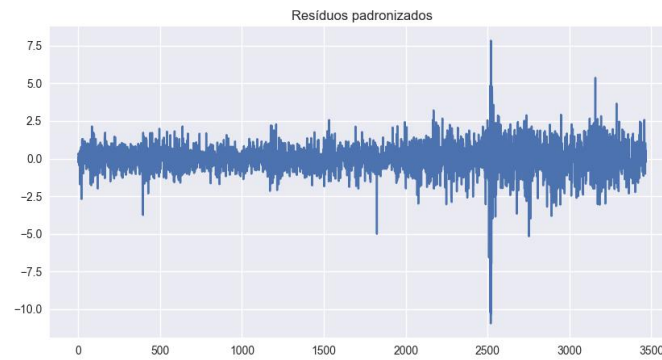


Figura 4: Gráfico dos resíduos padronizados.

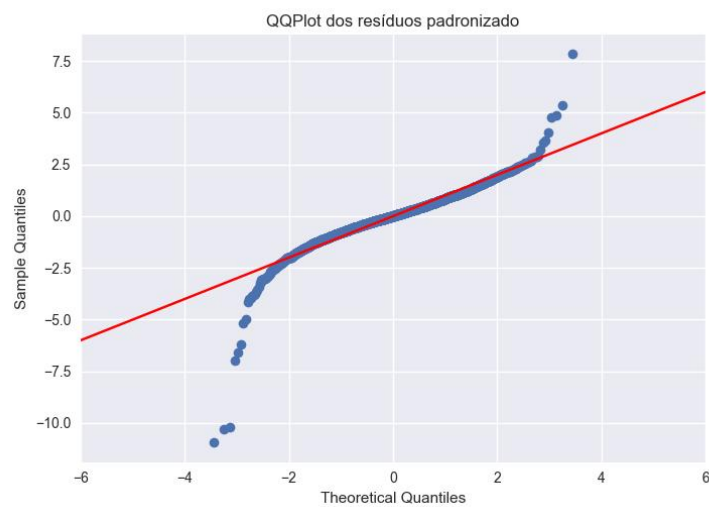


Figura 5: QQPlot dos resíduos padronizados.

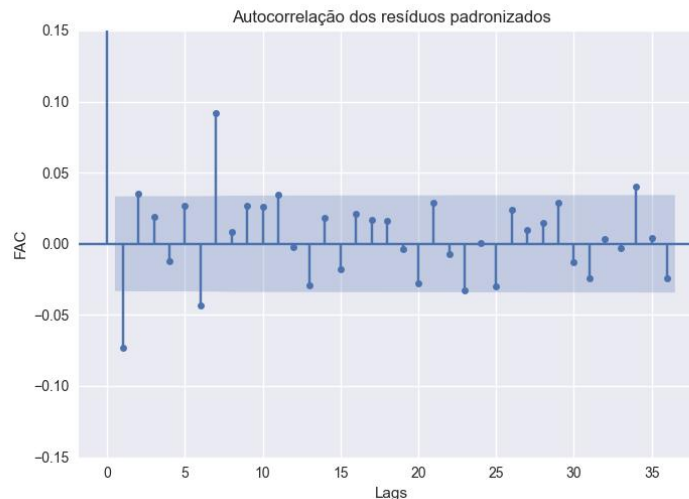


Figura 6: Gráfico da função de autocorrelação dos resíduos padronizados.

menta o grau de explicabilidade da variável dependente, dado que o modelo já tem acesso ao intercepto (β_0) e às variáveis econômicas (β_1, \dots, β_3), em especial ao próprio índice defasado que, como já apresentado, possui correlação fortíssima com o valor corrente. A Tabela 2 exibe o resultado do teste. Pode-se perceber que o p-valor do teste está próximo de 1%, um valor baixo o suficiente, ao meu ver, para rejeitar a hipótese nula e, consequentemente, concluir que a adição dos *google trends* no modelo auxiliam na explicação do fechamento diário do Ibovespa mesmo com a presença das 3 variáveis econômicas.

F	p-value
2.09276	0.01189

Tabela 2: Teste F para coeficientes associados aos *google trends*.

4. Conclusão

O objetivo deste trabalho é de estudar se notícias e interesse popular em relação à temas correlatos à economia e finanças podem impactar, ou explicar, o andamento do mercado de capitais brasileiro. Como medida quantitativa do interesse popular, foram utilizados os *google trends* diários e como aproximação do comportamento do mercado, os valores diários de fechamento do índice Ibovespa.

Uma vez analisados os resultados da metodologia proposta, baseada no modelo de regressão linear múltipla, foi averiguado, tanto via testes *t* para os coeficientes individuais de cada *google trend*, quando via testes *F* para avaliá-los conjuntamente, que essas métricas são estatisticamente significantes para explicar a variável dependente escolhida.

Sendo assim, é possível concluir que o interesse popular, medido através do volume de buscas de 10 termos na plataforma *Google*, é uma variável útil para explicar o comportamento do mercado de capitais brasileiro, representado pelo principal índice da bolsa nacional, o Ibovespa. É importante, no entanto, deixar claro que essa conclusão não se configura como recomendação de investimento no mercado brasileiro.

Seria possível estender este trabalho de algumas maneiras. Primeiro, refinar a obtenção dos *google trends* a fim de diminuir a relação do valor da métrica com o período de busca. Uma possível estratégia seria implementar a busca apresentada por Preis et al. [2013]. Ademais, também seria interessante expandir a análise para além da regressão linear, buscando, por exemplo, efeitos marginais dos *google trends* ou implementando metodologias que permitam não linearidades, como métricas de importância de variáveis associadas a modelos de árvore de decisão e *random forests*. Por fim, também seria possível investigar se os *google trends* permanecem significantes na presença de novas variáveis de controle, como índices de inflação, câmbio diário, dentre outras.

Referências

- Aggarwal, R. e Schirm, D. C. (1998). Asymmetric impact of trade balance news on asset prices. *Journal of International Financial Markets, Institutions and Money*, 8(1):83–100.
- Aslanidis, N., Bariviera, A. F., e López, Ó. G. (2022). The link between cryptocurrencies and google trends attention. *Finance Research Letters*, 47:102654.
- Hyndman, R. e Athanasopoulos, G. (2021). *Forecasting: principles and practice*.
- Kaneko, Y. (2021). A time-series analysis of how google trends searches affect cryptocurrency prices for decentralized finance and non-fungible tokens. In *2021 International Conference on Data Mining Workshops (ICDMW)*, p. 222–227. IEEE.
- Kristoufek, L. (2013). Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3(1):3415.
- Malkiel, B. G. (1989). Efficient market hypothesis. In *Finance*, p. 127–134. Springer.
- Preis, T., Moat, H. S., e Stanley, H. E. (2013). Quantifying trading behavior in financial markets using google trends. *Scientific reports*, 3(1):1–6.
- Wilcoxson, J., Follett, L., e Severe, S. (2020). Forecasting foreign exchange markets using google trends: Prediction performance of competing models. *Journal of Behavioral Finance*, 21(4): 412–422.
- Xu, S. Y. e Berkely, C. (2014). Stock price forecasting using information from yahoo finance and google trend. *UC Brekley*.