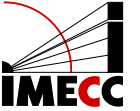




UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA  
DEPARTAMENTO DE MATEMÁTICA APLICADA



Augusto Plastino Duarte

## Um estudo empírico acerca da situação econômica brasileira atual através do uso da Teoria de Matrizes Aleatórias.

Monografia apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos para obtenção de créditos na disciplina Projeto Supervisionado, sob a orientação do(a) Prof. José Mario Martínez.

Campinas  
2019

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
1.1	Premissa . . . . .	2
1.2	Indicadores . . . . .	3
1.3	Orientações . . . . .	3
<b>2</b>	<b>Dados Utilizados</b>	<b>4</b>
<b>3</b>	<b>Indicadores</b>	<b>4</b>
3.1	Tipo <i>A</i> . . . . .	5
3.1.1	Teoria Matemática . . . . .	5
3.1.2	Interpretações . . . . .	7
3.2	Tipo <i>B</i> . . . . .	8
3.2.1	Interpretações . . . . .	8
<b>4</b>	<b>Resultados obtidos</b>	<b>9</b>
4.1	Tipo <i>A</i> . . . . .	9
4.2	Tipo <i>B</i> . . . . .	10
<b>5</b>	<b>Conclusões</b>	<b>12</b>

## Resumo

Apresentamos uma aplicação atual de indicadores de crises financeiras baseadas nas propriedades espectrais de dados recentes do mercado brasileiro. Escolhemos dados diários de fechamento de diversos índices de bolsa – de 2011 à 2019 – proporcionados pela BOVESPA, e construímos matrizes de covariância e correlação para cada dia, considerando os 150 dias anteriores. A primeira classe de indicadores é construída com base nas distâncias de Hellinger [1], que compara a distribuição empírica dos autovalores da matriz de covariância com matrizes referências, advindas de conjuntos conhecidos (*ensembles*) de matrizes aleatórias. A segunda classe de indicadores provém da análise espectral das matrizes construídas, relacionando certos valores puramente algébricos com sinais de correlação e volatilidade de mercado. Provedmos argumentos e referências nos quais embasamos o poder preditivo destes indicadores.

## 1 Introdução

### 1.1 Premissa

Temos como intuito inicial apresentar indicadores de crises financeiras futuras brasileiras, aplicando métodos descritos em [6] para o mercado brasileiro, no período de 2011 a 2019. Nos baseamos fortemente nesta referência, e repetiremos aqui a apresentação dos conceitos por nós utilizados.

Enfatizamos que os indicadores que reproduziremos não possuem a premissa de predizer ou profetizar crises econômicas com sequer um mínimo de certeza, e sim, afirmar (qualitativamente) se a probabilidade de uma crise econômica ocorrer está aumentando (ou não). Arrazoando uma analogia presente em [6], tais indicadores não intuem prever quando uma faísca vai reagir com o gás presente em um quarto e iniciar um incêndio, e sim medir se a quantidade de gás no quarto é adequada para que o incêndio possa acontecer. Faíscas como estas – eventos aleatórios, adversos e variados –, podem desencadear uma reação em cadeia, afetando negativamente o mercado como um todo.

A explosão de uma bolha econômica, a falência de uma ou múltiplas empresas importantes, grandes acontecimentos políticos – sejam na micropolítica do país ou da macropolítica do continente, dos parceiros econômicos do país, ou mesmo do planeta –, enfim, são inúmeros os acontecimentos que podem ser considerados como as faíscas da analogia supracitada. Ao medirmos a situação do mercado sobre como este deve se comportar diante destes acontecimentos, julgamos que isso é estatisticamente equivalente a prever uma crise econômica.

## 1.2 Indicadores

Acerca dos indicadores, estes se baseiam – dado o *dataset* adequado – em mensurar quantitativamente o nível de correlação do mercado – ou seja, de como (e quanto) quaisquer produtos de mercado (ativos, ações, etc.) estão correlacionados entre si – e de volatilidade (variância) de mercado. Extendemos esta análise, de ativos independentes como descrito em alguns *datasets* em [6], para índices de mercado, de tal forma que aqui trabalharemos puramente com índices da bolsa brasileira advindos da BOVESPA.

Propomos a apresentação de métodos modernos (estreados em [6]) construídos parcialmente com auxílio da teoria das matrizes aleatórias [4]. Note que esta não é a única forma de obter índices qualitativos e/ou quantitativos de correlação e volatilidade de mercado; índices como o *VIX*, que rastreia a volatilidade do *S&P500*, e diversos métodos estatísticos (e.g. coeficiente de Pearson) que calculam a correlação entre os retornos de 2 ou mais ativos financeiros são exemplos que firmam seus alicerces em outros campos matemáticos.

O campo teórico de matrizes aleatórias tem sido recentemente utilizado como um conjunto ferramental moderno na literatura econômetrica – como, por exemplo, provar a presença fontes de erro nas estimativas de autovetores de matrizes de covariância, e prover métodos que melhoram as otimizações de tais portfólios [7] –, e por isso cremos em sua importância e valor agregado em nossa aplicação.

## 1.3 Orientações

A volatilidade é, a princípio, definida pela dispersão (i.e. variância) dos retornos de um título financeiro, e usualmente implica diretamente no nível de risco de investimentos envolvendo este título. No decorrer deste artigo, discutiremos estas afirmações, além do relacionamento entre volatilidade e instabilidade no mercado financeiro. Vemos também a importância do relacionamento entre a volatilidade geral de mercado e o índice de correlação deste.

Argumentamos heurísticamente que esses indicadores, além de serem úteis para a predição de volatilidade de mercado, também podem nos dizer – recorrendo à analogia supracitada – o tamanho da explosão caso esta ocorra, visto que os indicadores, intuem, na analogia, a quantidade de gás, e na realidade, o quanto os diferentes submercados e setores "dependem" (em um sentido de correlação) uns dos outros.

Holisticamente, a alta volatilidade do mercado no geral usualmente é associada com períodos de crise econômica, ainda que hajam certos proponentes que argumentam a relação inversa. Discutiremos isto mais a fundo na conclusão, sob outras análises técnicas acadêmicas e sob os valores de nossos indicadores coletados. A princípio, esperamos que estes possam nos indicar empiricamente períodos de crises passadas do mercado brasileiro, e talvez indicar a direção a qual nosso mercado se encontra agora.

Há uma ampla literatura acerca da predição de crises financeiras que podemos recomendar como leitura complementar, e uma parte considerável destas na qual nos baseamos pode ser encontradas nas referências de [6]. O código por nós escrito e utilizado foi feito no *software* **GNU Octave**.

## 2 Dados Utilizados

Escolhemos dados provenientes da BOVESPA, a bolsa de valores brasileira; mais precisamente, trabalhamos com 7 índices públicos de setores, os quais se relacionam com distintos setores da economia brasileira: IMOB (imobiliário), ICON (consumo), IEEX (energia elétrica), IFNC (financeiro), INDX (industrial), IMAT (materiais básicos) e, por fim, IBOV, o índice econômico geral. Nossos dados são os valores de fechamento diários, de 01/06/2011 até 31/10/2019; um período maior nos seria mais interessante, mas a bolsa brasileira é muito jovem, e muitos destes índices foram criados recentemente, e portanto escolhemos a data de início na qual conseguiríamos trabalhar com todos estes.

Esperamos que os relacionamentos entre as séries temporais destes índices sejam minimamente desvendados pelos métodos propostos; esperamos um índice de correlação maior do que o usual, em média (o que não nos deve atrapalhar pois nos preocupamos com as taxas de acréscimo e decréscimo de nossos indicadores, e não com seus valores absolutos), visto que estamos incluindo um índice genérico (IBOV) em nossos dados.

Além disto, o fato de estarmos trabalhando com índices faz com que as taxas de correlação de mercado tenham mais ruído do que desejaríamos, visto os efeitos mediadores dos indicadores [6]; não julgamos que isto nos atrapalha em demasia, pois consideramos que este ruído é controlado, já que estamos trabalhando com um mercado pequeno (comparativamente) como o brasileiro, com índices menores e menos complexos (novamente, comparativamente). Admitidamente, esta suposição é completamente heurística, e mais pesquisas acerca de usos de ativos disjuntos brasileiros (vide ações referentes a símbolos pré-determinados) como *dataset* poderiam ser feitos e concluírem acerca desta suposição com maior rigor.

Os métodos a seguir foram aplicados nos *log-retornos* dos dados, por estes nos proporcionarem a informação da taxa de variação de nossos dados com maior pureza, desconsiderando escalas e valores absolutos.

## 3 Indicadores

No total, temos 6 indicadores que construímos e utilizamos, divididos em 2 classes, as quais nos referiremos como Tipo *A* e Tipo *B*. Ao falarmos, por exemplo, do segundo indicador do tipo *B*, nos referiremos a este como *B2*.

No total, temos 2085 dias em nosso *dataset*, ou seja, 2084 (1 a menos devido à transformação de retornos diários para *log-retornos* diários, o que exclui o primeiro dia) valores para as séries temporais de cada um dos indicadores utilizados.

A partir de uma determinada data de início (dia 21 para os indicadores do Tipo *A* e dia 1 para os indicadores do tipo *B* – mais detalhes sobre o rationale destas escolhas mais à frente), utilizamos os 90 dias seguintes (contando o inicial) para construímos um valor referente à metodologia do indicador em questão. Após isso, pulamos 15 dias, caracterizando o ponto seguinte, e repetimos a mesma metodologia para tal; fazemos este processo ponto a ponto até alcançarmos a data final – na qual, evidentemente, trabalharemos com os dias restantes que sejam possíveis, e desta forma, temos que o último ponto utiliza 84 dias, ao invés de 90.

Cremos que o uso de 90 dias para a construção de cada valor individual, ainda que menor que os 150 originalmente propostos em [6], seja suficiente para termos um efeito mediador que exclui ruídos diários e semanais, que poderiam arruinar nossas conclusões acerca da responsividade dos indicadores.

Os indicadores são construídos sobre uma das seguintes matrizes descritas a seguir: sendo  $T = 90$  (os dias em questão),  $N = 7$  (os *assets* em questão), temos uma matriz  $R(T, N)$ , sendo que o elemento  $R_{i,j}$  é o *log-retorno* do indicador  $j$  no dia  $i$  da janela em questão. A partir desta matriz, derivamos as seguintes matrizes:

- $R^*(T, N)$ , tal que  $R_{:,i}^* = R_{:,i} - \text{mean}(R_{:,i})$ , matriz com colunas normalizadas em torno de 0.

- $CV(N, N) = \frac{R^*(T, N) \times (R^*(T, N))^T}{T-1}$ , matriz de covariância.
- $CR(N, N)$ , tal que  $CR_{i,j} = \frac{CV_{i,j}}{\text{std}(CV_{i,:}) \cdot \text{std}(CV_{:,j})}$ , matriz de correlações.

Quando trabalhamos com matrizes de covariância, escalamos os autovalores destas, multiplicando-os pela inversa da média das variâncias de cada uma das colunas do *dataset*; em nosso caso, este fator de re-escala é de 4888.7. Tal multiplicação só será mais importante nos indicadores de Tipo *B*, visto que os indicadores de tipo *A* se baseiam na distribuição dos autovalores, e só será influenciada pelo fator de re-escala por conta da consideração do suporte máximo, que discutiremos à frente.

### 3.1 Tipo A

Os indicadores do tipo *A* se embasam na teoria de matrizes aleatórias – mais especificamente, nas distribuições dos autovalores de certas matrizes apropriadamente construídas, pertencentes a conjuntos (descritos pela especificação da construção da matriz) que nos referiremos como *ensembles*.

A teoria das matrizes aleatórias nos descreve, além de inúmeros outros resultados, como a distribuição de matrizes de *ensembles* clássicos e específicos se comporta quando as dimensões das matrizes tendem ao infinito [4]. Ainda que a teoria técnica por trás destes resultados seja vasta e altamente interessante, não nos preocuparemos muito com ela nestes experimentos, não por uma falta de rigor, mas pela ausência da necessidade, visto que podemos verificar empiricamente que os *ensembles* por nós utilizados (descritos inicialmente, para uso análogo, em [6]) proporcionam – a princípio – distribuições que julgamos interessantes e úteis em nossa aplicação, sem a obrigação do rigor matemático que explica os PORQUÊS e COMOS de essas distribuições ocorrerem naturalmente. É claro (e ficará ainda mais claro na análise de resultados) que um embasamento teórico mais profundo agregaria, em especial nas escolhas dos *ensembles*, mas isto não é cabível no escopo do estudo.

#### 3.1.1 Teoria Matemática

Para cada um dos indicadores deste tipo, computamos a distância de Hellinger [1] entre a distribuição empírica dos autovalores da matriz de covariância (própriamente re-escalada, como supracitado) do *dataset*, e a distribuição dos autovalores do *ensemble* utilizado. A distância de Hellinger  $\mathbb{D}$ , em suma, determina um valor entre 0 e 1 de similaridade entre duas distribuições de probabilidade, e é definido (em sua forma discreta) por:

$$\mathbb{D}^2(X, Y) = \sum_{i=1}^K (\sqrt{P(X_i)} - \sqrt{P(Y_i)})^2,$$

sendo  $X$  e  $Y$  as distribuições em questão e  $K$  o número de pontos utilizados na comparação discreta; não utilizamos a forma integral da distância de Hellinger (que compara as distribuições contínuas) por não termos a distribuição dos autovalores de  $CV$  e de certos *ensembles* em uma forma contínua.

Para todos os fins, calcularemos os histogramas probabilisticamente normalizados dos autovalores de  $CV$  e de uma matriz do *ensemble* em questão, cada um com 40 secções; por fim, devemos definir também o suporte de nosso cálculo, ou seja, o valor máximo de  $X$  no qual ainda calculamos as probabilidades para as duas distribuições. Discutiremos o valor escolhido para este suporte mais adiante.

Descobrimos que as dimensões ( $N_m = 77$ ,  $T_m = 90$ ) que utilizamos para as matrizes dos *ensembles* testados nos foram suficientemente grandes para que as distribuições de seus autovalores fossem consistentes (para nossos fins experimentais) a cada execução distinta do algoritmo.

Logicamente, quanto menor for a distância de Hellinger de duas distribuições, maior sua similaridade. A seguir, descreveremos os *ensembles* que utilizaremos para cada um dos 3 respectivos indicadores deste tipo.

- *C1*: *ensemble* cuja distribuição de autovalores é descrita pela Lei de Marchenko-Pastur [8]; tome  $X(N, T)$  cujos elementos sejam todos *i.i.d* (independentes e identicamente distribuídos), de média nula e variância finita, e  $Y = XX^T/T$ , a distribuição  $f(x)$  dos autovalores de  $Y$ , para  $N/T \rightarrow \sigma \in (0, 1)$ , conforme  $N, T \rightarrow \infty$ , é dada por:

$$f(x) = \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{2\pi\sigma x} I(x \in [\lambda_-, \lambda_+]),$$

sendo  $\lambda_- = (1 - \sqrt{\sigma})^2$  e  $\lambda_+ = (1 + \sqrt{\sigma})^2$ ; para nosso experimento, tomamos os elementos de  $X$  como  $\mathcal{N}(0, 1)$  *i.i.d*.

O valor de  $\lambda_+$  também será utilizado nos outros índices.

- *C2*: *ensemble* de matrizes  $Z(N, T)$  tais que seus elementos são todos  $\mathcal{N}(0, 1)$  *i.i.d*, mas suas colunas possuem um certo grau de correlação entre si, através da introdução de uma variável aleatória  $\mathcal{N}(0, 1)$  para os valores daquela coluna; consideramos um coeficiente  $\rho \in (0, 1)$  de multiplicação nesta introdução; o cálculo exato se dá pela multiplicação de cada uma das linhas de  $X_{i,:}$  por  $\sqrt{(1 - \rho^2)}$ , seguido pela soma desta linha com um vetor  $Z_0(1, T)$  cujos elementos são  $\mathcal{N}(0, 1)$  *i.i.d*, multiplicado por  $\rho$ .

Em outras palavras, tendo  $X(N, T)$  cujos elementos são todos  $\mathcal{N}(0, 1)$  *i.i.d*:

$$\forall i \in (1, N), Z_{i,:} = \sqrt{1 - \rho^2} X_{i,:} + \rho Z_0.$$

Tomamos  $\rho = 0.5$ , nos inspirando pela metodologia descrita em [6], no qual se calcula os coeficientes de correlação entre os *assets* do conjunto de dados.

- *C3*: *ensemble* com metodologia de construção exatamente igual a *C2*, mas as distribuições normais são substituídas pela distribuição de Student [5] com  $t = 3$ .

Construímos então os indicadores  $A1$ ,  $A2$  e  $A3$ , cada um representado em uma série temporal de distâncias de Hellinger entre a distribuição empírica  $\epsilon_n(x)$  dos autovalores da matriz de covariância  $CV$ , re-escalada, dos *log-retornos* de nosso *dataset* dentro da janela supracitada de 90 dias, com pulos de 15 dias, e a distribuição empírica  $\theta_n$ , dos autovalores de matrizes pertencentes aos respectivos *ensembles* *C1*, *C2* e *C3*.

Ademais, visto que, ao computarmos os autovalores de  $CV$  em uma janela, teríamos apenas 7 valores - que claramente não é capaz de descrever bem um histograma como desejado, para cada uma das janelas, calculamos também os autovalores da matriz de covariância, também re-escalada, dos *log-retornos* do *dataset* centrado em cada um dos 10 dias anteriores, com a janela de 90 dias apropriadamente transladada.

Sinonimamente, temos  $11 \cdot 7 = 77$  autovalores para descrever nossa distribuição empírica  $\epsilon_n(x)$  pelo histograma na janela inicialmente descrita.

O suporte da computação da distância de Hellinger entre as distribuições nos foi fixado como  $25\lambda_+$ , visto que - nos inspirando em [6] - estudos empíricos indicam que os autovalores de matrizes de covariância similares aos por nós utilizados podem alcançar até 25 vezes o suporte máximo da distribuição de Marchenko-Pastur correspondente.

Finalmente, utilizaremos uma técnica também descrita em [6] que nos permite ignorar autovalores pequenos - próximos de zero - no cálculo da distância de Hellinger, fixando uma base  $\beta = \frac{\lambda_+}{10}$ , tal que para valores de  $x < \beta$ , modificamos a distribuição advinda do dataset:

$$\epsilon_n^*(x) = \begin{cases} \min(\epsilon_n(x), \theta_n(x)) & \text{se } x \leq \beta \\ \epsilon_n(x) & \text{senão} \end{cases}$$

A lógica por trás desta técnica é que buscamos caracterizar, nestes indicadores, quando e quanto os autovalores de  $CV$  estão crescendo como um conjunto, em especial os maiores; os autovalores pequenos não nos interessam, por representarem ruído aleatório no mercado econômico, em termos de correlação e volatilidade; além disso, cremos que as probabilidades destes autovalores pequenos – que usualmente são acumuladas em nossos dados e *ensembles* descritos –, nestas distribuições (especialmente em  $\epsilon$ ), variam o suficiente para as considerarmos como ruidosas, plausivelmente perturbando o cálculo de nossos indicadores de uma maneira indesejada.

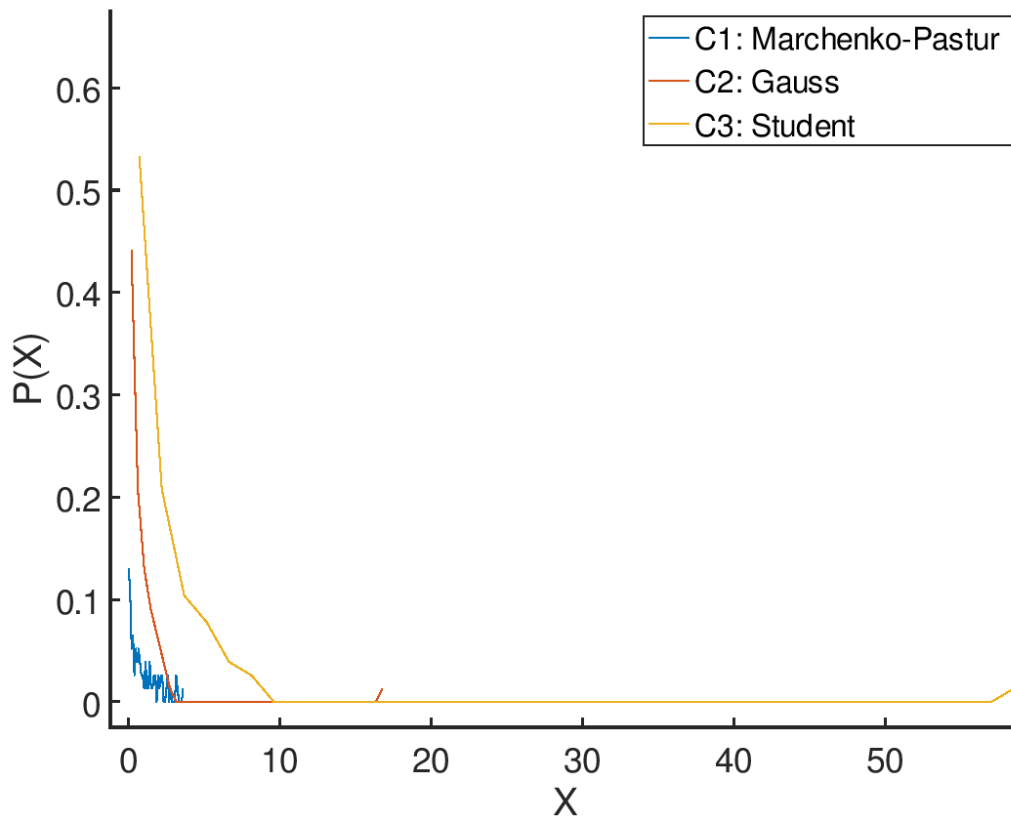
Resumidamente,  $An = \mathbb{D}(\epsilon_n^*, \theta_n)$ , e aglutinamos, para cada um destes 3 indicadores, seus valores para diferentes janelas de tempo em uma série temporal.

### 3.1.2 Interpretações

Montamos os *ensembles* acima de tal forma que  $A1$  e  $A2$  exprimem mercados calmos e controlados, enquanto  $A3$  se relaciona com um mercado turbulento, indicando a possibilidade de uma crise. Os *ensembles*  $C1$  e  $C2$  são caracterizados por distribuições de autovalores branda, sem raio espectral excessivo; já em  $A3$ , certos autovalores muito grandes estatisticamente aparecem com maior frequência.

Basicamente, como lembrete dos indicadores do tipo  $A$ , estamos comparando algébricamente o comportamento das variâncias internas de mercado com o comportamento das variâncias internas de matrizes pertencentes aos *ensembles* descritos.

Em [6], caracteriza-se  $A2$  como melhor e mais preciso que  $A1$  para descrever um mercado calmo e tranquilo, visto que mesmo mercados completamente calmos possuem consideráveis variâncias internas naturais i.e. que não indicam turbulência. Testaremos esta hipótese em nosso estudo, ainda que a teoria tenha sentido, heurísticamente.



(1) Distribuições empíricas dos autovalores dos *ensembles* descritos

Em suma, conseguimos tirar conclusões melhores caso combinarmos os indicadores em uma análise, por exemplo: caso  $A1$  e  $A2$  subam, isto quer dizer que o comportamento do mercado está se distanciando (descrito por  $\mathbb{D}$ ) de uma estrutura calma; se  $A3$  diminuir, concluímos que o mercado pode estar acumulando um caráter agitado e violento. Caso estes acontecimentos sejam coincidentes, a conclusão ressonante de que o mercado está mais próximo de uma crise é estatisticamente maior.

É importante esclarecer que esperamos que falsos positivos ocorram nestas análises; reiteramos que os nossos indicadores não possuem a pretensão de prever crises econômicas, mas sinalizar condições de mercado que favorecem o acontecimento de uma crise – o que não implica que a crise necessariamente vai ocorrer em um tempo hábil.

## 3.2 Tipo $B$

Os indicadores de tipo  $B$  consistem em 3 simples indicadores que calculamos sob o espectro das matrizes de covariância e de correlação de nosso *dataset*, para uma janela fixa.

- $B1$ : o raio espectral (maior autovalor) da matriz de covariância,  $CV$ .
- $B2$ : o traço da matriz de  $CV$ .
- $B3$ : o raio espectral da matriz de correlação,  $CR$ .

### 3.2.1 Interpretações

A análise dos indicadores de tipo  $B$  se baseia nos componentes correlação e volatilidade. Como descrito anteriormente, a correlação nos diz quanto os diferentes índices, em nosso *dataset*, se relacionam temporalmente, e a volatilidade nos diz, em poucas palavras, a variância de mercado. Há um relacionamento entre estes componentes, e também uma relação entre o comportamento destes componentes e a instabilidade de mercado; ressaltaremos ambos em mais detalhes na conclusão.

Devemos deixar claro que os autovalores determinam a “variância” dos pontos na direção dos autovetores no espaço vetorial. Desta forma, o indicador  $B1$  expressa a volatilidade de mercado, mas como ele está relacionado ao autovetor cujo, por definição, tem a maior variância, e denotando os autovetores como portfólios (neste caso, de nossos índices), o indicador  $B1$  também expressa, indiretamente, a correlação de mercado.

O indicador  $B2$  claramente expressa a “energia” total do sistema –, ou seja, a volatilidade integral do mercado.

Como os autovalores da matriz de correlação são não-correlacionados entre si (diferentemente da matriz de covariância, cujos autovalores são “não-covariados” entre si, mas não necessariamente não-correlacionados), o maior autovalor da matriz de correlação mede um sinal de correlação de mercado apenas.

Visto isso, podemos tirar certas conclusões sob análise da série temporal dos indicadores, por exemplo: quando os indicadores  $B1$  e  $B2$  se aproximam, isto profere que as correlações do mercado estão crescendo, pois um autovetor (o associado ao maior autovalor) está se tornando dominante; caso isto aconteça e  $B3$  não cresça de uma maneira clara e indicativa, isto pode significar que os índices de maior volatilidade estão perdendo correlação, e desta forma os índices de menor volatilidade estão se correlacionando (visto que a correlação total aumenta). Insistimos novamente que, assim como os indicadores de tipo  $A$ , os de tipo  $B$  também podem apresentar falsos positivos na análise de dados.

Poderíamos possivelmente extrair um quarto indicador analisando os cossenos entre os autovetores da matriz de covariância; se dois autovetores apontassem em direções parecidas, isto atestaria que eles estão variando de maneira muito similar, sendo assim altamente covariantes, o que indica alta correlação entre diferentes portfólios; isto significaria que diferentes estratégias levariam, algebricamente, a resultados



similares. Acerca dos usos deste indicador teórico, seja para análise de risco de mercado ou para outras aplicações, não exploraremos aqui.

## 4 Resultados obtidos

A seguir apresentamos resultados de nossa análise sobre as séries temporais dos índices que construímos. Achamos de suma importância destacar os seguintes adendos, antes de prosseguirmos:

1. As datas apresentadas são facilmente calculadas pela posição do ponto no eixo  $X$ , em relação ao número total de pontos daquela série temporal; tentamos apresentar, no máximo, o intervalo do mês no qual estamos observando i.e. não julgamos que os dias exatos agregam em nossas conclusões.
2. Nossa análise não é (e não pode ser, visto a inexequibilidade da conciliação dos campos político, econômico e estatístico, neste curto espaço) completamente técnica, e está sujeita à subjetividades (com, obviamente, as devidas *tentativas* de remoção de qualquer viés) dos autores. A discordância saudável do leitor é plausivelmente esperada e muito bem-vinda.
3. Repetimos novamente que os indicadores por nós construídos não intencionam (por construção) alcançar a perfeição de revisão empírica nem de predição, e com certeza estão sujeitos à classificação errônea, com uma taxa não-negligível de falsos positivos e falsos negativos.
4. Nossa análise pode aparentar estar passível da crítica de que estamos “*cherry-picking*”<sup>\*</sup> momentos do passado político-econômico brasileiro e/ou das séries temporais em prol de defender a metodologia e/ou observações políticas; no entanto – ainda que não pretendemos alegar que tomamos critérios PURAMENTE objetivos e baseadas APENAS em ponderações puramente científicas – argumentamos que o método é suficientemente sólido como descrito e defendido com maior rigor em [6].
5. Por fim, não é de nosso intuito atacar ou defender qualquer ideal, chapa, partido, ou organização política em particular nestas análises; de qualquer forma, a política interna e externa não são os únicos fatores a afetarem as condições de mercado de um país.
6. Ainda que estes adendos sejam – em nossa opinião – justificáveis, não extendemos, neste estudo, nenhuma conclusão feita além do espaço hipotético, e cada uma das breves análises feitas é de cunho heurístico (como supracitado, incluem subjetividades) e curto. Encorajamos fortemente o leitor interessado a re-interpretar nossas análises com os dados providos, encontrar outros padrões em nossas métricas (ou quaisquer outras) que sejam passíveis de interpretação (tendo ela algum valor de *business* ou não), e/ou re-implementar os métodos descritos para outro *dataset* de interesse.

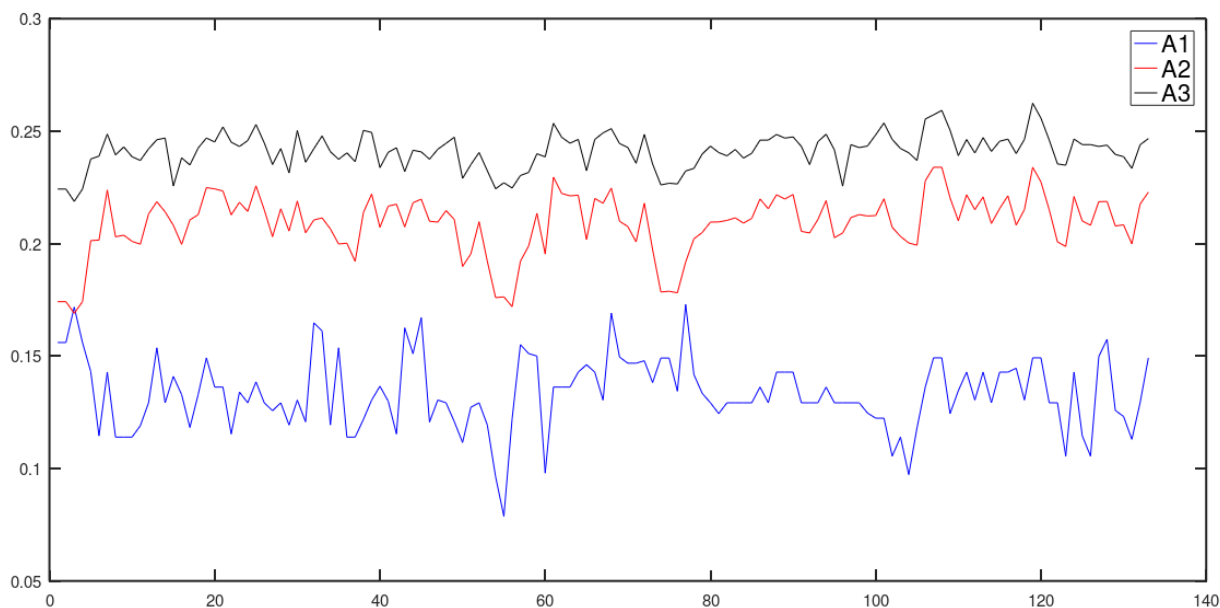
### 4.1 Tipo A

Obtivemos o valor aproximado de 89 para o suporte de nossas computações das distâncias de Hellinger. Nosso número total de pontos obtidos é 133 para cada indicador de tipo A.

A princípio, se considerarmos as advertências bem-embasadas em [6] de que o indicador  $A_2$  representa o comportamento de mercado com mais legibilidade e acurácia que  $A_1$ , podemos notar a dificuldade em encontrar – com assertividade – intervalos de tempo que caracterizem uma crise, analisando apenas  $A_2$  crescendo e  $A_3$  caindo, por notarmos que  $A_2$  e  $A_3$  majoritariamente se comportam de forma muito semelhante.

---

<sup>\*</sup>Clássica falácia anedótica-estatística de seleção observacional incompleta, que serve apenas como auto-confirmação de algum viés do autor.



(2) Séries temporais dos indicadores tipo A.

Creemos que isto se dá devido à incompatibilidade do mercado brasileiro com os *datasets* escolhidos em [6], no qual as discrepâncias entre  $A2$  e  $A3$  se mostraram menos ofuscadas. Conjecturamos com certa segurança que, com o *dataset* utilizado i.e. apenas 7 índices no período escolhido, não há uma grande volatilidade base caracterizando a economia saudável como acontece em [6]; isto também é evidenciado pela distância de Hellinger em  $A1$  ser consideravelmente e efetivamente menor que  $A2$  e  $A3$ .

Achamos provável que diferentes escolhas de hiperparâmetros e *ensembles* utilizados poderiam representar melhor uma economia saudável brasileira, que seguramente se comporta de maneira distinta aos *datasets* originalmente utilizados.

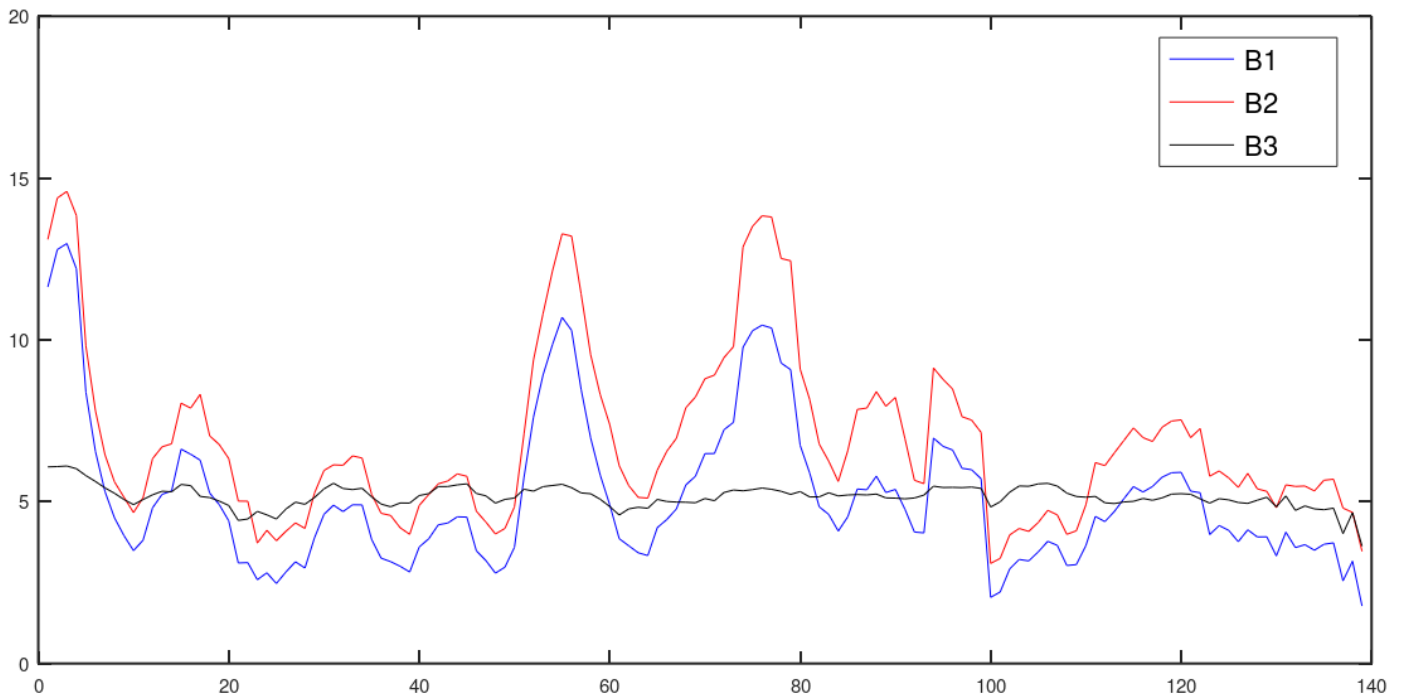
Isto não quer dizer que  $A2$  nos seja inútil – afinal, nos permitiu discorrer e conjecturar sobre como e por quê  $A2$  esteja, no mínimo parcialmente, sugerindo um mercado turbulento, ao invés de sua premissa original; além disto, julgamos que, caso  $A2$  suba e  $A3$  caia em concorrência, isto plausivelmente é um indicador mais forte de crise confinante, pois revela que autovalores maiores que o suporte dos autovalores de  $C2$  (desconsiderando seu raio espectral) estão entrando em ainda maior predominância. Infelizmente, não vimos isto acontecendo para nossas séries temporais construídas, de forma propositiva.

Enfim, a princípio vamos comparar  $A1$  e  $A3$ , tendo em mente que  $A1$  é mais suscetível à ruído, e portanto pode fornecer mais falsos positivos em nossa análise, por exemplo. Vemos fevereiro de 2015 como o melhor candidato nesta inspeção. Não encontramos grandes acontecimentos/escândalos políticos e/ou macroeconômicos que pudessem justificar facilmente a presença deste ponto. Contudo, a justificativa para esta recessão supostamente indicada pode ser explicadas pelo ruído de  $A1$ , ou causada por razões mais convolutas.

Outro candidato que soa promissor é fevereiro de 2014, e este nos traz um bom resultado, visto que a recessão político-econômica que nos afeta até a atualidade teve seu início em meados de 2014; acreditamos que este candidato assegura a funcionalidade de nosso *framework* construído.

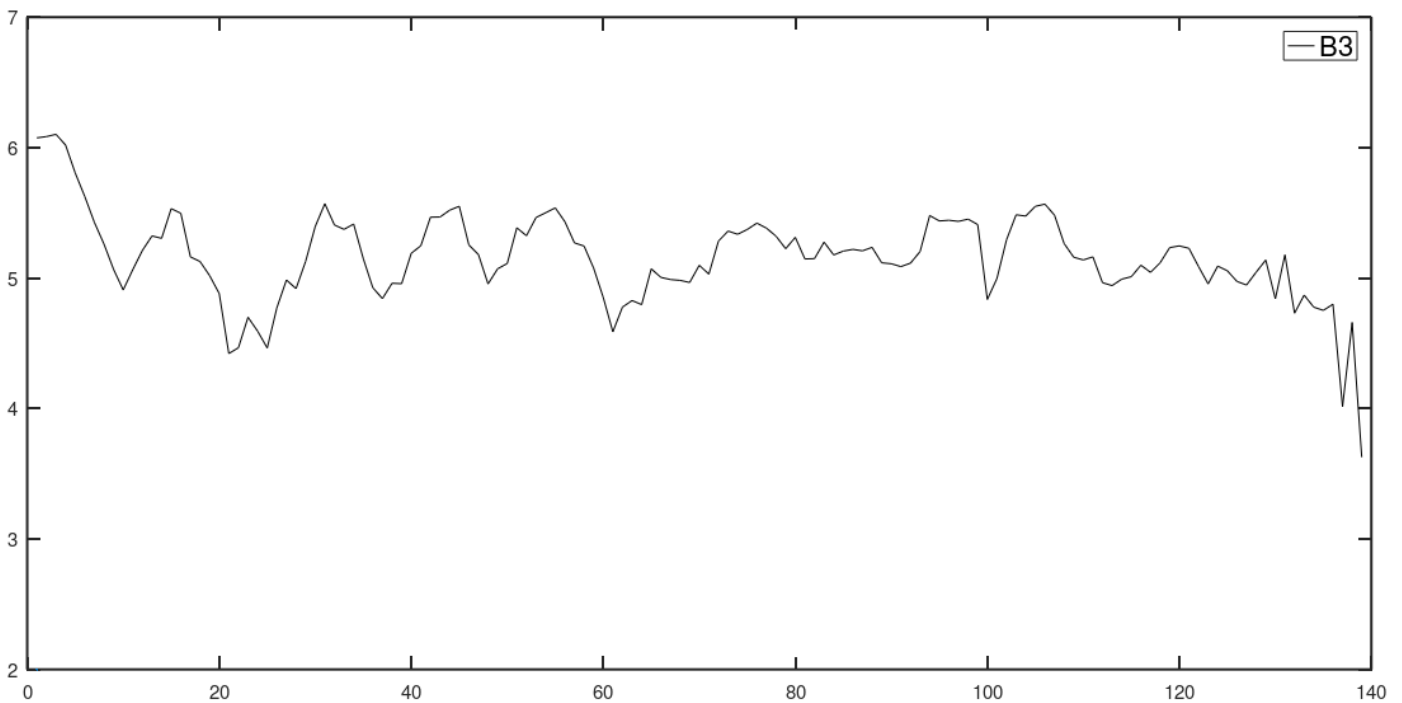
## 4.2 Tipo B

Obtivemos 139 pontos para cada indicador do tipo  $B$ . Esperamos uma boa funcionalidade destes índices, visto que o *approach* algébrico puro é um dos pilares importantes da teoria moderna de portfólios [9].

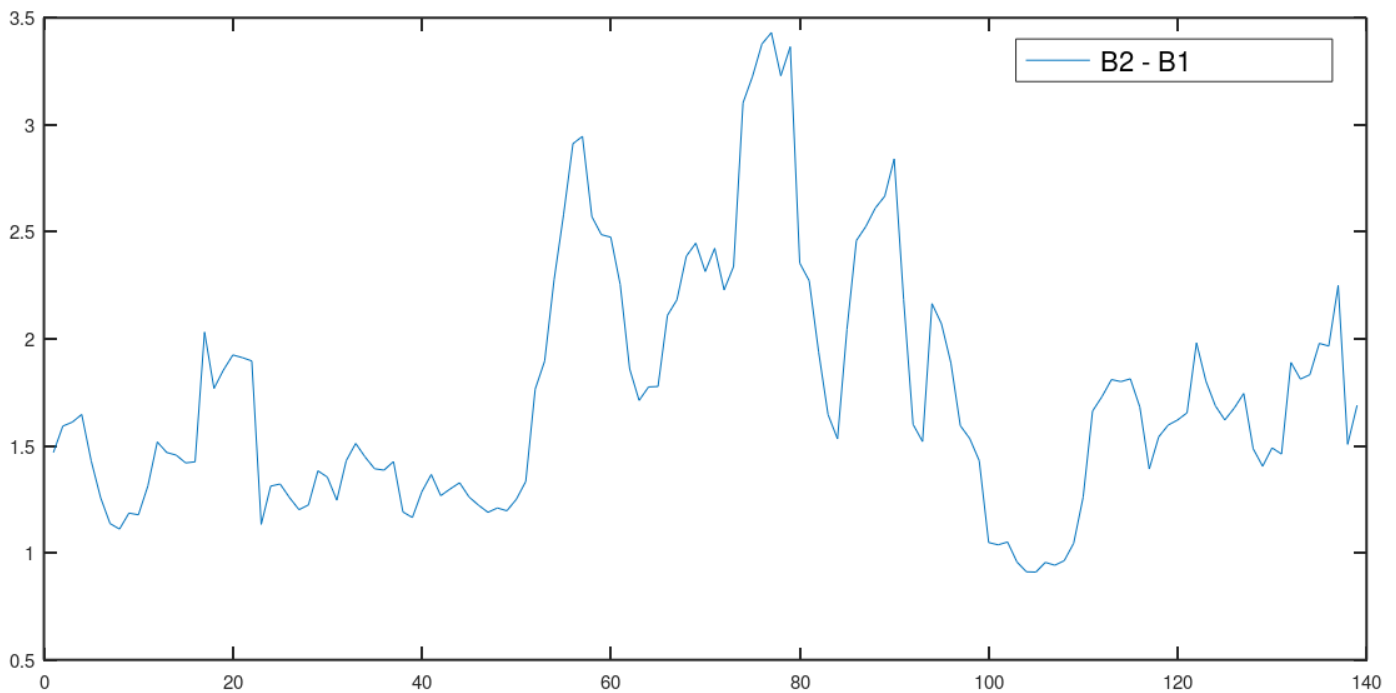


(3) Séries temporais dos indicadores tipo  $B$ .

É notável que o indicador  $B3$  aparenta apresentar monotonia, visto sua baixa variação, em relação aos outros dois, mas frisamos que os valores numéricos absolutos são de baixa importância em nossa análise, visto as diferentes origens dos dados e a ausência de uma interpretação clara sobre o valor numérico puro.



(4) Série temporal de  $B3$ , isolada das demais.



(5) Série temporal da diferença entre  $B2$  e  $B1$ .

Podemos ver que, em diversos períodos (final de 2014, início e final de 2016, por exemplo), as diferenças entre o traço e o raio espectral da matriz de covariâncias se acentua, indicando que não há extrema dominância de apenas um portfólio. Nestes casos citados, nas datas coincidentes, houve um crescimento notável de  $B3$ , significando um aumento da correlação interna entre os índices setoriais; é possível que, visto que estes períodos supracitados os índices de correlação e volatilidade estavam em alta, o mercado estava caracterizado por diferentes fluxos de investimento de alto risco nos diferentes setores.

Além disso, é notável a queda da diferença entre  $B2$  e  $B1$  em meados de junho de 2017; esta queda é inequivocadamente acompanhada de um considerável decréscimo de  $B3$ , indicando um ponto de inflexão do mercado, no qual índices setoriais tornaram-se voláteis separadamente (baixa correlação), indicando – pela nossa interpretação – uma potencial crise, possivelmente fruto de uma “força” externa como pontapé inicial; de fato, este período foi denotado por um dos auges da crise política atuante no país na época, protagonizados (neste intervalo) pela divulgação do acordo de delação premiada da *JBS*, dando maior descrédito ao relacionamento de empresas grandes brasileiras e chapas políticas.

Vemos o mesmo fenômeno acontecer – em aparente menor escala – em outros intervalos como no final de outubro de 2013, final de setembro de 2012, e meados de 2014.

## 5 Conclusões

Como dito anteriormente, a alta volatilidade do mercado é normalmente associada com períodos de crise econômica, mas existem alguns proponentes que argumentam a relação inversa, ou seja: baixa volatilidade de mercado também pode indicar uma crise econômica no futuro [3]; a argumentação se firma nas teorias de psicologia holística de mercado: a baixa volatilidade influencia investidores a tomarem posições mais arriscadas – as quais plausivelmente influenciarão negativamente a estabilidade de mercado.

Note que isto não desprova o consenso de que períodos de alta volatilidade estão fortemente associados de forma direta à crises econômicas. Pressupomos como provável a conciliação e realidade das duas frentes. No mercado americano, por exemplo, o índice *VIX*, que mede a volatilidade esperada implicada pelas opções do índice *SP500*, é apelidado como “índice do medo”, reforçando a ideia de que volatilidade

e instabilidade são conceitos que usualmente andam juntos. Nossas interpretações sob alguns resultados da parte  $B$  também associam alta volatilidade com turbulência.

O relacionamento entre a volatilidade e a correlação de mercado é menos direta e simples do que inicialmente aparenta, dependendo altamente do *dataset* e do momento do mercado, mas ambos os componentes são abrangidos tanto individualmente quanto de forma conjunta na análise espectral das matrizes de covariância e correlação.

Finalmente, acerca de nossos resultados, perfazemos que obtivemos um sucesso, no mínimo, parcial, ao explorar os métodos cuja praticidade é empiricamente “provada”, sejam eles mais clássicos ou embasados na moderna teoria de matrizes aleatórias. Reconhecemos, na maior parte, intervalos de tempo em nossas interpretações que condizem com períodos de conturbação política e econômica nacional. Não conseguimos resultados das datas mais recentes que consigam indicar com certa veemência a iminência de uma crise econômica.

Ainda que o escopo, *dataset* e metodologia interpretativa fossem vastamente diferentes dos percorridos em *mono*, vimos indícios da praticidade e usabilidade dos métodos aqui estudados, construídos e utilizados, para nosso *dataset* escolhido. Não iremos atestar que o método funciona com certidão, pois isto seria um argumento circular<sup>†</sup>, visto que já assumimos que ele é sólido nos embasando em [6]; em momentos como este, nos lembramos que o caráter deste estudo é meramente exploratório.

## Referências

- [1] HELLINGER, E. *Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen* Crelle's Journal Vol. 136 210-271 1909
- [2] AMADEZAK, R. *On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries* Electronic Journal of Probability Vol. 16(37) 1068-1095 2011
- [3] D, Jon., V. Marcela e Z. Iknur *Learning from History: Volatility and Financial Crises* SRC Discussion Paper No 57 2016
- [4] DUARTE, A. e SANTOS, S. *Um estudo sobre a variabilidade de problemas-teste lineares com dados gerados aleatoriamente* <https://www.ime.unicamp.br/~ra164505/IC/Relat%c3%b3rio%20Final.pdf>
- [5] HENRY, S. *The Characteristic Function of the Student-t Distribution* Financial Mathematics Research Report No. FMRR006-95 2010
- [6] KORNPORST, A. e DOUADY, R. *An Empirical Approach to Financial Crisis Indicators Based on Random Matrices* International Journal of Theoretical and Applied Finance Vol. 21(3) 1850022 2018 <https://arxiv.org/pdf/1506.00806.pdf>
- [7] GOLDBERG, L., PAPANICOLAU, A. e SHKOLNIK, A. *The Dispersion Bias* <https://ssrn.com/abstract=3071328>
- [8] MARCENKO, V.A. e PASTUR, L.A. *Distribution of Eigenvalues for some sets of Random Matrices* Math. USSR Sb Vol. 1(4) 457-483 1967
- [9] MARKOWITZ, H.M. *Portfolio Selection* . The Journal of Finance. 7 (1): 77–91 1952

---

<sup>†</sup>Falácia lógica na qual a conclusão de um argumento é utilizada como sua premissa i.e. a argumentação só é sólida se considerarmos a conclusão como verdadeira de antemão.