Monitoramento do mercado de ativos brasileiro: uma proposta de *pipeline* de dados para detecção de bolhas financeiras

Uiliam B. Bomfim[1], Flávia Maristela S. Nascimento[1]

[1] Instituto Federal da Bahia (IFBA) – Análise e Desenvolvimento de Sistemas

Salvador – BA – Brasil

wf3@outlook.com.br, flaviamsn@ifba.edu.br

Resumo. *Especulações e crises financeiras podem causar problemas econômicos, especialmente em países emergentes. Por isso, agentes econômicos buscam identificar bolhas financeiras antecipadamente, para que possam tomar medidas que proporcionem estabilidade econômica. Nesse cenário, apresentamos um pipeline de dados que facilita a identificação de bolhas financeiras na B3. A análise revela a eficácia do pipeline de dados que identificou dois períodos de bolhas financeiras no mercado brasileiro.*

Abstract. *Speculation and financial crisis can cause economic problems, especially in emerging countries. Therefore, economic agents looks forward to identifying financial bubbles beforehand, so that they can act towards economic stability. In this scenario, we present a data pipeline that allows to identify financial bubbles at B3 in a simple way. The analysis reveals the effectiveness of the data pipeline that identified two periods of financial bubbles in the Brazilian market.*

1. Introdução

Comportamentos especulativos e crises financeiras causam problemas à economia. Por isso, bancos centrais, reguladores e legisladores buscam identificar com antecedência os sinais de alerta para tomar medidas para a estabilidade financeira [Phillips e Shi 2020].

Do ponto de vista econômico, uma bolha é um desvio do valor fundamental de um ativo, que é determinado por meio de uma análise dos aspectos financeiros e econômicos associados a ele [Phillips e Shi 2020]. Isso inclui demonstrativos financeiros, perspectivas de crescimento, dividendos, taxas de juros e riscos - independentemente do preço de mercado atual [Bragagnolo 2020, Chaim e Laurini 2019]. Para Phillips et al. (2015a,b), uma bolha é um comportamento explosivo de preços, sendo transitória e apresentando fases de expansão e colapso. Durante a expansão, há má alocação de recursos com fundos direcionados à especulação em lugar de empresas produtivas [Shi e Phillips 2022].

Dado que há consenso de que crises especulativas causam danos à economia, apesar das características próprias de cada crise [Phillips e Shi 2020], é fundamental identificar capital especulativo para reguladores e agentes econômicos. Assim, após a crise de 2008, os bancos centrais passaram a preconizar ações preventivas contra bolhas [Espindola 2015], mesmo operando com limitações nas ferramentas de detecção.

Na literatura recente, observa-se esse esforço para identificar bolhas, com destaque para os algoritmos de *machine learning*, que constroem modelos computacionais capazes de aprender e tomar decisões autônomas com base em dados [Kufel et al. 2023]. Estudos mostram a eficácia de algoritmos com procedimentos recursivos na identificação e datação de bolhas financeiras [Phillips e Shi 2020]. Nesse

cenário, o algoritmo PSY é amplamente utilizado como diagnóstico de alerta precoce de comportamentos semelhantes a bolhas [Monschang e Wilfling 2020].

Na busca por identificação de bolhas, os estudos têm se voltado para a aplicação de algoritmos de *machine learning*. Contudo, constata-se uma lacuna na disponibilidade de ferramentas completas (ou scripts) que possam conduzir o processo integral de *pipeline* de dados. Esse processo envolve a transferência e transformação de dados provenientes de diversas fontes até um destino específico, com o propósito de gerar *insights* ou análises de negócios [Densmore 2021]. A utilização de *pipelines* de dados possibilita o tratamento abrangente dos dados, desde a sua ingestão até a etapa de visualização, tornando mais acessível a informação para os usuários finais.

Dada a importância para os agentes do mercado financeiro, reguladores e o impacto que bolhas especulativas podem causar na economia real, justifica-se criar um *pipeline* de dados automatizado para identificação de bolhas financeiras. Este trabalho tem por objetivo implementar um *pipeline* de dados para automatizar o processo de ingestão, transformação, carga dos dados e *machine learning*, além de disponibilizar um painel que facilita a identificação de bolhas financeiras pelos usuários finais, o que possibilita mitigar danos provocados por especulações financeiras

2. Estado da Arte

Na literatura existem diferentes estudos, que utilizam algoritmos de *machine learning* para identificar bolhas financeiras. Phillips et al. (2015a) propuseram um algoritmo inovador para lidar com o desafio econométrico de identificar bolhas em dados de longo prazo. Esse método frequentemente apresenta mecanismos de identificação de datas complexos. Posto isto, para este desafio, foi desenvolvido um novo teste baseado no *sup augmented Dickey-Fuller* (ADF) que oferece uma estratégia consistente para datar o início e o fim de múltiplas bolhas. Desse modo, simulações demonstram que esse teste melhora significativamente a capacidade de identificação de bolhas.

Escobari et al. (2017) investigaram períodos de bolha financeira nos principais mercados acionários da América Latina usando algoritmos baseados no *Augmented Dickey-Fuller*. Eles concluíram que há uma forte correlação entre os episódios de bolha nos mercados latino-americanos e as bolhas no S&P 500 - índice que representa as 500 maiores empresas dos EUA. Além disso, identificaram que as bolhas na América Latina começaram mais cedo e foram mais duradouras do que as observadas nos EUA durante a crise financeira de 2008.

Hu (2023) aplicou o algoritmo PSY com o novo procedimento *bootstrap* à famosa British Railway Mania da década de 1840. Os resultados dos testes fornecem evidências de comportamento explosivo nos preços das ações das ferrovias em 1835/1836 e 1846, que estão relacionados ao *boom* ferroviário em 1836 e à mais proeminente Railway Mania em meados da década de 1840, respectivamente.

Chen et al, (2023) utilizaram o algoritmo PSY para testar a presença e datar a origem e o término de bolhas de preços determinadas por fatores latentes em um sistema de grande dimensão que incorpora muitos mercados. As simulações mostram um bom desempenho do algoritmo em termos de taxas de detecção bem-sucedidas.

Akingbade et al. (2024) aplicou o algoritmo *Log-Periodic Power Law Singularity* (LPPLS), a uma amostra de bolhas positivas e negativas no histórico de preços do Bitcoin.

Eles demonstraram que, sempre que o modelo LPPLS é ajustado aos dados, o mesmo é capaz de gerar sinais de alerta precoce.

Os trabalhos apresentados nesta seção utilizam algoritmos de *machine learning* para detecção de bolhas, porém a utilização destas abordagens é de difícil implementação para usuários finais. Por outro lado, a proposta apresentada neste trabalho implementa o processo integral de *pipeline* de dados, disponibilizando, inclusive, um painel que permite aos usuários finais identificar bolhas financeiras com antecedência.

3. Proposta

3.1 Arquitetura da Solução

A Figura 1 descreve o *pipeline* de dados construído para este trabalho. Trata-se de uma sequência de fases de processamento que envolve ingestão, transformação, carga e *machine learning* e visualização de dados.

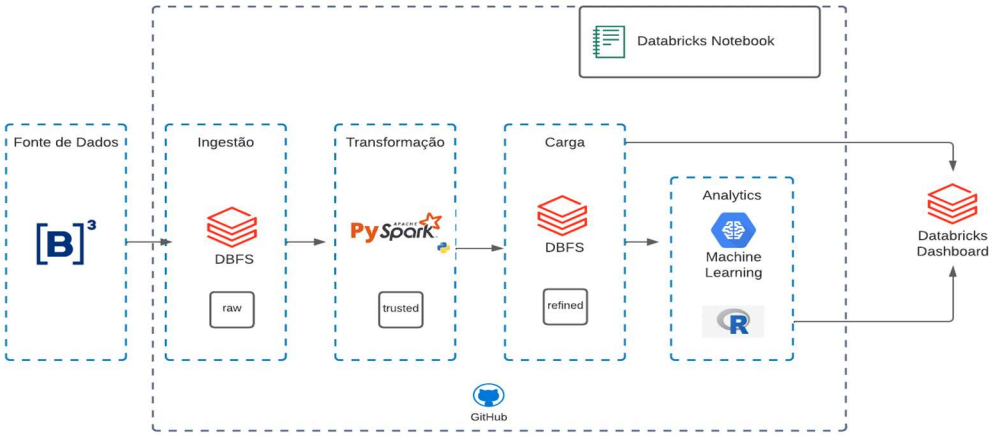

Figura 1: Fluxo de dados

1. Ingestão: a ingestão de dados foi realizada a partir do site da B3 (Brasil, Bolsa e Balcão), utilizando-se a técnica de *web scraping*.

2. Transformação: os dados brutos obtidos na fase de ingestão passaram por um processo de transformação visando a criação de uma estrutura apropriada para consulta e consumo pelos usuários. A transformação consistiu na conversão dos dados em *dataframes*[1] utilizando a biblioteca pandas. As etapas do processo incluíram a concatenação dos *dataframes* brutos, a limpeza para remoção de valores nulos, a seleção das colunas de data e preço, a conversão dos tipos de dados, e a subdivisão dos *dataframes* em três categorias distintas: um *dataframe* de dividendos, um *dataframe* de juros sobre capital próprio, e um *dataframe* combinando ambas as categorias.

3. Carga: Esta fase caracterizou-se por disponibilizar dados para análise dos usuários. Os *dataframes* resultantes da fase de transformação foram carregados no DBFS (*Databricks File System) no* formato *csv*.

4. *Machine Learning*: Nesta fase utilizou-se o algoritmo PSY, que é utilizado como um diagnóstico de alerta precoce de comportamento semelhante a bolhas devido à sua eficácia [Hu e Oxley 2018, Monschang e Wilfling (2021), Phillips et al. 2015a].

[1] O termo *dataframe* é utilizado aqui como uma estrutura de dados análoga a uma tabela de dados bidimensional

5. Visualização - Para esta fase, o principal objetivo foi traduzir informações complexas ou abstratas de maneira gráfica e intuitiva com o intuito de facilitar a análise e interpretação. Os gráficos incorporam informações relativas aos períodos dos dados de entrada e aos períodos de bolhas financeiras resultantes da fase de *machine learning*.

3.2 Modelo de *Machine Learning*

No modelo trabalhado, utilizamos o algoritmo proposto em Philips e Shi (2020), para identificar períodos de comportamento explosivo de preços, incluindo o início e fim do período. Este algoritmo usa evidências de comportamento explosivo de preços como *proxy* para detecção de bolhas financeiras. Observamos que o algoritmo PSY apresenta melhor desempenho que outros algoritmos, como *sup-ADF-style*, CUSUM, PWY, PSY*sign-based* [Monschang e Wilfling 2020, Phillips et al. 2015a,b], pois o PSY utiliza um procedimento de *bootstrap* projetado para mitigar o impacto potencial da heterocedasticidade.

Ademais, empregamos *Dividend Yield* e *Dividend* JSCP *Yield* como variáveis de entrada para estimação do modelo, conforme Caspi e Graham (2018), Escobari et al. (2017), Monschang e Wilfling (2021), Phillips et al. (2015a,b). O *Dividend Yield* compara os dividendos de uma empresa com o preço de suas ações, refletindo o retorno de dividendos. No Brasil, os Juros sobre Capital Próprio (JSCP) são uma alternativa que oferece vantagens fiscais ao distribuir lucros. Portanto,o *Dividend* JSCP *Yield* relaciona os JSCP pagos com o preço das ações, tornando-se crucial no contexto de peculiaridades contábeis brasileiras que demandam uma análise cuidadosa desses indicadores.

4. Resultados

4.1 Análise dos *Dividend yield*

Episódios de Crises das Empresas Listadas na B3: Dividend Yield

	start	end
1	2008-03-01	2008-03-01
2	2008-09-01	2009-04-01
3	2010-05-01	2012-09-01

Figura 2: Episódios de Crises das empresas Listadas na B3- *Dividend Yield*

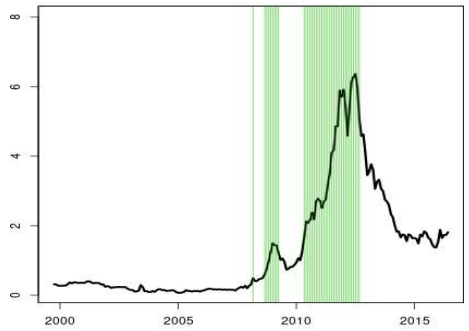

Figura 3: Episódios de Crises das Empresas listadas na B3 - *Dividend Yield*

Os resultados apresentados aqui foram baseados nos dados de *dividend yield* de todas empresas ativas listadas na B3, entre o período de 1996 e 2023, totalizando 3334 observações; (b) a Figura 2 traça a relação preço/dividendo das empresas na B3 e (c) as datas utilizadas referem-se à aprovação dos dividendos e não necessariamente da distribuição ou apuração e podem representar períodos mais longos de comportamento financeiro.

Quanto ao *Dividend Yield*, cabe destacar que, no início de 2008, houve um crescimento moderado, seguido por um aumento significativo entre o final de 2008 e

início de 2009. Esse período foi marcado pelo fluxo abundante de capitais para países emergentes - incluindo o Brasil - e por resultados positivos nas transações comerciais e correntes. No entanto, a crise do *subprime* interrompeu o ciclo virtuoso no país no final de 2009 [Lima e Deus 2013]. Um segundo crescimento do *Dividend Yield* ocorreu entre 2010 e 2012, coincidindo com o início da crise das *commodities*.

Destaque-se, ainda, que o modelo desenvolvido não detectou bolha financeira nos primeiros 12 anos da amostra. Esta ausência pode ser explicada pelo contexto macroeconômico brasileiro que conviveu nos anos 90 com baixo crescimento econômico, alta inflação [Pinheiro et. al. 1999] e um aumento expressivo no déficit da conta corrente [Frizo e Lima, 2014], além de um mercado de capitais incipiente [De Amorim et al. 2021]. Neste sentido, é possível inferir, de acordo com os resultados apresentados, que o *pipeline* é eficaz para automatizar o processo de identificação de bolhas financeiras das empresas listadas na B3.

4.2 Análise dos *Dividend JSCP yield*

Os resultados apresentados aqui foram baseados nos dados de de *Dividend JSCP yield* de todas empresas ativas listadas na B3 entre o período de 1997 e 2023, totalizando 4434 observações. Não foram encontrados dados para o ano de 1996. Assim como observamos nos resultados para *Dividend Yield*, as características do mercado brasileiro também podem influenciar na eficácia dos modelos de *machine learning*, especialmente nos 10 primeiros anos da amostra.

Episódios de Crises das Empresas Listadas na B3:
Dividend JSCP yield

	start	end
1	2008-03-01	2008-03-01
2	2008-09-01	2009-04-01
3	2010-05-01	2012-09-01

Figura 4: Episódios de Crises das empresas Listadas na B3 - *Dividend JSCP yield*

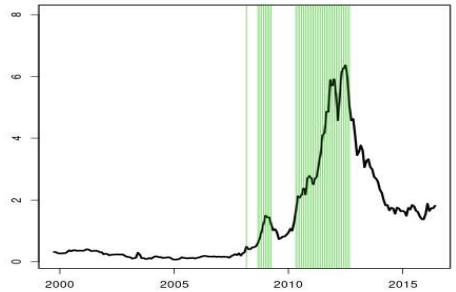

Figura 5: Episódios de Crises das Empresas listadas na B3 - *Dividend JSCP Yield*

Por sua vez, feita a análise do *Dividend JSCP yield,* observou-se uma tendência semelhante à do *Dividend Yield.* Isto se explica pelo fato de ambas as variáveis assimilarem funções diretamente proporcionais ao lucro das empresas. Os resultados demonstram dois períodos de crescimento. O primeiro está associado à abundância de fluxos de capitais entre 2003 e 2007 seguido pela crise do subprime em 2009. O segundo corresponde ao período que coincide com a crise das *commodities,* em 2010.

A título de conclusão, as análises dos dados de *Dividend* JSCP *yield* corroboram os resultados apresentados utilizando *Dividend yield. Estes resultados* apontam, em mais uma experiência, a eficácia do *pipeline* de dados para a identificação de bolhas financeiras, tomando como base as empresas listadas na B3.

4.3 Análise dos dados combinados (*Dividend JSCP yield* + *Dividend yield*)

Com caráter comprobatório, no mencionado modelo empregou-se dados de *Dividend yield* e *Dividend* JSCP *yield* (Juros Sobre Capital Próprio) de forma combinada. A amostra abrange as empresas listadas na B3, no período entre 1996 e 2023.

Episódios de Crises das Empresas Listadas na B3:
Total = Dividend + JSCP

	start	end
1	2008-03-01	2008-03-01
2	2008-09-01	2009-04-01
3	2010-05-01	2012-09-01

Figura 6: Episódios de Crises das empresas Listadas na B3 - Dados Combinados

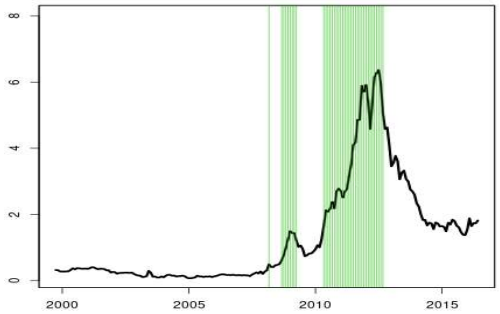

Figura 7: Episódios de Crises das Empresas listadas na B3 - Dados combinados

Assim como nos resultados seccionados, a análise conjunta do *Dividend Yield* e do *Dividend JSCP yield* revela dois períodos de crescimento. Este resultado era esperado porque ambas as variáveis são funções do lucro líquido das empresas. As análises dos dados combinados corroboram com os resultados apresentados aplicando-se *Dividend yield e Dividend* JSCP *yield* o que ressalta a eficácia do *pipeline* de dados para a identificação de bolhas financeiras nas empresas listadas na B3.

5. Considerações Finais

O presente trabalho apresentou um *pipeline* de dados automatizado para a detecção de bolhas financeiras no mercado de ativos brasileiros. O objetivo principal foi fornecer uma ferramenta eficaz para bancos centrais, reguladores fiscais e agentes econômicos identificarem precocemente sinais de comportamento especulativo e potenciais crises financeiras, visando a estabilidade econômica.

O *pipeline* desenvolvido mostrou-se eficaz na automação do processo de ingestão, transformação e carga de dados, além da aplicação do algoritmo PSY. A análise identificou dois períodos de bolhas financeiras no mercado brasileiro. A criação de um painel visual intuitivo permitiu uma análise fácil e rápida das empresas listadas na B3, facilitando a identificação de períodos de bolhas financeiras.

Apesar dos resultados positivos, devemos considerar alguns pontos. O contexto econômico brasileiro nas décadas de 90 e 2000, com baixo crescimento e alta inflação, pode ter afetado a detecção de bolhas nos primeiros 12 anos do estudo. Adicionalmente, a baixa liquidez e a concentração de ações no mercado nacional também representam desafios para o desempenho do algoritmo de *machine learning*. Por fim, o *pipeline* de dados proposto neste trabalho é uma contribuição significativa para a detecção de bolhas financeiras no mercado brasileiro. A automação do processo e a aplicação do algoritmo PSY oferecem uma ferramenta valiosa para agentes financeiros e reguladores, facilitando respostas mais eficientes a crises potenciais. Para trabalhos futuros é sugerido testar o modelo com diferentes variáveis financeiras e explorar outros algoritmos de *machine learning*. A inclusão de indicadores econômicos abrangentes e a adaptação do modelo a diversos contextos de mercado fortalecerão sua capacidade de identificar bolhas financeiras.

Referências

Akingbade, S., Gidea, M., Manzi, M., & Nateghi, V. (2024). Why topological data analysis detects financial bubbles?. Communications in Nonlinear Science and Numerical Simulation, 128, 107665.

Bragagnolo, G. (2020). Pecuária bovina no Brasil e disfuncionalidades do mercado financeiro: um estudo sobre os impactos no valor de mercado dos frigoríficos brasileiros de capital aberto decorrente do aumento da demanda chinesa em virtude da peste suína africana (Doctoral dissertation).

Caspi, I. e Graham, M. (2018). Testing for bubbles in stock markets with irregular dividend distribution. Finance Research Letters, 26, 89-94.

Chaim, P. e Laurini, M. (2019). Is Bitcoin a bubble?. Physica A: Statistical Mechanics and its Applications, 517, 222-232.

Chen, Y., Phillips, P., & Shi, S. (2023). Common bubble detection in large dimensionais financiais systems. Journal of Financial Econometrics, 21(4), 989-1063.

De Amorim, G., Lima, N. e Júnior, A. (2021). Distribuição de Dividendos e Valor de Empresas Listadas na B3. Advances in Scientific and Applied Accounting, p. 3-18.

Densmore, J. (2021). Data pipelines pocket reference. O'Reilly Media.

Escobari, D., Garcia, S. e Mellado, C. (2017). Identifying bubbles in Latin American equity markets: Phillips-Perron-based tests and linkages. Emerging Markets Review, 33, 90-101.

Espindola, R. (2015). A crise financeira e a política monetária no Brasil (Dissertation)

Frizo, P. e Lima, R. (2014). Efeitos da flutuação dos preços das commodities no fluxo de investimento estrangeiro direto no Brasil. Revista de Economia Contemporânea, 18, 393-408.

Hu, Y. (2023). A review of Phillips-type right-tailed unit root bubble detection tests. Journal of Economic Surveys, 37(1), 141-158.

Hu, Y.;Oxley, L. (2018). Bubble contagion: Evidence from Japan's asset price bubble of the 1980-90s. Journal of the Japanese and International Economies, 50, 89-95.

Lima, T. e Deus, L. (2013). A crise de 2008 e seus efeitos na economia brasileira. Revista Cadernos de Economia, 17(32), 52-65.

Monschang, V. e Wilfling, B. (2021). Sup-ADF-style bubble-detection methods under test. Empirical Economics, 61, 145-172.

Phillips, P., Shi, S. e Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. International economic review, 1043.

Phillips, P., Shi, S. e Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. International Economic Review, 56(4), 1079-1134.

Phillips, P., e Shi, S. (2018). Financial bubble implosion and reverse regression. Econometric Theory, 34(4), 705-753.

Phillips, P. e Shi, S. (2020). Real time monitoring of asset markets: Bubbles and crises. In Handbook of statistics (Vol. 42, pp. 61-80). Elsevier.

Pinheiro, A.; Giambiagi, F.; Gostkorzewicz, J. (1999). O desempenho macroeconômico do Brasil nos anos 90.

Shi, S. e Phillips, P. (2022). Econometric Analysis of Asset Price Bubbles (No. 2331). Cowles Foundation for Research in Economics, Yale University.