

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Francine Tatsch

**Sistema Multiagente Aplicado à  
Competitividade entre as Distribuidoras de  
Combustíveis no Rio Grande do Sul no  
Segmento de Varejo**

Porto Alegre

2019

Francine Tatsch

**Sistema Multiagente Aplicado à Competitividade entre  
as Distribuidoras de Combustíveis no Rio Grande do Sul  
no Segmento de Varejo**

Trabalho de Diplomação em Engenharia Física, realizado sob orientação do Dr. Carlo Requião da Cunha e apresentado à Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharela em Engenharia Física.

Orientador: Prof. Dr. Carlo Requião da Cunha

Porto Alegre

2019

Francine Tatsch

## **Sistema Multiagente Aplicado a Competitividade entre as Distribuidoras de Combustíveis no Rio Grande do Sul no Segmento de Varejo**

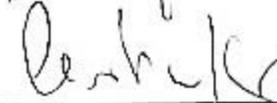
Trabalho de Diplomação em Engenharia Física realizado sob orientação do Dr. Carlo Requião da Cunha e apresentado a Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Bacharela em Engenharia Física.

Trabalho aprovado. Porto Alegre, 11 de Novembro de 2019:



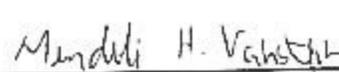
---

**Dr. Carlo Requião da Cunha**  
Orientador



---

**Dr. Cristiano Krug**  
UFRGS



---

**Dr. Mendeli Henning Vainstein**  
UFRGS

Porto Alegre  
2019

## AGRADECIMENTOS

A meus pais, Sergio Tatsch e Sirlei Adriane Tatsch, e a minha irmã, Sabrina Tatsch, por todo apoio, carinho e compreensão durante a minha caminhada na graduação.

A meus colegas de curso pela amizade e pelas valiosas discussões durante a graduação.

A meu orientador, Dr. Carlo Requião da Cunha, pelo apoio, paciência e ensinamentos compartilhados.

## RESUMO

O mercado de distribuição de combustíveis vem se desconcentrando nos últimos anos. Os novos entrantes estão introduzindo um grande desafio às distribuidoras que possuem marca já reconhecida: crescer em um mercado competitivo com mais jogadores e com margem de lucro cada vez menor, sendo que a tendência é ser bandeira branca - posto que não pertence a nenhuma rede de marca conhecida, como Raízen, Ipiranga e BR- e não sustentar a marca. Um dos principais dados que nos indicam como está evoluindo o mercado de combustíveis é o *Market Share*, o qual indica a participação no mercado das distribuidoras, isto é, a fatia de volume do mercado que cada uma está dominando.

Esse sistema de interação entre as distribuidoras em busca de maior representatividade no mercado consiste de agentes que exibem comportamento autônomo tomando decisões que levam à satisfação dos seus próprios objetivos, mas ao mesmo tempo interagem com os outros agentes presentes no sistema [1]. Neste trabalho, elaborou-se um modelo fenomenológico para a convergência de agentes em formação, partindo de posições aleatórias, de forma que eles se agrupem e sigam em uma determinada direção, em busca de *Market Share*. Nessa abordagem, essas interações podem ser simples, baseadas em atração e repulsão, ou mais complexas, consistindo em combinações dessas, dando-se entre vizinhos no espaço ou em uma rede subjacente, e os agentes não guardam memória de interações anteriores.

Dessa forma, o presente estudo tem por objetivo a simulação da evolução da competitividade entre as distribuidoras de combustíveis no mercado do Rio Grande do Sul - *Market Share* - sob a perspectiva de sistema multiagentes com o modelo de Vicsek como ferramenta estratégica.

**Palavras-chave:** Combustível, Distribuidoras, Competitividade, Sistema Multiagentes, Modelo de Vicsek, Sistemas dinâmicos, Movimento coletivo.

## ABSTRACT

The market of fuel distribution has recently become decentralized. New players are introducing a big challenge to the already established distributors: how to grow in a competitive market with more players, smaller profit margins and in an environment where the tendency is to be a white flag - since it does not belong to any known brand network, such as Raízen, Ipiranga and BR - and not to support the brand. One of the key data that tells us how the fuel market is evolving is Market Share, which indicates the distributors market share, that is, how much market volume each is dominating.

This system of interaction between distributors seeking to increase their representativeness in the market consists of agents who exhibit autonomous behavior making decisions that satisfy their own goals, but at the same time interact with other agents present in the system [1]. In this work, a model was elaborated that contemplates the construction of a phenomenological model for the convergence of agents in formation, starting from random positions, so that they group and follow in a certain direction, in search of *Market Share*. In this approach, these interactions may be simple, based on attraction and repulsion, or more complex, consisting of combinations of these, occurring between neighbors in space or an underlying network, and agents do not store memory of previous interactions.

Thus, the main objective of this work is to analyze the evolution of competitiveness among the fuel distributors in the market of Rio Grande do Sul - *Market Share*- under the perspective of multiagent systems with Vicsek Model as a strategic framework.

**Keywords:** Fuel, Distributors, Competitiveness, Multiagent System, Vicsek Model, Dynamical systems. Collective motion.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	11
1.1	Objetivos	12
<b>2</b>	<b>CONTEXTUALIZAÇÃO</b>	13
2.1	A concorrência no mercado de combustíveis	13
2.2	Linha do Tempo	14
2.3	<i>Market Share</i>	14
<b>3</b>	<b>ANÁLISE DOS DADOS</b>	16
3.1	Série temporal	16
3.1.1	Retornos	17
3.1.2	Densidade Espectral	18
3.1.3	Autocorrelação	19
3.1.4	Matriz de Correlação	20
3.1.5	Autovetores: clusterização de mercado	21
3.1.6	Momentos Padronizados	22
3.1.7	Probabilidade Cumulativa	25
3.2	Conclusões da Análise Estatística	26
<b>4</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	27
4.1	Modelo Baseado em Agentes	27
4.2	Modelo de Vicsek	29
4.2.1	Equação de Transporte de Boltzmann	30
4.2.2	Modelo Macroscópico - Limite Hidrodinâmico	33
<b>5</b>	<b>DESENVOLVIMENTO</b>	35
5.1	Simulação	35
5.2	Análise estatística da simulação	43
5.2.1	Conclusões da análise estatística	46
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	48
	<b>REFERÊNCIAS</b>	49

## LISTA DE ILUSTRAÇÕES

Figura 1 – Gráficos do <i>Market Share</i> mensal no RS - diesel, gasolina e etanol.	15
Figura 2 – Série temporal do <i>Market Share</i> em um período de 60 meses.	16
Figura 3 – Ganhos de Mercado para o período de janeiro de 2014 a dezembro de 2019.	17
Figura 4 – Ganhos de Mercado Normalizados para o período de janeiro de 2014 a dezembro de 2019.	18
Figura 5 – Densidade espectral de energia do <i>Market Share</i> .	19
Figura 6 – Gráfico de autocorrelação do conjunto de dados do <i>Market Share</i> no período de 2014 a 2018.	20
Figura 7 – Autovetores do conjunto de dados do <i>Market Share</i> no período de 2014 a 2018.	22
Figura 8 – Função densidade de probabilidade do <i>Market Share</i> para a) BR, b) Ipiranga, c) Raízen e d) Independentes.	24
Figura 9 – Momento normalizado em função do índice $k$ do conjunto de dados do <i>Market Share</i> no período de 2014 a 2018. A curva composta de círculos corresponde ao polinômio de grau 3 $b_3x^3 + b_2x^2 + b_1x + c$ ajustado a média dos dados do <i>Market Share</i> , sendo $b_1 = -0,404$ , $b_2 = 0,209$ , $b_3 = -6,99 \cdot 10^{-4}$ e $c = 0,975$ .	25
Figura 10 – Distribuição cumulativa do conjunto de dados do <i>Market Share</i> no período de 2014 a 2018.	26
Figura 11 – Simulação do <i>Market Share</i> com os seguintes fatores: $\xi$ (nível de ajuste) = 0,1, $\eta$ (nível de ruído) = 0,001, 4 agentes (BR, Ipiranga, Raízen e Independentes), 5 estratégias para a) 5200 interações b) zoom nas interações finais.	38
Figura 12 – Simulação do <i>Market Share</i> para diferentes níveis de ruído a) $\eta = 0,001$ , b) $\eta = 0,01$ , c) $\eta = 0,1$ , d) $\eta = 1,0$	38
Figura 13 – Simulação do <i>Market Share</i> viciando um agente em uma estratégia para a) 5200 interações; b-I e b-II) zoom de um intervalo das interações.	39
Figura 14 – Simulação do <i>Market Share</i> ponderando as estratégias para a) 5200 interações; b) zoom de um intervalo das interações.	40
Figura 15 – Simulação do <i>Market Share</i> ditando o perfil de negócio de cada agente.	42
Figura 16 – a) Interpolação dos pontos iniciais para análise estatística da reprodução do <i>Market Share</i> , b) zoom na etapa em que independentes supera o <i>Market Share</i> da Ipiranga e da BR, c) projeção para os próximos anos d) resultado final onde temos um agente agindo mais agressivamente.	43
Figura 17 – Comparativo dos dados reais versus dados simulados do <i>Market Share</i> .	44

Figura 18 – Momento normalizado em função do índice $k$ do conjunto de dados do	
<i>Market Share</i> simulado. A curva composta de círculos corresponde ao	
polinômio de grau 3 $b_3x^3 + b_2x^2 + b_1x + c$ ajustado a média dos dados	
simulados do <i>Market Share</i> , sendo $b_1 = 0,261$ , $b_2 = -0,830$ , $b_3 = 0,362$	
e $c = 0,982$ .	45
Figura 19 – Probabilidade cumulativa do conjunto de dados do Market Share.	46
Figura 20 – Clusterização obtida com os dados simulados.	46

## LISTA DE TABELAS

Tabela 1 – Evolução do <i>Market Share</i> de Combustíveis do Varejo no RS. . . . .	15
Tabela 2 – Matriz de Correlação dos dados do <i>Market Share</i> . . . . .	21
Tabela 3 – Parâmetros estatísticos do <i>Market Share</i> . . . . .	23
Tabela 4 – Parâmetros estatísticos do <i>Market Share</i> simulado. . . . .	44
Tabela 5 – Matriz de Correlação dos dados simulados do <i>Market Share</i> . . . . .	45

## LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional do Petróleo
ABM	Modelo Baseado em Agentes

## LISTA DE SÍMBOLOS

$\mu$	Valor Médio
$S(X)$	Obliquidade
$K(X)$	Curtose
$\omega$	Frequência angular
$F(\omega)$	Transformada de Fourier de $f(t)$
$F^*(\omega)$	Conjugado complexo da transformada de Fourier
$R(\tau)$	Autocorrelação
$\mathbf{p}$	Momento
$\mathbf{r}$	Posição
$f(\mathbf{r}, \mathbf{p}, t)$	Densidade de Probabilidade
$p_m$	Dados do <i>Market Share</i>
$r_m$	<i>Log-retorno</i>
$n_m$	<i>Log-retorno</i> normalizado
$S(\omega)$	Densidade Espectral
$\sigma$	Nível de Ruído
$dW_t$	Processo de Wiener

# 1 INTRODUÇÃO

Competitividade é um imperativo para o bom funcionamento dos mercados, em que a concorrência acarreta benefícios aos consumidores e aos empreendedores mais eficientes e inovadores. Sendo uma atividade econômica concorrencial no seu todo, a cadeia de combustíveis é tipicamente desenhada para proporcionar a maior competitividade possível, de forma a se buscar máximo bem-estar econômico, decorrente de:

- Eficiência Produtiva: incentivar a produtividade de tal forma que se pratique os menores custos a partir dos recursos disponíveis;
- Eficiência de Alocação de Recursos: sinalizar preços que reflitam as alternativas de suprimento e incentivem os investimentos necessários;
- Eficiência Dinâmica: favorecer a inovação que permita saltos produtivos no futuro.

No mercado de combustíveis o cenário de competitividade vem solidificando-se no RS com novos entrantes, cada qual com diferentes estratégias de negócios e ambições quanto à atuação na cadeia. Esses novos jogadores aumentam a competitividade e minimizam a possibilidade do exercício de poder de mercado via precificação abusiva. Por outro lado, uma maior fragmentação do mercado implica a necessidade de investir para a viabilização de diferenciais a fim de se destacar frente à concorrência e continuar a oferecer um modelo de negócio competitivo e rentável.

A questão é: em quais quesitos a revenda e o consumidor atribuem maior valor -excelência operacional, preço, tecnologia - e como está se projetando o crescimento do mercado de combustíveis no Rio Grande do Sul? Um dos principais dados que nos indicam como está evoluindo o mercado de combustíveis é o *Market Share*, o qual apresenta a participação no mercado das distribuidoras, isto é, a fração de volume que cada uma está dominando.

Como ferramenta para simular as interações estratégicas desse cenário será utilizado um sistema multiagentes com o modelo de Vicsek, o qual busca modelar sistemas a partir de uma coleção dinâmica de agentes autônomos heterogêneos -as distribuidoras, cada qual com suas estratégias e interesses, resultando em uma transição de fase de movimento browniano para movimento coletivo ordenado. Comportamentos coletivos de sistemas multiagentes são importantes pontos de partida no estudo de sistemas complexos, e o modelo de Vicsek é um modelo típico de sistemas multiagentes. Além de sua simplicidade, ele preserva alguns dos principais recursos de sistemas complexos, como comportamentos dinâmicos, interações locais e mudança de vizinhança.

## 1.1 Objetivos

Em linha com o problema apresentado, o presente trabalho tem como objetivo simular os dados experimentais do *Market Share* no segmento de varejo das distribuidoras de combustíveis no Rio Grande do Sul por meio de sistemas multiagentes autônomos descritos pelo modelo de Vicsek para que se possa modelar o processo de decisão desse cenário estratégico de competitividade entre elas. Na primeira etapa do trabalho, capítulo 3, foi feita uma análise da série temporal para entender os dados que possuímos; na segunda etapa, capítulo 4, o objetivo foi implementar um modelo e, na terceira, capítulo 5, verificar se as propriedades das séries que esse modelo gera são compatíveis com a série real.

## 2 CONTEXTUALIZAÇÃO

### 2.1 A concorrência no mercado de combustíveis

O mercado de distribuição de combustíveis é um ambiente que está se tornando cada vez mais competitivo e menos concentrado, o que resulta em margens de lucro comprimidas. BR, Ipiranga e Raízen dominam mais de 70% da distribuição de combustíveis do varejo no Rio Grande do Sul e apresentam um grande valor agregado à marca; os outros quase 30% são de independentes, os quais estão em fase de expansão.

Principais características das distribuidoras de combustíveis:

- BR, Ipiranga e Raízen: distribuidoras mais conhecidas no mercado e que possuem maior representatividade;
- Raízen (Shell): Investimento em excelência operacional, produto diferenciado e tecnologia. Realiza hoje pensando no futuro;
- Independentes: o *Share* de independentes inclui todos os novos entrantes do mercado, como Rodoil, Ale, Mega Petro, Vitol, Glencore, Total, Petronac, entre outros. Apresentam maior agressividade nos preços <sup>1</sup> e nas propostas de embandeiramento - propostas realizadas a postos bandeira branca para eles fazerem uso da sua marca-, pois vêm buscando expansão;
- Bandeira Branca: não ostenta marca conhecida e tradicional, o que o permite comprar produto de qualquer distribuidora com preços reduzidos - recebe ofertas diariamente de diferentes distribuidoras, todas com objetivo de ganhar volume e realizar futuro embandeiramento.

As soluções baseadas em agentes são projetadas para lidar com ambientes complexos e em constante mudança, onde há potencialmente várias opções e resultados em cada estágio do processo. Os agentes são simulados para serem autônomos na busca de seu objetivo, mas capazes de interagir com outros agentes, seja colaborativamente ou em competição. Sistemas multiagentes, portanto, podem oferecer benefícios comerciais mensuráveis para muitas organizações e um modelo básico de sistemas multiagentes é o chamado modelo Vicsek que possui algumas características principais de sistemas complexos, como comportamento dinâmico, interação local e mudança de vizinhança. As distribuidoras apresentadas acima, nesse estudo, serão os agentes a serem estudados.

---

<sup>1</sup> Preço de venda do produto aos postos.

## 2.2 Linha do Tempo

Os últimos anos apresentaram marcos históricos que tiveram impacto na distribuição de combustíveis e que devem ser levados em consideração em qualquer análise:

- 2010: Cosan e Shell anunciam *joint venture*<sup>2</sup>, Raízen, para integrarem seus negócios na distribuição de combustíveis e na indústria da cana-de-açúcar;
- 2015: Abertura de arbitragem para derivados;
- 2016: Contrato de fornecimento da Petrobras ajustado;
- 2016/17: Nova política de preços Petrobras; desenvolvimento de infraestrutura;
- 2017/18: Entrada de importadores;
- 2018: Greve dos caminhoneiros;
- 2018/2019: Ambiente competitivo e novos entrantes.
- 2019: Venda de 30 % do capital da BR Distribuidora

Esses marcos tiveram e estão tendo relevância no *Market Share*, principalmente dos novos entrantes que estão aumentando a competitividade.

## 2.3 *Market Share*

O *Market Share* de combustíveis do RS nos mostra a participação no mercado das distribuidoras, isto é, o volume de combustível comercializado por cada uma frente ao total no período. No gráfico da Figura 1 e na Tabela 1 está apresentado o *Market Share* RS 2. Nesses dados podemos verificar um claro crescimento da participação de independentes, redução da participação da BR e da Ipiranga e Raízen mantendo-se estável.

---

<sup>2</sup> Expressão de origem inglesa, que significa a união de duas ou mais empresas já existentes com o objetivo de realizar uma atividade econômica comum e visando, dentre outras motivações, o lucro

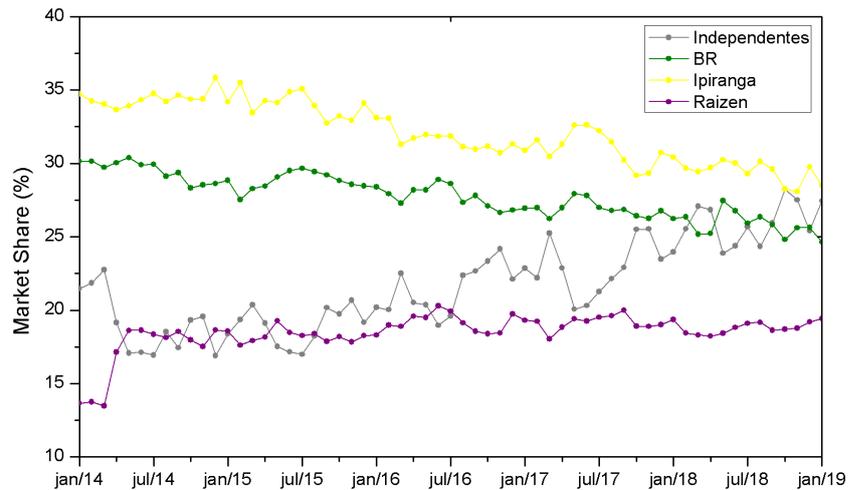


Figura 1 – Gráficos do *Market Share* mensal no RS - diesel, gasolina e etanol.

Tabela 1 – Evolução do *Market Share* de Combustíveis do Varejo no RS.

Bandeira	2014	2015	2016	2017	2018	mar/2019
BR	29,5 %	28,8%	27,7 %	26,9 %	25,9 %	25,5 %
Ipiranga	34,4 %	34,0 %	31,7 %	31,0 %	29,5 %	29,0 %
Raízen	17,1 %	18,2 %	19,1 %	19,1 %	18,8 %	18,9 %
Independentes	19,0 %	18,9 %	21,4 %	22,9 %	25,8 %	26,6 %

Todas as informações trazidas nessa seção são de extrema importância na simulação do cenário da competitividade das distribuidoras no RS pelo sistema multiagentes. Esse sistema irá ajudar a indicar quais ações/estratégias as distribuidoras realizam em interações com as outras e que melhor atendam a seus interesses - aumento do *Market Share*. Como modelo e ponto inicial será buscado reproduzir os fatos estilizados que a série temporal do *Market Share* apresenta.

## 3 ANÁLISE DOS DADOS

Esse capítulo tem como objetivo realizar uma primeira modelagem dos dados para entender as informações que possuímos do *Market Share*. Por meio dessa análise da série temporal será apresentado os retornos, retornos normalizados, periodograma, autocovariância, matriz de covariância, autovetores (clusterização de mercado), análise de momentos padronizados, distribuição cumulativa e estimativa de kernel. Para a realização desta análise estatística foram utilizados programação em *Python* e o programa *Origin* e as análises estatísticas escolhidas foram com base no artigo presente na referência [3].

### 3.1 Série temporal

Uma série temporal é dita contínua quando as observações são feitas continuamente no tempo. Definindo o conjunto  $T = \{t : t_1 < t < t_2\}$  a série temporal será denotada por  $\{X(t) : t \in T\}$ . Uma série temporal é dita discreta quando as observações são feitas em tempos específicos, geralmente equiespaçados. Definindo o conjunto  $T = \{t_1, \dots, t_n\}$  a série temporal será denotada por  $\{X_t : t \in T\}$ . A série temporal discreta do *Market Share* mensal pode ser visualizada na Figura 2.

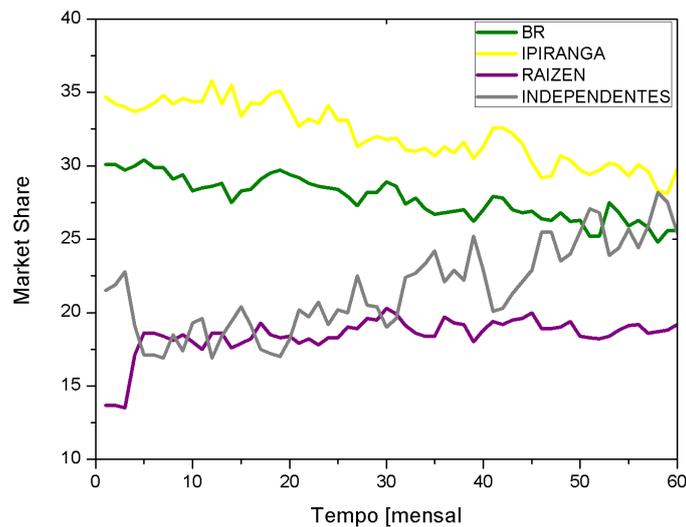


Figura 2 – Série temporal do *Market Share* em um período de 60 meses.

### 3.1.1 Retornos

Os retornos de *log* do *Market Share* foram calculados como  $r_m = \log(p_{m+1}) - \log(p_m)$ , onde  $p_m$  são os dados do *Market Share*, e são mostrados a Figura 3.

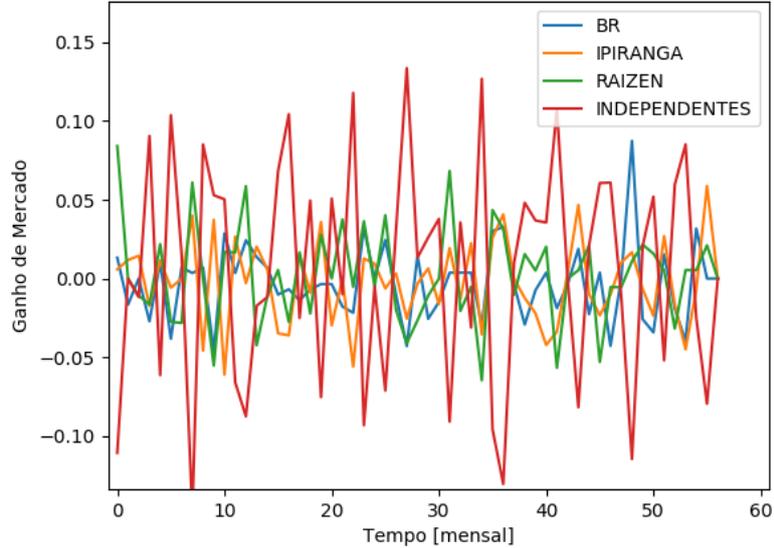


Figura 3 – Ganhos de Mercado para o período de janeiro de 2014 a dezembro de 2019.

Para que possamos ter uma comparação mais precisa entre as distribuidoras é importante a análise dos ganhos normalizados, pois os ganhos são ajustados para eliminar os efeitos da sazonalidade, receita e despesas que são incomuns ou influências únicas resultando em uma avaliação mais precisa da verdadeira integridade e desempenho de uma empresa.

Os retornos de *log* foram normalizados conforme a Equação 3.1 e são mostrados na Figura 4.

$$n_m = \frac{r_m - \langle r \rangle}{(\langle r^2 \rangle - \langle r \rangle^2)^{\frac{1}{2}}} \quad (3.1)$$

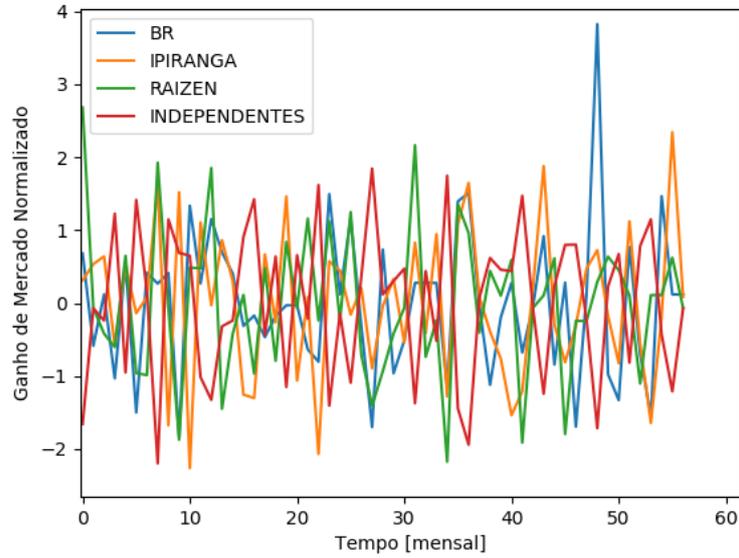


Figura 4 – Ganhos de Mercado Normalizados para o período de janeiro de 2014 a dezembro de 2019.

### 3.1.2 Densidade Espectral

A densidade espectral de energia descreve como a energia de um sinal ou uma série temporal é distribuída com a frequência e nos permite identificar periodicidades. Se  $f(t)$  é uma função integrável de energia finita, a densidade espectral  $S(\omega)$  do sinal será o quadrado da magnitude da transformada de Fourier do sinal, conforme Equação 3.2 [4].

$$S(\omega) = \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \right|^2 = \frac{F(\omega)F^*(\omega)}{2\pi} \quad (3.2)$$

No caso estudado estamos trabalhando com sinais discretos com valores  $f_n$ , portanto, têm-se uma densidade espectral de energia conforme a Equação 3.3. Aplicando aos dados, obtemos o gráfico de densidade espectral de energia da Figura 5, o qual nos indica que há uma periodicidade sazonal no *Market Share*.

$$S(\omega) = \left| \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} f_n e^{-i\omega n} \right|^2 = \frac{F(\omega)F^*(\omega)}{2\pi} \quad (3.3)$$

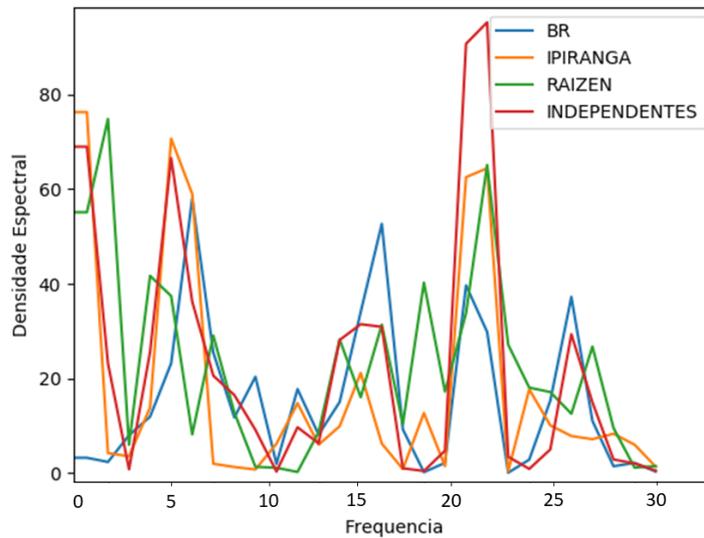


Figura 5 – Densidade espectral de energia do *Market Share*.

### 3.1.3 Autocorrelação

Uma importante ferramenta para identificar as propriedades de uma série temporal consiste no coeficientes de autocorrelação amostral. Autocorrelação é uma representação matemática do grau de similaridade entre uma determinada série temporal e uma versão defasada de si mesma em intervalos de tempo sucessivos. O teorema de Wiener-Khinchin estabelece uma importante relação entre a autocorrelação de um sinal e a sua densidade espectral; segundo esse teorema temos que a autocorrelação é dada pela transformada de Fourier inversa da densidade espectral, conforme a Equação 3.4

$$R(\tau) = F^{-1}(S(\omega)) \quad (3.4)$$

Segundo Damodaran (1997) <sup>[4]</sup> o teste de hipóteses de autocorrelação avalia a ligação entre as mudanças dos dados em períodos de tempo consecutivos, sendo uma medida de quanto às mudanças dos dados depende da mudança dos dados ao longo do período anterior. Quando a autocorrelação de hoje com a de ontem é zero, o sistema não possui memória e temos um mercado eficiente <sup>[5]</sup>.

O periodograma obtido para a série temporal do *Market Share* pode ser visualizado na Figura 6.

<sup>1</sup> Professor de finanças da *Stern School of Business*, na Universidade de New York, onde ensina finanças corporativas e avaliação de ações.

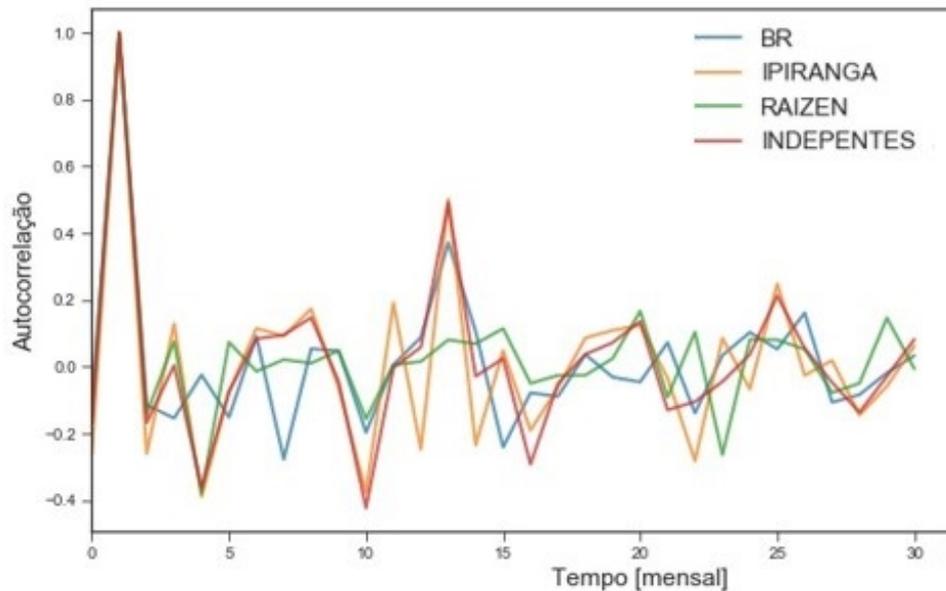


Figura 6 – Gráfico de autocorrelação do conjunto de dados do *Market Share* no período de 2014 a 2018.

A análise das autocorrelações nos indicam que temos um mercado não eficiente, pois apesar das autocorrelações das distribuidoras caírem rapidamente a zero, há oscilações menores e, posteriormente, o retorno de uma alta autocorrelação, ou seja, há memória.

### 3.1.4 Matriz de Correlação

A matriz de correlação indica como uma variável se comporta em um cenário onde outra está variando, visando identificar se existe alguma relação entre a variabilidade de ambas. Importante ressaltar que a existência de forte correlação não implica em um efeito de causalidade.

O coeficiente de correlação mais amplamente utilizado para medir o grau de relação entre as variáveis relacionadas linearmente é o coeficiente de correlação de *Pearson*. Considerando duas variáveis,  $x$  e  $y$ , o coeficiente de correlação de *Pearson*,  $r$ , é dado pela Equação 3.5.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.5)$$

O coeficiente de correlação de Pearson pode assumir valores entre -1 e +1. Quando  $r$  se aproxima de -1 ou +1, a relação linear entre as variáveis numéricas passa a ser mais forte; quando próximo de zero, a relação linear é fraca. O sinal indica se as variáveis estão diretamente ou inversamente correlacionadas. A matriz de correlação obtida sobre os dados do *Market Share* mês a mês pode ser visualizada na Tabela 2.

Tabela 2 – Matriz de Correlação dos dados do *Market Share*.

	BR	IPIRANGA	RAÍZEN	INDEPENDENTES
BR	1			
IPIRANGA	0,299	1		
RAÍZEN	0,383	0,290	1	
INDEPENDENTES	-0,688	-0,775	-0,716	1

Por meio da matriz de correlação verifica-se que BR, Ipiranga e Raízen possuem um fator bem próximo de -1 quando se analisa a correlação com independentes, o que nos indica que há um alto grau de relação inversamente proporcional entre o *Market Share* de independentes e de BR, Ipiranga e Raízen.

### 3.1.5 Autovetores: clusterização de mercado

Sistemas de recomendação e recuperação de informações comumente usam técnicas e metodologias de clusterização para reconhecer e extrair padrões em um conjunto de dados. Nesse tipo de análise, o objetivo é organizar objetos em grupos ou *clusters*, de forma que os objetos do mesmo *cluster* possuam uma semelhança significativa entre si por algum tipo de critério estabelecido [6].

O agrupamento de objetos é determinado em função de uma medida de distância. Desta forma, o objetivo de um algoritmo de clusterização é encontrar e classificar itens similares de maneira que a distância intra-*cluster* (soma das distâncias dos objetos de um *cluster* até o seu centro) seja a menor possível e a distância entre diferentes *clusters* seja maximizada [7].

Os autovetores calculados a partir da matriz de covariância representam as direções em que os dados acumulam a maior variância. Já os respectivos autovalores determinam a quantidade de variância naquela direção. Uma vez que obtemos os autovetores, podemos selecionar aqueles onde serão projetados os dados. Os autovetores selecionados são chamados componentes principais.

O critério para seleção dos autovetores é o total de variância que ele representa. Como já foi dito, essa variância é representada pelo autovalor de cada autovetor. A soma dos autovalores é igual a variância total dos dados. A clusterização de mercado pode ser visualizada na Figura 7, onde os extremos indicam os autovetores (0 a 3) e os eixos são os autovalores (-0,8 a +0,8). Por meio dela podemos verificar que BR, Ipiranga e Raízen formam um *cluster* separado das independentes, sendo que deste *cluster* a BR e a Raízen formam um *subcluster*.

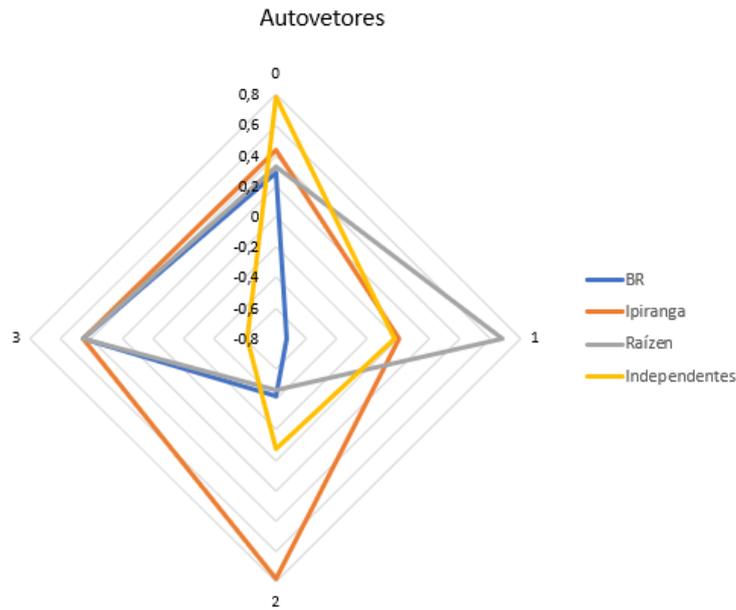


Figura 7 – Autovetores do conjunto de dados do *Market Share* no período de 2014 a 2018.

### 3.1.6 Momentos Padronizados

O  $k$ -ésimo momento padronizado é dado pela Equação 3.6.

$$m_k = \frac{\mu_k}{\sigma^k} = \frac{\langle (X - \mu)^k \rangle}{[\langle (X - \mu)^2 \rangle]^{\frac{k}{2}}} = \frac{\int_{-\infty}^{\infty} (x - \mu)^k P(x) dx}{[\langle (X - \mu)^2 \rangle]^{\frac{k}{2}}} \quad (3.6)$$

O primeiro momento padronizado é nulo, pois  $\langle (X - \mu) \rangle = 0$ . O segundo momento padronizado é igual a 1, pois  $\mu^2 = \sigma^2$ . O terceiro momento padronizado chama-se obliquidade e nos fornece uma ideia sobre a assimetria da distribuição. (Equação 3.7). O quarto momento padronizado chama-se curtose (Equação 3.8).

$$S(X) = \frac{\mu_3}{\sigma^3} \quad (3.7)$$

$$K(X) = \frac{\mu_4}{\sigma^4} \quad (3.8)$$

A curtose nos fornece informação apenas sobre a cauda, deixando o pico basicamente de fora. O excesso de curtose é definido como a própria curtose subtraída de 3. O excesso de curtose amostral, portanto, é dado pela Equação 3.9

$$J(X) = \frac{\mu_4}{\sigma^4} - 3 \quad (3.9)$$

Caso o excesso de curtose seja nulo, dizemos que a distribuição é mesocúrtica - exemplo clássico é a distribuição normal. Se ele for positivo, dizemos que ela é leptocúrtica

ou super-Gaussiana - neste caso, a distribuição possui cauda gorda e exemplos incluem a distribuição exponencial. Já se o excesso de curtose for negativo, a distribuição é dita platicúrtica ou sub-Gaussiana - estas são distribuições de cauda leve como a distribuição uniforme.

Na Tabela 3 está apresentado os principais momentos e outros parâmetros estatísticos relevantes calculados.

Tabela 3 – Parâmetros estatísticos do *Market Share*.

Bandeira	Média	Variância	Mediana	Curtose	Obliquidade
BR	27,678	1,889	27,818	-0,843	-0,042
Ipiranga	32,037	3,952	31,840	-1,074	-0,013
Independentes	21,560	9,589	20,675	-0,939	0,285
Raízen	18,725	0,413	18,551	-1,176	-0,297

Em estatística a estimativa de densidade do kernel é uma maneira não-paramétrica de estimar a função de densidade de probabilidade de uma variável aleatória. A estimativa da densidade de kernel é um problema fundamental de suavização de dados, no qual são feitas inferências sobre a população, com base em uma amostra de dados finitos.

Seja  $(x_1, x_2, \dots, x_n)$  uma amostra univariada independente e identicamente distribuída, extraída de alguma distribuição com uma densidade desconhecida  $f$ . Estamos interessados em estimar a forma dessa função. Seu estimador de densidade do kernel é dado por:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n L_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n L\left(\frac{x - x_i}{h}\right)$$

onde  $L$  é o kernel - uma função não negativa - e  $h > 0$  é um parâmetro de suavização chamado largura de banda.

A função de kernel utilizada foi o kernel Epanechnikov [8], o qual é o ideal no sentido de variância mínima e devido às suas propriedades matemáticas convenientes.  $K$ , portanto, é dado pela Equação 3.10 e o tamanho da janela utilizada é dado pela Equação 3.11. A função densidade de probabilidade pode ser visualizada na Figura 8.

$$L(u) = \frac{3}{4}(1 - u^2) \quad (3.10)$$

$$h = 1,06\sigma^2 N^{-0,2} \quad (3.11)$$

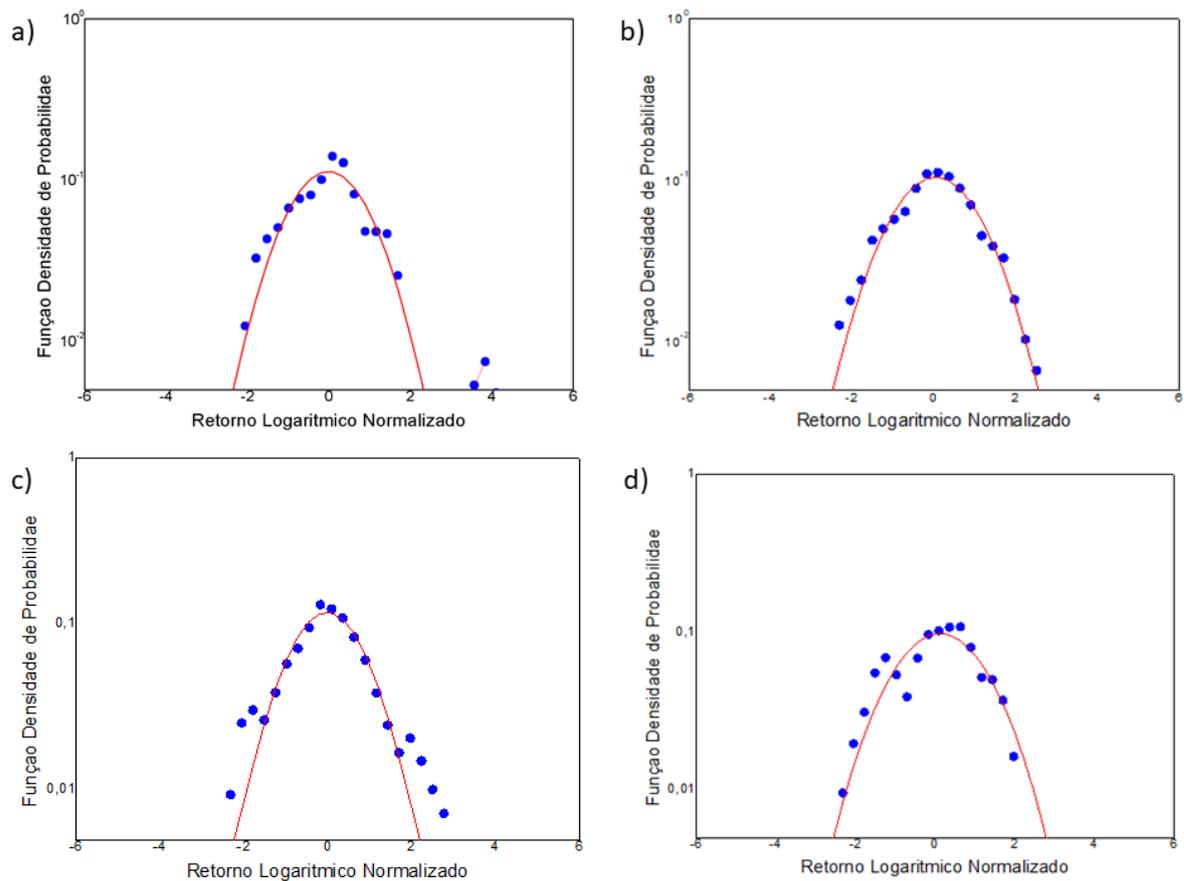


Figura 8 – Função densidade de probabilidade do *Market Share* para a) BR, b) Ipiranga, c) Raízen e d) Independentes.

Pelos parâmetros e gráficos apresentados verificamos que as distribuições se assemelham a uma gaussiana e apresentam alguns pontos que podemos destacar:

- Raízen, BR, Ipiranga e independentes apresentam distribuição platicúrtica. Dentre elas a Raízen apresenta um maior grau de achatamento da função de distribuição de probabilidade, isto é, possui um menor grau de concentração de valores da distribuição em torno do centro desta distribuição;
- Independentes apresentam uma distribuição com cauda mais pesada a direita, enquanto BR, Ipiranga e Raízen possuem essa cauda mais pesada a esquerda.

É comum fazer um gráfico do  $k^{\text{ésimo}}$  momento em função do índice  $k$  como mostrado na Figura 9. Assim, podemos visualmente contrastar o momento normalizado obtido dos dados experimentais para as quatro distribuidoras.

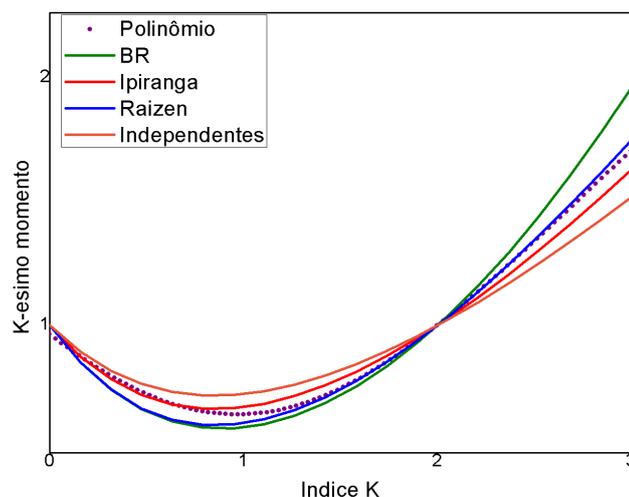


Figura 9 – Momento normalizado em função do índice  $k$  do conjunto de dados do *Market Share* no período de 2014 a 2018. A curva composta de círculos corresponde ao polinômio de grau 3  $b_3x^3 + b_2x^2 + b_1x + c$  ajustado a média dos dados do *Market Share*, sendo  $b_1 = -0,404$ ,  $b_2 = 0,209$ ,  $b_3 = -6,99 \cdot 10^{-4}$  e  $c = 0,975$ .

### 3.1.7 Probabilidade Cumulativa

A função de densidade de probabilidade para uma variável aleatória  $X$  é dada por:

$$\lim_{\Delta \rightarrow 0} F(X)\Delta = \lim_{\Delta \rightarrow 0} Pr(X \leq x \leq X + \Delta)$$

Assim, a probabilidade de se obter um valor dentro de um intervalo  $[a, b]$  é dado por:

$$Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

Na teoria de probabilidade e estatística, a função de distribuição cumulativa (*cumulative distribution function* - CDF) de uma variável aleatória de valor real  $X$  é a probabilidade de que  $X$  terá um valor menor ou igual a  $x$ . Essa função de distribuição cumulativa pode ser visualizada na Figura 10 e é dada por:

$$\frac{d}{dx}F(x) = f(x)$$

$$F(x) = \int_{-\infty}^x f(y)dy$$

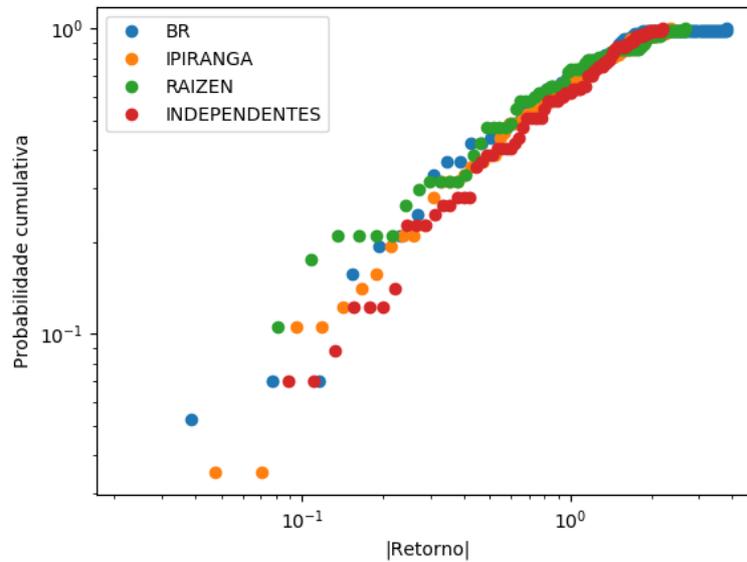


Figura 10 – Distribuição cumulativa do conjunto de dados do *Market Share* no período de 2014 a 2018.

Essa função de distribuição cumulativa nos indica que a Raízen possui a cauda menos longa de retornos, enquanto as independentes possuem a cauda mais longa. A Ipiranga está colada nas independentes e a BR está entre elas e a Raízen.

### 3.2 Conclusões da Análise Estatística

A análise estatística do *Market Share* nos permitiu entender os dados que possuímos e realizar as seguintes observações: *i*) há uma periodicidade sazonal no *Market Share*; *ii*) representam um mercado não eficiente; *iii*) Raízen possui a cauda menos longa de retornos, enquanto as independentes possuem a cauda mais longa; *iv*) BR, Ipiranga e Raízen formam um *cluster* separado das independentes e Raízen e BR formam um *subcluster*; *v*) a correlação entre independentes e as outras distribuidoras é sempre um alto valor negativo.

## 4 FUNDAMENTAÇÃO TEÓRICA

Esse capítulo tem como objetivo realizar a revisão teórica do Modelo Baseado em Agentes, o qual será a principal ferramenta utilizada para desenvolver a segunda e terceira etapa proposta para esse trabalho: simular o modelo e verificar se as propriedades da série que esse modelo gera são compatíveis com as séries reais.

Na simulação via modelo baseado em agentes será utilizado o modelo de Vicsek. Dados reais, como o *Market Share* fornecem uma evidência clara de uma transição de fase em direção ao movimento coletivo associado ao alinhamento local das estratégias em busca do aumento do *Market Share* semelhantes ao modelo de Vicsek. Uma teoria macroscópica baseada em modelos hidrodinâmicos auto-organizados previamente derivados é adaptada a esse contexto e fornece previsões consistentes com o movimento observado.

Importante ressaltar que a simulação foi realizada utilizando processo de Wiener que é um caso particular do processo markoviano no qual considera que a única informação necessária para determinar o estado futuro é o estado atual (não é necessário a memória do restante do processo). Entretanto, a mais adequada para essa simulação seria o processo de Lévy, porém, a simulação com esse processo iria exigir uma física mais avançada.

### 4.1 Modelo Baseado em Agentes

O modelo baseado em agentes (ABM) é um método de simulação para modelar sistemas dinâmicos, adaptativos e autônomos [9]. Os agentes são sociais, interagem com os outros, vivem em um ambiente e suas próximas ações são baseadas no estado atual do ambiente. Os agentes podem ter objetivos explícitos para maximizar ou minimizar e podem aprender e se adaptar com base na experiência [10].

Epstein e Axtell (1996) [11] caracterizam o ABM como um sistema dinâmico temporalmente discreto, descrito por equações genéricas simultâneas:

$$Agente_{t+1} = f(Agente_t, Meio_t) \quad (4.1)$$

$$Meio_{t+1} = g(Agente_t, Meio_t) \quad (4.2)$$

A notação das Equações 4.1 e 4.2 nos permite capturar as interações entre os agentes, a evolução dessa interação entre os agentes e o ambiente, ou simplesmente as transformações autônomas dos agentes e do ambiente [12]. Nesse contexto, o ABM é adequado para problemas do mundo real onde há vários agentes autônomos, heterogêneos

e interativos [13], o qual permite investigar como as interações entre esses agentes, e entre agentes e outros fatores externos se somam para formar os padrões vistos no mundo real [14].

Fundamentalmente, um sistema multiagente compreende um conjunto de agentes, no nosso caso as distribuidoras de combustíveis, e seu ambiente, os postos e o mercado em que essas distribuidoras se inserem, no qual os agentes são projetados para executar tarefas específicas - aumento do *Market Share*. A esse respeito, exige-se que os agentes individuais sejam autônomos, o que significa capacidade de aprendizado por meio de interações com o ambiente, bem como adaptação às mudanças no ambiente causadas pelas ações dos agentes internamente e pela dinâmica externa dos ambientes.

O trabalho de Wooldridge e Jennings (1995) mostra algumas propriedades que são essenciais para uma melhor caracterização do agente, como:

- **Autonomia:** característica que o agente inteligente deve ter para tomar decisões e ações importantes para a conclusão de uma tarefa ou objetivo sem a necessidade da interferência do ser humano ou qualquer outra entidade. Ser capaz de agir independentemente com seu ambiente através de seus próprios “sensores” ou com as suas próprias percepções com o objetivo de realizar alguma tarefa seja ela externa ou gerada por ele próprio;
- **Proatividade:** capacidade que o agente deve ter de tomar iniciativas. Eles não respondem simplesmente de acordo com o meio. Têm a capacidade de exibir comportamentos baseados em objetivos;
- **Reatividade:** capacidade de reagir rapidamente a alterações no ambiente. Percebe o meio e responde de modo oportuno;
- **Robustez:** ser capaz de tomar decisões fundamentadas em informações incompletas ou escassas, lidar com erros e ter uma capacidade de adaptação ou aprendizagem através da experiência;
- **Habilidade de comunicação:** capacidade de comunicação com repositórios de informações. Seja apenas repositório de dados, seja outro agente, ou seja, no próprio ambiente é fundamental uma constante troca de informações;
- **Raciocínio:** é talvez o aspecto mais importante que distingue um agente inteligente dos outros agentes. Afirmar que um agente tem raciocínio significa dizer que ele tem a capacidade de analisar e inferir baseado em seu conhecimento atual e nas suas experiências.

## 4.2 Modelo de Vicsek

Movimento coletivo é um fenômeno que ocorre em coleções de agentes, similares ou não, que interagem entre si, e movem-se com aproximadamente a mesma velocidade. Essas interações podem ser simples, baseadas em atração e repulsão, ou mais complexas, consistindo em combinações dessas, e se dando entre vizinhos no espaço ou em uma rede subjacente. Esse movimento é um fenômeno onipresente, observado em uma ampla gama de diferentes sistemas vivos e em uma escala ainda mais ampla, cardume de peixes, rebanhos de mamíferos, colônias de bactérias e migrações celulares, até o comportamento cooperativo de motores moleculares e biopolímeros no nível subcelular [15].

São basicamente dois os tipos de sistemas de comportamento coletivo: aqueles em que os agentes são autopropulsores [16], que é o caso de sistemas biológicos, e os sistemas em equilíbrio [17], caracterizados por grandes quantidades de partículas e bem descritos pela física estatística tradicional. No primeiro caso, ao considerar um sistema consistindo exclusivamente de partículas autopropulsoras, sem incluir as mudanças no ambiente, toda vez que uma partícula interage com outra, seu momento linear é alterado. Em contrapartida, no caso dos sistemas em equilíbrio, a distribuição de velocidades é preservada. Esses dois tipos de sistemas, apesar de suas diferenças, apresentam características similares em relação às mudanças de estado coletivo/macroscópico.

Em 1995 o físico húngaro Tamás Vicsek propôs um modelo para a flocação de partículas auto dirigidas [18]. Neste modelo existe apenas conservação de número de partículas dado que nenhum agente nasce ou morre durante a simulação da evolução do sistema. O momento, contudo, não é conservado e os agentes se movem em um meio dissipativo. Logo, como consequência da falta de conservação do momento, ele viola o princípio de invariância de Galileu. Este modelo exhibe auto-organização em várias escalas. [19]. No modelo em estudo, onde estamos trabalhando com as distribuidoras e suas estratégias de negócio, a escala que nos interessa é a macroscópica, a qual é descrita pela equação hidrodinâmica.

O sucesso do Modelo de Vicsek vem do fato de ele ser um modelo mínimo. Utilizando-se apenas de partículas auto-propelentes sujeitas a interações simples e locais é possível observar uma transição de fase de movimento desordenado para movimento coletivo ordenado. A ideia central no Modelo de Vicsek é que a cada passo do tempo, uma partícula se move na direção média do movimento de seus vizinhos, com algum ruído, não havendo a necessidade de um líder. Entretanto, as partículas cometem “erros” ao avaliar a direção do movimento de seus vizinhos, esses erros são considerados como sendo o ruído. Existem várias fontes de ruídos, mas eles podem ser divididos em dois grupos:

- Ruído Extrínseco: Tal como o ambiente. Cada empresa não vê muito bem a direção em que seus vizinhos estão se movendo. Esse ruído é chamado extrínseco porque tem a

ver com incertezas no mecanismo de estratégia e de "comunicação" empresa-empresa;

- Ruído Intrínseco: As empresas possuem livre-arbítrio em que, mesmo quando conhecem a direção de seus vizinhos, decidem mover-se em uma direção diferente. É chamado de ruído intrínseco, porque tem a ver com o mecanismo interno de tomada de decisão das empresas.

O modelo Vicsek identifica três itens mínimos que são suficientes para explicar o surgimento do movimento direcionado. (i) Partículas sempre se movem e fazem isso com um passo de velocidade de módulo constante (auto-propelência); (ii) Uma partícula tende a alinhar o sentido da sua velocidade com o das suas vizinhas (ordem); (iii) Toda a partícula tem um ruído que tende a desalinhar o sentido, ou que dificulta o alinhamento, da sua velocidade com o das suas vizinhas (desordem).

Para entender o modelo de Vicsek em um primeiro momento devemos revisar a Equação de Transporte de Boltzmann e encontrar seus primeiros momentos de forma a chegar na equação hidrodinâmica de Vicsek.

#### 4.2.1 Equação de Transporte de Boltzmann

A teoria desenvolvida para a equação de transporte de Boltzmann faz parte de uma das mais importantes áreas de equações diferenciais parciais, abrangendo teoria cinética, mecânica dos fluidos e mecânica estatística. Ela descreve o comportamento estatístico de um sistema termodinâmico que não está em estado de equilíbrio e foi desenvolvida no século XIX por Ludwig Boltzmann, um dos pais da mecânica estatística, em seus estudos em teoria cinética dos gases [20]. Tendo a função de descrever estatisticamente a distribuição de partículas em um gás, a equação de transporte de Boltzmann é utilizada no estudo de como um gás transporta determinadas quantidades físicas, como calor e massa.

O exemplo clássico desse sistema é um fluido com gradientes de temperatura no espaço, fazendo o calor fluir das regiões mais quentes para as mais frias, pelo transporte aleatório, mas tendencioso, das partículas que compõem esse fluido. Na literatura moderna, o termo equação de Boltzmann é frequentemente usado em um sentido mais geral, referindo-se a qualquer equação cinética que descreva a alteração de uma quantidade macroscópica em um sistema termodinâmico, como energia, carga ou número de partículas.

A equação surge não analisando as posições individuais e o momento de cada partícula no fluido, mas considerando uma distribuição de probabilidade para a posição e o momento de uma partícula típica - ou seja, a probabilidade de a partícula ocupar uma dada região muito pequena do espaço (matematicamente, o elemento de volume  $d^3\mathbf{r}$ ) centralizada na posição  $\mathbf{r}$  e possuir momento quase igual a um determinado vetor de momento  $\mathbf{p}$  (ocupando, assim, uma região muito pequena de espaço de momento  $d^3\mathbf{p}$ ,

em um instante de tempo). A equação é uma equação integro-diferencial não linear e a função desconhecida na equação é uma função de densidade de probabilidade de posição e de momento de partícula [21].

O conjunto de todas as posições possíveis  $\mathbf{r}$  e momento  $\mathbf{p}$  é chamado de espaço de fase do sistema. Todo o espaço é 6N-dimensional: um ponto nesse espaço é  $(\mathbf{r}, \mathbf{p}) = (x, y, z, p_x, p_y, p_z)$ , e cada coordenada é parametrizada pelo tempo  $t$ . O pequeno volume (elemento diferencial de volume) é escrito como:

$$d^3\mathbf{r}d^3\mathbf{p} = dx dy dz dp_x dp_y dp_z \quad (4.3)$$

Como a probabilidade de  $N$  moléculas que possuem todos  $\mathbf{r}$  e  $\mathbf{p}$  dentro de  $d^3\mathbf{r}d^3\mathbf{p}$  está em questão, na equação há uma quantidade  $f$  que fornece essa probabilidade por unidade de volume espaço-fase, no instante do tempo  $t$ . Esta é uma função de densidade de probabilidade:  $f(\mathbf{r}, \mathbf{p}, t)$ , definida de modo que:

$$dN = f(\mathbf{r}, \mathbf{p}, t)d^3\mathbf{r}d^3\mathbf{p} \quad (4.4)$$

Onde  $dN$  é o número de moléculas que possuem posições dentro de um elemento de volume  $d^3\mathbf{r}$  sobre  $\mathbf{r}$  e momento localizado em um elemento de espaço de momento  $d^3\mathbf{p}$  sobre  $\mathbf{p}$ , em um tempo  $t$ . A integração sobre uma região de espaço de posição e espaço de momento fornece o número total de partículas que possuem posições e momentos nessa região:

$$N = \int_{\text{momento}} d^3\mathbf{p} \int_{\text{posição}} d^3\mathbf{r} f(\mathbf{r}, \mathbf{p}, t) \quad (4.5)$$

Supõe-se que as partículas no sistema sejam idênticas (portanto, cada uma tem uma massa idêntica  $m$ ). Para um conjunto de mais de uma espécie química, é necessária uma distribuição para cada uma, conforme a seguinte equação:

$$\frac{df}{dt} = \left( \frac{\partial f}{\partial t} \right)_{\text{força}} + \left( \frac{\partial f}{\partial t} \right)_{\text{dif}} + \left( \frac{\partial f}{\partial t} \right)_{\text{col}}, \quad (4.6)$$

onde o termo "*força*" corresponde às forças exercidas sobre as partículas por uma influência externa, o termo "*dif*" representa a difusão das partículas e "*col*" é o termo de colisão - responsável pelas forças que atuam entre as partículas nas colisões.

Considere as partículas descritas por  $f$ , cada uma experimentando uma força externa  $F$ . Suponha que no momento  $t$  algum número de partículas tenha posição  $\mathbf{r}$  dentro do elemento  $d^3\mathbf{r}$  e momento  $\mathbf{p}$  dentro de  $d^3\mathbf{p}$ . Se uma força  $\mathbf{F}$  atua instantaneamente

em cada partícula, em seguida, no tempo  $t + \Delta t$  será a sua posição  $\mathbf{r} + \Delta\mathbf{r} = \mathbf{r} + \mathbf{p}\frac{\Delta t}{m}$  e impulso  $\mathbf{p} + \Delta\mathbf{p} = \mathbf{p} + \mathbf{F}\Delta t$ . Então, na ausência de colisões,  $f$  deve satisfazer:

$$f\left(\mathbf{r} + \frac{\mathbf{p}}{m}\Delta t, \mathbf{p} + \mathbf{F}\Delta t, t + \Delta t\right) d^3\mathbf{r}d^3\mathbf{p} = f(\mathbf{r}, \mathbf{p}, t)d^3\mathbf{r}d^3\mathbf{p} \quad (4.7)$$

Na Equação 4.7 foi utilizado o fato de que o elemento de volume do espaço de fase  $d^3\mathbf{r}d^3\mathbf{p}$  é constante, o que pode ser mostrado usando as equações de Hamilton (Teorema de Liouville). No entanto, como ocorrem colisões, a densidade de partículas no volume do espaço de fase  $d^3\mathbf{r}d^3\mathbf{p}$  muda, então

$$dN_{col} = \left(\frac{\partial f}{\partial t}\right)_{col} \Delta t d^3\mathbf{r}d^3\mathbf{p} \quad (4.8)$$

$$dN_{col} = f\left(\mathbf{r} + \frac{\mathbf{p}}{m}\Delta t, \mathbf{p} + \mathbf{F}\Delta t, t + \Delta t\right) d^3\mathbf{r}d^3\mathbf{p} - f(\mathbf{r}, \mathbf{p}, t)d^3\mathbf{r}d^3\mathbf{p} \quad (4.9)$$

$$dN_{col} = \Delta f d^3\mathbf{r}d^3\mathbf{p} \quad (4.10)$$

onde  $\Delta f$  é a variação total em  $f$ . Dividindo 4.10 por  $d^3\mathbf{r}d^3\mathbf{p}\Delta t$  e tomando os limites  $\Delta t \rightarrow 0$  e  $\Delta f \rightarrow 0$ , temos:

$$\left(\frac{df}{dt}\right) = \left(\frac{\partial f}{\partial t}\right)_{col} \quad (4.11)$$

O diferencial total de  $f$  é:

$$df = \frac{\partial f}{\partial t} dt + \left(\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz\right) + \left(\frac{\partial f}{\partial p_x} dp_x + \frac{\partial f}{\partial p_y} dp_y + \frac{\partial f}{\partial p_z} dp_z\right) \quad (4.12)$$

$$df = \frac{\partial f}{\partial t} dt + \nabla f \cdot \frac{\mathbf{p}}{m} dt + \frac{\partial f}{\partial \mathbf{p}} \cdot \mathbf{F} dt \quad (4.13)$$

Assim, obtemos a seguinte equação:

$$\frac{\partial f}{\partial t} + \nabla f \cdot \frac{\mathbf{p}}{m} + \frac{\partial f}{\partial \mathbf{p}} \cdot \mathbf{F} = \left(\frac{\partial f}{\partial t}\right)_{col} \quad (4.14)$$

Nesse contexto,  $F(\mathbf{r}, t)$  é o campo de força que atua sobre as partículas no fluido e  $m$  é a massa das partículas. O termo no lado direito é adicionado para descrever o efeito de colisões entre partículas; se for zero, as partículas não colidem. A equação de Boltzmann sem colisão, onde colisões individuais são substituídas por interações agregadas de longo alcance, por exemplo, interações de Coulomb, é frequentemente chamada de equação de Vlasov.

A Equação 4.14 pode ser escrita utilizando a velocidade das partículas  $\mathbf{v}$  ao invés do momento  $\mathbf{p}$ ; eles estão relacionados na definição de momento por  $\mathbf{p} = m\mathbf{v}$ .

$$\frac{\partial f}{\partial t} + \nabla_{\mathbf{r}} f \cdot \mathbf{v} + \nabla_{\mathbf{v}} f \cdot \mathbf{F}_m = \left( \frac{\partial f}{\partial t} \right)_{col} \quad (4.15)$$

Não é necessário resolver a equação de Boltzmann para a função de distribuição, a fim de determinar as variáveis macroscópicas de interesse físico. As equações diferenciais que governam as variações temporais e espaciais dessas variáveis macroscópicas podem ser derivadas diretamente da equação de Boltzmann sem resolvê-la. Essas equações diferenciais são conhecidas como equações de transporte macroscópicas e suas soluções, sob certas suposições, fornecem diretamente as variáveis macroscópicas. As variáveis macroscópicas estão relacionadas aos momentos da função de distribuição e as equações de transporte satisfeitas por essas variáveis podem ser obtidas tomando-se os vários momentos da equação de Boltzmann.

Os três primeiros momentos da equação de Boltzmann obtidos pela multiplicação por  $m_\alpha$ ,  $m_\alpha v$  e  $m_\alpha \frac{v^2}{2}$ , respectivamente, e integrando em todo o espaço de velocidade, nos dão a equação de conservação de massa, a equação de conservação de momento e a equação de conservação de energia. Para o presente trabalho se faz necessário obter esses momentos.

#### 4.2.2 Modelo Macroscópico - Limite Hidrodinâmico

O momento zero da equação de Boltzmann no espaço recíproco é dado por:

$$m \int d\mathbf{v} \frac{\partial f}{\partial t} + m \int d\mathbf{v} \mathbf{v} \cdot \nabla f + m \int d\mathbf{v} \mathbf{F}_m \cdot \nabla_{\mathbf{v}} f = m \int d\mathbf{v} \left( \frac{df}{dt} \right)_{col} \quad (4.16)$$

O termo de colisão contém termos de espalhamento para dentro e para fora deste volume, logo sua integral é nula. A primeira integral no lado esquerdo da equação é conhecida como termo transiente e produz:

$$\int d\mathbf{v} \frac{\partial f(\mathbf{r}, \mathbf{v}, t)}{\partial t} = \frac{\partial}{\partial t} \int d\mathbf{v} f(\mathbf{r}, \mathbf{v}, t) = \frac{\partial}{\partial t} \rho(\mathbf{r}, t) \quad (4.17)$$

onde  $\rho(\mathbf{r}, t)$  é a densidade das partículas. A segunda integral do lado esquerdo da equação é conhecida como termo de difusão e produz:

$$\int d\mathbf{v} \mathbf{v} \cdot \nabla f(\mathbf{r}, \mathbf{v}, t) = \nabla \cdot \int d\mathbf{v} \mathbf{v} f(\mathbf{r}, \mathbf{v}, t) = \nabla \cdot (\bar{v} \rho(\mathbf{r}, t)) \quad (4.18)$$

Já a integral restante representa um termo de deriva e produz:

$$\int d\mathbf{v} \mathbf{F}_m \cdot \nabla_{\mathbf{v}} f(\mathbf{r}, \mathbf{v}, t) = \int d\mathbf{k} \nabla_{\mathbf{v}} (\mathbf{F}_m f(\mathbf{r}, \mathbf{v}, t)) = \oint \oint_s f(\mathbf{r}, \mathbf{v}, t) \mathbf{F}_m \cdot d\mathbf{s} = -Q(\mathbf{r}, t), \quad (4.19)$$

onde usamos o teorema da divergência. Assim, o momento zero da equação de Boltzmann, também conhecido como Equação da Continuidade, é dado por:

$$\frac{\partial \rho}{\partial t}(\mathbf{r}, t) + \nabla \cdot (\bar{\mathbf{v}}\rho(\mathbf{r}, t)) = Q(\mathbf{r}, t) \quad (4.20)$$

Na Equação da Continuidade [4.20](#) ficamos com a função  $Q(\mathbf{r}, t)$ , a qual se faz necessário encontrar analiticamente para obter o primeiro momento, porém, não é trivial obter o mesmo, pois os agentes possuem mais graus de liberdade, não só posição e tempo, temos também as estratégias, por exemplo. Dessa equação da continuidade tiramos que possuímos a conservação do número de agentes (BR, Ipiranga, Raízen e Independentes) e do *Market Share* (a soma sempre dará 1), os quais entrariam na função  $Q(\mathbf{r}, t)$ .

## 5 DESENVOLVIMENTO

### 5.1 Simulação

Na simulação da competitividade entre as distribuidoras de combustíveis no RS via Modelo de Vicsek os agentes são as distribuidoras e as estratégias utilizadas são os fatores que geram o caminho a ser seguido em busca do aumento do *Market Share*. A atualização das estratégias adotadas pelas distribuidoras pode ser feita primeiramente considerando o ruído intrínseco onde o novo ângulo de orientação das distribuidoras é formado pela contribuição das demais e uma certa indecisão em seguir esta média é adicionada.

Para a simulação inicial utilizamos  $\xi$  (nível de ajuste) = 0,1,  $\eta$  (nível de ruído) = 0,001, 4 agentes (BR, Ipiranga, Raízen e Independentes), 5 estratégias (preço, excelência operacional, loja de conveniência, tecnologia e produto diferenciado) e 5.200 interações. Por meio dessa simulação obtemos a informação normalizada de o quão próximo a distribuidora está do consumidor (essa distância que a empresa está do consumidor é definida como prêmio durante a simulação). Assim, se as estratégias da empresa estão em linha com as do consumidor, o prêmio vai a zero; se as estratégias da empresa não estão em linha com a do consumidor, o prêmio aumenta (ou seja, o prêmio vai de 0 até infinito). Para que essa informação se reflita no *Market Share* foi utilizada uma função muito utilizada em *machine learning* que se chama *Softmax*.

A função *Softmax* é uma função que recebe como entrada um vetor de  $K$  números reais e o normaliza em uma distribuição de probabilidade que consiste em  $K$  probabilidades proporcionais às exponenciais dos números de entrada. Ou seja, antes da aplicação do *Softmax*, alguns componentes do vetor podem ser negativos ou maiores que 1, mas depois de aplicar o *Softmax* cada componente estará no intervalo  $(0, 1)$ , e os componentes adicionarão até 1 para que possam ser interpretados como probabilidades. Além disso, os componentes de entrada maiores corresponderão a probabilidades maiores. O *Softmax* é frequentemente usado em redes neurais para mapear a saída não normalizada de uma rede para uma distribuição de probabilidade sobre as classes de saída previstas. A função é definida pela fórmula da Equação 5.1 para  $i = 1, \dots, K$  e  $\mathbf{z} = (z_1, \dots, z_K) \in R^K$ . Em palavras: aplicamos a função exponencial padrão a cada elemento  $z_i$  do vetor de entrada  $\mathbf{z}$  e normalizamos esses valores dividindo pela soma de todas essas exponenciais; essa normalização garante que a soma dos componentes do vetor de saída  $\sigma(\mathbf{z})$  seja 1. Aplicando a equação a simulação realizada  $\mathbf{z}$  é o prêmio e  $\sigma(\mathbf{z})$  é o *Market Share*.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (5.1)$$

Temos, portanto, a função de mapeamento  $Market\ Share = e^{-prêmio}$  para que a variação do *Market Share* seja entre 0 e 1. Logo, se o prêmio for 0, a distribuidora tem a máxima fatia do mercado, mas quanto mais ela se afasta do consumidor, menor a sua fatia do mercado. A simulação é finalizada com a normalização do *Market Share*; logo, o *share* de cada uma vai de 0% até 100%, mas a soma destes é igual a 100%. A estratégia do consumidor definida inicialmente é escolhida aleatoriamente. Conforme ocorrem as interações cada empresa ajusta sua estratégia pela do vencedor. O resultado desta simulação pode ser verificado na Figura 11 o qual mostra o *Market Share* obtido. O código que gerou essa análise inicial foi feito com Linguagem Julia (possui um tempo de execução menor, é uma linguagem interativa e com sintaxe direta) e pode ser verificado a seguir:

```
# Constantes
xi = 0.1      # Nível de ajuste
eta = 0.001  # Nível Ruído
NOp = 5      # Número de opções/estratégias
NAg = 4      # Número de agentes: BR, Ipiranga, Raízen e Independentes
NIIt = 5200  # Número de interações

# 0 que o consumidor espera
# Escolhe aleatoriamente a estratégia que o consumidor espera e normaliza a mesma
consumer = rand(NOp)
consumer /= norm(consumer)

# Inicializa empresas (agentes)
# Direciona as estratégias utilizadas por cada agente
Agent = []
for a in [1:NAg...]
    global Agent

    # Estratégia deve ser normalizada
    Ag = rand(NOp)
    Ag /= norm(Ag)
    if a == 1
        Agent = Ag
    else
        Agent = hcat(Agent, Ag)
    end
end
end
```

```

# Joga
y = []
for it in [1:NIt...]
    global y

    # Calcula quão próximo empresa está do consumidor
    prêmio = [norm(Agent[:,ag]-consumer)/norm(consumer) for ag in [1:NAg...]]
    # Encontra marketshare = e^{-p}
    t = [exp(-p) for p in prêmio]
    # Normaliza
    S = sum(t)
    mshare = [100*m/S for m in t]

    # Encontra o vencedor com maior marketshare
    pos = sortperm(mshare,rev=true)
    top = Agent[:,pos[1]]

    # Cada empresa ajusta sua estratégia pela do vencedor
    for ag in [1:NAg...]
        Ag = Agent[:,ag]
        Ag = xi*top + (1-xi)*Ag
        Ag = eta*rand(NOp,1) + (1-eta)*Ag
        Ag /= norm(Ag)

        for n in [1:NOp...]
            Agent[n,ag] = Ag[n]
        end
    end

    end

    if it == 200
        y = mshare
    elseif it > 200
        y = hcat(y,mshare)
    end
end
end

```

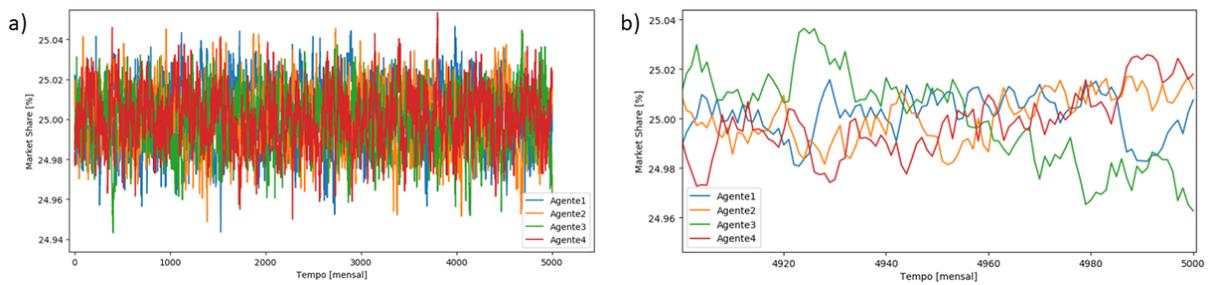


Figura 11 – Simulação do *Market Share* com os seguintes fatores:  $\xi$  (nível de ajuste) = 0,1,  $\eta$  (nível de ruído) = 0,001, 4 agentes (BR, Ipiranga, Raízen e Independentes), 5 estratégias para a) 5200 interações b) zoom nas interações finais.

Na Figura 12 podemos verificar diferentes configurações do *Market Share* para diferentes valores de ruído verificando a influência do mesmo sobre o resultado da simulação.

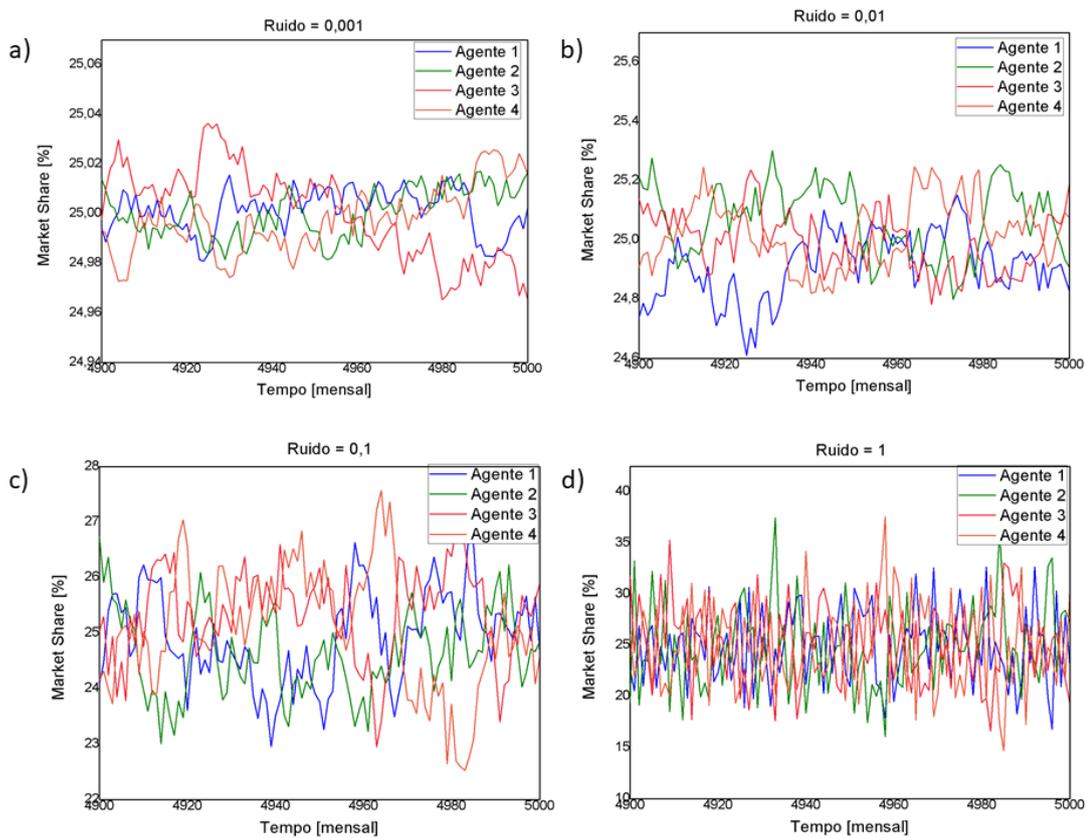


Figura 12 – Simulação do *Market Share* para diferentes níveis de ruído a)  $\eta = 0,001$ , b)  $\eta = 0,01$ , c)  $\eta = 0,1$ , d)  $\eta = 1,0$

Após a simulação inicial, verificou-se a necessidade de ajustes, uma vez que nem todos os agentes fazem uso de todas as estratégias. A exemplo das independentes que

possuem um aumento exponencial do *Market Share* e normalmente priorizam apenas a estratégia de manter o melhor preço do mercado. Para realizar esse ajuste foi colocado o último agente com uma estratégia viciada, isto é, ele inicia a simulação com uma estratégia e mantém a mesma durante todo o tempo. Além disso, os outros agentes recebem a informação incompleta, pois na interação em que ocorre o ajuste das estratégias, não é incluído esse agente com estratégia viciada. Para isso foi incluído no código a seguinte condição:

```

for ag in [1:NAg-1...]
    Ag = Agent[:,ag]
    Ag = xi*top + (1-xi)*Ag
    Ag = eta*rand(NOp,1) + (1-eta)*Ag
    Ag /= norm(Ag)

    for n in [1:NOp...]
        Agent[n,ag] = Ag[n]
    end
end
end

```

Porém, conforme podemos verificar pela Figura 13, apenas viciar um agente em uma estratégia não é o suficiente, pois os três agentes que interagiram ficaram com *Market Share* muito próximo (linhas sobrepostas na Figura 13) e o outro agente ficou muito distante.

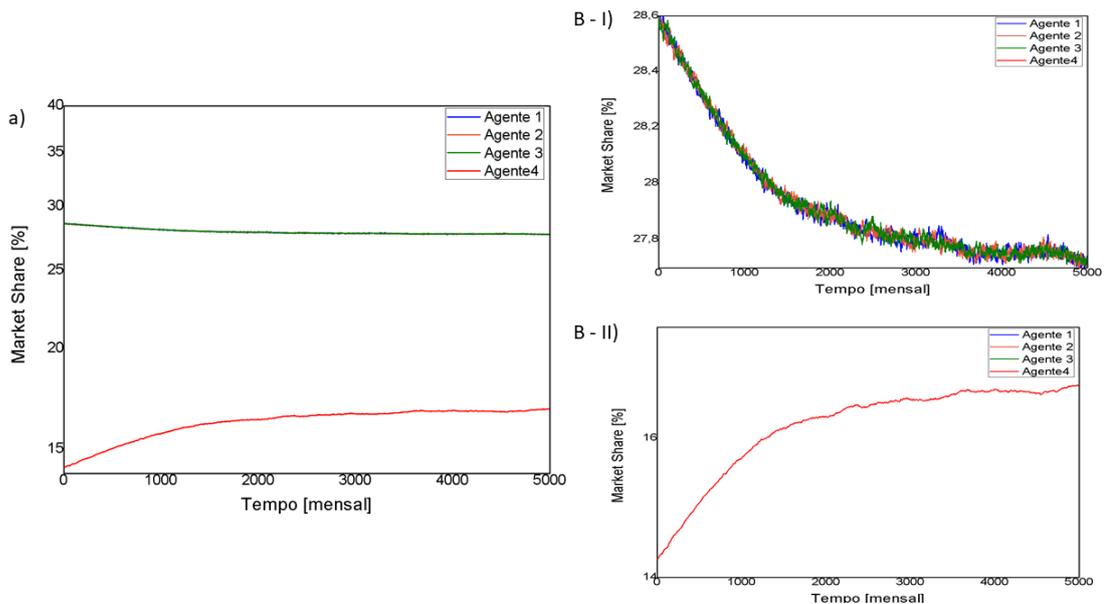


Figura 13 – Simulação do Market Share viciando um agente em uma estratégia para a) 5200 interações; b-I e b-II) zoom de um intervalo das interações.

A hipótese de que não se sabe o que o consumidor quer é invalidada; a conclusão é que o cliente pondera as estratégias de forma diferente. Se um agente tem maior *Market Share* que os outros, isto implica que existe uma estratégia que aproxima ele do interesse do consumidor. Para implementar isso no código, o ajuste das estratégias do último agente foi modificado para que suas estratégias se ajustem de modo a seguir a estratégia do consumidor (obter o menor prêmio), enquanto o ajuste das outras empresas se mantém baseado pela estratégia do primeiro colocado. O resultado obtido pode ser visualizado na Figura 14

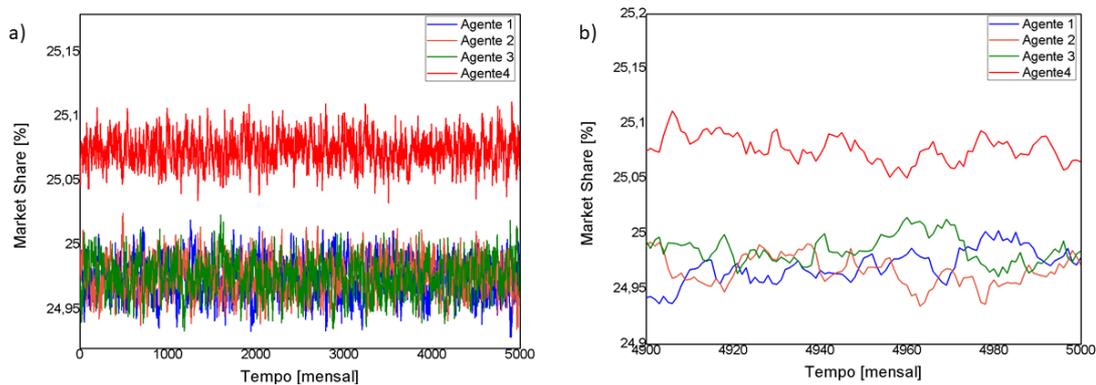


Figura 14 – Simulação do Market Share ponderando as estratégias para a) 5200 interações; b) zoom de um intervalo das interações.

Para essa simulação, a etapa do código em que cada empresa ajusta sua estratégia pela do vencedor foi alterada com a seguinte condição:

```

for ag in [1:NAg...]
    Ag = Agent[:,ag]
    if ag != NAg
        Ag = xi*top + (1-xi)*Ag
    else
        Ag = xi*consumer + (1-xi)*Ag
    end
    Ag = eta*rand(NOp,1) + (1-eta)*Ag
    Ag /= norm(Ag)

    for n in [1:NOp...]
        Agent[n,ag] = Ag[n]
    end
end
end

```

Como último ajuste da simulação os parâmetros  $\xi_n$  foram simulados de modo a ditar o perfil de negócio de cada agente, uma das distribuidoras sendo mais conservadora (Raízen), outra mais agressiva (independentes) e duas delas seguindo o modelo de Vicsek (BR e Ipiranga). Para isso foi utilizada a seguinte condição inicial, onde Raízen recebe a estratégia inicial randomicamente; Independentes recebe uma parte randomicamente e outra mais direcionada ao consumidor; e BR e Ipiranga recebem a estratégia inicial direcionada ao consumidor:

```
Raizen = 0.9*rand(NOp)
Indep = 0.9*rand(NOp) + 0.1*consumer
Br = 0.4*consumer
Ipiranga = 0.6*consumer
Agent = [Raizen Indep Br Ipiranga]
```

Para definir o perfil de negócio de cada agente foi utilizado três níveis de ajustes distintos:  $\xi_1$  é a tendência de seguir o líder,  $\xi_2$  de seguir o consumidor e  $\xi_3$  de seguir a média (Vicsek). Os parâmetros utilizados estão indicados no código a seguir:

```
for ag in [1:NAg...]
    Ag = Agent[:,ag]
    # Raizen: Conservadora
    if ag == 1
        xi1 = 0.1
        xi2 = 0.01
        xi3 = 0
    end
    # Independentes: Agressiva
    if ag == 2
        xi1 = 0.01
        xi2 = 0.1
        xi3 = 0
    end
    # Br: Vicsek
    if ag == 3
        xi1 = 0.01
        xi2 = 0.01
        xi3 = 0.5
    end
    # Ipiranga: Vicsek
    if ag == 4
```

```

xi1 = 0.01
xi2 = 0.01
xi3 = 0.5
end

Ag = xi1*top + xi2*consumer + xi3*avg + (1-xi1-xi2-xi3)*Ag
Ag = eta*rand(NOp,1) + (1-eta)*Ag # Informação incompleta.
Ag /= norm(Ag)

```

O resultado obtido pode ser verificado na Figura 15 e na Figura 16 é destacado o zoom das principais partes: a) interpolação dos pontos iniciais para análise estatística da reprodução do *Market Share*, b) zoom do momento em que independentes supera o *Market Share* da Ipiranga e da BR, c) projeção para os próximos anos, onde é possível observar uma transição de fase de movimento desordenado para movimento coletivo ordenado, tendo a Raízen uma estratégia mais conservadora, d) resultado final onde temos um dos agentes, as independentes, agindo mais agressivamente.

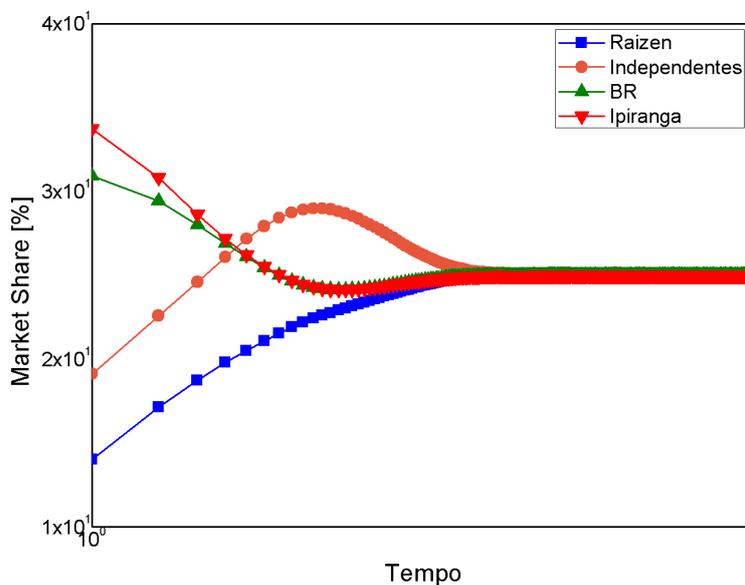


Figura 15 – Simulação do Market Share ditando o perfil de negócio de cada agente.

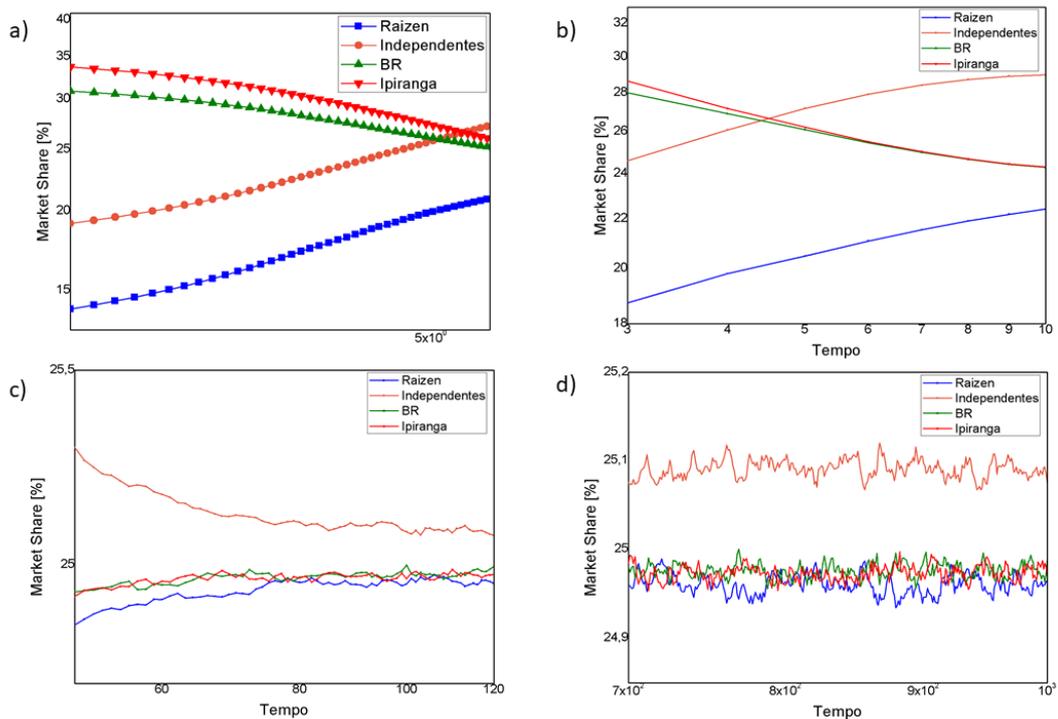


Figura 16 – a) Interpolação dos pontos iniciais para análise estatística da reprodução do *Market Share*, b) zoom na etapa em que independentes supera o *Market Share* da Ipiranga e da BR, c) projeção para os próximos anos d) resultado final onde temos um agente agindo mais agressivamente.

## 5.2 Análise estatística da simulação

Após obtidos os dados da simulação do Market Share, faz-se necessário validar os mesmos por meio da análise estatística para verificar se os resultados da simulação estão em linha com os experimentais. Essa simulação realizada busca indicar o movimento do mercado para os próximos anos, portanto, para comparar estatisticamente as informações foi realizado interpolação com os dados das interações iniciais, até o momento que o *Market Share* simulado reflete o experimental: independentes obtendo um maior *Market Share* que a Ipiranga e a BR. Para essa etapa foi realizado a média dos dados de 50 simulações. O resultado real versus simulado pode ser verificado na Figura [17](#).

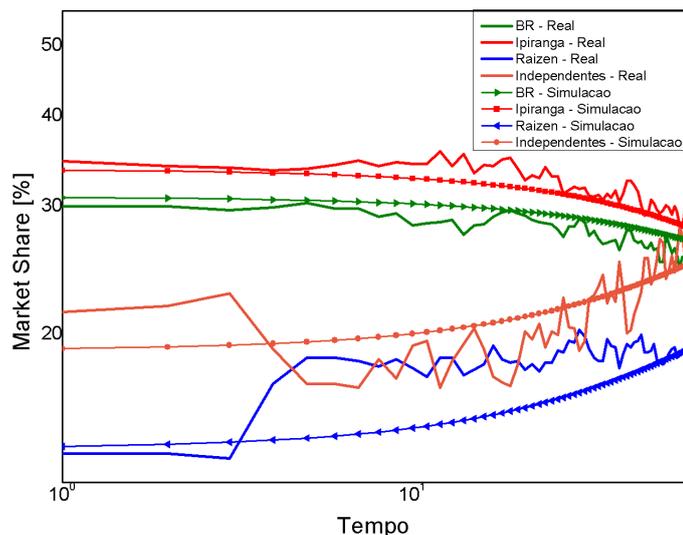


Figura 17 – Comparativo dos dados reais versus dados simulados do *Market Share*.

Na Tabela 4 estão apresentados os principais momentos e outros parâmetros estatísticos relevantes calculados. Por meio desses dados obtidos podemos verificar que em relação aos dados experimentais conseguimos obter valores muito próximos de média, de variância -BR e Ipiranga obtemos valores muito próximos; independentes, apesar da diferença, mantém-se como a distribuidora com maior variância-, de mediana e de obliquidade. A maior diferença encontra-se na curtose, pois a mesma possui relação com a cauda indicando o grau de assimetria da distribuição.

Tabela 4 – Parâmetros estatísticos do *Market Share* simulado.

Bandeira	Média	Variância	Mediana	Curtose	Obliquidade
BR	28,589	1,846	28,545	-1,231	-0,001
Ipiranga	30,716	3,151	30,681	-1,216	-0,035
Independentes	22,184	4,142	21,780	-1,391	0,236
Raízen	17,818	2,774	16,994	-1,227	-0,109

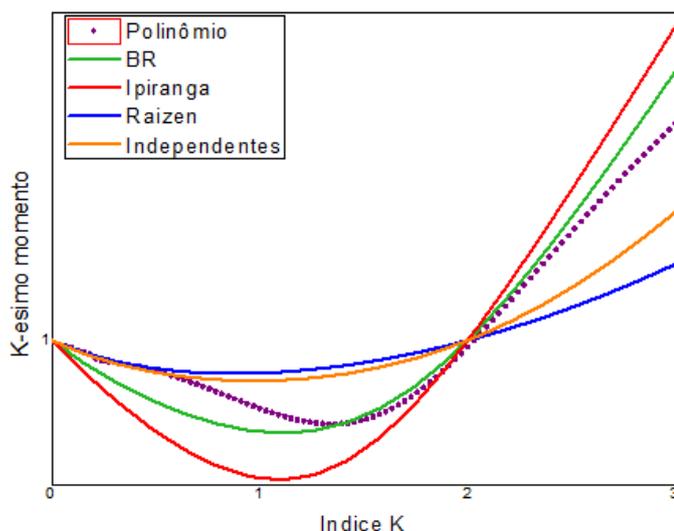


Figura 18 – Momento normalizado em função do índice  $k$  do conjunto de dados do *Market Share* simulado. A curva composta de círculos corresponde ao polinômio de grau 3  $b_3x^3 + b_2x^2 + b_1x + c$  ajustado a média dos dados simulados do *Market Share*, sendo  $b_1 = 0,261$ ,  $b_2 = -0,830$ ,  $b_3 = 0,362$  e  $c = 0,982$ .

Quando analisamos a Tabela 5 referente à matriz de correlação verificamos nossa maior diferença da simulação em relação aos dados reais, o qual acreditamos que deve ter ocorrido devido a simulação via processo de Wiener não considerar a memória, apenas o estado atual, assim como devido à simulação considerar um sistema em equilíbrio.

Tabela 5 – Matriz de Correlação dos dados simulados do *Market Share*.

	BR	IPIRANGA	RAÍZEN	INDEPENDENTES
BR	1			
IPIRANGA	0,440	1		
RAÍZEN	-0,704	-0,457	1	
INDEPENDENTES	-0,777	-0,569	0,982	1

Na Figura 19 podemos verificar a probabilidade cumulativa dos dados da simulação, a qual está com inclinação semelhante aos dados reais. O mesmo podemos verificar na clusterização presente na Figura 20, por meio do qual podemos verificar o *cluster* entre BR, Ipiranga e Raízen e um *subcluster* entre Raízen e BR.

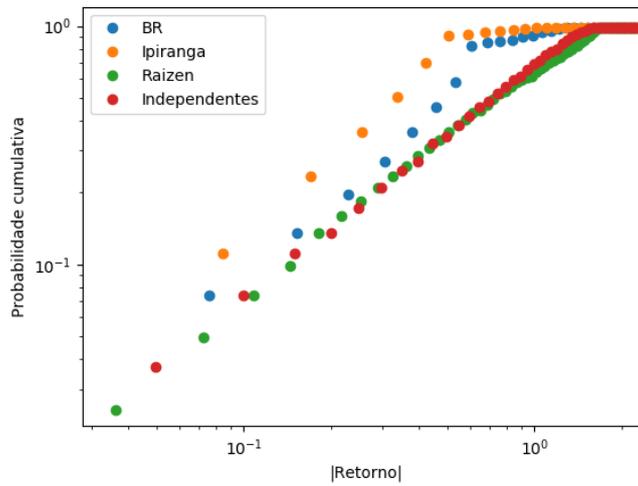


Figura 19 – Probabilidade cumulativa do conjunto de dados do Market Share.

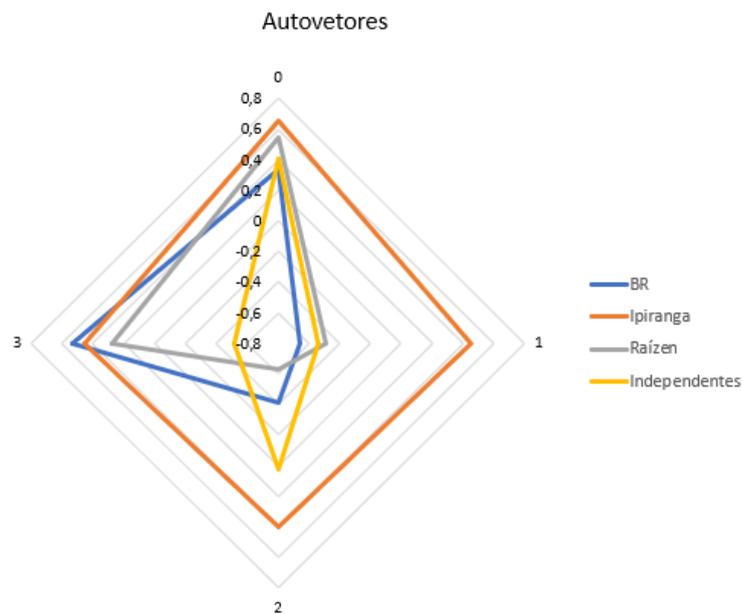


Figura 20 – Clusterização obtida com os dados simulados.

### 5.2.1 Conclusões da análise estatística

A análise estatística da simulação do *Market Share* nos permitiu validar as seguintes informações em relação aos dados experimentais:

- Plotando dados experimentais e simulados juntos verificamos grande semelhança no comportamento dos dados;

- Parâmetros estatísticos: valores de média muito semelhantes; maior variância é apresentada pelas independentes, as quais também apresentam uma distribuição com cauda mais pesada a direita;
- A probabilidade cumulativa apresenta declinação semelhante aos dados experimentais e nos indica que as independentes possuem a cauda mais longa. Apenas a Raízen apresentou um comportamento diferente em relação aos dados reais (probabilidade cumulativa com cauda mais longa em relação aos dados reais);
- BR, Ipiranga e Raízen formam um *cluster* separado das independentes;
- Há um alto grau de relação inversamente proporcional entre o *Market Share* das independentes e BR, Ipiranga, conforme verificado na matriz de correlação. Apenas não foi verificado coerência nos dados reais com os simulados da Raízen com as outras distribuidoras.

## 6 CONSIDERAÇÕES FINAIS

Conhecer as regras de funcionamento de movimentos coletivos na natureza é importante, pois elas podem ser aplicadas a diferentes mercados. Nesse sentido, a simulação realizada foi de grande relevância para dar início ao entendimento do complexo mercado em que as distribuidoras estão inseridas. Ao longo deste trabalho, verificamos grande semelhança entre o comportamento da série temporal real e simulada, assim como verificamos um conjunto de informações da análise estatística dos dados simulados que fecharam com os dados reais do *Market Share*, como média, variância, obliquidade, *clusterização*, correlação, probabilidade cumulativa, etc.

Quando realizamos a simulação seguindo o modelo de Vicsek todas tenderam a chegar na mesma direção após centenas de interações. Entretanto, quando estruturamos a simulação pensando no modelo de Vicsek é importante ressaltar que estamos usando um sistema que não considera a memória e também um sistema em equilíbrio. Portanto, para trabalhos futuros e aprimoramento dos resultados é de extrema importância considerar o modelo de Lévy ao invés de Wiener, pois o sistema simulado possui memória e não está em equilíbrio.

## REFERÊNCIAS

- [1] RUSSEL, S.; NORVING, P. *Artificial intelligence: a modern approach*. Norvig: Prentice Hall (2003).
- [2] <http://www.anp.gov.br/>. Acesso em 01 de Maio de 2019 às 21:12 h.
- [3] Cunha, C. R., Silva, R. *Relevant Stylized Facts About Bitcoin: Fluctuations, First Return Probability, and Natural Phenomena* Porto Alegre, RS (2019)
- [4] <https://www.mathstat.dal.ca/~stat5390/>. Acesso em 22 de Junho de 2019 às 19:02 h.
- [5] Sales, L. *Testando a hipótese de passeio aleatório no mercado de ações brasileiro* São Paulo (2017) 21-33
- [6] GAN, G.; MA, C.; WU, J. *Data Clustering: Theory, Algorithms, and Applications (ASA-SIAM Series on Statistics and Applied Probability)* [S.l.]: SIAM, 2007. (ASA-SIAM Series on Statistics and Applied Probability).
- [7] AMATRIAIN, X. *Data mining methods for recommender systems*. Boston, MA: Springer US (2011).
- [8] <https://bookdown.org/egarpor/NP-UC3M/kde-i-kde.html>. Acesso em 01 de Setembro de 2019 às 21:53 h.
- [9] Macal, C. M. *A guide for building hospital simulation models*. Argonne National Laboratory (2018).
- [10] Borshchev A, Filippov A. *Da dinâmica do sistema e eventos discretos à modelagem baseada em agentes práticos: razões, técnicas, ferramentas*. Oxford, Inglaterra (2004).
- [11] EPSTEIN, J. M.; AXTELL, R. *Growing artificial societies: social science from the bottom up*. Washington: Brookings Institution Press; Cambridge: MIT Press (1996).
- [12] Furtado B. *Policy Space: agent-based modeling*. IPEA (2018).
- [13] BILGE, U. *Agent based modelling and the global trade network*. In: GEYER, R.; CAIRNEY, P. Handbook on complexity and public policy. Cheltenham; Northampton: Edward Elgar Publishing (2015) 414-431.
- [14] HAMILL, L.; GILBERT, N. *Agent-Based modelling in economics*. United Kingdom: Wiley (2016).

- [15] Parrish J.K., Hamner W.M. *Animal Groups in Three Dimensions*, Cambridge University Press, Cambridge (1997).
- [16] VICSEK, T.; CZIRÓK, A.; BEN-JACOB, E.; COHEN, I.; SHOCHET, O. *Novel type of phase transition in a system of self-driven particles*. Phys. Rev. Lett., v. 75 (1995) 1226–1229.
- [17] NARAYAN, V.; RAMASWAMY, S.; MENON, N. *Long-Lived Giant Number Fluctuations in a Swarming Granular Nematic*. Science, v. 317, n. 5834 (2007) 105–108.
- [18] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Shochet. *Novel type of phase transition in a system of self-driven particles*. Physical Review Letters, vol. 75, n.6 (1995) 1-4.
- [19] CHATTERJEE, P. *The Flocking Transition : A Review of The Vicsek Model*. 2017.
- [20] BOLTZMANN, L. Weitere Studien uber das warmegleichgewicht unter gas molekulen. Sitzungsberichte Akademie der Wissenschaften, vol. 60 (1972).
- [21] GINELLI, F. *The Physics of the Vicsek model*. SUPA, Institute for Complex Systems and Mathematical Biology, King’s College, University of Aberdeen, (2016).