

Vinícius da Silva Garcia

Análise de ativos financeiros a partir da teoria de matrizes aleatórias

Porto Alegre

2022

Vinícius da Silva Garcia

Análise de ativos financeiros a partir da teoria de matrizes aleatórias

Trabalho apresentado ao curso de Engenharia Física do Instituto de Física e da Escola de Engenharia da UFRGS como requisito obrigatório para aprovação da disciplina de Projeto Final de Curso em Engenharia Física II.

Universidade Federal do Rio Grande do Sul – UFRGS

Instituto de Física e Escola de Engenharia

Engenharia Física

Orientador: Dr. Carlo Requião da Cunha

Porto Alegre

2022

Lista de ilustrações

Figura 1	– Gráfico da distribuição de autovalores de Marčenko-Pastur para algumas proporções $\frac{n}{m}$. No eixo x temos o espaço de autovalores λ e no eixo y temos a densidade de probabilidade $p(\lambda)$	18
Figura 2	– Distribuição de Tukey lambda para alguns valores de λ . Imagem de IkamusumeFan, licença CC BY-SA 4.0 < https://creativecommons.org/licenses/by-sa/4.0 >, obtida de Wikimedia Commons.	21
Figura 3	– No topo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. Embaixo, em roxo, temos os preços de fechamento do ETF SPY.	25
Figura 4	– No topo, em verde, temos os preços de fechamento do ETF SPY, subtraído de sua tendência de queda. Embaixo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. A correlação de Pearson entre essas séries temporais é de $\rho = -0.40$	26
Figura 5	– Em cada figura, no topo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. Embaixo, em roxo, temos os preços de fechamento do SPY. Em preto tracejado temos linhas verticais nas datas de interesse para facilitar a visualização. Na figura 5a temos destaque nas datas 24/02/2020 e 11/06/2020. Na figura 5b temos destaque nas datas 1º/09/2020 e 26/10/2020	27
Figura 6	– No topo, em verde, temos o melhor λ da distribuição de Tukey lambda ajustado aos retornos dos ativos, com uma linha tracejada em cinza indicando $\lambda = 0.14$, situação em que a distribuição é aproximadamente normal. Embaixo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. A correlação entre essas duas séries temporais é 0.39.	28
Figura 7	– No topo, em azul, temos a série temporal da correlação média entre os ativos. Embaixo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. A correlação entre essas duas séries temporais é 0.73.	29

Agradecimentos

Agradeço aos meus pais, à minha irmã e a meu cunhado por não só tolerarem, mas por incentivarem e cultivarem a minha curiosidade desde pequeno.

À minha namorada, pelo apoio diário durante toda a graduação. Por estar sempre disponível para conversar, discutir e divergir, fosse sobre física ou fosse sobre a vida.

Ao meu orientador, Carlo, pelas correções e direcionamentos durante o desenvolvimento deste trabalho. Também, tendo sido meu professor em várias disciplinas, agradeço pelos ensinamentos.

Aos amigos da UFRGS, por terem compartilhado o dia-a-dia do curso, com as alegrias e dificuldades que o caminho proporcionou.

Aos professores que passaram pelo meu caminho, pelas mais diferentes abordagens e visões de mundo que proporcionaram que eu expandisse meus horizontes.

Aos professores e outros profissionais do Instituto de Física e da Escola de Engenharia, pelo esforço e dedicação em tornar e manter o curso de Engenharia Física realidade.

À UFRGS, instituição de ensino federal pública e gratuita, pela graduação.

Resumo

Apesar de o mercado financeiro ser imprevisível, sendo os retornos de ativos muitas vezes modelados como variáveis aleatórias normalmente distribuídas, sabemos que diferentes ativos possuem correlações entre si e também que suas distribuições de retorno não são estritamente normais. A fim de investigar o comportamento de ativos financeiros – sobretudo durante crises – utilizamos como base a teoria de matrizes aleatórias, que entre outras coisas descreve o comportamento assintótico de conjuntos de variáveis aleatórias independentes normalmente distribuídas. Dessa teoria, utilizamos a distribuição de Marčenko-Pastur como uma referência à qual comparamos uma distribuição empírica relacionada aos retornos de ativos financeiros reais. A partir disso, verificamos que os retornos dos ativos divergem do que é esperado para variáveis aleatórias independentes. Quantificando esta divergência através da divergência de Kullback-Leibler, mostramos que ela é maior em períodos de crise. Comparando com a distribuição de Tukey-lambda, verificamos que essa divergência não é causada pelo aumento do peso das caudas nas crises. Os resultados indicam que a principal causa do crescimento da divergência é o crescimento da correlação dos ativos nestes períodos de crise.

Palavras-chaves: teoria de matrizes aleatórias. econofísica. crise financeira. correlação.

Abstract

Although the financial market is unpredictable, and asset returns are often modeled as normally distributed random variables, we know that different assets have correlations with each other and also that their return distributions are not strictly normal. In order to investigate the behavior of financial assets – especially during crises – we used the framework of random matrix theory, which among other things describes the asymptotic behavior of sets of normally distributed independent random variables. From this theory, we use the Marčenko-Pastur distribution as a benchmark to compare to an empirical distribution derived from the returns of real financial assets. From this, we found that asset returns behavior diverge from what is expected for independent random variables. By quantifying this divergence through the Kullback-Leibler divergence, we show that it is greater in periods of crisis. Comparing with the Tukey-lambda distribution, we verified that this divergence is not caused by the increase in the weight of the tails in the crises. The results indicate that the main cause of the divergence growth is the growth of the asset correlation in these periods of crisis.

Keywords: random matrix theory. econophysics. financial crisis. correlation.

Sumário

1	INTRODUÇÃO	13
1.1	Contexto socioeconômico	13
1.2	Ideia central	14
1.2.1	Organização	15
2	METODOLOGIA	17
2.1	Fundamentação teórica	17
2.1.1	Distribuição de Marčenko-Pastur	17
2.1.2	Distribuição de autovalores empírica	18
2.1.3	Divergência de Kullback-Leibler	20
2.1.4	Distribuição de Tukey lambda	20
2.2	Caracterização	21
3	RESULTADOS	25
4	CONCLUSÕES	31
4.1	Trabalhos futuros	31
	REFERÊNCIAS	33

1 Introdução

1.1 Contexto socioeconômico

O valor de um ativo financeiro pode variar, a depender de fatores diversos. Para um investidor, é de interesse que seus ativos valorizem com o passar do tempo, gerando lucro. Para sustentar tomadas de decisão, o investidor tem à sua disposição toda sorte de análise financeiras, as quais são vendidas ou oferecidas gratuitamente (como estratégia de marketing) por corretoras de valores, casas de *research* e analistas individuais. As análises vão desde recomendações de operações com ativos específicos, passando por análises dos mercados financeiros, até análises macroeconômicas e geopolíticas. Normalmente essas ofertas de análises acompanham a promessa de que a análise poderá resultar na compra de ativos mais lucrativos. Apesar destas promessas serem discutíveis, no mínimo estes produtos podem oferecer ideias e informações para auxiliar nas tomadas de decisão de investidores.

Ainda que a oferta de análises financeiras seja vasta, a grande maioria delas possui viés puramente econômico ou econométrico, sendo também estes os tipos mais comuns de análise em gestão de fundos de investimento¹. Análises interdisciplinares, englobando mas não se limitando a econometria tradicional, são mais raras (um exemplo de empresa que presta serviços mais amplos é a RAND Corporation²), mas podem oferecer novas visões de problemas econômicos.

Um dos campos de estudo interdisciplinares envolvendo economia é a econofísica, que engloba a abordagem de sistemas econômicos pela ótica da física, oferecendo ferramentas que complementam as abordagens mais tradicionais (1). O termo foi cunhado por Eugene Stanley, um dos autores do primeiro livro sobre o tema (2), referenciando e englobando trabalhos que já vinham sendo feitos nas décadas anteriores.

Apesar de ferramentas deste campo serem usadas (não necessariamente de maneira exclusiva) em alguns fundos quantitativos³, ainda é uma área que, mesmo sendo bem estabelecida na pesquisa científica, é incipiente no mercado. Não foram encontradas no Brasil análises sendo vendidas com claro cunho econofísico. Neste sentido, análises com este viés podem oferecer um diferencial tanto técnico quanto de *marketing* para o desenvolvimento de produtos.

¹ Veja alguns exemplos de estratégias de fundos de investimento, <www.investopedia.com/articles/investing/111313/multiple-strategies-hedge-funds.asp>, acesso em 16/04/2022.

² RAND Corporation <www.rand.org>, acesso em 16/04/2022

³ Glossário Mais Retorno: Econofísica. Disponível em <<https://maisretorno.com/portal/termos/e/econofisica>>, acesso em 17/04/2022.

Análises de crises financeiras podem ser especialmente valiosas em cenários de crise. Crises do mercado financeiro são estressantes para a sociedade, provocando impactos econômicos negativos a nível global (3). O desemprego nos anos subsequentes à crise do subprime de 2008 cresceu no mundo todo, com impactos mais intensos nos países desenvolvidos, subindo de 5% em dezembro de 2007 para 10% em outubro de 2009 nos Estados Unidos⁴. A crise da dívida pública da Zona do Euro atingiu Espanha, Grécia, Irlanda, Itália e Portugal, que tiveram problemas com os pagamentos e refinanciamento dos seus títulos de dívida, necessitando de apoio do Banco Central Europeu e do Fundo Monetário Internacional⁵. Análises mais completas que possam apoiar tomadores de decisões nestes momentos de grande estresse econômico devem ser relevantes.

1.2 Ideia central

Apesar de ser comum a fim de simplificação considerar que retornos de ativos financeiros seguem uma distribuição normal, é bem estabelecido na comunidade científica que isso nem sempre é verdade e frequentemente é verificável que os retornos seguem distribuições de caudas pesadas (4), como reconhecido por Mandelbrot em 1963 (5). Além disso, ativos financeiros normalmente não são independentes entre si. Retornos de diferentes ativos podem possuir correlações, as quais podem variar consideravelmente ao longo do tempo e de acordo com momentos de mercado (6).

A ideia central deste trabalho é realizar a caracterização do mercado financeiro à luz da teoria de matrizes aleatórias. Esta teoria, desenvolvida no contexto da física quântica (com exemplos de trabalho de Eugene Wigner (7, 8)) com contribuições na formalização por Marčenko e Pastur (9), é amplamente utilizada para o estudo de correlação em sistemas complexos cujas correlações são desconhecidas (10). Apesar da aplicação dessa teoria na compreensão da correlação entre ativos financeiros não ser inédita (11, 12, 13), não encontramos suas aplicações em produtos de análise financeira.

Para a análise a ser realizada neste trabalho, vamos selecionar séries temporais de vários ativos, montar matrizes de correlação entre eles, e a partir disso realizar comparações entre o que a teoria prevê para matrizes aleatórias e o que obtemos para essas séries temporais. Como sabemos que as séries temporais de ativos não são independentes entre si e podem possuir caudas pesadas, deve haver uma divergência entre os dados empíricos e a previsão teórica, e a medida dessa divergência pode oferecer uma maior compreensão das crises financeiras. Propõe-se também que esta divergência pode ser usada como um indicador financeiro.

⁴ International Labour Organisation 2011, Global employment trends 2011: the challenge of a jobs recovery, ILO, Geneva, <http://www.ilo.org/global/publications/books/WCMS_150440/lang-en/index.htm>, acesso em 04/03/2022.

⁵ Duignan, Brian. Great Recession. Encyclopedia Britannica, 16 de setembro de 2019, <<https://www.britannica.com/topic/great-recession>>, acesso em 04/03/2022.

1.2.1 Organização

Na seção 2.1 vamos estabelecer os conceitos teóricos necessários para a análise almejada. Na seção 2.1.1 introduziremos a distribuição de Marčenko-Pastur e seu papel na teoria de matrizes aleatórias e na seção 2.1.2 explicaremos como calcular uma distribuição empírica (comparável à de Marčenko-Pastur) usando o método da função de Green. Já na seção 2.1.3 vamos explicar a divergência de Kullback-Leibler, que será usada para compararmos a distribuição teórica de Marčenko-Pastur com a distribuição empírica. Na seção 2.1.4 apresentaremos a distribuição de Tukey lambda, que usaremos para estimar o peso das caudas dos retornos dos ativos.

Na sequência, na seção 2.2 apresentaremos os dados, as ferramentas e o passo-a-passo da análise, que engloba o período da crise do coronavírus.

Os resultados são apresentados no capítulo 3, onde comparamos séries temporais de algumas medidas construídas ao longo do trabalho.

Ao final, no capítulo 4, discutiremos os impactos dos resultados e apresentaremos as conclusões. Na seção 4.1 indicaremos possibilidades de trabalhos futuros.

2 Metodologia

2.1 Fundamentação teórica

2.1.1 Distribuição de Marčenko-Pastur

A partir de n períodos de tempo e m ativos financeiros, podemos calcular os retornos logarítmicos, utilizando os preços de fechamento de cada período, como:

$$r_{m,n} = \log(p_{m,n+1}) - \log(p_{m,n}) \quad (2.1)$$

para o n -ésimo período do m -ésimo ativo. Nós então calculamos os retornos normalizados como:

$$r'_{m,n} = \frac{r_{m,n} - \langle r_m \rangle}{\sqrt{\langle r_m^2 \rangle - \langle r_m \rangle^2}} \quad (2.2)$$

sendo a média feita no tempo.

Com isso, podemos construir matrizes de retornos \mathbf{X} definidas em $\mathbb{R}^{n \times m}$ com elementos $r'_{m,n}$. Delas, podemos obter a matriz de correlação definida em $\mathbb{R}^{m \times m}$ como $\mathbf{\Lambda} = \frac{1}{n\sigma} \mathbf{X}^T \mathbf{X}$, sendo σ o desvio padrão. Como os retornos estão normalizados para $\sigma = 1$, pode-se omiti-lo da equação. Esta matriz de correlação deve corresponder a uma transformação limitada em um espaço de Banach.

Para analisar séries temporais à luz da teoria de matrizes aleatórias, precisamos ter uma referência teórica de como matrizes aleatórias se comportam. A distribuição (por vezes chamada de lei) do semicírculo de Wigner, que estabelece a distribuição de autovalores de matrizes simétricas com valores independentes e aleatórios, não se aplica a matriz de correlação construída porque os elementos desta, por definição, não são independentes.

Para nosso cenário, então, podemos usar para comparação a distribuição de Marčenko-Pastur. No estudo de matrizes aleatórias, Marčenko e Pastur (9) provaram o teorema que estabelece o comportamento assintótico da distribuição de autovalores da correlação de matrizes aleatórias. Para variáveis aleatórias independentes e seguindo uma distribuição normal padrão, a densidade de autovalores da sua matriz de correlação em $\mathbb{R}^{m \times m}$ é dada por:

$$\rho(\lambda) = \frac{m}{2\pi n} \frac{\sqrt{(\lambda - \lambda_{\min})(\lambda_{\max} - \lambda)}}{\lambda}, \quad (2.3)$$

onde

$$\lambda_{\max, \min} = \left(1 \pm \sqrt{\frac{n}{m}}\right)^2 \quad (2.4)$$

quando $n, m \rightarrow \infty$ mantendo a proporção $\frac{n}{m}$ constante. Esse comportamento assintótico também pode ser obtido quando calculamos os autovalores de um ensemble de matrizes aleatórias de tamanho finito. Esta distribuição é a referência teórica com a qual iremos comparar nossos dados.

Na figura 1 podemos visualizar a distribuição de Marčenko-Pastur para alguns valores da proporção $\frac{n}{m}$. Fora do intervalo entre λ_{\min} e λ_{\max} definidos pela equação 2.4 a densidade de probabilidade é zero.

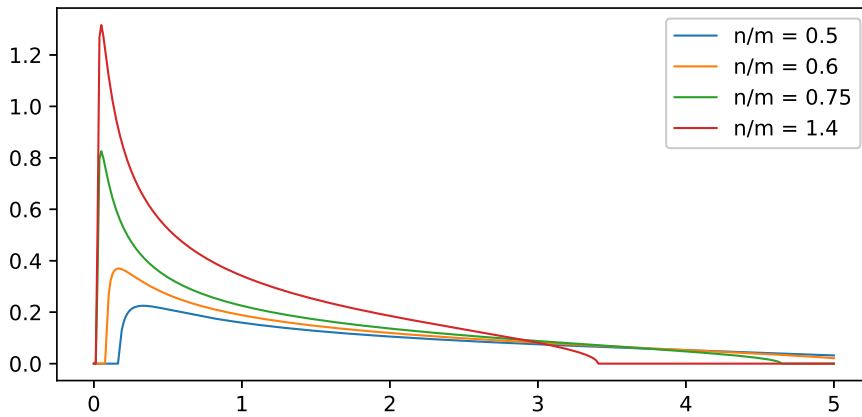


Figura 1 – Gráfico da distribuição de autovalores de Marčenko-Pastur para algumas proporções $\frac{n}{m}$. No eixo x temos o espaço de autovalores λ e no eixo y temos a densidade de probabilidade $p(\lambda)$.

2.1.2 Distribuição de autovalores empírica

Para que possamos fazer comparação com a teoria, precisamos estimar a distribuição de autovalores empírica. Para fazer essa comparação, alguns autores computam diretamente a distribuição experimental através de histogramas (11). Neste trabalho, porém, vamos seguir a abordagem de outros autores, utilizando o método de função de Green para encontrar a distribuição de autovalores (14), como se segue.

Dada a matriz de correlação $\mathbf{\Lambda} = \frac{1}{n} \mathbf{X}^T \mathbf{X}$ definida em $\mathbb{R}^{m \times m}$, o resolvente (equivalente a discretização de uma função de Green retardada) associado é calculado como (14):

$$\mathbf{G}_\eta^R(\lambda) = [(\lambda + i\eta) \mathbf{I} - \mathbf{\Lambda}]^{-1}, \quad (2.5)$$

onde $\eta \rightarrow 0^+$. Considerando $|m\rangle$ autovetor da matriz de correlação $\mathbf{\Lambda}$ com respectivos

autovalores λ_m , temos:

$$\begin{aligned}
\mathbf{G}_\eta^R(\lambda)|m\rangle &= (\lambda + i\eta)^{-1} \left[\mathbf{I} - \frac{\mathbf{\Lambda}}{\lambda + i\eta} \right]^{-1} |m\rangle \\
&= (\lambda + i\eta)^{-1} \sum_{j=0}^{\infty} \frac{1}{(\lambda + i\eta)^j} \mathbf{\Lambda}^j |m\rangle \\
&= (\lambda + i\eta)^{-1} \left(\sum_{j=0}^{\infty} \frac{1}{(\lambda + i\eta)^j} \lambda_m^j \right) |m\rangle \\
&= (\lambda + i\eta)^{-1} \left(1 - \frac{\lambda_m}{\lambda + i\eta} \right)^{-1} |m\rangle \\
&= \frac{1}{(\lambda - \lambda_m) + i\eta} |m\rangle,
\end{aligned} \tag{2.6}$$

onde a expressão da equação 2.5 foi substituída pela sua respectiva representação em série geométrica.

Usando a relação de proximidade, podemos escrever a função de Green na base de autovetores da matriz de correlação:

$$\begin{aligned}
\mathbf{G}_\eta^R(\lambda) &= \mathbf{G}_\eta^R(\lambda) \mathbf{1} \\
&= \sum_{m=0}^{M-1} [(\lambda + i\eta)\mathbf{I} - \mathbf{\Lambda}]^{-1} |m\rangle \langle m| \\
&= \sum_{m=0}^{M-1} |m\rangle \frac{1}{(\lambda - \lambda_m) + i\eta} \langle m| \\
&= \sum_{m=0}^{M-1} |m\rangle \mathbf{C}_m,
\end{aligned} \tag{2.7}$$

onde:

$$\mathbf{C}_m = \frac{\langle m|}{(\lambda - \lambda_m) + i\eta} \tag{2.8}$$

Assim, definimos a função de Green para o operador $\mathbf{\Lambda}$ como:

$$\mathbf{G}_\eta^R(\lambda) = \sum_{m=0}^{M-1} \frac{|m\rangle \langle m|}{(\lambda - \lambda_m) + i\eta} \tag{2.9}$$

Se tomamos o limite $\eta \rightarrow 0^+$:

$$\lim_{\eta \rightarrow 0^+} \text{Im} \left\{ \langle n | \mathbf{G}_\eta^R | n \rangle \right\} = -\pi \delta(\lambda - \lambda_n) \tag{2.10}$$

Assim, podemos calcular a densidade de probabilidade dos autovalores da matriz de correlação como:

$$\begin{aligned}
\rho(\lambda) &= \frac{1}{M} \sum_{n=0}^{M-1} \delta(\lambda - \lambda_n) \\
&= -\frac{1}{\pi M} \lim_{\eta \rightarrow 0} \text{Tr} \left\{ \text{Im} \left\{ \mathbf{G}_\eta^R(\lambda) \right\} \right\}
\end{aligned} \tag{2.11}$$

A partir da expressão de densidade de autovalores (equação 2.11) é possível computar a distribuição empírica sem a utilização de histogramas, que é o que foi feito para este trabalho.

2.1.3 Divergência de Kullback-Leibler

A fim de visualizar o quanto a distribuição de retornos se afasta de uma distribuição normal e também se afasta da independência entre ativos queremos comparar a densidade de autovalores empírica com a que seria esperada a partir da densidade descrita na equação 2.3. Para essa comparação, vamos utilizar a divergência de Kullback-Leibler. Esta divergência quantifica a diferença entre duas distribuições de probabilidade (15). Considerando duas distribuições de probabilidade P e Q contínuas com densidade de probabilidade $p(x)$ e $q(x)$ a divergência de Kullback-Leibler $D_{KL}(P||Q)$ de Q para P é definida como:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \quad (2.12)$$

Essa divergência é também chamada de entropia relativa e pode ser interpretada à luz da teoria da informação como uma medida do excesso de informação que se tem ao usar Q em vez da distribuição teórica P . Quando esta divergência é zero, ambas as distribuições possuem a mesma quantidade de informação.

É importante notar que esta não é uma métrica. Métricas, por definição, devem atender a alguns axiomas, incluindo o da simetria ($m(x, y) = m(y, x)$, sendo m uma métrica qualquer). Em geral a divergência de Kullback-Leibler de Q para P é diferente da divergência de P para Q , não sendo portanto uma métrica.

2.1.4 Distribuição de Tukey lambda

Como mencionado na seção 1.2, os retornos do mercado financeiro demonstram em suas distribuições empíricas caudas mais pesadas do que se esperaria de uma variável aleatória que segue distribuição normal. Um indicador para o peso das caudas pode ser encontrado ao realizarmos o ajuste dos dados à distribuição de Tukey lambda.

Esta distribuição, simétrica e centrada em zero, é definida em termos da função quantil e tem apenas um parâmetro de forma λ . Explicitando:

$$Q(p; \lambda) = \begin{cases} \frac{1}{\lambda} [p^\lambda - (1-p)^\lambda], & \text{se } \lambda \neq 0 \\ \log\left(\frac{p}{1-p}\right), & \text{se } \lambda = 0 \end{cases} \quad (2.13)$$

sendo p a probabilidade $Q(p)$ o quantil para determinada probabilidade.

$\lambda = -1$	Aproximadamente Cauchy
$\lambda = 0$	Exatamente logística
$\lambda = 0.14$	Aproximadamente normal

A densidade de probabilidade para a maior parte do espaço do parâmetro λ só pode ser encontrada numericamente. Essa distribuição é comumente usada para identificar qual a distribuição apropriada para determinado conjunto de dados. Isso pode ser feito porque se sabe que distribuições conhecidas são idênticas ou muito próximas à distribuição de Tukey lambda para determinados valores de λ (16):

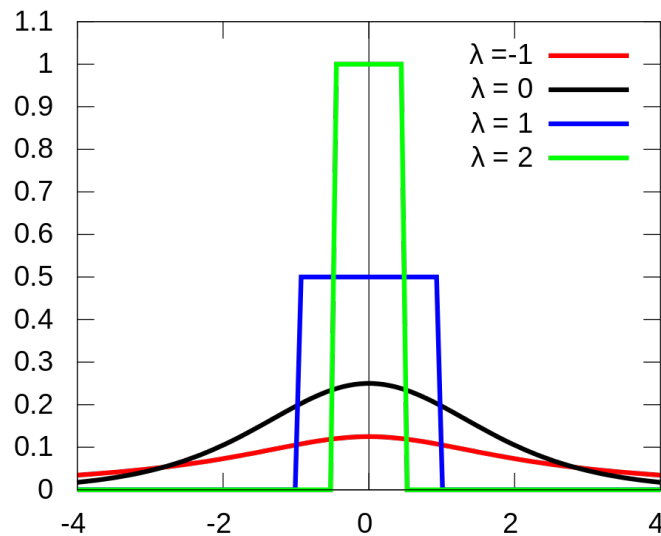


Figura 2 – Distribuição de Tukey lambda para alguns valores de λ . Imagem de IkamusumeFan, licença CC BY-SA 4.0 <<https://creativecommons.org/licenses/by-sa/4.0/>>, obtida de Wikimedia Commons.

Alguns exemplos podem ser observados na figura 2. É possível observar que a medida que aumenta o valor de λ as caudas ficam mais leves. A distribuição de Cauchy possui caudas mais pesadas que a distribuição logística que por sua vez possui caudas mais pesadas que a distribuição normal. Dessa forma, ao encontrarmos o lambda que melhor se ajusta aos nossos dados podemos usá-lo como uma medida do peso da cauda da distribuição e o quão diferente os dados ajustados são de uma normal.

2.2 Caracterização

O período escolhido para análise – de 19 de setembro de 2019 a 8 de março de 2022 – contém a crise do coronavírus em 2020, um acontecimento recente e com impactos presentes ainda hoje. Escolhemos para a análise as ações BAC, MSFT, BBY, GOOGL, XOM, AAPL, AMZN, ACN, AXP, T, JNJ, KMB, KHC, MCD, D, KO, INTC, MDLZ, WMT e V, listadas em bolsas dos Estados Unidos. Essas são as ações de algumas das maiores empresas do mundo, de vários setores (tecnologia, finanças, varejo, entre outras),

e fazem parte do índice S&P500. Escolhemos este mercado por ser o maior do mundo, que além de ser representativo oferece uma maior facilidade de aquisição de dados intradia. O índice, por sua vez, indexa 500 das maiores empresas listadas nos Estados Unidos, cobrindo 80% da capitalização de mercado disponível nesse país¹.

Para cada dia ao longo de uma janela de 622 dias de negociação, fizemos a aquisição de dados de preços de ativos minuto a minuto, construímos um ensemble de matrizes de correlação a partir de subconjuntos da série temporal minuto a minuto dos retornos logarítmicos normalizados dos ativos utilizando reamostragem *bootstrapping*. Com isso, foi possível encontrar a distribuição empírica de autovalores das matrizes de correlação e assim caracterizar cada dia através da comparação com a distribuição de Marčenko-Pastur.

A implementação dos algoritmos usados para este trabalho foram feitas em Python 3, utilizando as bibliotecas NumPy² e pandas³, incluindo a sub-biblioteca pandas-datareader⁴ deste último. Algumas bibliotecas auxiliares que valem citar incluem datetime para manipulação de datas e matplotlib para geração de gráficos. Também foi utilizada a linguagem de programação Julia para cálculos que demandaram mais poder computacional.

As séries temporais de preços intradia (preço de fechamento de cada período de 1 minuto) foram adquiridas através da API alphavantage⁵ para os ativos mencionados, entre os dias 19/09/2019 e 08/03/2022.

A partir dos retornos logarítmicos normalizados dos $M = 20$ ativos em um dia, fazemos um sorteio de subconjuntos de tamanho fixo, encontramos a matriz de correlação para cada subconjunto e então construímos a distribuição empírica de autovalores que pode ser comparada com a distribuição de Marčenko-Pastur. Como mencionado na seção 2.1, a distribuição de Marčenko-Pastur busca representar os autovalores de matrizes de correlação de variáveis aleatórias normalmente distribuídas e independentes entre si. Portanto, uma medida de divergência entre a distribuição de autovalores empírica e a distribuição teórica deve oferecer uma medida que abarca tanto a dependência entre os ativos quanto a diferença das distribuições dos retornos da distribuição normal.

Para realização dessa comparação, utilizamos a divergência de Kullback-Leibler, efetuando numericamente a integral de 2.12 com P sendo a distribuição teórica e Q a distribuição empírica de autovalores. Ao analisar a evolução temporal desta divergência, temos uma medida ao longo do tempo que contém ambos os efeitos mencionados acima e que podemos analisar em momentos de interesse.

A fim de analisar isoladamente os efeitos (i) da correlação entre os ativos e (ii) da aproximação ou afastamento dos retornos em relação à distribuição normal, fizemos –

¹ Mais informações em <<https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview>>

² Acesse a página do projeto NumPy em <<https://numpy.org/>>

³ Acesse a página do projeto pandas em <<https://pandas.pydata.org/>>

⁴ Para documentação de pandas-datareader, acesse <<https://pydata.github.io/pandas-datareader/>>

⁵ Acesse a página oficial da API Alpha Vantage em <alphavantage.co/>

paralelamente ao cálculo da divergência de Kullback-Leibler – a média da correlação entre todos os ativos a cada dia e um ajuste com a distribuição de Tukey lambda.

Com a medida de correlação média entre os ativos, podemos identificar períodos de crescimento ou decrescimento da correlação e como estes se relacionam com a divergência dos dados com a distribuição teórica. Já com o ajuste dos retornos à distribuição de Tukey lambda temos no λ um indicador do peso da cauda da distribuição de retornos.

Para podermos compreender os momentos de mercado dentro do intervalo definido, adquirimos a série temporal dos dados de preço de fechamento para um dos ETFs (fundo de índice) que replica o índice S&P 500, o SPY. Não utilizamos um portfólio com os ativos porque (i) construí-lo e rebalanceá-lo não seria trivial e (ii) o SPY é representativo das nossas ações (porque elas compõem o índice) e do mercado dos Estados Unidos como um todo, sendo suficiente para o propósito de ter uma referência (*benchmark*) para verificar momentos de quedas generalizadas do mercado.

3 Resultados

No gráfico presente na figura 3 é possível verificar que a divergência da distribuição de autovalores empírica para a teórica é sempre diferente de zero, tendo como valor mínimo cerca de 0.5. A distribuição teórica de autovalores que estamos comparando, como mencionado na seção 2.1.1, é o que seria esperado para matrizes aleatórias com distribuição normal e com elementos independentes. Isso demonstra que, dentro do nosso conjunto de dados, os retornos logarítmicos de ativos financeiros não são realmente comparáveis a variáveis aleatórias normalmente distribuídas e independentes. Além disso, a variação de intensidade desta divergência ao longo do tempo oferece informações adicionais.

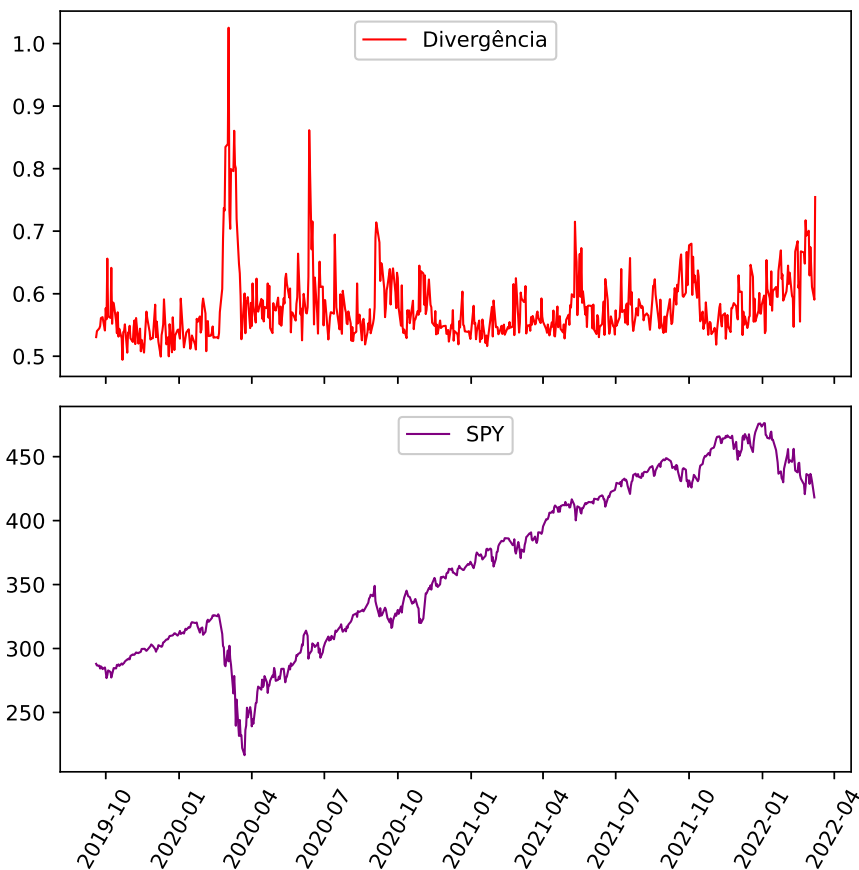


Figura 3 – No topo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. Embaixo, em roxo, temos os preços de fechamento do ETF SPY.

Para o período analisado parece haver uma relação entre as quedas do SPY e aumento da divergência. Para buscarmos uma medida dessa relação, subtraímos a tendência de subida gradual dos preços do SPY, como pode ser observado na figura 4, e calculamos a correlação de Pearson entre a divergência e o SPY com a tendência subtraída,

e encontramos $\rho = -0.40$. Esta correlação não é forte, mas junto com o gráfico é uma evidência de que a divergência está relacionada com as crises e com movimentos do mercado em geral.

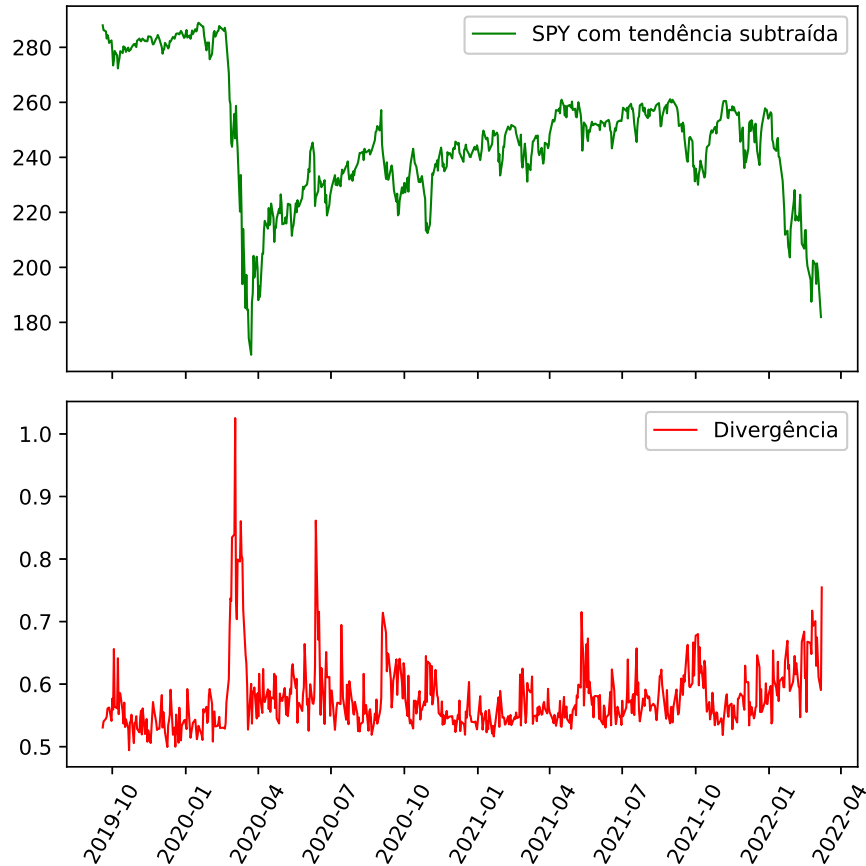


Figura 4 – No topo, em verde, temos os preços de fechamento do ETF SPY, subtraído de sua tendência de queda. Embaixo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. A correlação de Pearson entre essas séries temporais é de $\rho = -0.40$.

No período da crise do coronavírus, é notável o crescimento da divergência simultaneamente com a queda generalizada do mercado (representado pelos preços do SPY). Na figura 5 temos um subconjunto do gráfico da figura 3 que mostra um intervalo de tempo menor contendo (i) a crise da Covid-19 e (ii) um momento de volatilidade em 2020. Este momento de volatilidade foi interpretado pela mídia à época como sendo causado por preocupações decorrentes também da Covid-19¹. Foram colocadas retas verticais em datas próximas a inícios de quedas nos preços do SPY, que coincidem com o início de aumentos da divergência.

Como o aumento da divergência coincide com a queda nos preços, ele não tem nenhum poder preditivo. Entretanto, pode nos ajudar a caracterizar as crises. Como já

¹ Ver www.cnbc.com/2020/09/20/stock-market-futures-open-to-close-news.html, acesso em 17/04/2022.

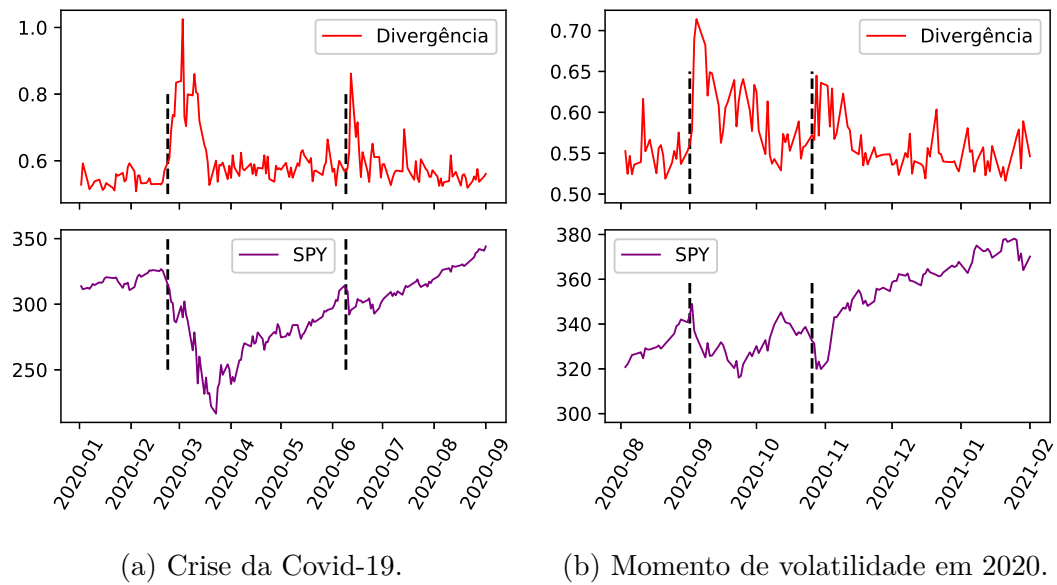


Figura 5 – Em cada figura, no topo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. Embaixo, em roxo, temos os preços de fechamento do SPY. Em preto tracejado temos linhas verticais nas datas de interesse para facilitar a visualização. Na figura 5a temos destaque nas datas 24/02/2020 e 11/06/2020. Na figura 5b temos destaque nas datas 1^o/09/2020 e 26/10/2020

apresentado anteriormente, o aumento da divergência pode nos fornecer pistas de que há um aumento da correlação entre os ativos ou de que os retornos dos ativos estão com muitos valores extremos. Para tentar isolar esses efeitos, fizemos separadamente outras duas análises.

Na figura 6 apresentamos a série temporal da divergência de Kullback-Leibler e a série temporal do λ da distribuição de Tukey lambda que melhor ajustou os dados para cada dia. Contraintuitivamente, nos períodos de aumento da divergência tivemos uma diminuição dos pesos das caudas (representado pelo aumento do λ). Isto evidencia que a principal contribuição para divergência não é o peso das caudas da distribuição de retornos. Devemos então investigar o outro fator que sabemos que pode causar uma divergência da teoria: a correlação entre os ativos.

Na figura 7, apresentamos a série temporal da divergência e a série temporal da correlação média para cada dia. É interessante notar que momentos de rápido crescimento da correlação coincidem com aumentos substanciais da divergência. As quedas na divergência, porém, não são acompanhadas por quedas abruptas da correlação; pelo contrário, a correlação possui um período de relaxação prolongado após períodos de queda do mercado. Isso parece indicar que o crescimento da correlação média é um dos fatores que está contido no aumento da divergência. Por outro lado, talvez a estrutura de correlações também influencie na divergência já que a divergência retorna ao patamar anterior à crise antes da correlação, não apresentando período de relaxação.

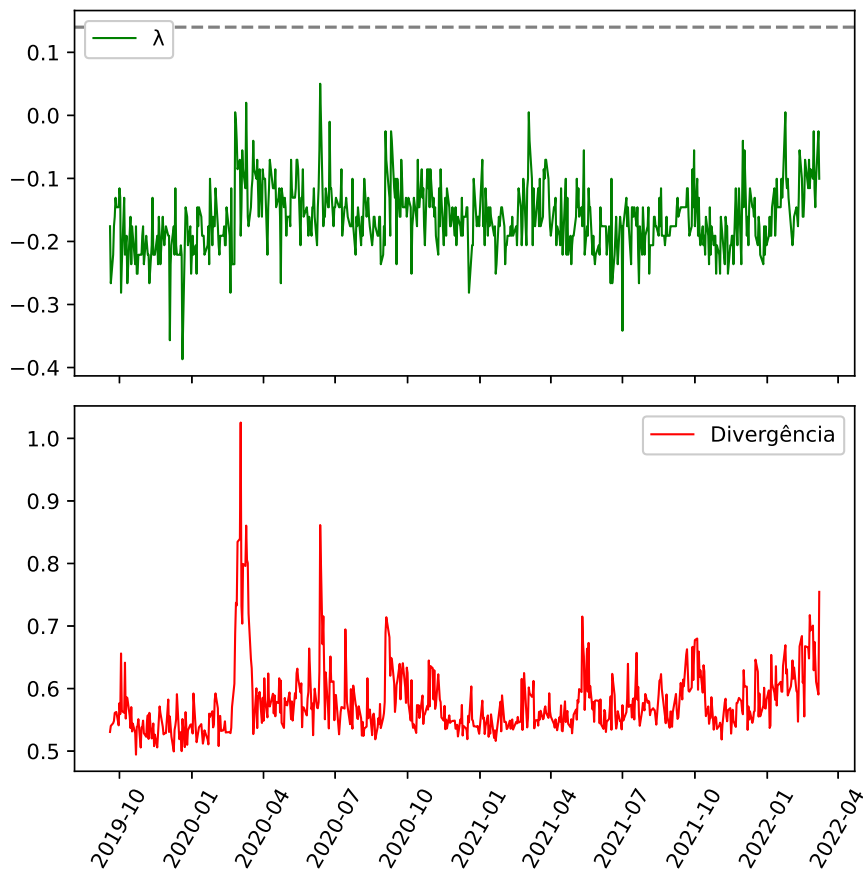


Figura 6 – No topo, em verde, temos o melhor λ da distribuição de Tukey lambda ajustado aos retornos dos ativos, com uma linha tracejada em cinza indicando $\lambda = 0.14$, situação em que a distribuição é aproximadamente normal. Embaixo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. A correlação entre essas duas séries temporais é 0.39.

É importante esclarecer que a distribuição de Marčenko-Pastur empírica oferece informações sobre o perfil de correlações de um conjunto de dados. A queda da divergência logo após as crises pode indicar que a correlação remanescente é menos generalizada, por exemplo. Este é um ponto que pode ser explorado em análises futuras.

Fica clara com esta análise a estreita relação entre a correlação e a distribuição de Marčenko-Pastur empírica. Também é evidenciado que a divergência de Kullback-Leibler é bastante sensível às crises, sem a persistência temporal observada na correlação média.

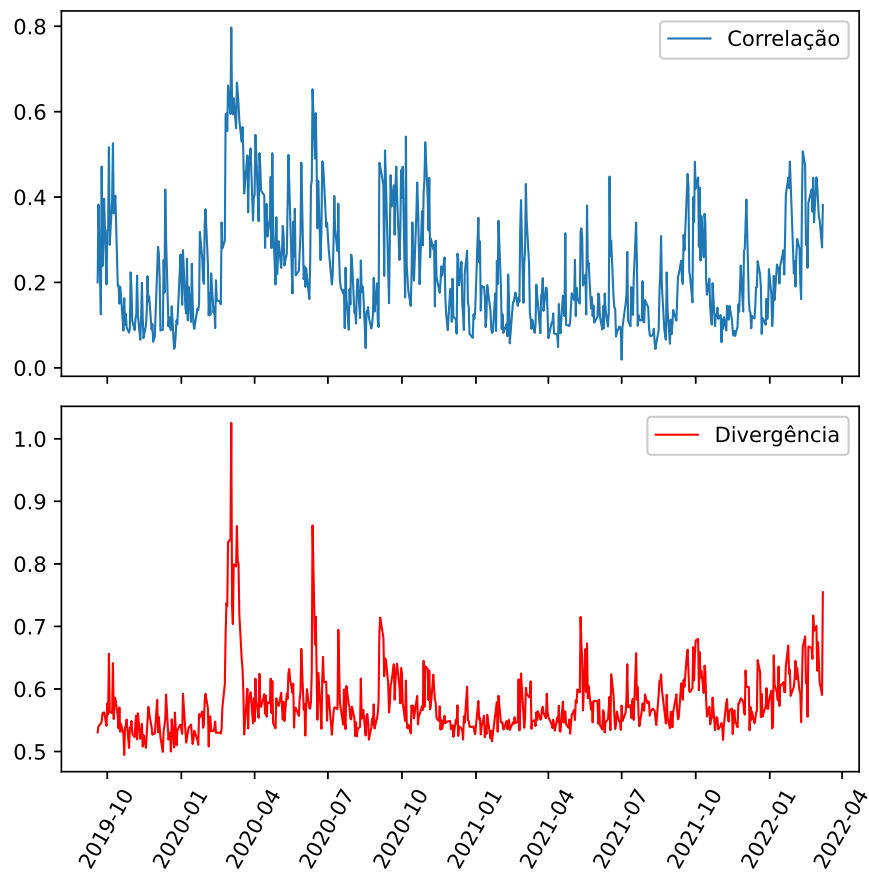


Figura 7 – No topo, em azul, temos a série temporal da correlação média entre os ativos. Embaixo, em vermelho, temos a divergência de Kullback-Leibler da distribuição de autovalores empírica para a distribuição de autovalores teórica. A correlação entre essas duas séries temporais é 0.73.

4 Conclusões

A série temporal da divergência da distribuição de autovalores da matriz de correlação dos ativos para a distribuição teórica demonstrou que os retornos de ativos financeiros não são como variáveis aleatórias independentes entre si. Ao longo de crises a divergência é ainda maior, reagindo fortemente ao aumento da correlação entre os ativos. Dessa forma, a divergência é um indicador bastante sensível à correlação e pode oferecer uma compreensão clara dela.

A divergência de Kullback-Leibler é mais especial, no contexto deste trabalho, que a correlação média. Isso se deve ao fato dela representar uma distância estatística (mas não métrica) entre um conjunto de dados e a previsão teórico, dentro da estrutura da teoria de matrizes aleatórias. Nossos resultados apontam a relevância desta teoria e abrem caminho para se investir na exploração dos autovalores e autovetores empíricos e sua comparação com os teóricos.

Além disso, esta divergência pode, como proposta no nosso trabalho, servir como um indicador para compor com um conjunto de outros indicadores para ajudar tomadores de decisão.

Com a análise feita, podemos concluir que a teoria de matrizes aleatórias oferece uma estrutura promissora para análise de correlações, sendo potencialmente útil para integrar produtos de análise. Tiramos também algumas conclusões mais específicas:

1. O peso das caudas das distribuições de retorno *intraday* não contribui de maneira notável na divergência.
2. A correlação parece ser o fator principal que afeta a divergência.
3. A divergência não possui comportamento idêntico à correlação média, devendo conter também informações sobre o perfil de correlações.
4. A divergência pode servir como indicador dos momentos de queda do mercado, tendo em vista que ela não persiste em patamares altos quando os preços param de cair.

4.1 Trabalhos futuros

A teoria de matrizes aleatórias demonstrou e demonstra ser frutífera. Apesar de termos explorado uma abordagem específica aqui, ela pode oferecer ainda mais ferramentas e opções de análises para se agregar em um produto. Uma possibilidade é construir um filtro que remove os autovetores referentes aos autovalores que caem fora do previsto pela

teoria a fim de reconstruir a matriz de correlação para efetuar otimização de portfólios (17).

Outra análise interessante que pode ser feita no futuro é o estudo da evolução temporal dos maiores autovalores e autovetores para a compreensão da estabilidade (ou não) de *clusters* de correlação entre ativos (18). A partir deste estudo pode ser possível verificar como choques no mercado influenciam ativos individualmente ou como blocos.

Podemos também explorar a influência do perfil de correlações na divergência da distribuição de autovalores empírica para a teórica. Isso poderia oferecer ainda mais clareza sobre o tipo de informação que a divergência oferece.

Neste contexto, este trabalho explorou a teoria de matrizes aleatórias, mostrando como ela pode ser uma ferramenta poderosa para análises de mercado, devendo ter no futuro um espaço cada vez maior em produtos de análise financeira.

Referências

- 1 CHAKRABORTI, A. et al. Econophysics review: I. empirical facts. *Quantitative Finance*, Routledge, v. 11, n. 7, p. 991–1012, 2011. Disponível em: <<https://doi.org/10.1080/14697688.2010.539248>>. Citado na página 13.
- 2 MANTEGNA, H. E. S. R. N. *An introduction to econophysics: correlations and complexity in finance*. 0. ed. Cambridge University Press, 2000. ISBN 9780521620086,0521620082. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=dfe41b3351c63731ab5dc9f3e7520666>>. Citado na página 13.
- 3 FALLON, P. R.; LUCAS, R. E. B. The Impact of Financial Crises on Labor Markets, Household Incomes, and Poverty: A Review of Evidence. *The World Bank Research Observer*, v. 17, n. 1, p. 21–45, 05 2002. ISSN 0257-3032. Disponível em: <<https://doi.org/10.1093/wbro/17.1.21>>. Citado na página 14.
- 4 CONT, R. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, v. 1, n. 2, p. 223–236, 2001. Citado na página 14.
- 5 MANDELBROT, B. The variation of certain speculative prices. *The Journal of Business*, v. 36, 1963. Disponível em: <<https://EconPapers.repec.org/RePEc:ucp:jnlbus:v:36:y:1963:p:394>>. Citado na página 14.
- 6 ANG, A.; CHEN, J. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, v. 63, n. 3, p. 443–494, 2002. ISSN 0304-405X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304405X02000685>>. Citado na página 14.
- 7 WIGNER, E. P. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, Annals of Mathematics, v. 62, n. 3, p. 548–564, 1955. ISSN 0003486X. Disponível em: <<http://www.jstor.org/stable/1970079>>. Citado na página 14.
- 8 WIGNER, E. P. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, Annals of Mathematics, v. 67, n. 2, p. 325–327, 1958. ISSN 0003486X. Disponível em: <<http://www.jstor.org/stable/1970008>>. Citado na página 14.
- 9 MARČENKO, V. A.; PASTUR, L. A. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, IOP Publishing, v. 1, n. 4, p. 457–483, apr 1967. Disponível em: <<https://doi.org/10.1070/sm1967v001n04abeh001994>>. Citado 2 vezes nas páginas 14 e 17.
- 10 SANDOVAL, L.; FRANCA, I. D. P. Correlation of financial markets in times of crisis. *Physica A: Statistical Mechanics and its Applications*, Elsevier BV, v. 391, n. 1-2, p. 187–208, jan 2012. Disponível em: <<https://doi.org/10.1016%2Fj.physa.2011.07.023>>. Citado na página 14.
- 11 PLEROU, V. et al. A random matrix approach to cross-correlations in financial data. *Physical review. E, Statistical, nonlinear, and soft matter physics*, v. 65, p. 066126, 07 2002. Citado 2 vezes nas páginas 14 e 18.

- 12 PLEUROU, V.; GOPIKRISHNAN, P.; STANLEY, H. E. Two-phase behaviour of financial markets. *Nature*, v. 421, n. 6919, p. 130–130, Jan 2003. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/421130a>>. Citado na página 14.
- 13 Kulkarni, V.; Deo, N. Correlation and volatility in an Indian stock market: A random matrix approach. *European Physical Journal B*, v. 60, n. 1, p. 101–109, nov. 2007. Citado na página 14.
- 14 CUNHA, C. R. da; SILVA, R. da. *On the quantum behavior and clustering properties of correlated financial portfolios*. 2020. Citado na página 18.
- 15 KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *The Annals of Mathematical Statistics*, Institute of Mathematical Statistics, v. 22, n. 1, p. 79–86, 1951. ISSN 00034851. Disponível em: <<http://www.jstor.org/stable/2236703>>. Citado na página 20.
- 16 KARIAN, E. J. D. Z. A. *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*. [S.l.]: CRC Press LLC, 2000. Citado na página 21.
- 17 LALOUX, L. et al. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, v. 3, p. 391, 07 2000. Citado na página 32.
- 18 ETEROVIC, N. A. *A Random Matrix Approach to Portfolio Management and Financial Networks*. Tese (Doutorado) — University of Essex, 2016. Citado na página 32.