

# Análise de propriedades das séries temporais dos ativos que compõem o índice IBOVESPA

César Daltoé Berci\*, Celso Pascoli Bottura†

## Resumo

Diversas características das séries temporais financeiras são de interesse tanto do ponto de vista acadêmico, a partir do qual pretende-se analisar a dinâmica dos dados e suas propriedades numéricas, bem como de investidores, que por sua vez utilizam-se desse conhecimento na intenção de obter lucro em suas transações financeiras. Através da aplicação de diversas ferramentas de análise, fazendo uso de uma capacidade de computação massiva, foram avaliados os ativos que compõem o índice IBOVESPA, quanto as suas propriedades numéricas e estatísticas. Dada a relevância e abrangência das séries temporais analisadas, os resultados obtidos a partir desta análise podem servir como base para a caracterização das séries temporais financeiras.

## Palavras-chave

Séries temporais, IBOVESPA, Retorno de ativos.

# Analysis of time series properties for the assets that compose the IBOVESPA index

## Abstract

Several characteristics of financial time series are of interest both from an academic point of view, which is intended to analyze the dynamics of the data and its numerical properties, as well as from investors point of view, who use this knowledge to generate profit in their financial transactions. By applying several analysis tools and using a massive computing capacity, the numerical and statistical properties of the assets that compose the IBOVESPA index were evaluated. Given the relevance and scope of the analyzed time series, the results obtained from this analysis can serve as a basis for the characterization of financial time series.

## Keywords

Time series, IBOVESPA, Asset returns

## I. INTRODUÇÃO

O presente estudo apresenta uma avaliação criteriosa das propriedades estatísticas das séries temporais financeiras, estabelecendo a partir desta análise, resultados consistentes de forma a subsidiar a modelagem de tais séries temporais, independentemente dos métodos e modelos que se pretenda aplicar e da finalidade a qual se propõe a aplicação de desses modelos.

Utiliza-se aqui uma vasta quantidade de ferramentas, provendo métodos e meios necessários à construção das conclusões objetivadas pela análise proposta. O leitor encontrará descrições com razoável nível de detalhamento da maioria dessas ferramentas, de forma que seja possível contextualizar e reproduzir os resultados aqui enunciados.

O objetivo final da análise, ainda que sua intangibilidade o torne incerto por sua própria natureza, é identificar se e como os valores passados de um ativo podem influenciar seus valores futuros. Tal questionamento ganhou espaço na

discussão acadêmica por volta de 1900, com a tese de Louis Bachelier, intitulada *Théorie de la spéculation* [1]. Desde então o estudo de séries temporais financeiras é objeto de inúmeras pesquisas.

Existem controvérsias entre diferentes autores que podem ser divididos em dois grandes grupos, aqueles que baseiam-se no modelo de mercado eficiente [2], e aqueles que defendem a teoria de que as séries temporais financeiras comportam-se como caminhos aleatórios.

Não há, porém, formas determinísticas de se afirmar sem margem de dúvida qual a natureza de algo essencialmente aleatório, pois ao fazê-lo ele deixaria de ser o que de fato é, aleatório. Todavia pretende-se manter o objetivo no campo do intangível, pois próximo a ele encontrar-se-há padrões e parâmetros capazes de ampliar a compreensão dos fenômenos envolvidos nas análises.

Das características dos dados e parâmetros avaliados, tem-se a aleatoriedade como fator predominante, de tal sorte que toda avaliação traz consigo condicionantes, que podem restringir as conclusões a grupos específicos de séries temporais. Visando obter conclusões mais abrangentes, optou-se por avaliar uma base de dados que possuía, por sua própria natureza, a abrangência desejada.

Uma descrição detalhada dessa base de dados é apresentada

\* UNICAMP - Universidade Estadual de Campinas; † UNICAMP - Universidade Estadual de Campinas

E-mails: cesar@berci.com.br; cpascolib@gmail.com

Data de envio: 19/11/2020

Data de aceite: 31/05/2021

<http://dx.doi.org/10.18226/23185279.v9iss2p05>

na próxima sessão, seguida por uma avaliação dos modelos de probabilidade das distribuições empíricas, onde é apresentada uma alternativa de distribuição de probabilidade, analítica e de variância finita, para representar essas distribuições empíricas.

A avaliação de uma grande quantidade de dados, conforme proposto neste trabalho, requer a aplicação de técnicas capazes de condensar a informação ou expressá-la utilizando medidas de resumo, tornando possível uma análise comparativa de forma abrangente. O formato preferencial adotado para expressar uma grande quantidade de dados é utilização de gráficos, que concatenam um extenso volume de dados em uma única informação visual condensada.

Posteriormente, demais propriedades, como a homoscedasticidade e estacionaridade das séries temporais também são avaliadas através de testes analíticos. Na sequência é estudada a correlação inter ativos e intra ativos, confrontando, inclusive, alguns padrões empíricos, bem conhecidos por investidores como o efeito *leverage* [3].

Esses padrões também são analisados nas sessões seguintes, onde é avaliada a dependência temporal e a sazonalidade das séries de retorno, que apresentam impacto direto na teoria de mercado eficiente.

## II. BASE DE DADOS

Os dados utilizados nesta análise foram extraídos do sistema *BM&FBovespa*, tendo sido avaliados os ativos que compõem o índice IBOVESPA. Esta escolha é, por óbvio, a mais adequada, pois esse índice é utilizado como *proxy* para o índice do mercado de capitais brasileiro [4]. Muito embora existam críticas a esta abordagem, como em [5], entende-se que esta seja uma base de dados significativa e suficientemente abrangente para a análise proposta.

O *BM&FBovespa* define os critérios para o cálculo do índice IBOVEPA, conforme descrito em [6].

Em consulta realizada em Abril de 2018, o índice era composto por 64 ativos. Os dados desses ativos, disponibilizados pelo sistema IBOVESPA, foram coletados para o período entre 08 de março de 2015 até 06 de março de 2018. A quantidade de amostras obtidas e os ativos que compõem o índice estão representados na Tabela I:

TABLE I: Ativos que compõem o índice IBOVESPA

	#	N	Erros		#	N	Erros
<b>ABEV3</b>	1	740	0	<b>ITSA4</b>	33	739	0
<b>B3SA3</b>	2	740	1	<b>ITUB4</b>	34	739	0
<b>BBAS3</b>	3	740	0	<b>JBSS3</b>	35	740	2
<b>BBDC3</b>	4	740	0	<b>KLBN11</b>	36	740	1
<b>BBDC4</b>	5	740	1	<b>KROT3</b>	37	740	1
<b>BBSE3</b>	6	740	1	<b>LAME4</b>	38	740	0
<b>BRAP4</b>	7	740	1	<b>LREN3</b>	39	740	2
<b>BRFS3</b>	8	740	1	<b>MGLU3</b>	40	740	1
<b>BRKM5</b>	9	740	1	<b>MRFG3</b>	41	740	1
<b>BRML3</b>	10	740	0	<b>MRVE3</b>	42	740	1
<b>CCRO3</b>	11	740	0	<b>MULT3</b>	43	740	0
<b>CIEL3</b>	12	740	0	<b>NATU3</b>	44	740	2
<b>CMIG4</b>	13	740	0	<b>PCAR4</b>	45	740	1
<b>CPFE3</b>	14	740	1	<b>PETR3</b>	46	740	1
<b>CPLE6</b>	15	740	0	<b>PETR4</b>	47	740	1
<b>CSAN3</b>	16	740	2	<b>QUAL3</b>	48	740	0
<b>CSNA3</b>	17	740	0	<b>RADL3</b>	49	740	1
<b>CYRE3</b>	18	740	0	<b>RAIL3</b>	50	263	0
<b>ECOR3</b>	19	740	0	<b>RENT3</b>	51	740	0
<b>EGIE3</b>	20	740	0	<b>SANB11</b>	52	740	1
<b>ELET3</b>	21	740	1	<b>SAPR11</b>	53	91	0
<b>ELET6</b>	22	740	1	<b>Sbsp3</b>	54	740	0
<b>EMBR3</b>	23	740	1	<b>SMLS3</b>	55	110	1
<b>ENBR3</b>	24	740	1	<b>SUZB3</b>	56	740	2
<b>EQTL3</b>	25	739	0	<b>TAEE11</b>	57	740	0
<b>ESTC3</b>	26	740	1	<b>TIMP3</b>	58	740	0
<b>FIBR3</b>	27	740	1	<b>UGPA3</b>	59	740	1
<b>FLRY3</b>	28	740	1	<b>USIM5</b>	60	740	0
<b>GGBR4</b>	29	740	1	<b>VALE3</b>	61	740	1
<b>GOAU4</b>	30	740	1	<b>VIVT4</b>	62	740	2
<b>HYPE3</b>	31	740	1	<b>VVAR11</b>	63	740	0
<b>IGTA3</b>	32	740	0	<b>WEGE3</b>	64	740	2

Na Tabela I, a coluna “N” representa a quantidade de “Pregões” do ativo disponíveis no sistema IBOVESPA para o período analisado. A coluna “Erros” traduz a quantidade de não conformidades identificadas nos dados, mais especificamente, é a somatória dos casos onde se observa um valor máximo inferior ao fechamento, ou um valor mínimo superior ao fechamento, por fim, a coluna # é um índice de indexação utilizado para o reconhecimento do ativo em gráficos e tabelas. Essas situações, apesar de tratadas como erros na Tabela I, representam uma situação normal para o mercado, pois podem refletir transações realizadas no “*after market*”. Assim, para esses casos, optou-se por apenas ajustar os dados não conformes.

Para cada um dos ativos foram coletados os preços de abertura, fechamento, máximo, mínimo e volume financeiro negociado, considerando pregões diários entre 08/04/2015 e 06/04/2018, totalizando 740 amostras de cinco séries temporais por ativo. Em resumo, estão sendo avaliadas 305 séries temporais com 740 amostras cada, totalizando 225700 amostras.

As seguintes grandezas são definidas para cada uma das séries temporais:

- 1)  $o_t$  é o preço de abertura realizado pelo ativo no pregão  $t$ ;
- 2)  $c_t$  é o preço de fechamento realizado pelo ativo no pregão  $t$ ;
- 3)  $l_t$  é o preço mínimo do ativo negociado no pregão  $t$ ;
- 4)  $h_t$  é o preço máximo do ativo negociado no pregão  $t$ ;
- 5)  $v_t$  é o volume total negociado no pregão.

A título de ilustração a Figura 1 mostra todos os dados disponíveis do ativo RAIL3, incluindo as 5 séries temporais,

dispostas em dois gráficos distintos, um para o volume e um gráfico de candles das cotações.



Fig. 1: Histórico de cotações do ativo RAIL3

#### A. Validação prévia dos dados

A validação prévia dos dados foi realizada com base em inspeção visual das séries, plotadas conforme mostrado na Figura 1, e na quantidade de dados de cada série.

Dos sessenta e quatro ativos, quatro foram omitidos das análises envolvendo cálculo de correlações, em virtude da escassez de dados, ou de anomalias neles presente que possam enviesar a análise de alguma forma. Desses, três apresentam séries truncadas, sendo eles: RAIL3, SAPR11 e SMLS3, enquanto a quarta série, LRN3, apresenta variações anômalas nos dados que podem prejudicar a modelagem, conforme observa-se na Figura 2.

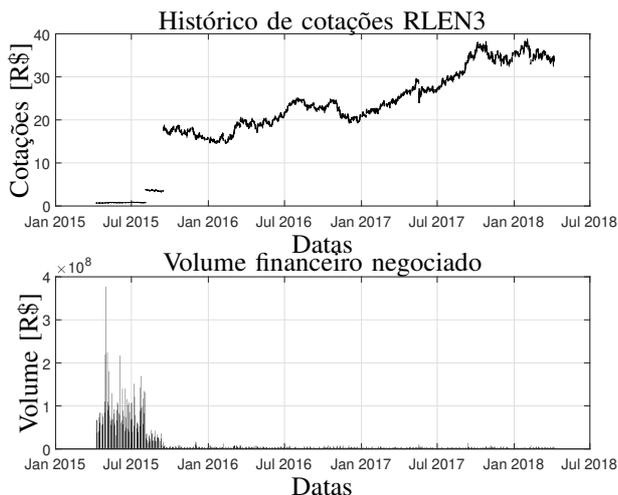


Fig. 2: Histórico de cotações do ativo RLEN3

Apesar de não serem consideradas nas comparações envolvendo modelagem, as propriedades dessas quatro séries também serão avaliadas nos contextos onde essa análise é possível e pertinente, como na determinação da distribuição empírica dos retornos, por exemplo. Para a série LRN3, apenas, serão desconsiderados os dados dos pregões posteriores

ao dia 15 de setembro de 2015, onde as anomalias foram observadas.

#### B. Retornos

Consideremos a série temporal que representa os retornos do preço de fechamento dos ativos, descritas pela seguinte equação:

$$r_t = \frac{c_t - c_{t-1}}{c_{t-1}} \approx \ln(c_t) - \ln(c_{t-1}) \quad (1)$$

A relevância econômica do preço de fechamento de um ativo, computada em pregões diários, é relativamente maior do ponto de vista do investidor quando comparado as outras grandezas sob análise. Atribui-se essa constatação ao fato de que são esses valores que impõem o preço de mercado das empresas e commodities negociados em bolsas de valores. Assim, muitas das análises desenvolvidas nesse trabalho, bem como na maioria dos documentos a respeito deste tema, tem como foco as séries de retornos dos preços de fechamento.

A fim de ilustrar o comportamento das séries de retornos de forma abrangente, tomemos como medida de resumo dessas grandezas os dados mostrados na Figura 3. Nela são plotados no eixo vertical os valores médios dos retornos dos 64 ativos da base de dados, incluindo os intervalos de 1, 2 e 3 desvios padrão, e no eixo horizontal tem-se o número do ativo conforme indexação dada na tabela I.

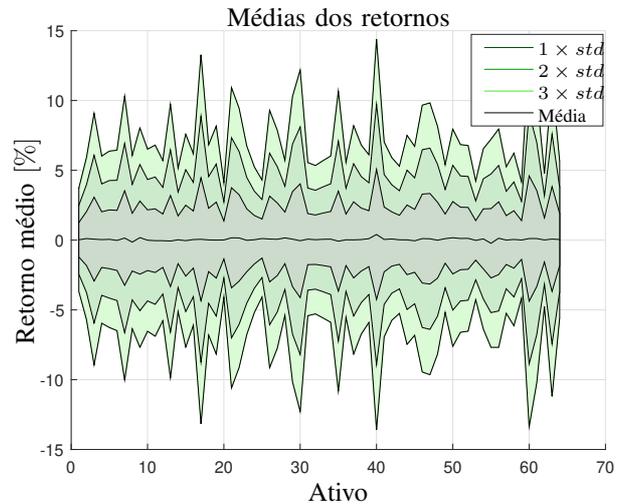


Fig. 3: Média e desvio padrão dos retornos

Observa-se na figura 3 que o valor médio dos retornos é muito próximo a zero para todos os ativos da base de dados e suas dispersões, apesar de bastante distintas entre diferentes ativos, não ultrapassam 15% (considerando o intervalo de 3 desvios padrão da média). Empiricamente pode-se reconhecer na figura a volatilidade dos ativos.

### III. DISTRIBUIÇÕES EMPÍRICAS

Pretendendo avaliar as distribuições de probabilidade teóricas que regem o comportamento dos preços dos ativos, consideremos a série de retornos do preço de fechamento como representante suficiente para explicar essas propriedades, deixando para o final da análise avaliar se essa representatividade de fato é verdadeira.

Estudos similares concluem que a hipótese gaussiana não é a melhor alternativa para representar os dados [7], e sim, prevalece sobre ela a hipótese de Mandelbrot [8], que sugere que os dados seguem uma distribuição  $\alpha$ -estável com  $\alpha < 2$ . Essa hipótese, de que as distribuições  $\alpha$ -estáveis são as que melhor representam as distribuições de probabilidade das séries temporais financeiras [9] é amplamente aceita atualmente.

O presente estudo além de avaliar as hipóteses de normalidade para os ativos da base de dados, contrastando com a hipótese de Mandelbrot, apresenta uma alternativa de distribuição para representar os dados, que melhor se adequam à leptocurtose e às caudas pesadas das distribuições empíricas.

A fim de avaliar os parâmetros básicos das distribuições empíricas, foi considerado inicialmente o gráfico da Figura 4. Nele são plotados os valores dos coeficientes de assimetria e curtose dos retornos de todos os ativos, bem como linhas de referência para comparação com os valores típicos de uma distribuição normal.

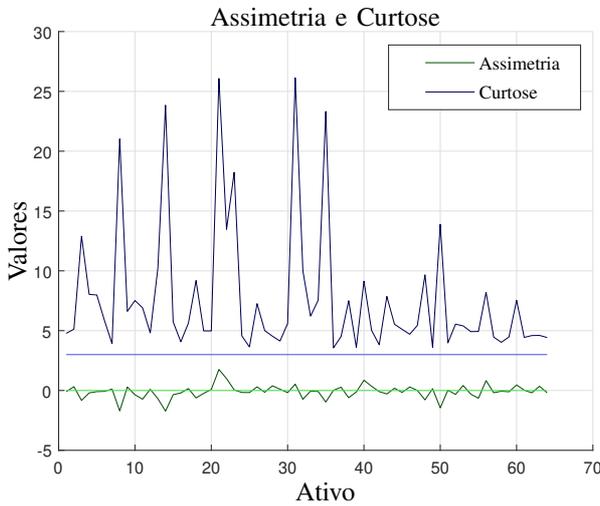


Fig. 4: Coeficientes de assimetria e curtose dos retornos dos 64 ativos

Nota-se no gráfico a proximidade do coeficiente de assimetria com linha base de referência, porém ao contrário do observado por Taylor [10], esses coeficientes não apresentam predominância de sinal, indicando que os retornos são, aproximadamente simétricos. Porém os coeficientes de curtose são todos elevados, apenas alguns ativos apresentaram coeficientes inferiores a 5, enquanto outros têm valores superiores a 25. Isso indica que a distribuição empírica dos retornos é predominantemente leptocúrtica.

A. Testes de normalidade

Consideremos uma amostra  $X$  de uma variável aleatória  $\mathcal{X}$ , cuja distribuição empírica é hipoteticamente normal. Os testes de normalidade aqui descritos têm por objetivos julgar a veracidade dessa hipótese com base nos dados amostrais disponíveis.

B. Teste de Kolmogorov Smirnov

O teste de Kolmogorov Smirnov tem por objetivo avaliar as seguintes hipóteses:

$$\begin{cases} H_0 : \text{Os dados seguem uma distribuição normal} \\ H_1 : \text{Os dados não seguem uma distribuição normal} \end{cases}$$

Considerando a função cumulativa empírica de probabilidade da amostra  $F_X(x)$ , a qual pretende-se comparar com a função  $F_n(x)$  (função cumulativa normal). A estatística utilizada no teste é dada por [11]:

$$D_n = \sup_x |F_X(x) - F_n(X)| \tag{2}$$

A função  $F_X(x)$  é definida pela seguinte equação:

$$F_X(x) = \frac{1}{n} \sum_{i=1}^n nI_{[\infty, x]}(x_i) \tag{3}$$

onde  $I$  é uma função indicadora que retorna 1 apenas quando  $x_i < x$  e zero caso contrário.

Como a função empírica é discreta e a função hipotética é contínua, deve-se obter o valor crítico do teste através das seguintes equações:

$$D_n^+ = \sup_x |F_X(x_i) - F_n(x_i)| \tag{4}$$

$$D_n^- = \sup_x |F_X(x_i) - F_n(x_{i-1})| \tag{5}$$

$$D_n = \max(D_n^+, D_n^-) \tag{6}$$

Por fim, se  $D_n$  é maior que o valor crítico  $C_v$ , rejeita-se a hipótese  $H_0$ , que aqui representa a normalidade dos dados. O valor crítico, por sua vez, deve ser calculado utilizando a distribuição hipotética, no caso, normal.

C. Teste de Kolmogorov-Smirnov, método de Stephens

O método elaborado por M. A. Stephens propõe a utilização da estatística de teste modificada [12]:

$$D^* = D_n \left( \sqrt{n} - 0,01 + \frac{0,85}{\sqrt{n}} \right) \tag{7}$$

Os valores críticos para a estatística do teste  $D^*$  foram obtidas por Stephens através de simulações de Monte Carlo. A avaliação do teste é feita com base nos p-valores<sup>1</sup> ( $p$ ) são calculados através de interpolações dos valores dados na equação 8 [13].

$$p = \begin{cases} > 0,15 & D^* < 0,775 \\ 0,15 + \frac{(D^* - 0,775)(0,10 - 0,15)}{0,819 - 0,775} & 0,775 \leq D^* < 0,819 \\ 0,10 + \frac{(D^* - 0,819)(0,05 - 0,10)}{0,895 - 0,819} & 0,819 \leq D^* < 0,895 \\ 0,05 + \frac{(D^* - 0,895)(0,025 - 0,05)}{0,995 - 0,895} & 0,895 \leq D^* < 0,995 \\ 0,025 + \frac{(D^* - 0,995)(0,01 - 0,025)}{1,035 - 0,995} & 0,995 \leq D^* < 1,035 \\ < 0,01 & D^* > 1,035 \end{cases} \tag{8}$$

onde  $p$  é o p-valor do teste.

<sup>1</sup>O p-valor, descreve a probabilidade de se obter uma estatística de teste igual ou maior do que aquela observada em uma amostra, tomando como base a hipótese nula.

**D. Teste de Kolmogorov-Smirnov, método de Marsaglia**

Este método, proposto por Marsaglia, Tsang e Wang [14], utiliza uma matriz quadrada  $\mathbf{H}$  de dimensão  $m = 2k - 1$ . A equação para o cálculo da estatística é dada por:

$$P[D_n < d] = \frac{n!}{n^n} t_{kk} \tag{9}$$

onde  $d = (k - h)/n$ ,  $k$  é um número positivo e  $0 \leq h < 1$ .

Os valores utilizados no cálculo são obtidos através do algoritmo 1.

**Algorithm 1** Método de Marsaglia

```

1:  $k =$  Arredondar para cima  $nD_n$ 
2:  $m = 2k - 1$ 
3:  $h = k - nD_n$ 
4: for  $i=1:m-1$  do
5:    $\mathbf{H}(i, 1) = (1 - h_i)/i!$ 
6: end for
7:  $\mathbf{H}(m, c) = \mathbf{H}(r, 1)^T, r = 1, \dots, m - 1, c = m - r + 1$ 
8: for  $i = 1 : m - 1$  do
9:   for  $j = 2 : m$  do
10:    if  $i - j + 1 \geq 0$  then
11:       $\mathbf{H}(i, j) = 1/(i - j + 1)!$ 
12:    else
13:       $\mathbf{H}(i, j) = 0$ 
14:    end if
15:  end for
16: end for
17: if  $h \leq 0.5$  then
18:    $\mathbf{H}(m, 1) = (1 - 2h^m)/m!$ 
19: else
20:    $\mathbf{H}(m, 1) = (1 - 2h^m + \max(0.2h - 1)^m)/m!$ 
21: end if
22:  $p\text{-valor} = 1 - \frac{n!}{n^n} \mathbf{H}^n(k, k)$ 

```

**E. Teste de Kolmogorov-Smirnov, modificação de Lilliefors**

Hubert W. Lilliefors propôs uma correção ao método de Kolmogorov-Smirnov [15], para o caso onde os parâmetros da distribuição hipotética são desconhecidos. Posteriormente, Gerard E. Dallal e Leland Wilkinson, determinaram aproximações numéricas para os p-valores do método de Lilliefors [16] utilizando uma formulação analítica.

O método de Lilliefors é baseado no cálculo dos valores de  $D\alpha$ ,  $a_1$ ,  $b_1$ ,  $c_1$ ,  $a_2$ ,  $b_2$  e  $c_2$ . Dado o nível de significância  $\alpha$ , e seu respectivo valor crítico  $C_\alpha$ , considere  $D\alpha$ :

$$D\alpha = C_\alpha$$

As constantes  $a_1$ ,  $b_1$ ,  $c_1$ ,  $a_2$ ,  $b_2$  e  $c_2$  são calculadas com base nas seguintes equações:

$$a_1 = -7,01256(n + 2,78019) \tag{10}$$

$$b_1 = 2,99587\sqrt{n + 2,78019} \tag{11}$$

$$c_1 = 2,1804661 + 0.974598/\sqrt{n} + 1,67997/n \tag{12}$$

$$a_2 = -7,90289126054n^{0,98} \tag{13}$$

$$b_2 = 3,18037017572ln^{0,49} \tag{14}$$

$$c_2 = 2,2947256 \tag{15}$$

Para obtenção do valor crítico para  $\alpha = 0.10$ , é necessário avaliar a dimensão da amostra. Para  $n \leq 100$ , utiliza-se:

$$a = a_1, b = b_1, c = c_1$$

Caso contrário, para  $n > 100$ :

$$a = a_2, b = b_2, c = c_2$$

O valor crítico  $D10$  é dado por:

$$D10 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

O p-valor do teste ( $p$ ) é calculado utilizando a equação a seguir:

$$p = \begin{cases} 0,10 & D_n = D10 \\ e^{aD_n^2 + bD_n + c - 2,3025851} & D10 \leq D_n < D15 \\ 0,15 + (D_n - D15)[(0,10 - 0,15)/(D10 - D15)] & D15 \leq D_n < D20 \\ 0,20 + (D_n - D20)[(0,15 - 0,20)/(D15 - D20)] & D_n \geq D20 \\ > 0,20 & D_n \leq D10 \end{cases} \tag{16}$$

**F. Teste de Anderson-Darling**

O teste de Anderson-Darling é utilizado para determinar se a amostra segue uma distribuição contínua  $F(x)$  [17], utilizando a estatística de teste baseada na seguinte integral:

$$A^2 = n \int_{-\infty}^{\infty} \frac{|F_n(x) - F(x)|}{F(x)(1 - F(x))} dF(x) \tag{17}$$

onde  $F_n(x)$  é a distribuição acumulada empírica da amostra.

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ \frac{k}{n} & x_i \leq x < x_{i+1} \\ 1 & x \geq x_n \end{cases} \tag{18}$$

os valores  $x_i$ , são os valores ordenados da amostra  $X$ .

Utilizando os valores de  $U_i = F(x_i)$ , pode-se calcular a estatística de Anderson-Darling através da equação:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1)(\ln(U_i)) + (2(n - i) + 1) \ln(1 - U_i)] \tag{19}$$

Stephens [12], [13] propôs a seguinte equação para o cálculo da estatística  $A^2$ , com o intuito de diferenciar o tamanho das amostras:

$$A^{*2} = A^2 (1 + 0,75/n + 2,25/n^2) \tag{20}$$

O p-valor do teste pode ser calculado utilizando o algoritmo 2 [13].

**Algorithm 2** Método de Anderson-Darling

```

1:  $S = 0$ 
2: for  $i = 1 : n$  do  $S = S + (2i - 1)(\ln u_i + \ln(1 - u_i))$ 
3: end for
4:  $A = -n - (1/n) \cdot S$ 
5:  $A^* = A(1 + 0,75/n + 2,25/n^2)$ 
6: if  $A^* \leq 0,2$  then
7:   p - valor =  $\frac{1}{\exp(-13,436 + 101,14A^* - 223,73A^{*2})}$ 
8: else if  $A^* \leq 0,34$  then
9:   p - valor =  $\frac{1}{\exp(-8,318 + 42,798A^* - 59,938A^{*2})}$ 
10: else if  $A^* \leq 0,6$  then
11:   p - valor =  $\exp(0,9177 - 4,279A^* - 1,38^{*2})$ 
12: else if  $A^* \leq 153,467$  then
13:   p - valor =  $\exp(1,293 - 5,709A^* - 0,0186A^{*2})$ 
14: else
15:   p - valor = 0
16: end if

```

**Algorithm 3** Método de Cramer-Von Mises

```

1:  $S = 0$ 
2: for  $i = 1 : n$  do
3:    $S = S + (u_i - (2i - 1)/2n)^2$ 
4: end for
5:  $W = (1/12n) + S$ 
6:  $W^* = W(1 + 0,5/n)$ 
7: if  $W^* < 0,0275$  then
8:   p - valor =  $\frac{1}{\exp(13,953 + 775,5W^* - 12542,61W^{*2})}$ 
9: else if  $W^* \leq 0,051$  then
10:   p - valor =  $\frac{1}{\exp(-5,903 + 179,546W^* - 1515,29W^{*2})}$ 
11: else if  $W^* < 0,092$  then
12:   p - valor =  $\exp(0,886 - 31,62W^* + 10,897W^{*2})$ 
13: else
14:   p - valor =  $1,111 - 34,242W^* + 12,832W^{*2}$ 
15: end if

```

**G. Teste de Cramer-Von Mises**

Este método foi proposto de forma independente por Harald Cramér [18] e Richard Edler von Mises [19] e tem por objetivo avaliar a qualidade da aproximação comparando as funções de distribuição empírica e teórica.

A estatística do teste originalmente proposta  $\omega^2$  é dada por:

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x) \quad (21)$$

onde  $F_n(x)$  é a distribuição empírica da amostra e  $F(x)$  a distribuição teórica.

Considerando uma sequência  $x_i$  ordenada de valores da amostra  $X$ , a estatística do teste pode ser calculada utilizando a equação [20]:

$$T = n\omega^2 = \frac{1}{12n} + \sum_{i=1}^n \left[ \frac{2i-1}{2n} - F(x_i) \right]^2 \quad (22)$$

os valores de  $T$  são comparados com valores tabulados por Pearson e Hartley [21] e o cálculo do p-valor pode ser realizado utilizando o algoritmo 3, proposto por Stephens [12], [13]:

**H. Teste de Shapiro-Wilk**

O teste de Shapiro-Wilk [22] é utilizado para avaliar a hipótese de que os dados amostrais sejam oriundos de uma população normalmente distribuída. A estatística do teste é dada por:

$$W = \frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (23)$$

onde  $x_i$  é uma sequência ordenada dos valores da amostra  $X$ , e  $\bar{x}$  é a sua média amostral.

Os coeficientes  $a_i$  são elementos do seguinte vetor:

$$\mathbf{a} = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{(\mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m})^{1/2}} \quad (24)$$

onde  $\mathbf{m} = [m_1, m_2, \dots, m_n]$  é um vetor de valores esperados baseados em uma distribuição normal e  $\mathbf{V}$  é a matriz de covariância desses valores. O vetor  $\mathbf{m}$  também é conhecido como *Blom-score*. Para uma amostra de tamanho  $n$  e posto  $r$ , o vetor  $\mathbf{m}$  é definido por:

$$\mathbf{m} = \Phi \left( \frac{r - 3/8}{n + 1/4} \right) \quad (25)$$

O algoritmo 4 [13] mostra uma implementação baseada na aproximação de Royston para aplicação do teste de Shapiro-Wilk [23].

---

**Algorithm 4** Método de Shapiro-Wilk

---

```

1: for  $i = 1 : n$  do
2:    $\hat{m}_i = \phi^{-1} \left( \frac{i-0,375}{n+0,25} \right)$ 
3: end for
4:  $m = \hat{\mathbf{m}}^T \hat{\mathbf{m}}$ 
5:  $u = 1/\text{sqrtn}$ 
6:  $a_n = -2,70606u^5 + 4,43469u^4 - 2,07119u^3 - 0,14798u^2 + 0,22116u + \hat{m}_n/\sqrt{m}$ 
7:  $a_{n-1} = -3,58263u^5 + 5,68263u^4 - 1,75246u^3 - 0,29376u^2 + 0,04298u + \hat{m}_n/\sqrt{m}$ 
8:  $a_1 = -a_n$ 
9:  $a_2 = -a_{n-1}$ 
10: for  $i = 3 : (n-2)$  do
11:    $\epsilon = \frac{m-2\hat{m}_n^2-2\hat{m}_{n-1}^2}{1-sa_n^2-2a_{n-1}^2}$ 
12:    $a_i = \hat{m}_i/\sqrt{\epsilon}$ 
13: end for
14:  $x = \ln(n)$ 
15:  $\mu = 1,5861 - 0,31082x - 0,083751x^2 + 0,0038915x^3$ 
16:  $\sigma = \exp(-0,4803 - 0,082676x + 0,0030302x^2)$ 
17:  $W = \frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ 
18:  $Z = \frac{\ln(1-W) - \mu}{\sigma}$ 
19: p - valor =  $P[Z \geq z]$ 

```

---

*I. Teste de Shapiro-Francia*

Esse teste foi proposto por S. S. Shapiro e R. S. Francia em 1972, como uma alternativa de simplificação ao teste de Shapiro-Wilk [24].

A estatística do teste de Shapiro-Francia é baseada no coeficientes de correlação de Pearson entre os dados amostrais e as estatísticas teóricas da distribuição normal, e seu cálculo é realizado através da equação 26 [13].

$$W' = \frac{(\sum_{i=1}^n b_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (26)$$

onde  $y_i$  é uma seqüência ordenada de estatísticas e  $b_i$  são os componentes do vetor  $\mathbf{b}$ , definido com base em uma aproximação do vetor  $\mathbf{m}$  (equação 25).

$$\mathbf{b} = \frac{\hat{\mathbf{m}}^T}{\sqrt{\hat{\mathbf{m}}^T \hat{\mathbf{m}}}} \quad (27)$$

Royston [25] propôs uma aproximação para o cálculo dos p-valores do teste de Shapiro-Francia, conforme descrito no algoritmo 5.

---

**Algorithm 5** Método de Shapiro-Francia

---

```

1:  $y_i =$  ordenação dos valores da amostra  $X$ 
2: for  $i = 1 : n$  do
3:    $\hat{m}_i = \phi^{-1} \left( \frac{i-0,375}{n+0,25} \right)$ 
4: end for
5:  $W' = \frac{(\sum_{i=1}^n \hat{m}_i X_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n \hat{m}_i^2}$ 
6:  $u = \ln(n)$ 
7:  $v = \ln(u)$ 
8:  $\mu = -1,2725 + 1,0521(v - u)$ 
9:  $\sigma = 1,0308 + 0,26758(v + 2/u)$ 
10:  $Z = \frac{\ln(1-W') - \mu}{\sigma}$ 
11: p - valor =  $P[Z \geq z]$ 

```

---

*J. Teste de Jarque-Bera*

O teste de Jarque-Bera[26] é um teste de qualidade da aproximação baseado em momentos. A estatística do teste, equação (28), tem por objetivo comparar os momentos amostrais com as expectativas de momentos de uma distribuição normal.

$$JB = \frac{n}{6} \left( S^2 + \frac{1}{4} (K - 3)^2 \right) \quad (28)$$

As quantidades  $S$  e  $K$  na equação 28 são o coeficiente de assimetria e curtose amostrais, respectivamente. O valor de  $JB$  é sempre positivo e quanto mais afastado de zero maior a evidência de que os dados não seguem uma distribuição normal.

Os p-valores do teste são obtidos de uma distribuição  $\chi^2$  com dois graus de liberdade.

*K. Teste de D'Agostino-Pearson*

Proposto por Ralph D'Agostino e E. S. Pearson [27], esse é também um teste baseado em momentos que mede a perda de normalidade (*departure from normality*) nos dados amostrais.

A estatística do teste é dada por:

$$DAP = Z^2(g_1) + Z^2(g_2) \quad (29)$$

onde  $g_1$  é o coeficiente de assimetria amostral e  $g_2$  é o coeficiente de curtose amostral ( $K - 3$ ).

A estatística do teste segue uma distribuição assintótica  $\chi^2$  com dois graus de liberdade, que também é utilizada para o cálculo dos p-valores.

*L. Aplicação dos testes de normalidade*

A aplicação dos 10 diferentes testes de normalidade para os 64 ativos da base de dados está resumida no gráfico da figura 5, onde são mostrados os valores médios dos p-valores, obtidos pelos 10 testes para a série de retornos de cada um dos ativos da base de dados.

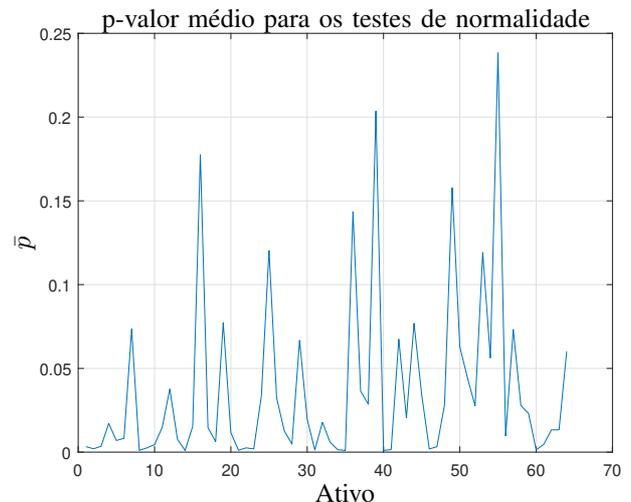


Fig. 5: p-valor médio:  $\bar{p}$ , para 10 diferentes testes de normalidade

O gráfico mostra que nenhum dos ativos obteve um p-valor superior à 0,25, sendo que a maioria fica na faixa de rejeição,

abaixo de 0,05 para o nível de significância adotado,  $\alpha = 5\%$ . O ativo com maior  $\bar{p}$  é o ativo SMLS3 ( $\bar{p} = 0,2383$ ).

A figura 6 mostra as razões de aceitação da hipótese de normalidade<sup>2</sup>, complementando assim a análise proposta na figura 5. Nele é possível verificar que nenhum ativo passa por todos os testes de normalidade, sendo o que obteve melhor resultado foi o ativo LREN3 ( $\bar{p} = 0,2036$ ), único a obter sucesso em 70% dos testes.

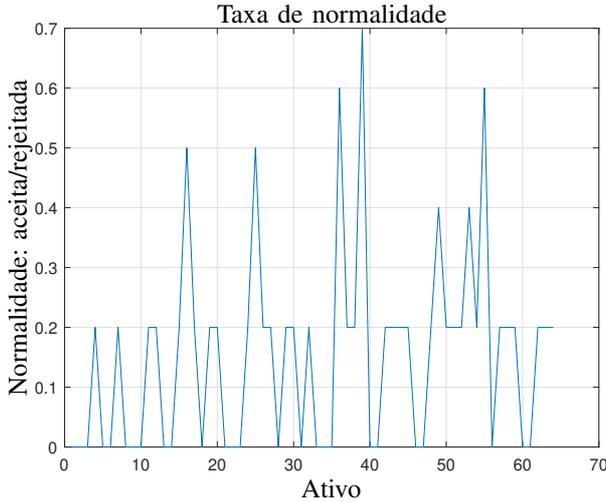


Fig. 6: Razão entre a quantidade de testes que aceitam e que rejeitam a normalidade

A tabela II traz um resumo dos gráficos das figuras 5 e 6. Se tomarmos como medida de resumo geral da base de dados o valor médio da coluna  $\bar{p}$ , da tabela II, é obtido o valor: 0,0373, que situa-se abaixo do nível crítico de rejeição da hipótese gaussiana. Apesar desse resultado não ter valor analítico de teste, ele é um parâmetro empírico útil para o julgamento proposto, demonstrando haver fortes evidências para rejeição da hipótese Gaussiana para os retornos dos ativos.

TABLE II: Resumo geral dos testes de normalidade

Teste	$\bar{p}$	$\zeta$
Kolmogorov-Smirnov (KS)	0,15195	0,59
KS Stephens Modification	0,02512	0,11
KS Marsaglia Method	0,14802	0,59
KS Lilliefors Modification	0,02137	0,11
Anderson-Darling Test	0,00968	0,05
Cramer-Von Mises Test	0,01203	0,08
Shapiro-Wilk Test	0,00276	0,02
Shapiro-Francia Test	0,00133	0,00
Jarque-Bera Test	0,00021	0,00
DAgostino & Pearson Test	0,00094	0,00

M. Outliers

Dentre os ativos, o que obteve a menor estatística no teste de Kolmogorov-Smirnov, portanto apresenta menor evidência de não normalidade, tomando esse critério como referência, é o ativo CSAN3. Os retornos desse ativo são mostrados no gráfico de controle da figura 7. Esse gráfico foi construído utilizando limites de controle de  $3,5S_X$ , e sugere a existência

<sup>2</sup>Valor calculado através da razão entre o número de testes onde obtêm-se sucesso e o número total de testes

de 3 outliers, localizados nos dia 04 de abril de 2016, 11 de novembro de 2016 e 18 de maio de 2017.

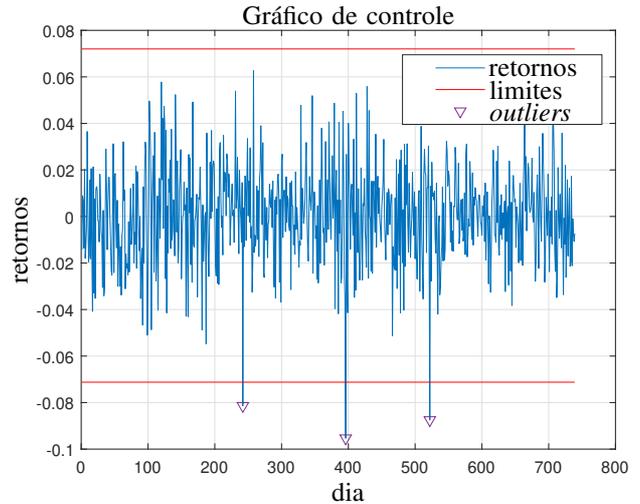


Fig. 7: Gráfico de controle para os retornos do ativo CSAN3

A saber, os pontos no gráfico de controle da figura 7, localizados externamente a linha de controle, ocorreram em datas onde o mercado sofreu fortes influências políticas, o que poderia justificar o comportamento, aparentemente anômalo, dos dados.

A existência desses pontos também fica explícita no gráfico de normalidade, conforme figura 8, onde nota-se a existência de três pontos distantes da linha de normalidade na parte inferior do gráfico.

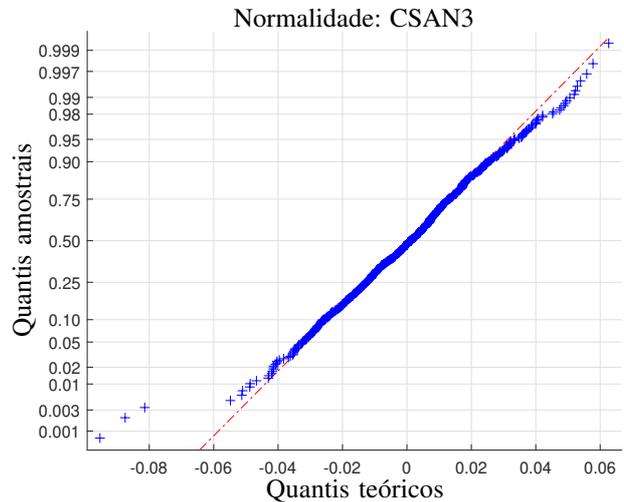


Fig. 8: CSAN3, Gráfico de normalidade da série de retornos.

1) *Comparação entre as séries ajustada e original:* A série ajustada dos retornos para o ativo CSAN3 é aquela onde esses 3 outliers foram removidos dos dados originais. Nessa condição são obtidos os seguintes parâmetros para a distribuição empírica dos retornos:

- média  $\mu_r = 7,5557 \cdot 10^4 \cong 0$
- Coeficiente de assimetria: 0,0869, relativamente também muito pequeno.
- Curtose: 3,0447, próximo a uma distribuição normal.

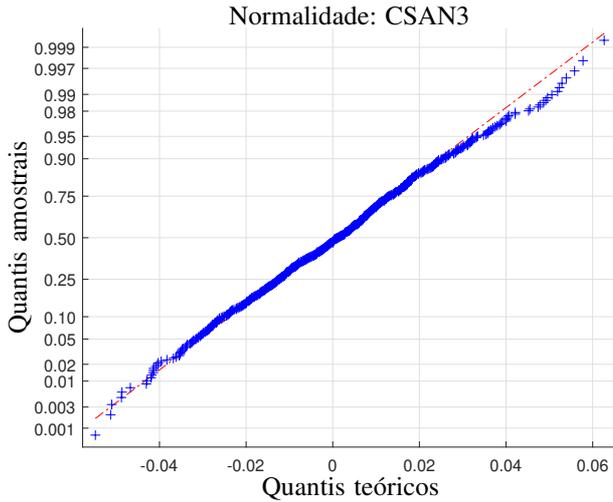


Fig. 9: CSAN3, Gráfico de normalidade da série ajustada de retornos.

O gráfico da figura 9 indica evidências de que os dados sejam provenientes de uma distribuição normal, corroborando com o resultado do teste de Kolmogorov-Smirnov.

Adicionalmente, a aplicação de 10 diferentes testes de normalidade a ambas as séries de retornos, original e ajustada, do ativo CSAN3, confirma a hipótese de normalidade para um nível de significância de 5% apenas para a série ajustada, conforme mostrado nas tabelas III e IV.

TABLE III: Testes de normalidades dos retornos do ativo CSAN3

Teste	Estatística	p-valor	Resultado
Kolmogorov-Smirnov (KS)	0,7290	0,66252	Não rejeitar $H_0$
KS Método de Stephens	0,7296	0,15000	Não rejeitar $H_0$
KS Método de Marsaglia	0,7290	0,65235	Não rejeitar $H_0$
KS Método de Lilliefors	0,0268	0,20000	Não rejeitar $H_0$
Anderson-Darling	0,7177	0,04245	Rejeitar $H_0$
Cramer-Von Mises	0,1160	0,06799	Não Rejeitar $H_0$
Shapiro-Wilk	0,9903	0,00009	Rejeitar $H_0$
Shapiro-Francia	0,9895	0,00008	Rejeitar $H_0$
Jarque-Bera	40,5870	0,00000	Rejeitar $H_0$
DAgostino & Pearson	22,7867	0,00001	Rejeitar $H_0$

TABLE IV: Testes de normalidades dos retornos ajustados do ativo CSAN3

Teste	Estatística	p-valor	Resultado
Kolmogorov-Smirnov (KS)	0,5959	0,86969	Não rejeitar $H_0$
KS Método de Stephens	0,5964	0,15000	Não rejeitar $H_0$
KS Método de Marsaglia	0,5959	0,86178	Não rejeitar $H_0$
KS Método de Lilliefors	0,0220	0,20000	Não rejeitar $H_0$
Anderson-Darling	0,3859	0,39098	Não rejeitar $H_0$
Cramer-Von Mises	0,0595	0,38409	Não rejeitar $H_0$
Shapiro-Wilk	0,9976	0,37814	Não rejeitar $H_0$
Shapiro-Francia	0,9980	0,46682	Não rejeitar $H_0$
Jarque-Bera	0,9882	0,61011	Não rejeitar $H_0$
DAgostino & Pearson	1,0847	0,58137	Não rejeitar $H_0$

Utilizando como medida de resumo dos 10 testes, o valor médio dos p-valores:  $\bar{p} = 0,4883 \gg 0,5$  e o percentual de testes com sucesso  $\varsigma$ , são obtidos os seguintes resultados:

- Série ajustada:  $\bar{p} = 0,4883$ ,  $\varsigma = 100\%$
- Série original:  $\bar{p} = 0,1775$ ,  $\varsigma = 50\%$

Fica demonstrado que existem evidências relativamente fortes de normalidade para a série de retornos ajustados, não

podendo ser feita a mesma afirmação para a série original. Esse resultado, por outro lado, confirma a existência de amostras que não podem ser explicadas através da hipótese Gaussiana.

2) *Análise dos extremos das caudas*: Pretendendo verificar a capacidade da distribuição ajustada de representar os valores extremos da série temporal (caudas da distribuição de probabilidades), é apresentada a tabela V, que mostra resultado da avaliação das frequências dos pontos extremos.

TABLE V: Frequências esperadas e observadas com base na distribuição normal (retornos CSAN3)

$\alpha$	Freq. esperada	Freq. observada	Erro
0,005	4	12	2,165088
0,010	7	20	1,759134
0,015	11	23	1,082544
0,020	15	27	0,811908
0,025	18	35	0,920162
0,050	37	55	0,487145
0,100	74	100	0,351827
0,200	148	173	0,169147

Tomando como medida de resumo deste teste o erro médio  $\bar{e}$ , dado pela definição 1.

*Definição 1 (Erro de aproximação das caudas  $\bar{e}$ ):*

Considerando a diferença entre as frequências esperada:  $\phi_e(\alpha)$ , e realizada  $f_r(\alpha)$ , para os valores extremos de uma amostra de tamanho  $n$ , ambas definidas por:

$$\phi_e(\alpha) = \alpha \cdot n$$

$$f_r(\alpha) = freq(X \geq F_X^{-1}(1 - \alpha/2) \cup X \leq F_X^{-1}(\alpha/2))$$

onde  $freq(A)$  é a contagem do número de eventos no conjunto  $A$  e  $F_X(x)$  é a CDF teórica. O erro de aproximação das caudas é então definido por:

$$\bar{e} = \frac{1}{3} \left[ \frac{abs(\phi_e(0,005) - f_r(0,005))}{\phi_e(0,005)} + \frac{abs(\phi_e(0,010) - f_r(0,010))}{\phi_e(0,010)} + \frac{abs(\phi_e(0,015) - f_r(0,015))}{\phi_e(0,015)} \right] \quad (30)$$

Da equação (30) obtém-se o valor:  $\bar{e} = 1,6689$ .

### N. Distribuição $\alpha$ -estável

A fim de avaliar a hipótese de Mandelbrot, foram estimadas distribuições  $\alpha$ -estáveis para todos os ativos da base de dados, utilizando para tanto o método da máxima verossimilhança na obtenção dos parâmetros dessas distribuições.

1) *Análise do parâmetro  $\alpha$* : Para uma distribuição  $\alpha$ -estável, o parâmetro  $0 < \alpha < 2$  representa o nível de concentração das observações no entorno da média, sendo que para  $\alpha = 2$ , temos uma distribuição gaussiana. Esse parâmetro pode também ser tomado como medida da perda de normalidade dos dados, pois o distanciamento do parâmetro  $\alpha$  em relação ao valor 2 é proporcional ao afastamento da hipótese Gaussiana.

A figura 10 mostra os valores de  $\alpha$  obtidos para cada um dos ativos da base de dados.

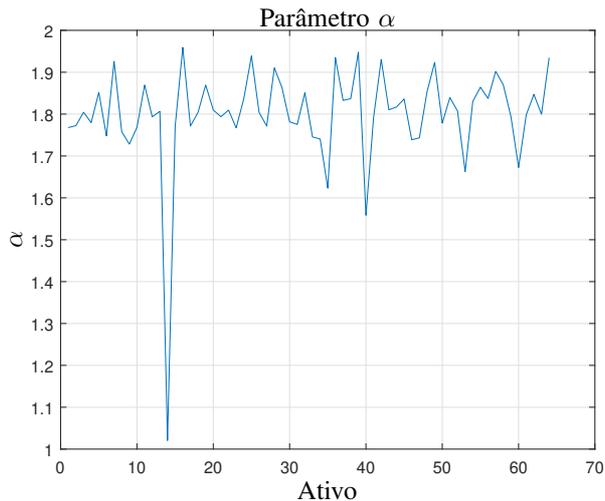


Fig. 10: Parâmetro  $\alpha$  das distribuições  $\alpha$ -estáveis.

O gráfico da figura 10 mostra que todos os ativos obtiveram um valor inferior a 2 para o parâmetro  $\alpha$ , o que corrobora com a hipótese de leptocurtose das distribuições empíricas e consequente perda de normalidade. A única discrepância é relacionada com o ativo CPFE3, que apresenta um comportamento anômalo na sua série de retornos, conforme mostra o gráfico da figura 11.

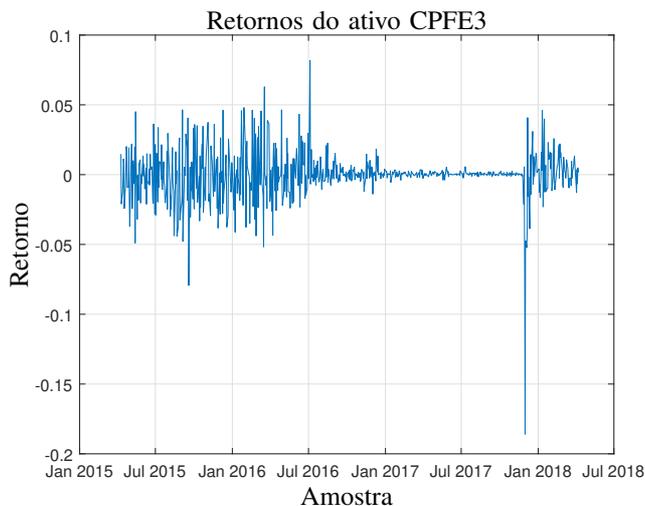


Fig. 11: Retornos do ativo CPFE3.

Supondo que esse comportamento não possa ser explicado apenas pelas séries temporais, recorre-se então a uma análise dos fatos históricos, envolvendo a companhia CPFL Energia.

Em julho de 2016 a chinesa State Grid, maior empresa do setor elétrico mundial, anunciou a compra de 54,6% da CPFL Energia. A negociação, porém, só seria concluída em novembro de 2017, no dia da OPA (Oferta Pública de Ações), quando a State Grid pagou R\$ 27,69 por cada ação da empresa, cerca de 35% a mais do que o seu valor de mercado à época.

A operação iniciada em julho de 2016 desenquadrando o ativo CPFE3 dos 25% de ações do "free float", gerando uma situação que forçou a State Grid a vender 20% das ações CPFE3, que acabara de comprar, cotadas na ocasião acerca de 13% abaixo do valor pago pela empresa, ou negociar os 5,25% restantes.

Esse período, entre julho de 2016 e novembro de 2017, marcou o ativo CPFE3 pela incerteza e desinteresse dos investidores, refletindo em um comportamento altamente anômalo da série de retornos. Essa conclusão fica evidenciada nos dados pela análise do parâmetro  $\alpha$  das distribuições  $\alpha$ -estáveis.

2) *Análise dos p-valores médios:* O comportamento anômalo dos retornos também se reflete no gráfico dos p-valores médios, mostrado na figura 12. Nesse gráfico pode-se observar que o p-valor médio obtido para o ativo CPFE3 é inferior a significância (0,05), portanto, rejeita-se a hipótese de Mandelbrot para essa série de retornos.

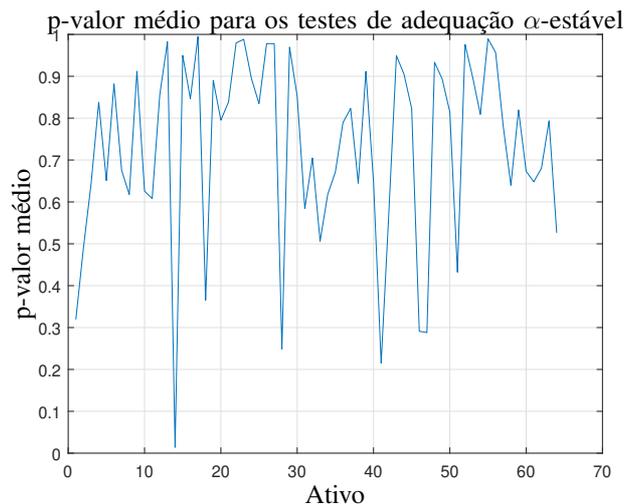


Fig. 12: p-valor médio para os testes KS e CvM

Para criar uma base de comparação consistente, foi considerada a aplicação dos testes de qualidade da aproximação, KS e CvM, para as funções de densidade Gaussianas, da mesma forma como foi aplicado para as distribuições  $\alpha$ -estáveis. O resultado é mostrado no gráfico da figura 13.

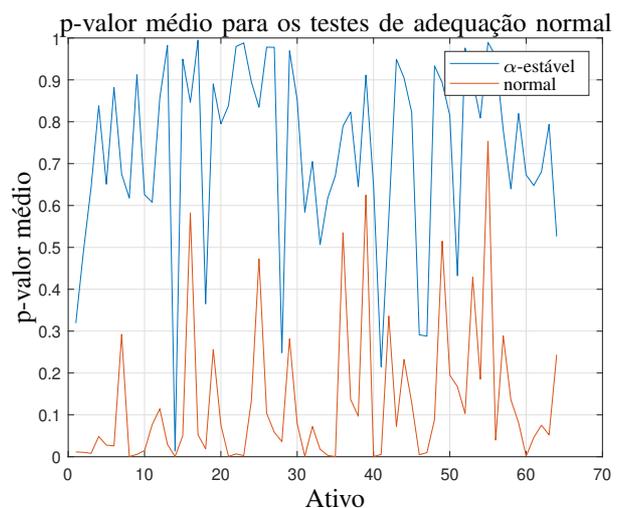


Fig. 13: p-valor médio para os testes KS e CvM, obtido através de uma aproximação normal

O gráfico demonstra que os p-valores obtidos pela aproximação  $\alpha$ -estável são todos superiores aos da aproximação Gaussiana, portanto, concluí-se novamente

que a hipótese de Mandelbrot prevalece sobre a hipótese Gaussiana. Completando essa análise, o p-valor médio geral, obtido utilizando distribuições  $\alpha$ -estáveis foi: 0,7301 e para as distribuições Gaussianas foi obtido o valor 0,1338, substancialmente menor.

Ainda com relação aos p-valores, o único ativo para o qual foi rejeitada a hipótese de que os dados tenham sido extraídos de uma população  $\alpha$ -estável é o ativo CPFE3. Ambos os testes obtiveram um p-valor menor do que o nível de significância para esse ativo. Esse resultado carrega em si duas conclusões possíveis, uma delas é que se trata de mais uma evidência da anomalia de comportamento atribuída aos retornos do ativo CPFE3, outra é que as distribuições  $\alpha$ -estáveis não são capazes de modelar todos os possíveis comportamentos dos ativos.

O. Análise dos erros de representação das caudas

A figura 14 traz uma análise comparativa entre os erros  $\bar{e}$  para os valores extremos, cometidos pelas distribuições normal e  $\alpha$ -estável.

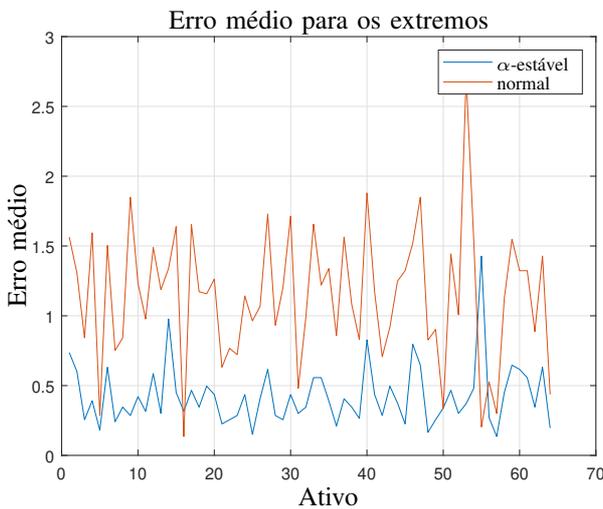


Fig. 14: Erros médios para os pontos extremos, cometidos pela aproximação normal

Analisando o gráfico da figura 14, pode-se observar que existem apenas dois ativos para os quais o erro  $\bar{e}$  é menor para a distribuição normal, sendo eles: CSAN3 e SMLS3. Para os demais a distribuição  $\alpha$ -estável supera a distribuição normal nesse critério, obtendo um erro médio geral de 0,4259, bastante inferior ao resultado da distribuição normal que é 1,1432.

P. Testes de adequação à distribuição 2-Gaussiana

Para todos os 64 ativos da base de dados foram estimados os parâmetros da distribuição 2-Gaussiana [28] para as séries de retornos, utilizando o método da máxima verossimilhança. Nesta sessão, suas propriedades e parâmetros são avaliados relativamente aos das demais distribuições.

1) Análise dos p-valores: Considerando a aplicação dos testes de Kolmogorov-Smirnov e Cramer-Von Mises, foram obtidos os p-valores médios para cada um dos ativos. O resultado é resumido no gráfico da figura 15.

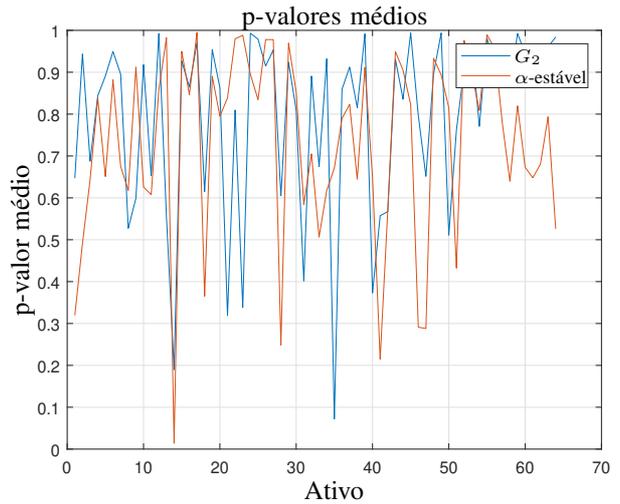


Fig. 15: p-valores médios para os testes KS e CvM, obtidos para as distribuições 2-Gaussiana e  $\alpha$ -estável

O gráfico da figura 15 mostra que há semelhança entre os resultados obtidos para as duas distribuições, tendo a distribuição 2-Gaussiana obtido resultados relativamente superiores.

Quantitativamente, a média dos p-valores para a distribuição 2-Gaussiana é 0,7909 enquanto para a distribuição  $\alpha$ -estável é 0,7301, sendo então 8,33% superior. O número de ativos para os quais a distribuição 2-Gaussiana apresenta um p-valor superior, representa 62,5% do total de ativos na base de dados. Por fim, os testes não rejeitaram a hipótese de que a distribuição empírica é 2-Gaussiana para nenhum dos ativos.

Q. Análise das caudas

A capacidade de representar os extremos das caudas, medida aqui pelo valor de  $\bar{e}$  também obteve resultados melhores para a distribuição 2-Gaussiana, conforme pode-se observar no gráfico da figura 16, cuja a análise quantitativa revela que o valor médio dos erros obtidos pela distribuição 2-Gaussiana foi de 0,3249. Esse valor apresenta uma diferença de 31,06% em relação ao valor obtido para a distribuição  $\alpha$ -estável, a favor da distribuição 2-Gaussiana. Para a totalidade dos ativos, o erro  $\bar{e}$  da distribuição 2-Gaussiana é inferior ao da distribuição  $\alpha$ -estável para 62,63% dos ativos analisados.

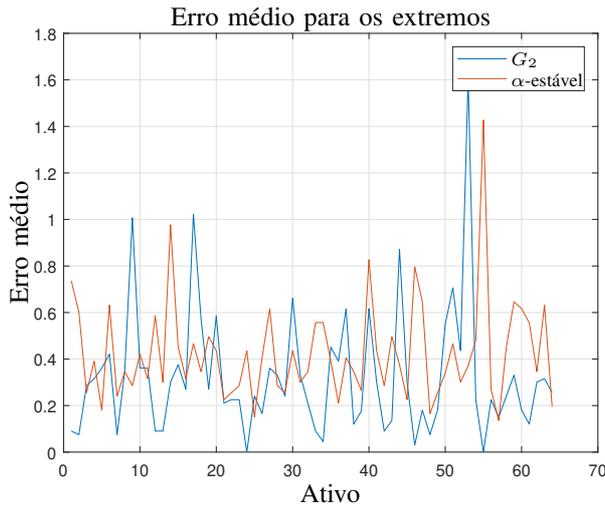


Fig. 16: Erros médios para os pontos extremos, cometidos pelas distribuições 2-Gaussiana e  $\alpha$ -estável

#### IV. PARÂMETROS DAS DISTRIBUIÇÕES

As tabelas VI, VII e VIII mostram as distribuições teóricas aproximadas a partir das distribuições empíricas dos dados, juntamente com algumas de suas propriedades mais relevantes.

Na tabela VI são caracterizadas as distribuições Gaussianas. Nela são mostradas as médias  $\mu$  e desvios padrão  $\sigma$ , estimados para cada ativo. Para essas distribuições os momentos de ordens mais elevadas são constantes e independem da média e da variância, sendo a assimetria nula e a curtose igual a 3.

As distribuições  $\alpha$ -estáveis são mostradas na tabela VII, onde estão descritos os quatro parâmetros populacionais que as compõem,  $\alpha$ ,  $\beta$ ,  $\gamma$  e  $\delta$ , juntamente com a média populacional, calculada a partir do primeiro momento da distribuição. Como a distribuição  $\alpha$ -estável não tem variância finita, os momentos de ordem superior foram excluídos da análise.

Na tabela VIII são mostrados os parâmetros das distribuições 2-Gaussiana ( $\mu_1$ ,  $\sigma_1$ ,  $\mu_2$  e  $\sigma_2$ ) enquanto na tabela IX são mostrados os quatro primeiros momentos da distribuição, calculados através das equações:

$$Var[X] = E[X^2] - E[X]^2$$

$$Assimetria = \frac{E[(X - E[X])^3]}{Var[X]^{2/3}}$$

$$Curtose = \frac{E[(X - E[X])^4]}{Var[X]^2}$$

A análise das tabelas mostra como as distribuições 2-normais flexibilizam a curtose e a assimetria para se adequar aos dados amostrais, mantendo, porém, a média muito próxima da média amostral, conclusão essa que pode ser observada comparando o parâmetro  $\mu$  das distribuições Gaussianas (estimado pela média amostral), com o valor esperado das distribuições 2-normais.

Para as distribuições  $\alpha$ -estáveis, apesar de modelarem de forma mais adequada a leptocurtose e as caudas pesadas, pode-se observar uma discrepância maior entre as médias populacionais e os valores amostrais.

TABLE VI: Distribuições normais dos retornos

Ativo	$\mu$	$\sigma$
ABEV3	0,00024	0,01234
B3SA3	0,00109	0,01932
BBAS3	0,00069	0,03003
BBDC3	0,00034	0,01996
BBDC4	0,00056	0,02106
BBSE3	-0,00023	0,02154
BRAP4	0,00146	0,03388
BRFS3	-0,00144	0,02070
BRKM5	0,00178	0,02617
BRML3	-0,00008	0,02174
CCRO3	-0,00040	0,02281
CIEL3	-0,00043	0,01912
CMIG4	-0,00067	0,03263
CPFE3	0,00023	0,01712
CPLE6	-0,00040	0,02541
CSAN3	0,00040	0,02046
CSNA3	0,00059	0,04405
CYRE3	0,00012	0,02271
ECOR3	0,00000	0,02722
EGIE3	0,00010	0,01364
ELET3	0,00164	0,03587
ELET6	0,00156	0,03093
EMBR3	-0,00014	0,02252
ENBR3	0,00027	0,01736
EQTL3	0,00112	0,01397
ESTC3	0,00078	0,03072
FIBR3	0,00059	0,02603
FLRY3	0,00159	0,01848
GGBR4	0,00069	0,03408
GOAU4	-0,00046	0,04076
HYPE3	0,00070	0,01828
IGTA3	0,00028	0,01771
ITSA4	0,00058	0,01879
ITUB4	0,00068	0,01985
JBSS3	-0,00078	0,03580
KLBN11	0,00020	0,01848
KROT3	0,00017	0,02726
LAME4	0,00037	0,02257
LREN3	0,00103	0,02002
MGLU3	0,00403	0,04661
MRFG3	0,00047	0,02326
MRVE3	0,00080	0,01942
MULT3	0,00021	0,01757
NATU3	0,00018	0,02495
PCAR4	-0,00052	0,02255
PETR3	0,00109	0,03186
PETR4	0,00094	0,03245
QUAL3	-0,00008	0,02718
RADL3	0,00106	0,01763
RAIL3	0,00169	0,02593
RENT3	0,00117	0,02241
SANB11	0,00126	0,02217
SAPR11	-0,00020	0,01416
SBSP3	0,00087	0,02153
SMLS3	-0,00230	0,02483
SUZB3	0,00134	0,02606
TAE11	0,00003	0,01748
TIMP3	0,00046	0,02059
UGPA3	0,00006	0,01389
USIM5	0,00111	0,04491
VALE3	0,00112	0,03437
VIVT4	0,00009	0,01586
VVAR11	0,00075	0,03756
WEGE3	0,00040	0,01862

TABLE VII: Distribuição  $\alpha$ -estável dos retornos

Ativo	$\alpha$	$\beta$	$\gamma$	$\delta$	$E[X]$
ABEV3	1,767793	-0,176771	0,007598	0,000604	0,000092
B3SA3	1,772329	0,292508	0,011869	0,000018	0,001316
BBAS3	1,804339	0,213597	0,017466	-0,000057	0,001127
BBDC3	1,779584	-0,089983	0,011896	0,000605	0,000219
BBDC4	1,851926	-0,036389	0,012993	0,000649	0,000537
BBSE3	1,747989	0,063161	0,012814	-0,000428	-0,000090
BRAP4	1,926020	0,150646	0,022891	0,000995	0,001398
BRFS3	1,757995	-0,106947	0,011526	-0,000801	-0,001294
BRKM5	1,728290	0,363309	0,014886	-0,000110	0,002349
BRML3	1,768459	0,245823	0,012989	-0,000775	0,000440
CCRO3	1,869376	-0,641388	0,014328	0,001401	-0,000511
CIEL3	1,793401	0,128929	0,011913	-0,000792	-0,000275
CMIG4	1,806683	0,079832	0,019385	-0,000869	-0,000384
CPF3	1,020408	0,024074	0,004382	0,000235	0,003524
CPL6	1,770282	-0,036749	0,015360	0,000014	-0,000227
CSAN3	1,959159	-0,730123	0,013952	0,001059	0,000405
CSNA3	1,771406	0,226225	0,026475	-0,001215	0,001033
CYRE3	1,804082	0,009028	0,013599	0,000383	0,000422
ECOR3	1,869388	-0,024126	0,017722	0,000220	0,000131
EGIE3	1,809802	0,010628	0,008486	0,000054	0,000082
ELET3	1,793800	0,487913	0,019964	-0,001265	0,002005
ELET6	1,809531	0,570953	0,018289	-0,001131	0,002089
EMBR3	1,767172	0,003085	0,012399	-0,000036	-0,000021
ENBR3	1,836735	-0,114197	0,011109	0,000623	0,000291
EQTL3	1,938922	-0,682210	0,009545	0,001658	0,001032
ESTC3	1,804082	-0,151080	0,018725	0,001243	0,000344
FIBR3	1,771429	-0,063661	0,015878	0,000991	0,000612
FLRY3	1,910496	0,680884	0,012338	0,000576	0,001765
GGBR4	1,863260	0,247272	0,022244	-0,000227	0,000973
GOAU4	1,781602	-0,031165	0,024687	-0,000179	-0,000454
HYPE3	1,775308	0,076104	0,009903	0,000638	0,000916
IGTA3	1,851086	0,064026	0,011027	0,000330	0,000498
ITSA4	1,745405	0,106364	0,011140	0,000315	0,000816
ITUB4	1,740573	0,114310	0,011605	0,000249	0,000822
JBSS3	1,623071	0,255846	0,017109	-0,002645	0,000299
KLBN11	1,934694	0,111960	0,012618	0,000072	0,000218
KROT3	1,832777	0,307646	0,017375	-0,001037	0,000401
LAME4	1,836735	-0,155143	0,014117	0,001043	0,000468
LREN3	1,947783	-0,474103	0,013733	0,001470	0,000934
MGLU3	1,558438	0,133826	0,023144	0,001384	0,003959
MRFG3	1,792054	0,425669	0,014399	-0,001199	0,000877
MRVE3	1,930509	-0,165158	0,013177	0,001038	0,000800
MULT3	1,810323	0,385271	0,010850	-0,000712	0,000572
NATU3	1,816883	0,263224	0,015323	-0,000835	0,000358
PCAR4	1,836223	0,155071	0,014261	-0,000782	-0,000200
PETR3	1,738776	0,125203	0,019343	-0,000056	0,000997
PETR4	1,743293	0,030361	0,019463	0,000653	0,000905
QUAL3	1,852623	-0,226068	0,016767	0,000929	0,000036
RADL3	1,923374	0,498209	0,011950	0,000501	0,001221
RAIL3	1,778301	0,484121	0,014535	0,000690	0,003245
RENT3	1,839494	0,059870	0,014534	0,001054	0,001278
SANB11	1,807236	-0,128591	0,013694	0,001837	0,001287
SAPR11	1,662825	0,369464	0,008050	-0,001327	0,000414
SBSP3	1,830285	-0,103925	0,013619	0,001366	0,000980
SMLS3	1,864195	-0,801961	0,015808	0,000165	-0,002581
SUZB3	1,837374	0,498397	0,016154	-0,000544	0,001559
TAE11	1,902041	-0,509600	0,011591	0,000729	-0,000187
TIMP3	1,869388	-0,074364	0,013541	0,000641	0,000431
UGPA3	1,793610	-0,080000	0,008676	0,000273	0,000040
USIM5	1,672366	0,159633	0,025234	-0,000528	0,001750
VALE3	1,798139	0,166633	0,021629	0,000275	0,001457
VIVT4	1,847031	-0,157927	0,010187	0,000485	0,000090
VVAR11	1,799644	0,311710	0,023447	-0,001439	0,000940
WEGE3	1,934693	-0,257006	0,012541	0,000758	0,000426

TABLE VIII: Distribuições 2-Gaussiana dos retornos

Ativo	$\mu_1$	$\sigma_1$	$\mu_2$	$\sigma_2$
ABEV3	0,00126	0,00999	-0,00078	0,02247
B3SA3	-0,0011	0,01577	0,00332	0,03505
BBAS3	-0,00019	0,024	0,00164	0,05574
BBDC3	0,00111	0,01676	-0,00047	0,03647
BBDC4	0,00045	0,018	0,00068	0,0383
BBSE3	-0,00088	0,01756	0,00044	0,03956
BRAP4	0,00236	0,05849	0,00056	0,03407
BRFS3	-0,00288	0,03907	-0,00011	0,015
BRKM5	0,00466	0,04816	-0,0009	0,02124
BRML3	-0,00184	0,0175	0,00173	0,03986
CCRO3	0,00178	0,0219	-0,00268	0,04003
CIEL3	-0,00139	0,01639	0,00056	0,03456
CMIG4	-0,00053	0,02811	-0,00082	0,0596
CPF3	0,00021	0,00285	0,00024	0,03271
CPL6	-0,00006	0,02146	-0,00074	0,04631
CSAN3	0,00095	0,02205	-0,00016	0,03446
CSNA3	0,00424	0,07971	-0,00292	0,038
CYRE3	-0,00085	0,01883	0,00113	0,04156
ECOR3	-0,00028	0,0482	0,00028	0,02542
EGIE3	0,00039	0,02442	-0,00019	0,01237
ELET3	-0,00196	0,02799	0,00567	0,06711
ELET6	-0,00274	0,02542	0,00613	0,05638
EMBR3	-0,00019	0,01746	-0,00009	0,04248
ENBR3	-0,00029	0,03102	0,00082	0,01565
EQTL3	0,00007	0,02372	0,00218	0,01458
ESTC3	-0,00085	0,05539	0,00234	0,02718
FIBR3	0,00133	0,02265	-0,00018	0,04708
FLRY3	-0,00149	0,01725	0,00467	0,03207
GGBR4	-0,00171	0,03287	0,00311	0,05956
GOAU4	0,001	0,03614	-0,00197	0,07353
HYPE3	-0,00009	0,01318	0,00158	0,03483
IGTA3	-0,00014	0,01598	0,00071	0,03178
ITSA4	-0,00023	0,01504	0,00141	0,03459
ITUB4	0,00108	0,03692	0,00029	0,01513
JBSS3	0,00202	0,07018	-0,00313	0,02334
KLBN11	0,00019	0,03085	0,00022	0,02032
KROT3	-0,00265	0,02513	0,00302	0,04812
LAME4	0,00107	0,02043	-0,00034	0,04042
LREN3	0,00204	0,02171	0,00001	0,03355
MGLU3	0,00716	0,09003	0,00119	0,03
MRFG3	0,00377	0,04209	-0,00282	0,01868
MRVE3	0,00151	0,02036	0,00009	0,03304
MULT3	0,0018	0,03158	-0,00133	0,01549
NATU3	-0,00178	0,02357	0,00222	0,04409
PCAR4	0,00091	0,04012	-0,00191	0,02057
PETR3	0,00181	0,05893	0,00036	0,02386
PETR4	0,00165	0,02557	0,00023	0,05975
QUAL3	-0,00124	0,04895	0,00103	0,02426
RADL3	-0,00136	0,01858	0,00348	0,02955
RAIL3	-0,00058	0,0216	0,00419	0,04773
RENT3	0,00288	0,03953	-0,00053	0,02079
SANB11	0,00013	0,0399	0,00236	0,01965
SAPR11	0,00345	0,02531	-0,00399	0,0093
SBSP3	0,00159	0,02	0,00013	0,03822
SMLS3	0,00279	0,02301	-0,00754	0,04272
SUZB3	0,00507	0,04646	-0,00225	0,02308
TAE11	-0,00258	0,02999	0,00261	0,01725
TIMP3	0,00097	0,01881	-0,00005	0,03659
UGPA3	0,00033	0,01204	-0,00022	0,02506
USIM5	0,00515	0,08413	-0,00279	0,03206
VALE3	-0,00024	0,02999	0,00249	0,06189
VIVT4	0,00046	0,01432	-0,00029	0,02833
VVAR11	-0,00189	0,03218	0,00344	0,06784
WEGE3	0,00102	0,03259	-0,00022	0,01788

TABLE IX: Distribuições 2-Gaussiana dos retornos - propriedades

Ativo	$E[X]$	$\sigma_X$	Assimetria	Curtose
ABEV3	0,00024	0,01234	-0,16532	4,32919
B3SA3	0,00111	0,01934	0,22434	4,28551
BBAS3	0,00073	0,03036	0,06235	4,41444
BBDC3	0,00032	0,02008	-0,07665	4,26894
BBDC4	0,00057	0,02116	0,01032	4,22199
BBSE3	-0,00022	0,02165	0,06096	4,34694
BRAP4	0,00146	0,03386	0,03937	3,72919
BRFS3	-0,0015	0,02097	-0,14672	4,64173
BRKM5	0,00188	0,02646	0,20986	4,33391
BRML3	-0,00006	0,02184	0,16495	4,35577
CCRO3	-0,00045	0,02293	-0,15572	3,85565
CIEL3	-0,00042	0,01915	0,0964	4,19593
CMIG4	-0,00068	0,03295	-0,00839	4,21348
CPFE3	0,00023	0,01642	0,00204	5,91009
CPLE6	-0,0004	0,02553	-0,02589	4,25296
CSAN3	0,00039	0,02046	-0,03404	3,5265
CSNA3	0,00066	0,0443	0,15166	4,17388
CYRE3	0,00014	0,02284	0,0852	4,29886
ECOR3	0	0,02725	-0,01718	3,95686
EGIE3	0,0001	0,01369	0,03738	4,04961
ELET3	0,00185	0,03656	0,21784	4,45284
ELET6	0,0017	0,03124	0,2762	4,26192
EMBR3	-0,00014	0,02296	0,00429	4,51629
ENBR3	0,00027	0,01738	-0,05701	4,05721
EQTL3	0,00112	0,01396	-0,10172	3,6041
ESTC3	0,00074	0,03089	-0,09458	4,11747
FIBR3	0,00058	0,02613	-0,05404	4,16618
FLRY3	0,00159	0,01847	0,26839	3,86039
GGBR4	0,0007	0,0341	0,11259	3,84369
GOAU4	-0,00049	0,04099	-0,06618	4,11666
HYPE3	0,00075	0,01864	0,10011	4,67914
IGTA3	0,00028	0,01779	0,04263	4,06634
ITSA4	0,00059	0,01888	0,0889	4,38982
ITUB4	0,00069	0,01995	0,0424	4,52167
JBSS3	-0,00056	0,03707	0,16617	4,90508
KLBN11	0,0002	0,01847	-0,00114	3,4679
KROT3	0,00018	0,02729	0,17608	3,9581
LAME4	0,00036	0,02265	-0,05506	4,05283
LREN3	0,00103	0,02001	-0,06235	3,50131
MGLU3	0,00418	0,04754	0,1501	4,90547
MRFG3	0,00048	0,02326	0,2794	4,29589
MRVE3	0,0008	0,01942	-0,04908	3,60486
MULT3	0,00023	0,01765	0,16133	4,10636
NATU3	0,00022	0,02508	0,13189	3,91372
PCAR4	-0,0005	0,02259	0,10876	4,01431
PETR3	0,00108	0,0318	0,04931	4,54683
PETR4	0,00094	0,0325	-0,04536	4,42889
QUAL3	-0,00011	0,02734	-0,07532	4,09679
RADL3	0,00106	0,01762	0,17498	3,5413
RAIL3	0,0018	0,0263	0,17788	4,28508
RENT3	0,00118	0,0224	0,12843	3,95273
SANB11	0,00124	0,02227	-0,09124	4,1091
SAPR11	-0,00027	0,01399	0,56499	4,49496
SBSP3	0,00086	0,02158	-0,05793	3,97315
SMLS3	-0,00238	0,02481	-0,32886	3,82738
SUZB3	0,00141	0,0262	0,24831	4,05228
TAEE11	0,00002	0,01749	-0,21872	3,72535
TIMP3	0,00046	0,02058	-0,04302	4,01437
UGPA3	0,00005	0,0139	-0,03719	4,17056
USIM5	0,00118	0,04519	0,19502	4,64534
VALE3	0,00112	0,03442	0,07346	4,14858
VIVT4	0,00009	0,01588	-0,04213	4,05342
VVAR11	0,00078	0,03764	0,13374	4,18846
WEGE3	0,0004	0,0186	0,05369	3,86449

## V. HOMOSCEDASTICIDADE

Pode-se definir a heterocedasticidade como a condição oposta à Homoscedasticidade<sup>3</sup>, que é a suposição básica de que a variância seja a mesma em todos os níveis [29], ou ainda, para toda amostra  $X_i$  retirada de uma população, observa-se a mesma variância.

## A. Teste de Barlett

A hipótese de homoscedasticidade dos dados para um conjunto de amostras pode ser testada utilizando diferentes métodos, dentre os quais estão o teste de Barlett [30], que consiste em verificar a hipótese:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_T^2 \quad (31)$$

Dispondo de  $T$  amostras com tamanhos  $n_i$  e variâncias amostrais  $S_i^2$ . A construção do teste exige o cálculo das seguintes grandezas:

$$S_e^2 = \frac{\sum_{i=1}^T (n_i - 1) S_i^2}{n - T} \quad (32)$$

$$M = (n - T) \ln S_e^2 - \sum_{i=1}^T (n_i - 1) \ln S_i^2 \quad (33)$$

$$C = 1 + \frac{1}{3(T-1)} \left[ \sum_{i=1}^T \left( \frac{1}{n_i - 1} \right) - \left( \frac{1}{n - T} \right) \right] \quad (34)$$

A estatística do teste é então dada por  $M/C$ , que segue uma distribuição aproximada  $\chi^2$ .

## B. Resultados

A aplicação do teste de Barlett para verificação da hipótese de homoscedasticidade das séries temporais dos retornos (uma para cada um dos ativos da base de dados), revela que há evidências suficientes para rejeitar a hipótese de homoscedasticidade dos dados para os retornos de todos os ativos, conforme ilustra o gráfico da Figura 17, que traz um resumo dos resultados obtidos para cada uma das séries temporais.

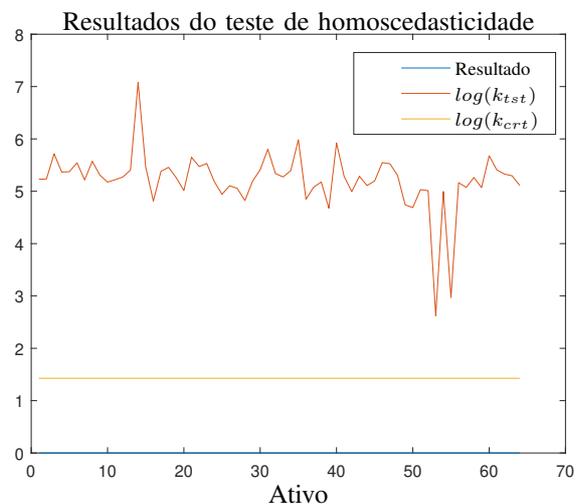


Fig. 17: Resumo dos testes de homoscedasticidade

<sup>3</sup>A palavra Homoscedasticidade tem origem na combinação de dois vocábulos gregos: ὅμοιος (ómoios ou homos) o mesmo ou igual e σκεδάννυμι (skedánummi) dispersão ou espalhamento

É evidente no gráfico que, as estatísticas de teste  $k_{tst}$  são todas superiores aos valores críticos  $k_{crt}$ , donde conclui-se que as séries temporais apresentam fortes evidências de heterocedasticidade.

### VI. ESTACIONARIEDADE

Se o valor esperado de um processo estocástico for constante em relação ao tempo,  $E[X_t] = E[X_{t+\delta}] = \mu_X$ , e sua autocorrelação depender apenas da diferença de tempo entre duas variáveis aleatórias,  $Cov(X_t, X_{t+\delta}) = \gamma_\delta$ , então esse processo é dito (fracamente) estacionário [31], [32], [33].

Dado ao fato de que nenhuma das séries de retorno passa pelo teste de homoscedasticidade, pode-se concluir que os modelos de probabilidade individuais de cada uma delas variam ao longo do tempo, dado que sua variância assim o faz. Portanto as séries temporais analisadas não podem ser consideradas estacionárias.

Na literatura voltada ao estudo de séries temporais financeiras, porém, é comum supor que as séries de retornos sejam fracamente estacionárias [32], utilizando como evidência o resultado dos testes de raízes unitárias. Esta suposição é necessária para que seja possível a aplicação de diversos modelos, mas não sendo verdadeira todos os resultados da modelagem podem ser comprometidos.

1) *Testes de estacionariedade fraca (raízes unitárias):* A avaliação da estacionariedade de séries temporais visa sustentar a hipótese necessária para aplicação de diversas ferramentas de modelagem e análise.

Na análise econométrica essa premissa tem implicações práticas, como na determinação de tendências [34]. Neste contexto, os método de teste comumente aplicados, utilizam a hipótese de existência de raízes unitárias no modelo autorregressivo, para determinação da estacionariedade do processo [35], [36], [33], [34], [37].

Em linhas gerais, temos uma série temporal  $y_t$  para qual é estimado o seguinte modelo:

$$y_t = \phi y_{t-1} + \epsilon_t$$

onde:

$$\epsilon_t \approx \mathcal{N}(\mu, \sigma^2) \forall t$$

$$E[\epsilon_t \epsilon_s] = 0, \forall t \neq s$$

sendo  $\mu$  e  $\sigma$  são a média e variância de  $\epsilon$ . Para que a série  $y_t$  seja fracamente estacionária, a restrição  $|\phi| < 1$  deve ser atendida.

a) *Teste de Dickey-Fuller aumentado:* Inicialmente sugerido por Dickey e Fuller em 1979 [38], e comumente aplicado na forma do teste conhecido com o ADF (Augmented Dickey-Fuller), o teste ADF visa estudar a estacionariedade da série temporal com base no seguinte modelo:

$$y_t = c + \delta t + \phi y_{t-1} + \beta \delta y_{t-1} + \dots + \beta \Delta y_{t-p} + \epsilon_t \quad (35)$$

onde:

- $\Delta$  é o operador de diferença para trás;
- $p$  é a ordem do modelo autorregressivo;
- $\delta$  é o intercepto ou *drift*.

As hipóteses do modelo são:

$$H_0 : \phi = 1 \quad (36)$$

$$H_1 : \phi < 1 \quad (37)$$

A estatística do teste, equação 38, é comparada com os valores críticos que foram estabelecidos por Dickey e Fuller utilizando simulações de Monte Carlo. Quando a estatística do testes é menor que o valor crítico (valores negativos, pois, o teste ADF não é simétrico), a hipótese  $H_0$  é rejeitada. Quanto menores os resultados de  $T$ , maiores as evidências para se rejeitar  $H_0$ . Rejeitando a hipótese  $H_0$  em favor de  $H_1$ , o teste indica que há evidências de estacionariedade fraca nos dados da série temporal.

$$T = \frac{\hat{\phi}}{Se(\hat{\phi})} \quad (38)$$

onde  $Se(\hat{\phi})$  é um estimador de desvio padrão para o parâmetro  $\phi$ .

b) *Teste de Phillips-Perron:* O teste de Phillips-Perron é uma generalização do teste de Dickey-Fuller, que utiliza a mesma distribuição assintótica, porém a estatística do teste é calculada de forma diferente [37]. Este teste utiliza para o cálculo da estatística o modelo descrito na equação 35, e os respectivos resíduos dados por  $r_t$ . A estatística de Phillips-Perron é então calculada a partir da seguinte equação:

$$Z = n\hat{\phi}_n - \frac{n^2\hat{\sigma}^2}{2S_n^2} (\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}) \quad (39)$$

onde  $\hat{\sigma}^2$  é um estimador de variância de  $\phi$  e:

$$\hat{\gamma}_{j,n} = \frac{1}{n} \sum_{i=j+1}^n r_i r_{i-j}$$

$$\hat{\lambda}_n^2 = \hat{\gamma}_{0,n} + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1}\right) \hat{\gamma}_{j,n}$$

$$S_n^2 = \frac{1}{n-k} \sum_{i=1}^n r_i^2$$

O teste considera as mesmas hipóteses  $H_0$  e  $H_1$  do teste de Dickey-Fuller e a mesma distribuição hipotética para obtenção do valor crítico. Nota-se ainda que, caso o processo seja não correlacionado observa-se a igualdade:  $\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}$ , logo, a estatística  $Z$  será igual a estatística do teste ADF.

c) *Teste KPSS:* Esse teste foi desenvolvido por Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt e Yongcheol Shin e tem uma concepção diferente dos demais. Em seu artigo [39], os autores afirmam que existem evidências empíricas de raízes unitárias em diversas séries temporais financeiras avaliadas, porém os testes ADF e de Phillips-Perron falham em rejeitar a hipótese nula para esses casos.

Alternativamente, o teste KPSS utiliza um modelo que contém um caminho aleatório, conforme equação a seguir:

$$y_t = c_t + \delta t + \epsilon_t \quad (40)$$

$$c_t = c_{t-1} + \epsilon'_t \quad (41)$$

onde:

$$\epsilon'_t \approx \mathcal{N}(0, \sigma_a)$$

A hipótese do teste é:

- $H_0 : \sigma_a^2 = 0$
- $H_1 : \sigma_a^2 > 0$

Quando o caminho aleatório tiver variância nula, então a série temporal será estacionária, assim, pode-se também julgar as hipóteses como segue:

- $H_0 =$  A série é estacionária
- $H_1 =$  A série apresenta raiz unitária

A estatística do teste é calculada utilizando a equação:

$$LM = \sum_{t=1}^N \frac{S_t^2}{N^2 \hat{\sigma}_a^2} \quad (42)$$

onde  $e_t$  é a série dos resíduos e:

$$S_t = \sum_{i=1}^t e_i, \quad t = 1, 2, \dots, T$$

A estatística do teste converge assintoticamente para um Movimento Browniano com valores críticos tabelados [39].

### A. Resultados dos testes de raízes unitárias

Visando ampliar então o estudo sobre a estacionariedade das séries, as figuras 18 e 19, mostram os valores críticos e as estatísticas dos testes de estacionariedade ADF e KPSS, para os retornos dos 64 ativos analisados.

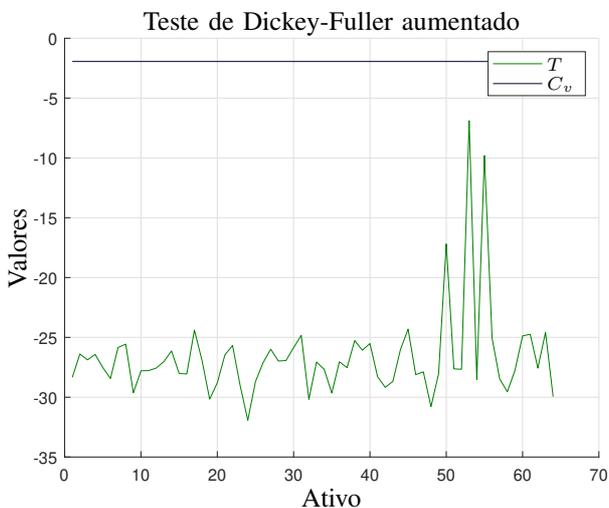


Fig. 18: Resultados do teste de estacionariedade aumentado de Dickey-Fuller

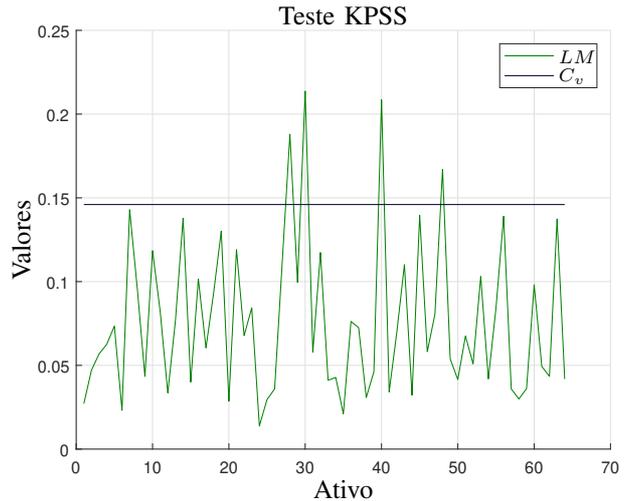


Fig. 19: Resultados do teste de estacionariedade KPSS

O teste de Dickey-Fuller aumentado retornou estatísticas significativamente mais negativas do que as estatísticas de teste. Esta observação permite concluir que há fortes evidências para rejeitar a hipótese nula em favor da alternativa, concluindo então que todas as séries são estacionárias.

Contrastando-se à ele, o teste KPSS manteve os valores das estatísticas do teste próximos dos valores críticos. É possível observar também que para 4 séries o teste indica evidências suficientes para rejeitar  $H_0$ , o que nesse caso significa rejeitar a hipótese da série ser estacionaria.

Os resultados dos testes KPSS e DFA, apesar da diferenças entre eles, levam a conclusão que as séries são estacionárias (restam dúvidas a respeito apenas de 4 delas, rejeitadas pelo testes KPSS). Esse resultado, porém, é oposto a análise da homoscedasticidade.

Para fins práticos, dada a formulação dos testes, conclui-se ser possível modelar os retornos através de processos autorregressivos. Essa conclusão, porém, contrasta com a evidência de não estacionariedade das séries, seja no sentido fraco ou no sentido amplo da definição, o que coloca em dúvida os resultados obtidos desta modelagem.

## VII. CORRELAÇÕES

A correlação, medida através de coeficientes de correlação, expressa a dependência de probabilidade entre duas variáveis aleatórias. Os coeficientes de correlação descrevem tanto a força dessa dependência mútua quanto sua direção.

### A. Correlação de Pearson

Conhecido como coeficiente de correlação linear, essa medida expressa a dependência linear entre duas variáveis aleatórias [40] e é o coeficiente mais utilizado para medir a correlação [41].

O cálculo amostral do coeficiente de Pearson, considerando duas amostras  $x$  e  $y$  de mesmo tamanho  $n$ , é dado pela seguinte equação:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (43)$$

A equação 43 representa a razão entre a covariância amostral e o produto das variâncias amostrais:

$$r = \frac{Cov(x, y)}{S_x S_y}$$

O parâmetro populacional equivalente, ou seja, a correlação de Pearson para as variáveis aleatórias  $X$  e  $Y$  é dado por:

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y} \quad (44)$$

### B. Correlação de Spearman

O coeficiente de Spearman mede a relação monotônica entre duas variáveis aleatórias [42]. Nesse cálculo, pressupõe-se que as grandezas tendem a variar de forma conjunta monotonicamente, porém, não com a taxa de variação constante. A equação 45 estabelece o coeficiente de correlação de Spearman amostral para duas amostras  $x$  e  $y$  de mesmo tamanho  $n$ :

$$r = \frac{\sum_{i=1}^n (x'_i - \bar{x}') (y'_i - \bar{y}')}{\sqrt{\sum_{i=1}^n (x'_i - \bar{x}')^2 (y'_i - \bar{y}')^2}} \quad (45)$$

onde  $x'$  e  $y'$  são os postos das observações  $x$  e  $y$ .

Popularmente o coeficiente de Spearman é calculado para duas amostras com base na seguinte equação:

$$r = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (46)$$

### C. Correlação de Kendall

O coeficiente de correlação de Kendall  $\tau$ , avalia a associação monotônica entre duas variáveis, porém, não necessariamente contínua.[43]. O calculo do coeficiente é realizado através da equação 47, considerando duas amostras  $x$  e  $y$ , ambas de tamanho  $n$ .

$$\tau = \frac{\sum_{i=1}^n \sum_{j=1}^n \text{sign}(x_i - x_j) \text{sign}(y_i - y_j)}{n(n - 1)} \quad (47)$$

Esse coeficiente faz uso do conceito de concordância, sendo que, dado o par  $(x_i, y_j)$ , este será concordante quando  $x_i < x_j$  e  $y_i < y_j$  ou  $x_i > x_j$  e  $y_i > y_j$ , portanto  $(x_i - x_j)(y_i - y_j) > 0$ .

### D. Intervalos de confiança da correlação

Supondo que  $x$  e  $y$  são amostras de duas variáveis aleatórias normais, com função de densidade conjunta (função de densidade normal padrão bivariada):

$$f_{XY} = \frac{1}{2\pi(1 - \rho^2)} e^{-\frac{x^2 + y^2 - 2\rho xy}{2(1 - \rho^2)}}$$

A fim de julgar se existe correlação entre as variáveis  $X$  e  $Y$ , consideremos as seguintes hipóteses:

- $H_0 : \rho = 0$ , Variáveis não são correlacionadas;
- $H_1 : \rho \neq 0$ , Existe algum nível de correlação entre as variáveis.

A estatística amostral de teste é definida por [44]:

$$Q = r \sqrt{\frac{n - 2}{1 - r^2}} \quad (48)$$

Essa estatística segue uma distribuição T-Student com  $n - 2$  graus de liberdade.

Alternativamente, Sir Ronald Aylmer Fisher desenvolveu uma transformação para obter uma estatística de teste que tende rapidamente a uma distribuição normal com o aumento do tamanho da amostra  $n$ . A estatística de Fisher é dada na equação 49.

$$z = \frac{1}{2} \ln \left( \frac{1 + r}{1 - r} \right) \quad (49)$$

$Z$  tem uma distribuição normal com média  $\bar{z}$  e variância  $\sigma_z$ .

$$\bar{z} = \frac{1}{2} \ln \left( \frac{1 + \rho}{1 - \rho} \right)$$

$$\sigma_z = \frac{1}{\sqrt{n - 3}}$$

### E. Coeficientes de correlação

Inicialmente, a avaliação das correlações exige um ajuste nos ativos da base de dados. Alguns ativos representam ações de uma mesma empresa, separados apenas em ações ordinárias e preferenciais. Avaliar a correlação entre esses ativos não acrescenta informação a presente análise, senso assim, as seguintes ações preferências serão excluídas desta análise:

- PETR4
- BBDC4
- ELET6

A base de dados resultante, utilizada nesta análise, está descrita na tabela X, que também associa números de série aos ativos, com finalidade de possibilitar a visualização gráfica dos resultados através desta indexação.

Serão avaliadas as correlações entre todos os pares de ativos na base de dados resultante, através dos coeficientes de Pearson, Spearman e Kendall. Os resultados serão julgados através de um teste onde, um p-valor menor do que 0,05 (intervalo de confiança de 5%), indica que existe uma correlação diferente de zero entre as amostras.

TABLE X: Base de dados para estudo das correlações

#	Ativo	#	Ativo	#	Ativo
1	ABEV3	20	ELET3	39	MRVE3
2	B3SA3	21	EMBR3	40	MULT3
3	BBAS3	22	ENBR3	41	NATU3
4	BBDC3	23	EQTL3	42	PCAR4
5	BBSE3	24	ESTC3	43	PETR3
6	BRAP4	25	FIBR3	44	QUAL3
7	BRFS3	26	FLRY3	45	RADL3
8	BRKM5	27	GGBR4	46	RENT3
9	BRML3	28	GOAU4	47	SANB1
10	CCRO3	29	HYPE3	48	SBSP3
11	CIEL3	30	IGTA3	49	SUZB3
12	CMIG4	31	ITSA4	50	TAE1
13	CPFE3	32	ITUB4	51	TIMP3
14	CPLE6	33	JBSS3	52	UGPA3
15	CSAN3	34	KLBN1	53	USIM5
16	CSNA3	35	KROT3	54	VALE3
17	CYRE3	36	LAME4	55	VIVT4
18	ECOR3	37	MGLU3	56	VVAR1
19	EGIE3	38	MRFG3	57	WEGE3

1) *Coefficientes de Pearson*: Considerando como medida de correlação o coeficiente de Pearson, as figuras 20, 21 e 22 mostram as correlações entre os retornos dos ativos com maior coeficiente.

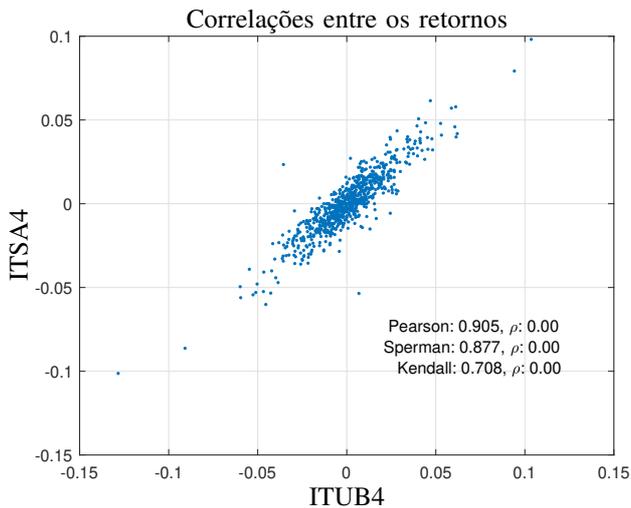


Fig. 20: Correlação entre os retornos dos ativos ITSA4 e ITUB4

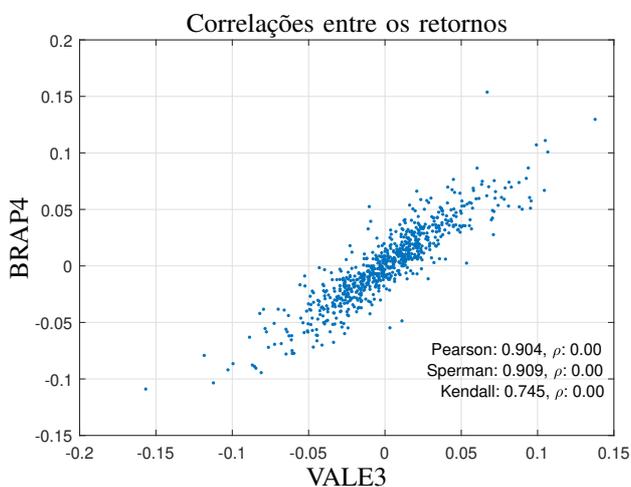


Fig. 21: Correlação entre os retornos dos ativos BRAP4 e VALE3

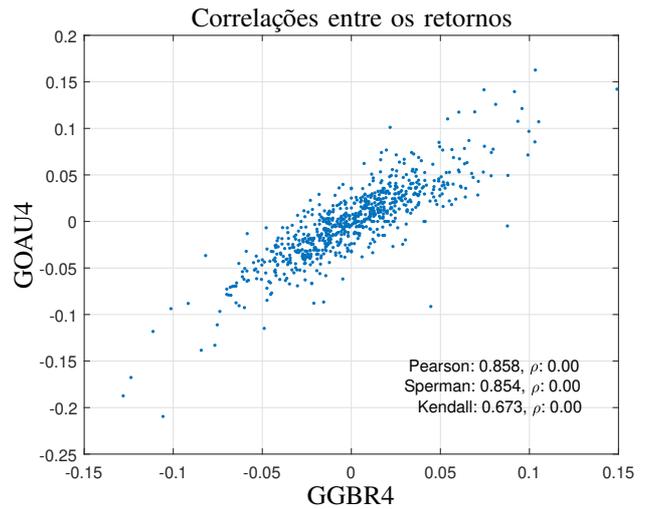


Fig. 22: Correlação entre os retornos dos ativos GGBR4 e GOAU4

Para esses três pares de ativos, observar-se que  $\rho$  (p-valor do teste de hipótese para a correlação) é menor do que 0,05 (5%), confirmando que não se pode rejeitar a hipótese de que existe correlação entre as amostras. A análise dos gráficos corrobora com essa conclusão, pois a distribuição dos pares de retornos também indica a existência de correlação.

A coleção completa de ativos contém 57 amostras, que combinadas em pares, resulta em 1596 coeficientes de correlação. Essa quantidade pode ser analisada com o auxílio de um gráfico tridimensional. Consideremos a figura 23, que mostra uma superfície composta pelos retornos dos pares de ativos, nela os dados são dispostos de forma que eixos horizontais  $x$  e  $y$  representem o número de índice do ativo, entre 0 e 57 e o eixo vertical  $z$  representa o valor do coeficiente de correlação para o par de ativos indexados por  $x$  e  $y$ . Nota-se que existem picos na superfície, tanto positivos quanto negativos, e ainda parte dos retornos encontra-se próximo ao plano  $z = 0$ .

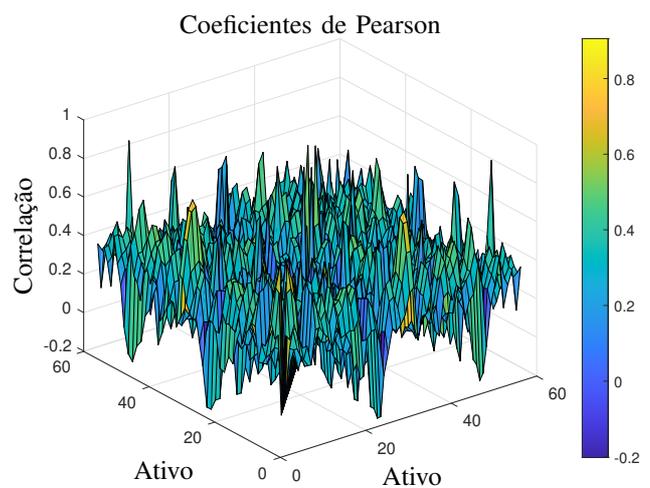


Fig. 23: Coeficientes de correlação de Pearson entre os retornos dos 1596 pares de ativos da base de dados

As figuras 24 e 25 resumem os 1596 coeficientes de correlação, através de um gráfico de distribuição de frequência e um gráfico contendo os dados plotados em linha. Nesse

último perde-se a referência direta ao par de ativos que gera o valor de correlação, por outro lado proporciona uma visualização clara e direta do comportamento desses índices.

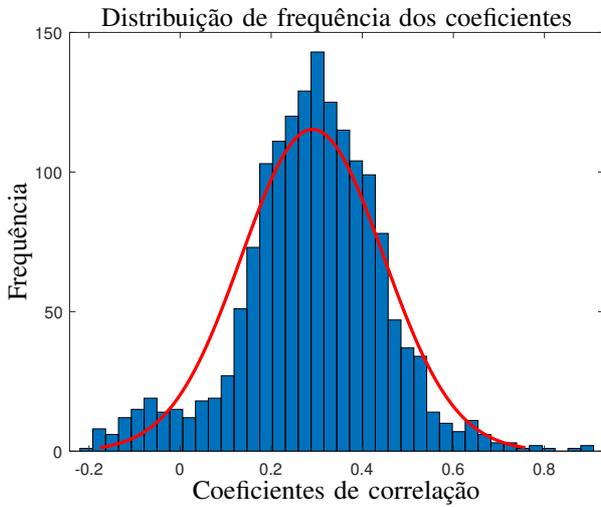


Fig. 24: Distribuição de frequências dos coeficientes de Pearson

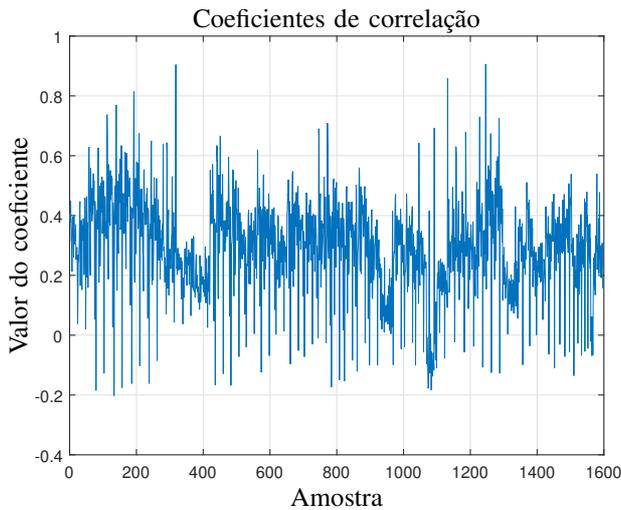


Fig. 25: Coeficientes de correlação de Pearson

Em resumo, fica demonstrado que os coeficientes tem média em torno de 0,3 sendo predominantemente positivos, sendo que para 93,86% dos pares de ativos (1498), há evidências de correlação entre os dados. Do total, 94,74% dos ativos apresentam coeficiente de correlação positivos.

A figura 26 mostra os retornos do par com maior correlação negativa. Apesar de uma concentração de dados no entorno da origem, que contribui com a hipótese de não correlação nos dados, os pontos extremos encontram-se em quadrantes opostos, contrariando essa hipótese.

O p-valor do teste ( $\rho$ ), mostra que é possível aceitar a hipótese de correlação linear nos dados, a partir do coeficiente de Pearson, porém não é possível chegar a mesma conclusão considerando os coeficientes de Spearman e Kendall.

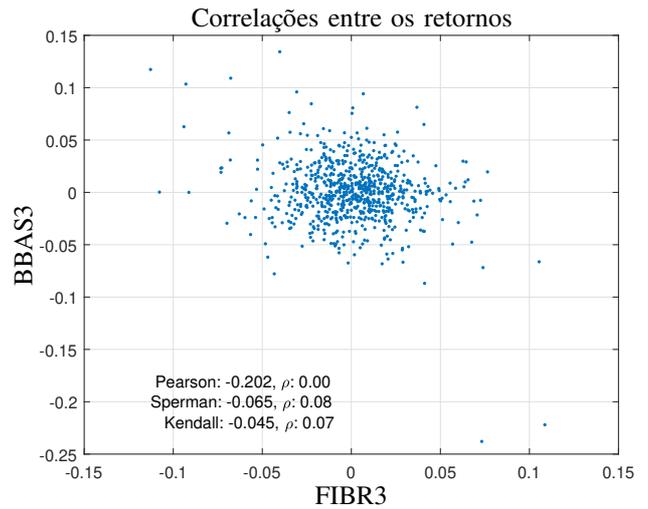


Fig. 26: Correlação entre os retornos dos ativos BBAS3 e FIBR3

2) *Coefficientes de Spearman*: Os coeficientes de Pearson medem a correlação linear entre os dados amostrais, enquanto o coeficiente de Spearman tenta mensurar a relação monotônica, não necessariamente linear, por tanto podendo haver outras formas de correlação nos dados, não evidenciadas pelos coeficientes de Pearson. As figuras 27, 28 e 29 resumem os coeficientes de Spearman para a base de dados.

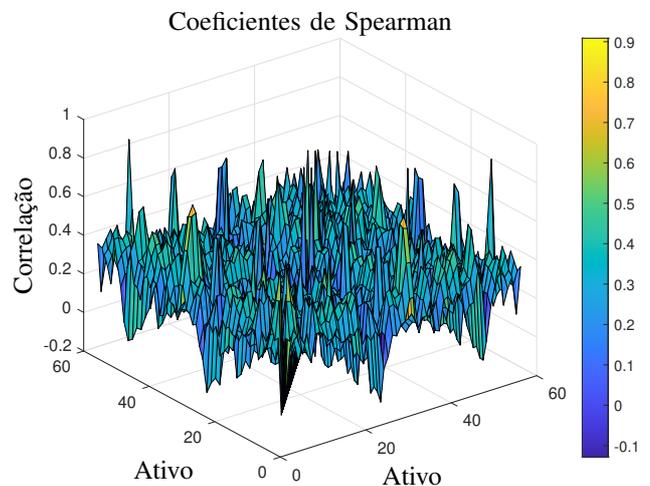


Fig. 27: Coeficientes de correlação de Spearman entre os retornos dos 1596 pares de ativos da base de dados

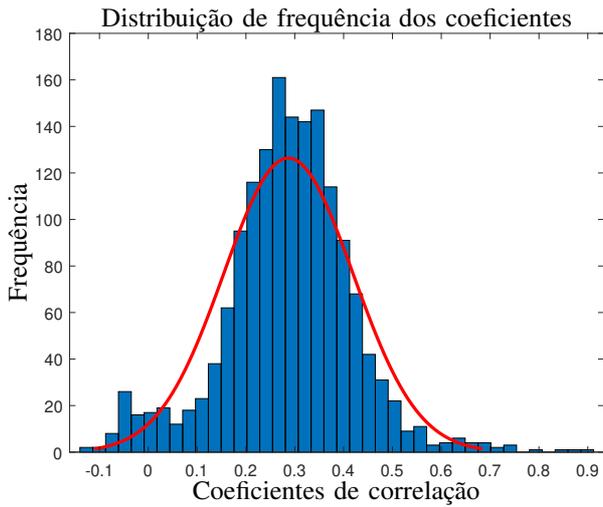


Fig. 28: Distribuição de frequências dos coeficientes de Spearman

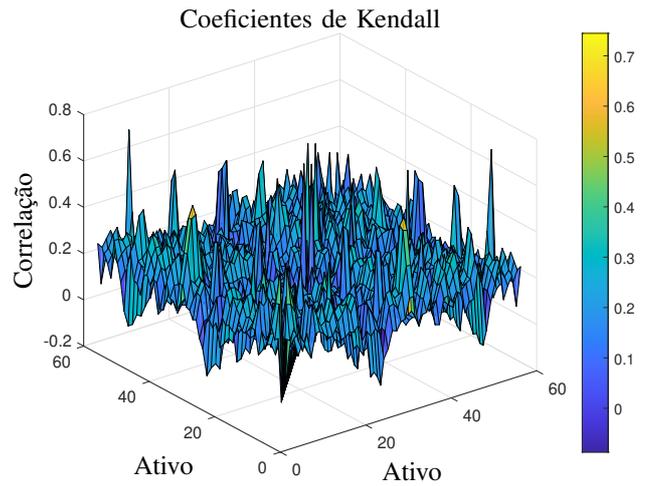


Fig. 30: Coeficientes de correlação de Spearman entre os retornos dos 1596 pares de ativos da base de dados

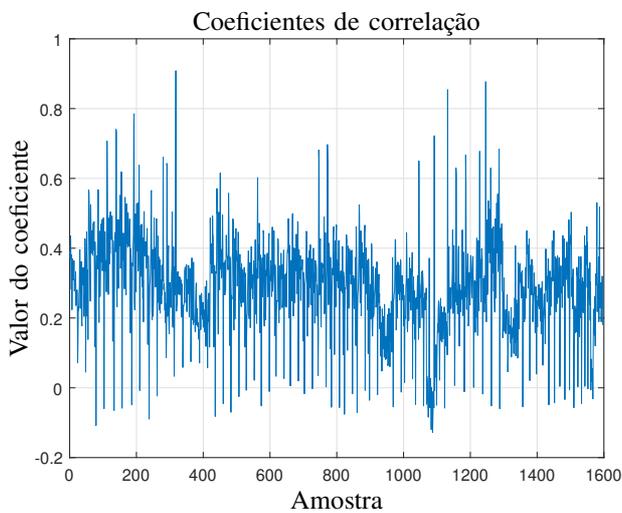


Fig. 29: Coeficientes de correlação de Spearman

Em resumo, 1490 amostras ou 93,36% dos coeficientes de Spearman apresentam evidências de correlação entre os dados e 1534 (96,12%) coeficientes indicam uma correlação positiva.

3) *Coeficientes de Kendall*: Por fim, as figuras 30, 31 e 32 mostram o resumo dos coeficientes de correlação de Kendall.

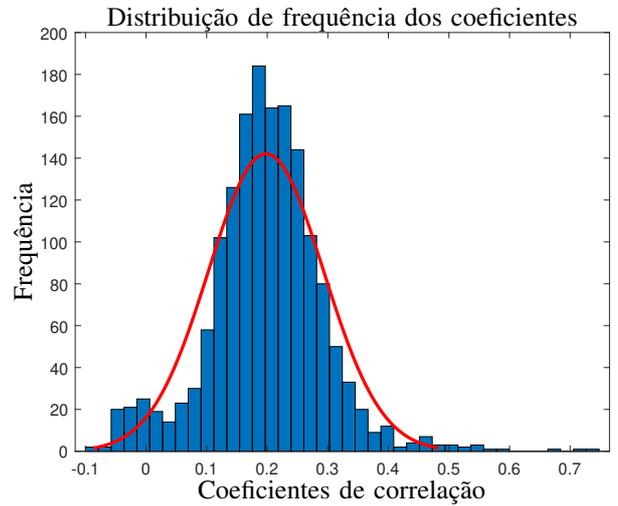


Fig. 31: Distribuição de frequências dos coeficientes de Kendall

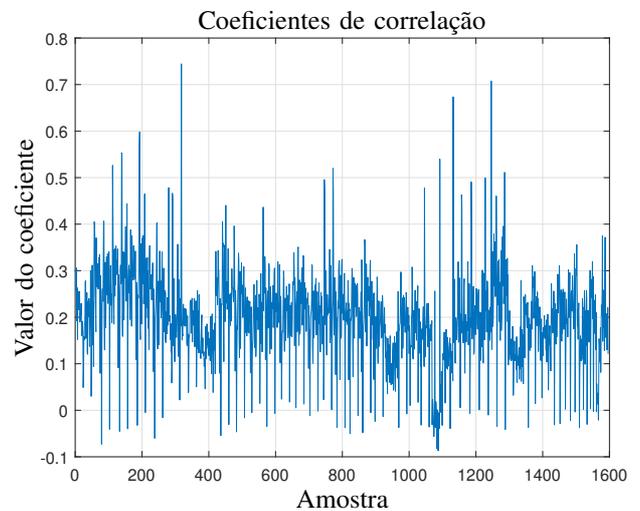


Fig. 32: Coeficientes de correlação de Kendall

Em resumo, 1490 amostras ou 93,36% dos coeficientes de

Kendall apresentam evidências de correlação entre os dados e 1534 (96,12%) coeficientes indicam uma correlação positiva, os mesmos números obtidos analisando os coeficientes de Spearman.

F. Resultados Gerais

A fim de analisar os resultados dos três coeficientes, de maneira integrada, é apresentado um resumo no gráfico da figura 33. Os dados revelam que os coeficientes de Pearson e Spearman são relativamente muito parecidos, enquanto o coeficiente de Kendal apresenta um discrepância maior, com um pico de frequência próximo a 0,2, concentrando um maior número de coeficientes com valores mais baixos.

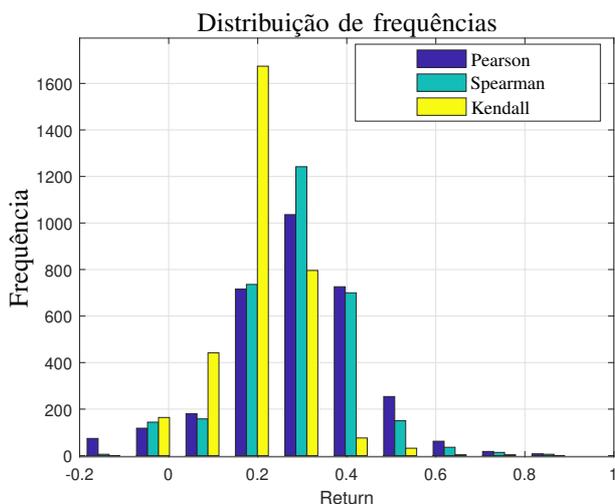


Fig. 33: Coeficientes de correlação de Kendall

A quantidade de coeficientes próximos de zero também é similar para os três indicadores, donde conclui-se que há fortes evidências de que esse grupo de ativos seja de fato, composto por dados não correlacionados. Esse resultado é significativo para a análise proposta, pois demonstra propriedades observáveis do mercado de capitais.

A fim de avaliar essa hipótese, é apresentada a tabela XI que mostra os pares de ativos com coeficientes de correlação menores que 0,05 considerando as três formas de cálculo. A última coluna mostra o menor p-valor entre os três coeficientes. Para esses pares de ativos, há fortes evidências da não existência de qualquer tipo de correlação entre os retornos.

TABLE XI: Ativos com coeficientes de correlação menores do que 0,05

Ativos	Pearson	Spearman	Kendall	p-valor
SUZB3 ABEV3	0,019917	-0,04905	0,030065	0,182854
SUZB3 BRAP4	0,005041	0,021707	0,02211	0,368588
FIBR3 CIEL3	0,018226	-0,00286	-0,00497	0,620825
SUZB3 CIEL3	-0,02514	0,034807	0,014848	0,34471
FIBR3 CPFE3	0,032791	-0,00723	0,023986	0,330207
SUZB3 CPFE3	0,018226	-0,03559	0,017738	0,334007
SUZB3 CSAN3	0,047504	0,015044	-0,00172	0,197076
FIBR3 CSNA3	0,03704	0,025587	0,022782	0,314633
SUZB3 CSNA3	0,015567	-0,04781	0,001449	0,194183
SUZB3 EGIE3	-0,02464	-0,01707	-0,00529	0,503667
FLRY3 EMBR3	-0,03911	0,015044	0,035247	0,151882
SUZB3 ENBR3	-0,04229	-0,0252	-0,0077	0,250875
FIBR3 EQTL3	-0,00603	0,048579	0,01043	0,187125
SUZB3 EQTL3	-0,02259	-0,04832	0,017056	0,189477
CIEL3 FIBR3	0,008657	-0,00956	-0,00497	0,795353
CPFE3 FIBR3	0	0,03074	0,023986	0,330207
CSNA3 FIBR3	0	-0,04977	0,022782	0,176489
EQTL3 FIBR3	0	0,007024	0,01043	0,671476
GGBR4 FIBR3	0	-0,0252	0,022878	0,35214
GOAU4 FIBR3	0	0,001411	0,013962	0,570375
JBSS3 FIBR3	0	0	0,032883	0,181135
MGLU3 FIBR3	0	0	0,018917	0,442034
MRFG3 FIBR3	0	0	0,012265	0,618568
TAE1 FIBR3	0	0	-0,00474	0,847158
UGPA3 FIBR3	0	0	0,022078	0,369226
VIVT4 FIBR3	0	0	0,029399	0,231817
EMBR3 FLRY3	0	0	0,035247	0,151882
KLBN1 FLRY3	0	0	0,015568	0,526933

Nota-se na tabela XI que alguns ativos aparecem com maior frequência, tendo destaque o ativo FIBR3, que leva a hipótese de que alguns ativos podem ser, em linhas gerais, menos correlacionados com os demais, o que também tem implicações tangíveis na análise de investimentos. A fim de analisar essa hipótese, as médias dos coeficientes são mostradas na figura 34.

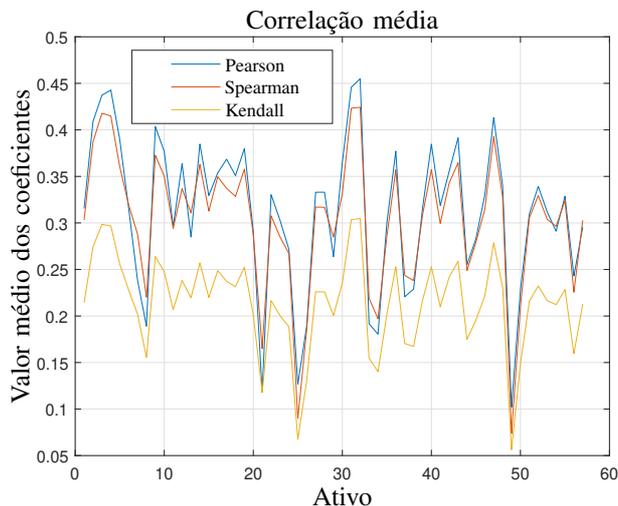


Fig. 34: Média dos valores absolutos dos coeficientes de correlação

Analisando o gráfico, pode-se observar três pontos com relativo destaque, que ocorrem nas posições 21, 25 e 49 no eixo x, que representam os ativos: EMBR3, FIBR3 e SUZB3 de acordo com a tabela X. Esses pontos representam os ativos com menor correlação média com os demais. A figura 35 ilustra um desses ativos, onde são mostrados os coeficientes de correlação calculados com relação aos demais ativos.

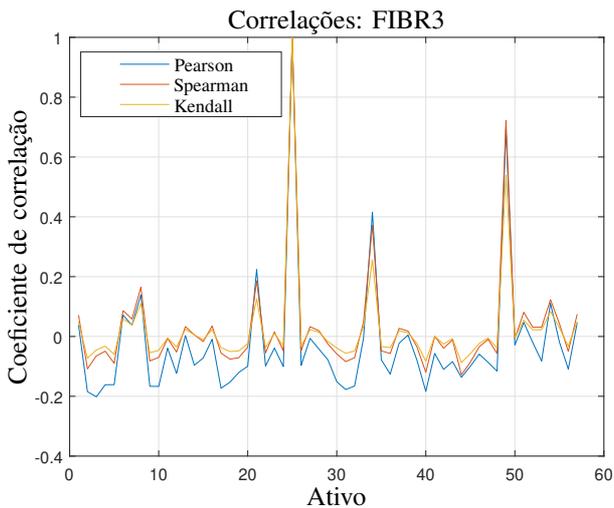


Fig. 35: Coeficientes de correlação calculados para os retornos do ativo FIBR3

Para a série de retornos do ativo FIBR3, o teste de hipóteses para um intervalo de confiança de 5% mostra que, considerando o coeficiente de Pearson pode-se supor que existe alguma correlação nos dados para 36 ativos, enquanto o teste de Kendall indica correlação com apenas 19 e o teste de Spearman com 16.

Em resumo, os coeficientes concordam na correlação do ativo FIBR3 com 13 ativos, sendo eles: B3SA3, BBSE3, BRKM5, BRML3, ECOR3, EMBR3, ITSA4, KLBN1, MULT3, QUAL3, RADL3, SUZB3 e VALE3. Pode-se então assumir que não há evidências de correlação entre os retornos da série FIBR3 e os demais ativos analisados.

A análise oposta, buscando os pontos máximos no gráfico da figura 34, mostra que o ativo com maior correlação média é o ITUB4, cujas correlações médias são mostradas no gráfico da figura 36.

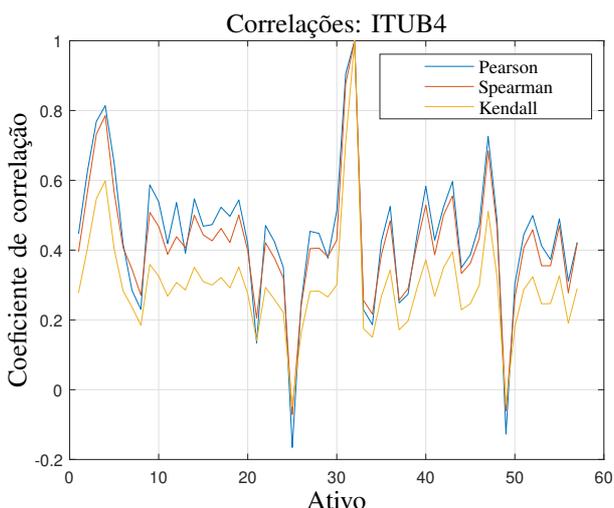


Fig. 36: Coeficientes de correlação calculados para os retornos do ativo ITUB4

Analisando os resultados do teste de hipótese para os coeficientes calculados para o ativo ITUB4, verifica-se que não há evidências de correlação nos dados apenas para o coeficiente de Spearman em relação ao ativo FIBR3 e para

os coeficientes de Spearman e Kendall para o ativo SUZB3. O coeficiente de Pearson indica evidências de correlação entre o ativo ITUB4 e todos os demais ativos da base de dados.

Em resumo, a análise dos dados revela não haver consenso nos resultados das correlações, de tal sorte que não se pode afirmar que as séries temporais financeiras apresentam, ou não apresentam, correlação de qualquer natureza, em um âmbito geral.

Contrastando com a hipótese de Fama [7], de que os retornos dos preços de fechamentos dos ativos se comportam como caminhos aleatórios, nota-se que a oscilação de ativos como ITUB4 está claramente correlacionada com as demais, assim, apesar de aleatório deve-se considerar o modelo que contemple a distribuição de probabilidade conjunta entre esse ativo e os demais com ele correlacionados.

## VIII. CORRELAÇÕES ENTRE SÉRIES DO MESMO ATIVO

Uma hipótese utilizada por investidores do mercado financeiro é de que o volume negociado impacta no retorno dos ativos. Essa suposição é objeto de estudo de diversos trabalhos acadêmicos [45], visando em geral avaliar a correlação entre o volume financeiro negociado e algum parâmetro do ativo, como a volatilidade, por exemplo. Esse tipo de hipótese tem especial interesse por parte dos investidores pois, se verdadeira, possibilitaria a realização de lucros com base em sua análise.

### A. Correlação entre volume financeiro e retorno

Propõe-se aqui avaliar a correlação direta entre o volume financeiro no pregão  $t$  e o retorno do ativo no pregão  $t + 1$ . Para tanto é apresentado o gráfico da figura 37, que mostra os coeficientes de correlação, calculados utilizando os três métodos anteriormente discutidos, para as séries de retorno e volume financeiro (um passo atrás dos retornos).

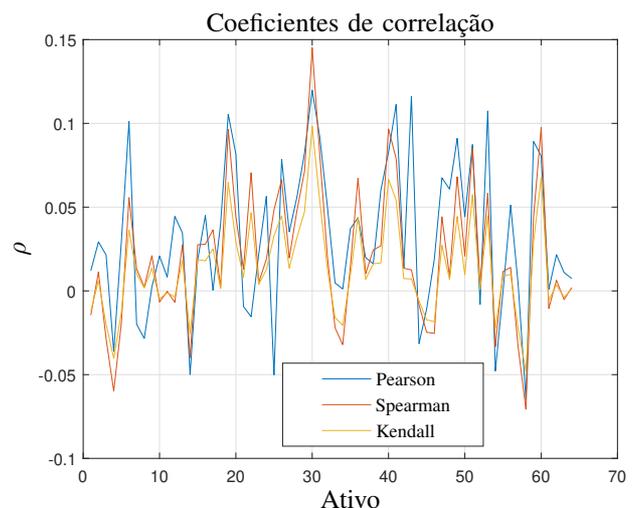


Fig. 37: Correlação entre os retornos e os volumes financeiros negociados um passo atrás

Pode-se observar que os coeficientes de correlação são muito pequenos para todos os ativos, sendo seus valores muito menores em módulo que a unidade. A presente análise leva a conclusão, com base nos valores dos coeficientes calculados,

que não há correlação entre os retornos no pregão  $t$  e os volumes negociados no pregão  $t - 1$ .

Analisando agora a correlação entre os retornos e os volumes financeiros para as séries sem defasagem, obtém-se os valores resumidos no gráfico da figura 38. Esse gráfico, apesar de apresentar coeficientes superiores àqueles observados no gráfico da figura 37, demonstra não haver evidências de forte correlação entre os volumes financeiros negociados e os retornos dos ativos.

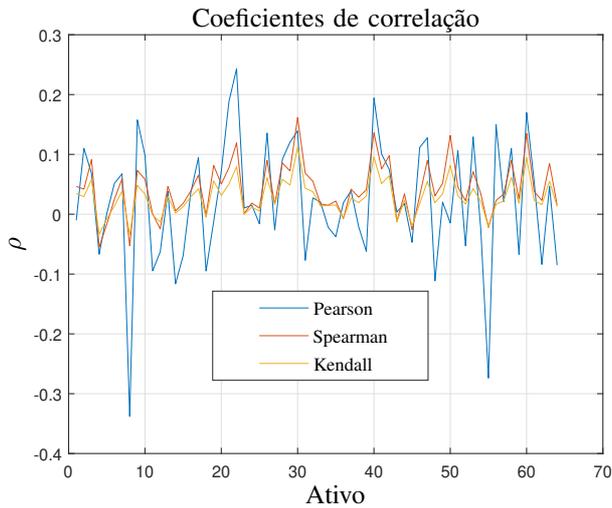


Fig. 38: Correlação entre os retornos e os volumes financeiros negociados

**B. Correlação entre a volatilidade e o retorno**

Considerando como medida de volatilidade a variação máxima dos preços dos ativos negociados durante o pregão  $t$ , dada pela equação 50. O gráfico da figura 39 mostra os coeficientes de correlação entre a série de retornos  $r_t$  e a série de volatilidades um passo atrás  $\hat{\sigma}_{t-1}$ . Podendo-se constatar no gráfico que não há evidências de correlação entre os retornos no pregão  $t$  e a volatilidade do pregão  $t - 1$ .

$$\hat{\sigma}_t = h_t - l_t \tag{50}$$

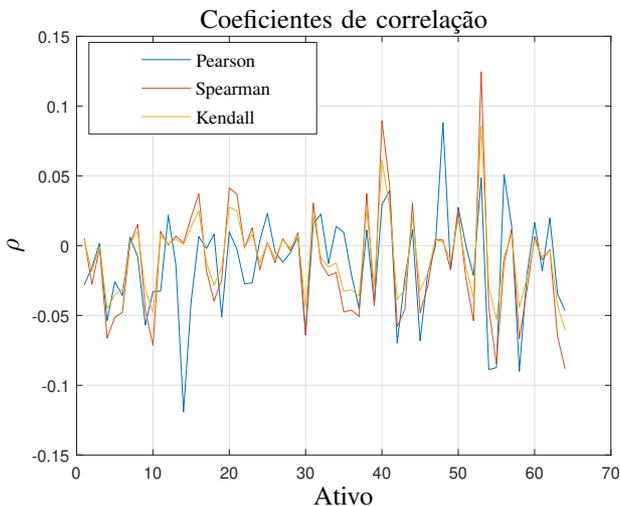


Fig. 39: Correlação entre os retornos e as volatilidades um passo atrás

Um fenômeno bem conhecido pelos investidores, chamado efeito *leverage*, estabelece que a volatilidade de um ativo, em um dado período, é diretamente afetada pelo retorno por ele obtido no período anterior [46]. Visando verificar esse fenômeno é apresentado o gráfico da figura 40. Nele são mostrados os coeficientes de correlação entre a volatilidade no pregão  $t$  e o retorno do pregão anterior  $t - 1$ .

Assim como no trabalho desenvolvido em [47], apesar de utilizada uma abordagem mais simples, chega-se à mesma conclusão, de que não há evidências suficientes nos dados para justificar a existência de correlação entre a volatilidade e o retorno anterior. Dessa forma, o estudo proposto contesta a existência do efeito *leverage*, pois se existe ele não pode observado.

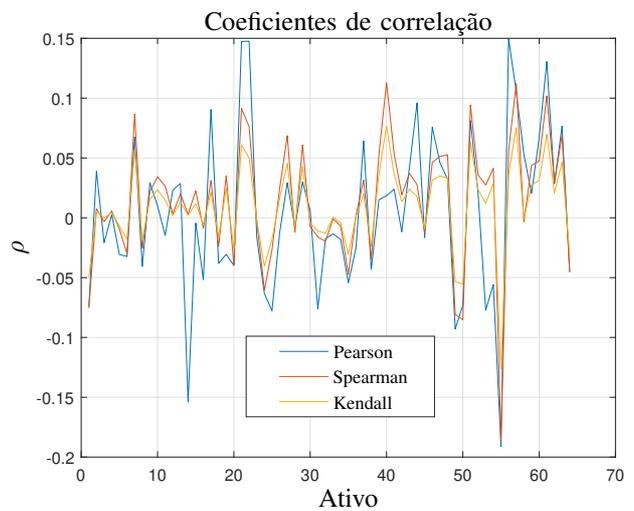


Fig. 40: Correlação entre a volatilidade e o retorno um passo atrás

**IX. DEPENDÊNCIA TEMPORAL**

Nesta sessão, tem-se por objetivo julgar o quanto o retorno no instante  $t$ , depende dos valores do retorno nos instantes de tempo anteriores, ou seja, pretende-se avaliar a probabilidade  $P[r_t | r_{t-1}, r_{t-2}, \dots, r_{t-n}]$ . Esse estudo, assim como outros aqui propostos, tem por finalidade averiguar a previsibilidade dos retornos, ou ainda, a quantidade de informação a seu respeito contida na séries temporais.

**A. Análise das autocorrelações**

Inicialmente é apresentada a figura 41, que mostra um resumo dos coeficientes de autocorrelação dos retornos. O gráfico ilustra o valor médio das autocorrelações para cada intervalo avaliado, e uma variação de 95% calculada com base na dispersão em torno dessa média.

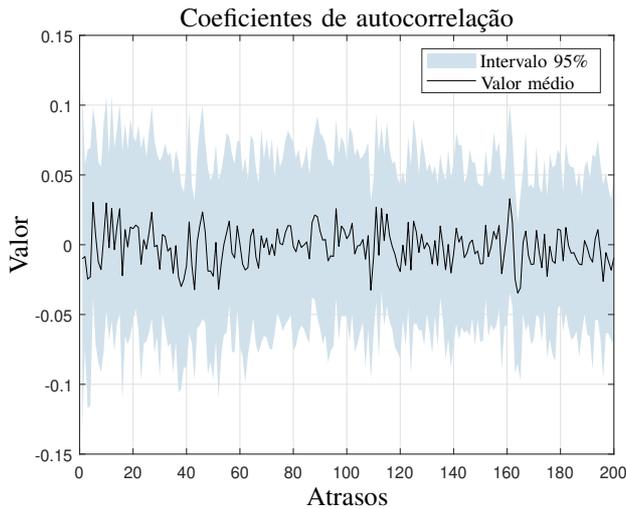


Fig. 41: Valores médios dos coeficientes de autocorrelação

O gráfico da figura 41 mostra que os coeficientes de autocorrelação são todos muito próximos de zero, mesmo considerando a dispersão. Dando sequência à análise das médias dos coeficientes, consideremos o gráfico da figura 42. Nele são mostradas as frequências dos coeficientes de autocorrelação.

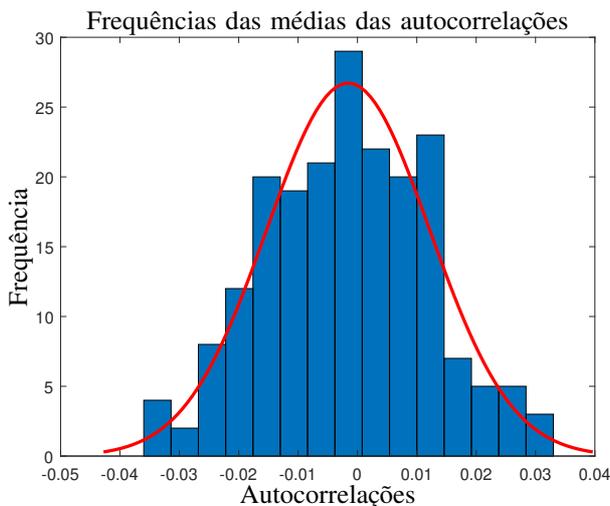


Fig. 42: Diagrama de frequências para as médias das autocorrelações

A curva da figura 42 assemelha-se a uma distribuição Gaussiana, o que indicaria que os coeficientes de autocorrelação estão aleatoriamente distribuídos ao redor da média com distribuição empírica aproximadamente Gaussiana. A hipótese de normalidade é comprovada com o auxílio do teste de Kolmogorov-Smirnov, que comprova a hipótese por não haver indícios suficientes para rejeitar  $H_0$ .

Resultado do teste:

- Lógico: Não rejeitar a hipótese de normalidade;
- Estatística do teste: 0,04;
- Valor crítico: 0,10;
- $\rho_{valor}$ : 0,95.

Atendida a hipótese de normalidade, pretende-se então avaliar a média geral dos coeficientes de autocorrelação, para

a partir desse resultado inferir se existe correlação temporal entre as amostras. Esta análise é realizada utilizando o teste de hipótese bilateral para a média:

- $H_0 : \mu = \mu_0$
- $H_1 : \mu \neq \mu_0$

onde  $\mu$  é o parâmetro populacional que descreve a média dos coeficientes de autocorrelação e  $\mu_0$  é o valor amostral calculado.

Para um nível de significância  $\alpha = 5\%$  tem-se o seguinte resultado:

- Lógico: Não rejeitar  $H_0$ ;
- $\rho_{valor}$ : 0,11;
- Intervalo de confiança de 95% para a média:  $[-0,00350, 0003]$ ;
- Desvio padrão populacional estimado: 0,0137.

Pode-se então concluir que os coeficientes de autocorrelação são normalmente distribuídos em torno de uma média nula com desvio padrão  $\sigma = 1,96$ . Esse resultado corrobora com a hipótese de que as séries dos retornos são formadas por sequências de amostras independentes. Essa conclusão é tomada com base na análise do resumo dos dados, conforme demonstrado, e contraria outros estudos como o apresentado em [48].

A fim de analisar o comportamento dos coeficientes de autocorrelação em cada série de retornos individualmente, optou-se por classificar as autocorrelações, estudando então os extremos dessa classificação na intenção de inferir, a partir desse estudo, o comportamento de todos os ativos. Para tanto tomemos o coeficientes  $\gamma$ :

$$\gamma = \sum_{\tau=1}^{200} |R_{\tau}| \quad (51)$$

onde  $R_{\tau}$  é o coeficiente de autocorrelação de ordem  $\tau$ ,

A figura 43 resume o resultado do cálculo do coeficiente  $\gamma$  para todos os ativos da base de dados. Os valores mostrados tomam como base os mesmos coeficientes apresentados na figura 41.

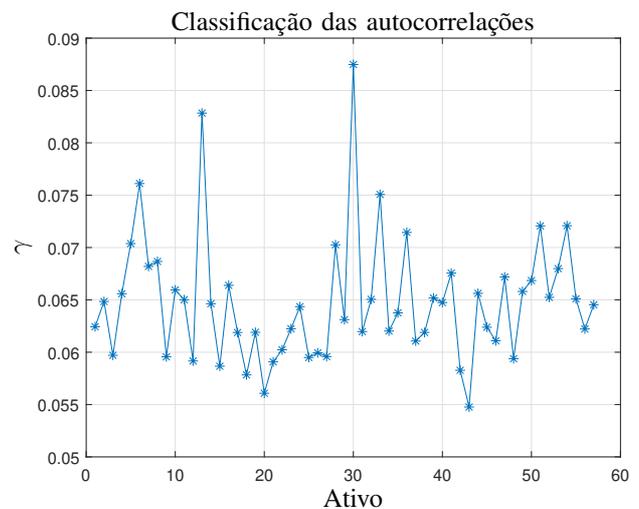


Fig. 43: Coeficientes  $\gamma$  para classificação dos ativos quanto as suas autocorrelações

O valor de  $\gamma$  é então utilizado para classificar as séries temporais em função dos coeficientes de autocorrelação  $r_{\tau}$ .

Utilizando essa classificação, as figuras 44, 45 e 46 mostram os coeficientes das séries com maiores índices de autocorrelação.

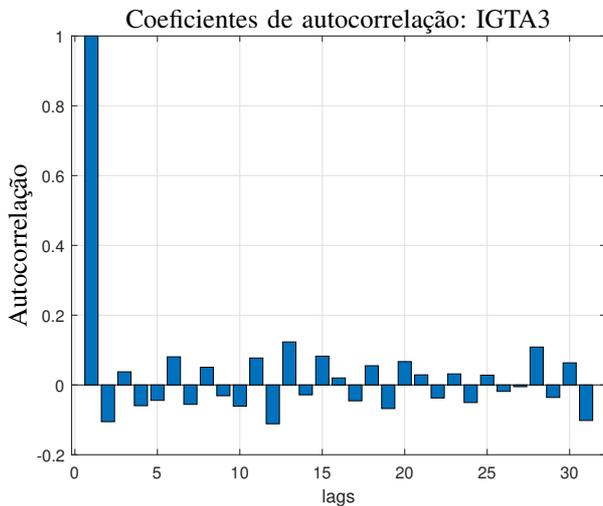


Fig. 44: Coeficientes de autocorrelação para os retornos do ativo IGTA3

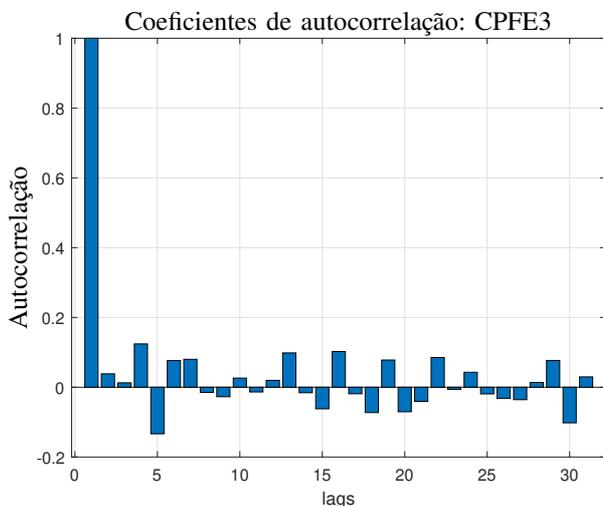


Fig. 45: Coeficientes de autocorrelação para os retornos do ativo CPFE3

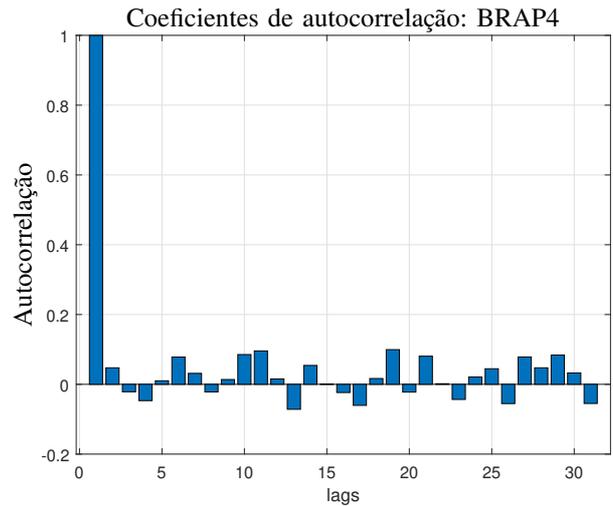


Fig. 46: Coeficientes de autocorrelação para os retornos do ativo BRAP4

O padrão dos coeficientes de autocorrelação para essas séries de retornos apresenta pouca semelhança com uma senoide amortecida, dificultando a conclusão de que os processos apresentam algum tipo de comportamento autorregressivo linear. Tal fato também é observado nas séries cuja classificação das autocorrelações é menor, como para o ativo PETR3, que apresenta o menor coeficiente  $\gamma$  entre as séries estudadas.

Além da forma geral dos coeficientes, observa-se também que esses são todos próximos de zero. Esse resultado sugere que uma análise da sua significância deve ser realizada para julgar se existe algum padrão observável, ou os valores representam apenas a aleatoriedade dos dados.

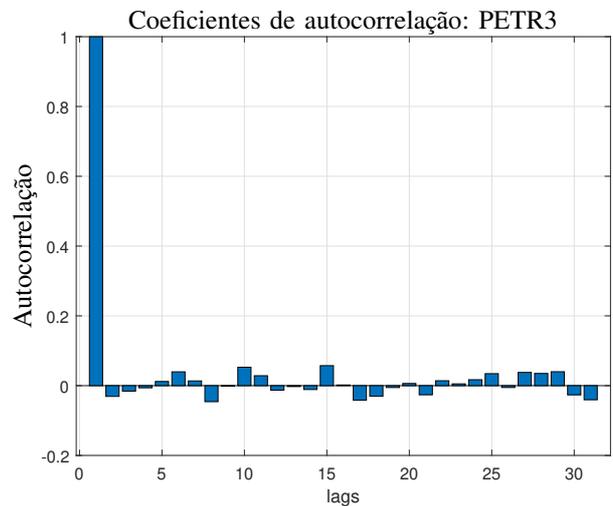


Fig. 47: Coeficientes de autocorrelação para os retornos do ativo PETR3

Visando verificar a existência de um padrão nos coeficientes de autocorrelação, a tabela XII mostra os valores médios e as dispersão dos coeficientes de autocorrelação para os 10 ativos com maior classificação segundo o critério  $\gamma$ . Também está mostrado na tabela o resultado lógico do teste KS (Kolmogorov-Smirnov) que avalia a hipótese de normalidade dos dados e do teste bilateral para média, que avalia a hipótese  $H_0 : \mu = 0$ .

TABLE XII: Dados dos 10 ativos com maiores classificações segundo o critério  $\gamma$ 

	Média	Desvio padrão	Teste KS	Teste de média
IGTA3	-9,2888e-05	0,065607	0	0
CPFE3	0,0079899	0,064076	0	0
BRAP4	0,017229	0,051451	0	0
JBSS3	-0,010466	0,060577	0	0
VALE3	-0,0005591	0,049304	0	0
TIMP3	-0,0022687	0,049281	0	0
LAME4	-0,0041428	0,04795	0	0
BBSE3	-0,0068587	0,05227	0	0
GOAU4	0,026906	0,037494	0	1
BRKM5	0,0013772	0,04528	0	0

Os resultados demonstram que não se pode rejeitar a hipótese de que os coeficientes de autocorrelação apresentem uma distribuição Gaussiana. Esse resultado corrobora com a hipótese de que os coeficientes de correlação sejam aleatórios e não possuam um padrão observável de comportamento. Adicionalmente, o teste das médias demonstra que não se pode rejeitar a hipótese de que os coeficientes apresentem uma média nula, exceto para o ativo GOAU4, cuja média amostral é 0,026906 sendo ainda relativamente próxima de zero.

As mesmas conclusões são obtidas da análise dos dados da tabela XIII, que mostra os dados dos 10 ativos com menor classificação segundo o critério  $\gamma$ .

TABLE XIII: Dados dos 10 ativos com menores classificações segundo o critério  $\gamma$ .

	Média	Desvio padrão	Teste KS	Teste de média
PETR3	0,0031139	0,028682	0	0
ELET3	0,007761	0,031653	0	0
ECOR3	0,0047194	0,035643	0	0
PCAR4	0,0088196	0,034576	0	0
CSAN3	0,0013971	0,033287	0	0
EMBR3	-0,0038492	0,034349	0	0
CMIG4	0,0099355	0,0343	0	0
SBS3	-0,013479	0,032567	0	1
FIBR3	0,007543	0,034083	0	0
BRML3	0,0023375	0,036276	0	0

Dessa forma, não existem evidências fortes o bastante para validar a hipótese de correlação serial ou periodicidade nos dados, com base na avaliação dos coeficientes de autocorrelação. A conclusão desta análise é que prevalece então a alternativa de que os dados sejam meramente aleatórios.

Por fim são apresentados os gráficos de *Lag Plot* para os ativos anteriormente examinados (IGTA3, CPFE3, BRAP4 e PETR4), na tentativa de identificar padrões que indiquem a presença de possíveis não aleatoriedades e correlações seriais. Essa abordagem gráfica oferece uma alternativa empírica para a análise das autocorrelações.

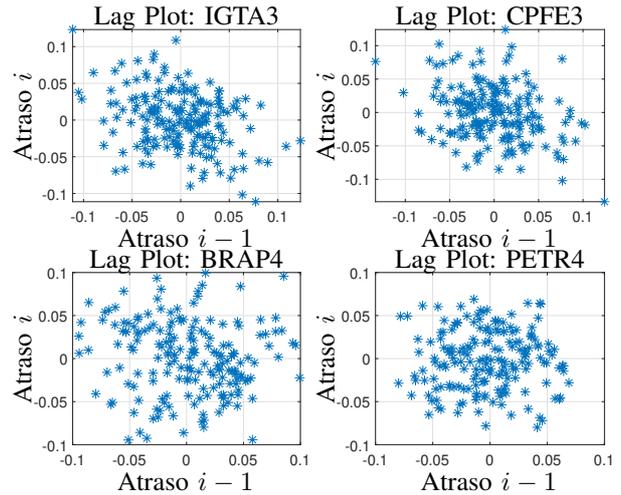


Fig. 48: Lag plot para os coeficientes de autocorrelação

Os gráficos da figura 48 mostram a presença de alguns poucos *outliers*, de amplitude relativamente baixa, e nenhum padrão mensurável nos coeficientes de autocorrelação. A existência de padrões mensuráveis surgiria nos gráficos na forma de padrões geométricos observáveis, o que ocorre.

### B. Teste de Run

O teste de *Run* pode ser utilizado para determinar se um conjunto de dados foi gerado a partir de um processo aleatório [49]. Este teste é apresentado com o mesmo intuito da análise anterior, oferecendo uma alternativa para aprofundar a investigação a respeito da previsibilidade das séries temporais financeiras.

**Definição 1 (Run):** Dado um valor de referência  $v$ , *run* é uma sequência de amostras, todas com valor superior ou inferior à  $v$ . A quantidade de amostras na sequência é o tamanho do *run*.

Para um processo aleatório, a probabilidade do valor de uma amostra  $t$  ser maior ou menor do que o da amostra  $t-1$  segue uma distribuição binomial, sendo esta suposição a base para o teste de *run*, que consiste em avaliar as seguintes hipóteses:

- $H_0$ : A sequência foi produzida por um processo aleatório
- $H_1$ : A sequência não foi produzida por um processo aleatório

A estatística do testes é calculada utilizando a equação 52.

$$Z = \frac{R - \bar{R}}{s_R} \quad (52)$$

onde:

- $Z$  é a quantidade total de *runs* na sequência;
- $\bar{R}$  é quantidade esperada de *runs*;
- $s_R$  é o desvio padrão teórico para número de *runs*.

A quantidade esperada de *runs* e o seu desvio padrão são calculados com base na distribuição binomial, utilizando as equações 53 e 54.

$$\bar{R} = \frac{2n_1n_2}{n_1 + n_2} + 1 \quad (53)$$

$$s_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 + 1)} \quad (54)$$

onde  $n_1$  e  $n_2$  representam os números de amostras positivas e negativas na série respectivamente.

A estatística do teste deve ser comparada com o valor crítico correspondente a uma distribuição normal para o nível de significância desejado. Para testes com pequenas amostras existem tabelas para determinação do valor crítico, tomando como base os valores de  $n_1$  e  $n_2$  [50].

Visando avaliar a aleatoriedade das séries dos retornos, foi aplicado um teste de *run* para todos os ativos da base de dados ajustada. Os resultados desses testes estão resumidos no gráfico da figura 49, onde a estatística do teste  $Z$  pode ser comparada com o valor crítico do teste  $Z_{1-\alpha/2}$ . Foi utilizada uma significância  $\alpha = 0,05$  para rejeição da hipótese nula, de aleatoriedade na sequência de dados e a comparação para determinação da quantidade de *runs* é feita utilizando o valor médio da série.

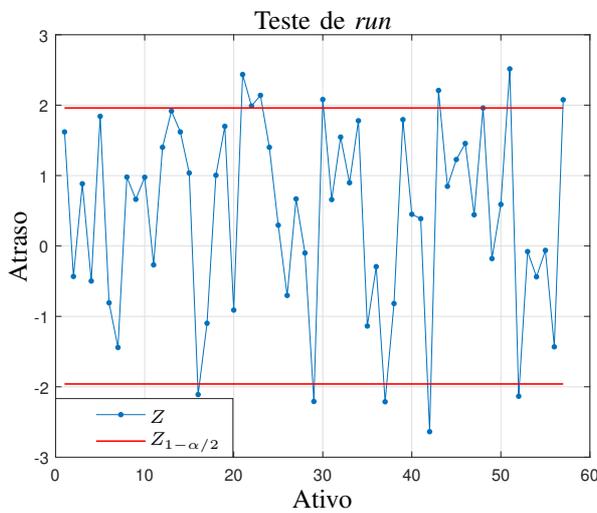


Fig. 49: Estatísticas do teste de *run* para os retornos dos ativos

O gráfico indica que para 12 ativos da base de dados, existem evidências de que os retornos não são aleatórios para o nível de significância adotado. Esses ativos são mostrados na tabela XIV, onde é possível também verificar o número de *runs* encontrado na série e o esperado, estatística do teste e  $\rho_{valor}$ .

TABLE XIV: Ativos para os quais o teste de *run* rejeita a hipótese de aleatoriedade da sequência.

	<i>n runs</i>	esperado	$Z$	$\rho_{valor}$
CSNA3	341	370	-2,1107	0,034732
EMBR3	404	370	2,4361	0,014793
ENBR3	398	370	1,9883	0,046716
EQTL3	400	370	2,1394	0,032339
HYPE3	340	370	-2,2082	0,027165
IGTA3	399	370	2,0807	0,037378
MGLU3	339	369	-2,2127	0,026874
PCAR4	334	370	-2,6374	0,0083143
PETR3	401	370	2,2091	0,027096
TIMP3	405	370	2,5154	0,011837
UGPA3	341	370	-2,135	0,03269
WEGE3	399	370	2,0769	0,037734

Examinando então a autocorrelação dos ativos listados na tabela XIV, no intuito de comprovar os resultados obtidos. Nota-se que o ativo que apresentou a maior evidência de não aleatoriedade nos dados segundo o teste de *run* foi o PCAR4.

Para este ativo é apresentado o gráfico de *Lag Plot* para os coeficientes de autocorrelação:

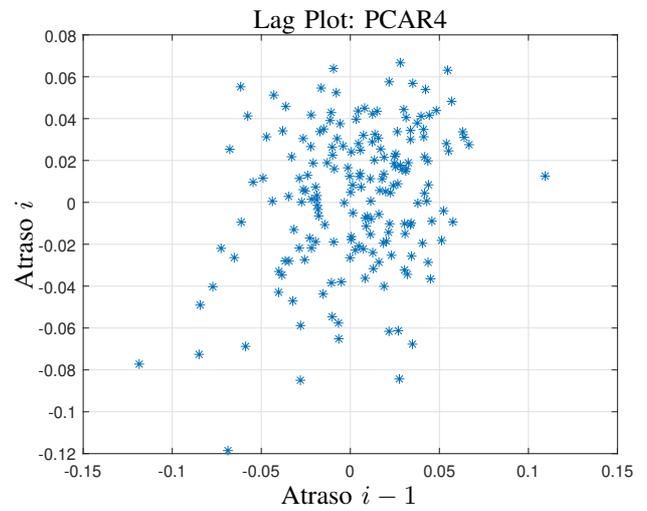


Fig. 50: Lag plot para os coeficientes de autocorrelação do ativo PCAR4

O gráfico da figura 50 mostra que, a menos de alguns poucos *outliers*, os coeficientes de autocorrelação estão dispersos no entorno da origem sem formar um padrão definido. Esse gráfico é um forte indicio de aleatoriedade nos dados (inexistência de padrão de correlação serial), porém, o teste de *run* rejeita a hipótese de aleatoriedade. Temos então uma indefinição com relação a este comportamento, não podendo aceitar prontamente os resultados do teste de *run*.

As figuras 51, 52 e 53, mostram os gráficos dos coeficientes de autocorrelação para os 12 ativos da tabela XIV, para os quais foi rejeitada a hipótese de aleatoriedade pelo teste de *run*. Esses gráficos corroboram com a conclusão anterior, de que não há padrão observável nos coeficientes de autocorrelação, mesmo quando o teste de *run* rejeita a hipótese de aleatoriedade.

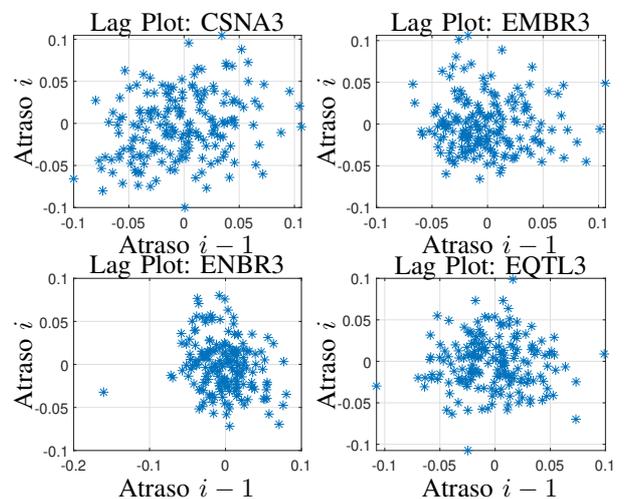


Fig. 51: Lag plot para os coeficientes de autocorrelação

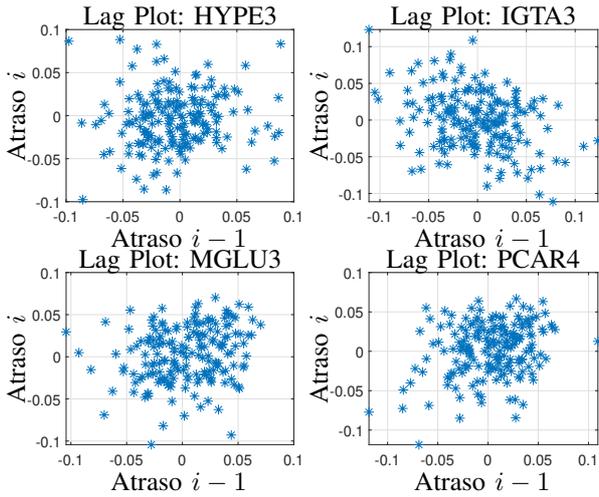


Fig. 52: Lag plot para os coeficientes de autocorrelação

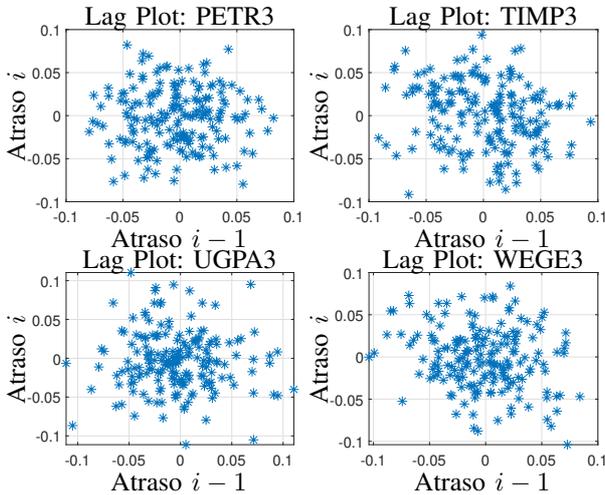


Fig. 53: Lag plot para os coeficientes de autocorrelação

Mesmo a aplicação do teste de *run* tendo rejeitado a hipótese de aleatoriedade nos dados, indicando a existência de algum tipo de correlação serial, outras análises, como a dos gráficos *Lag Plot*, contrariam essa hipótese. Dessa forma pode-se concluir que não há evidências suficientes que comprovem a existência de correlações seriais para os retornos dos ativos analisados. Como tomou-se nesse estudo uma base de dados suficientemente abrangente para representar o comportamento do mercado de capitais, essa conclusão pode ser estendida para qualquer ativo.

X. ANÁLISE NO DOMÍNIO DA FREQUÊNCIA

Visando avaliar a existência de periodicidades (sazonalidades) nos retornos dos ativos, a partir da análise no domínio da frequência, foi empregada a transformada rápida de Fourier às series temporais, para construção do espectro de frequências. Essa aplicação consiste em transformar a sequência de valores reais  $r(t) \in \mathcal{R}$ , dependentes do tempo, em uma sequência de valores complexos  $R(\omega) \in \mathcal{C}$  dependentes da frequência. A título de ilustração o gráfico da figura 54 representa o espectro

de frequências para o ativo PETR3<sup>4</sup>.

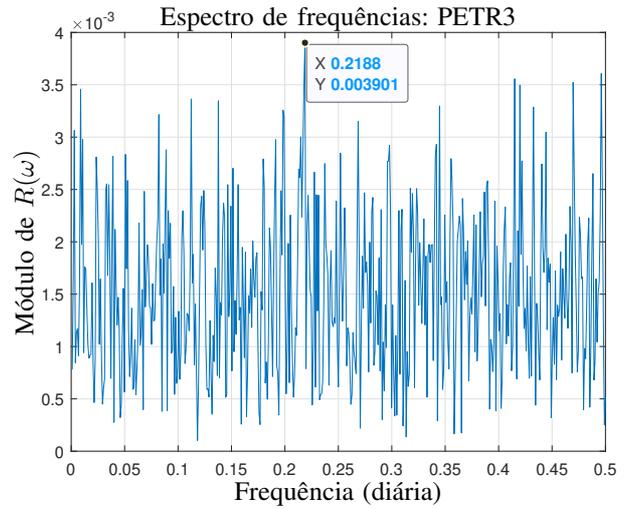


Fig. 54: Espectro de frequências para os retornos do ativo PETR3

Esse gráfico representa o espectro de frequências calculado para os retornos do ativo PETR3. Pode-se observar que não existem pontos no gráfico que expressem com clareza nenhum tipo de periodicidade nos valores de  $r(t)$ . O ponto máximo do sinal ocorre para uma frequência de 0,2188 (a cada 4,57 dias), porém esse pico é relativamente pequeno para representar de forma consistente alguma sazonalidade.

Resultado semelhante ao do gráfico da figura 54 é observado para os demais ativos da base de dados, e por não agregarem nova informação a análise não são apresentados. A avaliação dos espectros de frequência, de todos os ativos, não demonstra evidências de qualquer tipo de periodicidade nas séries temporais, corroborando com as análises anteriores.

XI. CONCLUSÕES

A partir das análises desenvolvidas, são obtidas as seguintes conclusões gerais:

- Os dados não apresentam indícios fortes de normalidade na sua distribuição;
- A distribuição 2-Gaussiana é aquela que melhor representa a distribuição empírica dos retornos.
- As séries temporais apresentam simetria, porém são todas leptocúrticas;
- A fortes evidências de heterocedasticidade, para todos os ativos;
- Os testes de raízes unitárias indicam estacionariedade das séries de retorno, contrariando a condição de homoscedasticidade.
- A análise dos coeficientes de correlação indica que existem ativos descorrelacionados da grande maioria dos demais, enquanto outros são fortemente correlacionados com diversos ativos, assim, a existência de correlação é uma propriedade de cada par de ativos e não extensível como propriedade geral das séries de retornos;

<sup>4</sup>A escala de frequências no gráfico é diária, da qual entende-se o inverso do período de um dia, ou seja:  $1/dia$ .

- Não há evidências significativas de correlação entre os retornos dos ativos e os volumes financeiros negociados um passo atrás, contrariando uma hipótese conhecida dos investidores do mercado financeiro.
- Demais correlações entre volatilidade e retornos também não foram constatadas na análise dos dados, contrariando a existência do suposto efeito *leverage*.
- A análise das autocorrelações não mostra nenhum padrão de correlação serial ou de dependência temporal nas séries de retornos. Também não há evidências de qualquer tipo de periodicidade;
- Os testes de *run* comprovam a aleatoriedade para a maioria dos ativos, aleatoriedade essa corroborada pela avaliação das autocorrelações;
- O espectro de frequências corrobora com os resultados da análise das autocorrelações, demonstrando que não há indícios de sazonalidade nas séries de retornos.

Em resumo, conclui-se que as séries de retornos são provenientes de um processo aleatório, com distribuição leptocúrtica, simétrica, não havendo autocorrelação ou dependência temporal nos dados. Dessa forma, fica demonstrado que o modelo mais adequado para representar o preço de um ativo é o caminho aleatório da equação 55, corroborando com o resultado obtido por Fama em [7] e acrescentando uma distribuição de probabilidades mais adequada ao modelo:

$$p_{t+1} = p_t (1 + \epsilon_t) \quad (55)$$

onde  $\epsilon_t$  é uma variável aleatória com distribuição 2-Gaussiana:

$$\epsilon_t = \mathcal{G}_2(\mu_1, \sigma_1, \mu_2, \sigma_2)$$

Alternativamente  $\epsilon_t$  pode apresentar uma distribuição de probabilidades conjunta, abrangendo as correlações entre os retornos de diferentes ativos.

Todavia, esse resultado não pode ser tomado como único e definitivo, pois a natureza aleatória das séries temporais financeiras não o permite, sendo o resultado aqui obtido fruto de uma criteriosa e minuciosa análise dos dados, com base em diversas ferramentas e métodos. Porém, pretendendo atribuir outro modelo diferente deste, deve-se provar, através de uma análise com o mesmo rigor, sua adequação aos dados.

## XII. BIBLIOGRAFIA

- [1] Louis Bachelier, “Théorie de la spéculation,” *Annales de l’Ecole Normale Supérieure*, pp. 21–86, 1900.
- [2] Eugene F. Fama, “Efficient capital markets: A review of theory and empirical work,” *The Journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
- [3] E.F. Brigham and J.F. Houston, *Fundamentals of Financial Management*, Cengage Learning, 2015.
- [4] Ricardo Goulart Serra; André Taue Saito; Luiz Paulo Lopes Fávero, “Nova metodologia do ibovespa, betas e poder explicativo dos retornos das ações,” *Revista de Contabilidade das Organizações*, vol. 21, pp. 71–85, 2016.
- [5] Cristina Elizabeth M. Hagler; Ricardo Dias de Oliveira Brito, “Sobre a eficiência dos índices de ações brasileiros,” *Revista de Administração*, vol. 42, no. 1, pp. 74–85, 2007.
- [6] BM&FBovespa, *Metodologia do Índice Bovespa*, Apr. 2015.
- [7] Eugene F. Fama, “The behaviour of stock market prices,” *Journal of Business*, pp. 34–105, 1965.
- [8] Eugene F. Fama, “Mandelbrot and the stable paretian hypothesis,” *The Journal of Business*, vol. 36, no. 4, pp. 420–429, 1963.
- [9] J. Huston McCulloch, “13 financial applications of stable distributions,” in *Statistical Methods in Finance*, vol. 14 of *Handbook of Statistics*, pp. 393 – 425. Elsevier, 1996.
- [10] S.J. Taylor, *Modelling Financial Time Series*, World Scientific, 2008.
- [11] G.W. Corder and D.I. Foreman, *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*, Wiley, 2011.
- [12] R.B. D’Agostino, *Goodness-of-Fit-Techniques*, CRC Press, 2017.
- [13] Metin Öner and İpek Deveci Kocakoç, “Jmasm 49: A compilation of some popular goodness of fit tests for normal distribution: Their algorithms and matlab codes (matlab),” *Journal of Modern Applied Statistical Methods*, vol. 16, pp. 547–575, 2017.
- [14] George Marsaglia, Wai Wan Tsang, and Jingbo Wang, “Evaluating Kolmogorov’s Distribution,” *Journal of Statistical Software*, vol. 8, no. 118, 2003.
- [15] Hubert W. Lilliefors, “On the kolmogorov-smirnov test for normality with mean and variance unknown,” *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.
- [16] Gerard E. Dallal and Leland Wilkinson, “An analytic approximation to the distribution of lilliefors’s test statistic for normality,” *The American Statistician*, vol. 40, no. 4, pp. 294–296, 1986.
- [17] T. W. Anderson and D. A. Darling, “Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes,” *Ann. Math. Statist.*, vol. 23, no. 2, pp. 193–212, 06 1952.
- [18] Harald Cramér, “On the composition of elementary errors,” *Scandinavian Actuarial Journal*, vol. 1928, no. 1, pp. 13–74, 1928.
- [19] R. Von Mises and Verlag von Julius Springer, *Wahrscheinlichkeit, Statistik und Wahrheit: Einführung in die neue Wahrscheinlichkeitslehre und ihre Anwendung*, Schriften zur Wissenschaftlichen Weltauffassung. Verlag von Julius Springer, 1928.
- [20] T. W. Anderson, “On the distribution of the two-sample cramer-von mises criterion,” *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1148–1159, 09 1962.
- [21] E.S. Pearson and H.O. Hartley, *Biometrika Tables for Statisticians*, vol. 2, University Press, Cambridge, 1972”.
- [22] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.
- [23] J. P. Royston, “An extension of shapiro and wilk’s w test for normality to large samples,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 31, no. 2, pp. 115–124, 1982.
- [24] S. S. Shapiro and R. S. Francia, “An approximate analysis of variance test for normality,” *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 215–216, 1972.
- [25] Patrick Royston, “A pocket-calculator algorithm for the shapiro-francia test for non-normality: An application to medicine,” *Statistics in Medicine*, vol. 12, no. 2, pp. 181–184, 1993.
- [26] Carlos M. Jarque and Anil K. Bera, “A test for normality of observations and regression residuals,” *International Statistical Review / Revue Internationale de Statistique*, vol. 55, no. 2, pp. 163–172, 1987.
- [27] Ralph D’Agostino and E. S. Pearson, “Tests for departure from normality. empirical results for the distributions of b2 and sqrt(b1),” *Biometrika*, vol. 60, no. 3, pp. 613–622, 1973.
- [28] César Daltoé Berci and Celso Pascoli Bottura, “Distribuições d-normais,” *Em preparação*, 2020.
- [29] Wilton de Oliveira Bussab, *Estatística Básica*, Editora Saraiva, 3ª edição edition, 23013.
- [30] Frank Jonk Massey Wilfrid J. Dixon (Author), *Introduction to Statistical Analysis*, McGraw-Hill, New York, 4ª edição edition, 1957.
- [31] R.D. Yates and D.J. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers*, Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers. Wiley, 2014.
- [32] Ruey S. Tsay, *Analysis of Financial Time Series*, Wiley & Sons, 2010.
- [33] James Douglas Hamilton, *Time series analysis*, Princeton Univ. Press, Princeton, NJ, 1994.
- [34] Charles R. Nelson and Charles R. Plosser, “Trends and random walks in macroeconomic time series: Some evidence and implications,” *Journal of Monetary Economics*, vol. 10, no. 2, pp. 139 – 162, 1982.
- [35] C.W.J. Granger and P. Newbold, “Spurious regressions in econometrics,” *Journal of Econometrics*, vol. 2, no. 2, pp. 111 – 120, 1974.
- [36] Denis Kwiatkowski, Peter C.B. Phillips, Peter Schmidt, and Yongcheol Shin, “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?,” *Journal of Econometrics*, vol. 54, no. 1, pp. 159 – 178, 1992.
- [37] Russell Davidson and James G. MacKinnon, *Econometric theory and methods*, Oxford Univ. Press, New York, NY [u.a.], 2004.
- [38] David A. Dickey and Wayne A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427–431, 1979.
- [39] Peter Schmidt Denis Kwiatkowski, Peter C.B. Phillips and Yongcheol Shin, “Testing the null hypothesis of stationarity against the alternative of a unit root,” *Journal of Econometrics*, vol. 54, pp. 159–178, 1992, North-Holland.

- [40] Karl Pearson, *The life, letters and labours of Francis Galton*, vol. 1, Cambridge [Eng.]University press., 1930.
- [41] Nian Shong Chok, "Pearson's versus spearman's and kendall's correlation coefficients for continuous data," M.S. thesis, University of Pittsburgh, 2010.
- [42] C. Spearman, "The proof and measurement of association between two things," *The American Journal of Psychology*, vol. 15, no. 1, pp. 72–101, 1904.
- [43] M.G. Kendall, *Rank Correlation Methods*, Griffin books on statistics. Griffin, 1962.
- [44] Jean Dickinson Gibbons and Subhabrata Chakraborti, *Nonparametric Statistical Inference*, pp. 977–979, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [45] Hung Xuan Do, Robert Brooks, Sirimon Treepongkaruna, and Eliza Wu, "How does trading volume affect financial return distributions?," *International Review of Financial Analysis*, vol. 35, pp. 190 – 206, 2014.
- [46] Jean-Philippe Bouchaud, Andrew Matacz, and Marc Potters, "Leverage effect in financial markets: The retarded volatility model," *Phys. Rev. Lett.*, vol. 87, pp. 228701, Nov 2001.
- [47] Stephen Figlewski and Xiaozu Wang, "Is the 'leverage effect' a leverage effect?," *SSRN Electronic Journal*, 11 2000.
- [48] Tara Keshar Nanda Baidya Paulo Henrique Soto Costa, "Propriedades estatísticas das séries de retornos das principais ações brasileiras," *Pesquisa Operacional*, vol. 21, no. 1, pp. 61–87, June 2001.
- [49] J.V. Bradley and Prentice Hall, *Distribution-free Statistical Tests*, Prentice-Hall, 1968.
- [50] William Mendenhall and James Reinmuth, *Statistics for Management and Economics*, Duxbury Press, fourth edition edition, 1982.



**César Daltoé Berci** Possui graduação em Engenharia Elétrica pela Universidade Estadual do Oeste do Paraná (2005), Mestrado em Engenharia Elétrica pela Universidade Estadual de Campinas (2008) e atualmente cursando Doutorado em Engenharia Elétrica nessa universidade. Tem experiência nas áreas de Engenharia Elétrica e Mecânica, atuando na indústria, no ensino superior e prestando consultorias, com ênfase em desenvolvimento de equipamentos eletrônicos, simulação computacional de sistemas mecânicos, térmicos, elétricos e eletrônicos,

análise computacional de dados, desenvolvimento de software embarcado, aplicativos mobile, WEB e software PC, desenvolvimento de gabinetes mecânicos e mecanismos, análise, comissionamento e projetos de sistemas elétricos de potência, diligenciamento técnico voltado ao setor elétrico, desenvolvimento de sistemas inteligentes, automação industrial, predial e residencial.



**Celso Pascoli Bottura** possui graduação em Engenharia Aeronáutica pelo Instituto Tecnológico de Aeronáutica (1962), mestrado em Engenharia Mecânica pela *Purdue University* (1964) e doutorado em Engenharia Elétrica pela Universidade Estadual de Campinas (1973). Atualmente é professor Titular Aposentado da Faculdade de Engenharia Elétrica e Computação da mesma universidade. Tem experiência na área de Engenharia Elétrica, com ênfase em Eletrônica Industrial, Sistemas e Controles Eletrônicos, atuando principalmente nos seguintes

temas: controle, máquina de indução, caos, controle automático e controle de máquinas, identificação multivariável, jogos dinâmicos e controle multi-agente, modelagem computacional de dados e controle inteligente no espaço de estado, observadores de estado e controladores geométricos inteligentes de sistemas dinâmicos não-lineares.