

Análise da flutuação da popularidade de ações e suas relações com volume de negociações

Augusto Castello Branco Bastos

Mat 2018057582

03/11/2020

Comentado [C2]: Por capa na forma de artigo. Ver a formatação sugerida na nossa revista do PPGC – UFMG por exemplo: <https://revistas.face.ufmg.br/index.php/contabilidadevistaerevista/about/submissions>

1) Introdução

O advento da internet e a consolidação como principal meio de comunicação do século XXI tornou possível o aumento da velocidade e disponibilidade de informações. A quantidade de dados disponíveis tornou o processamento de informações o maior problema na comunicação. Dessa forma os mecanismos de pesquisa tornaram-se cada vez mais necessários para obter o conteúdo relevante encontrado de forma rápida e acessível. O Google é o principal buscador da internet possuindo mais de 90% do mercado¹ e disponibiliza o volume de buscas de cada palavra chave através da ferramenta Google Trends.

O mercado financeiro sempre foi extremamente dependente do fluxo de informações global para a confecção de análises, projeções e obtenção de demonstrações financeiras por exemplo. O aumento da velocidade da disponibilização dos dados reduziu o intervalo entre a divulgação da informação e o início da análise pelo investidor, porém passou a ser necessária uma busca pela informação e pelos dados necessários na internet.

O objetivo desse estudo é analisar o indicador do Google Search Volume (GSV) das ações e as suas relações com os movimentos no mercado financeiro, traçando relações entre as variações de buscas mensais e a identificação .

2) Problema de Pesquisa e Sua Importância

A teoria dos mercados eficientes de Fama (1970) afirma que o preço das ações é corrigido no momento em que o fato relevante é gerado. Essa teoria não leva em conta o intervalo de tempo entre a geração da informação, no caso, a publicação na internet, busca e interpretação do investidor para tomada de decisão.

A pesquisa é baseada na teoria de que os mercados não são eficientes, uma vez que, do momento da divulgação da notícia até a reação do investidor existe um intervalo de tempo necessário para análise da informação e de sua influência no mercado. Existem estudos tratando da influência direta de buscas pelos Tickers das ações e sua influência no curto prazo (ver, e.g., Bijl et al., 2016; Challet & Ayed, 2013; Kristoufek, 2013; Mayer, 2018).

O estudo, porém, busca desenvolver a análise de buscas de dados que traçam a estratégia de médio prazo dos investidores qualificados, uma vez que essas teriam um menor volume em buscas mais complexas e maior necessidade de informações. A partir disso seria possível a partir de uma análise das estatísticas mensais dos tickers no GSV prever o comportamento das ações?

3) Objetivos

O objetivo do trabalho foi buscar as séries temporais mensais dos tickers de ações negociadas no Índice Ibovespa e comparar as flutuações dessas séries com as flutuações de volume e preços mensais das ações na bolsa, buscando encontrar relações que permitam a projeção de volume e retorno dessas séries.

A partir desse objetivo geral foram traçados dois objetivos específicos:

Analisar as séries utilizando modelos VAR para analisar a relação entre os dados de tickers das ações no âmbito de busca de notícias e pesquisas para modelar o comportamento do volume de negociações das ações respectivas.

Analisar o comportamento das ações utilizando modelos VEC para mudanças de tendências mensais nos retornos dos ativos respectivos.

Para análise da relevância de movimentação de informações foram utilizadas as séries de pesquisas dos 77 tickers que compõem o índice ibovespa no terceiro trimestre de 2020 separadas em períodos mensais. As séries obtidas são da categoria geral de

pesquisas na web e na categoria de pesquisas específicas de notícias sobre esses tickers além de regionalizadas entre pesquisas no Brasil e no mundo para cada uma dessas séries buscando analisar diferenças entre os comportamentos e seus diferentes impactos no volume e retornos projetados.

Comentado [C3]: Quais os objetivos específicos?

4) Revisão de Literatura

A análise da relação do GSV com o mercado financeiro foi inicialmente feita com o objetivo de analisar a relação comportamental da população buscando uma relação macroeconômica entre o sentimento geral da população e os movimentos da economia. Preis (2013) apresentaram resultados relevantes para a previsão de crises no mercado, conforme o perfil de buscas da população se tornava mais conservador. Esse estudo inicial apontou que a população muda seus padrões de busca na internet segundo a situação da economia e um portfólio simulado apresentou um retorno incremental de 326% em relação à média de carteiras aleatórias.

Os principais grupos de investidores atuando na bolsa de valores hoje são investidores individuais 22%, investidores estrangeiros 48% e investidores institucionais 25%; B3, (2020). É possível afirmar que a análise de relação direta do indicador GSV seria majoritariamente relevante para o grupo de investidores individuais conforme verificado por (Bijl et al., 2016) e Yoshinaga (2019).

Também segundo as pesquisas os resultados encontrados apontam que a proxy do aumento do GSV tem relação direta com o aumento do volume de negociações e também da redução do retorno indicando que o aumento da popularidade de um ativo traz a perda de aderência com seu valor fundamental conforme proposto por De Long (1990) ao caracterizar os noise traders.

Não foram encontrados estudos sobre a análise da relação entre as séries GSV de tickers utilizando uma estratégia de Long-Short para o mercado brasileiro.

Comentado [C4]: A revisão de literatura tem que ser mais ampla e embasada em mais artigos.

4) Metodologia

4.1 Dado obtidos

Os dados do GSV foram obtidos a partir da função gtrends do software R. as séries são obtidas a partir da pesquisa de uma palavra chave em um intervalo de tempo predeterminado. É possível especificar a categoria que os dados se referem, sendo possível buscar em todas as categorias ou especificar entre;

- "web" – pesquisas gerais no buscador tendo como resultado sites em que a palavra pode ser encontrada
- "news" – buscas de notícias em sites categorizados como sites de notícias
- "images" – buscas de imagens relacionadas ao conteúdo
- "youtube" – vídeos no youtube relacionados as palavras chaves

O estudo em questão utilizou as séries de todas juntas as categorias e as séries de notícias com o objetivo de avaliar se os efeitos seriam diferentes em cada caso.

Os dados obtidos são padronizados em um índice de 100 pontos, sendo esse o valor para o pico de buscas e todos os outros como uma fração aproximada em valor inteiro. Esse fator pode omitir variações menores entre os índices e prejudicar a análise de variação em períodos com menor volume total de buscas.

Os dados de cotação das ações foram obtidos a partir do site yahoo finance com um comando de importação BatchGetSymbols. Os dados foram agrupados em séries mensais de médias e de desvios padrão intramês do volume negociado para avaliar se

as variações de buscas estavam relacionadas ao aumento do volume ou de sua variação.

4.2 Análise de Estacionaridade das séries

As séries obtidas foram tiveram sua estacionaridade testada pelo teste de Dick Fuller. Os resultados obtidos apontaram que a maior parte das séries tanto do GSV quanto das médias de volume e de preço não eram estacionárias. Os resultados dos testes podem ser encontrados no anexo adf1.txt.

O resultado já era esperado e as séries foram então transformadas em séries de primeiras diferenças para obtenção das variações inermensais. Essa transformação foi realizada com o objetivo de capturar as variações entre os indicadores e dessa forma obter análises dos efeitos de choques no comportamento do volume sobre as mudanças no GSV.

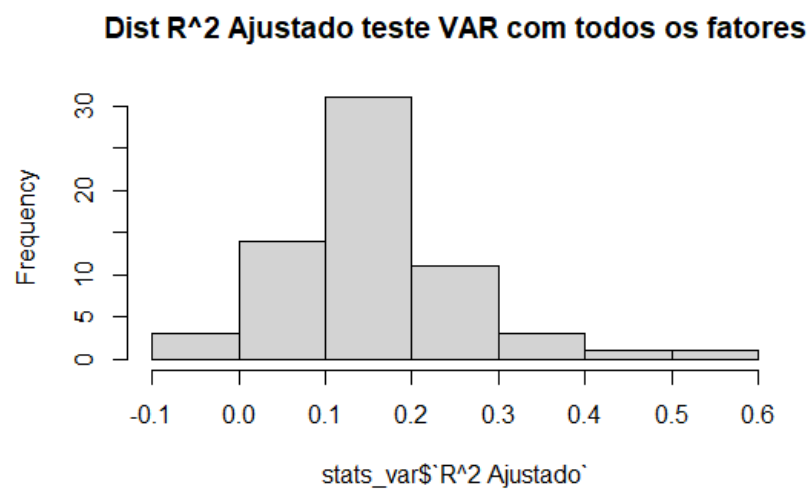
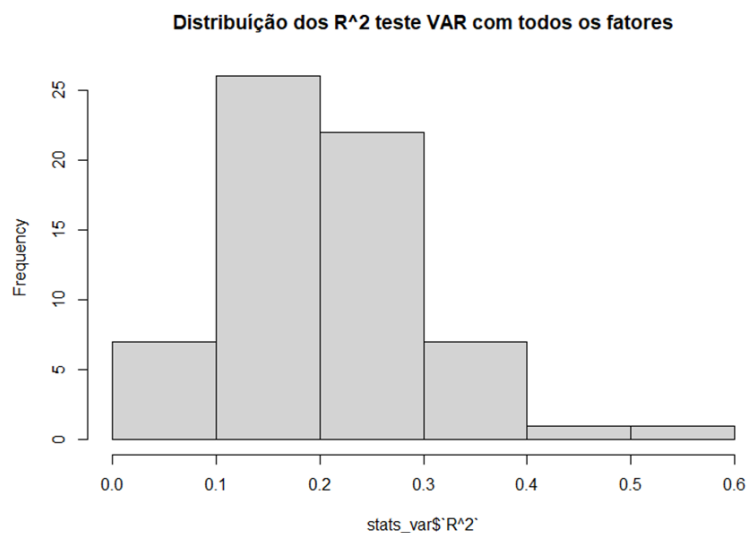
A nova série foi novamente testada para o teste de estacionaridade e todos os resultados obtidos foram estacionários. Os resultados dos testes podem ser encontrados no anexo adf2.txt.

4.3 Estimação de Modelo Auto Regressivo com todos os fatores para Desvios de Volume

As séries de primeiras diferenças obtidas foram analisadas utilizando o método de Vetores Auto Regressivos com o objetivo de avaliar a Causalidade de Granger entre as séries e a possibilidade de prever os choques de volume a partir do GSV. As saídas dos testes se encontram no anexo var1.txt

Todos os modelos encontrados tiveram a normalidade dos resíduos testados para o nível de significância de 5% e todos tiveram estatísticas testes de Jarque Bera inferiores aos níveis de significância. Os resultados compilados se encontram na tabela Testes modelo VAR 1. As saídas dos testes se encontram no anexo var1.txt

Os modelos não obtiveram resultados satisfatórios na maioria dos casos, a média dos R^2 foi de 0,280 e dos R^2 ajustados de 0,155. O próximos dois histogramas demonstram a distribuição dos R^2 obtidos



4.4 Estimação de Modelo Auto Regressivo com fatores com valor $p < 0,1$ para Desvios de Volume

Foram reestimados 24 modelos com o objetivo de avaliar se os modelos obtidos na seção anterior em que o valor p foi significativo resultariam em um melhor ajuste no modelo caso fosse realizada uma estimação considerando apenas os fatores com valor p significativo a 10%.

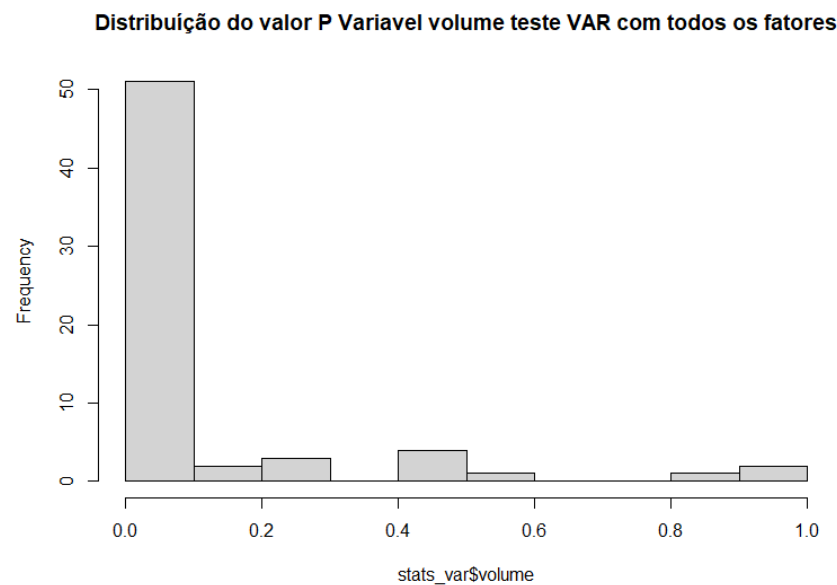
Essa teoria partiu da hipótese que os desvios em pesquisas de uma certa categoria e região que possuem maior relação com os movimentos poderiam ter sua relação prejudicada por desvios de outra categoria ou região com comportamento diferente.

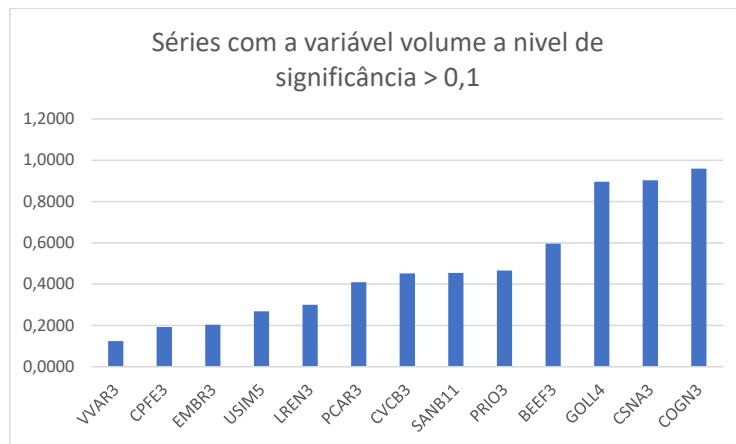
Os resíduos dos modelos encontrados foram novamente testados para normalidade sobre o teste de Jarque Bera e todos os resultados obtidos apontaram a normalidade dos resíduos. Os valores p obtidos tiveram comportamentos diversos para todas as variáveis, com a mediana próxima a 0 e média de aumento igual a 0,0375.

O R^2 ajustado para os modelos em que havia um ou mais fatores do GSV relevantes obtidos sofreu pouca mudança, com mediana da melhora percentual de 0,1% e média de queda percentual de 3,7%. Dessa forma podemos concluir que a retirada de fatores do GSV do modelo pouco influenciou para mensuração do modelo

5) Resultados e discussão

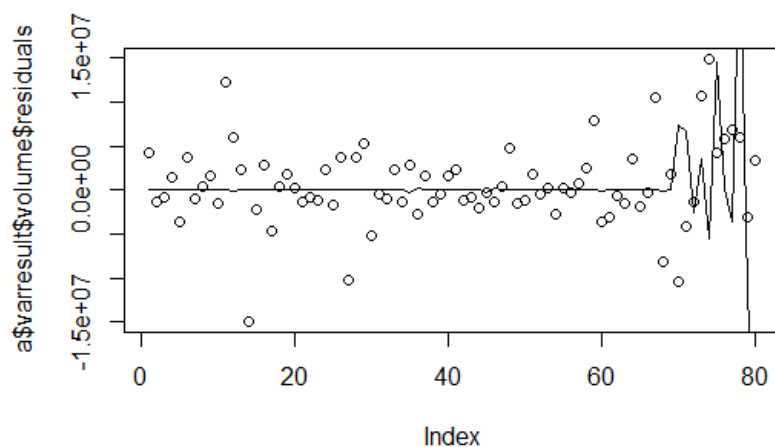
Os resultados obtidos identificaram forte significância da série de volumes nos testes p da variável. O resultado já era esperado, porém algumas séries se destacam por não possuir esse valor significativo:



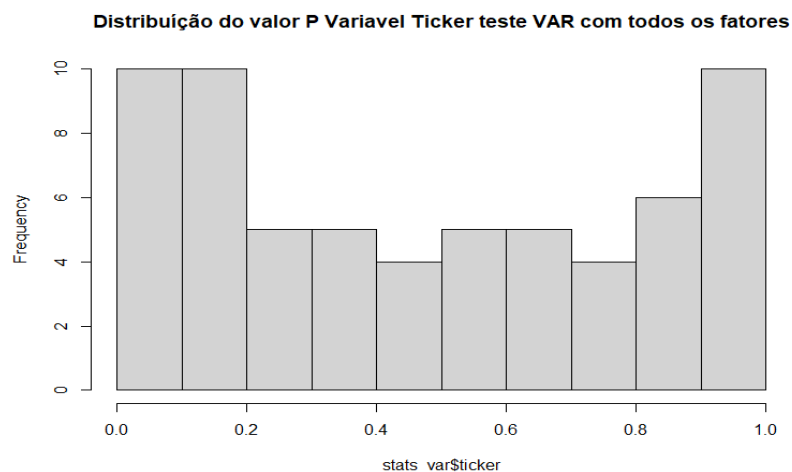


É interessante observar o caso específico de COGN3 que obteve o maior valor de ajustamento do R^2 . As séries de desvios no GSV de notícias obtiveram os maiores valores p de todos os modelos simulados enquanto o valor p da série de volumes teve o pior desempenho de todos os valores estudados

Porém ao analisar a série versus os seus resíduos podemos observar que o valor previsto é em maioria igual a 0 com perturbações observadas nos últimos períodos. O gráfico abaixo confronta o valor estimado como linha e os resíduos como pontos no gráfico:

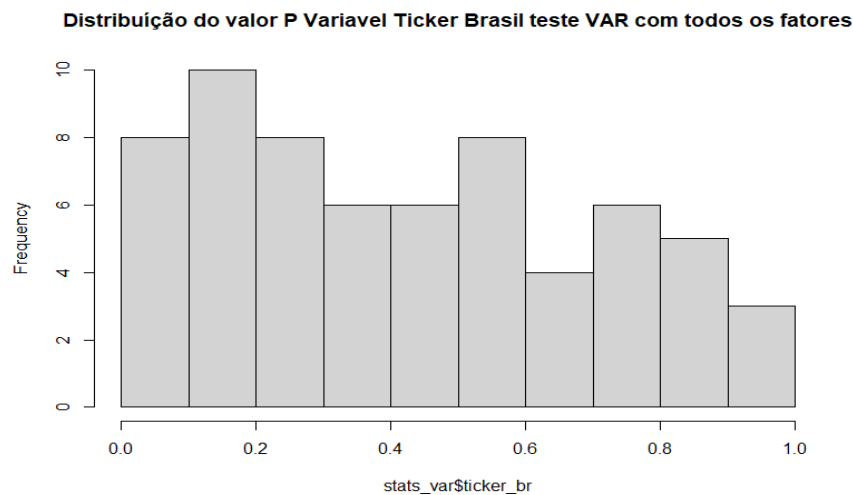


Os resultados para a variáveis GSV no modelo de previsão de volume obtiveram distribuições de resultados semelhantes.



A variável Ticker apresentou 10 valores significativos e 54 não significativos a um nível de significância de 10% no primeiro teste. Não houve melhora significativa dos resultados no segundo teste.

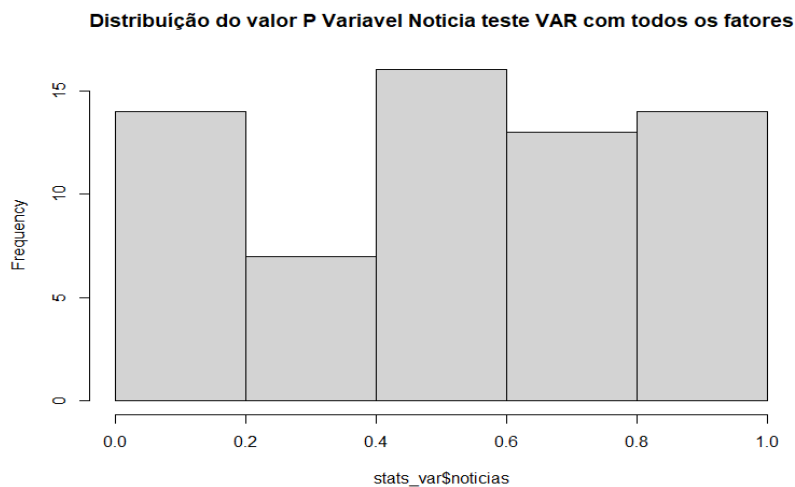
Esse resultado evidencia que a hipótese proposta de significância da relação dos índices mensais do ticker de buscas globais pelo ticker no buscador Google considerando todos os países traduzidos pelo índice GSV não possui relação mensurável em um modelo VAR considerando as flutuações mensais com os desvios de volume de negociação das ações.



A variável Ticker Brasil apresentou 8 valores significativos e 54 não significativos a um nível de significância de 10% no primeiro teste. Não houve melhora significativa dos resultados no segundo teste.

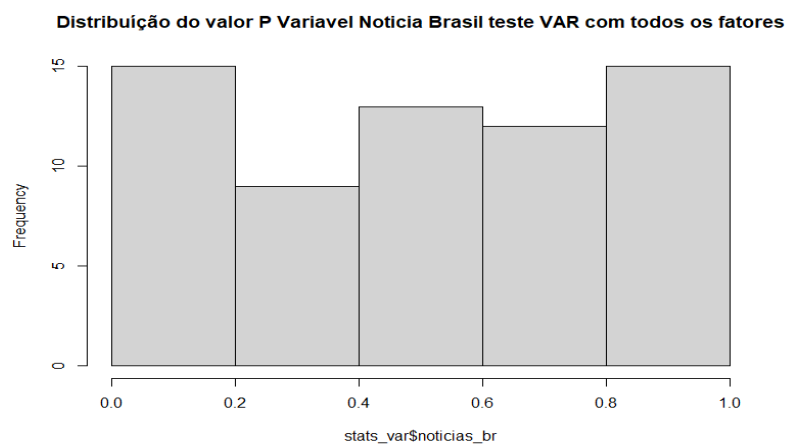
Esse resultado evidencia que a hipótese proposta de significância da relação dos índices mensais do ticker de buscas brasileiras no buscador Google considerando todos os países traduzidos pelo índice GSV não possui relação mensurável em um modelo VAR considerando as flutuações mensais com os desvios de volume de negociação das ações.

É interessante ressaltar, porém, analisar que a distribuição dos resultados foi mais concentrada em valores inferiores diferente dos outros gráficos. Outros estudos tratando especificamente essa relação podem ser realizados futuramente.



A variável Notícias apresentou 8 valores significativos e 56 não significativos a um nível de significância de 10% no primeiro teste. Não houve melhora significativa dos resultados no segundo teste.

Esse resultado evidencia que a hipótese proposta de significância da relação dos índices mensais de buscas globais por notícias do ticker da ação no buscador Google considerando todos os países traduzidos pelo índice GSV não possui relação mensurável em um modelo VAR considerando as flutuações mensais com os desvios de volume de negociação das ações mensalmente.



A variável Notícias Brasil apresentou 10 valores significativos e 56 não significativos a um nível de significância de 10% no primeiro teste. Não houve melhora significativa dos resultados no segundo teste.

Esse resultado evidencia que a hipótese proposta de significância da relação dos índices mensais de buscas no Brasil por notícias do ticker da ação no buscador Google considerando todos os países traduzidos pelo índice GSV não possui relação mensurável em um modelo VAR considerando as flutuações mensais com os desvios de volume de negociação das ações mensalmente.

6) Conclusão

Podemos concluir que o estudo de relação entre os desvios de indicadores GSV dos tickers das ações e a variação dos desvios da média de volumes negociados na bolsa de valores brasileira não possuem relação mensurável em um modelo VAR.

Os resultados encontrados sugerem que existe significância em alguns casos que precisam ser estudados mais detalhadamente, vários fatores obtiveram indicadores significativos e podem ser estudados com o objetivo contribuir para análises quantitativas de ações.

É interessante para trabalhos futuros a avaliação do indicador como fator de mensuração para correção de erros em modelos econométricos ou de avaliação de risco para robôs de estratégias quantitativas.

Ainda é interessante explorar essa relação utilizando séries diárias ou semanais do GSV onde os resultados encontrados podem ser diferentes.

7) Referências Bibliográficas

Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.

B3. (2020). Participação dos investidores. http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/participacao-dos-investidores/volume-total/

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International - Review of Financial Analysis*, 45, 150–156.

Yoshinaga, C. ; Rocco, F. (2019) - Atenção do Investidor: O Volume de Buscas no Google é Capaz de Prever os Retornos de Ações? – BBR 17 , 523 – 539

De Long, B., Shleifer, A., & Summers, L. (1990). Noise trader risk in financial markets. *The Journal of Political Economy*, 98(4), 703–738.

Challet, D., & Ayed, A. B. H. (2013). Predicting financial markets with Google Trends and not so random keywords. arXiv preprint, 1307.4643. <https://arxiv.org/pdf/1307.4643.pdf>

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3, 3415.

Preis , T.; Moat2 , H. ; Stanley , H. - Quantifying Trading Behavior in Financial Markets Using Google Trends - SCIENTIFIC REPORTS | 3 : 1684 |

Os Anexos se encontram em arquivos .txt identificados na parte de metodologia com o objetivo de viabilizar o envio do arquivo por email.

Tabela de resultados dos testes

Ação	R^2	R^2 Ajustado	volume	ticker	Valor P			noticias	noticias_br	Jarque Bera
					ticker_br					
ABEV3	0,3408	0,2968	0,0000		0,8292	0,6475	0,6518	0,3341	0,0000	
BBAS3	0,2896	0,2423	0,0000		0,1719	0,0780	0,4356	0,2277	0,0000	
BBDC3	0,1471	0,0902	0,0095		0,5531	0,6057	0,8460	0,0511	0,0000	
BBDC4	0,1076	0,0481	0,0552		0,0271	0,1354	0,5296	0,2292	0,0000	
BBSE3	0,3520	0,3088	0,0000		0,3271	0,1484	0,6585	0,9630	0,0000	
BEEF3	0,0427	-0,0212	0,5957		0,9081	0,3968	0,6091	0,8906	0,0000	
BRAP4	0,1485	0,0918	0,0145		0,1340	0,4360	0,5296	0,4473	0,0000	
BRFS3	0,2448	0,1945	0,0000		0,7827	0,5890	0,2625	0,7927	0,0000	
BRKM5	0,2659	0,2170	0,0033		0,3726	0,1883	0,1400	0,3269	0,0000	
BRML3	0,3434	0,2996	0,0017		0,9042	0,5414	0,5178	0,0080	0,0000	
BTOW3	0,2301	0,1788	0,0236		0,0123	0,0211	0,3062	0,6015	0,0000	
CCRO3	0,2660	0,2171	0,0004		0,6580	0,5972	0,4703	0,1745	0,0000	
CIEL3	0,1257	0,0674	0,0312		0,8006	0,1589	0,2179	0,4104	0,0000	
CMIG4	0,1859	0,1317	0,0641		0,9459	0,2634	0,7052	0,9309	0,0000	
COGN3	0,5747	0,5463	0,9593		0,5623	0,1734	0,0000	0,0000	0,0000	
CPFE3	0,1005	0,0406	0,1925		0,2296	0,2216	0,2205	0,8555	0,0000	
CSNA3	0,2128	0,1604	0,9027		0,1922	0,5430	0,4414	0,0139	0,0000	
CVCB3	0,3744	0,3326	0,4521		0,1113	0,1974	0,0687	0,0021	0,0000	
CYRE3	0,1648	0,1091	0,0148		0,0380	0,0559	0,3712	0,2134	0,0000	
ECOR3	0,1927	0,1389	0,0045		0,1324	0,4184	0,5937	0,5705	0,0000	
EGIE3	0,2726	0,2241	0,0000		0,2061	0,2487	0,9981	0,5882	0,0000	
ELET3	0,2185	0,1664	0,0001		0,2114	0,0268	0,7068	0,6101	0,0000	
ELET6	0,1601	0,1041	0,0203		0,9560	0,7202	0,9573	0,9783	0,0000	
EMBR3	0,0876	0,0268	0,2028		0,8785	0,4919	0,0304	0,6955	0,0000	
ENBR3	0,1756	0,1207	0,0005		0,7905	0,8109	0,3541	0,8055	0,0000	
EQTL3	0,2437	0,1932	0,0001		0,1815	0,8803	0,6864	0,7041	0,0000	
EZTC3	0,1609	0,1049	0,0076		0,0048	0,0154	0,9153	0,6357	0,0000	
FLRY3	0,1820	0,1275	0,0016		0,8628	0,7658	0,1688	0,0885	0,0000	
GGBR4	0,2437	0,1933	0,0021		0,4641	0,2761	0,0208	0,8520	0,0304	
GOAU4	0,1081	0,0486	0,0355		0,7601	0,9371	0,7685	0,5965	0,0000	
GOLL4	0,1863	0,1321	0,8957		0,5355	0,7233	0,1483	0,9115	0,0000	
HGTX3	0,4590	0,4230	0,0000		0,9362	0,3174	0,0029	0,0500	0,0000	
HYPE3	0,2128	0,1603	0,0001		0,4638	0,5489	0,4036	0,8594	0,0000	
IGTA3	0,2932	0,2461	0,0001		0,0433	0,1173	0,5066	0,5916	0,0000	
ITUB4	0,1851	0,1308	0,0127		0,1635	0,3017	0,6895	0,2031	0,0000	
JBSS3	0,1764	0,1215	0,0210		0,1765	0,1022	0,8894	0,5110	0,0000	
KLBN11	0,2187	0,1666	0,0103		0,0236	0,0120	0,8564	0,8902	0,0000	
LAME4	0,1936	0,1398	0,0001		0,6227	0,9349	0,7912	0,2198	0,0000	
LREN3	0,0379	-0,0263	0,2999		0,2711	0,3779	0,5039	0,5610	0,0000	
MGLU3	0,0984	0,0382	0,0123		0,5820	0,5780	0,9153	0,7608	0,0000	
MRFG3	0,2099	0,1573	0,0021		0,0386	0,4695	0,1848	0,1595	0,0000	
MRVE3	0,1601	0,1041	0,0026		0,6385	0,5370	0,7463	0,4636	0,0000	
MULT3	0,2287	0,1772	0,0000		0,6899	0,7043	0,6145	0,5295	0,0000	
PCAR3	0,1653	0,1096	0,4094		0,0039	0,0033	0,6781	0,5979	0,0000	
PETR3	0,0893	0,0286	0,0553		0,9531	0,8869	0,4704	0,3761	0,0000	
PETR4	0,1339	0,0761	0,0052		0,3380	0,2495	0,4308	0,3359	0,0000	
PRI03	0,2167	0,1645	0,4661		0,3234	0,9419	0,0023	0,0027	0,0000	
QUAL3	0,3441	0,3004	0,0000		0,9691	0,8896	0,9877	0,9011	0,0000	
RADL3	0,1849	0,1305	0,0009		0,1120	0,1589	0,5201	0,6809	0,0000	
RAIL3	0,2604	0,1987	0,0219		0,3386	0,2441	0,0403	0,0237	0,0000	
RENT3	0,1717	0,1165	0,0004		0,9752	0,6899	0,9588	0,7481	0,0000	
SANB11	0,0576	-0,0052	0,4544		0,5368	0,6904	0,7745	0,8408	0,0000	
SBSF3	0,2048	0,1518	0,0007		0,9788	0,5253	0,9857	0,0865	0,0000	
SULA11	0,1621	0,1063	0,0617		0,4599	0,2341	0,1683	0,9561	0,0000	
SUZB3	0,2113	0,1587	0,0031		0,4598	0,3777	0,5102	0,4507	0,0000	
TAEF11	0,1089	0,0495	0,0223		0,2947	0,3374	0,9305	0,9063	0,0000	
TOTS3	0,2652	0,2162	0,0023		0,0792	0,2131	0,5624	0,0431	0,0000	
UGPA3	0,2342	0,1831	0,0000		0,8914	0,8165	0,8858	0,6661	0,0000	
USIM5	0,0945	0,0341	0,2685		0,7152	0,4237	0,2157	0,7664	0,0000	
VALE3	0,3168	0,2713	0,0025		0,0668	0,0469	0,0051	0,0061	0,0000	
VIVT4	0,2620	0,2128	0,0000		0,8175	0,7035	0,5691	0,4168	0,0000	
VVAR3	0,3365	0,2923	0,1243		0,6934	0,4257	0,0011	0,0000	0,0000	
WEGE3	0,1458	0,0889	0,0008		0,9779	0,7478	0,8733	0,9860	0,0000	
YDUQ3	0,1279	0,0698	0,0068		0,1977	0,1373	0,8349	0,7561	0,0000	
Média	0,2081	0,1551	0,1056		0,4746	0,4228	0,5131	0,4978	0,0005	