

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Aplicação do processo de Hawkes multivariado para prever o movimento do preço médio de livro de ofertas**

**Luis Henrique Claudino Silva**

Dissertação de Mestrado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-C<sup>2</sup>MC)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Luis Henrique Claudino Silva**

## Aplicação do processo de Hawkes multivariado para prever o movimento do preço médio de livro de ofertas

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Vanderlei Bonato

**USP – São Carlos**  
**Dezembro de 2021**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

C615a Claudino Silva, Luís Henrique  
Aplicação do processo de Hawkes multivariado para  
prever o movimento do preço médio de livro de  
ofertas / Luís Henrique Claudino Silva; orientador  
Vanderlei Bonato. -- São Carlos, 2021.  
107 p.

Dissertação (Mestrado - Programa de Pós-Graduação  
em Ciências de Computação e Matemática  
Computacional) -- Instituto de Ciências Matemáticas  
e de Computação, Universidade de São Paulo, 2021.

1. Processo de Hawkes. 2. Livro de Ofertas. I.  
Bonato, Vanderlei, orient. II. Título.

**Luis Henrique Claudino Silva**

Applying the multivariate Hawkes process to predict the  
movement of mid-price in a order book

Dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Master in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Vanderlei Bonato

**USP – São Carlos**  
**December 2021**



*Este trabalho é dedicado aos meus pais, Luis Carlos da Silva e Sandra Aparecida Claudino, que sempre me apoiaram desde o início do meu mestrado para que este sonho fosse concluído com sucesso.*



# AGRADECIMENTOS

---

---

Agradecimentos principais aos meus pais e familiares que sempre estiveram do meu lado sempre prontos para ajudar e apoiando nos momentos de felicidade e dificuldade. Ao professor Dr. Vanderlei Bonato pelo apoio, conselho, sugestões e principalmente pela confiança depositada em mim para a condução deste trabalho.

Meu muito obrigado aos membros do LCR: Caio, Carlos, Cláudio, Erinaldo, Fernanda, Flaviana, Henrique, Julian, Leandro, Magna, Rafael, Romeu, Valdemar e Veronica. Com alguns de vocês tive pouco contato, outros mais, porém cada um de vocês me fez sentir cada vez mais acolhido e agradeço pelas ajudas nas matérias, conversas e risadas. Agradecer também à Elisabeth pela estadia na sua pensão, morar lá foi como estar em casa, além de todo o pessoal que se hospedou lá que não deixaram ser apenas um lugar para dormir.

Agradecimento especial ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro, que permitiu a minha estadia em São Carlos durante dois anos para estar presente fisicamente na universidade para obter créditos de aulas e desenvolver meu projeto.



*“Seja você quem for, seja qual for a posição social que você tenha na vida, a mais alta ou a mais baixa, tenha sempre como meta muita força, muita determinação e sempre faça tudo com muito amor e com muita fé em Deus, que um dia você chega lá. De alguma maneira você chega lá.”*

*(Ayrton Senna)*



# RESUMO

SILVA, L. H. C. **Aplicação do processo de Hawkes multivariado para prever o movimento do preço médio de livro de ofertas.** 2021. 107 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Com a modernização do mercado financeiro, além da automatização de várias tarefas, uma grande quantidade de dados ficou disponível para operadores e pesquisadores utilizarem na criação de modelos e indicadores para entender o comportamento de investidores e seu impacto nos preços. Entre as várias formas de previsão, o Processo de Hawkes tem sido explorado como uma forma simples e flexível de analisar a dinâmica da microestrutura do mercado financeiro por meio da interação de eventos. Este trabalho implementa o processo de Hawkes multivariado para fazer previsões do movimento do preço-médio entre a melhor oferta de compra e de venda do livro de ofertas de ativos. Os resultados demonstraram uma precisão de aproximadamente 70% para previsão durante um período de 23 dias de um ativo de grande liquidez. Além disso, este trabalho apresenta a exploração de um algoritmo de negociação usando o método de Hawkes e apresenta a comparação dos resultados com estimativas feitas por meio de uma rede neural. Por fim, este projeto mostra os principais desafios para tornar este processo de Hawkes viável para problemas de previsão de preço em aplicações em tempo real. Por exemplo, a aceleração da computação de funções necessárias para os ajustes de parâmetros que ocorre periodicamente.

**Palavras-chave:** Processo de Hawkes, Livro de Ofertas.



# ABSTRACT

SILVA, L. H. C. **Applying the multivariate Hawkes process to predict the movement of mid-price in a order book.** 2021. 107 p. Dissertação (Mestrado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

With the modernization of the financial market and automation of various tasks, a large amount of data has become available for investors and researchers to create models and indicators to understand the market behavior and its impact on prices. Among the various forms of forecasting, the Hawkes Process has been explored as a simple and flexible way to analyze the dynamics of financial market microstructure through the interaction of events. This paper implements the multivariate Hawkes process to predict the mid-price movement between the best bid and ask in an order book of an asset. The results demonstrated a forecasting accuracy of approximately 70% over 23 days for a highly liquid asset. In addition, this paper presents the exploration of a trading algorithm using Hawkes' method and presents a comparison of the results with estimates made using a neural network. Finally, this project highlights the main challenges in making this Hawkes process feasible for price prediction problems in real-time applications. For example, the acceleration of the computation of functions needed for parameter settings that occurs from time to time.

**Keywords:** Hawkes Process, Order Book.



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Exemplo de um livro de ofertas com ofertas de compra no lado esquerdo em azul e com ofertas de venda no lado direito em vermelho. . . . .	30
Figura 2 – Chegada de uma oferta limitada no valor de R\$18,00 e volume 100 no lado da compra do livro de ofertas. . . . .	31
Figura 3 – Chegada de uma oferta a mercado de compra de 200 ações no valor de R\$18,02 no livro de ofertas e consumindo as ofertas de venda disponíveis. . . . .	32
Figura 4 – Cancelamento de ofertas registradas no livro de ofertas. . . . .	33
Figura 5 – Exemplo de um livro de ofertas real. Retirado de Nelogica (2019). . . . .	34
Figura 6 – Execução de um Processo de Hawkes Univariado com os parâmetros configurados da seguinte forma: $\mu = 1.2$ , $\alpha = 0.6$ e $\beta = 0.8$ . Retirado de (CHEN, 2017). . . . .	38
Figura 7 – Execução de um Processo de Hawkes Bivariado com os parâmetros configurados da seguinte forma: $\mu_1 = 0.1$ , $\mu_2 = 0.5$ , $\alpha_{11} = 0.1$ , $\alpha_{12} = 0.7$ , $\alpha_{21} = 0.5$ , $\alpha_{22} = 0.2$ , $\beta_{11} = 1.2$ , $\beta_{12} = 1.0$ , $\beta_{21} = 0.8$ , $\beta_{22} = 0.6$ . Retirado de (CHEN, 2017). . . . .	40
Figura 8 – Fluxo de execução do trabalho. . . . .	43
Figura 9 – Fluxograma demonstrando o fluxo para a mudança da estrutura de dados do Livro de Ofertas. Onde, é verificado qual lado será modificado e é aplicado uma das modificações realizadas por uma oferta. . . . .	48
Figura 10 – Estratégia de negociação. . . . .	58
Figura 11 – Matriz de confusão das previsões usando a estratégia simples do dia de negociações 01/10/2019. . . . .	61
Figura 12 – Métricas do desempenho da previsão do Processo de Hawkes no dia de negociações 01/10/2019. . . . .	62
Figura 13 – Ações realizadas pela estratégia atualizada no dia de negociações 01/10/2019. . . . .	63
Figura 14 – Métricas das previsões realizadas pelo processo de Hawkes usando a simples no mês de Outubro de 2019. . . . .	64
Figura 15 – Número de operações de compra/venda feita pela estratégia simples no mês de Outubro de 2019. . . . .	65
Figura 16 – Número de operações de compra/venda feita pela estratégia alterada no mês de Outubro de 2019. . . . .	65
Figura 17 – Lucro e prejuízo diário feito pela estratégia de Chen (2017) no mês de Outubro de 2019. . . . .	66

Figura 18 – Lucro e prejuízo acumulado feito pela estratégia de Chen (2017) no mês de Outubro de 2019. . . . .	66
Figura 19 – Lucro e prejuízo diário feito pela estratégia aprimorada no mês de Outubro de 2019. . . . .	67
Figura 20 – Lucro e prejuízo acumulado feito pela estratégia aprimorada no mês de Outubro de 2019. . . . .	67
Figura 21 – Comparação das métricas de previsão da Rede Neural de Bileki (2021) e o Processo de Hawkes. . . . .	69
Figura 22 – Lucratividade média dos algoritmos de previsão da Rede Neural de Bileki (2021) e do Processo de Hawkes. . . . .	69
Figura 23 – Métricas de precisão dos algoritmos de previsão de lançamento de moeda e do Processo de Hawkes. . . . .	70
Figura 24 – Lucratividade média dos algoritmos de previsão de lançamento de moeda e do Processo de Hawkes. . . . .	71
Figura 25 – Métricas de precisão das estratégias de 1 e 2 níveis na primeira semana de Outubro de 2019. . . . .	72
Figura 26 – Lucro e prejuízo diário feito pelas estratégias de 1 e 2 níveis na primeira semana de Outubro de 2019. . . . .	72
Figura 27 – Lucro e prejuízo acumulado feito pelas estratégias de 1 e 2 níveis na primeira semana de Outubro de 2019. . . . .	73
Figura 28 – Ações realizadas pela estratégia simples no dia de negociações 01/10/2019.	81
Figura 29 – Ações realizadas pela estratégia simples no dia de negociações 02/10/2019.	82
Figura 30 – Ações realizadas pela estratégia simples no dia de negociações 03/10/2019.	82
Figura 31 – Ações realizadas pela estratégia simples no dia de negociações 04/10/2019.	83
Figura 32 – Ações realizadas pela estratégia simples no dia de negociações 07/10/2019.	83
Figura 33 – Ações realizadas pela estratégia simples no dia de negociações 08/10/2019.	84
Figura 34 – Ações realizadas pela estratégia simples no dia de negociações 09/10/2019.	84
Figura 35 – Ações realizadas pela estratégia simples no dia de negociações 10/10/2019.	85
Figura 36 – Ações realizadas pela estratégia simples no dia de negociações 11/10/2019.	85
Figura 37 – Ações realizadas pela estratégia simples no dia de negociações 14/10/2019.	86
Figura 38 – Ações realizadas pela estratégia simples no dia de negociações 15/10/2019.	86
Figura 39 – Ações realizadas pela estratégia simples no dia de negociações 16/10/2019.	87
Figura 40 – Ações realizadas pela estratégia simples no dia de negociações 17/10/2019.	87
Figura 41 – Ações realizadas pela estratégia simples no dia de negociações 18/10/2019.	88
Figura 42 – Ações realizadas pela estratégia simples no dia de negociações 21/10/2019.	88
Figura 43 – Ações realizadas pela estratégia simples no dia de negociações 22/10/2019.	89
Figura 44 – Ações realizadas pela estratégia simples no dia de negociações 23/10/2019.	89
Figura 45 – Ações realizadas pela estratégia simples no dia de negociações 24/10/2019.	90
Figura 46 – Ações realizadas pela estratégia simples no dia de negociações 25/10/2019.	90

Figura 47 – Ações realizadas pela estratégia simples no dia de negociações 28/10/2019.	91
Figura 48 – Ações realizadas pela estratégia simples no dia de negociações 29/10/2019.	91
Figura 49 – Ações realizadas pela estratégia simples no dia de negociações 30/10/2019.	92
Figura 50 – Ações realizadas pela estratégia simples no dia de negociações 31/10/2019.	92
Figura 51 – Ações realizadas pela estratégia atualizada no dia de negociações 01/10/2019.	93
Figura 52 – Ações realizadas pela estratégia atualizada no dia de negociações 02/10/2019.	94
Figura 53 – Ações realizadas pela estratégia atualizada no dia de negociações 03/10/2019.	94
Figura 54 – Ações realizadas pela estratégia atualizada no dia de negociações 04/10/2019.	95
Figura 55 – Ações realizadas pela estratégia atualizada no dia de negociações 07/10/2019.	95
Figura 56 – Ações realizadas pela estratégia atualizada no dia de negociações 08/10/2019.	96
Figura 57 – Ações realizadas pela estratégia atualizada no dia de negociações 09/10/2019.	96
Figura 58 – Ações realizadas pela estratégia atualizada no dia de negociações 10/10/2019.	97
Figura 59 – Ações realizadas pela estratégia atualizada no dia de negociações 11/10/2019.	97
Figura 60 – Ações realizadas pela estratégia atualizada no dia de negociações 14/10/2019.	98
Figura 61 – Ações realizadas pela estratégia atualizada no dia de negociações 15/10/2019.	98
Figura 62 – Ações realizadas pela estratégia atualizada no dia de negociações 16/10/2019.	99
Figura 63 – Ações realizadas pela estratégia atualizada no dia de negociações 17/10/2019.	99
Figura 64 – Ações realizadas pela estratégia atualizada no dia de negociações 18/10/2019.	100
Figura 65 – Ações realizadas pela estratégia atualizada no dia de negociações 21/10/2019.	100
Figura 66 – Ações realizadas pela estratégia atualizada no dia de negociações 22/10/2019.	101
Figura 67 – Ações realizadas pela estratégia atualizada no dia de negociações 23/10/2019.	101
Figura 68 – Ações realizadas pela estratégia atualizada no dia de negociações 24/10/2019.	102
Figura 69 – Ações realizadas pela estratégia atualizada no dia de negociações 25/10/2019.	102
Figura 70 – Ações realizadas pela estratégia atualizada no dia de negociações 28/10/2019.	103
Figura 71 – Ações realizadas pela estratégia atualizada no dia de negociações 29/10/2019.	103
Figura 72 – Ações realizadas pela estratégia atualizada no dia de negociações 30/10/2019.	104
Figura 73 – Ações realizadas pela estratégia atualizada no dia de negociações 31/10/2019.	104
Figura 74 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 01/10/2019. . . . .	105
Figura 75 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 02/10/2019. . . . .	106
Figura 76 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 03/10/2019. . . . .	106
Figura 77 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 04/10/2019. . . . .	107
Figura 78 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 07/10/2019. . . . .	107



# LISTA DE ALGORITMOS

---

---

Algoritmo 1 – Simulação de Monte Carlo com N execuções independentes para estimar a probabilidade do preço-médio subir baseado nas observações até o tempo $t^*$ .	53
Algoritmo 2 – Simulação de Monte Carlo com N execuções independentes para estimar a probabilidade do preço-médio subir dois centavos baseado nas observações até o tempo $t^*$ . . . . .	54
Algoritmo 3 – Algoritmo de simulação da ocorrência do próximo evento do tipo k em um Processo de Hawkes M-variado . . . . .	57



# LISTA DE TABELAS

---

---

Tabela 1 – Layout da organização dos dados do arquivo NEG (Negócios) informado pela B3. . . . .	44
Tabela 2 – Layout da organização dos dados dos arquivos OFER_CPA e OFER_VDA informado pela B3. . . . .	45
Tabela 3 – Layout do arquivo de mensagens "Message". . . . .	46
Tabela 4 – Exemplo da estrutura de um arquivo de Mensagens criado pelo algoritmo Get_Data a partir do Layout da Tabela 3. . . . .	47
Tabela 5 – Descrição dos dados das colunas do arquivo do Livro de Ofertas criado pelo algoritmo Get_Data. . . . .	49
Tabela 6 – Exemplo da estrutura de um Livro de Ofertas criado pelo algoritmo exibindo 2 níveis de preço. . . . .	49



# SUMÁRIO

---

---

<b>1</b>	<b>INTRODUÇÃO</b>	<b>25</b>
1.1	Contextualização	25
1.2	Objetivo	26
1.3	Justificativa	26
1.4	Organização do Trabalho	27
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>29</b>
2.1	Considerações iniciais	29
2.2	Livro de Ofertas	29
2.3	Processos Pontuais	34
2.3.1	<i>Definições Básicas e Propriedades</i>	34
2.3.2	<i>Processo Pontual Auto-Excitante</i>	36
2.3.3	<i>Processo de Hawkes Univariado</i>	37
2.3.3.1	<i>Função de Log-Verossimilhança</i>	38
2.3.3.2	<i>Função de verossimilhança do Processo de Hawkes Univariado</i>	38
2.3.4	<i>Processo de Hawkes Multivariado</i>	39
2.3.4.1	<i>Função de verossimilhança do Processo de Hawkes Multivariado</i>	41
2.4	Trabalhos relacionados	41
<b>3</b>	<b>DESENVOLVIMENTO</b>	<b>43</b>
3.1	Considerações iniciais	43
3.2	Obtenção dos dados dos arquivos da B3	43
3.3	Geração dos arquivos de Mensagens e de Livro de Ofertas	45
3.4	Dinâmica do Livro de Ofertas com o Processo de Hawkes	49
3.4.1	<i>Ajustes dos parâmetros do Processo de Hawkes</i>	50
3.4.2	<i>Simulação de Monte Carlo</i>	52
3.4.2.1	<i>Simulação da ocorrência do próximo evento</i>	55
3.4.3	<i>Implementação de estratégia de negociação</i>	57
<b>4</b>	<b>RESULTADOS</b>	<b>59</b>
4.1	Considerações iniciais	59
4.2	Período de testes	59
4.3	Ajustes dos Parâmetros	59
4.4	Resultados da execução diária de negociação	60

4.5	Resultados das execuções do mês de Outubro de 2019 . . . . .	63
4.6	Comparação com técnica de Rede Neural . . . . .	68
4.7	Comparação com previsões aleatórias . . . . .	70
4.8	Adição do segundo nível na Simulação de Monte Carlo . . . . .	71
5	CONCLUSÕES . . . . .	75
5.1	Trabalhos Futuros . . . . .	76
	REFERÊNCIAS . . . . .	77
ANEXO A	DADOS DAS EXECUÇÕES DIÁRIAS DA ESTRATÉGIA SIMPLES PARA O MÊS DE OUTUBRO DE 2019 . . . . .	81
ANEXO B	DADOS DAS EXECUÇÕES DIÁRIAS DA ESTRATÉGIA APRIMORADA PARA O MÊS DE OUTUBRO DE 2019 . . . . .	93
ANEXO C	EXECUÇÕES DIÁRIAS DA ESTRATÉGIA USANDO 2 NÍVEIS PARA O MÊS DE OUTUBRO DE 2019 . . . . .	105

---

# INTRODUÇÃO

---

## 1.1 Contextualização

Muitas pessoas e instituições utilizam a bolsa de valores como forma de expansão e de proteção de seu patrimônio. De acordo com o relatório da Federação Mundial de Bolsas de Valores ([World Federation of Exchanges, 2018](#)), o mercado global possuía um valor de capitalização de 74.4 trilhões de dólares no final do último trimestre de 2018. A bolsa de valores é um mercado constituído por ofertas de compra e venda de diversos instrumentos (ativos) financeiros, dentre eles ações de empresas, commodities, moedas, índices e etc., em que negociações ocorrem quando há uma combinação de valores entre essas ofertas. Há diversas estratégias que podem ser adotadas para operar nesse tipo de mercado financeiro, sendo a maioria baseada em duas grandes fontes de informação, que são a análise técnica e a análise fundamentalista.

A análise fundamentalista leva em consideração o estudo de atributos fundamentais de uma empresa, tais como tamanho da empresa, taxa de capitalização, receitas, despesas, ativos e passivos ([VANSTONE; FINNIE, 2009](#)). Por outro lado, na análise técnica acredita-se que todos os fundamentos que influenciam o movimento do mercado já estão implicitamente incorporados no preço ([MURPHY, 1999](#)). Este tipo de análise busca identificar padrões acreditando que estas informações são suficientes para determinar o comportamento futuro do mercado ([TEIXEIRA; OLIVEIRA, 2010](#)).

Com a modernização dos sistemas de suporte a transações financeiras de bolsas de valores, uma grande quantidade de dados tornou-se disponível para ajudar o usuário na tomada de decisão. Devido ao grande volume de dados e também à redução significativa da janela de tempo em que as oportunidades permanecem disponíveis para negociação, há uma constante busca por algoritmos e métodos computacionais inteligentes que apoiem essa tomada de decisão. Um dado muito utilizado e rico em informações é o livro de ofertas, onde aparecem todas as

ofertas de compra e venda disponíveis no mercado em tempo real. Isso gerou uma necessidade na criação de modelos que aprendam e saibam reproduzir o seu comportamento para entender as várias dinâmicas presentes no mercado financeiro.

## 1.2 Objetivo

O objetivo principal deste projeto é implementar o Processo de Hawkes com finalidade de medir sua capacidade e identificar as suas limitações na previsão de preço médio do livro de ofertas em aplicações de tempo real.

Os objetivos específicos deste trabalho são:

- Criar uma base de dados que contenha livro de ofertas fiéis aos gerados pela bolsa de valores;
- Identificar o custo de computação para a calibração das variáveis do processo de Hawkes;
- Explorar a influência de diferentes tipos de variáveis presentes no livro de ofertas;
- Explorar o Processo de Hawkes em algoritmos de negociação que inclua custos transacionais;
- Comparar resultados com métodos baseados em aprendizado de máquina.

## 1.3 Justificativa

Uma das formas de realizar a modelagem da dinâmica do livro de ofertas é a estatística que assume que a chegada de eventos do mercado são diretamente regidas por processos estocásticos (GOULD *et al.*, 2013). Em um primeiro momento, Processos de Poisson tornaram-se a escolha de muitos autores por sua facilidade analítica para encontrar soluções de modelos teóricos, como pode ser visto em Cont, Stoikov e Talreja (2010), Daniels *et al.* (2003) e Huang e Kercheval (2012). Porém, uma clusterização dos eventos do mercado observada empiricamente por Biais, Hillion e Spatt (1995) e Hall e Hautsch (2008) sugere que tal comportamento não poderia ser explicado com o uso de Processos de Poisson.

O Processo de Hawkes, introduzido por Hawkes (1971b) tem sido usado como uma alternativa de modelagem a este problema na área financeira (HAWKES, 2018; BACRY; MASTROMATTEO; MUZY, 2015). Nele há propriedades de autoexcitação, em que eventos do mesmo tipo influenciam novas ocorrências, e de clusterização das ocorrências. Cabe ressaltar, que o processo de Hawkes é também amplamente aplicado em outras áreas da ciência, como sismologia (OGATA, 1988), neurociência (JOHNSON, 1996; CHORNOBOY; SCHRAMM; KARR, 1988), criminologia (MOHLER *et al.*, 2013; SANTITISSADEEKORN *et al.*, 2020) e estudos de redes sociais (ZADEH; SHARDA, 2015; ZIPKIN *et al.*, 2016).

## 1.4 Organização do Trabalho

O Capítulo 2 abrange os principais conceitos de um Livro de Ofertas, sua dinâmica na interação com as ofertas de compra e venda, depois introduz a teoria sobre processos pontuais auto-excitantes como o Processo de Hawkes univariado e multivariado e realiza uma apresentação dos diferentes trabalhos presentes na literatura para modelagem do comportamento de preços de um Livros de Ofertas utilizando Processo de Hawkes.

No Capítulo 3 é detalhado como são realizadas todas as etapas para obter os dados do *log* da B3 para construir os arquivos de Mensagens e do Livro de Ofertas. Tais dados são utilizados pelo algoritmo de simulação de Monte Carlo que realiza a previsão de subida ou descida do preço-médio e também pela estratégia de negociação que simula negociações de um ativo da B3.

O Capítulo 4 apresenta todos os resultados de precisão e lucratividade obtidos durante as simulações diárias no período do mês de Outubro de 2019, além de realizar comparações com um algoritmo de previsão baseado em Redes Neurais e um algoritmo que utiliza uma variável aleatória que simula o lançamento de uma moeda para realizar previsões. Além de comparar implementações do algoritmo de previsão do Processo de Hawkes que prevê o movimento de um centavo do preço médio com outro que prevê o movimento de dois centavos.

Por último, o Capítulo 5 traz as considerações finais com relação ao que foi apresentado neste projeto e aponta possíveis trabalhos futuros.



---

## REVISÃO BIBLIOGRÁFICA

---

### 2.1 Considerações iniciais

Neste capítulo serão apresentados os conceitos necessários para a compreensão deste projeto. A Seção 2.2 define um Livro de Ofertas, eventos que modificam seu estado e informações obtidas da sua estrutura. A Seção 2.3 apresenta alguns conceitos de probabilidade e de Processos Pontuais para levar ao entendimento do Processo de Hawkes. A Seção 2.4 apresenta os principais trabalhos relacionados a este trabalho de pesquisa.

### 2.2 Livro de Ofertas

O *continuous double auction* é um mecanismo de *trading* amplamente utilizado em bolsa de valores do mercado financeiro. O leilão (*auction*) é chamado de “*continuous double*”, pois os participantes do mercado podem submeter ordens na forma de *bids* (Ordens de Compra) e *asks* (Ordens de Venda) em qualquer ponto no tempo (ZHENG, 2013).

Um Livro de Ofertas Limitadas (*Limit Order Book*), é um registro que contém todas as ordens de compra e venda enviadas pelos participantes para o mercado que ainda não foram negociadas. Essas ofertas são denominadas limitadas pois delimitam um preço pelo qual é desejado para realizar uma negociação e possuem de forma geral as seguintes informações (ABERGEL *et al.*, 2016):

- Sinal : se é uma ordem de compra ou de venda;
- Preço : é o valor mínimo no caso de uma oferta de venda ou o valor máximo no caso de uma oferta de compra pelo qual um participante aceita realizar uma negociação;
- Volume : quantidade de ativos que o emissor da oferta deseja comprar ou vender;
- *Timestamp* : marcação do tempo de chegada da ordem ao sistema da bolsa de valores.

As ordens são organizadas em filas de compra e de venda, que são ordenadas por preço. Os maiores valores de compra se encontram no topo da fila de compra, enquanto os menores valores de venda se encontram na base da fila de venda. Para cada preço presente nas filas de compra e venda pode existir outra prioridade que é a ordem de chegada, onde as ordens mais antigas têm prioridade na execução.

A Figura 1 exemplifica o formato de um livro de ofertas, onde os retângulos em azul representam as ofertas de compra e em vermelho as ofertas de venda com o respectivo volume, separados por níveis de preço. O *spread* é a diferença, positiva, entre o melhor preço de venda e o melhor preço de compra, muito utilizado como um custo operacional para que determinada ordem seja executada. O *mid-price* é dado pela média aritmética simples entre o melhor preço de venda e de compra, no caso abaixo seria de R\$18,01.

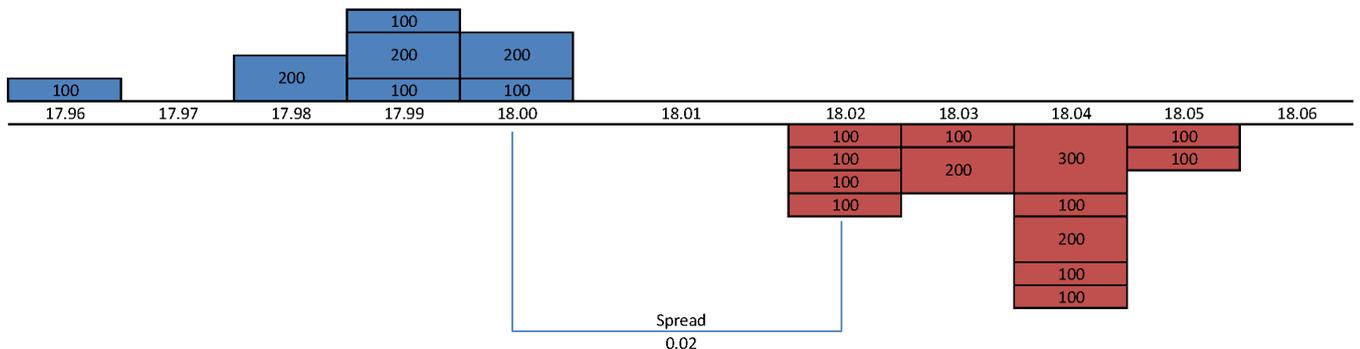


Figura 1 – Exemplo de um livro de ofertas com ofertas de compra no lado esquerdo em azul e com ofertas de venda no lado direito em vermelho.

A Figura 2 exemplifica a chegada de uma ordem limitada de compra com volume de 100 ações ao preço de R\$ 18,00 no livro de ofertas. Este tipo de ordem é chamada de limitada pois é delimitado um certo preço para realizar a negociação. Caso não haja ações do lado oposto do livro naquele preço, ela vai esperar alguma ordem chegar. Caso exista uma quantidade de ações menor do que a solicitada pela ordem, ela irá realizar uma negociação parcial e o restante vai esperar uma agressão do lado oposto do livro.

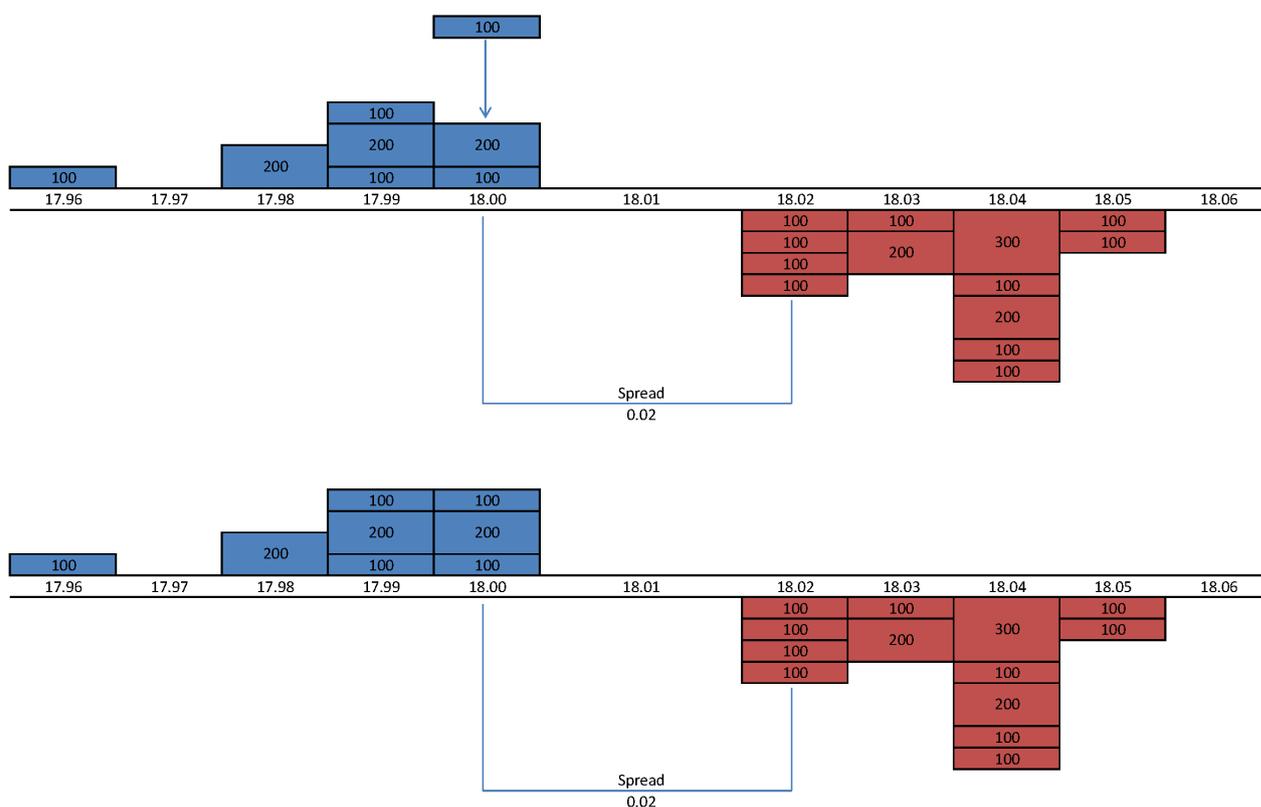


Figura 2 – Chegada de uma oferta limitada no valor de R\$18,00 e volume 100 no lado da compra do livro de ofertas.

A Figura 3 exemplifica a chegada de uma ordem a mercado de compra com volume de 200 ações ao preço de R\$18,02 no livro de ofertas. As ordens classificadas como “a mercado” são ordens para realizar uma negociação imediatamente na quantidade de ações desejada no melhor preço disponível na outra fila. Neste caso, como era desejado a compra de 200 ações ao preço de R\$18,02, duas ofertas de 100 ações na fila de venda foram negociadas. Caso não fosse possível realizar a compra total de ações naquele preço, o sistema iria realizar negociações com ações de um nível de preço acima até que toda a ordem fosse completada.

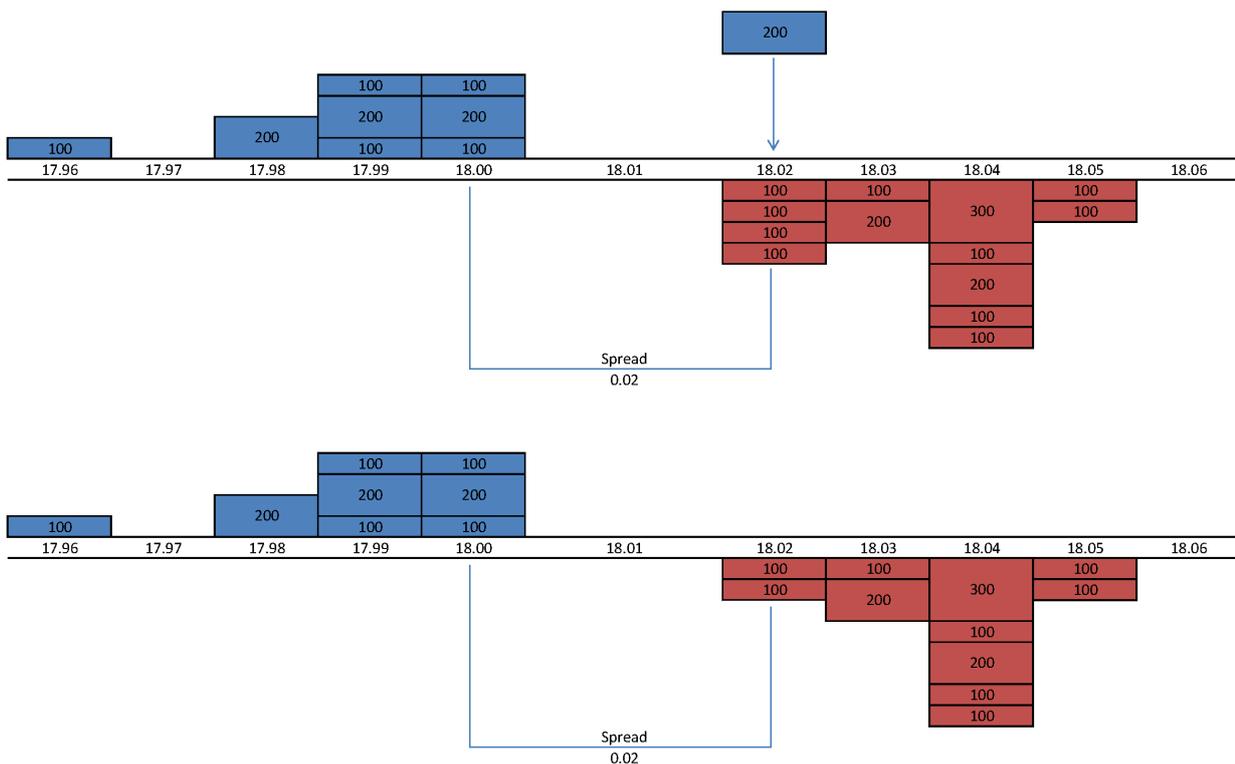


Figura 3 – Chegada de uma oferta a mercado de compra de 200 ações no valor de R\$18,02 no livro de ofertas e consumindo as ofertas de venda disponíveis.

A Figura 4 mostra o ato de cancelar ordens, neste caso a ordem é retirada com aval do operador que a colocou e o livro de ofertas é reorganizado.

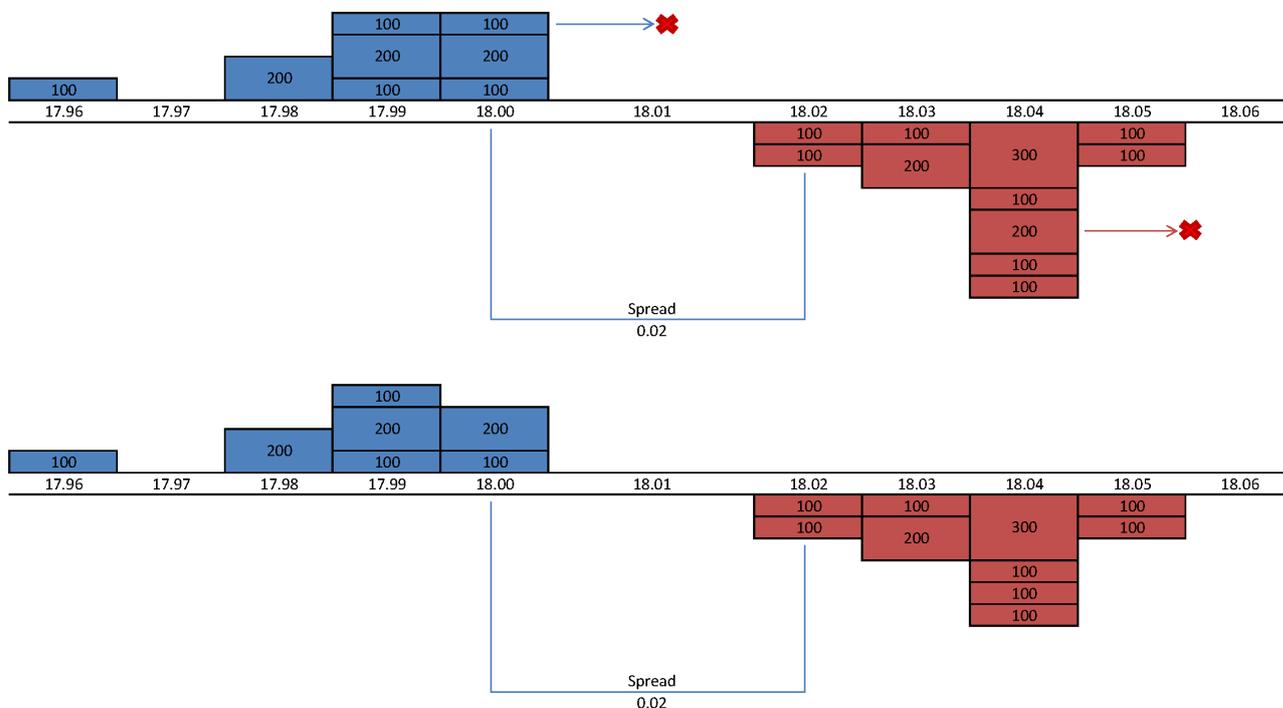


Figura 4 – Cancelamento de ofertas registradas no livro de ofertas.

A Figura 5 apresenta um exemplo real de um livro de ofertas completo onde é possível visualizar todas as ofertas que o compõem, ordenadas por preço e depois por tempo de chegada dentro da mesma faixa de preço, com as mais antigas posicionadas no topo. Cada faixa de preço é conhecida como nível, onde os melhores preços ficam no nível 1, os segundos melhores preços ficam no nível 2 e assim por diante. Neste tipo de livro de ofertas onde pode-se visualizar todas as ofertas disponíveis na fila é conhecido como livro de ofertas aberto.

Ativo	Último	Varia...	Hora	Volume	Negó...	Máximo	Mínimo	Abertura	Fechame...	
DOLFUT	2.683,00	2,21%	11:22:56	24.492,87M	13.461	2.689,00	2.628,00	2.632,50	2.625,00	
						(5,27k) 63%	37% (3,06k)			
Agente	Qtde	Compra	Venda	Qtde	Agente					
Bradesco	15	2.682,50	2.683,00	5	UBS Brasil					
UBS Brasil	5	2.682,50	2.683,50	10	J.P. Morgan					
Bradesco	5	2.682,50	2.683,50	10	ICAP					
Deutsche Bank	5	2.682,50	2.683,50	5	UBS Brasil					
BES	20	2.682,50	2.683,50	5	UBS Brasil					
BES	5	2.682,50	2.683,50	5	UBS Brasil					
Morgan Stanley	5	2.682,50	2.683,50	5	XP					
Alpes	5	2.682,50	2.683,50	5	UBS Brasil					
Alpes	5	2.682,00	2.683,50	5	Alpes					
Alpes	5	2.682,00	2.684,00	55	Liquidez					
Alpes	5	2.682,00	2.684,00	5	UBS Brasil					
UBS Brasil	5	2.682,00	2.684,00	5	UBS Brasil					
XP	5	2.682,00	2.684,00	10	XP					
XP	5	2.682,00	2.684,00	5	XP					
Alpes	10	2.682,00	2.684,00	5	UBS Brasil					
UBS Brasil	5	2.682,00	2.684,00	5	Alpes					
UBS Brasil	5	2.682,00	2.684,00	5	Alpes					
UBS Brasil	5	2.682,00	2.684,00	5	ICAP					
Tullett	10	2.682,00	2.684,50	5	UBS Brasil					
ICAP	5	2.682,00	2.684,50	5	J.P. Morgan					
Bradesco	15	2.682,00	2.684,50	5	UBS Brasil					
Alpes	5	2.681,50	2.684,50	5	UBS Brasil					

Figura 5 – Exemplo de um livro de ofertas real. Retirado de [Nelogica \(2019\)](#).

## 2.3 Processos Pontuais

Qualquer tentativa de modelar um fenômeno do mundo real deve levar em conta a possibilidade de aleatoriedade. Na maioria das vezes, a grandeza que está sendo estudada não será previsível com antecedência, mas irá exibir uma variação inerente que pode ser levada em consideração pelo modelo. Nesta seção, serão apresentados alguns conceitos básicos sobre Processos Pontuais.

### 2.3.1 Definições Básicas e Propriedades

Um processo pontual é uma coleção de pontos localizados randomicamente em um espaço matemático, como a linha real ou o espaço Euclidiano. [Daley e Vere-Jones \(2003\)](#) define um Processo Pontual como um objeto aleatório que surge ou está associado a um Processo Estocástico. De acordo com [Last e Brandt \(1995\)](#), processos pontuais na linha real formam um caso especial importante que é particularmente suscetível de estudo, pois os pontos são ordenados de uma forma natural em que o processo pode ser descrito completamente por intervalos entre os pontos.

O conjunto de todas as saídas possíveis de um experimento é conhecido como Espaço Amostral e é denotado por  $\Omega$ . Cada saída do experimento é conhecido como Evento ([ROSS, 2010](#)). Se o experimento consiste no arremesso de uma moeda, então os eventos possíveis são  $E_1 = \text{Cara}$  e  $E_2 = \text{Coroa}$  e o Espaço Amostral será  $\Omega = \{\text{Cara}, \text{Coroa}\}$ . Considerando um

espaço amostral  $\Omega$ , para cada evento  $E$ , a probabilidade de ocorrência do evento  $P(E)$  é definida por um número que satisfaça as seguintes condições:

1.  $0 \leq P(E) \leq 1$ ;
2.  $P(\Omega) = 1$ ;
3. Para qualquer sequência de eventos  $E_1, E_2, \dots$ , que sejam mutuamente exclusivos (eventos para os quais a intersecção  $E_n E_m = \emptyset$  quando  $n \neq m$ ), então  $P\left(\bigcup_{i=1}^{\infty} E_n\right) = \sum_{i=1}^{\infty} P(E_n)$ .

Com isso, um Processo Pontual pode ser definido da seguinte forma:

**Definição 1.** (SIGMAN, 1995, p. 1) Dado um Espaço de Probabilidade  $(\Omega, \mathcal{F}, P)$ . Um processo pontual no  $\mathbf{R}^+$  é uma sequência de variáveis aleatórias positivas  $\{T_k\}_{k=1,2,\dots}$ , tais que para todo  $k$ ,  $T_k \leq T_{k+1}$ .

Onde,  $\Omega$ , como já mencionado, é o espaço amostral.  $P$  é a função que retorna a probabilidade de um evento específico ocorrer.  $T_k$  representa os tempos de ocorrência de um tipo de evento aleatório.  $\mathcal{F}$  é chamado de filtração natural (COCULESCU; NIKEGHBALI, 2010), que é um subconjunto que contém as ocorrências de eventos no passado que se quer observar. Assim, quando for utilizada a notação  $\mathcal{F}_t^N$  para a filtração natural, ela se refere ao conjunto de eventos de um processo pontual  $N$  que ocorreram até um tempo  $t$  específico.

**Definição 2.** (SIGMAN, 1995, p. 1, (1.2)) O processo com caminhos amostrais contínuos à direita, é chamado processo de contagem associado com o processo pontual  $\{T_k\}_{k=1,2,\dots}$  e definido pela seguinte equação:

$$N(t) = \sum_{k=1,2,\dots} \mathbb{1}_{\{T_k \leq t\}}. \quad (2.1)$$

Ou seja, um processo de contagem pode ser interpretado como um contador acumulativo do número de chegadas de eventos aleatórios em um sistema e é representado por uma função degrau.

**Definição 3.** (LAST; BRANDT, 1995, p. 8, (1.5.1)) Para um processo pontual  $\{T_k\}_{k=1,2,\dots}$ , o ponto de explosão é definido como:

$$T_{\infty} = \lim_{k \rightarrow \infty} T_k.$$

Assim, o processo pontual é chamado não-explosivo se  $T_k \rightarrow \infty$  se e somente se  $k \rightarrow \infty$ , ou seja, para um determinado período finito de tempo o número de pontos deve ser finito.

**Proposição 1.** (LAST; BRANDT, 1995, p. 8, Observação 1.5.1) Um processo pontual é não explosivo se e apenas se  $N(t) < \infty$  para todo  $t < \infty$ .

Ou seja, o número de eventos que ocorrem em qualquer intervalo finito também é finito.

**Definição 4.** (GRANDELL, 1997, p. 52, Definição 3.3) Um processo pontual  $N(t)$  é dito ser (estritamente) estacionário se  $N(t+s) - N(t)$  possui a mesma distribuição para todo  $t \geq 0$ .

Em outras palavras, para um processo pontual ser estacionário, o número de ocorrências em um intervalo apenas depende do tamanho do intervalo.

**Definição 5.** (DALEY; VERE-JONES, 2003, p. 47, Definição 3.3.II) Um processo pontual é chamado de simples quando  $P\{N(t) = 0 \text{ ou } 1 \text{ para todo } t\} = 1$ .

Modelando um processo pontual como os tempos de ocorrência dos eventos de mercado, a propriedade de simplicidade do processo irá excluir a possibilidade de múltiplos eventos ocorrendo ao mesmo tempo. Assim, para qualquer ocorrência  $\{t_k\}_{k=1,2,\dots}$ , tem-se que  $t_k < t_{k+1}$  para todo  $k$ , e processo de contagem associado é uma função degrau com pulo de tamanho 1.

Uma forma de medir a taxa de ocorrência de pontos é feita pela seguinte fórmula apresentada por Khinchin, Andrews e Quenouille (2013):

$$\lambda(t) = \lim_{h \rightarrow 0^+} \frac{P(N(t+h) - N(t) > 0)}{h}. \quad (2.2)$$

A taxa de ocorrência de eventos, também chamada de função intensidade, está relacionada com a probabilidade de ocorrer pelo menos um evento em um período curto de tempo e isso é dependente do comportamento do processo de contagem associado.

### 2.3.2 Processo Pontual Auto-Excitante

**Definição 6.** (SNYDER; MILLER, 2012, p. 287) Um processo pontual é chamado auto-excitante se a intensidade  $\lambda$  depende não apenas do tempo  $t$ , mas também de todo o passado do processo pontual.

Ou seja, para possuir essa característica o comportamento do processo é dependente não só do estado atual mas também de todas as ocorrências passadas. Com isso, a função intensidade  $\lambda$  para um processo pontual auto-excitante pode ser definida de forma análoga à Função 2.2.

**Definição 7.** (LAST; BRANDT, 1995, p. 10, (1.6.4)) Seja  $N(t)$  um processo pontual com filtração natural  $\mathcal{F}_t^N$ . O processo contínuo à esquerda chamado de função intensidade estocástica do processo pontual é definido por:

$$\lambda(t|\mathcal{F}_t^N) = \lim_{h \rightarrow 0^+} \frac{P\{N(t+h) - N(t) > 0 | \mathcal{F}_t^N\}}{h}. \quad (2.3)$$

De acordo com Daley e Vere-Jones (2003, p. 232), a razão de usar a continuidade à esquerda está conectada com previsibilidade: "Se a intensidade condicional possui um descontinuidade em um ponto do processo, então seu valor naquele ponto deve ser definido pela história anterior ao ponto e não para o que acontece no ponto em si".

### 2.3.3 Processo de Hawkes Univariado

O Processo de Hawkes é um processo do tipo auto-excitante e é definido por:

**Definição 8.** (HAWKES, 1971b, p. 84, (7) e (8)) Um processo pontual  $N(t)$  é chamado Processo de Hawkes quando  $\lambda(t)$  é um processo estocástico caracterizado pela integral de Stieltjes:

$$\lambda(t) = \mu + \int_{s < t} \gamma(t-s) dN(s). \quad (2.4)$$

Onde,  $\mu > 0$  é a intensidade base que representa a taxa mínima de ocorrência de eventos e  $\gamma(u) \geq 0$  é o *Kernel* de excitação que representa o comportamento da variação da intensidade quando ocorre um evento. Quando  $\gamma(u)$  for uma função constante igual à zero, então ele torna-se um Processo de Poisson.

Este trabalho será focado em Processos de Hawkes com *Kernels* que são funções exponenciais por simplicidade. Para o caso univariado, ele possuirá a forma:

$$\gamma(u) = \alpha e^{-\beta u}. \quad (2.5)$$

Substituindo a função (2.5) na função (2.4), a função intensidade do processo de Hawkes univariado (HAWKES, 1971b, p. 84, (7) e (8)) pode ser escrito da seguinte forma:

$$\lambda(t) = \mu + \int_0^t \alpha e^{-\beta(t-s)} dN(s) = \mu + \sum_{\{k:t_k < t\}} \alpha e^{-\beta(t-t_k)}. \quad (2.6)$$

Com  $0 < \alpha < \beta$ ,  $\alpha$  representa o aumento instantâneo da intensidade quando ocorre um evento e  $\beta$  é a taxa de decaimento da intensidade após a ocorrência de um evento. Caso não ocorram eventos durante um período, o valor da função intensidade tende ao valor de  $\mu$  e permanece constante. A Figura 6 ilustra a função intensidade (gráfico de cima) e o processo de contagem associado (gráfico de baixo) do Processo de Hawkes Univariado que possui os parâmetros configurados com os valores  $\mu = 1.2$ ,  $\alpha = 0.6$  e  $\beta = 0.8$ . A função intensidade começa e mantém a intensidade base de  $\mu = 1.2$  até a ocorrência do primeiro evento. Depois de cada evento, a intensidade pula com um tamanho de  $\alpha = 0.6$  e imediatamente começa a decair à uma taxa determinada por  $\beta = 0.8$ . O processo de contagem associado é uma função degrau que incrementa seu valor por 1 a cada nova ocorrência.

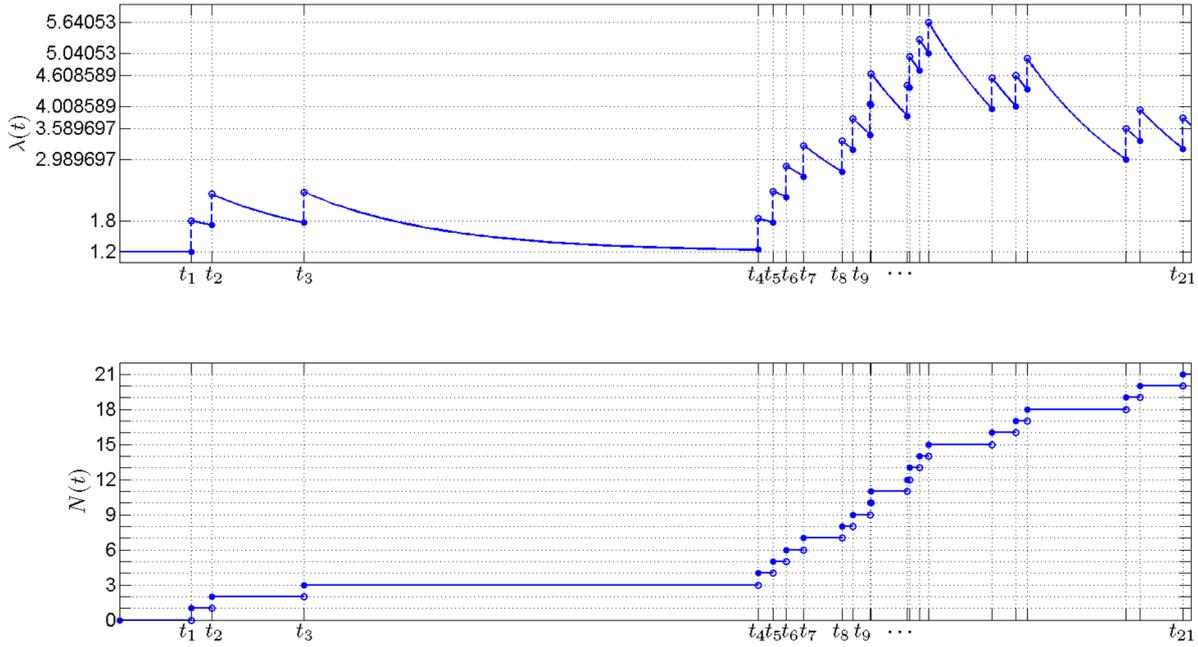


Figura 6 – Execução de um Processo de Hawkes Univariado com os parâmetros configurados da seguinte forma:  $\mu = 1.2$ ,  $\alpha = 0.6$  e  $\beta = 0.8$ . Retirado de (CHEN, 2017).

### 2.3.3.1 Função de Log-Verossimilhança

Funções de verossimilhança desempenham um papel fundamental na inferência estatística, especialmente em métodos de calibração de um conjunto de parâmetros para uma dada observação. Diferentemente de funções probabilísticas que descrevem futuras saídas para um conjunto de parâmetros fixo, as funções de verossimilhança são usadas para achar o conjunto de parâmetros dada uma saída já estabelecida. Para processos pontuais, a função de log-verossimilhança é dado pelo seguinte teorema:

**Teorema 1.** (RUBIN, 1972, p. 550-551, Teorema 2) Seja  $0 = t_0 < t_1 < \dots < t_n \leq T$  os eventos do processo pontual  $\{T_k\}_{k=1,2,\dots,n}$ , que é dependente do conjunto de parâmetros denotado por  $\theta$ , em um dado intervalo  $[0, T]$ . Adotando as notações de (RUBIN, 1972, p. 550, Definição 2), onde  $\{T_k\}_{k=1,2,\dots,n}$  é denotado por  $\omega$  e com  $\theta$  subscripto indicando dependência, a função de log-verossimilhança é:

$$\ln L(\theta|\omega) = - \int_0^T \lambda_\theta(t|\omega) dt + \int_0^T \ln \lambda_\theta(t|\omega) dN(t|\omega). \quad (2.7)$$

### 2.3.3.2 Função de verossimilhança do Processo de Hawkes Univariado

Substituindo a função (2.6) na função (2.7), tem-se a função de log-verossimilhança utilizada para obter os parâmetros  $\theta = (\mu, \alpha, \beta)$  do Processo de Hawkes Univariado dados os

eventos  $0 < t_1 < \dots < t_n$  no intervalo de  $[0, t_n]$ :

$$\ln L(\theta | t_1, t_2, \dots, t_n) = \mu t_n + \frac{\alpha}{\beta} [R(n) - (n-1)] + \sum_{i=1}^n \ln[\mu + \alpha R(i)]. \quad (2.8)$$

Onde  $R(1) = 0$  e:

$$R(k) = \sum_{i=1}^{k-1} e^{-\beta(t_k - t_i)}$$

para  $k \geq 2$ .

### 2.3.4 Processo de Hawkes Multivariado

Processos de Hawkes Multivariados são conhecidos por serem processos com excitação mútua, pois além de possuírem a característica de auto-excitação do caso univariado onde eventos de um tipo aumentam a probabilidade de ocorrer mais eventos do mesmo tipo, eles também possuem a característica de excitação cruzada entre diferentes dimensões. Eventos de um tipo aumentam a probabilidade de ocorrer eventos de tipos diferentes.

**Definição 9.** (BRÉMAUD, 1981, p. 19-20) Seja  $\{T_k\}_{k=1,2,\dots}$  um processo pontual definido em  $(\Omega, \mathcal{F}, P)$ , e seja  $\{Z_k\}_{k=1,2,\dots}$  uma sequência de variáveis aleatórias M-valoradas, também definida em  $(\Omega, \mathcal{F}, P)$ . Então, a dupla sequência  $\{(T_k, Z_k)\}_{k=1,2,\dots}$  é chamada um processo pontual M-variado.

Assim, para uma realização  $\{t_k, z_k\}_{k=1,2,\dots}$  de um processo pontual M-variado,  $t_k$  denota o tempo de ocorrência do k-ésimo evento e  $z_k$  indica o seu tipo. Portanto, é dito ser do tipo  $m \in \{1, 2, \dots, M\}$  se e somente se  $z_k = m$ .

**Definição 10.** (BRÉMAUD, 1981, p. 19) Para qualquer  $m \in \{1, 2, \dots, M\}$  e  $t \geq 0$ ,  $N^m(t)$  conta o número de ocorrências do tipo m até o tempo t e é dado pela seguinte equação:

$$N^m(t) = \sum_{k=1,2,\dots} \mathbb{1}_{T_k \leq t} \cdot \mathbb{1}_{Z_k = m}. \quad (2.9)$$

O processo M-vetor é o processo de contagem associado à um Processo de Hawkes Multivariado e contém todos os processos de contagem para os M tipos como observado na seguinte equação:

$$\mathbf{N}(t) = (N^1(t), N^2(t), \dots, N^M(t)). \quad (2.10)$$

Deste modo,  $\mathbf{N}(t)$  realiza a contagem de todos os eventos de todos os tipos que ocorreram.

**Definição 11.** (COX; LEWIS *et al.*, 1972, p. 414) Seja  $\mathbf{N}(t) = (N^1(t), N^2(t), \dots, N^M(t))$  um processo pontual M-variado. As funções de intensidade estocástica para o processo, com  $m \in \{1, 2, \dots, M\}$ , podem ser definidas por:

$$\lambda^m(t | \mathcal{F}_t^{\mathbf{N}}) = \lim_{h \rightarrow 0^+} \frac{P\{N^m(t+h) - N^m(t) > 0 | \mathcal{F}_t^{\mathbf{N}}\}}{h}. \quad (2.11)$$

Onde,  $\mathcal{F}_t^{\mathbf{N}}$  é a filtração natural do processo pontual M-variado  $\mathbf{N}(t)$  que contém o histórico de eventos ocorridos de todos os M tipos. Dessa forma, define-se o Processo de Hawkes multivariado (HAWKES, 1971b) como:

$$\begin{aligned}\lambda^m(t) &= \mu_m + \sum_{n=1}^M \int_0^t \alpha_{mn} e^{-\beta_{mn}(t-s)} dN^n(s) \\ &= \mu_m + \sum_{n=1}^M \sum_{k: t_k^n < t} \alpha_{mn} e^{-\beta_{mn}(t-t_k^n)}.\end{aligned}\tag{2.12}$$

Onde, para  $m, n = 1, 2, \dots, M$ ,  $\mu_m > 0$  é a intensidade base que representa a taxa mínima de ocorrência de eventos do tipo m,  $\alpha_{mn} \geq 0$  é o aumento instantâneo da intensidade do tipo m quando ocorre um evento do tipo n e  $\beta_{mn} \geq 0$  é a taxa de decaimento da intensidade do tipo m após a ocorrência de um evento do tipo n. A figura 7 mostra um exemplo de execução de um Processo de Hawkes Multivariado com  $M = 2$ , os gráficos de cima apresentam o valor da função intensidade para cada um dos dois tipos e os gráficos de baixo mostram os processos de contagem associados para cada um dos dois tipos. Neste exemplo é possível observar a ocorrência de ambos efeitos de auto-excitação e de excitação cruzada nas funções intensidades, o primeiro pulso ocorrido em  $\lambda^2(t)$  é ocasionado pelo primeiro ponto  $t_1^1$  de  $N^1(t)$  e o tamanho do pulso é determinado por  $\alpha_{21} = 0.5$ . Além disso, esse mesmo evento realiza um pulso de tamanho  $\alpha_{11} = 0.1$  em  $\lambda^1(t)$ .

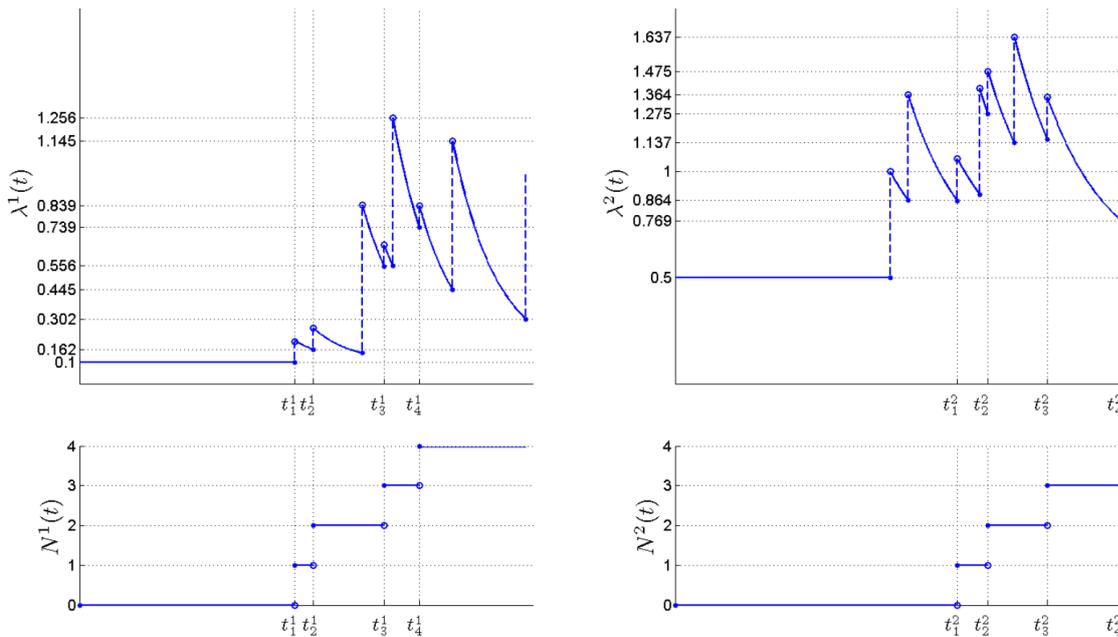


Figura 7 – Execução de um Processo de Hawkes Bivariado com os parâmetros configurados da seguinte forma:  $\mu_1 = 0.1$ ,  $\mu_2 = 0.5$ ,  $\alpha_{11} = 0.1$ ,  $\alpha_{12} = 0.7$ ,  $\alpha_{21} = 0.5$ ,  $\alpha_{22} = 0.2$ ,  $\beta_{11} = 1.2$ ,  $\beta_{12} = 1.0$ ,  $\beta_{21} = 0.8$ ,  $\beta_{22} = 0.6$ . Retirado de (CHEN, 2017).

### 2.3.4.1 Função de verossimilhança do Processo de Hawkes Multivariado

De acordo com Daley e Vere-Jones (2003), a função de log-verossimilhança do Processo de Hawkes Multivariado pode ser escrita como o somatório das funções de log-verossimilhança dos  $M$  tipos individualmente:

$$\ln L(\theta|\omega) = \sum_{m=1}^M \ln L^m(\theta|\omega). \quad (2.13)$$

Onde para uma realização  $\omega$  que contém todos os pontos  $\{t_k^m\}_{k=1,2,\dots}$  para cada uma das dimensões  $m = 1, 2, \dots, M$  no intervalo  $[0, T]$ :

$$\begin{aligned} \ln L^m(\theta|\{t_k^m\}_{k=1,2,\dots,M}) &= \int_0^T \lambda_\theta^m(t|\omega) dt + \int_0^T \ln \lambda_\theta^m(t|\omega) dN^m(t) \\ &= -\mu_m T - \sum_{n=1}^M \frac{\alpha_{mn}}{\beta_{mn}} \sum_{\{k:t_k^n < T\}} \left[ 1 - e^{-\beta_{mn}(T-t_k^n)} \right] + \sum_{\{k:t_k^m < T\}} \ln \left[ \mu_m + \sum_{n=1}^M \alpha_{mn} R_{mn}(k) \right]. \end{aligned} \quad (2.14)$$

Com  $R_{mn}(k)$ :

$$R_{mn}(k) = \sum_{\{i:t_i^n < t_k^m\}} e^{-\beta_{mn}(t_k^m - t_i^n)}, \quad R_{mn}(0) = 0.$$

## 2.4 Trabalhos relacionados

Esta seção possui a finalidade de mostrar algumas aplicações do processo de Hawkes na área financeira.

Alan G. Hawkes introduziu pelos trabalhos (HAWKES, 1971a; HAWKES, 1971b; HAWKES, 1972) uma categoria de modelos para processos pontuais estocásticos chamados "processos pontuais auto-excitantes e mutuamente excitantes", onde a principal propriedade era que a ocorrência de qualquer evento aumentava a probabilidade de novos eventos ocorrerem. Como apontado por Bacry, Mastromatteo e Muzy (2015), essas características ajudaram na sua adoção por seus parâmetros terem uma interpretação direta e bastante simples durante a modelagem da dinâmica de mercados.

Ao modelar a dinâmica de um Livro de Ofertas utilizando um processo bivariado para simular a interação entre o tempo de ocorrência de negociações e mudanças do *mid-price*, Bowsher (2007) encontrou relações como o aumento da intensidade de eventos de mudanças do preço-médio com o aumento da intensidade de ocorrências de negociações e vice-versa. Com isso, ele conseguiu por meio da estimativa da intensidade de mudanças do preço-médio realizar uma aproximação da volatilidade instantânea do preço do mercado.

Toke e Pomponio (2011) criaram um modelo para simular a ocorrência de *trades-through*, transações que agridem o livro de ofertas de modo a alcançar pelo menos o segundo nível de preços no Livro de Ofertas. Os resultados apontaram que a influência cruzada entre *trades-through* no lado da compra e venda é fraca e que apesar de a calibração variar muito de um dia para outro, os padrões gerais se mantiveram e que um simples processo de Hawkes bivariado se ajusta perfeitamente nas observações empíricas dos dados *tick por tick* da Euronext. Assim, a simplicidade do modelo poderia levar a futuros trabalhos práticos em estratégias de negociações baseadas em *trades-through* e modelagem do livro de ofertas.

O trabalho de Bacry *et al.* (2013) utiliza o processo de Hawkes para modelar os saltos positivos e negativos do preço para mais de um ativo, conseguindo reproduzir o ruído da microestrutura do mercado entre os ativos. Isso levou Bacry a criar um novo modelo (BACRY; MUZY, 2014) para modelar as dinâmicas do preço e de negociações em alta frequência utilizando informações de tempo de chegada de ordens a mercado e movimentos do preço, conseguindo encontrar correlações na interação desses tipos de eventos.

Lu e Abergel (2018) utilizou Processo de Hawkes não linear multi-dimensional com *kernels* exponenciais para descrever a dinâmica de eventos de inserção, liquidação e cancelamento de ordens que movimentam ou não o preço médio tanto do lado da compra como na venda. Devido às distinções entre eventos que ativam ou não uma mudança imediata do preço, o modelo conseguiu refletir as dinâmicas do preço do mercado.

O trabalho de Chen (2017) propôs o uso de um processo de Hawkes de 4 dimensões para modelar eventos de aumento e diminuição de volumes dos lados da compra e venda no primeiro nível de preços do livro de ofertas e assim prever as probabilidades do movimento do preço-médio usando simulação de Monte Carlo. O algoritmo apresentou retornos positivos em testes utilizando dados reais da INTC e da MSFT presentes na NASDAQ.

O processo de Hawkes já foi utilizado para quantificar a reflexividade do mercado, que é uma teoria social onde o comportamento do mercado foge do seu estado de equilíbrio, causando tendências que não correspondem aos verdadeiros fundamentos e que pode causar momentos de "boom-burst" (SHAIKH, 2013). O trabalho de Filimonov e Sornette (2012) utilizou um processo de Hawkes para modelar mudanças de preços devidos a interações endógenas ao mercado e exógenas geradas por notícias para obter a taxa média de derivações de eventos  $n$ . No evento de "flash-crash" de 6 de maio de 2010 observou um aumento precursor de  $n$  chegando ao valor crítico de 1, sugerindo uma forte influência endógena derivada da ação de reflexividade do mercado. Outras aplicações na área financeira podem ser vistas em Hawkes (2018) e Bacry, Mastromatteo e Muzy (2015).

---

## DESENVOLVIMENTO

---

### 3.1 Considerações iniciais

Neste capítulo serão apresentados a estrutura e extração de dados dos arquivos da B3, a etapa de pré-processamento para gerar os arquivos de mensagens e de livro de ofertas, ajustes dos parâmetros e descrição dos algoritmos utilizados para fazer previsões do movimento do preço médio, como apresentado na Figura 8.



Figura 8 – Fluxo de execução do trabalho.

### 3.2 Obtenção dos dados dos arquivos da B3

Os dados são obtidos do log de transações da B3, com resolução de 1 milissegundo em relação ao tempo de chegada de ofertas. Para cada dia de negociação a B3 disponibiliza três arquivos de logs para cada ativo negociado na bolsa de valores com as iniciais NEG, OFER\_CPA e OFER\_VDA, onde:

- NEG: Contém as informações de negociações ocorridas;
- OFER\_CPA: Contém as informações de eventos do lado que contém ofertas de compra (*bid*);

- OFER\_VDA: Contém as informações de eventos do lado que contém ofertas de venda (*ask*);

O arquivo NEG possui os dados organizados conforme a Tabela 1:

Tabela 1 – Layout da organização dos dados do arquivo NEG (Negócios) informado pela B3.

Nome da Coluna	Descrição do valor da coluna
Data Sessão	Data da Sessão, é o dia de negociação do arquivo gerado. Ex: 2019-07-03
Símbolo do Instrumento	Símbolo do Instrumento, é o código do ativo. Ex: PETR4
Nr.Negócio	Número do negócio é um número de identificação que incrementa a cada nova negociação. Ex: 0000327780
Preço Negócio	Preço em que foi realizado a negociação. Ex: 00000000027.160000
Quantidade	Quantidade negociada. Ex: 00000000000000100
Hora	Horário da negociação (formato HH:MM:SS.NNN). Ex: 16:27:23.972
Ind.Anulação	Indicador de Anulação: 1 - ativo / 2 - cancelado
Data Oferta Compra	Data de geração da oferta de compra. Ex: 2019-07-03
Seq.Oferta Compra	Número sequencial da oferta de compra, valor único e identificador da ordem. Ex: 000082638623279
GenerationID - Of.Compra	Número de geração (GenerationID) da Oferta de compra. Quando ocorre mais de uma mudança com a ordem este número é atualizado para identificar um novo estado da oferta. Ex: 000000048602374
Condição Oferta de Compra	Verifica se a ordem de compra é o iniciador da agressão ao livro de ofertas. Pode conter os seguintes valores: 0 - Oferta Neutra - é aquela que entra no mercado e não fecha com oferta existente. 1 - Oferta Agressora - é aquela que ingressa no mercado para fechar com uma oferta existente. 2 - Oferta Agredida - é a oferta (existente) que é fechada com uma oferta agressora.
Data Oferta Venda	Data de geração da oferta de venda. Ex: 2019-07-03
Seq.Oferta Venda	Número sequencial da oferta de venda. Ex: 000082638621198
GenerationID - Of.Venda	Número de geração (GenerationID) da Oferta de venda. Quando ocorre mais de uma mudança com a ordem este número é atualizado para identificar um novo estado da oferta. Ex: 000000048602375
Condição Oferta de Venda	Verifica se a ordem de venda é o iniciador da agressão ao livro de ofertas. Pode ser: 0 - Oferta Neutra - é aquela que entra no mercado e não fecha com oferta existente. 1 - Oferta Agressora - é aquela que ingressa no mercado para fechar com uma oferta existente. 2 - Oferta Agredida - é a oferta (existente) que é fechada com uma oferta agressora.
Indicador de direto	Código que identifica se o negócio direto foi intencional: 1 - Intencional / 0 - Não Intencional
Corretora Compra	Código de identificação da corretora de compra. Ex: 00000015
Corretora Venda	Código de identificação da corretora de venda. Ex: 00000085

Os arquivos de OFER\_CPA e OFER\_VDA possuem os dados organizados conforme a Tabela 2:

Tabela 2 – Layout da organização dos dados dos arquivos OFER\_CPA e OFER\_VDA informado pela B3.

Nome da Coluna	Descrição do valor da coluna
Data Sessão	Data da Sessão, é o dia de negociação do arquivo gerado. Ex: 2019-07-03
Símbolo do Instrumento	Símbolo do Instrumento, é o código do ativo. Ex: PETR4
Sentido Of.Compra / Venda	Indicador de sentido da ordem, em qual lado será feita a modificação: 1 - compra / 2 - venda
Seqüência	Número sequencial da oferta, é um valor único e identificador da ordem. Ex: 000082408193128
GenerationID - Of.Compra / Venda	Número de geração (GenerationID) da Oferta de Compra. Quando ocorre mais de uma mudança com a ordem este número é atualizado para identificar um novo estado da oferta. Ex: 000000006432071
Cód. do Evento da Of.Compra / Venda	Código do Evento da Ordem, indica qual é o tipo de ação da ordem: 1 - New / 2 - Update / 3 - Cancel - Solicitado pelo participante 4 - Trade / 5 - Reentry - Processo interno (quantidade escondida) / 6 - New Stop Price / 7 - Reject 8 - Remove - Removida pelo Sistema (final de dia ou quando é totalmente fechada) 9 - Stop Price Triggered / 11 - Expire - Oferta com validade expirada.
Hora Prioridade	Hora de registro da oferta no sistema (no formato, HH:MM:SS.NNN), utilizada como indicadora de prioridade. Ex: 09:45:00.964000
Ind. de Prioridade Of.Compra / Venda	Indicador de Prioridade. Ex: 2147483647
Preço Of. Compra / Venda	Preço da Oferta. Ex: 00000000028.940000
Qtd. Total Of. Compra / Venda	Quantidade Total de ações da Oferta. Se tiver alteração ela reflete a nova quantidade. Ex: 00000000000003900
Qtd. Negociada Of. Compra / Venda	Quantidade Negociada. Ex: 00000000000000400
Data Oferta Compra / Venda	Data de Inclusão da Oferta. Pode ser uma data anterior à Data da Sessão, quando se tratar de uma Oferta com Validade. Ex: 2019-06-21
Data de Entrada Of.Compra / Venda	Data/Hora de Entrada da Oferta (formato: DD/MM/AAAA HH:MM:SS). Ex: 2019-06-21 10:41:03
Estado Of. Compra / Venda	Indicador de estado da ordem: 0 - Novo / 1 - Negociada parcialmente / 2 - Totalmente executada 4 - Cancelada / 5 - Modificada / 8 - Rejeitada / C - Expirada
Condição Oferta	Código que identifica a condição da oferta. Pode ser: 0 - Oferta Neutra - é aquela que entra no mercado e não fecha com oferta existente. 1 - Oferta Agressora - é aquela que ingressa no mercado para fechar com uma oferta existente. 2 - Oferta Agredida - é a oferta (existente) que é fechada com uma oferta agressora.
Corretora	Código que identifica univocamente a corretora. Ex: 00000386

### 3.3 Geração dos arquivos de Mensagens e de Livro de Ofertas

Para a realização dos experimentos foram gerados dois arquivos novos, que são o arquivo Mensagens e o arquivo Livro de Ofertas. Para isso foi desenvolvido o algoritmo Get\_Data que recebe como entrada os três arquivos de log da B3 (NEG, OFER\_VDA e OFER\_CPA), o código da ação que se deseja extrair a informação, horário inicial e final da extração e o tipo e quantidade de níveis do Livro de Ofertas.

O algoritmo Get\_Data é dividido em duas etapas. A primeira etapa realiza o filtro dos dados obtidos nos arquivos originais da B3 (NEG, OFER\_VDA e OFER\_CPA) para obter apenas dados referentes ao código da ação informado. Com isso são gerados três arquivos intermediários NEG\_FILTERED, OFER\_VDA\_FILTERED e OFER\_CPA\_FILTERED e a união desses dois últimos gera um outro arquivo intermediário OFER\_MERGED.

Na seqüência, o arquivo NEG\_FILTERED é ordenado pela linha 6 da tabela 1 do

mais antigo para o mais novo. No caso dos valores serem os mesmos, utiliza-se a linha 3 da tabela 1 do menor para o maior gerando o arquivo intermediário NEG\_ORDERED. O arquivo OFER\_MERGED será ordenado do mais antigo para o mais novo de acordo com a linha 12 da tabela 2. Caso haja empate, utiliza-se a linha 7 da tabela 2 e por último a linha 5 da tabela 2 do maior para o menor. Assim, obtém-se o arquivo OFER\_ORDERED. Até este momento os arquivos intermediários estão com dados na formatação dos arquivos originais da B3 (NEG, OFER\_VDA e OFER\_CPA) sem alterações.

Com os dados filtrados e ordenados nos arquivos OFER\_ORDERED e NEG\_ORDERED, o algoritmo Get\_Data irá realizar a segunda etapa, onde ele irá iterar por todas as linhas desses arquivos verificando qual é o tipo de evento e em qual lado do livro de ofertas ocorre a modificação. Em seguida, realiza a inserção nos arquivos de Mensagens e de Livro de Ofertas. O arquivo de mensagens contém as informações provenientes do arquivo OFER\_ORDERED para eventos de inserção, modificação e exclusão e do arquivo NEG\_ORDERED para eventos de negociação. A Tabela 3 apresenta o formato e descrição dos campos do arquivo de mensagens e a Tabela 4 apresenta um exemplo com várias linhas contendo de valores no arquivo de mensagens.

Tabela 3 – Layout do arquivo de mensagens "Message".

Coluna	Descrição
instante	É o tempo da ocorrência do evento em segundos com precisão de milissegundos com referência à meia noite. Ex: 11:00:34.456 será 39634.456
lado	É o indicador de sentido da ordem, em que lado do livro de ofertas aquele evento vai alterar. 1 - compra 2 - venda 4 - negociação (negociações retiram ordens dos dois lados)
tipo	É o valor do código do evento da ordem, correspondendo à ação que aquela ordem realizará no livro de ofertas. 1 - Inserção, 2 - Alteração e 3 - Cancelamento.
price	Preço da oferta ajustado no padrão LOBSTER, ou seja, multiplicado por 10000. Ex: 145,50 é ajustado para 1455000
vol	Quantidade total do volume de ativos financeiros da oferta, se tiver alteração ela reflete a nova quantidade.
agg_c	Se for negociação (lado 4), este campo é indicador de agressividade do lado do bid, 1 - é uma ordem de bid que agrediu o ask 2 - é uma ordem de bid que foi agredida pelo o ask. Caso contrário, será 0.
agg_v	Se for negociação (lado 4), este campo é indicador de agressividade do lado do ask, 1 - é uma ordem de ask que agrediu o bid 2 - é uma ordem de ask que foi agredida pelo o bid. Caso contrário, será 0.
cross_trade	Indicador de cross-trade, ocorre quando a corretora faz a negociação entre seus clientes e reporta para a bolsa.
corretora_c	Código de identificação da corretora da ordem de bid. Se esse evento possui lado igual a 1 ou 4, caso contrário não terá valor sinalizado por -1.
corretora_v	Código de identificação da corretora da ordem de ask. Se esse evento possui lado igual a 2 ou 4, caso contrário não terá valor sinalizado por -1.
order_id_c	Código de identificação da ordem de bid. Se esse evento possui lado igual a 1 ou 4, caso contrário não terá valor sinalizado por -1.
order_id_v	Código de identificação da ordem de ask. Se esse evento possui lado igual a 2 ou 4, caso contrário não terá valor sinalizado por -1.

Tabela 4 – Exemplo da estrutura de um arquivo de Mensagens criado pelo algoritmo Get\_Data a partir do Layout da Tabela 3.

instante	lado	tipo	price	vol	agg_c	agg_v	cross_trade	corretora_c	corretora_v	order_id_c	order_id_v
36555.739	1	1	272300	100	0	0	0	308	-1	82621897804	-1
36555.748	2	2	281200	5000	0	0	0	-1	308	-1	82621897690
36555.81	2	1	278400	100	0	0	0	-1	238	-1	82621897816
36555.871	1	1	278400	300	0	0	0	238	-1	82621897853	-1
36555.871	4	1	278400	300	1	2	0	238	39	82621897853	82621886949
36555.889	1	1	278400	100	0	0	0	40	-1	82621897856	-1
36555.889	4	1	278400	100	1	2	0	40	39	82621897856	82621886949
36555.89	2	1	278800	100	0	0	0	-1	40	-1	82621897857

O fluxograma 9 mostra como são inseridas as informações na estrutura de dados do Livro de Ofertas, onde cada oferta possui as seguintes informações:

- Identificação (ID) da oferta proveniente da coluna Sequência do arquivo OFER\_ORDERED;
- Preço proveniente da coluna Preço Oferta de Compra / Venda do arquivo OFER\_ORDERED;
- Volume proveniente da subtração das colunas Qtd.Total Of.Compra / Venda e Qtd.Negociada da Oferta de Compra / Venda do arquivo OFER\_ORDERED;
- Corretora do arquivo OFER\_ORDERED;

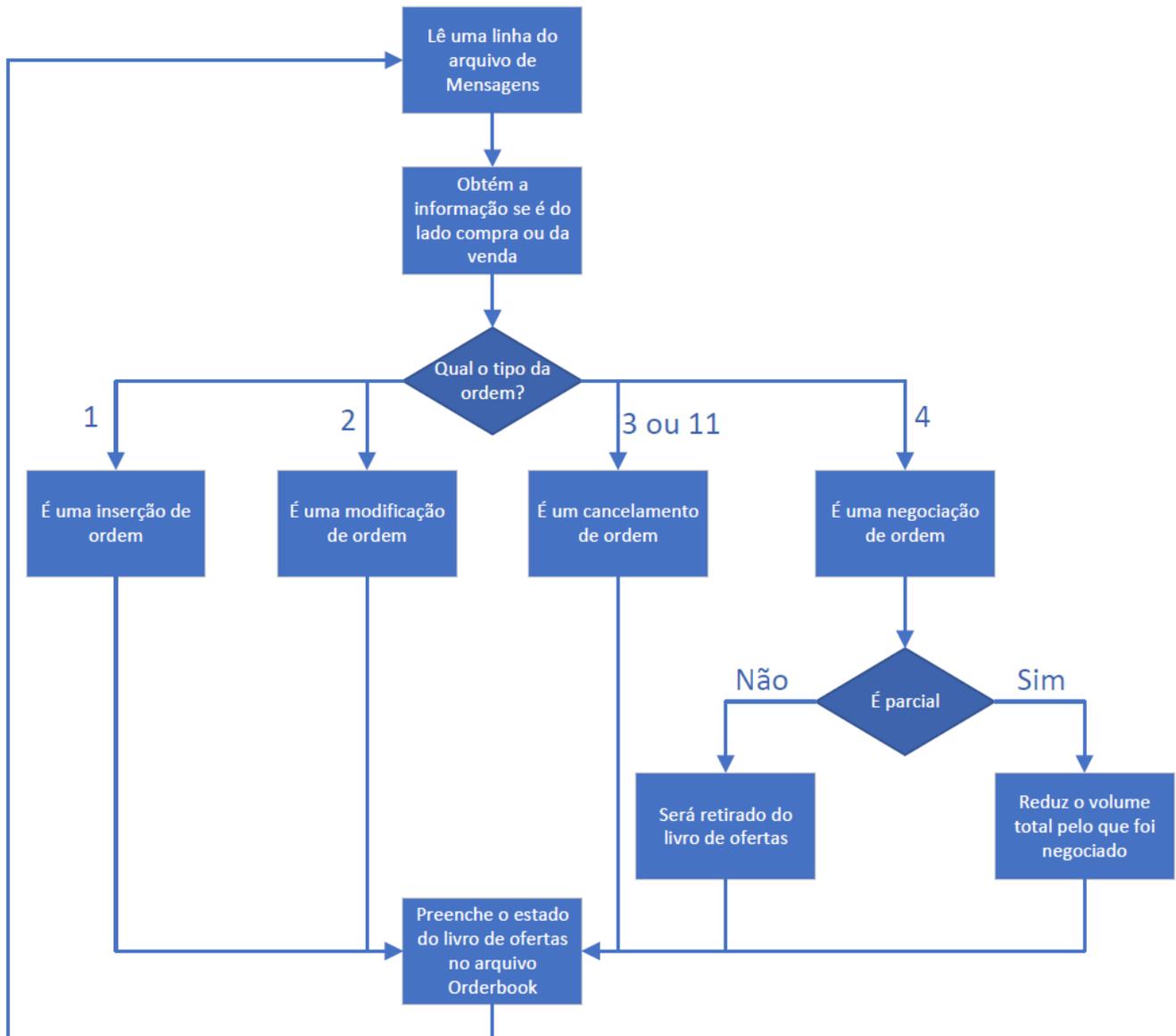


Figura 9 – Fluxograma demonstrando o fluxo para a mudança da estrutura de dados do Livro de Ofertas. Onde, é verificado qual lado será modificado e é aplicada uma das modificações realizadas por uma oferta.

Para cada leitura da mensagem do arquivo OFER\_ORDERED era verificado em qual lado seria realizada a modificação pelo valor da coluna Sentido da Oferta (linha 3 da Tabela 2), para depois com o valor da coluna Código do Evento (linha 6 da Tabela 2) verificar qual tipo de ação é realizada (inserção, modificação, cancelamento ou negociação). Com essas informações basta procurar a oferta na estrutura de dados pelo Número de Sequência (linha 4 da Tabela 2) para modificar seu preço ou volume para casos de modificação e negociação parcial ou excluir a oferta da estrutura de dados caso seja um cancelamento ou negociação total ou ainda criar uma nova em caso de inserção.

A tabela 5 realiza uma descrição dos campos do arquivo do Livro de Ofertas e a tabela 6 mostra um exemplo da estrutura do arquivo criado com 3 níveis de preço.

Tabela 5 – Descrição dos dados das colunas do arquivo do Livro de Ofertas criado pelo algoritmo Get\_Data.

Nome da coluna	Descrição dos dados da coluna
price_bid_X	É o preço da oferta de bid no nível X multiplicado por 10000.
volume_bid_X	É o volume da oferta de bid no nível X.
price_ask_X	É o preço da oferta de ask no nível X multiplicado por 10000.
volume_ask_X	É o volume da oferta de ask no nível X.

Tabela 6 – Exemplo da estrutura de um Livro de Ofertas criado pelo algoritmo exibindo 2 níveis de preço.

price_bid_1	volume_bid_1	price_ask_1	volume_ask_1	price_bid_2	volume_bid_2	price_ask_2	volume_ask_2
72600	300	72700	1165	72500	5165	72800	2860
72600	300	72700	1170	72500	5165	72800	2860
72600	300	72700	1030	72500	5165	72800	2860
72600	300	72700	1000	72500	5165	72800	2860
72600	300	72700	965	72500	5165	72800	2860

### 3.4 Dinâmica do Livro de Ofertas com o Processo de Hawkes

O preço médio é determinado pelo melhor preço do lado da compra e da venda, como consequência, a mudança deste preço reflete a perspectiva do mercado na precificação de um ativo. Para modelar este comportamento foi reproduzido o modelo de um Processo de Hawkes Multivariado de [Chen \(2017\)](#).

Para modelar a dinâmica do Livro de Ofertas é usado um Processo de Hawkes Multivariado de 4 variáveis, ou seja, a Função [2.12](#) configurado com  $M = 4$ :

$$\lambda^m(t) = \mu_m + \sum_{n=1}^4 \sum_{\{k:t_k^n < t\}} \alpha_{mn} e^{-\beta_{mn}(t-t_k^n)}. \quad (3.1)$$

Onde,  $\{t_k^n\}_{k=1,2,\dots}$  são os pontos em  $N^n(t)$  e  $m, n = 1, 2, 3, 4$ .

$\mathbf{N}(t)$  corresponde ao processo de contagem para os eventos de decremento do volume do lado da compra (*bid*), incremento do volume do lado da compra (*bid*), decremento do volume do lado da venda (*ask*) e incremento do volume do lado da venda (*ask*).

Os processos  $N^1(t)$  e  $N^3(t)$  geram eventos que decrementam o volume no lado da compra e venda, respectivamente, e apenas modelam a chegada de ordens a mercado. Isto se deve ao fato de as ordens de cancelamento serem difíceis de modelar, pois normalmente estão relacionadas com tentativas de *traders* de alta frequência explorando o mercado por informações extras. [Chen \(2017\)](#) considera tal comportamento um ruído da microestrutura do mercado e não uma mudança

no volume que possua um significado relevante, que é mostrado no trabalho de [Gould et al. \(2013\)](#).

Os processos  $N^2(t)$  e  $N^4(t)$  geram eventos que realizam um incremento do volume no lado da compra e venda, respectivamente, e consistem de ordens limitadas no primeiro nível de preço que no futuro forem observadas que são pelo menos parcialmente executadas antes de acontecer um cancelamento.

### 3.4.1 Ajustes dos parâmetros do Processo de Hawkes

Antes de fazer o ajuste dos parâmetros, o arquivo da etapa anterior é processado para retirar informações dos eventos que ocorreram no dia de negociação da seguinte forma:

- Evento `inc_bid`: todo evento de inserção do lado do *bid* que gerou negociação no futuro é salvo, contendo o respectivo *timestamp* e volume da oferta;
- Evento `inc_ask`: todo evento de inserção do lado do *ask* que gerou negociação no futuro é salvo, contendo o respectivo *timestamp* e volume da oferta;
- Evento `dec_bid`: todo evento de negociação que agrediu o lado do *bid* é salvo contendo o respectivo *timestamp* e volume da oferta;
- Evento `dec_ask`: todo evento de negociação que agrediu o lado do *ask* é salvo contendo o respectivo *timestamp* e volume da oferta;

Este modelo assume que os eventos de mercado são submetidos um por vez, excluindo a possibilidade de submissão de mais de uma ordem ao mesmo tempo. Porém, o tempo de chegada de um evento registrado pelo mercado pode coincidir dependendo da precisão do tempo usado. Como os dados da B3 possuem uma precisão de milisegundos, essa coincidência ocorre mais vezes do que os dados da NASDAQ usado por ([CHEN, 2017](#)) que são registrados em nanosegundos. Portanto, para garantir a característica de um Processo Pontual simples da Definição 5, os eventos do mesmo tipo que possuem o mesmo tempo de ocorrência serão agrupados pelo volume e considerados apenas um evento.

[Chen \(2017\)](#) considerou que para a simulação de 30 minutos do mercado seriam utilizados 1 hora de eventos anterior a este período para ajuste dos parâmetros supondo que o mercado continua possuindo o mesmo padrão de comportamento. Por exemplo, para ajustar o Processo de Hawkes Multivariado para simular o comportamento do Livro de Ofertas no período das 11:00 horas até às 11:30 horas são utilizados os eventos capturados no período das 10:00 horas até às 11:00 horas. Para simular os próximos 30 minutos, a janela de 1 hora de obtenção de eventos avançará 30 minutos e um novo ajuste é feito.

A estimativa dos parâmetros é feita utilizando a rotina *fmincon* no MATLAB que resolve otimizações numéricas para problemas com funções objetivas não-lineares multivariadas com

restrições, configurada para utilizar o algoritmo de ponto interior. Assim, a estimativa do conjunto de parâmetros  $\theta = (\mu_{1 \times 4}, \alpha_{4 \times 4}, \beta_{4 \times 4})$  dadas às observações de cada período  $\{t_k^n\}_{k=1,2,\dots}^{n=1,2,\dots,4}$  é realizada minimizando a seguinte função:

$$\min f(\theta) = -\ln L(\theta | \{t_k^n\}_{k=1,2,\dots}^{n=1,2,\dots,4}). \quad (3.2)$$

Onde, a função a ser minimizada é oriunda da Função 2.14 com  $M = 4$  tipos de eventos:

$$\ln L(\theta | \{t_k^n\}_{k=1,2,\dots}^{n=1,2,3,4}) = \sum_{m=1}^4 \ln L^m(\theta | \{t_k^n\}_{k=1,2,\dots}^{n=1,2,3,4}). \quad (3.3)$$

Para cada  $m = 1, 2, 3, 4$  é calculado:

$$\begin{aligned} \ln L^m(\theta | \{t_k^n\}_{k=1,2,\dots}^{n=1,2,3,4}) = & -\mu_m T - \sum_{n=1}^M \frac{\alpha_{mn}}{\beta_{mn}} \sum_{\{k:t_k^n < T\}} \left[ 1 - e^{-\beta_{mn}(T-t_k^n)} \right] \\ & + \sum_{\{k:t_k^n < T\}} \ln \left[ \mu_m + \sum_{n=1}^M \alpha_{mn} R_{mn}(k) \right]. \end{aligned} \quad (3.4)$$

Onde,  $T$  é o tamanho da janela e  $R_{mn}(k)$  é definido recursivamente como:

$$R_{mn}(k) = e^{-\beta_{mn}(t_k^m - t_{k-1}^m)} R_{mn}(k-1) + \sum_{\{i:t_{k-1}^m \leq t_i^n < t_k^m\}} e^{-\beta_{mn}(t_k^m - t_i^n)}. \quad (3.5)$$

Com a condição inicial:

$$R_{mn}(0) = 0. \quad (3.6)$$

Os parâmetros buscados ao executar a rotina *fmincon* estarão sujeitos às seguintes restrições:

$$\mu_m > 0, \quad \alpha_{mn} \geq 0, \quad \beta_{mn} \geq 0, \quad \rho(\Gamma) < 1.$$

Onde,  $\rho(\Gamma)$  é o raio espectral de  $\Gamma$  que pode ser obtido resolvendo numericamente os autovalores da matriz  $\Gamma$  e que é definido por Hawkes (1971b) da seguinte forma:

$$\Gamma = \left\{ \frac{\alpha_{mn}}{\beta_{mn}} \right\}_{m,n=1,2,3,4}.$$

Além disso, para cada período dentro da janela de tempo  $T$ , o algoritmo deve realizar a média aritmética dos volumes para cada um dos quatro tipos de eventos separadamente para utilizar na Simulação de Monte Carlo na Sessão 3.4.2. Onde, são obtidas as constantes  $s_M^b$  e  $s_M^a$

relacionados à quantidade média de ações que são retiradas do volume quando ocorre um evento de decréscimo do volume do lado da compra (*bid*) e do lado da venda (*ask*), respectivamente. E também, são obtidas as constantes  $s_L^b$  e  $s_L^a$  realizando a média aritmética dos eventos de acréscimo do volume do lado da compra (*bid*) e do lado da venda (*ask*), respectivamente.

Diferentemente de [Chen \(2017\)](#) que utiliza apenas duas variáveis  $s_M$  e  $s_L$ . Neste trabalho foram incluídas essas quatro variáveis, pois notou-se que ao utilizar a mesma variável para ambos os lados pode ocorrer uma tendência das previsões para uma determinada direção quando um dos lados está mais ativo (muitos eventos) que o outro lado refletido por meio da grande diferença entre a média dos volumes.

### 3.4.2 Simulação de Monte Carlo

O Algoritmo 1 mostra como ocorre as  $N$  execuções independentes da simulação de Monte Carlo que retorna a probabilidade do preço-médio subir na próxima mudança. O Algoritmo 1 recebe como entrada as seguintes informações:

- $q^a(t^*)$  é o volume do primeiro nível do lado do *ask* no tempo que está sendo observado  $t^*$ ;
- $q^b(t^*)$  é o volume do primeiro nível do lado do *bid* no tempo que está sendo observado  $t^*$ ;
- $s_M^b$  é quantidade média de ações que são retiradas do volume quando ocorre um evento de decréscimo do volume do lado da compra (*bid*);
- $s_M^a$  é quantidade média de ações que são retiradas do volume quando ocorre um evento de decréscimo do volume do lado da venda (*ask*);
- $s_L^b$  é quantidade média de ações que são inseridas no volume quando ocorre um evento de acréscimo do volume do lado da compra (*bid*);
- $s_L^a$  é quantidade média de ações que são inseridas no volume quando ocorre um evento de acréscimo do volume do lado da venda (*ask*);
- $\mu_{M \times 1}$ ,  $\alpha_{M \times M}$  e  $\beta_{M \times M}$  que são os parâmetros do Processo de Hawkes Multivariado obtido no processo de ajuste;
- $\mathbf{X}_{M \times M}^{t^*}$  é a influência do histórico de eventos apresentado na subseção 3.4.2.1;
- $N$  é o número de simulações que serão feitas no Monte Carlo.

Para simular o movimento do preço médio de dois centavos foram necessários adicionar mais as seguintes informações utilizadas no Algoritmo 2:

- $q_2^a(t^*)$  é o volume do segundo nível do lado do *ask* no tempo que está sendo observado  $t^*$ ;

- $q_2^b(t^*)$  é o volume do segundo nível do lado do *bid* no tempo que está sendo observado  $t^*$ ;
- $s_{L2}^b$  é quantidade média de ações que são inseridas no volume do segundo nível de preço quando ocorre um evento de acréscimo do volume do lado da compra (*bid*);
- $s_{L2}^a$  é quantidade média de ações que são inseridas no volume do segundo nível de preço quando ocorre um evento de acréscimo do volume do lado da venda (*ask*);

**Algoritmo 1** – Simulação de Monte Carlo com N execuções independentes para estimar a probabilidade do preço-médio subir baseado nas observações até o tempo  $t^*$

---

```

1: função alg_sim_monte_carlo( $q^a(t^*), q^b(t^*), s_M^b, s_M^a, s_L^b, s_L^a, \mu_{M \times 1}, \alpha_{M \times M}, \beta_{M \times M}, \mathbf{X}_{M \times M}^{t^*}, N$ )
2:   Inicializa  $y \leftarrow 0$   $\triangleright$  y conta o número de execuções que terminaram com movimento de
   subida
3:   para  $i \leftarrow 1$  até  $N$  faça
4:      $(Q^a(0), Q^b(0)) \leftarrow (q^a(t^*), q^b(t^*))$   $\triangleright$  Seta o estado inicial do volume do ask e bid
5:      $\mathbf{X}(0) \leftarrow \mathbf{X}(t^*)$   $\triangleright$  Seta o estado inicial do histórico de eventos
6:      $t \leftarrow 0$ 
7:     enquanto  $Q^a(0) > 0$  e  $Q^b(0) > 0$  faça  $\triangleright$  Fica no loop enquanto nenhum dos lados
   zerar
8:        $s, k \leftarrow \text{alg\_gerador\_de\_eventos}(\mu, \beta, \mathbf{X}(t))$   $\triangleright$  Obtém o tempo s e tipo k do
   evento gerado
9:       se  $k == 1$  então
10:          $Q^b(t+s) \leftarrow Q^b(t) - s_M^b$   $\triangleright$   $k == 1$  é um evento de decréscimo do bid
11:       senão se  $k == 2$  então
12:          $Q^b(t+s) \leftarrow Q^b(t) + s_L^b$   $\triangleright$   $k == 2$  é um evento de acréscimo do bid
13:       senão se  $k == 3$  então
14:          $Q^a(t+s) \leftarrow Q^a(t) - s_M^a$   $\triangleright$   $k == 3$  é um evento de decréscimo do ask
15:       senão
16:          $Q^a(t+s) \leftarrow Q^a(t) + s_L^a$   $\triangleright$   $k == 4$  é um evento de acréscimo do ask
17:       fim se
18:       para  $m = 1$  até 4 faça
19:         para  $n = 1$  até 4 faça
20:            $X_{mn}(t+s) \leftarrow X_{mn}(t) \cdot e^{-\beta_{mn}s}$   $\triangleright$  Realiza o decaimento exponencial pela
   passagem do tempo
21:         se  $k == n$  então
22:            $X_{mn}(t+s) \leftarrow X_{mn}(t+s) + \alpha_{mn}$   $\triangleright$  Realiza o pulo da ocorrência de um
   novo evento do tipo k
23:         fim se
24:       fim para
25:     fim para
26:      $t \leftarrow t + s$   $\triangleright$  Adianta o tempo em s segundos
27:   fim enquanto
28:   se  $Q^a(t) \leq 0$  então  $\triangleright$  Se zerou o lado do ask soma 1 no contador
29:      $y \leftarrow y + 1$ 
30:   fim se
31: fim para
32: return  $y/N$   $\triangleright$  Retorna a probabilidade do movimento ser de subida
33: fim função

```

---

**Algoritmo 2** – Simulação de Monte Carlo com  $N$  execuções independentes para estimar a probabilidade do preço-médio subir dois centavos baseado nas observações até o tempo  $t^*$

```

1: Inicializa  $y \leftarrow 0$                                 ▷  $y$  contador do número de execuções de subida
2: para  $i \leftarrow 1$  até  $N$  faça
3:    $(Q^a(0), Q^b(0)) \leftarrow (q^a(t^*), q^b(t^*))$     ▷ Seta o estado inicial do volume do ask e bid do
primeiro nível
4:    $(Q_2^a(0), Q_2^b(0)) \leftarrow (q_2^a(t^*), q_2^b(t^*))$  ▷ Seta o estado inicial dos volumes do segundo nível
5:    $\mathbf{X}(0) \leftarrow \mathbf{X}(t^*)$                         ▷ Seta o estado inicial do histórico de eventos
6:    $t \leftarrow 0$ 
7:   enquanto  $Q_2^a(0) > 0$  e  $Q_2^b(0) > 0$  faça ▷ Continua enquanto não zerar o segundo nível
8:      $s, k \leftarrow alg\_gerador\_de\_eventos(\boldsymbol{\mu}, \boldsymbol{\beta}, \mathbf{X}(t))$  ▷ Obtém o tempo  $s$  e tipo  $k$  do evento
gerado
9:     se  $k == 1$  então                                ▷ é um evento de decréscimo do bid
10:      se  $Q^b(0) > 0$  então                            ▷ decresce o volume do primeiro nível enquanto tiver
11:         $Q^b(t+s) \leftarrow Q^b(t) - s_M^b$ 
12:      senão                                           ▷ decresce o volume do segundo nível
13:         $Q_2^b(t+s) \leftarrow Q_2^b(t) - s_M^b$ 
14:      fim se
15:      senão se  $k == 2$  então                            ▷ é um evento de acréscimo do primeiro nível do bid
16:         $Q^b(t+s) \leftarrow Q^b(t) + s_L^b$ 
17:      senão se  $k == 3$  então                            ▷ é um evento de decréscimo do ask
18:        se  $Q^a(0) > 0$  então                            ▷ decresce o volume do primeiro nível enquanto tiver
19:           $Q^a(t+s) \leftarrow Q^a(t) - s_M^a$ 
20:        senão                                           ▷ decresce o volume do segundo nível
21:           $Q_2^a(t+s) \leftarrow Q_2^a(t) - s_M^a$ 
22:        fim se
23:        senão se  $k == 4$  então                            ▷ é um evento de acréscimo do primeiro nível do ask
24:           $Q^a(t+s) \leftarrow Q^a(t) + s_L^a$ 
25:        senão se  $k == 5$  então                            ▷ é um evento de acréscimo do segundo nível do bid
26:           $Q_2^b(t+s) \leftarrow Q_2^b(t) + s_{L2}^b$ 
27:        senão                                           ▷ é um evento de acréscimo do segundo nível do ask
28:           $Q_2^a(t+s) \leftarrow Q_2^a(t) + s_{L2}^a$ 
29:        fim se
30:        para  $m = 1$  até 4 faça
31:          para  $n = 1$  até 4 faça
32:             $X_{mn}(t+s) \leftarrow X_{mn}(t) \cdot e^{-\beta_{mn}s}$     ▷ Realiza o decaimento exponencial pela
passagem do tempo
33:            se  $k == n$  então
34:               $X_{mn}(t+s) \leftarrow X_{mn}(t+s) + \alpha_{mn}$     ▷ Realiza o pulo da ocorrência de um
novo evento do tipo  $k$ 
35:            fim se
36:          fim para
37:        fim para
38:         $t \leftarrow t + s$                                 ▷ Adianta o tempo em  $s$  segundos
39:      fim enquanto
40:      se  $Q_2^a(t) \leq 0$  então                            ▷ Se zerou todo o lado do ask soma 1 no contador
41:         $y \leftarrow y + 1$ 
42:      fim se
43:    fim para
44:  return  $y/N$                                           ▷ Retorna a probabilidade do movimento ser de subida

```

## 3.4.2.1 Simulação da ocorrência do próximo evento

Nesta seção será mostrada como ocorre a criação de um evento utilizado na linha 8 do Algoritmo 1. Para realizar o cálculo da função intensidade do Processo de Hawkes Multivariado é feita a soma sobre todos os pontos, o que possui complexidade  $O(n^2)$ . Porém, utilizando o *kernel* exponencial não é necessário saber toda a história para calcular a função intensidade por conta da Propriedade Markoviana apresentada por [Abergel et al. \(2016\)](#), assim a Função 2.12 pode ser reescrita como:

$$\lambda^m(t) = \mu_m + \sum_{n=1}^M X_{mn}(t). \quad (3.7)$$

Onde,

$$X_{mn}(t) = \int_{s < t} \alpha_{mn} e^{-\beta_{mn}(t-s)} dN^n(s). \quad (3.8)$$

Note que,

$$\begin{aligned} X_{mn}(t) &= \int_{s \leq 0} \alpha_{mn} e^{-\beta_{mn}(t-s)} dN^n(s) + \int_0^t \alpha_{mn} e^{-\beta_{mn}(t-s)} dN^n(s) \\ &= e^{-\beta_{mn}(t)} \int_{s \leq 0} \alpha_{mn} e^{-\beta_{mn}(0-s)} dN^n(s) + \int_0^t \alpha_{mn} e^{-\beta_{mn}(t-s)} dN^n(s) \\ &= e^{-\beta_{mn}(t)} X_{mn}(0^+) + \int_0^t \alpha_{mn} e^{-\beta_{mn}(t-s)} dN^n(s). \end{aligned} \quad (3.9)$$

Então, a Função 2.12 pode ser reescrita como:

$$\lambda^m(t) = \mu_m + \sum_{n=1}^M \left[ e^{-\beta_{mn}(t)} X_{mn}(0^+) + \int_0^t \alpha_{mn} e^{-\beta_{mn}(t-s)} dN^n(s) \right]. \quad (3.10)$$

$X_{mn}(0^+)$  pode ser considerado uma representação do histórico inicial da influência de todos eventos do tipo n na função intensidade do tipo m. O tempo atual  $t$  é assumido como zero, então a Função Intensidade será representada por:

$$\lambda^m(t) = \mu_m + \sum_{n=1}^M e^{-\beta_{mn}(t)} X_{mn}(0^+). \quad (3.11)$$

Assim, para cada novo cálculo da Função Intensidade quando ocorre um novo evento não será necessário fazer o cálculo de todo o histórico novamente. Existem duas formas para escolher o estado inicial do parâmetro  $\mathbf{X}_{M \times M}$ :

- Começar sem história onde todos os valores da matriz  $\mathbf{X}_{M \times M}$  são zero;
- Utilizar um período fixo onde os eventos localizados dentro desse período são levados em consideração na influência da função intensidade;

Como apontado por [Chen \(2017\)](#), utilizar todos os eventos que ocorreram durante um dia é uma forma ineficiente pois com o passar do tempo mais e mais eventos são acumulados. Por outro lado, o efeito de cada evento é decrescente no tempo por conta do decaimento exponencial do *kernel* com o passar do tempo. Suponha que o tempo atual é zero, qualquer evento que aconteceu antes desse tempo possui o seguinte efeito remanescente na função intensidade  $\lambda^m(0)$ :

$$\alpha_{mn} e^{-\beta_{mn}(0-\tau_{mn}^\varepsilon)} = \varepsilon. \quad (3.12)$$

O cálculo do tempo máximo que um evento deve estar no passado para possuir o efeito remanescente de  $\varepsilon$  e ser contabilizado no histórico é:

$$\tau_{mn}^\varepsilon = \frac{\ln\left(\frac{\varepsilon}{\alpha_{mn}}\right)}{\beta_{mn}}. \quad (3.13)$$

O valor de  $\tau_{mn}^\varepsilon$  é obtido para todos os tipos  $m, n = 1, 2, 3, 4$  e se obtém o valor máximo  $\tau^\varepsilon$  para que todos os eventos contabilizados no cálculo tenham um efeito maior que  $\varepsilon$ :

$$\tau^\varepsilon = \min_{m,n} \tau_{mn}^\varepsilon = -\max_{m,n} |\tau_{mn}^\varepsilon|. \quad (3.14)$$

Assim, a Equação 3.9 pode ser transformada da seguinte forma para representar os valores iniciais de  $\mathbf{X}_{M \times M}^t$  no tempo  $t$  dados os eventos ocorridos no período  $t - \tau^\varepsilon$  e  $t$ :

$$X_{mn}^t = \sum_{\{k: t-\tau^\varepsilon \leq t_k^n < t\}} \alpha_{mn} e^{-\beta_{mn}(t-t_k^n)}. \quad (3.15)$$

Com isso, criou-se o Algoritmo 3 replicado do Algoritmo 5 do trabalho de [Chen \(2017\)](#), que com o histórico já calculado de  $\mathbf{X}_{M \times M}$  por meio de 3.15 e os parâmetros  $\boldsymbol{\mu}_{M \times 1}$  e  $\boldsymbol{\beta}_{M \times M}$  encontrados na etapa de ajuste de parâmetros, a simulação para criação de novos eventos futuros é feita da seguinte forma:

---

**Algoritmo 3** – Algoritmo de simulação da ocorrência do próximo evento do tipo  $k$  em um Processo de Hawkes  $M$ -variado

---

```

1: função alg_gerador_de_eventos( $\boldsymbol{\mu}_{M \times 1}, \boldsymbol{\beta}_{M \times M}, \mathbf{X}_{M \times M}$ )
2:   Inicializa  $s = 0$  ▷ Tempo atual
3:   enquanto True faça ▷ Continua no loop enquanto não aceitar um evento
4:      $\bar{\lambda} \leftarrow \sum_{m=1}^M \lambda^m(s) = \sum_{n=1}^M [\mu_m + e^{-\beta_{mn}(s)} X_{mn}]$  ▷ Obtém a função intensidade no instante atual
5:      $u \leftarrow \text{uniforme}(0, 1)$ 
6:      $w \leftarrow -\ln(u)/\bar{\lambda}$  ▷ Cria o período de tempo  $w$  com distribuição exponencial
7:      $s \leftarrow s + w$  ▷ Gera o próximo ponto candidato
8:      $D \leftarrow \text{uniforme}(0, 1)$  ▷ Probabilidade de o ponto ser aceito
9:     se  $D\bar{\lambda} \leq \sum_{m=1}^M \lambda^m(s) = \sum_{n=1}^M [\mu_m + e^{-\beta_{mn}(s)} X_{mn}]$  então
10:        $k = 1$ 
11:       enquanto  $D\bar{\lambda} > \sum_{m=1}^k \lambda^m(s)$  faça ▷ Procura pelo primeiro  $k$  em que retorne falso na condição
12:          $k \leftarrow k + 1$ 
13:       fim enquanto
14:       return  $s, k$  ▷ O próximo evento será do tipo  $k$  no tempo  $s$ 
15:     fim se
16:   fim enquanto
17: fim função

```

---

### 3.4.3 Implementação de estratégia de negociação

Para testar a eficiência e lucratividade do método de previsões do Processo de Hawkes foi replicado um algoritmo simples de negociações do trabalho de [Chen \(2017\)](#), apresentado em azul na Figura 10. Para isso utilizou-se os arquivos de Mensagens e de Livro de Ofertas, onde para cada linha dos arquivos observa-se a diferença dos melhores preços de compra e venda é de um centavo. Se essa condição for satisfeita, a simulação de Monte Carlo gera  $N$  execuções e no final retorna a probabilidade " $p$ " do preço-médio subir, caso contrário, passa para a próxima linha.

Com isso, é feita a verificação se a probabilidade é maior que o *threshold* " $\eta$ " que define a probabilidade mínima para que seja considerado um sinal de subida. Para definir um sinal de descida é utilizado o *threshold* " $1 - \eta$ ", pois as simulações que finalizaram com o lado do *bid* zerado estão contabilizadas no complemento da probabilidade  $p$ .

Com o tipo de sinal definido é verificado qual posição está sendo tomada. Se é um sinal de subida e no momento o algoritmo não possui ações ou esteja comprado, será realizada a compra de uma ação. Caso o algoritmo estiver vendido apostando na queda do preço, a posição é zerada. Se é um sinal de descida e no momento o algoritmo não possui ações ou esteja vendido, será realizada a venda de uma ação. Caso o algoritmo estiver comprado apostando na subida do

preço, a posição é zerada. Após a efetivação de uma ordem de compra ou venda, o algoritmo entra em um *cooldown* de  $\Delta t$  segundos para evitar uma supervalorização da estratégia.

Porém, essa estratégia de negociações de Chen (2017) apresentou alguns problemas. Como exemplo, zerar uma posição já acumulada por alguma previsão no sentido contrário que surgia antes de ocorrer a mudança do preço-médio perdendo dinheiro para no *spread*, por exemplo, comprando várias ações pelo preço de R\$13,96 e vendendo por R\$13,95. Outro problema que pode ocorrer é logo após uma mudança do preço-médio, quando o *spread* retorna para uma diferença de 1 centavo normalmente um dos lados possui poucas ações onde uma ou duas negociações zerariam o volume, podendo gerar um sinal contrário finalizando uma posição correta que estava formada e poderia dar continuidade ao movimento do mercado.

Por isso, foram adicionadas mais alguns passos e verificações apresentadas em verde na Figura 10 para amenizar esses efeitos. Adicionou-se um *cooldown* de  $\Delta s$  segundos condicionada a mudança do preço-médio e também um novo contador de previsões que ao perceber o início de uma tendência de três previsões consecutivas gera um sinal na direção da tendência para evitar o zeramento da posição pela influência de apenas um sinal contrário.

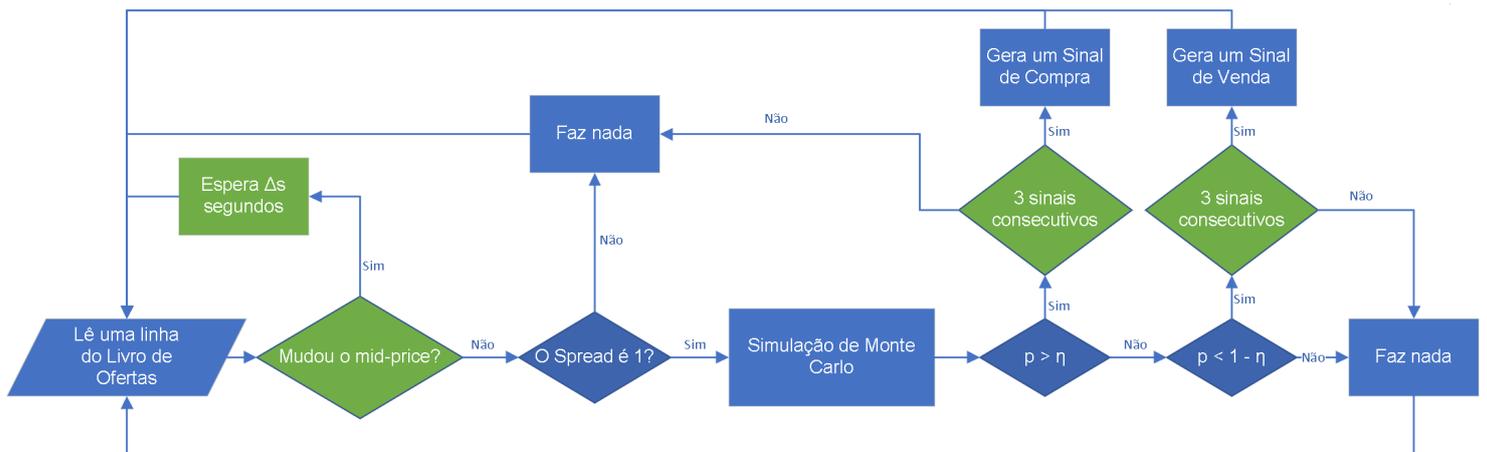


Figura 10 – Estratégia de negociação.

---

## RESULTADOS

---

### 4.1 Considerações iniciais

Neste capítulo serão apresentados informações do período escolhido para testes, como é feito o ajuste dos parâmetros, os resultados dos testes de execuções do algoritmo de negociações com o Processo de Hawkes, comparações com um algoritmo utilizando Redes Neurais apresentado por [Bileki \(2021\)](#), com a previsão usando uma variável aleatória que simula o lançamento de uma moeda e a extensão do Processo de Hawkes com dois níveis.

### 4.2 Período de testes

Os resultados foram obtidos a partir de experimentos conduzidos em um período de 23 dias de negociações do ativo PETR4 (Petrobras), compreendendo todo o mês de Outubro de 2019. Este mês utilizado apresentou uma volatilidade anualizada de 23,45% comparando com as volatilidades anualizadas de 29,71% do período de Abril de 2019 até Setembro de 2019 e de 79,57% do período de Novembro de 2019 até Abril de 2020, isso mostrou que foi utilizado um período com mais estabilidade dos preços, evitando deste modo avaliar dados de momentos de grandes volatilidade causados por eventos.

### 4.3 Ajustes dos Parâmetros

A rotina *fmincon* do MATLAB foi configurada da seguinte forma:

- *Algorithm*: o algoritmo de minimização utilizado foi o *interior-point*, o mesmo configurado no trabalho de [Chen \(2017\)](#);
- *MaxIter*: número de iterações máxima configurada com 100;

- *MaxFunEvals*: número máximo de avaliação de funções configurada com 100000.

O ajuste de parâmetros de cada um dos nove períodos de 1 hora ( $T = 3600s$ ) com janela deslizante de 30 minutos, como descrito na Subseção 3.4.1, é feita entre às 10:00 e 15:00 que são utilizados para a simulação de negociações no período das 11:00 até 15:30. A simulação das negociações é realizada neste período para evitar as instabilidades dos leilões de abertura e fechamento.

Durante a realização dos ajustes, percebeu-se uma demora de 12 horas chegando até 2 dias para obter todos os parâmetros de todos os períodos. Isso deve-se pois para cada avaliação de função das iterações da rotina *fmincon* do MATLAB é realizado o cálculo da função 2.14 de log-verossimilhança. Analisando esta função, percebe-se que esse cálculo possui complexidade  $O(n^3)$  que é dependente do número de eventos de todos os tipos dentro do intervalo de 1 hora ou 3600 segundos e da quantidade de tipos de eventos que estão sendo observados.

## 4.4 Resultados da execução diária de negociação

Em relação a estratégia de negociação, foi adotado o algoritmo da Seção 3.4.3 configurado com as seguintes informações:

- $N$ : número de execuções dentro da Simulação de Monte Carlo configurado com o valor 200. Assim, a influência de cada execução na porcentagem total da previsão é de 0,5%;
- $\eta$ : *threshold* para que um sinal ser aceito é de 0.97. Assim, um sinal apenas é utilizado se mais de 97% das execuções da Simulação de Monte Carlo indicarem o mesmo sentido, baseado nos estudos da variação do  $\eta$  apresentado no trabalho de Chen (2017);
- $\Delta t$ : tempo de *cooldown* após realizar uma ação de compra ou venda foi configurado com o valor de 10 segundos;
- $\varepsilon$ : impacto mínimo da influência de evento no passado foi configurado com o valor 0.001 como configurado no trabalho de Chen (2017).

Para o algoritmo com a estratégia atualizada apresentada na Seção 3.4.3 é configurado também o parâmetro  $\Delta s$  com o valor de 1 segundo de *cooldown* após a ocorrência de uma mudança do preço-médio. Após a execução são criados os arquivos CM e Intraday. O primeiro possui as informações da matriz de confusão das previsões de subida e descida e o segundo possui informações ao longo do tempo do preço-médio, sinais de compra e venda realizados pelo algoritmo, o valor do Lucro/Prejuízo e quantidade de ações que o algoritmo possui.

Com o arquivo CM é possível criar a matriz de confusão apresentada pela Figura 11, onde o eixo y são os valores preditos pelo algoritmo, o eixo x são os valores que realmente

aconteceram. Verdadeiro indica um sinal positivo do algoritmo para subida do preço-médio e falso é um sinal negativo do algoritmo para a descida do preço-médio.

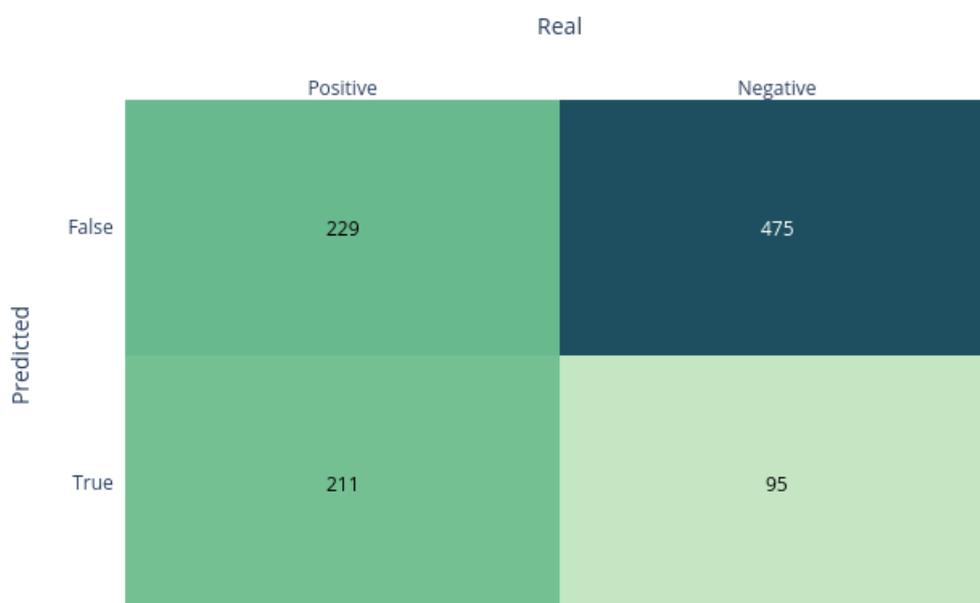


Figura 11 – Matriz de confusão das previsões usando a estratégia simples do dia de negociações 01/10/2019.

O gráfico com as métricas de acurácia (coluna azul), *recall* (coluna laranja), precisão (coluna verde) e F-measure (coluna roxa) apresentada na Figura 12 também é criado utilizando o arquivo *CM*. Onde:

- Acurácia é calculada utilizando a razão entre a quantidade de classificações verdadeiras e a quantidade total de classificações e mostra o desempenho geral do algoritmo;
- Precisão é calculada utilizando a razão entre as classificações verdadeiras e o real verdadeiro para uma determinada classe. Isso é útil para saber o quanto o algoritmo é preciso na estimativa de subida;
- *Recall* é calculado utilizando a razão das classificações real verdadeiras e a soma das classificações reais verdadeiras e falsos negativos, assim indicando o quão é sensível o algoritmo;
- *F-measure* é calculado realizando a média harmônica da precisão e *recall* para verificar quão acurado é o algoritmo. São atribuídos mais pesos para falsos negativos e falsos

positivos, enquanto não deixa um grande número de verdadeiros negativos influenciarem o resultado, como ocorre no cálculo da acurácia.

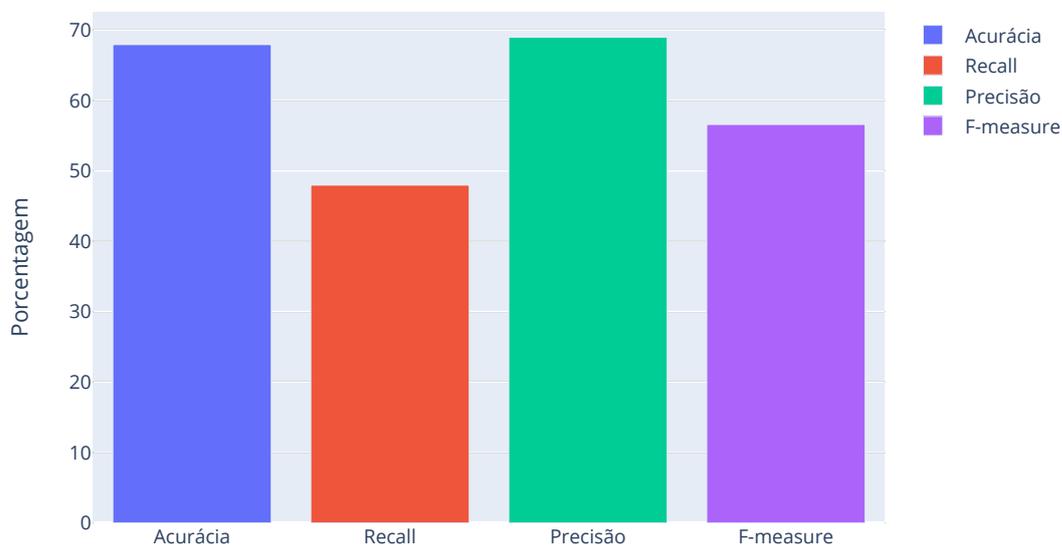


Figura 12 – Métricas do desempenho da previsão do Processo de Hawkes no dia de negociações 01/10/2019.

Os dados presentes no arquivo *Intraday data* gera a Figura 13 com o eixo x representando a passagem do tempo em segundos. O gráfico superior apresenta o movimento do preço-médio em reais e os símbolos verdes e vermelhos correspondem aos momentos em que ocorrem um sinal de compra ou de venda, respectivamente. O gráfico central mostra a variação do lucro/prejuízo em reais. O gráfico inferior mostra a posição em quantidade de ações, onde positivo significa comprado e negativo vendido. Para analisar os dados de *intraday* de outros dias de negociação são referenciados no anexo A que possui os resultados utilizando a estratégia simples e no anexo B utilizando a estratégia modificada.

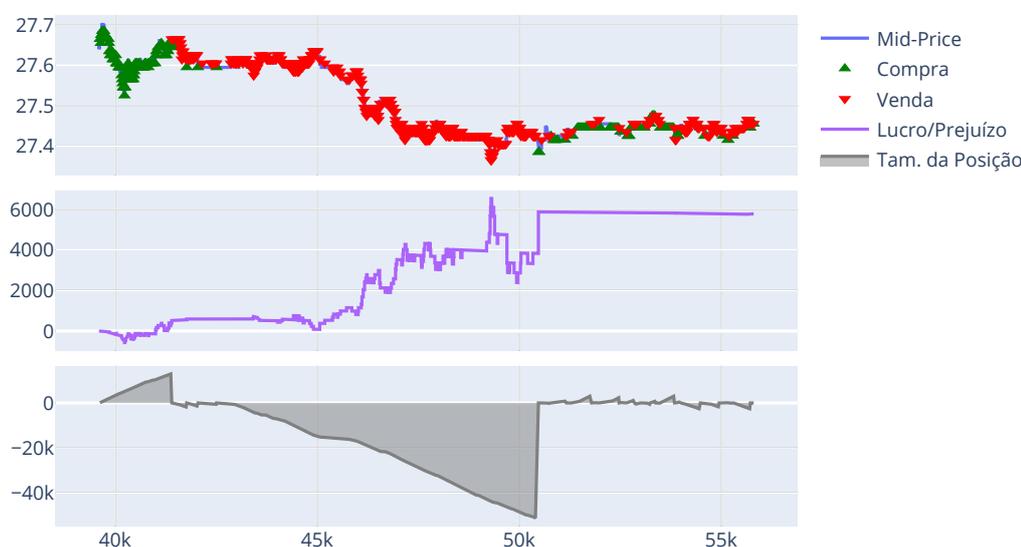


Figura 13 – Ações realizadas pela estratégia atualizada no dia de negociações 01/10/2019.

Um problema encontrado durante as execuções diárias foi o tempo de demora nas execuções, que levava em torno de 1 dia chegando até a extrapolar 1 semana para a execução finalizar. Isto se deve ao fato de que para realizar as previsões é utilizado a Função Intensidade do Processo de Hawkes Multivariado. É observado que a função 2.12 possui uma complexidade  $O(n^2)$  e seu tempo de execução é dependente do número de tipo sendo observados e da quantidade de eventos que são observados. Apesar da simplificação apresentada pela função 3.11, ainda é necessário realizar o cálculo das influências passadas que estão sendo usadas como estado inicial dos algoritmos 3 e 1 por meio da função 3.15. Lembrando que o cálculo da função 3.15 é feito para cada Simulação de Monte Carlo em cada linha do arquivo de Livro de Oferta.

## 4.5 Resultados das execuções do mês de Outubro de 2019

A Figura 14 mostra as métricas de acurácia (coluna azul), *recall* (coluna laranja), precisão (coluna verde) e F-measure (coluna roxa) de cada estratégia para os dias do mês de Outubro de 2019, onde negociações foram feitas no período das 11:00 horas até as 15:30 para evitar as instabilidades dos leilões de abertura e fechamento.

Percebe-se que as estatísticas de ambas estratégias não se alteram muito com a acurácia

e a precisão, ficando em torno de 60% à 80% com alguns dias ficando acima de 80%. As Figuras 17 e 19 mostram o ganho ou perda por dia dos algoritmos de Chen (2017) e o aprimorado neste trabalho supondo que os algoritmos começam o dia com nenhum dinheiro, enquanto as Figuras 18 e 20 apresentam o lucro ou prejuízo acumulado com o passar do tempo. Observa-se uma característica conservadora da estratégia atualizada com relação ao algoritmo de Chen (2017) fazendo com que o prejuízo de R\$ 4617,00 diminuisse para um prejuízo de R\$ 1667,00 sem contabilizar custos de transação. Isto se deve porque a estratégia atualizada realiza negociações somente quando há uma tendência de 3 sinais iguais para criar um sinal de compra ou venda. Isso gera menos eventos e a queda no número de eventos pode ser observada nas Figuras 15 e 16.

Este fenômeno de alta taxa de precisão e baixa lucratividade se deve ao fato do Algoritmo 1 que realiza previsões acertar o próximo movimento do preço médio, porém o movimento de 1 centavo não garante lucro. Por exemplo, se a previsão verifica que vai subir e compra uma ação, e após a subida do preço médio realizar a venda, ele estará vendendo no mesmo preço que comprou.

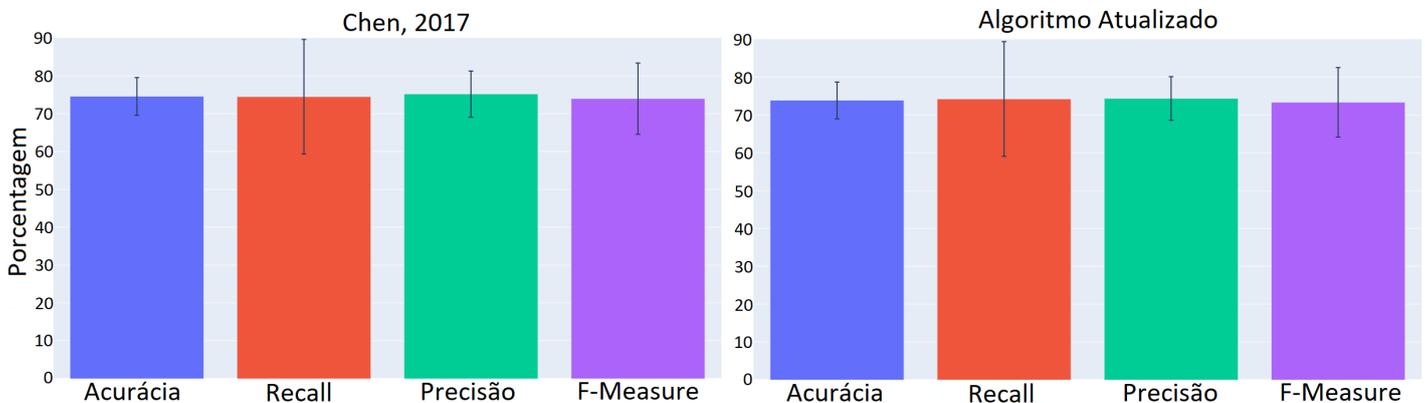


Figura 14 – Métricas das previsões realizadas pelo processo de Hawkes usando a simples no mês de Outubro de 2019.

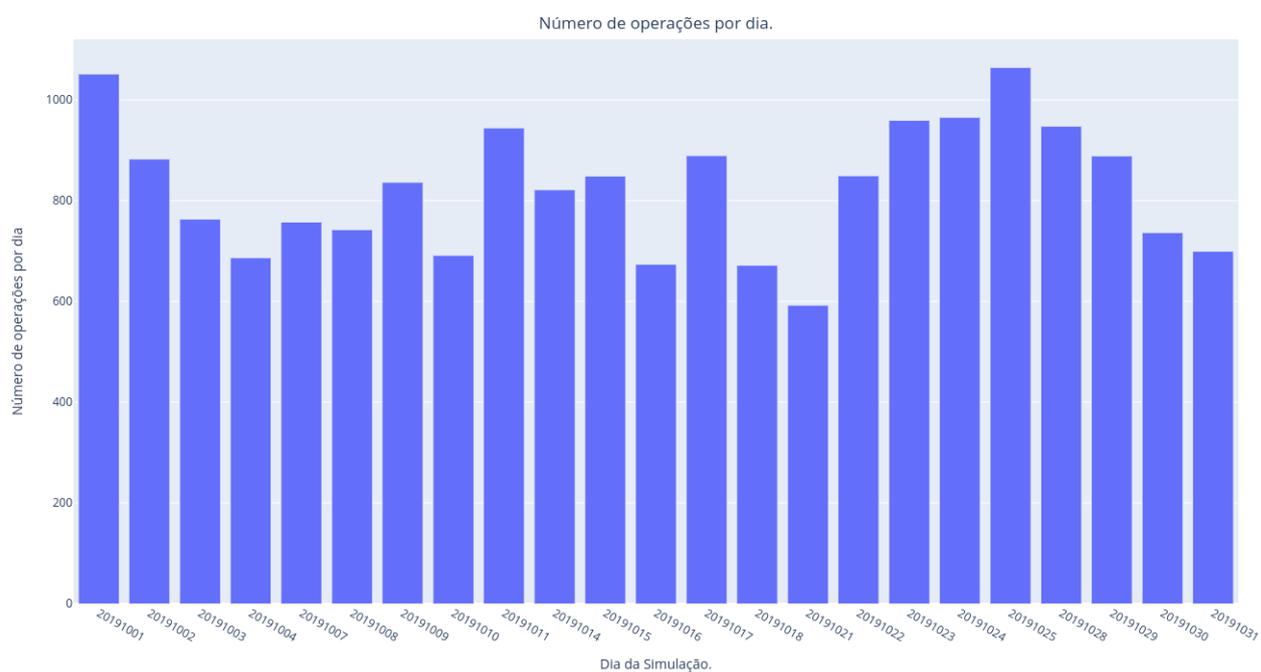


Figura 15 – Número de operações de compra/venda feita pela estratégia simples no mês de Outubro de 2019.

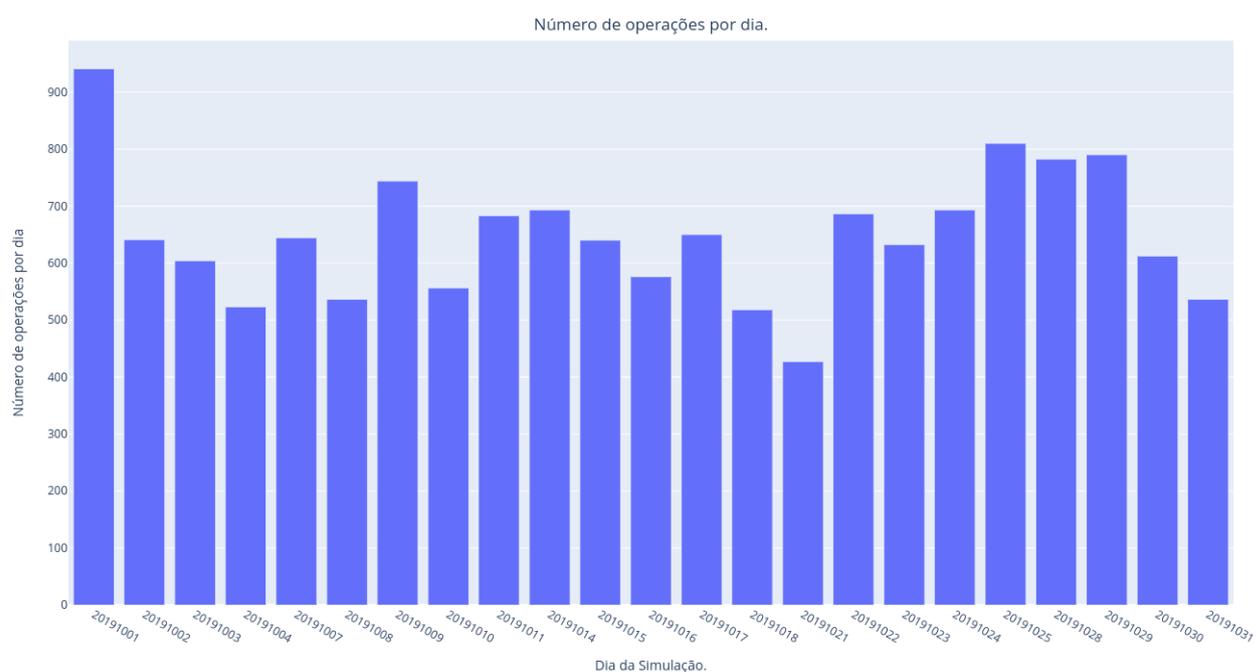


Figura 16 – Número de operações de compra/venda feita pela estratégia alterada no mês de Outubro de 2019.



Figura 17 – Lucro e prejuízo diário feito pela estratégia de Chen (2017) no mês de Outubro de 2019.

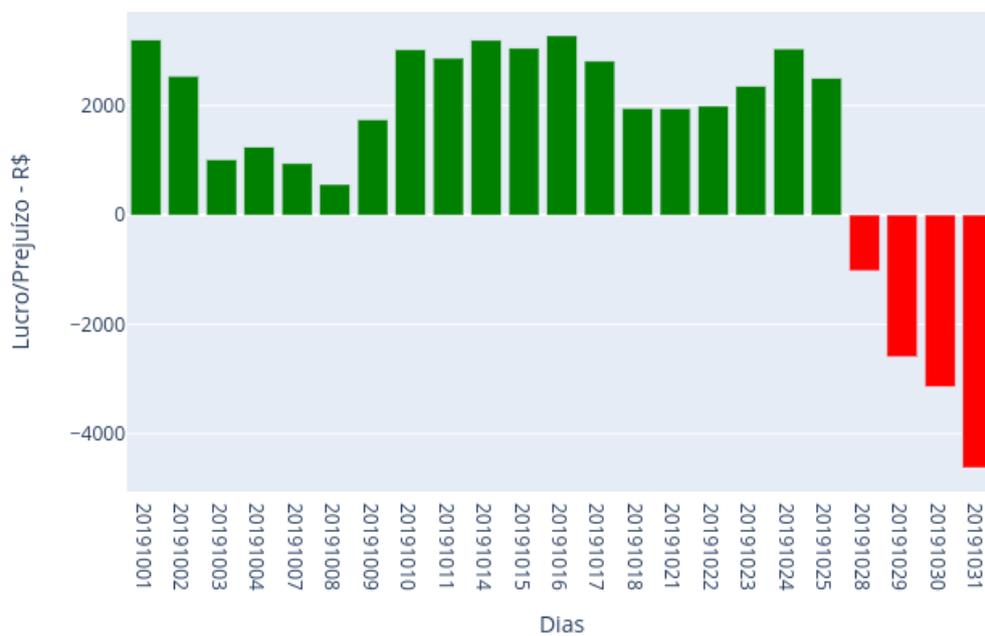


Figura 18 – Lucro e prejuízo acumulado feito pela estratégia de Chen (2017) no mês de Outubro de 2019.



Figura 19 – Lucro e prejuízo diário feito pela estratégia aprimorada no mês de Outubro de 2019.

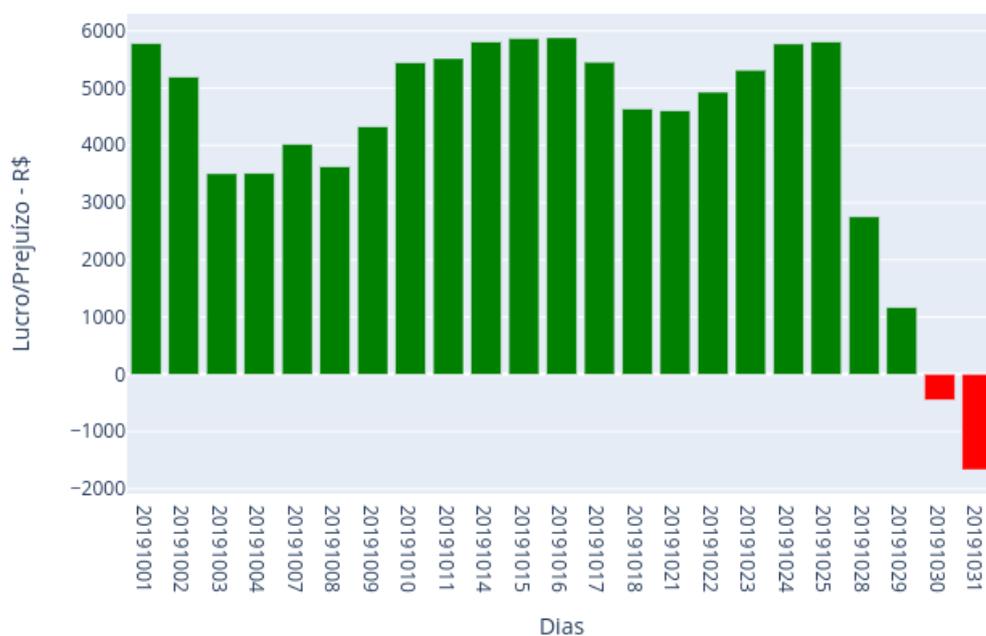


Figura 20 – Lucro e prejuízo acumulado feito pela estratégia aprimorada no mês de Outubro de 2019.

## 4.6 Comparação com técnica de Rede Neural

A fim de comparar a eficiência da previsão gerada pela Simulação de Monte Carlo com soluções implementadas via estimativas geradas por redes neurais, foram realizados testes no mesmo período de Outubro de 2019 com a rede neural de [Bileki \(2021\)](#). Essa rede neural é uma solução híbrida que usa a convolução de uma rede CNN (*Convolutional Neural Network*) para extrair vetores de características acoplado a um classificador baseado em árvore de decisão.

A forma da rede neural de [Bileki \(2021\)](#) realizar a rotulagem do movimento futuro do preço-médio é feita considerando a relação  $l_t$  da média dos  $k$  preços-médios antes e após um instante  $t$  e um limiar  $\alpha$  para definir se o sentido do movimento é para cima, para baixo ou estacionário. O cálculo do preço-médio  $p$  no momento  $t$  é feito com a média aritmética do melhor preço lado do bid e do ask. A média  $m_-$  dos valores dos  $k$  preços-médios anteriores ao tempo  $t$  e a média  $m_+$  dos valores dos  $k$  preços-médios posteriores ao tempo  $t$  são calculadas pelas seguintes equações:

$$m_-(t) = \frac{1}{k} \sum_{i=0}^k p_{t-i} \quad (4.1)$$

$$m_+(t) = \frac{1}{k} \sum_{i=1}^k p_{t+i} \quad (4.2)$$

Com isso, obtém-se a relação  $l_t$  entre as médias  $m_-(t)$  e  $m_+(t)$ :

$$l_t = \frac{m_-(t) - m_+(t)}{m_+(t)} \quad (4.3)$$

Após o cálculo da Função 4.3 é feita uma comparação do limiar  $\alpha$ , se  $l_t > \alpha$  o movimento é classificado como subida, se  $l_t < -\alpha$  é classificado como descida e caso não satisfaça nenhuma dessas condições é classificada como estacionária. Os valores de  $k$  e  $\alpha$  foram definidos individualmente para cada um dos dias de teste para possuir uma distribuição de classificações parecida com o algoritmo de previsão do processo de Hawkes.

A comparação foi feita apenas levando em consideração previsões de subida e descida, ignorando as previsões rotuladas como estacionárias pela rede neural. Isso se deve ao fato do Hawkes não possuir tal classe. Foram utilizados 20 dias do mês de Outubro de 2019 gerando a Figura 21 que mostram as médias e desvio padrão para cada uma das métricas de Acurácia, Recall, Precisão e F-measure para ambos algoritmos.

Foi utilizado a estratégia simples para verificar se utilizar apenas a informação da previsão para gerar o sinal de compra ou venda era suficiente para implementação em um cenário real e evitar que a estratégia atualizada desenvolvida visando melhorar a efetividade apenas da previsão do Processo de Hawkes não gerasse uma vantagem sobre a Rede Neural. Assim, ao utilizar a

estratégia simples de negociações no período das 11:00 horas até as 15:30 sem considerar custos de transações, o algoritmo de previsões da Rede Neural apresentou uma média de lucro/prejuízo de R\$ -35655,80 e uma média de 32831 negociações por dia, enquanto o algoritmo de previsões usando Processo de Hawkes apresentou uma média de lucro/prejuízo de R\$ -146,95 e uma média de 699 negociações por dia.

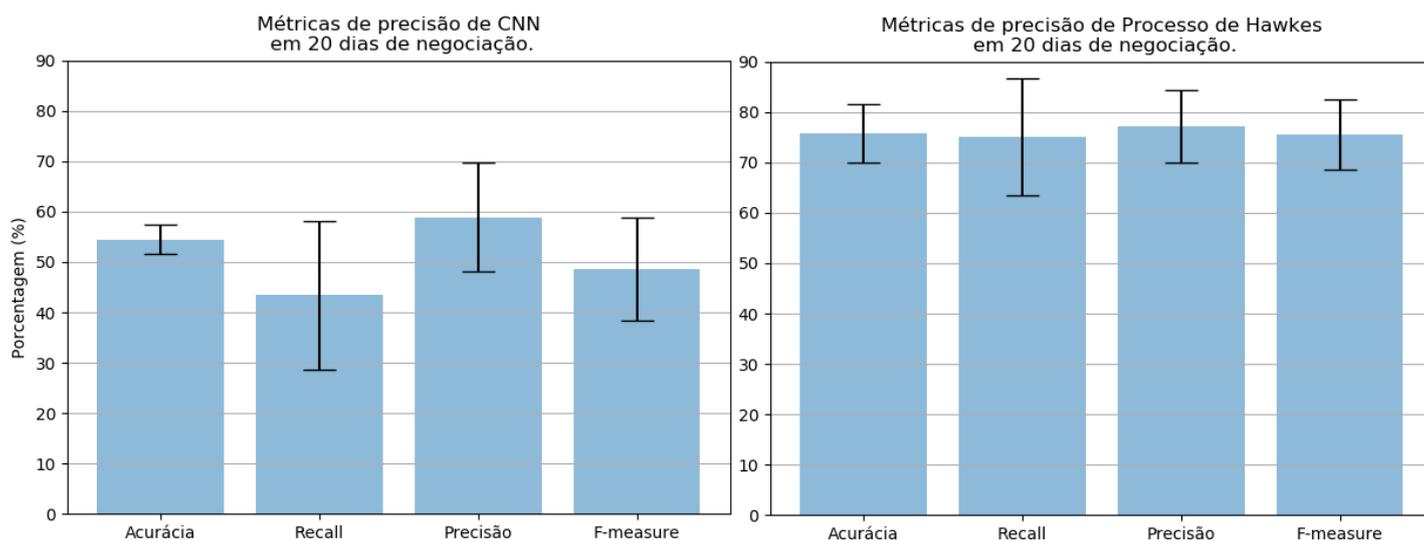


Figura 21 – Comparação das métricas de previsão da Rede Neural de [Bileki \(2021\)](#) e o Processo de Hawkes.

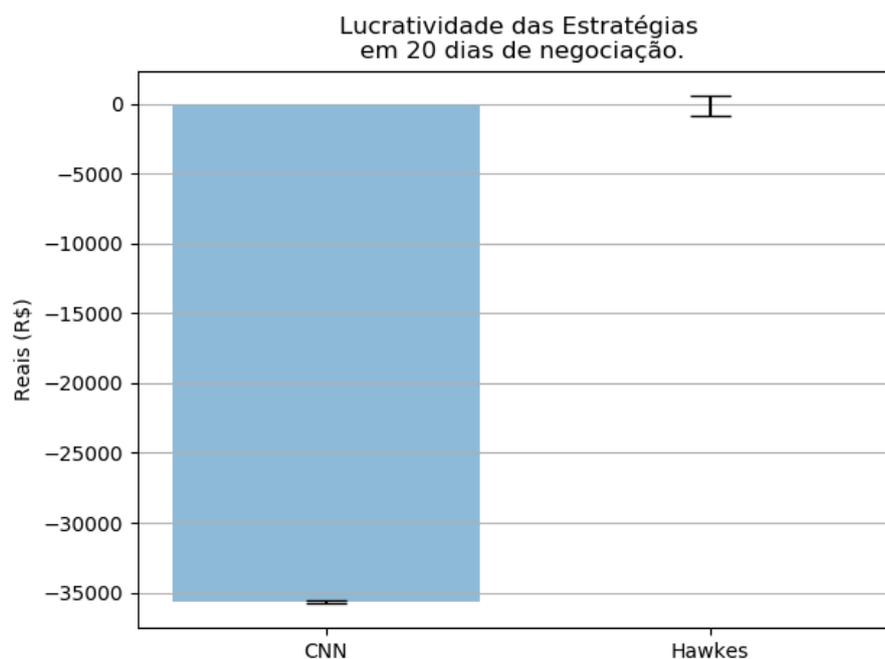


Figura 22 – Lucratividade média dos algoritmos de previsão da Rede Neural de [Bileki \(2021\)](#) e do Processo de Hawkes.

Esses resultados mostram que a capacidade preditiva para indicar a direção da próxima mudança do preço-médio do algoritmo utilizando a modelagem do livro de ofertas com Processo de Hawkes é superior à da Rede Neural proposta por [Bileki \(2021\)](#). Os resultados mostraram que a abordagem com Rede Neural é muito mais agressiva na geração de sinais que o Processo de Hawkes que tem uma abordagem mais conservadora, porém ambos acabaram no prejuízo muito por conta da utilização das previsões de forma direta pelo estratégia simples sem auxílio de outros indicadores.

## 4.7 Comparação com previsões aleatórias

O Processo de Hawkes apesar da boa taxa de acerto apresentou prejuízos no longo prazo, portanto foi conduzido um teste comparando com as previsões de uma variável aleatória que simulava o lançamento de uma moeda, que possui um esforço computacional muito menor. Para isso, em cada momento em que o Processo de Hawkes realizava uma previsão também era verificado a saída da variável aleatória para subida ou descida do movimento do preço médio.

A Figura 23 mostra que o Processo de Hawkes possui uma precisão e acurácia média perto de 80% o que é bem superior ao 50% da variável aleatória. A Figura 24 mostra que o Processo de Hawkes apesar de gerar um prejuízo, ele é em média inferior ao da variável aleatória e ainda possui dias de lucro em contraste à variável aleatória que é constante em obter prejuízos.

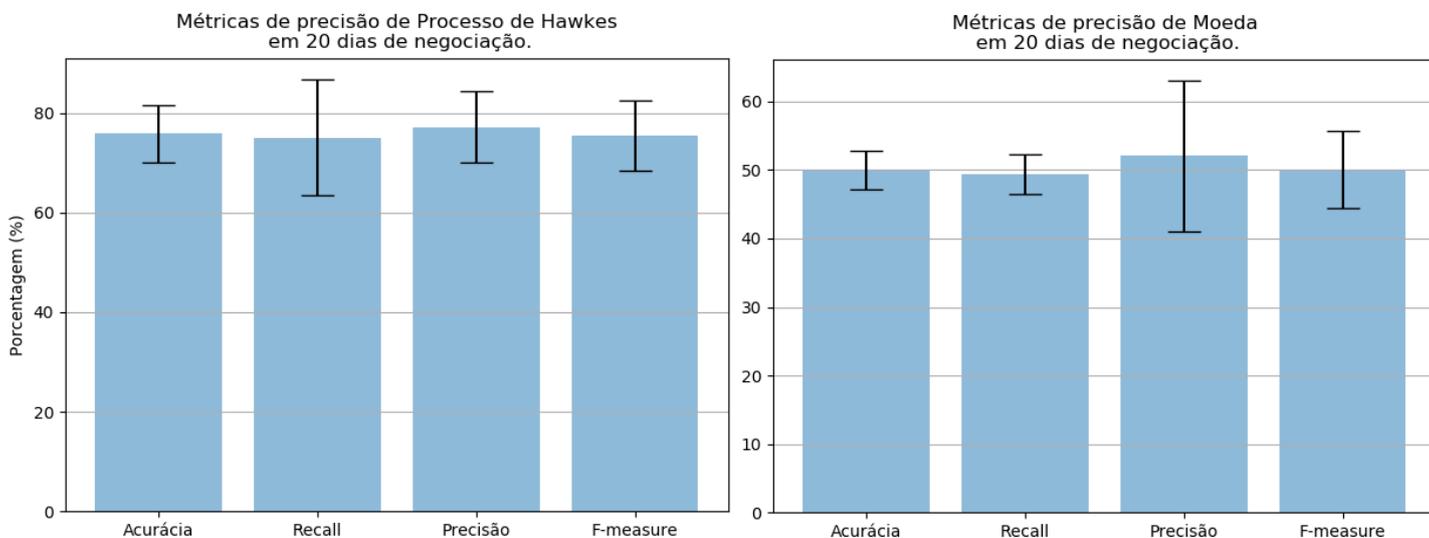


Figura 23 – Métricas de precisão dos algoritmos de previsão de lançamento de moeda e do Processo de Hawkes.

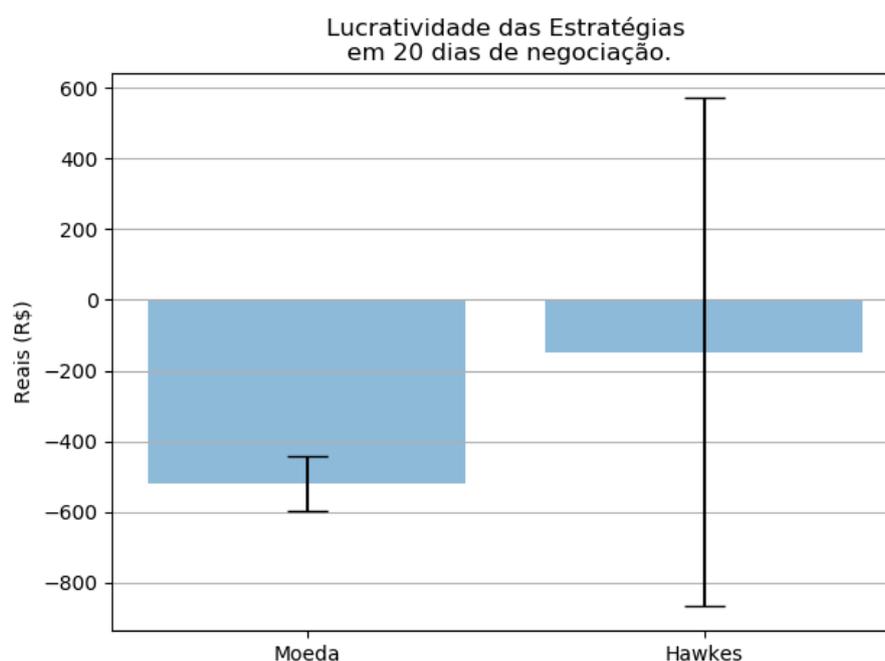


Figura 24 – Lucratividade média dos algoritmos de previsão de lançamento de moeda e do Processo de Hawkes.

## 4.8 Adição do segundo nível na Simulação de Monte Carlo

O algoritmo de previsão usando Processo de Hawkes apresentado neste trabalho realiza a previsão da mudança do preço médio em um centavo, porém a transição de um centavo não é lucrativa o que faz com que o algoritmo seja preciso mas não lucrativo. Suponha que em um determinado ativo a melhor oferta de compra esteja com o valor de R\$ 29,95 e a melhor oferta de venda esteja com o valor de R\$ 29,96, quando é feita a previsão de subida o algoritmo irá realizar uma compra no valor de R\$ 29,96, se o preço médio sobe um centavo então os novos melhores valores para compra e venda serão R\$ 29,97 e R\$ 29,96, respectivamente.

Com isso, percebe-se que ao realizar a venda terá vendido no mesmo preço em que comprou, porém para que haja lucro é necessário no mínimo um movimento de dois centavos do preço médio para começar a ter um centavo de lucro por ação. Por isso, foram realizados testes da implementação de previsão utilizando dois níveis nos primeiros 5 dias do mês de Outubro de 2019.

A Figura 25 mostrou que houve uma queda na precisão e acurácia média na mudança das abordagens, porém as Figuras 26 e 27 revelaram que a estratégia com dois níveis foi mais constante em ganhos, enquanto a estratégia com um nível teve o ganho inicial que depois foi decaindo com o tempo. Este comportamento aponta que a inserção de mais níveis pode ser mais

interessante na identificação por movimentos do mercado que são mais lucrativos do que de um centavo.

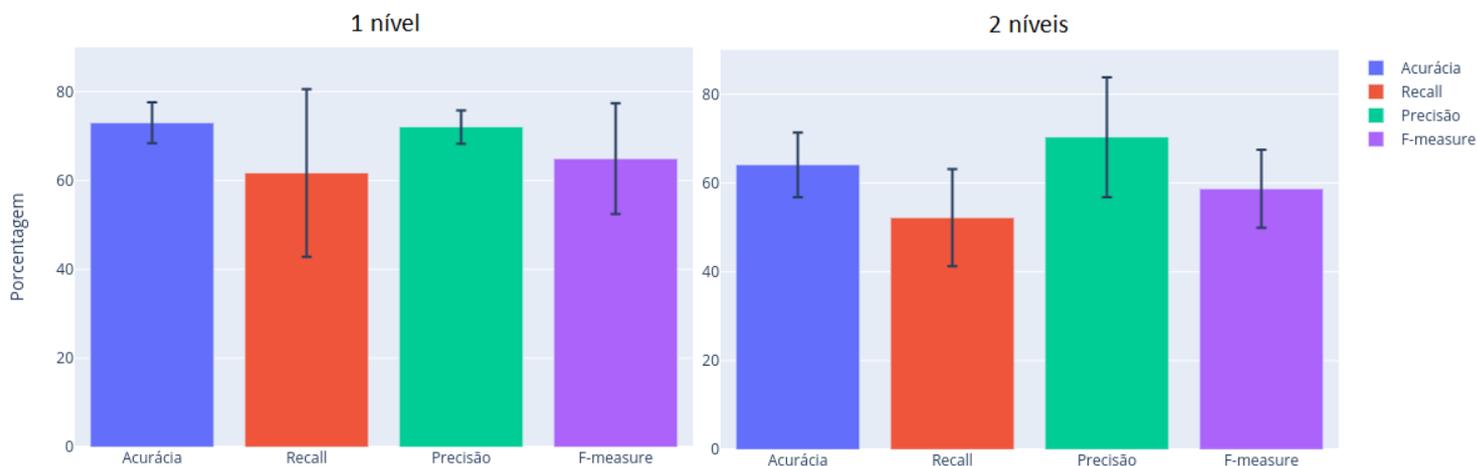


Figura 25 – Métricas de precisão das estratégias de 1 e 2 níveis na primeira semana de Outubro de 2019.

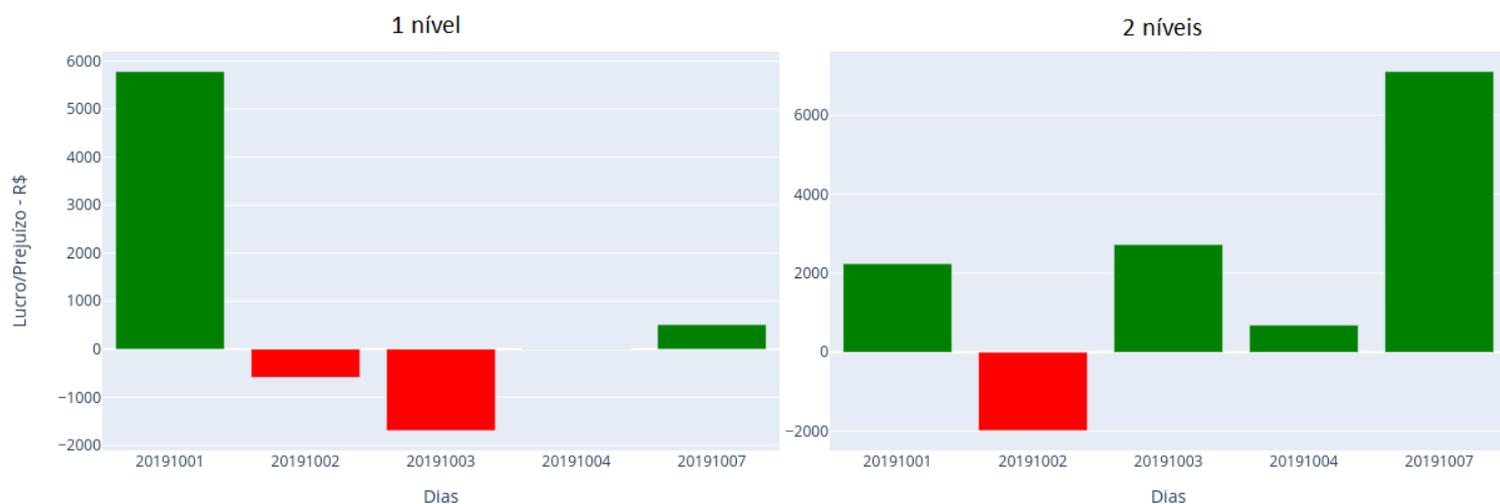


Figura 26 – Lucro e prejuízo diário feito pelas estratégias de 1 e 2 níveis na primeira semana de Outubro de 2019.

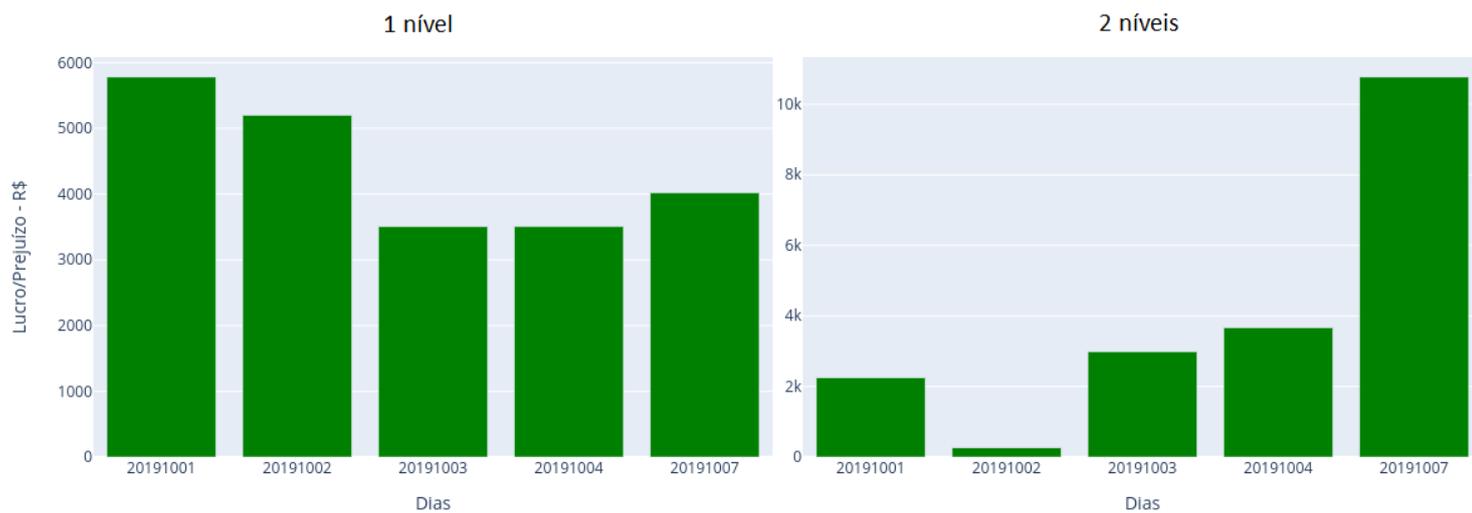


Figura 27 – Lucro e prejuízo acumulado feito pelas estratégias de 1 e 2 níveis na primeira semana de Outubro de 2019.



---

## CONCLUSÕES

---

Neste trabalho foi criado um algoritmo de leitura e extração dos dados da B3 para criação de arquivos de mensagem com apenas informações relevantes dos eventos, além da construção do arquivo de livro de ofertas.

Após isso, foi implementado um algoritmo para ajustes dos parâmetros que posteriormente seriam usados por um algoritmo de previsão do movimento do preço-médio utilizando a modelagem de um livro de ofertas utilizando Simulação de Monte Carlo com Processo de Hawkes Multivariado como apresentado na Seção 3.4.2. Durante o período de execuções para obter os dados apresentados no Capítulo 4, foi observado uma demora excessiva tanto durante os ajustes dos parâmetros quanto na simulação diária de negociações devido à complexidade de computação das funções de log-verossimilhança (Equação 2.12) e de intensidade (Equação 3.4) do processo de Hawkes Multivariado, sendo necessário a implementação de técnicas de aceleração para tornar possível uma futura implementação em tempo real, como os apresentados na Seção 5.1.

Com a adição dos algoritmos de estratégia, foi possível realizar testes para validar a lucratividade do uso das previsões em um ambiente de simulação de negociações e também realizar uma comparação com uma técnica utilizando redes neurais apresentado por Bileki (2021).

Dados os resultados obtidos na Seção 4, observou-se que o modelo foi capaz de representar bem a dinâmica de um livro de ofertas conseguindo uma precisão média de aproximadamente 70% nas previsões de mudanças do preço-médio superando a precisão média de aproximadamente 50% da Rede Neural.

Porém, pela simplicidade dos algoritmos de negociações sem nenhum suporte de outra técnica e as saídas das previsões dos algoritmos estarem relacionados com a mudança de 1 centavo, que não garante lucro como explicado na Seção 4.5, não foi possível obter uma boa lucratividade no decorrer dos dias apesar da boa precisão. Com isso, foi feito um estudo na Seção

4.7 para verificar se a utilização de uma variável aleatória teria um desempenho melhor, pois não seria necessário realizar todo esse trabalho se uma previsão aleatória fosse melhor e verificou-se que o algoritmo utilizando o Processo de Hawkes obteve um resultado melhor tanto na métricas de previsão como em lucratividade.

Após isso, foram realizados testes na Seção 4.8 com a inserção do segundo nível de preços nas previsões utilizando o Algoritmo 2 criado como uma extensão do trabalho de Chen (2017) que apenas prevê o movimento de um centavo. Assim, observou-se que o algoritmo teve uma baixa nas métricas de precisão e acurácia, o que era esperado pois torna-se mais difícil acertar quando é aumentado o tamanho da faixa de mudança de preço prevista. Porém, a lucratividade foi maior e mais constante do que o algoritmo que fazia a previsão do movimento de um centavo, o que evidencia um caminho promissor de melhora no algoritmo.

## 5.1 Trabalhos Futuros

Há diversas oportunidades de melhoria deste trabalho. Para atacar o problema de tempo de execução, existem alguns trabalhos como Guo *et al.* (2013) que acelerou o processamento em FPGA (*Field-Programmable Gate Array*) para o cálculo da função intensidade, obtendo um *speedup* de até 94 vezes em relação à implementação em uma CPU (*Central Process Unit*) de um núcleo e de 12 vezes CPU de oito núcleos. Guo e Luk (2014) apresenta resultados para a implementação em FPGA do cálculo da log-verossimilhança obtendo *speed-up* de até 79 vezes em relação à implementação em uma CPU de um núcleo e de 13 vezes para uma CPU de oito núcleos.

Além da necessidade de sistemas computacionais que permitam o ajuste de parâmetros em uma janela de tempo que viabilize a sua aplicação em negociações de tempo real, existe também a necessidade de melhorias significativas no algoritmo de negociação, como a fusão de informações oriunda de outros indicadores técnicos, como médias móveis, índices de força relativa (IFR) ou VWAP (Volume Weighted Average Price). Outros testes podem ser feitos como adicionar a informação da previsão do Processo de Hawkes com uma rede neural, substituir o Processo de Hawkes por um Modelo Bayesiano, ajustar os parâmetros utilizando dados de outros dia ou semanas para um mesmo período tempo do dia de negociações ou até adicionar novos eventos além dos mostrados nesse trabalho.

Este trabalho fez testes simulando um dia de negociações de forma simples supondo não ter custos operacionais e que ao comprar e vender os ativos é possível ser feito pelo melhor preço disponível no tempo observado. Então, será interessante fazer o estudo, criação e testes em um ambiente mais próximo ao real e que leve em consideração comportamentos reativos do mercado quando são realizadas negociações que geram movimentos de preço, como mostram os trabalhos de Carvalho *et al.* (2017) e Leite (2020).

## REFERÊNCIAS

---

---

- ABERGEL, F.; ANANE, M.; CHAKRABORTI, A.; JEDIDI, A.; TOKE, I. M. **Limit Order Books**. [S.l.]: Cambridge University Press, 2016. (Physics of Society: Econophysics and Sociophysics). Citado nas páginas 29 e 55.
- BACRY, E.; DELATTRE, S.; HOFFMANN, M.; MUZY, J.-F. Modelling microstructure noise with mutually exciting point processes. **Quantitative finance**, Taylor & Francis, v. 13, n. 1, p. 65–77, 2013. Citado na página 42.
- BACRY, E.; MASTROMATTEO, I.; MUZY, J.-F. Hawkes processes in finance. **Market Microstructure and Liquidity**, World Scientific, v. 1, n. 01, 2015. Citado nas páginas 26, 41 e 42.
- BACRY, E.; MUZY, J.-F. Hawkes model for price and trades high-frequency dynamics. **Quantitative Finance**, Taylor & Francis, v. 14, n. 7, p. 1147–1166, 2014. Citado na página 42.
- BIAIS, B.; HILLION, P.; SPATT, C. An empirical analysis of the limit order book and the order flow in the paris bourse. **the Journal of Finance**, Wiley Online Library, v. 50, n. 5, p. 1655–1689, 1995. Citado na página 26.
- BILEKI, G. A. **Uma abordagem com modelo de aprendizado de máquina híbrido para previsão de movimentos de preço médio de ativos pelo livro de ofertas**. Dissertação (Mestrado) — Universidade de São Paulo, 2021. Citado nas páginas 16, 59, 68, 69, 70 e 75.
- BOWSHER, C. G. Modelling security market events in continuous time: Intensity based, multivariate point process models. **Journal of Econometrics**, Elsevier, v. 141, n. 2, p. 876–912, 2007. Citado na página 41.
- BRÉMAUD, P. **Point processes and queues: martingale dynamics**. [S.l.]: Springer, 1981. v. 50. Citado na página 39.
- CARVALHO, L. F.; VIEIRA, F. V.; RIBEIRO, K. C. d. S.; BORGES, W. G. Os efeitos alavancagem e feedback na volatilidade do mercado acionário brasileiro. **Enfoque: Reflexão Contábil**, v. 36, n. 2, p. 19–37, maio 2017. Disponível em: <<https://periodicos.uem.br/ojs/index.php/Enfoque/article/view/31375>>. Citado na página 76.
- CHEN, Y. **Modelling Limit Order Book Dynamics Using Hawkes Processes**. Tese (Doutorado) — Florida State University, 2017. Citado nas páginas 15, 16, 38, 40, 42, 49, 50, 52, 56, 57, 58, 59, 60, 64, 66 e 76.
- CHORNOBOY, E.; SCHRAMM, L.; KARR, A. Maximum likelihood identification of neural point process systems. **Biological cybernetics**, Springer, v. 59, n. 4-5, p. 265–275, 1988. Citado na página 26.
- COCULESCU, D.; NIKEGHBALI, A. Filtrations. In: CONT, R. (Ed.). **Encyclopedia of Quantitative Finance**. Chichester, UK: Wiley & Sons, 2010. p. 1–5. ISBN 9780470061602. Disponível em: <<https://www.zora.uzh.ch/id/eprint/137263/>>. Citado na página 35.

- CONT, R.; STOIKOV, S.; TALREJA, R. A stochastic model for order book dynamics. **Operations research**, Informs, v. 58, n. 3, p. 549–563, 2010. Citado na página 26.
- COX, D.; LEWIS, P. *et al.* Multivariate point processes. In: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA. **Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory**. [S.l.], 1972. Citado na página 39.
- DALEY, D. J.; VERE-JONES, D. **An introduction to the theory of point processes. Vol. I**. 2. ed. [S.l.]: Springer, 2003. Citado nas páginas 34, 36, 37 e 41.
- DANIELS, M. G.; FARMER, J. D.; GILLEMOT, L.; IORI, G.; SMITH, E. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. **Physical review letters**, APS, v. 90, n. 10, p. 108102, 2003. Citado na página 26.
- FILIMONOV, V.; SORNETTE, D. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. **Physical Review E**, APS, v. 85, n. 5, p. 056108, 2012. Citado na página 42.
- GOULD, M. D.; PORTER, M. A.; WILLIAMS, S.; MCDONALD, M.; FENN, D. J.; HOWISON, S. D. Limit order books. **Quantitative Finance**, Routledge, v. 13, n. 11, p. 1709–1742, 2013. Disponível em: <<https://doi.org/10.1080/14697688.2013.803148>>. Citado nas páginas 26 e 50.
- GRANDELL, J. **Mixed poisson processes**. [S.l.]: CRC Press, 1997. v. 77. Citado na página 36.
- GUO, C.; LUK, W. Accelerating parameter estimation for multivariate self-exciting point processes. In: ACM. **Proceedings of the 2014 ACM/SIGDA international symposium on Field-programmable gate arrays**. [S.l.], 2014. p. 181–184. Citado na página 76.
- GUO, C.; LUK, W.; VINKOVSKAYA, E.; CONT, R. Customisable pipelined engine for intensity evaluation in multivariate hawkes point processes. **ACM SIGARCH Computer Architecture News**, ACM, v. 41, n. 5, p. 59–64, 2013. Citado na página 76.
- HALL, A. D.; HAUTSCH, N. Order aggressiveness and order book dynamics. In: **High Frequency Financial Econometrics**. [S.l.]: Springer, 2008. p. 133–165. Citado na página 26.
- HAWKES, A. Spectra of some mutually exciting point processes with associated variables. **Stochastic point processes**, Wiley, New York, p. 261–271, 1972. Citado na página 41.
- HAWKES, A. G. Point spectra of some mutually exciting point processes. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 33, n. 3, p. 438–443, 1971. Citado na página 41.
- \_\_\_\_\_. Spectra of some self-exciting and mutually-exciting point processes. **Biometrika**, v. 58, n. 1, p. 83–90, 04 1971. Citado nas páginas 26, 37, 40, 41 e 51.
- \_\_\_\_\_. Hawkes processes and their applications to finance: a review. **Quantitative Finance**, Routledge, v. 18, n. 2, p. 193–198, 2018. Citado nas páginas 26 e 42.
- HUANG, H.; KERCHEVAL, A. N. A generalized birth–death stochastic model for high-frequency order book dynamics. **Quantitative Finance**, Taylor & Francis, v. 12, n. 4, p. 547–557, 2012. Citado na página 26.
- JOHNSON, D. H. Point process models of single-neuron discharges. **Journal of computational neuroscience**, Springer, v. 3, n. 4, p. 275–299, 1996. Citado na página 26.

KHINCHIN, A. Y.; ANDREWS, D.; QUENOUILLE, M. H. **Mathematical methods in the theory of queuing**. [S.l.]: Courier Corporation, 2013. Citado na página 36.

LAST, G.; BRANDT, A. **Marked Point Processes on the real line: the dynamical approach**. [S.l.]: Springer Science & Business Media, 1995. Citado nas páginas 34, 35 e 36.

LEITE, A. W. O. **Análise da eficiência do mercado de ações brasileiro em alta frequência**. Dissertação (Mestrado) — Universidade de Brasília, 2020. Citado na página 76.

LU, X.; ABERGEL, F. High-dimensional hawkes processes for limit order books: modelling, empirical analysis and numerical calibration. **Quantitative Finance**, Taylor & Francis, v. 18, n. 2, p. 249–264, 2018. Citado na página 42.

MOHLER, G. *et al.* Modeling and estimation of multi-source clustering in crime and security data. **The Annals of Applied Statistics**, Institute of Mathematical Statistics, v. 7, n. 3, p. 1525–1539, 2013. Citado na página 26.

MURPHY, J. J. **Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications**. [S.l.]: New York Institute of Finance, 1999. Citado na página 25.

NELOGICA. **Ferramentas ligadas ao livro de ofertas (Book)**. 2019. <<https://www.nelogica.com.br/produtos/profit-plus/book>>. Acessado em: 15 de Janeiro de 2019. Citado nas páginas 15 e 34.

OGATA, Y. Statistical models for earthquake occurrences and residual analysis for point processes. **Journal of the American Statistical association**, Taylor & Francis, v. 83, n. 401, p. 9–27, 1988. Citado na página 26.

ROSS, S. **Introduction to Probability Models**. 10. ed. [S.l.]: Elsevier Science, 2010. Citado na página 34.

RUBIN, I. Regular point processes and their detection. **IEEE Transactions on Information Theory**, IEEE, v. 18, n. 5, p. 547–557, 1972. Citado na página 38.

SANTITISSADEEKORN, N.; LLOYD, D. J.; SHORT, M. B.; DELAHAIES, S. Approximate filtering of conditional intensity process for poisson count data: Application to urban crime. **Computational Statistics & Data Analysis**, Elsevier, v. 144, p. 106850, 2020. Citado na página 26.

SHAIKH, A. On the role of reflexivity in economic analysis. **Journal of Economic Methodology**, Taylor & Francis, v. 20, n. 4, p. 439–445, 2013. Citado na página 42.

SIGMAN, K. **Stationary marked point processes: an intuitive approach**. [S.l.]: Chapman & Hall, 1995. Citado na página 35.

SNYDER, D. L.; MILLER, M. I. **Random point processes in time and space**. [S.l.]: Springer Science & Business Media, 2012. Citado na página 36.

TEIXEIRA, L. A.; OLIVEIRA, A. L. I. de. A method for automatic stock trading combining technical analysis and nearest neighbor classification. **Expert systems with applications**, Elsevier, v. 37, n. 10, p. 6885–6890, 2010. Citado na página 25.

TOKE, I. M.; POMPONIO, F. Modelling trades-through in a limited order book using hawkes processes. **Economics discussion paper**, n. 2011-32, 2011. Citado na página 42.

VANSTONE, B.; FINNIE, G. An empirical methodology for developing stockmarket trading systems using artificial neural networks. **Expert systems with Applications**, Elsevier, v. 36, n. 3, p. 6668–6680, 2009. Citado na página 25.

World Federation of Exchanges. **The World Federation of Exchanges Publishes 2017 Full Year Market Highlights**. 2018. Acessado em: 15 de Janeiro de 2019. Disponível em: <<https://www.world-exchanges.org/news/articles/the-world-federation-of-exchanges-publishes-2017-full-year-market-highlights>>. Citado na página 25.

ZADEH, A. H.; SHARDA, R. Hawkes point processes for social media analytics. In: **Reshaping Society through Analytics, Collaboration, and Decision Support**. [S.l.]: Springer, 2015. p. 51–66. Citado na página 26.

ZHENG, X. **High frequency dynamics of order flow**. Dissertação (Mestrado) — Oxford University, 2013. Citado na página 29.

ZIPKIN, J. R.; SCHOENBERG, F. P.; CORONGES, K.; BERTOZZI, A. L. Point-process models of social network interactions: Parameter estimation and missing data recovery. **European journal of applied mathematics**, v. 27, n. 3, p. 502–529, 2016. Citado na página 26.

## DADOS DAS EXECUÇÕES DIÁRIAS DA ESTRATÉGIA SIMPLES PARA O MÊS DE OUTUBRO DE 2019

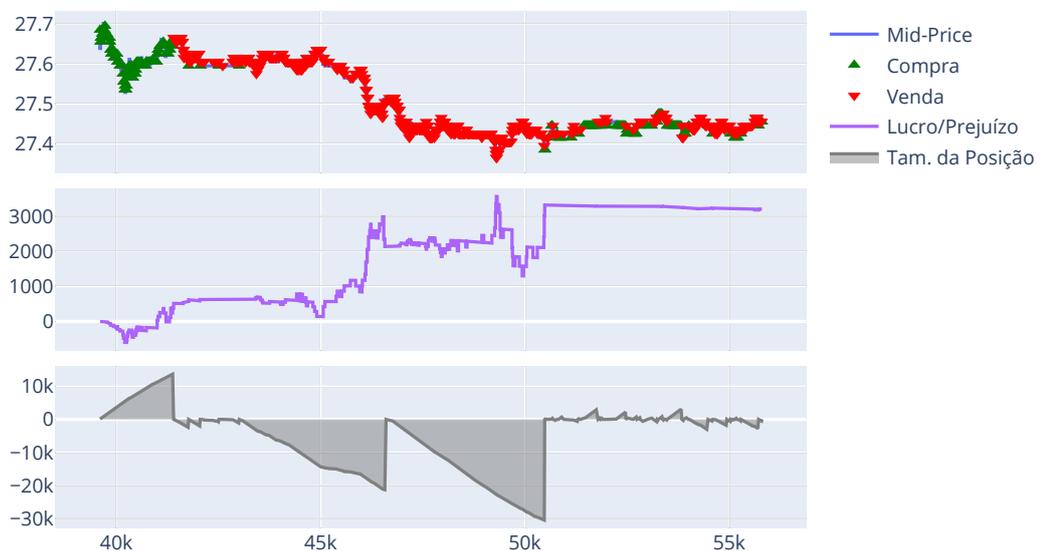


Figura 28 – Ações realizadas pela estratégia simples no dia de negociações 01/10/2019.

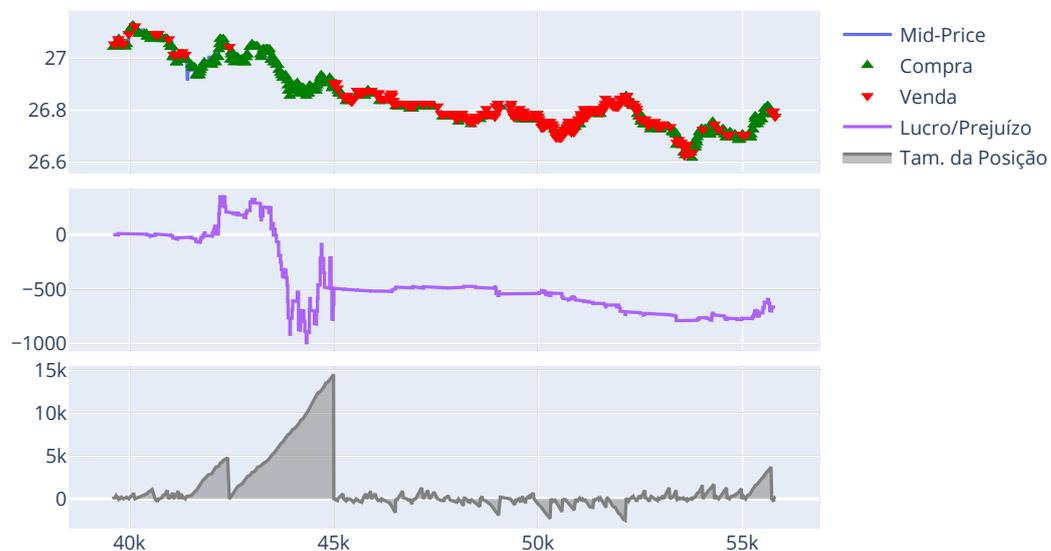


Figura 29 – Ações realizadas pela estratégia simples no dia de negociações 02/10/2019.

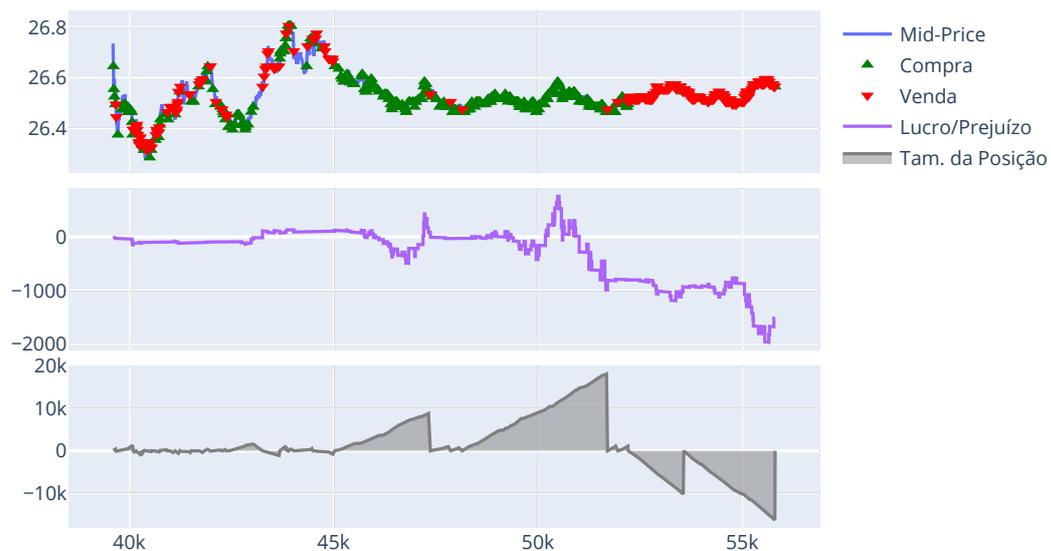


Figura 30 – Ações realizadas pela estratégia simples no dia de negociações 03/10/2019.

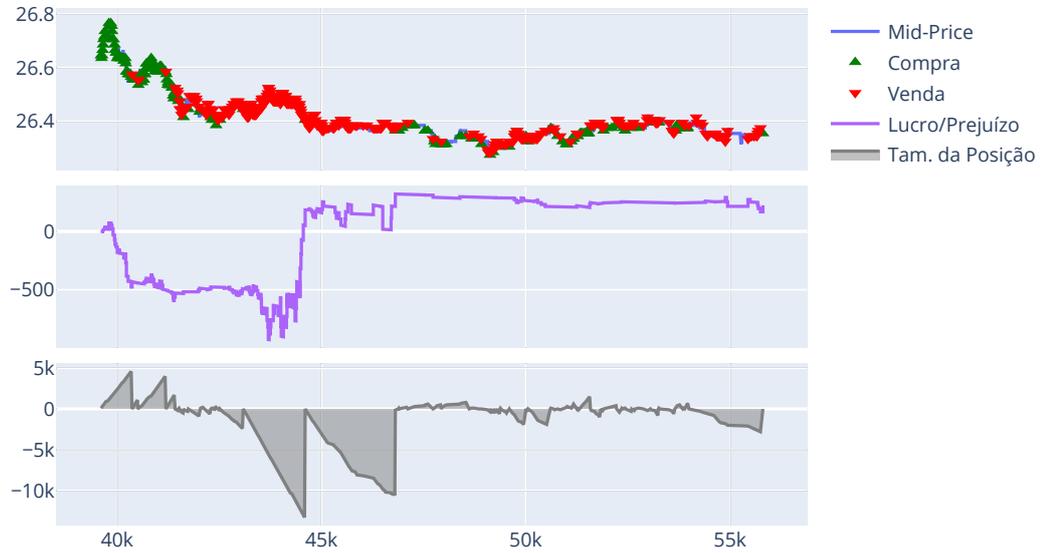


Figura 31 – Ações realizadas pela estratégia simples no dia de negociações 04/10/2019.

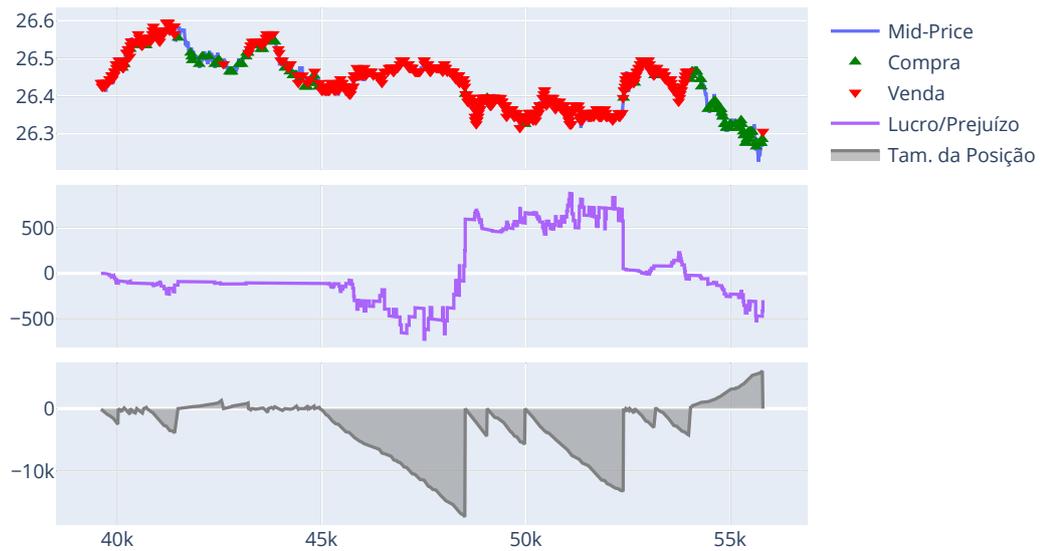


Figura 32 – Ações realizadas pela estratégia simples no dia de negociações 07/10/2019.

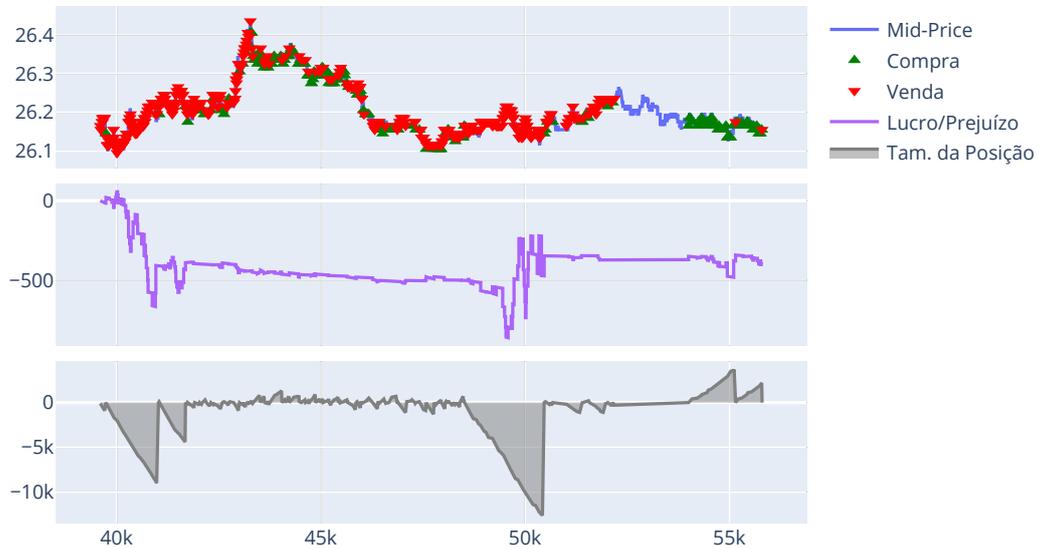


Figura 33 – Ações realizadas pela estratégia simples no dia de negociações 08/10/2019.

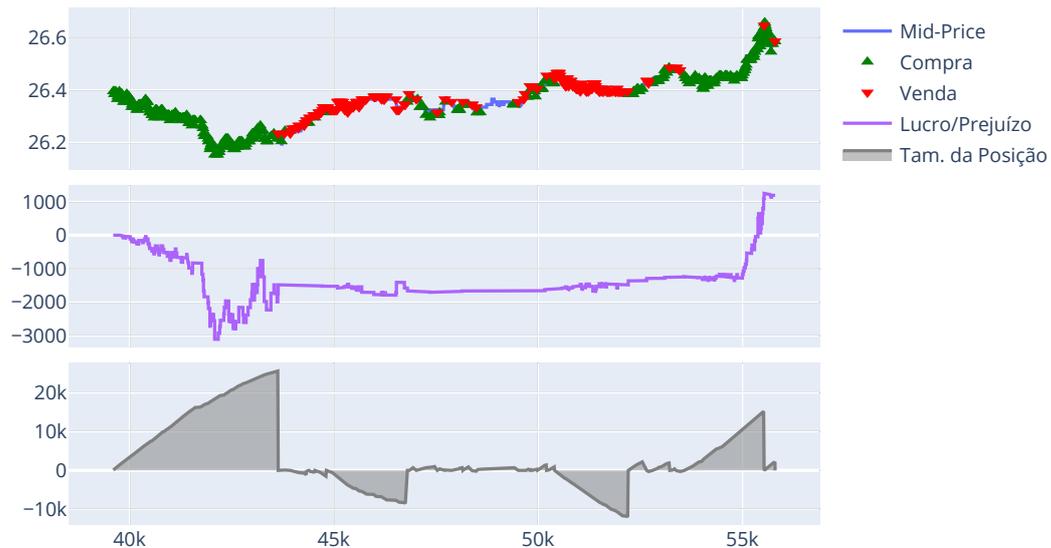


Figura 34 – Ações realizadas pela estratégia simples no dia de negociações 09/10/2019.

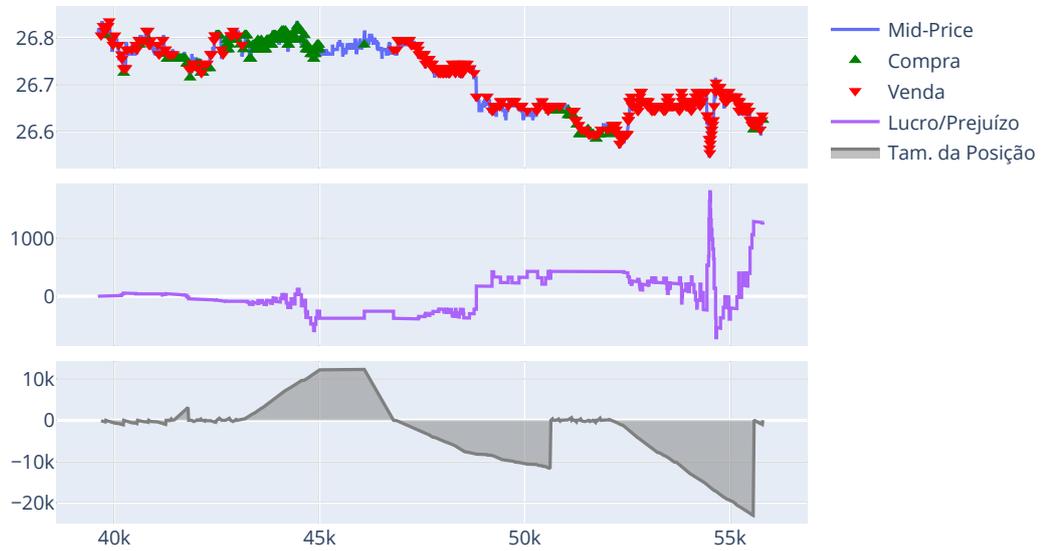


Figura 35 – Ações realizadas pela estratégia simples no dia de negociações 10/10/2019.



Figura 36 – Ações realizadas pela estratégia simples no dia de negociações 11/10/2019.

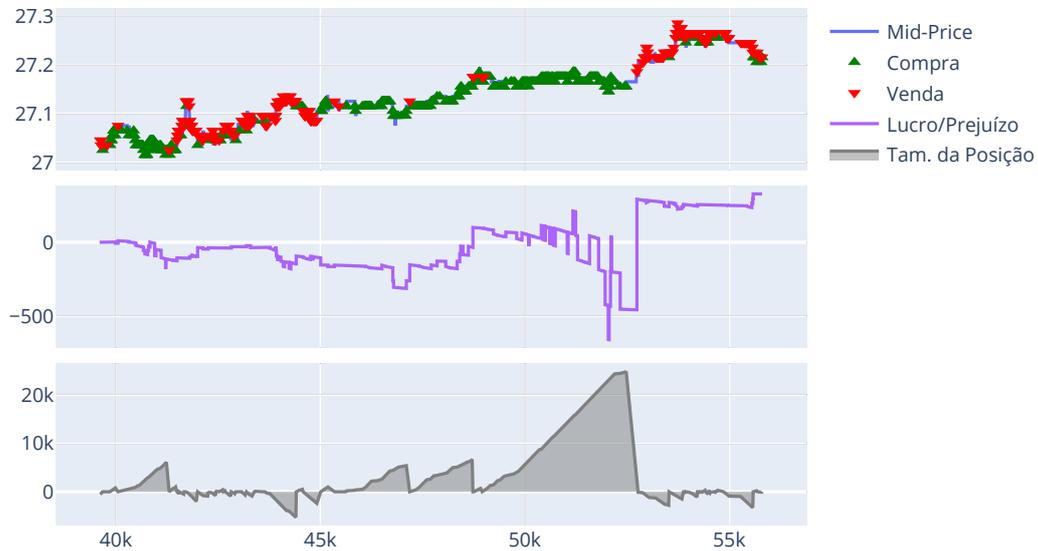


Figura 37 – Ações realizadas pela estratégia simples no dia de negociações 14/10/2019.



Figura 38 – Ações realizadas pela estratégia simples no dia de negociações 15/10/2019.

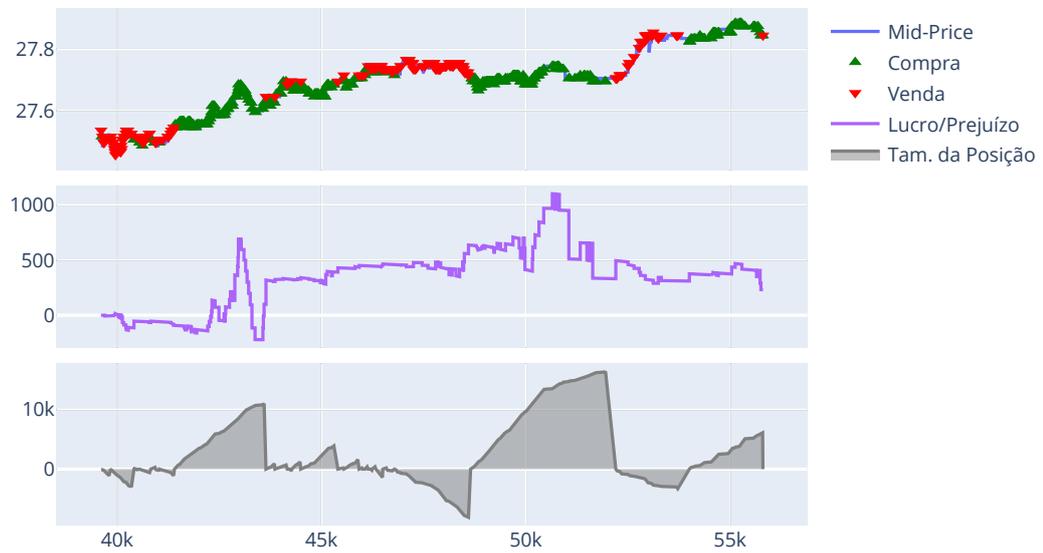


Figura 39 – Ações realizadas pela estratégia simples no dia de negociações 16/10/2019.

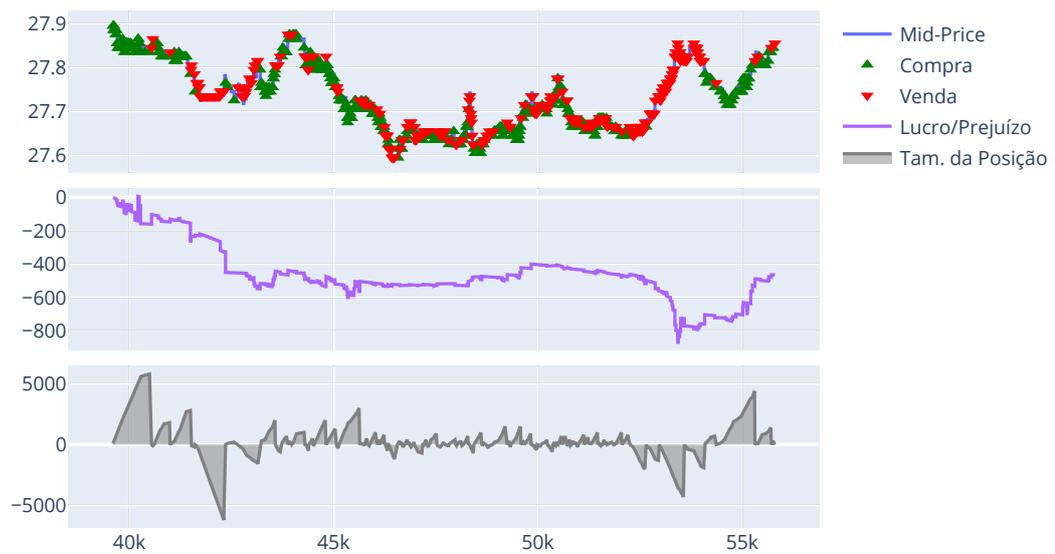


Figura 40 – Ações realizadas pela estratégia simples no dia de negociações 17/10/2019.

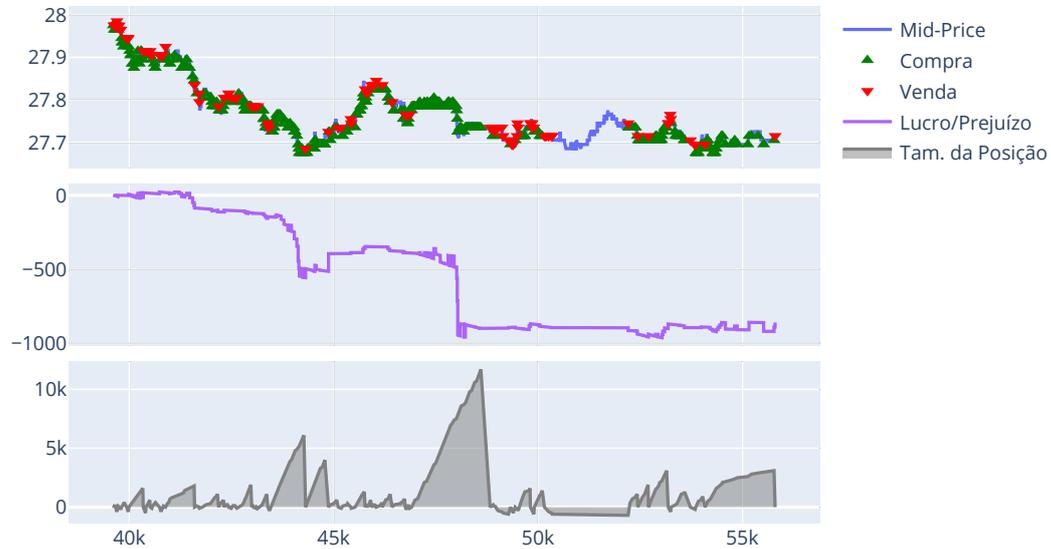


Figura 41 – Ações realizadas pela estratégia simples no dia de negociações 18/10/2019.

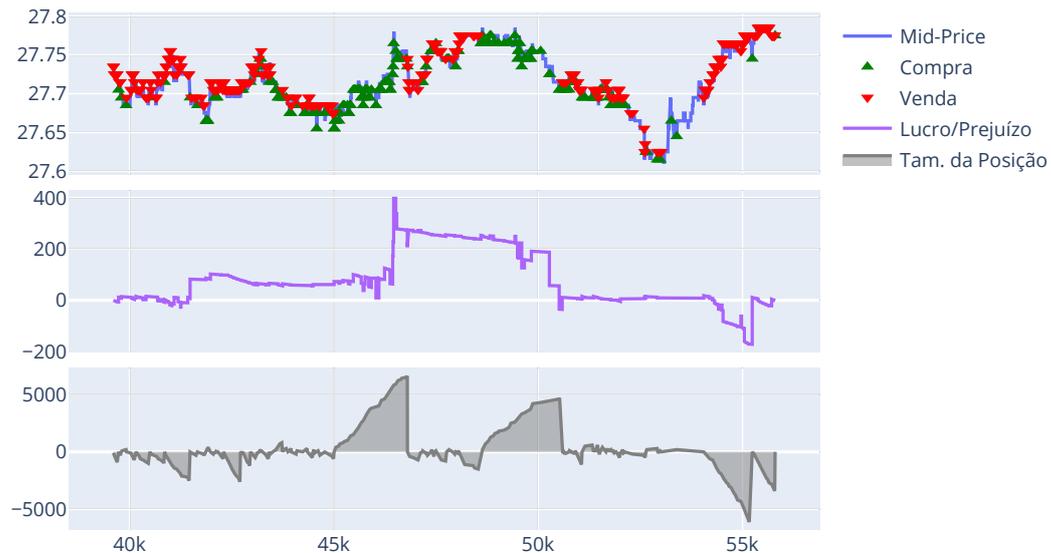


Figura 42 – Ações realizadas pela estratégia simples no dia de negociações 21/10/2019.



Figura 43 – Ações realizadas pela estratégia simples no dia de negociações 22/10/2019.

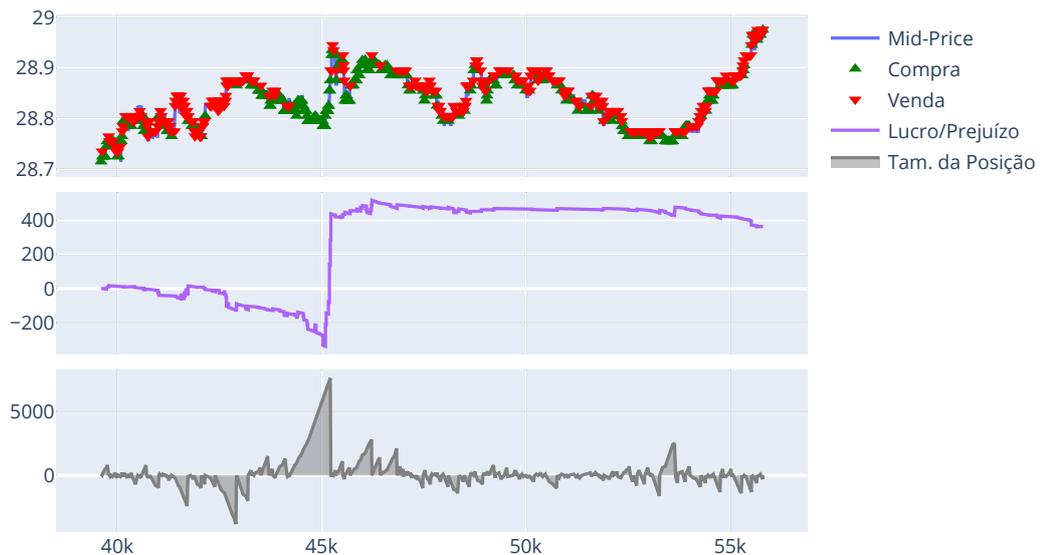


Figura 44 – Ações realizadas pela estratégia simples no dia de negociações 23/10/2019.

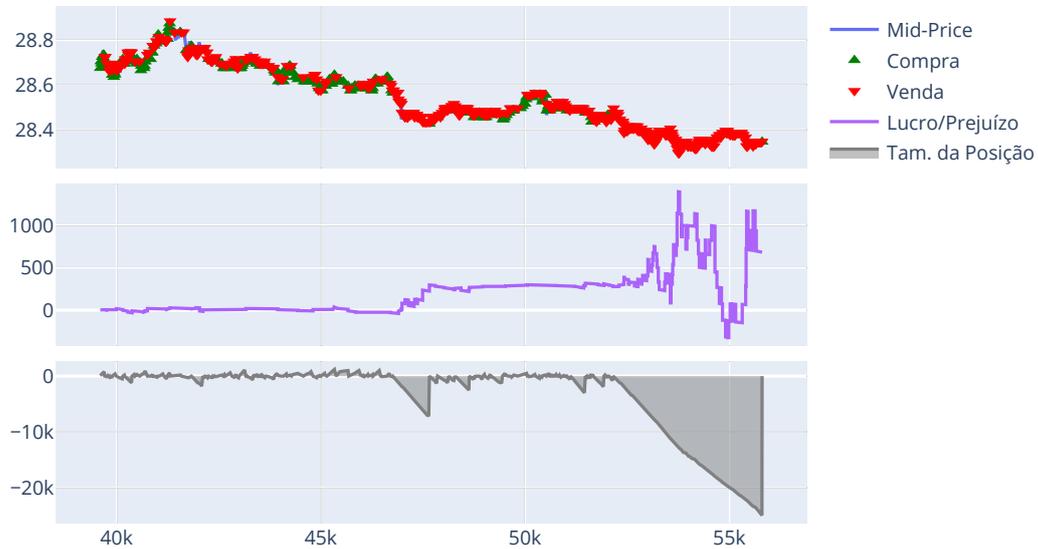


Figura 45 – Ações realizadas pela estratégia simples no dia de negociações 24/10/2019.

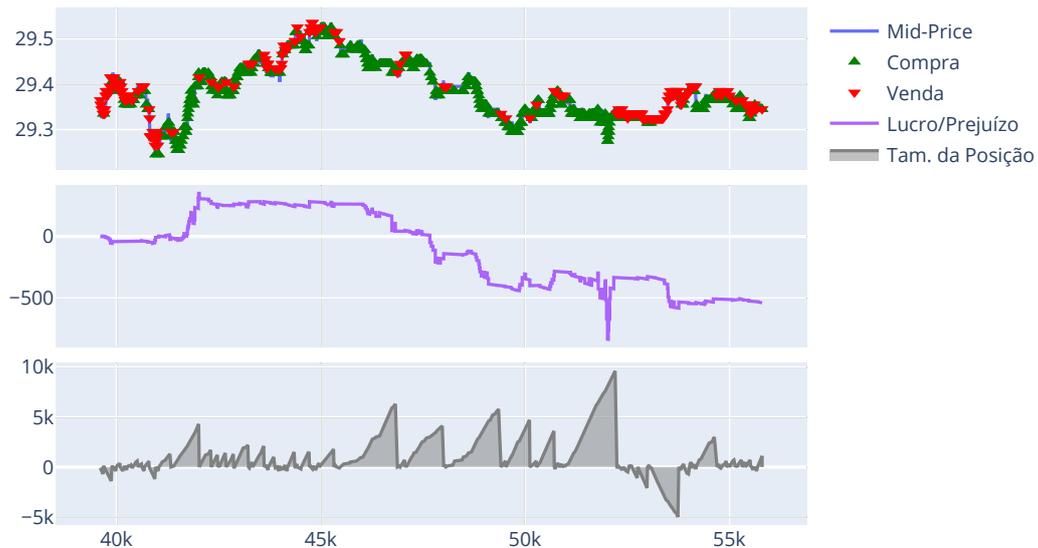


Figura 46 – Ações realizadas pela estratégia simples no dia de negociações 25/10/2019.



Figura 47 – Ações realizadas pela estratégia simples no dia de negociações 28/10/2019.

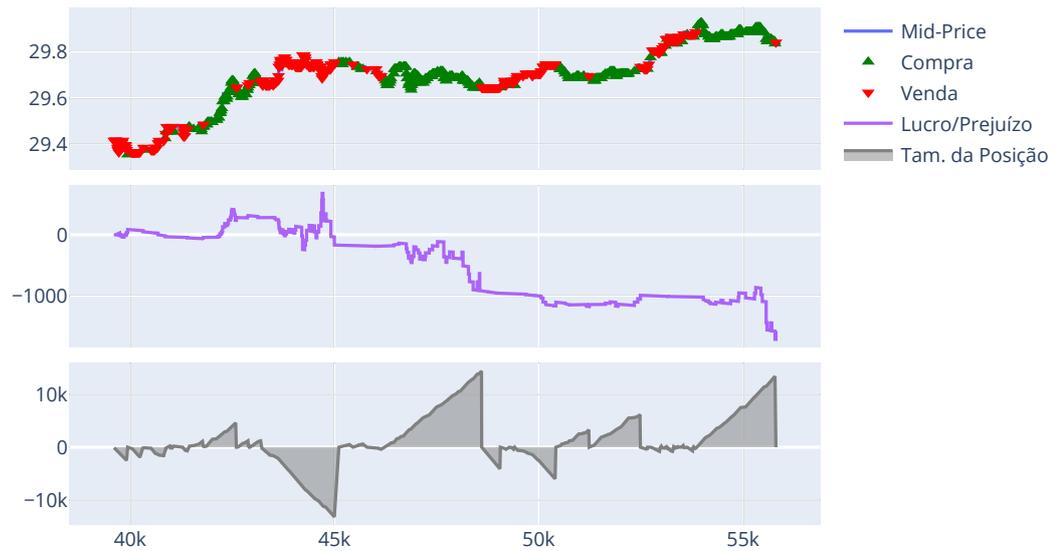


Figura 48 – Ações realizadas pela estratégia simples no dia de negociações 29/10/2019.

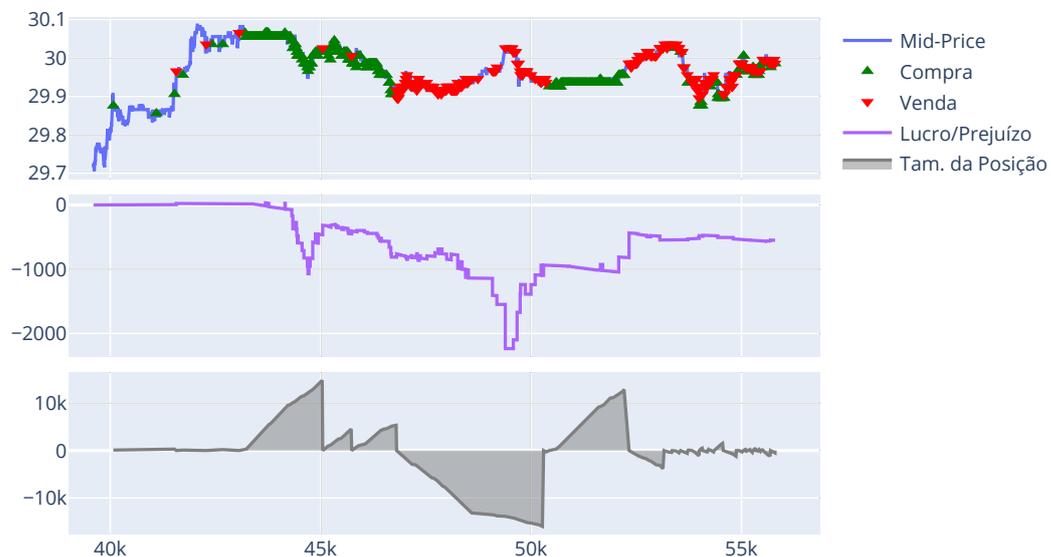


Figura 49 – Ações realizadas pela estratégia simples no dia de negociações 30/10/2019.



Figura 50 – Ações realizadas pela estratégia simples no dia de negociações 31/10/2019.

## DADOS DAS EXECUÇÕES DIÁRIAS DA ESTRATÉGIA APRIMORADA PARA O MÊS DE OUTUBRO DE 2019

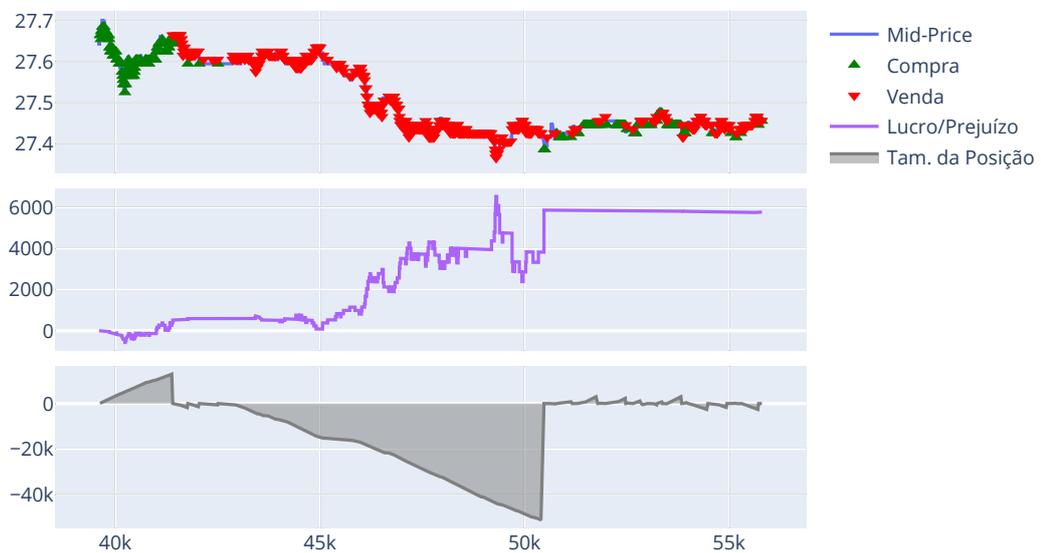


Figura 51 – Ações realizadas pela estratégia atualizada no dia de negociações 01/10/2019.



Figura 52 – Ações realizadas pela estratégia atualizada no dia de negociações 02/10/2019.

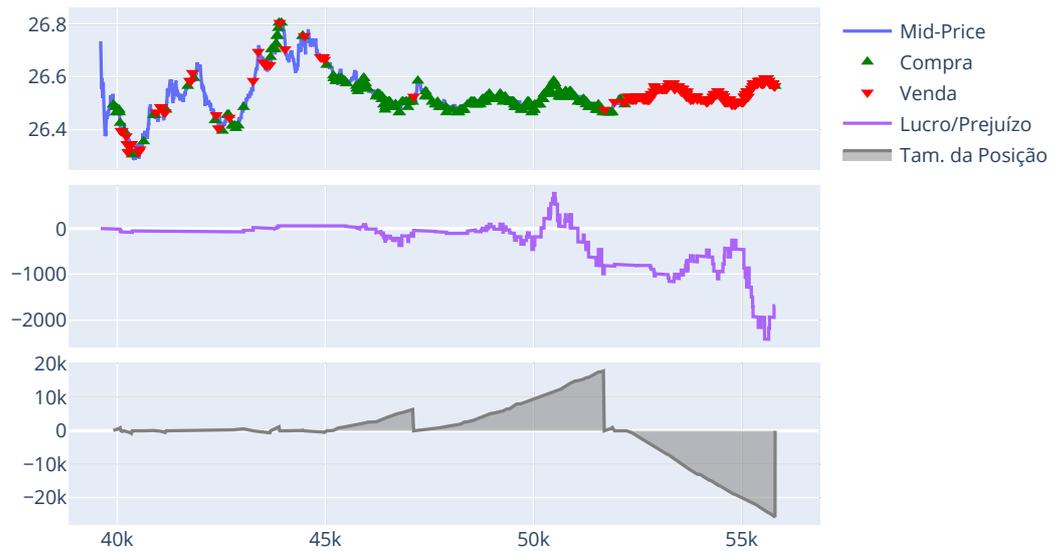


Figura 53 – Ações realizadas pela estratégia atualizada no dia de negociações 03/10/2019.

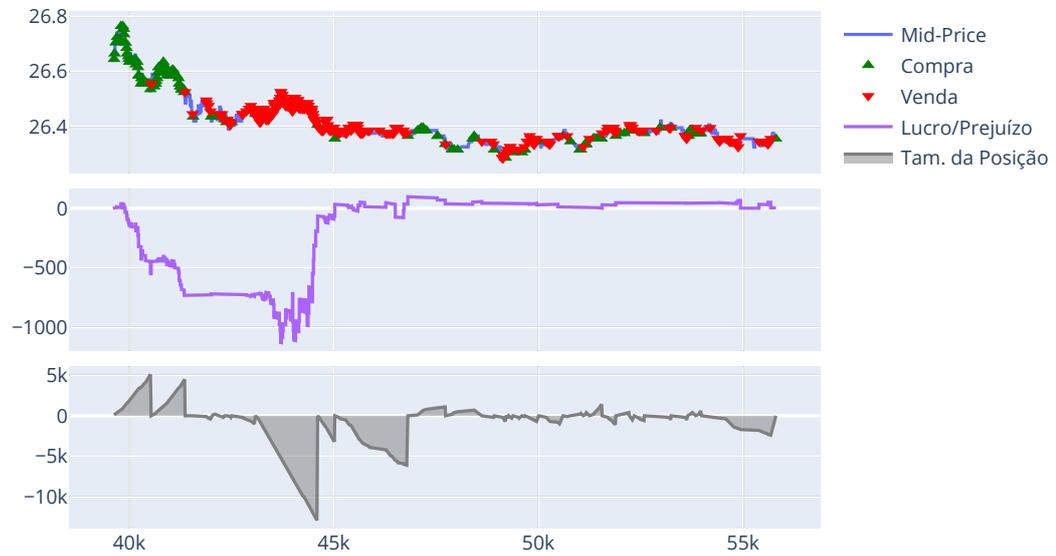


Figura 54 – Ações realizadas pela estratégia atualizada no dia de negociações 04/10/2019.

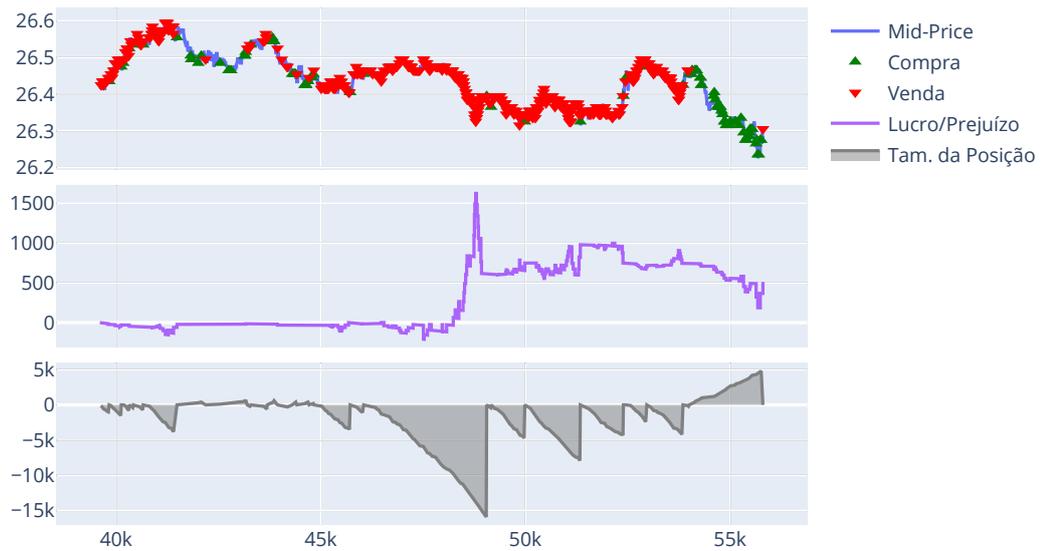


Figura 55 – Ações realizadas pela estratégia atualizada no dia de negociações 07/10/2019.

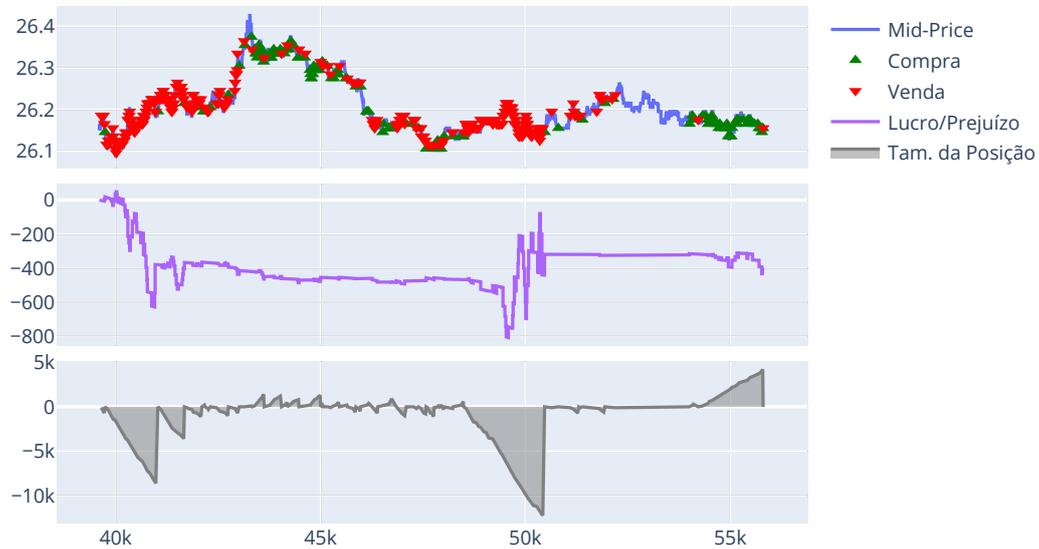


Figura 56 – Ações realizadas pela estratégia atualizada no dia de negociações 08/10/2019.

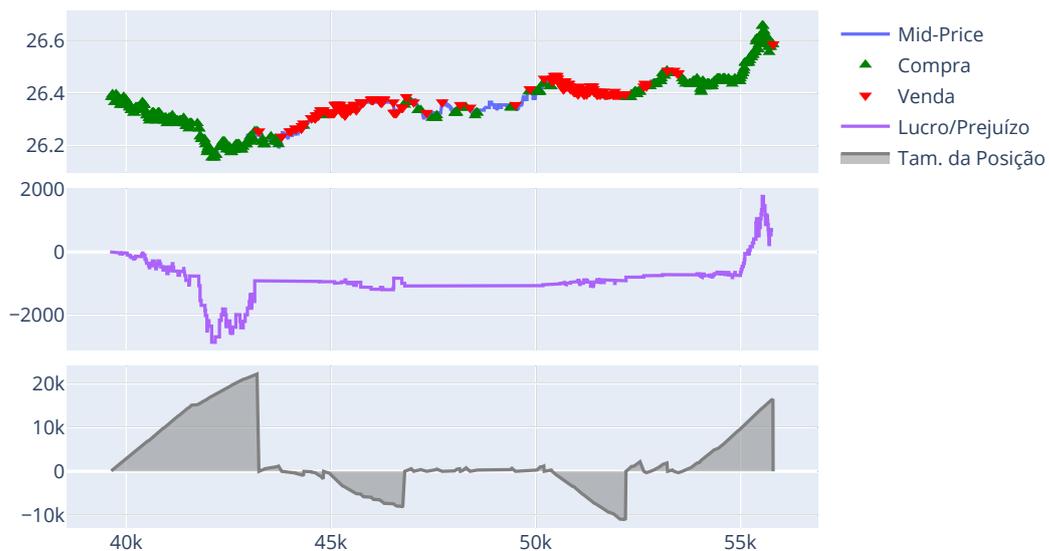


Figura 57 – Ações realizadas pela estratégia atualizada no dia de negociações 09/10/2019.

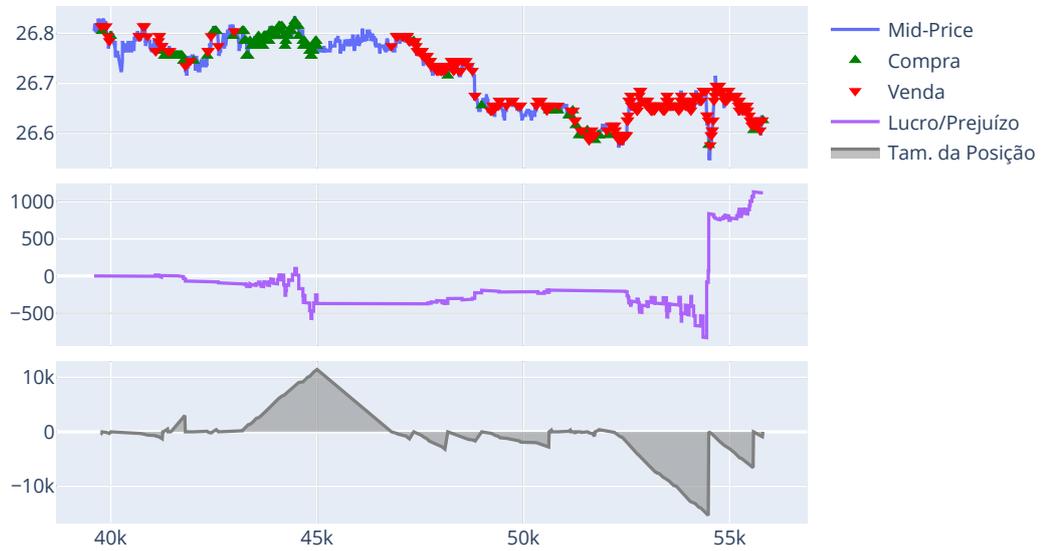


Figura 58 – Ações realizadas pela estratégia atualizada no dia de negociações 10/10/2019.



Figura 59 – Ações realizadas pela estratégia atualizada no dia de negociações 11/10/2019.

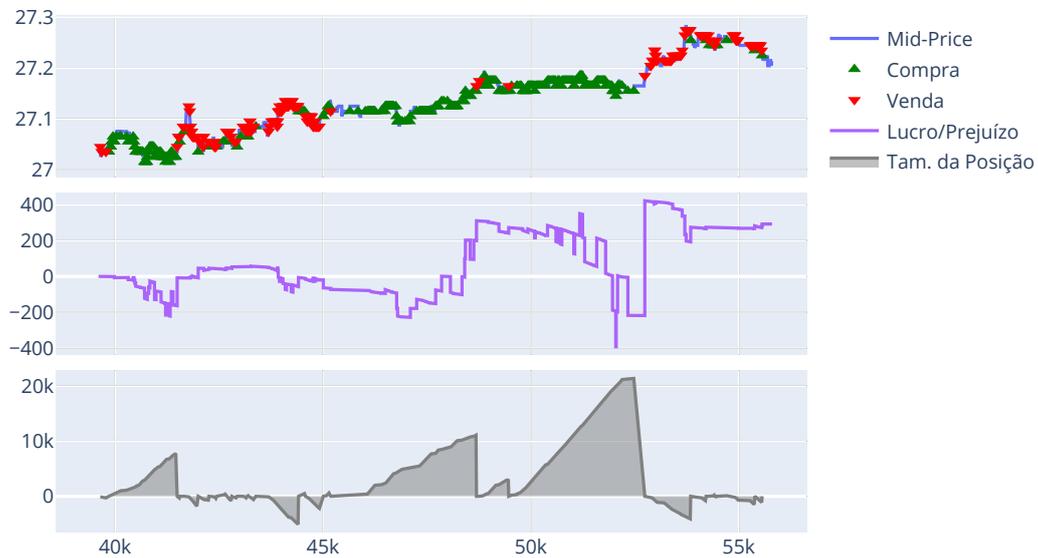


Figura 60 – Ações realizadas pela estratégia atualizada no dia de negociações 14/10/2019.

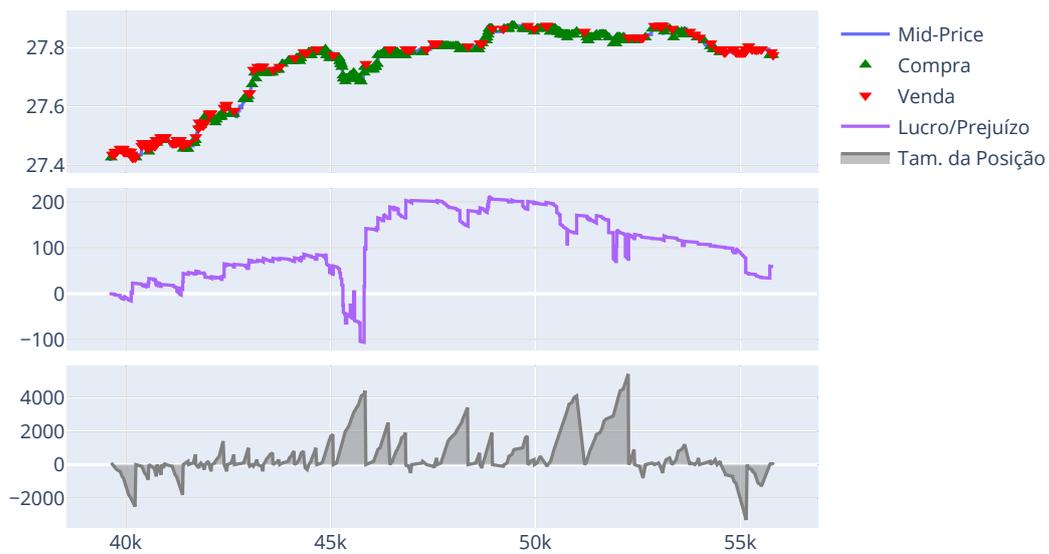


Figura 61 – Ações realizadas pela estratégia atualizada no dia de negociações 15/10/2019.

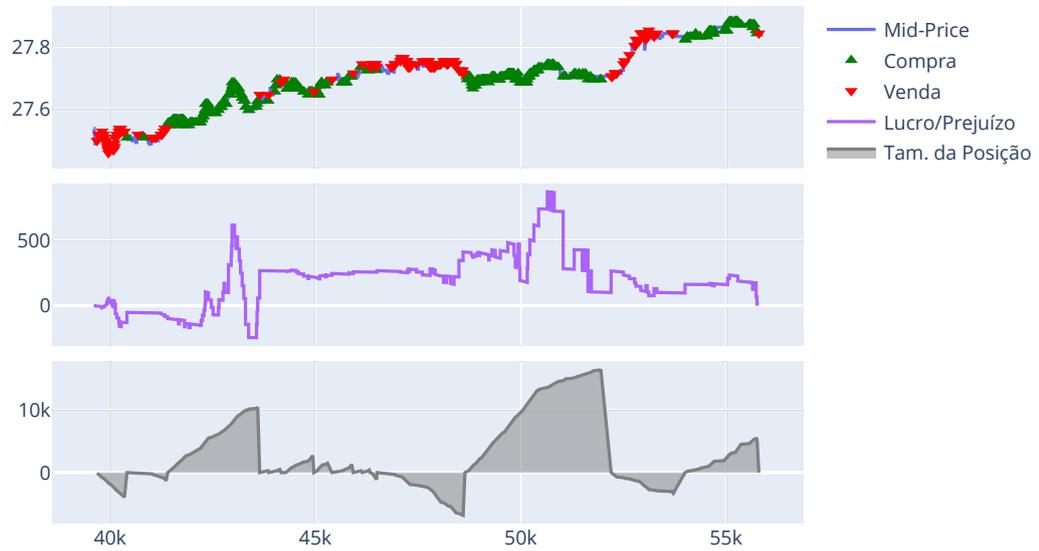


Figura 62 – Ações realizadas pela estratégia atualizada no dia de negociações 16/10/2019.

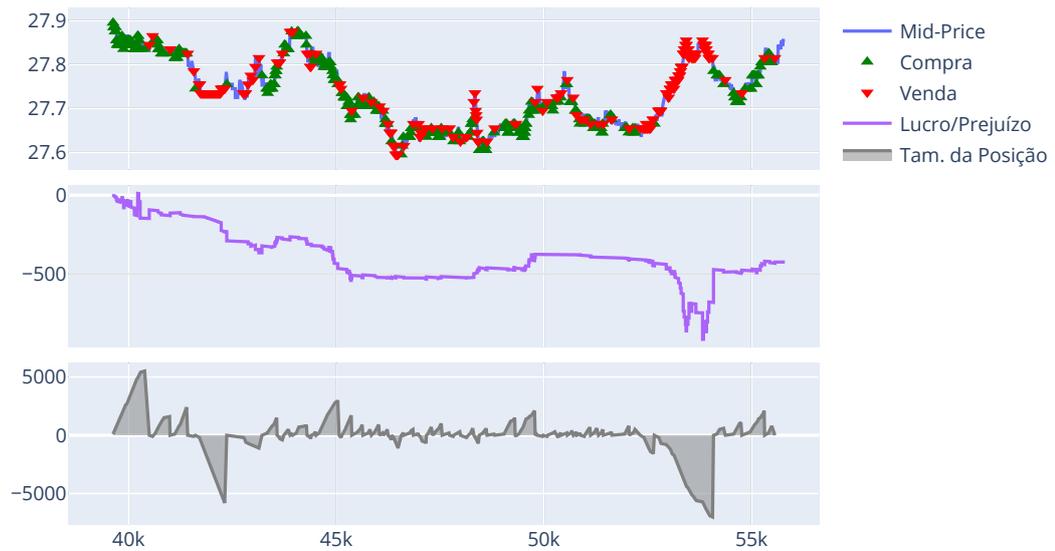


Figura 63 – Ações realizadas pela estratégia atualizada no dia de negociações 17/10/2019.



Figura 64 – Ações realizadas pela estratégia atualizada no dia de negociações 18/10/2019.



Figura 65 – Ações realizadas pela estratégia atualizada no dia de negociações 21/10/2019.



Figura 66 – Ações realizadas pela estratégia atualizada no dia de negociações 22/10/2019.

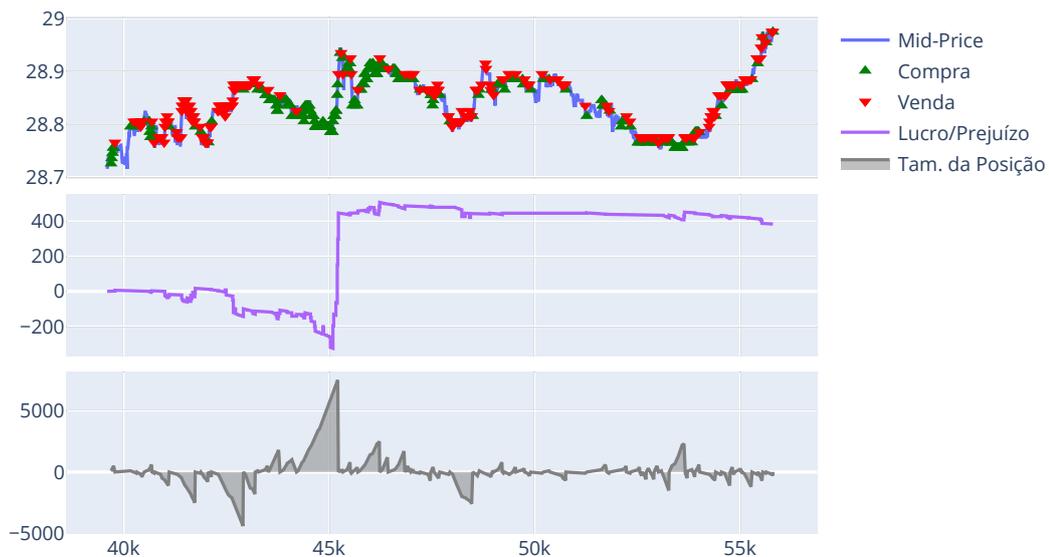


Figura 67 – Ações realizadas pela estratégia atualizada no dia de negociações 23/10/2019.

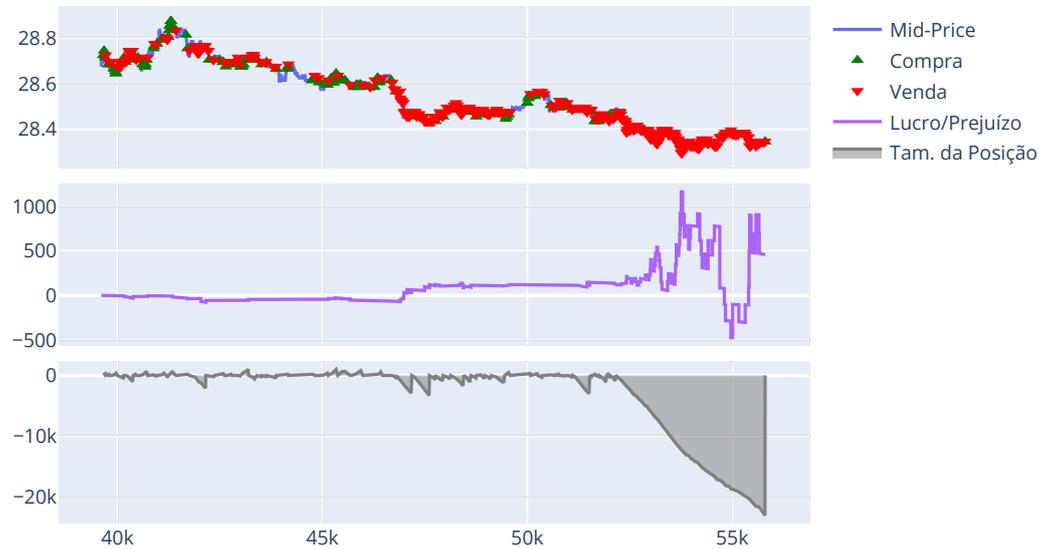


Figura 68 – Ações realizadas pela estratégia atualizada no dia de negociações 24/10/2019.



Figura 69 – Ações realizadas pela estratégia atualizada no dia de negociações 25/10/2019.

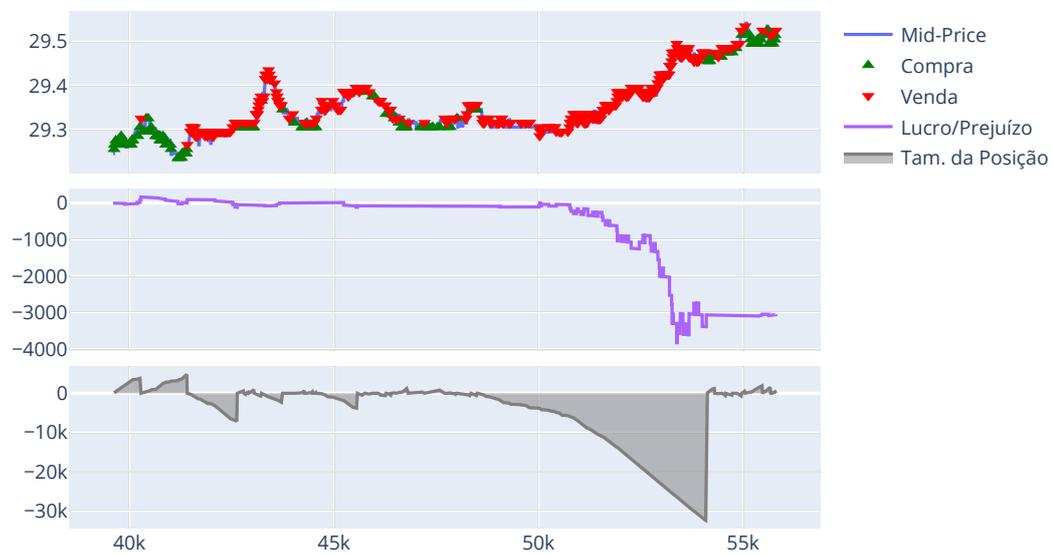


Figura 70 – Ações realizadas pela estratégia atualizada no dia de negociações 28/10/2019.

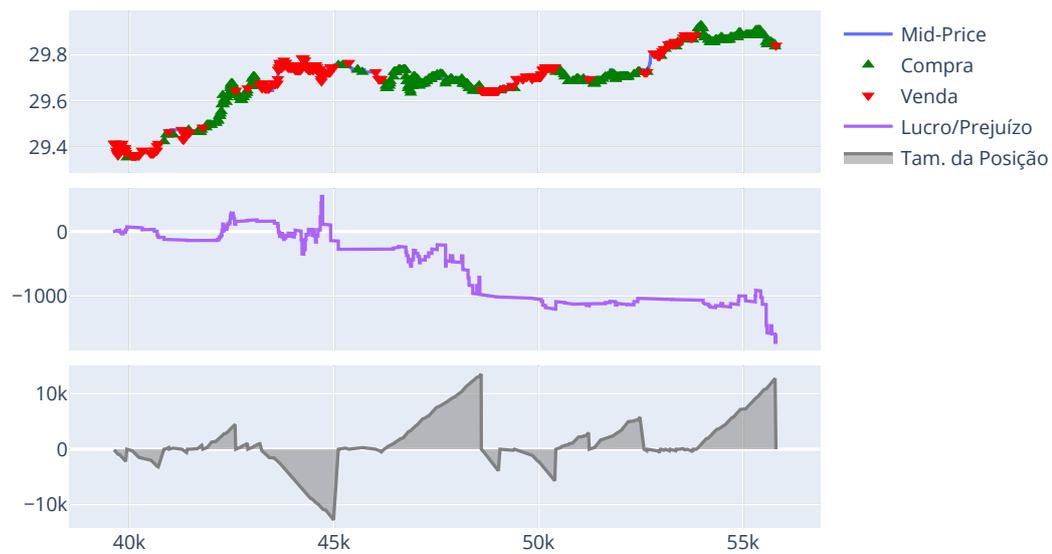


Figura 71 – Ações realizadas pela estratégia atualizada no dia de negociações 29/10/2019.

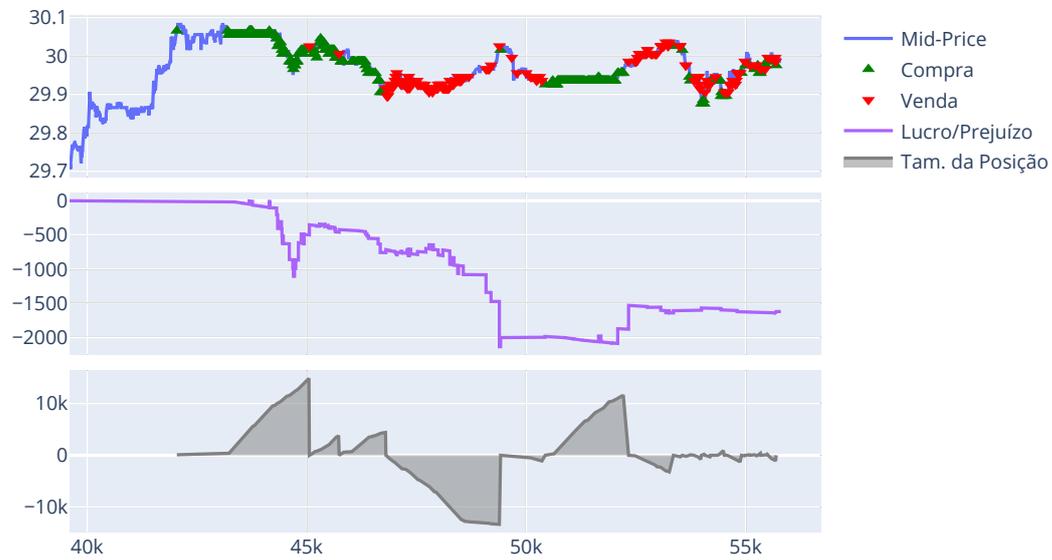


Figura 72 – Ações realizadas pela estratégia atualizada no dia de negociações 30/10/2019.



Figura 73 – Ações realizadas pela estratégia atualizada no dia de negociações 31/10/2019.

## EXECUÇÕES DIÁRIAS DA ESTRATÉGIA USANDO 2 NÍVEIS PARA O MÊS DE OUTUBRO DE 2019

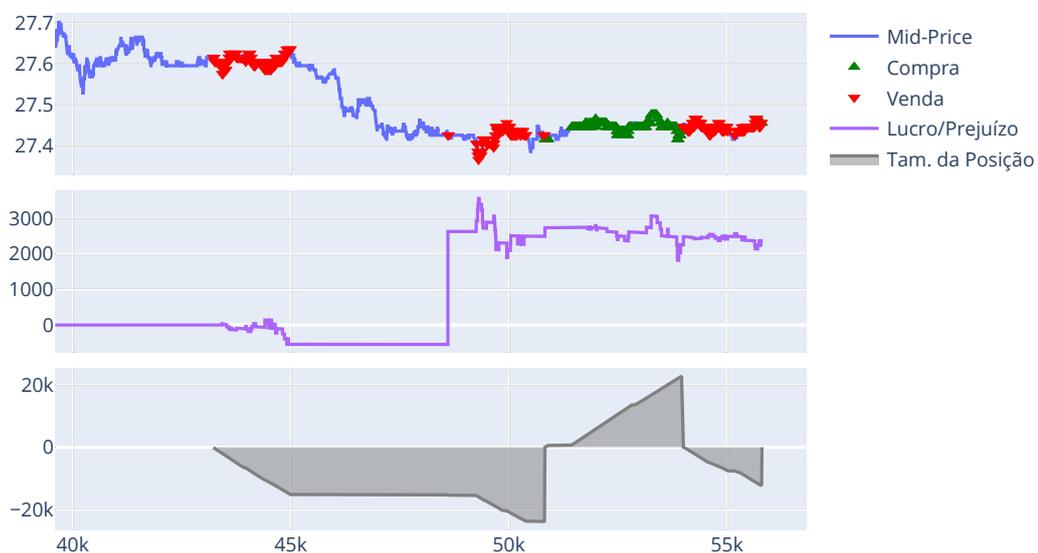


Figura 74 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 01/10/2019.



Figura 75 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 02/10/2019.

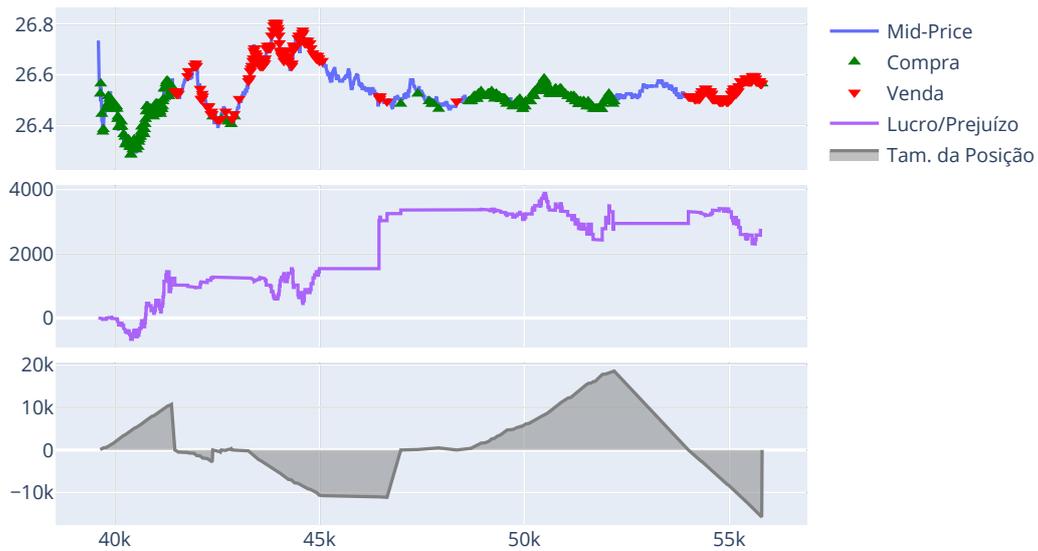


Figura 76 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 03/10/2019.

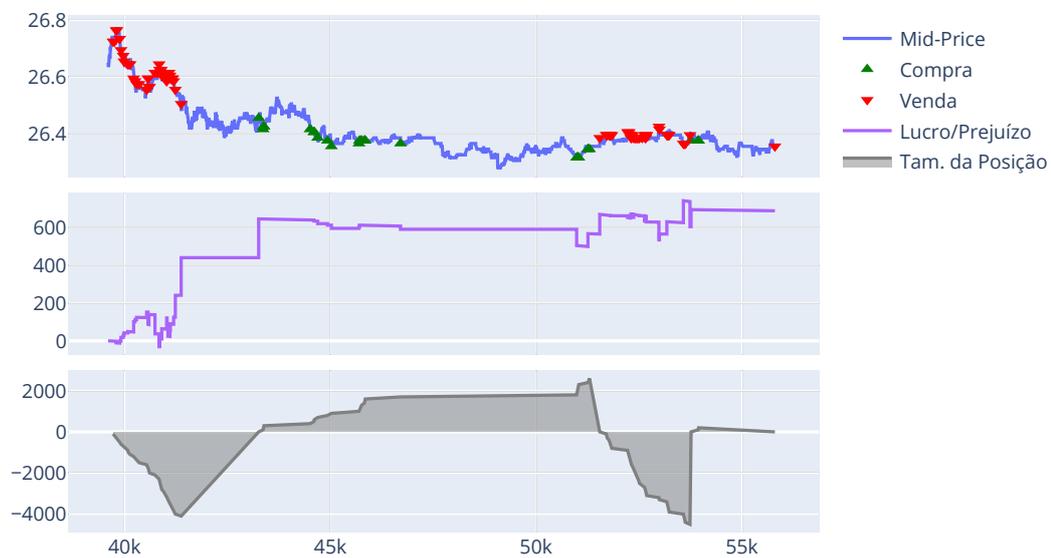


Figura 77 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 04/10/2019.

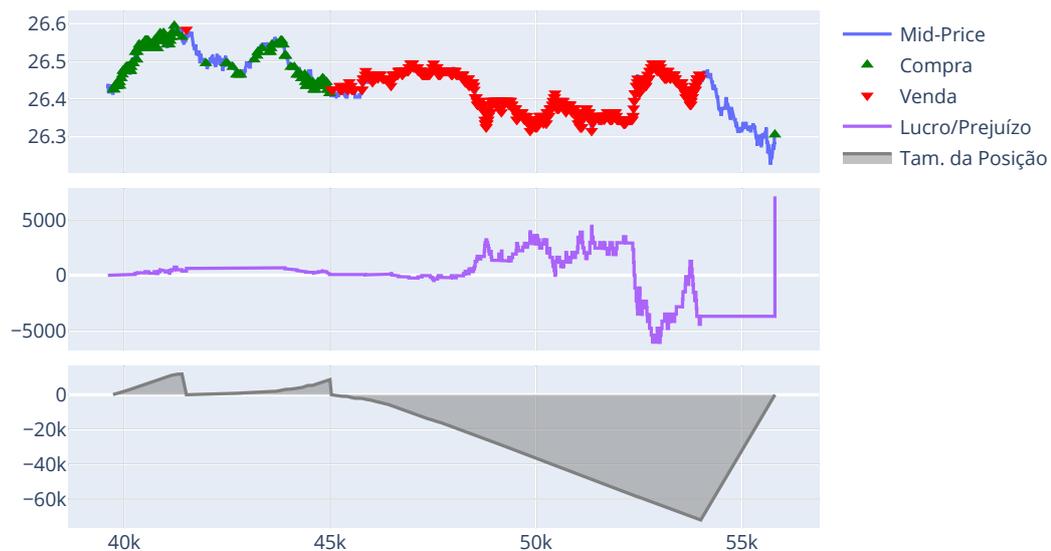


Figura 78 – Ações realizadas pela estratégia utilizando 2 níveis no dia de negociações 07/10/2019.

