



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

**Análise da Eficiência do Mercado de Ações Brasileiro em Alta Frequência.**

Arthur W. Oliveira Leite  
Mestrado em Estatística

Orientador  
Prof. Dr. Jose Augusto Fiorucci

Dezembro de 2020

## **AGRADECIMENTOS**

Agradeço, sobretudo, a Deus e ao meu Senhor Jesus Cristo, razão da minha existência e penhor da minha herança na eternidade. Agradeço a minha esposa, Sílvia Leite, pelo apoio e incentivo durante todo o tempo de dedicação. Agradeço ainda ao meu orientador, José Augusto Fiorucci, pelo acompanhamento e direcionamento da pesquisa, e aos amigos Nilson Assis e Hugo Oliveira, pela ajuda na revisão textual e melhoria da escrita.

## SUMÁRIO

AGRADECIMENTOS.....	ii
LISTA DE FIGURAS.....	v
LISTA DE FIGURAS - APÊNDICE.....	viii
LISTA DE TABELAS.....	x
LISTA DE TABELAS - APÊNDICE.....	xi
RESUMO.....	xii
ABSTRACT.....	xiii
1. INTRODUÇÃO.....	1
2. PASSEIO ALEATÓRIO.....	4
2.1 TEORIA DO PASSEIO ALEATÓRIO.....	4
3. MOVIMENTO BROWNNIANO.....	7
3.1 MOVIMENTO BROWNIANO GEOMÉTRICO.....	9
3.2 APROXIMAÇÃO DISCRETA DO MOVIMENTO BROWNIANO GEOMÉTRICO.....	11
3.3 ESTIMATIVA DA VOLATILIDADE $\sigma$ .....	13
4. OPÇÕES DE COMPRA E VENDA.....	15
4.1 MODELO DE PRECIFICAÇÃO DE BLACK, SCHOLES E MERTON.....	16
4.2 PRECIFICAÇÃO DAS OPÇÕES A PARTIR CURVA DE DISTRIBUIÇÃO DO ATIVO SUBJACENTE.....	18
5. OPERAÇÕES EM ALTA FREQUÊNCIA (HFT) E EFICIÊNCIA DO MERCADO INTRADIÁRIO.....	22
5.1 EFICIÊNCIA DO MERCADO INTRADIÁRIO.....	22
5.2 DINÂMICA DO MERCADO.....	23
6. DESCRIÇÃO DOS DADOS UTILIZADOS E RESULTADOS ENCONTRADOS.....	25
6.1 ANÁLISE DOS DADOS.....	27
7. MODELOS DE PREDIÇÃO PARA A VARIAÇÃO DOS PREÇOS E ESTRATÉGIA DE NEGOCIAÇÃO.....	36
7.1 MODELO DE REGRESSÃO LOGÍSTICA UTILIZADO.....	37
7.2 REGRESSÃO LOGÍSTICA PARA 1 (UM) MOVIMENTO ANTERIOR.....	39
7.2.1 Coeficientes e Probabilidades.....	40
7.3 REGRESSÃO LOGÍSTICA PARA 2 (DOIS) MOVIMENTOS ANTERIORES.....	41

7.3.1	Coeficientes e Probabilidades .....	42
7.4	REGRESSÃO LOGÍSTICA PARA 3 (TRÊS) MOVIMENTOS ANTERIORES .....	46
7.4.1	Coeficientes e Probabilidades .....	47
7.5	CAPACIDADE PREDITIVA DOS MODELOS .....	53
7.6	ESTRATÉGIA DE OPERAÇÕES PARA APROVEITAMENTO DA CAPACIDADE PREDITIVA DOS MODELOS.....	56
7.7	ESTRATÉGIA DE OPERAÇÃO DESCONSIDERANDO CUSTOS OPERACIONAIS	59
7.8	ESTRATÉGIA DE OPERAÇÃO CONSIDERANDO CUSTOS OPERACIONAIS .....	63
7.8.1	Diferença entre os preços e a lucratividade do modelo .....	64
7.8.2	Análise dos Resultados encontrados .....	65
7.8.3	Análise dos Custos operacionais.....	69
8.	MODELO PARA A MOVIMENTAÇÃO DOS PREÇOS DOS ATIVOS E PRECIFICAÇÃO DE OPÇÕES.....	75
8.1	UM EXEMPLO HIPOTÉTICO.....	77
8.2	ESTIMATIVA DA VOLATILIDADE $\sigma$ PARA OS DADOS DIÁRIOS.....	83
8.3	APLICAÇÃO DO MODELO PARA OS DADOS REAIS .....	84
8.4	UTILIZAÇÃO DO NOVO MODELO PARA A PRECIFICAÇÃO DE OPÇÕES .....	90
9.	CONCLUSÃO .....	94
	REFERÊNCIAS .....	97
	APÊNDICE .....	99

## LISTA DE FIGURAS

Figura 1 - Distribuição log-normal de $S(t)$ . .....	10
Figura 2 - Distribuições de $Ei \sim N(a\Delta t, \sigma^2\Delta t)$ para $a = 0,12$ , $\sigma = 0,4$ e $\Delta t = 0,1$ . .....	12
Figura 3 - Payoffs de uma Call e uma Put com Strike igual a 20 em relação ao valor de $S(T)$ . .....	16
Figura 4 - Probabilidades de execução da call e da put em $T = 1$ . .....	19
Figura 5 - Densidade de $S(t)$ , payoffs e os payoffs ponderados pela probabilidade da call e da put. ....	20
Figura 6 - Flutuação do preço entre os níveis Bid e Ask. ....	24
Figura 7 - Simulação computacional de um movimento browniano geométrico. ....	27
Figura 8 - Boxplot de $P(+)$ para diferentes períodos. ....	28
Figura 9 - Boxplot de $P(+ -)$ para diferentes períodos. ....	29
Figura 10 - Boxplot de $P(+++)$ para diferentes períodos. ....	29
Figura 11 - Número de dias estatisticamente significativos (Teste 1) .....	31
Figura 12 - Número de dias estatisticamente significativos. (Teste 2) .....	32
Figura 13 - Probabilidade Condicional para até dois movimentos à priori. ....	34
Figura 14 - Relação entre $y$ , $odd$ e $p$ . ....	39
Figura 15 - Coeficientes e significâncias do modelo logístico para 1 (um) movimento anterior. ....	40
Figura 16 - Probabilidades do modelo logístico para 1 (um) movimento anterior. ....	41
Figura 17 - Coeficientes e significâncias do modelo logístico para 2 (dois) movimentos anteriores. Dados até 15 segundos. ....	43
Figura 18 - Coeficientes e significâncias do modelo logístico para 2 (dois) movimentos anteriores. Dados de 20 segundos a 5 minutos .....	44
Figura 19 - Probabilidades do modelo logístico para 2 (dois) movimentos anteriores. ....	46
Figura 20 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados tick a tick a 3 segundos .....	48
Figura 21 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados 5 a 15 segundos. ....	49
Figura 22 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados 20 a 45 segundos. ....	50

Figura 23 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados 1 a 5 minutos.....	51
Figura 24 - Probabilidades do modelo logístico para 3 (três) movimentos anteriores.Dados tick a tick até 15 segundos.....	52
Figura 25- Probabilidades do modelo logístico para 3 (três) movimentos anteriores.Dados de 20 segundos até 5 minutos .....	53
Figura 26 - Percentuais de acerto para cada modelo logístico.....	54
Figura 27 - Estratégia de operação. ....	57
Figura 28 - Estratégias de operação bem-sucedidas .....	58
Figura 29 - Estratégias de operação malsucedidas .....	59
Figura 30 - Lucros Percentuais para os dados tick a tick e de até 15 segundos.....	61
Figura 31 - Lucros Percentuais para os dados de 20 segundos até 5 minutos .....	62
Figura 32 - Resultados Percentuais para os dados tick a tick e de até 15 segundos.....	67
Figura 33 - Resultados Percentuais para os dados de 20 segundos até 5 minutos..	68
Figura 34 - Custos Percentuais para os dados tick a tick e de até 15 segundos .....	69
Figura 35 - Custos Percentuais para os dados de 20 segundos até 5 minutos .....	70
Figura 36 - Lucros, Custos e Resultados Percentuais Acumulados Aritmeticamente para os dados tick a tick até 5 segundos. ....	72
Figura 37 - Lucros, Custos e Resultados Percentuais Acumulados Aritmeticamente para os dados de 10 a 45 segundos. ....	73
Figura 38 - Lucros, Custos e Resultados Percentuais Acumulados Aritmeticamente para os dados de 1 a 5 minutos. ....	74
Figura 39 - Movimentação diária para 20 simulações diferentes. ....	78
Figura 40 - Histograma e densidade aproximada de 100.000 simulações diárias. ...	79
Figura 41 - Densidade de $\log(S_n)$ e de uma distribuição $N(2,3; 0,09452)$ .....	79
Figura 42 - Densidade de $S_n$ e a distribuição log-normal do movimento Browniano geométrico.....	80
Figura 43 - Distribuição normal que ajustou $\log(S_n)$ , $N(2,3; 0,09452)$ , e a distribuição de $\ln S(t) \sim N(\ln S_0 + at, \sigma^2 t)$ .....	81
Figura 44 - Distribuições log-normais que ajustaram $S_n$ em cada simulação e a distribuição de $S(t)$ .....	82
Figura 45 - Distribuições normais que ajustaram $\log(S_n)$ em cada simulação e a distribuição de $\ln S(t)$ .....	83

Figura 46 - Histograma e densidade aproximada de 40.000 simulações para o dia 31/01/2018. ....	86
Figura 47 - Densidade de $\log(S_n)$ e da curva normal ajustada pelos parâmetros da média a variância de $\log(S_n)$ .....	87
Figura 48 - Densidade de $S_n$ e a distribuição log-normal do movimento Browniano geométrico.....	88
Figura 49 - Distribuição log-normal ajustada de $S_n$ e a distribuição do movimento Browniano geométrico.....	88
Figura 50 - Distribuições log-normais que ajustaram $S_n$ em cada simulação e a distribuição de $S(t)$ .....	90
Figura 51 - Densidade de $S_n$ e a distribuição log-normal do movimento Browniano geométrico.....	91
Figura 52 - Densidades, payoffs e os payoffs ponderados pela probabilidade da call e da put. ....	92

## LISTA DE FIGURAS - APÊNDICE

Figura A 1 - Boxplot de P(+) para diferentes períodos (ABEV3). .....	99
Figura A 2 - Boxplot de P(+) para diferentes períodos (BBDC4). .....	99
Figura A 3 - Boxplot de P(+) para diferentes períodos (VALE3). .....	100
Figura A 4 - Boxplot de P(+ -) para diferentes períodos (ABEV3) .....	100
Figura A 5 - Boxplot de P(+ -) para diferentes períodos (BBDC4). .....	101
Figura A 6 - Boxplot de P(+ -) para diferentes períodos (VALE3). .....	101
Figura A 7 - Número de dias estatisticamente significativos (Teste 1) - ABEV3 .....	102
Figura A 8 - Número de dias estatisticamente significativos (Teste 1) - BBDC4.....	102
Figura A 9 - Número de dias estatisticamente significativos (Teste 1) - VALE3.....	103
Figura A 10 - Número de dias estatisticamente significativos. (Teste 2) - ABEV3 ..	103
Figura A 11 - Número de dias estatisticamente significativos. (Teste 2) - BBDC4 ..	104
Figura A 12 - Número de dias estatisticamente significativos. (Teste 2) - VALE3...	104
Figura A 13 - Probabilidade Condicional para até dois movimentos à priori (ABEV3). .....	107
Figura A 14 - Probabilidade Condicional para até dois movimentos à priori (BBDC4). .....	107
Figura A 15 - Probabilidade Condicional para até dois movimentos à priori (VALE3). .....	108
Figura A 16 - Percentuais de acerto para cada modelo logístico (ABEV3). .....	108
Figura A 17 - Percentuais de acerto para cada modelo logístico (BBDC4). .....	109
Figura A 18 - Percentuais de acerto para cada modelo logístico (VALE3). .....	109
Figura A 19 - Lucros Percentuais para os dados tick a tick e de até 15 segundos (ABEV3) .....	110
Figura A 20 - Lucros Percentuais para os dados de 20 segundos até 5 minutos (ABEV3). .....	111
Figura A 21 - Lucros Percentuais para os dados tick a tick e de até 15 segundos (BBDC4). .....	112
Figura A 22 - Lucros Percentuais para os dados de 20 segundos até 5 minutos (BBDC4). .....	113
Figura A 23 - Lucros Percentuais para os dados tick a tick e de até 15 segundos (VALE3). .....	114



Figura A 24 - Lucros Percentuais para os dados de 20 segundos até 5 minutos (VALE3).....	115
Figura A 25 - Resultados Percentuais para os dados tick a tick e de até 15 segundos (ABEV3) .....	116
Figura A 26 - Resultados Percentuais para os dados de 20 segundos até 5 minutos (ABEV3) .....	117
Figura A 27 - Resultados Percentuais para os dados tick a tick e de até 15 segundos (BBDC4).....	118
Figura A 28 - Resultados Percentuais para os dados de 20 segundos até 5 minutos (BBDC4).....	119
Figura A 29 - Resultados Percentuais para os dados tick a tick e de até 15 segundos (VALE3).....	120
Figura A 30 - Resultados Percentuais para os dados de 20 segundos até 5 minutos (VALE3).....	121

## LISTA DE TABELAS

Tabela 1 - Resultado do Teste 1 para todos os dias agrupados. ....	32
Tabela 2 - Resultado do Teste 2 para todos os dias agrupados. ....	33
Tabela 3 - Estimativas das probabilidades condicionais dos movimentos reversos e dos movimentos de continuação. ....	35
Tabela 4 - Variáveis explicativas dependentes até 3 movimentos anteriores .....	37
Tabela 5 - Relação entre $y$ , $odd$ e $p$ .....	38
Tabela 6 - Limiar Preditivo da probabilidade $p$ e valor predito .....	54
Tabela 7 - Acurácia dos Modelos .....	55
Tabela 8 – Medianas e Médias dos Lucros Diários .....	62
Tabela 9 - Custos Operacionais B3 - Day Trade. Fonte B3 [19] .....	63
Tabela 10 – Medianas e Médias dos Resultados Diários .....	68
Tabela 11 – Medianas e Médias dos Custos Diários .....	70
Tabela 12 - Médias diárias dos Lucros, Custos operacionais e dos Resultados .....	71
Tabela 13 - $P(+ -)$ , $P(+ +)$ , $\mu_d$ e $\mu_u$ utilizados nas simulações .....	81
Tabela 14 - Intervalo, $n$ , $P(+ +)$ , $P(+ -)$ , $\mu_d$ e $\mu_u$ utilizados nas simulações .....	89
Tabela 15 - $P(+)$ - GBM e $P(+)$ - 20 dias anteriores .....	90

## LISTA DE TABELAS - APÊNDICE

Tabela A 1 - Resultado do Teste 1 para todos os dias agrupados (ABEV3)	105
Tabela A 2 - Resultado do Teste 1 para todos os dias agrupados (BBDC4) .....	105
Tabela A 3 - Resultado do Teste 1 para todos os dias agrupados (VALE3) .....	105
Tabela A 4 - Resultado do Teste 2 para todos os dias agrupado (ABEV3).....	106
Tabela A 5 - Resultado do Teste 2 para todos os dias agrupado (BBDC4). .....	106
Tabela A 6 - Resultado do Teste 2 para todos os dias agrupado (VALE3).....	106

## RESUMO

Este estudo busca demonstrar de forma empírica que a hipótese de eficiência do mercado não é válida para dados em alta frequência. Analisa-se como a eficiência do mercado pode ser avaliada para dados intradiários e, a partir de dados do mercado brasileiro, apresenta-se um estudo que questiona essa eficiência. Propõe-se uma estratégia de operações para obter vantagens financeiras a partir da ineficiência observada na movimentação dos ativos. Apresenta-se ainda o movimento Browniano geométrico, comumente utilizado como modelo de precificação, e sugere-se um novo modelo que incorpore essa ineficiência. A partir dele, realiza-se uma estimação mais precisa para os preços de opções de mercado europeias dentro de um mesmo dia.

**Palavras - chave:** mercado brasileiro, eficiência do mercado, passeio aleatório, probabilidade preditiva, regressão logística, estratégia de operações, precificação de opções.

## **ABSTRACT**

This study seeks to demonstrate empirically that the market efficiency hypothesis is not valid for high frequency data. It analyzes how the efficiency of the market can be assessed for intraday data. Is presented a study, based on data from the Brazilian Market, that questions this efficiency. An operations strategy is proposed to obtain financial advantages from the inefficiency observed in the movement of assets. The geometric Brownian movement, commonly used as a pricing model, is also presented, and a new model is suggested that incorporates this inefficiency. From it, a more accurate estimate is made for the prices of European market options within the same day.

**Keywords:** Brazilian market, market efficiency, random walk, predictive probability, logistic regression, operations strategy, option pricing.



## 1. INTRODUÇÃO

O conceito de eficiência é fundamental no mundo financeiro. O termo é usado para descrever um mercado em que toda a informação relevante está embutida no preço dos ativos negociados [1].

Num mercado eficiente, os valores futuros dos ativos não podem ser previstos, desde que os agentes não possuam acesso a informações privilegiadas. Considera-se ainda que as informações importantes estão disponíveis livremente para todos os participantes [2].

Considerando que um mercado é suficientemente competitivo, os investidores não podem esperar obter retornos superiores ao do próprio mercado a partir de suas estratégias de investimento. Apesar dessa ideia ser bastante clara e simples atualmente, no começo do século passado, ela era inovadora e pouco estudada pela comunidade científica [1].

O conceito de eficiência de mercado foi antecipado no final do século XIX por Bachelier. Em seu trabalho, ele descreve que os eventos passados, presentes e futuros de um ativo já estão refletidos em seu valor de mercado. Descreve ainda que, tendo em vista que o mercado não prediz suas próprias flutuações, espera-se que elas sejam mais ou menos prováveis e que essa probabilidade possa ser descrita matematicamente [1].

Essa ideia deu origem à formulação matemática que antecipou não apenas a derivação do processo de Einstein-Weiner do movimento Browniano, como muitos dos resultados analíticos que foram desenvolvidos pelos acadêmicos financeiros na segunda metade daquele século [1].

Nessa linha, pode-se citar o desenvolvimento da teoria do passeio aleatório para o movimento dos ativos financeiros, definida mais precisamente por Fama em 1965. Ela lançou sérias dúvidas sobre inúmeros outros métodos existentes para se prever a movimentação dos preços dos ativos - métodos, aliás, de grande popularidade fora dos meios acadêmicos [2].

Dois desses métodos largamente difundidos na análise e predição dos valores futuros dos ativos são denominados de método gráfico e método fundamentalista.

O primeiro estabelece que a movimentação passada do preço de determinado ativo irá influenciar probabilisticamente a movimentação futura desses preços. Assim,

sucessivas mudanças no preço de determinado ativo seriam dependentes, existindo uma probabilidade condicional dos preços futuros em relação aos passados [2].

O método fundamentalista, ou de valor intrínseco, por outro lado, estabelece que, em determinado instante, qualquer ativo possui um valor intrínseco (valor de equilíbrio) que depende do potencial de ganho desse ativo. Tal potencial é determinado por fatores fundamentais ligados à empresa subjacente, tais como qualidade da gestão empresarial, quantidade de vendas, dividendos e perspectivas futuras para a indústria e cenário econômico [2, 7].

A partir do conhecimento do valor intrínseco de um ativo, seria possível determinar qual a tendência de comportamento de seu valor de mercado, tendo em vista que o valor futuro tenderia a se aproximar de seu valor de equilíbrio.

Num mercado eficiente, a competição entre os diferentes participantes levaria a uma situação em que os preços atuais dos ativos já iriam refletir todas as informações dos eventos passados e o que se espera dos eventos futuros. Em outras palavras, o preço atual sempre seria uma boa estimativa do valor intrínseco do ativo [2].

O presente estudo busca questionar a hipótese de eficiência do mercado e do passeio aleatório para os dados intradiários a partir da verificação empírica de que o conhecimento passado da série de preços possa sim ser utilizado para aumentar os ganhos esperados.

Na Seção 2, mostra-se o processo estocástico denominado passeio aleatório e sua aplicação ao mercado financeiro. Na Seção 3, detalha-se o movimento browniano e sua variação mais utilizada, o movimento browniano geométrico. Na Seção 4, trata-se das opções de compra e venda de mercado e do modelo de precificação de Black, Scholes e Merton. Na seção 5, descreve-se as operações em alta frequência (HFT) e a hipótese de eficiência do mercado para dados intradiários. Na Seção 6, traz-se uma discriminação dos dados utilizados e os resultados encontrados para a verificação da hipótese de eficiência do mercado.

Na Seção 7, propõe-se modelos de predição para a variação dos preços dos ativos e uma estratégia de negociação para o aproveitamento de sua capacidade preditiva. Na Seção 8, trata-se de um novo modelo de precificação de ativos e mostra como utilizá-lo para se obter uma estimação mais precisa dos preços de opções



européias dentro de um mesmo dia. Por fim, na seção 9, apresenta-se a conclusão do presente estudo.

## 2. PASSEIO ALEATÓRIO

Passeio aleatório é definido como um processo estocástico formado pela soma sucessiva de variáveis aleatórias independentes e identicamente distribuídas (i.i.d), sendo um dos tópicos mais estudados na teoria da probabilidade [16].

A ideia por trás desse processo foi concebida ainda no século XVI, como um modelo para os jogos de azar, estando intimamente ligada ao nascimento da própria teoria da probabilidade [17].

Sobre a reta dos números inteiros  $\mathbb{Z}$ , o passeio aleatório é discriminado pela soma da seguinte sequência de variáveis aleatórias i.i.d [16]:

$$S_n = x + X_1 + \dots + X_n, \quad x \in \mathbb{Z}, \quad X_n \in \{\pm 1\} \quad (1)$$

Esse processo pode ser visto como uma cadeia de Markov com espaços de estados definidos pelos inteiros  $x = 0, \pm 1, \pm 2, \dots$  e que, para algum  $0 < p < 1$ ,

$$P_{x,x+1} = p = 1 - P_{x,x-1} \quad (2)$$

Em que  $P_{x,y}$  é a probabilidade de transição entre os estados  $x$  e  $y$ . Se  $p = 1/2$ , ele passa a ser chamado de passeio aleatório simétrico [11].

Nos números inteiros, o passeio aleatório pode ser interpretado como o montante obtido por um jogador que, em cada movimento, ganha ou perde uma unidade monetária. Nesse caso, se  $p > 1/2$ , é possível mostrar que existe uma probabilidade positiva de que a fortuna do jogador irá crescer indefinidamente [11].

Por outro lado, para  $p \leq 1/2$ , é demonstrado que o jogador irá zerar os seus recursos financeiros, com probabilidade 1, diante de um adversário infinitamente rico [11]. Essa característica é conhecida como "Problema da Ruína do Jogador". Nota-se que, para o passeio aleatório simétrico, em que  $p = 1/2$ , o jogador também irá à falência com probabilidade 1.

### 2.1 TEORIA DO PASSEIO ALEATÓRIO

A teoria ou hipótese do passeio aleatório afirma que a movimentação dos ativos nos mercados financeiros segue um processo estocástico do tipo passeio aleatório [2].

Essa teoria, juntamente com o conceito de eficiência do mercado, tornou-se ícone da economia financeira moderna e continua a despertar a imaginação de acadêmicos e profissionais de investimento [17].

Sua aceitação implica que uma série de preços não possui memória, ou seja, o valor passado da série não pode ser utilizado para prever o valor futuro [2]. Em outros termos, o comportamento futuro de uma série de preços não será mais previsível do que uma série de números aleatórios acumulados [7].

No mundo real, o valor intrínseco de um ativo nunca pode ser determinado com exatidão. Sempre existirão discordâncias entre os participantes no que se refere à estimação desse valor, ocasionando diferenças entre o valor atual e o próprio valor intrínseco [2].

A teoria do passeio aleatório afirma que, apesar dessa incerteza, as iniciativas dos diferentes participantes farão com que os preços dos ativos passem aleatoriamente sobre seus valores intrínsecos [2].

É evidente que novas informações podem acarretar mudança desse valor. Entretanto, os efeitos dessa nova informação não são refletidos instantaneamente nos preços e, na prática, observam-se duas implicações decorrentes da imprecisão e incerteza em torno dessa nova informação [2]:

- a) primeiramente, os preços irão variar demasiadamente acima ou abaixo do valor intrínseco;
- b) segundo, haverá um atraso para o ajuste completo do preço atual para o novo valor intrínseco. Esse atraso será, por si mesmo, uma variável aleatória independente.

Assim, sucessivas mudanças nos preços dos ativos serão independentes, comportando-se como um mercado do tipo passeio aleatório [2].

É improvável que a hipótese de passeio aleatório preveja uma exata descrição do comportamento dos preços dos ativos. No entanto, embora sucessivas mudanças nos preços possam não ser estritamente independentes, essa dependência pode ser tão pequena que se torna irrelevante [2].

Nesse sentido, a hipótese de passeio aleatório somente pode ser considerada válida desde que o conhecimento do comportamento passado da série não possa ser utilizado para aumentar os ganhos esperados [2].

Ao longo dos anos, um grande número de pesquisas empíricas foi realizado para avaliar a teoria do passeio aleatório. As principais abordagens dessas pesquisas foram testar a hipótese de sucessivas mudanças nos preços serem independentes. Os resultados encontrados sempre foram consistentes no sentido de validação dessa teoria [2].

No entanto, estudos mais recentes encontraram sim rentabilidades superiores aos riscos assumidos em alguns modelos de operações [7]. O consenso desses estudos é de que os preços são sim previsíveis, ao menos em parte do tempo.

Foi verificado que modelos de previsão que utilizam dados intradiários apresentam melhores rentabilidades do que aqueles que utilizam dados de períodos diários ou superiores [7].

Isso ocorre pois, quando um certo padrão de movimentação é verificado, ele é incorporado por outros operadores em seus sistemas de negociação, provocando uma adoção generalizada que eleva os preços dos ativos para cima e para baixo até o ponto desse padrão ser eliminado [7].

### 3. MOVIMENTO BROWNNIANO

O movimento Browniano desempenha um papel central em várias áreas, como na teoria da probabilidade, teoria dos processos estocásticos, física e em finanças, sendo considerado uma aproximação muito boa de vários fenômenos reais [14]. É ainda um dos mais úteis processos estocásticos na teoria de probabilidade aplicada [11].

Seu nome foi dado em homenagem ao botânico Robert Brown (1820) que, em suas pesquisas, descreveu o movimento de pequenas partículas totalmente imersas em um líquido ou gás [11].

No início do século XX, Louis Bachelier, Albert Einstein e Norbert Wiener iniciaram o desenvolvimento matemático do movimento Browniano [14].

Em 1900, Bachelier capturou muitas das propriedades essenciais desse movimento. A sua principal contribuição foi mostrar que os movimentos dos preços dos ativos financeiros poderiam ser bem descritos por esse processo, descobrindo assim uma nova área de aplicação dos processos estocásticos [14].

O seu trabalho lançou as bases para a teoria moderna do movimento Browniano, a partir do desenvolvimento de um modelo matemático aplicável às flutuações dos preços dos ativos financeiros. Ele argumentou que as pequenas flutuações dos preços dos ativos seriam normalmente distribuídas e independentes em relação aos seus valores passados [10].

Em 1905, desconhecendo o trabalho realizado por Bachelier, Einstein desenvolveu o seu modelo matemático para o movimento Browniano inspirado na teoria cinética da matéria [10].

Ele mostrou que o movimento Browniano poderia ser explicado considerando que uma partícula imersa está sendo continuamente bombardeada pelas moléculas ao seu redor [11]. Provou ainda que, sob as premissas da teoria molecular do calor, corpos de diâmetro da ordem de  $10^{-3}$  mm, suspensos num líquido, deveriam executar um movimento aleatório observável [10].

Em 1923, foi Norbert Wiener quem deu uma definição mais concisa desse processo, sendo o primeiro a utilizar uma base matemática rigorosa para descrever o movimento Browniano. Devido a sua contribuição, atualmente o movimento Browniano também é conhecido como processo de Wiener [10, 11, 14].

O movimento Browniano pode ser entendido a partir de um passeio aleatório simétrico sobre a reta dos números inteiros [11]. Considerando inicialmente uma cadeia de Markov com espaços de estados definidos pelos inteiros  $x = 0, \pm 1, \pm 2, \dots$ , estabelece-se a seguinte probabilidade de transição:

$$P_{x,x+1} = 1/2 = 1 - P_{x,x-1}$$

Agora, aumenta-se a velocidade desse processo, de forma que, a cada intervalo  $\Delta t$ , seja dado um passo de tamanho  $\Delta x$  cada vez menores. Denotando  $X(t)$  como a posição no tempo  $t$ , tem-se [11]:

$$X(t) = \Delta x(X_1 + \dots + X_{\lfloor \frac{t}{\Delta t} \rfloor})$$

Em que:

$$X_i = \pm 1$$

$\lfloor \frac{t}{\Delta t} \rfloor$  é o maior inteiro menor ou igual a  $t/\Delta t$ .

$X_i$  é independente com  $P\{X_i = 1\} = P\{X_i = -1\} = 1/2$

Como  $E[X_i] = 0$ ,  $Var(X_i) = E[X_i^2] = 1$ . Pode-se concluir ainda que:

$$E[X(t)] = 0 \text{ e que } Var(X(t)) = (\Delta x)^2 \lfloor \frac{t}{\Delta t} \rfloor$$

Agora, tomando  $\Delta x = \sigma\sqrt{\Delta t}$ , para uma constante positiva  $\sigma$ , e fazendo  $\Delta t \rightarrow 0$ , é possível mostrar que o processo estocástico  $\{X(t), t \geq 0\}$  possui as seguintes características [11]:

i.  $X(0) = 0$ ;

ii.  $\{X(t), t \geq 0\}$  possui incrementos estacionários e independentes;

iii. para todo  $t > 0$ ,  $X(t)$  é normalmente distribuída com média 0 e variância  $\sigma^2 t$ .

Esse processo, por definição, é um movimento Browniano. Pode ser provado que  $\{X(t), t \geq 0\}$  é uma função contínua. Para  $\sigma^2 = 1$ , o processo é denominado movimento Browniano padrão e será denotado por  $B(t)$ . Nota-se que qualquer movimento Browniano pode ser transformado no processo padrão pela seguinte atribuição  $B(t) = X(t)/\sigma$  [11]. Ainda, para uma variável  $Y$  normalmente distribuída com média 0 e variância 1, ou seja  $Y \sim N(0,1)$ , tem-se que  $B(t) = Y\sqrt{t}$ .

Tendo em vista que o movimento Browniano pode ser interpretado a partir de um limite aplicado ao passeio aleatório simétrico, o "Problema da Ruína do Jogador" também é aplicável nesse processo. Ou seja, se os ganhos de um jogador seguir um

movimento Browniano, ele inevitavelmente irá à falência frente a um adversário infinitamente rico.

### 3.1 MOVIMENTO BROWNIANO GEOMÉTRICO

A utilização do movimento Browniano para descrever o comportamento dos preços dos ativos financeiros permite que eles atinjam valores negativos, o que não é uma propriedade desejável, sendo considerado uma deficiência óbvia na utilização do processo para esse fim [10, 14].

Para remediar esse problema, Paul Samuelson, em 1960, introduziu um modelo denominado Browniano geométrico (GBM, do inglês *Geometric Brownian Motion*), no qual o preço de determinado ativo  $S(t)$  é dado por [10]:

$$S(t) = S(0) \exp(at + \sigma B(t)) \quad (3)$$

Em que  $B(t)$  é um movimento Browniano padrão e os parâmetros  $a$  e  $\sigma$  são constantes. Aplicando o *log* em ambos os lados do modelo, chega-se a:

$$\ln S(t) = \ln S(0) + at + \sigma B(t) \quad (4)$$

Como  $B(t) \sim N(0, t)$ , pode-se concluir que  $\ln S(t)$  é normalmente distribuída com média  $(\ln S(0) + at)$  e variância  $\sigma^2 t$ , ou seja:

$$\ln S(t) \sim N(\ln S(0) + at, \sigma^2 t) \quad (5)$$

Em outros termos, afirma-se que  $S(t)$  possui uma distribuição *log-normal*. Uma variável que possui uma distribuição *log-normal* pode assumir qualquer valor entre 0 a infinito. Ao contrário da normal, essa distribuição é enviesada, de modo que sua média, mediana e moda são diferentes entre si [18]. A Figura 1 ilustra a distribuição de  $S(t)$  para  $S(0) = 10$ ,  $a = 0,12$ ,  $\sigma = 0,4$ , em  $t = 1$ .

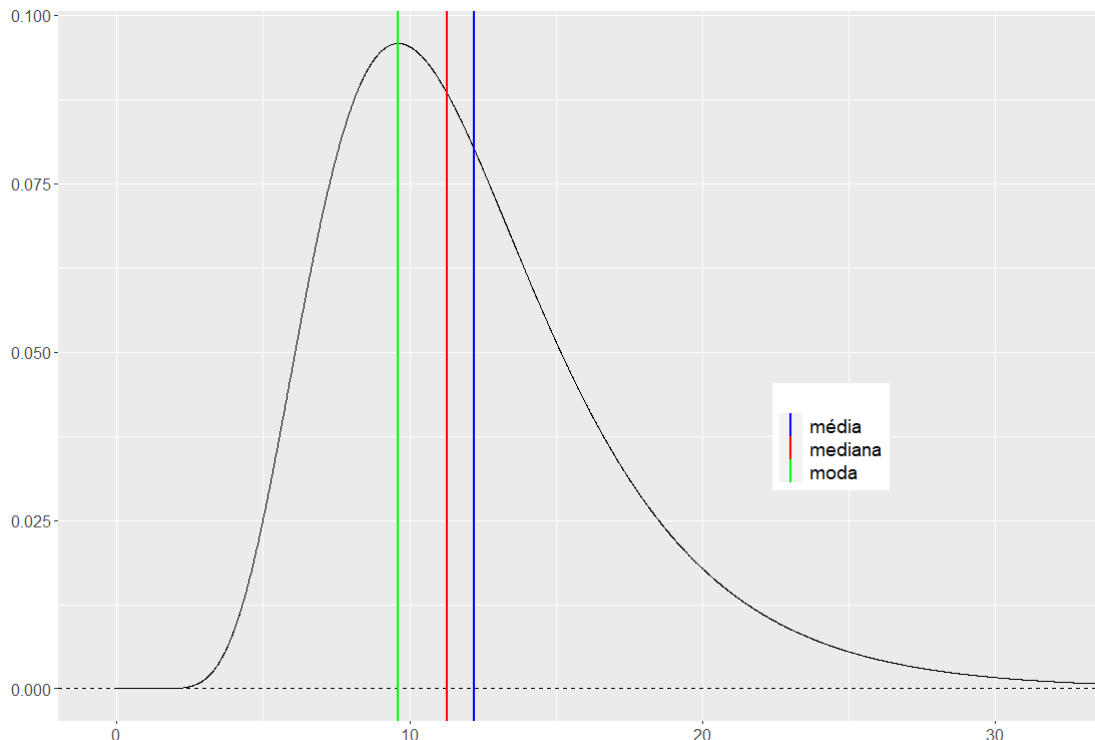


Figura 1 - Distribuição log-normal de  $S(t)$ .

Dado que  $B(t) \sim N(0, t)$ , tem-se também que  $E[e^{\sigma B(t)}] = \exp\left(\frac{1}{2}\sigma^2 t\right)$ . Assim, para  $a = \mu - \frac{1}{2}\sigma^2$ , pode-se concluir que  $E(S(t)) = S(0)e^{\mu t}$ . O parâmetro  $\mu$  é conhecido como taxa de crescimento esperada e o parâmetro  $\sigma$  representa a volatilidade, medindo o desvio padrão do *log-retorno* [10].

De fato, o desvio padrão de  $\log(S(t+h)/S(t))$  é dado por  $\sigma\sqrt{h}$  [10]. Pode ser mostrado ainda que a variância de  $S(t)$  é dada por  $S(0)^2 e^{2\mu t} (e^{\sigma^2 t} - 1)$  [18].

O movimento Browniano geométrico é útil na modelagem dos preços dos ativos no tempo quando as variações percentuais são independentes e identicamente distribuídas [11].

A mudança do movimento Browniano aritmético para o geométrico foi uma forma simples de garantir que os preços dos ativos sejam sempre positivos. Entretanto, tal alteração influenciou profundamente todo o desenvolvimento matemático, tendo em vista a não linearidade da função exponencial. Para a análise desse novo processo, foi necessário a utilização dos cálculos de Itô [10].



Aplicando a fórmula de Itô para o movimento Browniano geométrico e considerando  $a = \mu - \frac{1}{2}\sigma^2$ , pode-se concluir que  $S(t)$  satisfaz a seguinte equação diferencial estocástica [10, 11]:

$$dS(t) = \mu S(t)dt + \sigma S(t)dB(t) \quad \text{ou} \quad dS(t)/S(t) = \mu dt + \sigma dB(t) \quad (6)$$

Essa é uma representação muito mais intuitiva, sendo o modelo mais utilizado do comportamento do preço de ações [18]. Vê-se que a variação percentual de  $S(t)$  segue uma tendência linear  $\mu dt$  que é perturbada por um ruído estocástico  $\sigma dB(t)$ . O parâmetro  $\mu$  pode ser visto como a taxa média de retorno e pode-se concluir ainda que  $S(t)$  é um martingale se  $\mu$  for igual a zero [10, 14].

O parâmetro  $\sigma$  representa a volatilidade e é uma medida do risco de determinado ativo, tendo em vista que, quanto maior o seu valor, maior a flutuação de  $S(t)$  [10]. O quadrado desse parâmetro,  $\sigma^2$ , é chamado de taxa de variância [18].

Ao se considerar o movimento Browniano Geométrico uma boa representação do comportamento de ativos financeiros, o conhecimento dessa equação diferencial se torna essencial para as estratégias de compra e venda [10].

A utilização desse processo como modelo para a movimentação de ativos tornou-se cada vez mais aceita no decorrer dos anos. Em 1973, Black, Scholes e Merton utilizaram-no para o desenvolvimento de um trabalho relacionado à precificação de derivativos financeiros. Esse estudo é considerado um dos mais célebres já realizados na área e rendeu a dois de seus autores o prêmio Nobel de economia do ano de 1997 [10, 14].

### 3.2 APROXIMAÇÃO DISCRETA DO MOVIMENTO BROWNIANO GEOMÉTRICO

A versão em tempo discreto para o movimento Browniano geométrico pode ser representada por:

$$S(t + \Delta t) = S(t) \exp(a\Delta t + \sigma Y\sqrt{\Delta t}) \quad (7)$$

ou ainda

$$S_i = S_{i-1} \exp(a\Delta t + \sigma Y\sqrt{\Delta t}), \quad i = 1, 2, \dots \quad (8)$$

Em que  $Y$  é uma variável aleatória com distribuição normal de média 0 e variância 1, ou seja  $Y \sim N(0,1)$ . Em cada etapa  $i$ , o preço anterior será multiplicado pela exponencial de  $E_i = a\Delta t + \sigma Y\sqrt{\Delta t}$ , em que  $E_i \sim N(a\Delta t, \sigma^2 \Delta t)$ . Considerando um

tempo total  $T$  e tomando  $n$  incrementos dentro dele, de tal forma que  $\Delta t = T/n$ , tem-se, a partir de  $t = 0$ :

$$i) S(\Delta t) = S(0) \exp(E_1);$$

$$ii) S(2\Delta t) = S(\Delta t) \exp(E_2) = S(0) \exp(E_1) \exp(E_2) = S(0) \exp(E_1 + E_2)$$

Ocorre que, ao somar duas distribuições normais independentes, o resultado é uma distribuição também normal cuja média e variância são a soma das médias e variâncias das distribuições originais [18]. Portanto,  $E_1 + E_2 \sim N(a2\Delta t, \sigma^2 2\Delta t)$ , ou seja  $E_1 + E_2 = a2\Delta t + \sigma Y \sqrt{2\Delta t}$ .

Continuando o processo até a etapa  $n$ , tem-se:

$$S(n\Delta t) = S(0) \exp(E_1 + \dots + E_n),$$

em que  $E_1 + \dots + E_n \sim N(an\Delta t, \sigma^2 n\Delta t) \leftrightarrow N(aT, \sigma^2 T)$ . Assim,

$$S(n\Delta t) = S(0) \exp(an\Delta t + \sigma Y \sqrt{n\Delta t}),$$

ou ainda:

$$S(T) = S(0) \exp(aT + \sigma Y \sqrt{T}) \quad (9)$$

É importante observar que, em cada etapa, a probabilidade de alta do ativo será a probabilidade de  $E_i \sim N(a\Delta t, \sigma^2 \Delta t)$  ser maior do que 0. A Figura 2 mostra a distribuição de  $E_i$  com  $a = 0,12$ ,  $\sigma = 0,4$  e  $\Delta t = 0,5$ . A probabilidade de  $E_i$  ser maior do que 0 foi de 58,4%, indicada pela área sob cada curva. A linha em vermelho mostra a média da distribuição.

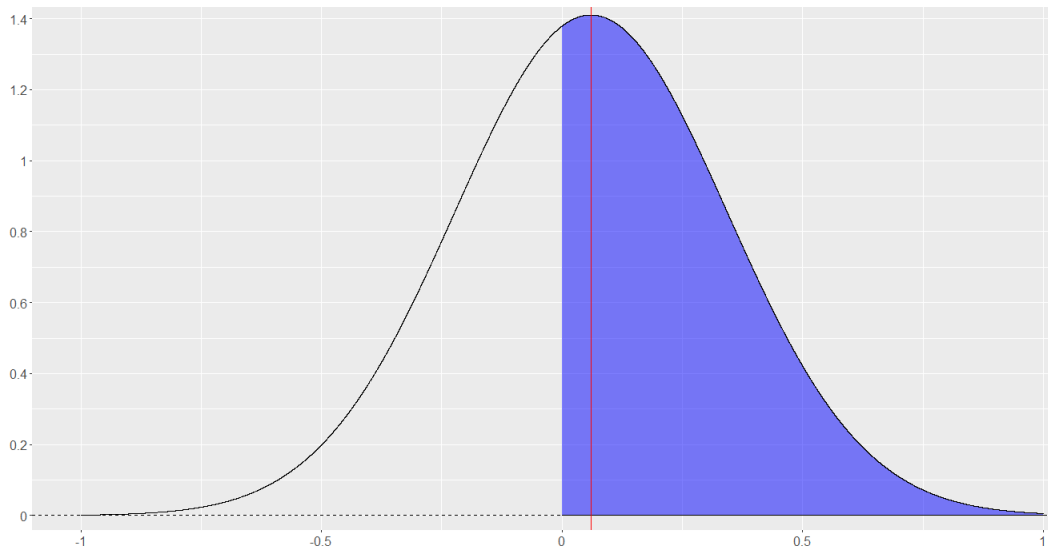


Figura 2 - Distribuições de  $E_i \sim N(a\Delta t, \sigma^2 \Delta t)$  para  $a = 0,12$ ,  $\sigma = 0,4$  e  $\Delta t = 0,1$ .

Salienta-se que, no modelo,  $a = \mu - \frac{1}{2}\sigma^2$ , em que  $\mu$  e  $\sigma$  são a taxa de retorno esperado por unidade de tempo e a volatilidade do preço da ação, respectivamente [18]. Assim, a probabilidade de alta em cada etapa será diretamente influenciada pelos valores desses parâmetros, e somente será igual a 1/2 quando  $a$  for igual a zero e, portanto,  $\mu = \frac{1}{2}\sigma^2$ .

A seguinte representação em tempo discreto, derivada da respectiva equação diferencial estocástica, também pode ser utilizada:

$$\Delta S(t) = \mu S(t)\Delta t + \sigma S(t)Y\sqrt{\Delta t} \quad \text{ou} \quad \Delta S(t)/S(t) = \mu\Delta t + \sigma Y\sqrt{\Delta t} \quad (10)$$

$\Delta S(t)$  discrimina a mudança no preço  $S(t)$  num pequeno intervalo de tempo  $\Delta t$ . A aproximação discreta do retorno da ação nesse intervalo é representada por  $\Delta S(t)/S(t)$ , que possui um valor esperado  $\mu\Delta t$  e um componente estocástico  $\sigma Y\sqrt{\Delta t}$ . A variância do retorno será dada por  $\sigma^2\Delta t$  [18].

Portanto,  $\Delta S(t)/S(t)$  tem uma distribuição aproximadamente normal com média  $\mu\Delta t$  e desvio padrão  $\sigma\sqrt{\Delta t}$ , ou seja [18]:

$$\Delta S(t)/S(t) \sim N(\mu\Delta t, \sigma^2\Delta t) \quad (11)$$

### 3.3 ESTIMATIVA DA VOLATILIDADE $\sigma$

A volatilidade  $\sigma$  de uma ação é uma medida da incerteza sobre os retornos oferecidos por ela. Ela pode ser definida como o desvio padrão do retorno oferecido pela ação em 1 ano quando o retorno é expresso utilizando capitalização contínua [18].

Quando  $\Delta t$  for pequeno, vê-se que  $\sigma^2\Delta t$  é aproximadamente igual à variância do retorno referente ao preço da ação nesse intervalo. Isso significa que  $\sigma\sqrt{\Delta t}$  será aproximadamente o desvio padrão da mudança percentual do preço da ação em  $\Delta t$  [18]. Assim, dado um desvio padrão de  $\sigma$ , referente a todo um período  $T$ , o desvio padrão da mudança percentual em  $\Delta t$  será dado por  $\sigma\sqrt{1/n}$ , em que  $n = T/\Delta t$ .

Para estimar a volatilidade de um preço de ação empiricamente, em geral, observa-se o seu preço em intervalos de tempo fixos (por exemplo, em todos os dias, semanas ou meses do ano) e define-se [18]:

$n + 1$ : Número de observações

$S_i$ : Preço da ação ao final do  $i$ -ésimo intervalo, com  $i = 0, 1, \dots, n$ .

$\tau$ : Intervalo de tempo em anos.

$$u_i = \ln (S_i/S_{i-1}) , \text{ para } i = 1, 2, \dots, n. \quad (12)$$

A estimativa normal  $s$  do desvio padrão de  $u_i$  será dada por [18]:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2}, \text{ em que } \bar{u} \text{ é a média de } u_i. \quad (13)$$

Assim,  $\sigma$  é estimado como  $\hat{\sigma} = s/\sqrt{\tau}$ . Pode ser demonstrado que o erro padrão dessa estimativa é aproximadamente igual a  $\hat{\sigma}/\sqrt{2n}$  [18].

#### 4. OPÇÕES DE COMPRA E VENDA

Derivativos financeiros podem ser definidos como instrumentos cujos valores dependem, ou derivam, de outras variáveis mais básicas [18]. Opções de compra e venda, também conhecidas como *calls* e *puts*, são dois desses derivativos negociados nas bolsas de valores ao redor do mundo.

Esses derivativos dão ao titular o direito, respectivamente, de comprar ou vender um ativo subjacente até uma determinada data por um preço específico. Em ambos os casos, a data limite para o exercício do direito é denominada data de vencimento ( $T$ ), enquanto o preço no qual o direito poderá ser exercido é conhecido como preço de exercício ou *strike* ( $K$ ) [18].

Se o exercício do direito da opção puder ser realizado somente na data de vencimento, elas são denominadas *opções europeias*; se esse direito puder ser exercido a qualquer tempo, até aquela data, elas são chamadas de *opções americanas* [14, 18].

No presente estudo, serão consideradas somente as *opções europeias*. Em geral, elas são mais fáceis de analisar do que as americanas, sendo que algumas das propriedades das opções americanas são deduzidas a partir das propriedades de sua contraparte europeia [18]. Além disso, com essa abordagem, será possível derivar soluções explícitas e fórmulas compactas para o problema de precificação desses derivativos [14].

O detentor da opção não é obrigado a exercê-la. Assim, no caso de uma *call*, se o preço do ativo  $S(t)$  for menor do que  $K$  na data de vencimento, o seu detentor não irá querer exercer esse direito, pois poderá comprar o ativo diretamente no mercado por um preço  $S(T) < K$  [14].

Em se tratando de uma *put*, se na data de vencimento o preço do ativo  $S(t)$  for maior do que  $K$ , o seu detentor também não irá querer exercê-la, pois poderá vender o ativo subjacente diretamente por um preço  $S(T) > K$ .

As seguintes equações mostram os ganhos, também denominados *payoffs*, obtidos pelo detentor de uma opção de compra ou venda na data de vencimento em razão do valor do ativo  $S(T)$  e do *strike*  $K$  [14]:

$$\text{payoff call} = g_{\text{call}}(S(T)) = \max(0, S(T) - K) = \begin{cases} S(T) - K, & \text{se } S(T) > K, \\ 0, & \text{se } S(T) \leq K \end{cases} \quad (14)$$

$$\text{payoff put} = g_{\text{put}}(S(T)) = \max(0, K - S(T)) = \begin{cases} K - S(T), & \text{se } S(T) < K, \\ 0, & \text{se } S(T) \geq K \end{cases} \quad (15)$$

Para um *strike*  $K = 20$ , a Figura 3 ilustra esses respectivos ganhos em função dos possíveis valores do ativo subjacente na data de vencimento  $T$ .

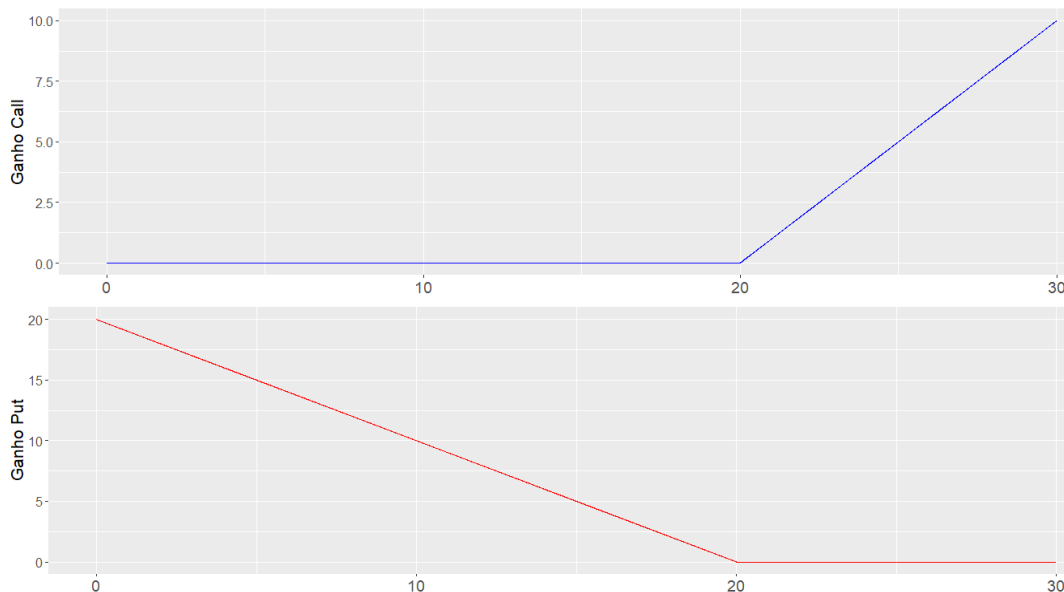


Figura 3 - Payoffs de uma Call e uma Put com Strike igual a 20 em relação ao valor de  $S(T)$ .

Salienta-se que, ao adquirir uma *call* ou *put*, o comprador não sabe de antemão qual será o valor do ativo subjacente na data de vencimento. Dessa forma, é importante estabelecer um preço o qual um comprador estaria disposto a pagar por esses derivativos no tempo  $t = 0$ .

#### 4.1 MODELO DE PRECIFICAÇÃO DE BLACK, SCHOLES E MERTON

Black, Scholes e Merton (1973), com o fim de adequadamente precificar os derivativos financeiros, propuseram a estruturação de um portfólio livre de risco, composto de uma posição no derivativo e de outra posição no ativo subjacente [18]. O retorno desse portfólio seria igual à taxa livre de risco do mercado ( $r$ ).

Quando um portfólio apropriado é estabelecido, os ganhos ou perdas na ação sempre compensam os ganhos ou perdas no derivativo, de forma que o valor total do portfólio, ao final de um breve período, é certamente conhecido [18].

Isso ocorre pelo fato dos preços das ações e derivativos serem afetados pela mesma fonte de incerteza, que é a movimentação do ativo principal. Assim, em um breve período, o preço do derivativo estará perfeitamente correlacionado com o preço da ação [18].

Os seguintes pressupostos foram estabelecidos [18]:

- i. O preço da ação segue um movimento Browniano geométrico com  $\mu$  e  $\sigma$  constantes.
- ii. A venda a descoberto de títulos é permitida.
- iii. Não há custos de transação ou impostos.
- iv. Todos os títulos são perfeitamente divisíveis.
- v. Não há dividendos durante a vida do derivativo.
- vi. A negociação de títulos é contínua.
- vii. A taxa de juros livre de risco,  $r$ , é constante e a mesma para todas as datas de vencimentos.

Eles concluíram que o preço  $f$  de um derivativo utilizado para a montagem desse portfólio seria obtido pela solução da seguinte equação diferencial, que se tornou conhecida como equação diferencial de Black - Scholes - Merton [18]:

$$\frac{\partial f}{\partial t} + rS \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf$$

Essa equação tem muitas soluções, correspondentes a todos os diferentes derivativos que podem ser definidos utilizando  $S$  como a variável subjacente. Entretanto, a solução mais famosa são as denominadas fórmulas de Black - Scholes - Merton para precificação de opções de compra e venda europeias. Elas são definidas, respectivamente, por [18]:

$$c = S_0 N(d1) - Ke^{-rT} N(d2) \quad \text{e} \quad p = Ke^{-rT} N(-d2) - S_0 N(-d1) \quad (16)$$

em que:

$$d1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d1 - \sigma\sqrt{T}$$

A função  $N(x)$  é a distribuição de probabilidade acumulada para uma variável com distribuição normal padrão, ou seja,  $N(x) = P(X \leq x)$ , para  $x \sim N(0,1)$ . As

variáveis  $c$  e  $p$  são os preços respectivos das opções de compra e venda europeias [18].

$S_0$  é o preço da ação no tempo zero,  $K$  é o preço de exercício ou *strike*,  $r$  é a taxa de juros livre de risco com capitalização contínua,  $\sigma$  é a volatilidade do preço da ação e  $T$  é o tempo até o vencimento da opção [18].

#### 4.2 PRECIFICAÇÃO DAS OPÇÕES A PARTIR CURVA DE DISTRIBUIÇÃO DO ATIVO SUBJACENTE

Da seção 3.1, tem-se que  $\ln S(t)$  é normalmente distribuída com média  $(\ln S(0) + at)$  e variância  $\sigma^2 t$ , ou seja:

$$\ln S(t) \sim N(\ln S(0) + at, \sigma^2 t)$$

Portando,  $S(t)$  possui uma distribuição *log-normal*. Para  $a = \mu - \frac{1}{2}\sigma^2$ , a média e variância de  $S(t)$  são encontradas, respectivamente, por  $S(0)e^{\mu t}$  e  $S(0)^2 e^{2\mu t} (e^{\sigma^2 t} - 1)$ .

É interessante definir a probabilidade do *payoff* da opção ser positivo na data de vencimento. Essa probabilidade pode ser entendida como a probabilidade da opção ser exercida (probabilidade de execução), pois, caso o *payoff* seja nulo, o detentor da opção não irá exercê-la, já que poderá negociar o ativo subjacente diretamente no mercado por um valor mais vantajoso.

Em se tratando de uma *call*, a probabilidade do *payoff* ser positivo será  $P(S(T) > K)$ , ou seja, a probabilidade do preço do ativo na data de vencimento ser maior do que o preço de exercício. No caso de uma *put*, será  $P(S(T) < K)$ , a probabilidade do preço do ativo na data de vencimento ser menor do que o preço de exercício.

A partir da curva de distribuição *log-normal* de  $S(T)$ , denominada agora  $f(S(T))$ , é possível encontrar facilmente essas probabilidades por meio das seguintes equações:

$$P(S(T) > K) = \int_K^{\infty} f(s) ds \quad \text{e} \quad P(S(T) < K) = \int_0^K f(s) ds$$



Daqui até o final desta seção, considera-se duas opções, *call* e *put*, relativas a um ativo subjacente com  $S(0) = 10$ ,  $\mu = 0,2$  e  $\sigma = 0,4$  e cujos preços de exercícios são  $K_{call} = 18$  e  $K_{put} = 6$ . A análise sempre será realizada no vencimento  $T = 1$ .

A Figura 4 ilustra as probabilidades de execução de ambas as opções por meio, respectivamente, das áreas em azul e em vermelho sob a curva de  $f(S(T))$ . A probabilidade de execução da *call*,  $P(S(T) > K_{call})$ , foi de 12,11%, enquanto que da *put*,  $P(S(T) < K_{put})$ , foi de 5,74%.

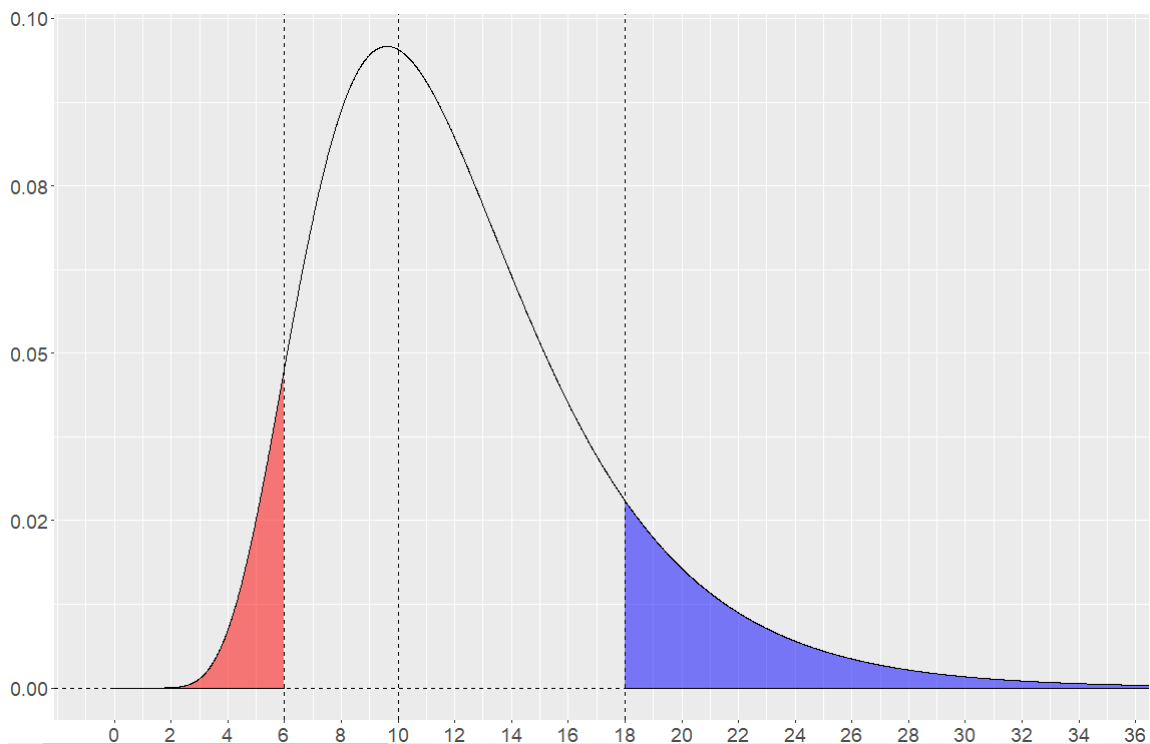


Figura 4 - Probabilidades de execução da call e da put em  $T = 1$ .

Define-se agora os *payoffs ponderados pela probabilidade* das opções. Eles serão calculados pelo produto dos respectivos *payoffs* pela densidade de  $S(t)$  na data de vencimento. Assim, denominando os *payoffs ponderados pela probabilidade* das opções de compra e venda por  $w_{call}$  e  $w_{put}$ , tem-se que:

$$w_{call}(S(T)) = g_{call}(S(T)) \cdot f(S(T)) \quad (17)$$

$$w_{put}(S(T)) = g_{put}(S(T)) \cdot f(S(T)) \quad (18)$$

A Figura 5 mostra, de forma consolidada, a densidade de  $S(t)$ , os *payoffs* e os *payoffs ponderados pela probabilidade* das opções que estão sendo tratadas. Na

curva de densidade, foram preenchidas as áreas representativas das probabilidades de execução de cada uma das opções.

Os *payoffs* e os *payoffs ponderados pela probabilidade* foram indicados pelos gráficos em azul, para a opção de compra, e em vermelho, para a opção de venda.

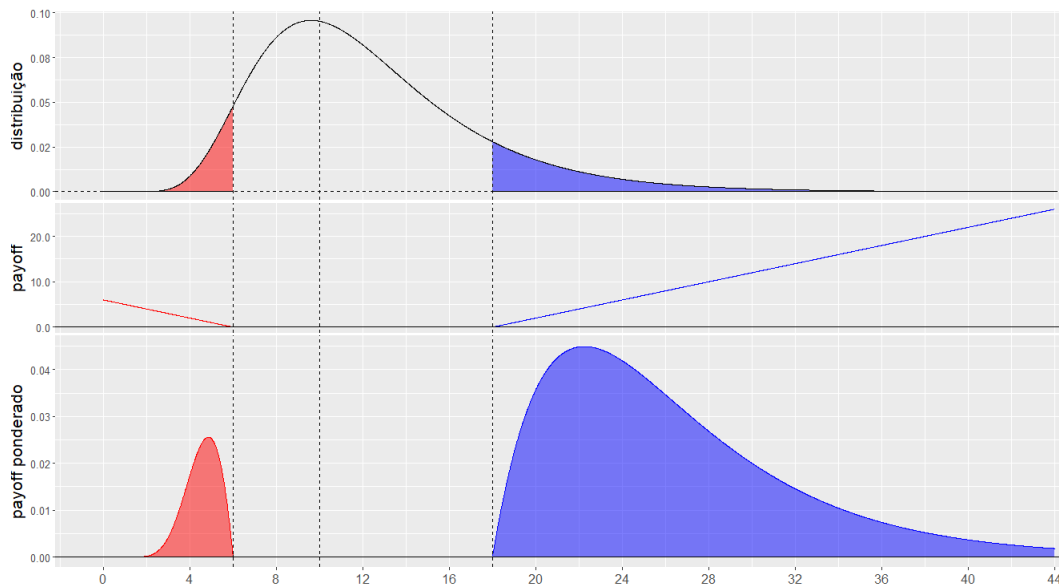


Figura 5 - Densidade de  $S(t)$ , *payoffs* e os *payoffs ponderados pela probabilidade* da *call* e da *put*.

Finalmente, os valores esperados da *call* e da *put* na data de vencimento, denotados por  $V_{call}$  e  $V_{put}$ , serão calculados pelos valores esperados dos respectivos *payoffs*. Eles podem também ser interpretados como as áreas sob os gráficos de  $w_{call}(S(T))$  e  $w_{put}(S(T))$ :

$$V_{call} = \int_{-\infty}^{\infty} g_{call}(s) \cdot f(s) ds = \int_{K_{call}}^{\infty} w_{call}(s) ds$$

$$V_{put} = \int_{-\infty}^{\infty} g_{put}(s) \cdot f(s) ds = \int_0^{K_{put}} w_{put}(s) ds$$

Acontece que esses valores são os esperados na data de vencimento das opções. Assim, para o cálculo dos respectivos preços em  $t = 0$ , é necessário descontá-los pela taxa de juros livre de risco,  $r$ , capitalizada continuamente. Em nosso modelo, essa taxa é equivalente ao retorno esperado,  $\mu$ . Dessa forma, os preços  $V_{call}$  e  $V_{put}$  das opções, em  $t = 0$ , serão dados por:

$$P_{call} = e^{-\mu T} * V_{call} \quad \text{e} \quad P_{put} = e^{-\mu T} * V_{put} \quad (19)$$

Para os dados aqui utilizados, foram encontrados  $V_{call} = 0,5170$ ,  $P_{call} = e^{-0,2} * 0,5170 = 0,4233$ ,  $V_{put} = 0,051$  e  $P_{put} = e^{-0,2} * 0,051 = 0,04175$ . Salienta-se que esses preços são exatamente os mesmos que seriam fornecidos pela aplicação direta das fórmulas de Black - Scholes - Merton, descritas na seção 4.1.

## 5. OPERAÇÕES EM ALTA FREQUÊNCIA (HFT) E EFICIÊNCIA DO MERCADO INTRADIÁRIO

As operações em alta frequência, do inglês *High Frequency Trading* (HFT), são baseadas na redução significativa do tempo necessário para o envio, execução e cancelamento de ordens de compra e venda de ativos financeiros [4]. Nos mercados modernos, esses tipos de operações desempenham um papel importante no fornecimento de liquidez e o seu crescimento deu-se sobretudo em razão da redução dos custos de transação [3].

Estima-se que no mercado de ações americanos, entre 2005 e 2010, por exemplo, o volume movimentado pelos HFTs foi em torno de 50% ou mais do volume total, sendo responsáveis ainda pela queda de 50% na diferença entre os níveis de oferta e demanda dos ativos negociados nesse período [5].

Os estudos empíricos relativos às HFTs possuem duas vertentes. A primeira pode ser caracterizada pela análise dos efeitos dessas operações no mercado. Assim, pesquisadores utilizam conjuntos de dados que explicitamente identificam as HFTs, exploram suas estratégias de operações, testam se essas estratégias são rentáveis e se essas operações melhoram ou não as estimativas dos preços. A segunda vertente analisa mudanças na estrutura de mercado que facilitam as atividades dos HFTs, examinando ainda suas consequências [6].

O presente trabalho encaixa-se nessa segunda vertente de estudos, avaliando a hipótese de eficiência do mercado nos preços intradiários de forma a facilitar o uso dos HFTs caso essa hipótese seja descartada.

### 5.1 EFICIÊNCIA DO MERCADO INTRADIÁRIO

Se a hipótese de eficiência dos preços dos ativos financeiros e do passeio aleatório for verdadeira, a probabilidade de um movimento direcional do preço num instante  $t$  não deve ser relacionada ao movimento num instante  $t-1$ . Assim, a probabilidade deve manter-se mesmo quando mais informações a priori forem adicionadas, tendo em vista que os preços do passado não podem ser utilizados para prever os movimentos futuros [7].

A eficiência do mercado dentro de um mesmo dia implica que ele nunca irá reagir de forma exagerada dentro desse intervalo [7]. No entanto, na prática, é verificado que novas informações são assimiladas pelo mercado de forma ineficiente no curto prazo, com seus preços variando demasiadamente acima ou abaixo do valor intrínseco, antes de se estabilizarem nele [2].

Se o mercado for realmente eficiente, qualquer esforço dedicado para se analisar e buscar estratégias de ganho serão desperdiçados e, apesar da possibilidade de se obter algumas taxas de retornos significativas, esses ganhos serão, na média, proporcionais aos riscos. No entanto, se um simples modelo de operações puder ser utilizado para aumentar a rentabilidade do mercado, isso já seria suficiente para contestar a hipótese de eficiência [7].

Diversos estudos empíricos já foram realizados para verificar a eficiência do mercado para dados intradiários. Para o menor nível possível de dados (*tick a tick*), estudos indicaram uma probabilidade preditiva acima de 70% [8,9].

No entanto, a explicação para essa grande probabilidade preditiva pode ser a variação dos preços entre os níveis de oferta e demanda de determinado ativo (conhecidos como *bid* e *ask*). Esse comportamento é conhecido como dinâmica do mercado [7].

## 5.2 DINÂMICA DO MERCADO

No mercado financeiro, existe uma pequena região de preço, relacionado ao valor subjacente dos ativos, chamada de *spread bid-ask*, Figura 6. *Bid* é o maior preço que um participante do mercado está disposto a pagar por um ativo financeiro, enquanto *ask* é o menor preço que outro participante está disposto a receber por esse mesmo ativo [7].

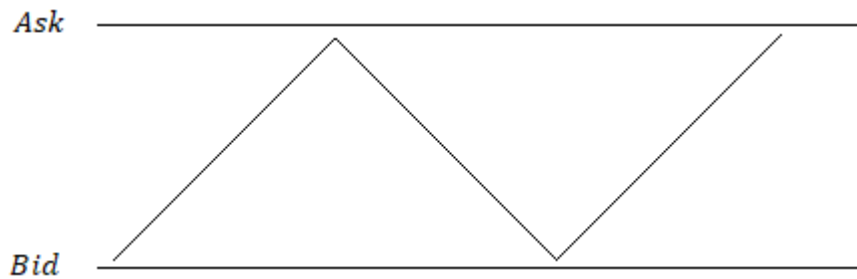


Figura 6 - Flutuação do preço entre os níveis Bid e Ask.

A flutuação dos valores do *bid* e *ask* depende da oferta e demanda. Maior demanda leva os preços para cima, enquanto uma maior oferta leva para baixo. As negociações de mercado somente são realizadas quando o comprador decide pagar o valor do *ask* ou quando o vendedor aceita receber o valor do *bid* [7].

O *spread bid-ask* indica a liquidez de um ativo, de forma que, quanto menor a diferença entre esses preços, maior é a liquidez. Além disso, num mercado eficiente, o *bid* e *ask* flutuam randomicamente [7].

É interessante observar que, em ativos muito negociados, existem milhares de ações disponíveis em cada *bid* e *ask*. No curto prazo, essas ordens funcionam como barreiras, de forma que o mercado somente irá se mover em alguma direção depois que o *bid* ou *ask* for eliminado. [7]

## 6. DESCRIÇÃO DOS DADOS UTILIZADOS E RESULTADOS ENCONTRADOS

Para analisar a eficiência do mercado de ações brasileiro, foram escolhidos 4 (quatro) dos ativos mais negociados na principal bolsa de valores do Brasil, B3 - Brasil Bolsa Balcão S.A., quais sejam ABEV3, BBDC4, PETR4 e VALE3. Devido a grande liquidez desses ativos, eles foram considerados excelentes candidatos para o estudo da eficiência na movimentação dos preços.

Em sítio específico da B3 [20], foram baixados os dados intradiários dos ativos para todos os dias em que houve negociação no ano de 2018 (do dia 2 de janeiro ao dia 28 de dezembro) totalizando 245 dias de negociação. Os preços dos negócios efetuados foram então analisados *tick a tick* e em períodos de 1, 3, 5, 10 e 30 segundos e de 1 e 5 minutos.

Como os dados foram estudados para cada dia de negociação separadamente, a análise para períodos superiores a 5 minutos se tornou prejudicada, pois a quantidade de preços assumidos pelo ativo no decorrer dos períodos iria diminuir consideravelmente.

Por exemplo, para um período de 30 minutos, considerando o horário de funcionamento do pregão das 10h às 17h, o ativo iria assumir somente 14 preços diferentes no decorrer do dia (um para cada período durante o pregão).

Com relação aos preços dos negócios, eles foram considerados como sendo os preços de fechamento de cada período analisado. Um possibilidade também seria adotar os preços médios ponderados pelo volume de cada intervalo [7].

Entretanto, tendo em vista o interesse na aplicabilidade dos resultados da pesquisa em operações reais, foi mais adequado adotar os preços de fechamento, tendo em vista ser muito mais provável que um investidor concretize suas operações de mercado nesses valores do que nos preços médio ponderados, por exemplo.

Em razão do enorme volume de negociação dos ativos estudados, em geral, sempre foi verificado alguma negociação em cada período. Entretanto, quando nenhuma negociação foi verificada, o preço do período foi considerado como sendo o mesmo do período anterior.

Para verificar a eficiência ou não do mercado, foi comparada a probabilidade de um movimento direcional em um instante  $t$  com essa mesma probabilidade, dado os movimentos anteriores do ativo. Sob a hipótese de eficiência, maiores informações

a priori não melhorariam a probabilidade preditiva do mercado. A violação da eficiência permitiria a existência de tendências previsíveis baseada nos dados a priori [7].

As seguintes notações serão utilizadas:  $P(+|-)$ , denotando a probabilidade de um movimento para cima, dado um movimento anterior para baixo;  $P(++)$ , denotando a probabilidade de um movimento para cima, dado um movimento anterior para cima;  $P(+|-,-)$ , denotando a probabilidade de um movimento para cima, dado dois movimentos anteriores para baixo e assim por diante [7].

Foram calculadas as probabilidades condicionais para até 5 (cinco) movimentos anteriores. Além disso, tendo em vista que o preço de fechamento de um dia é em geral diferente do preço de abertura do dia seguinte, as probabilidades condicionais de cada dia foram calculadas separadamente. Foi então utilizado um intervalo normal de 95% de confiança para determinar o número de dias cujos resultados foram estatisticamente significantes.

Se a movimentação dos preços seguir o passeio aleatório, espera-se que a probabilidade de um movimento para cima seja igual à probabilidade de um movimento para baixo, dado as informações a priori. Sob a hipótese da eficiência, espera-se, portanto, a igualdade entre essas probabilidades [7].

De forma a ilustrar a eficiência estatística de movimentos do tipo passeio aleatório, a Figura 7 mostra um total de 10.000 dados gerados de um movimento browniano geométrico com  $\alpha = 0,14$  e  $\sigma = 0,2$  e  $S(0) = 20$ .

Para esses dados, foi estimada uma probabilidade de um movimento para cima,  $P(+)$ , de 50,24% e de um movimento para cima, dado um movimento anterior para baixo,  $P(+|-)$ , de 50,19%. Vê-se, portanto, uma aproximação entre essas medidas, como esperado.



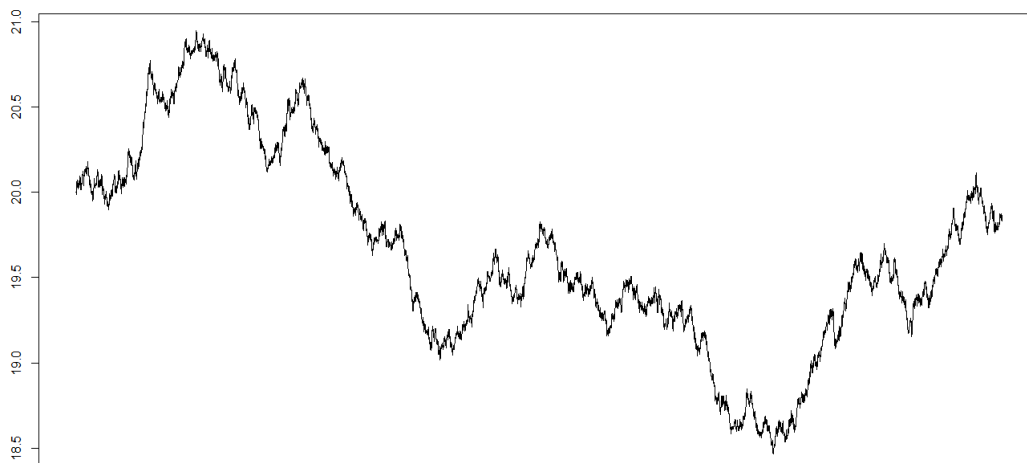


Figura 7 - Simulação computacional de um movimento browniano geométrico.

Por fim, para não sobrecarregar a presente leitura, foi realizada uma análise escrita minuciosa dos resultados obtidos tão somente para o ativo PETR4. Com relação aos demais ativos, os resultados se encontram anexados em apêndices ao final do trabalho.

## 6.1 ANÁLISE DOS DADOS

A eficiência (ou ineficiência) do mercado será avaliada a partir das estimativas das probabilidades condicionais de movimentos para cima e para baixo dos preços, dado os movimentos anteriores, em uma análise inspirada no estudo realizado por Rechenhth [7] para dados do mercado americano do ano de 2005.

Para cada período considerado e em cada dia, foram calculados a probabilidade de um movimento para cima,  $P(+)$ , a probabilidade de um movimento para cima, dado um movimento para baixo,  $P(+|-)$  e a probabilidade de um movimento para cima, dado um movimento para cima,  $P(+|+)$ . As Figuras 8, 9 e 10 mostram os resultados encontrados.

Pode-se observar que  $P(+)$  variou em torno de 50% para todos os períodos. Vê-se também um aumento da variância à medida que os períodos se tornaram cada vez maiores.

Entretanto,  $P(+|-)$  e  $P(+|+)$  apresentaram resultados bastante diferentes em cada período. Com relação a  $P(+|-)$ , os dados *tick a tick* forneceram estimativas das

probabilidades preditivas acima dos 80%. Essas estimativas foram diminuindo gradativamente, aproximando-se cada vez mais dos 50%.

Por outro lado, para  $P(+|+)$ , as estimativas das probabilidades preditivas ficaram abaixo dos 20%, para os dados *tick a tick*. Essas estimativas foram aumentando gradativamente, tornando-se também cada vez mais próximas dos 50%.

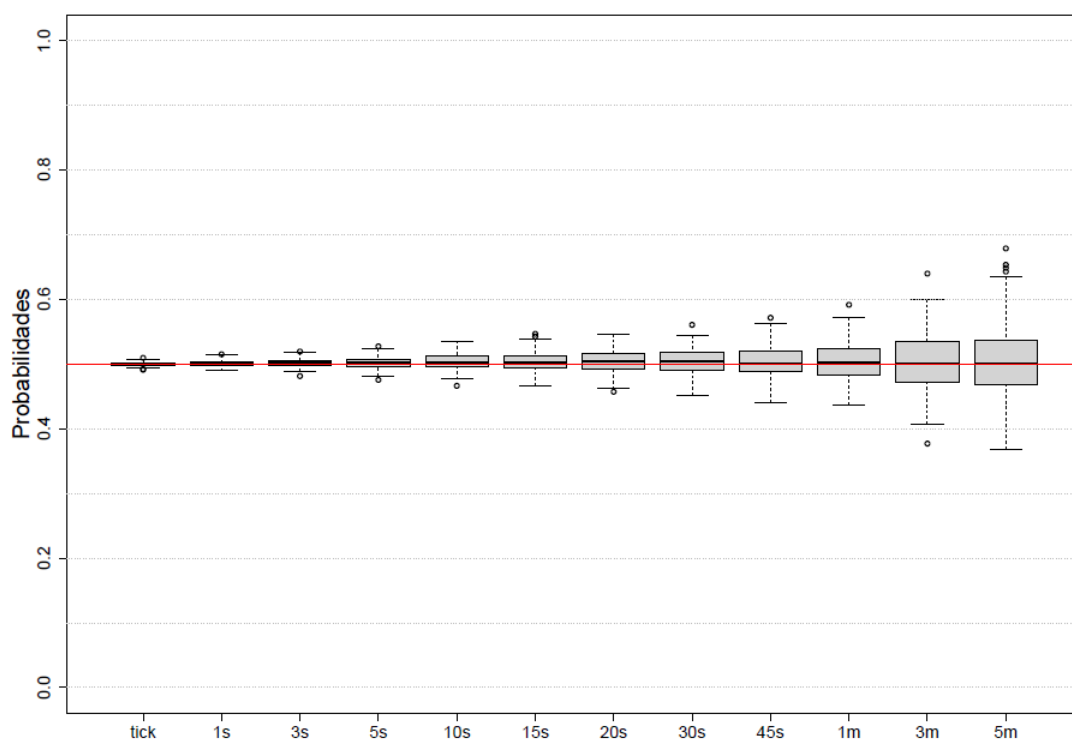


Figura 8 - Boxplot de  $P(+)$  para diferentes períodos.

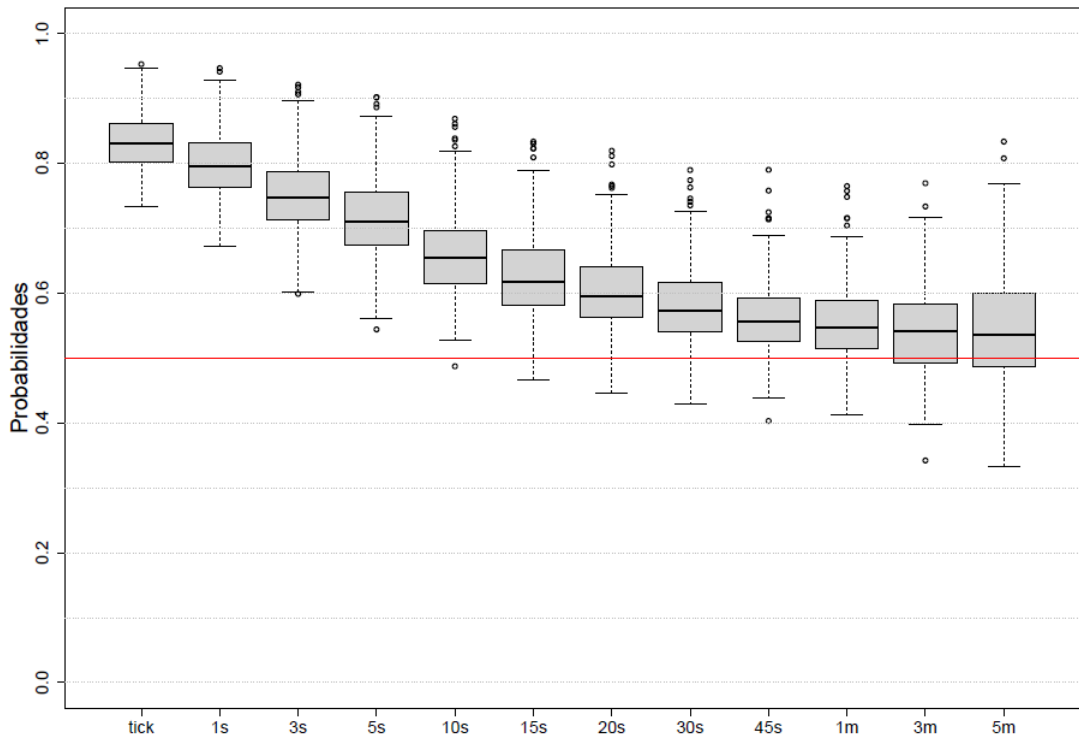


Figura 9 - Boxplot de  $P(+/-)$  para diferentes períodos.

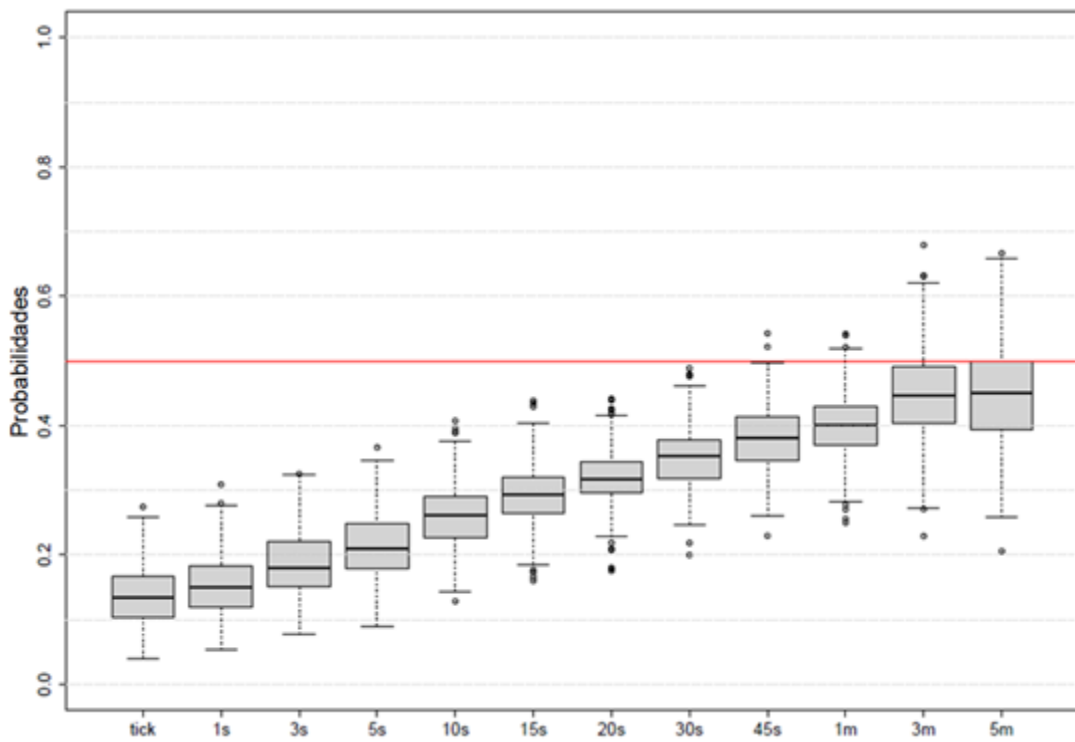


Figura 10 - Boxplot de  $P(+++)$  para diferentes períodos.

Para cada dia de negociação, foram realizados testes estatísticos  $Z$ , definidos a seguir como *Teste 1* e *Teste 2*, para comparação de duas proporções, com um intervalo de 95% de confiança, para testar hipóteses de independência ou não do movimento dos ativos.

Teste 1:

$H_0$ :  $P(+)=P(+|-)$ ; aceitar a independência

$H_1$ :  $P(+)\neq P(+|-)$ ; rejeitar a independência

Teste 2:

$H_0$ :  $P(+)=P(+|+)$ ; aceitar a independência

$H_1$ :  $P(+)\neq P(+|+)$ ; rejeitar a independência

Com relação ao *Teste 1*, o número de dias cuja hipótese de independência foi rejeitada está mostrado na Figura 11. Em todos os períodos, as diferenças significativas sempre ocorreram na banda superior da distribuição, indicando  $P(+|-) > P(+)$ . Nota-se que, para os dados *tick a tick* e de até 5 segundos, todos os 245 dias rejeitaram a hipótese de independência. Para os dados de 15 até 30 segundos, a hipótese nula foi rejeitada para a maioria dos dias.

Vê-se uma diminuição gradual do número de dias que rejeitaram a hipótese de independência à medida que os períodos aumentaram. Para os dados de 45 segundos e 1 minuto, somente 55 e 31 dias, respectivamente, tiveram a hipótese de independência rejeitadas. Para o período de 3 minutos, somente 2 dias rejeitaram a hipótese de independência, enquanto para os dados de 5 minutos, a hipótese nula não foi rejeitada em nenhum dos dias.

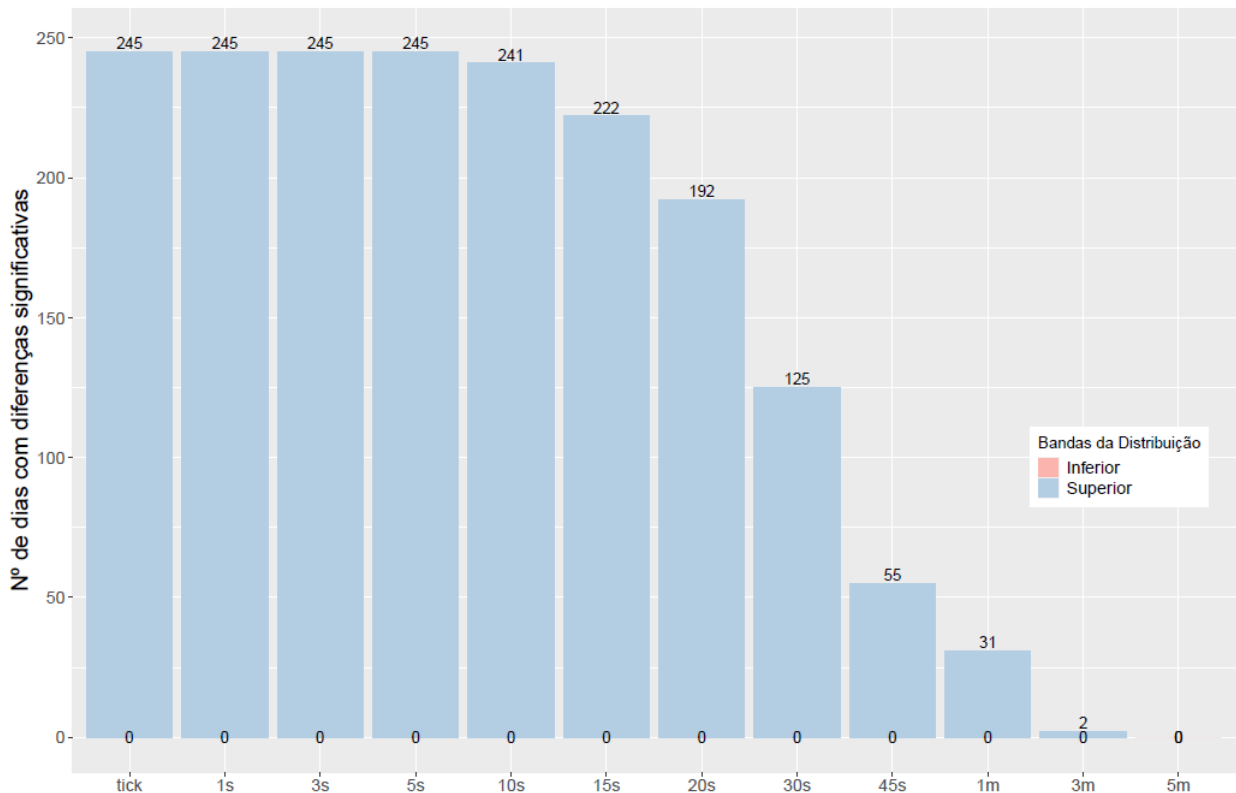


Figura 11 - Número de dias estatisticamente significativos (Teste 1)

Com relação ao *Teste 2*, o número de dias cuja hipótese de independência foi rejeitada está mostrado na Figura 12. Em todos os períodos, diferentemente do que foi verificado anteriormente, as diferenças significativas sempre ocorreram na banda inferior da distribuição, indicando  $P(+|+) < P(+)$ . Para os dados *tick* a *tick* e de até 5 segundos, todos os 245 dias rejeitaram a hipótese de independência. Para os dados de 10, 15, 20 e 30 segundos, a hipótese nula foi rejeitada para 241, 221, 191 e 122 dias, respectivamente.

Vê-se também uma diminuição gradual do número de dias que rejeitaram a hipótese de independência à medida que os períodos aumentaram. Para os dados de 45 segundos e 1 minuto, somente 58 e 29 dias, respectivamente, tiveram a hipótese de independência rejeitadas. Para o período de 3 minutos, somente 1 dia rejeitou a hipótese de independência, enquanto para dos dados de 5 minutos, a hipótese nula não foi rejeitada em nenhum dos dias.

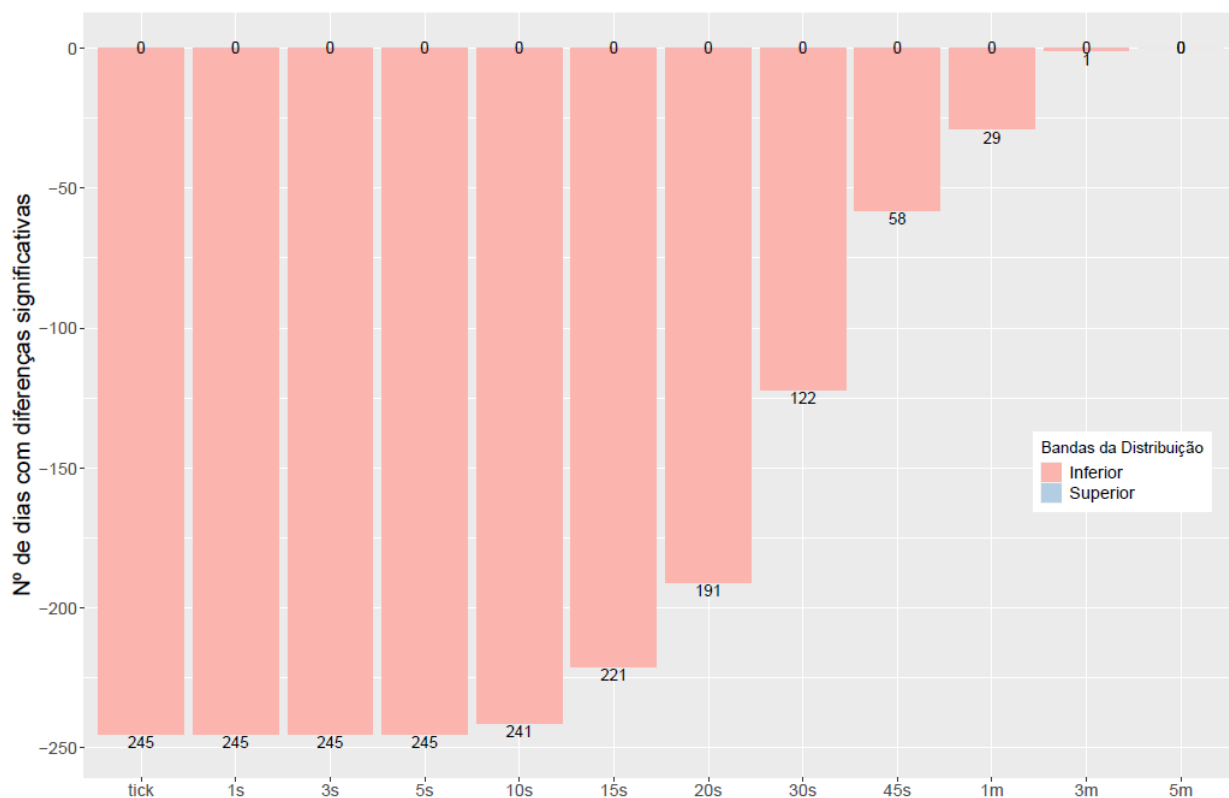


Figura 12 - Número de dias estatisticamente significativos. (Teste 2)

Os testes foram repetidos para avaliar as proporções dos dados agrupados de todos os 245 dias de negociação. A Tabela 1 mostra os resultados encontrados para o Teste 1. Vê-se que a hipótese de independência foi rejeitada em todos os períodos, sempre na banda superior da distribuição, indicando  $P(+)$  ser menor do que  $P(+|-)$ .

Tabela 1 - Resultado do Teste 1 para todos os dias agrupados.

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
tick	713,59	0
1s	419,12	0
3s	263,33	0
5s	188,54	0
10s	104,53	0
15s	68,26	0
20s	48,99	0
30s	30,71	0
45s	20,29	0
1m	14,59	0
3m	6,77	6,47E-12
5m	4,16	1,57E-05

A Tabela 2 mostra os resultados encontrados para o *Teste 2*. Novamente, a hipótese de independência foi rejeitada em todos os períodos, no entanto, sempre na banda inferior da distribuição, indicando  $P(+)$  ser maior do que  $P(+|+)$ .

*Tabela 2 - Resultado do Teste 2 para todos os dias agrupados.*

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
<i>tick</i>	-714,44	0
<i>1s</i>	-417,97	0
<i>3s</i>	-261,88	0
<i>5s</i>	-187,24	0
<i>10s</i>	-103,52	0
<i>15s</i>	-67,45	0
<i>20s</i>	-48,35	0
<i>30s</i>	-30,38	5,74E-203
<i>45s</i>	-20,15	1,34E-90
<i>1m</i>	-14,51	5,20E-48
<i>3m</i>	-6,96	1,75E-12
<i>5m</i>	-4,22	1,25E-05

Na Figura 13, estão sendo mostradas estimativas das probabilidades condicionais dados um e dois movimentos a priori. Nota-se, em todos os períodos, uma maior probabilidade de ocorrer movimentos reversivos, indicando que o mercado tende a caminhar em sentido oposto ao último movimento observado. Vê-se ainda uma aproximação gradual das probabilidades condicionais em relação a  $P(+)$ , à medida que a duração do intervalo considerado aumenta.

Uma observação a ser feita é que as altas probabilidades reversivas são influenciadas de forma muito mais contundente somente pelos últimos movimentos observados. Por exemplo,  $P(+|--)$  e  $P(+|+-)$  foram próximos entre si, de forma que o penúltimo movimento observado, sendo positivo ou negativo, não teve tanta influência na determinação dessas maiores estimativas das probabilidades. Raciocínio semelhante pode ser realizado em relação a  $P(+|+-)$  e  $P(+|++)$ .

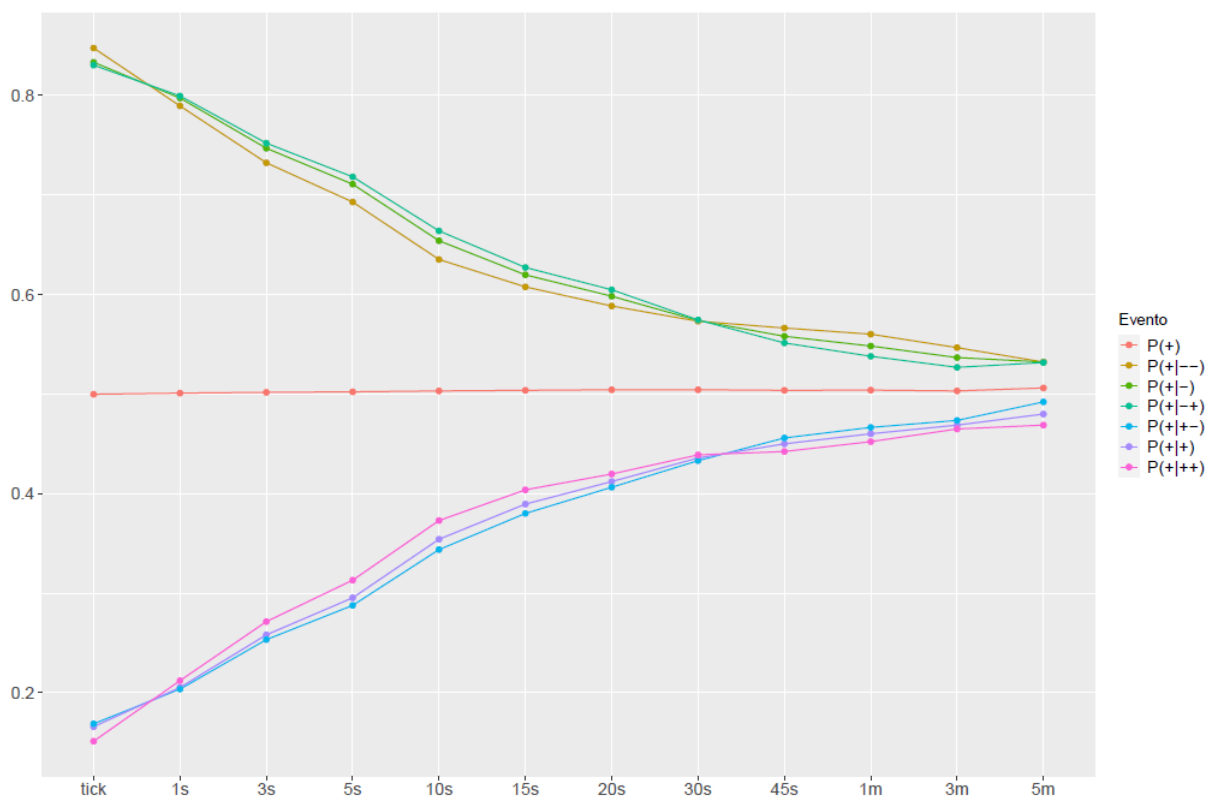


Figura 13 - Probabilidade Condicional para até dois movimentos à priori.

Comportamento semelhante pode ser observado na Tabela 3, que mostra as probabilidades condicionais e respectivos desvios padrões de movimentos contínuos e reversos para dados *tick a tick* e de 5, 10, 20, 30 e 45 segundos.

A probabilidade elevada de reversão, principalmente para os menores intervalos, pode ser explicada pela dinâmica do mercado e está relacionada à variação dos preços entre os níveis de oferta e demanda, *bid* e *ask* [7]. A alta probabilidade reversiva nesses períodos reflete a flutuação dos preços entre esses dois níveis.

Na prática, entretanto, verifica-se que, para os ativos mais líquidos da B3, o nível da oferta ou demanda é quebrado em poucos segundos, de forma que a alta probabilidade preditiva para os intervalos maiores não pode ser explicada tão somente pela variação dos preços entre esses níveis.

Pode-se concluir portanto que o mercado é potencialmente ineficiente no curto prazo e que informações a priori impactam a movimentação dos preços durante certo intervalo de tempo [7]. Em outras palavras, a hipótese do passeio aleatório não é satisfeita e não pode ser utilizada para descrever de forma satisfatória o movimento do ativo, para períodos de até 5 minutos.



Tabela 3 - Estimativas das probabilidades condicionais dos movimentos reversos e dos movimentos de continuação.

Evento	tick		5s		10s		20s		30s		45s	
	Prob.	sd	Prob.	sd	Prob.	sd	Prob.	sd	Prob.	sd	Prob.	sd
P(+ -)	0,83	0,0003	0,71	0,0008	0,65	0,0011	0,60	0,0015	0,57	0,0018	0,56	0,0022
P(+ --)	0,85	0,0007	0,69	0,0016	0,64	0,0019	0,59	0,0024	0,57	0,0028	0,57	0,0033
P(+ ---)	0,80	0,0020	0,66	0,0029	0,61	0,0033	0,58	0,0038	0,57	0,0043	0,57	0,0050
P(+ ----)	0,78	0,0045	0,64	0,0050	0,60	0,0052	0,58	0,0059	0,58	0,0066	0,58	0,0075
P(+ -----)	0,74	0,0101	0,63	0,0085	0,59	0,0083	0,58	0,0091	0,57	0,0102	0,58	0,0117
P(+ +)	0,17	0,0003	0,30	0,0008	0,35	0,0011	0,41	0,0015	0,44	0,0018	0,45	0,0022
P(+ ++)	0,15	0,0007	0,31	0,0016	0,37	0,0019	0,42	0,0024	0,44	0,0028	0,44	0,0032
P(+ +++)	0,20	0,0020	0,35	0,0029	0,40	0,0032	0,44	0,0037	0,45	0,0042	0,45	0,0049
P(+ ++++)	0,23	0,0046	0,38	0,0049	0,41	0,0051	0,45	0,0057	0,45	0,0063	0,44	0,0073
P(+ +++++)	0,26	0,0100	0,39	0,0080	0,43	0,0079	0,45	0,0085	0,45	0,0094	0,45	0,0110

## 7. MODELOS DE PREDIÇÃO PARA A VARIAÇÃO DOS PREÇOS E ESTRATÉGIA DE NEGOCIAÇÃO

Após se verificar que, no curto prazo, os movimentos anteriores dos preços impactam os movimentos seguintes, serão propostos modelos de regressão para estimar a probabilidade de alta/baixa de cada movimento em função dos movimentos passados em cada período considerado.

Utilizou-se o modelo de regressão logístico, em razão da sua adequação na estimação das probabilidades relacionadas a variáveis dependentes categóricas. Para isso, a variação do preço, que é a medida de interesse, foi parametrizada como sendo "0" para movimentos negativos e "1" para movimentos positivos.

As seguintes variáveis explicativas foram criadas para representar os movimentos anteriores: (+), para um movimento anterior para cima; (-), para um movimento anterior para baixo; (-+), para dois movimentos anteriores, sendo o primeiro para cima e o segundo para baixo; (+-), para dois movimentos anteriores, sendo o primeiro para baixo e o segundo para cima; e assim sucessivamente. Cada uma dessas variáveis também assumiu os valores "0", quando não verificada, e "1", quando verificada.

Nota-se ainda que muitas das variáveis explicativas são dependentes entre si e tiveram que ser retiradas dos modelos de previsão. A Tabela 4 mostra a dependência existente entre as variáveis dos modelos para até 3 (três) movimentos anteriores.

Dessa forma, a análise de regressão só pôde ser realizada considerando uma quantidade fixa de movimentos anteriores como variáveis independentes. Por exemplo: somente as variáveis (-) e (+); ou somente as variáveis (--), (-+), (+-) e (++); e assim por diante. Adicionalmente, em cada modelo, o intercepto foi relacionado a todos os movimentos anteriores negativos.

Tabela 4 - Variáveis explicativas dependentes até 3 movimentos anteriores

Dependências entre um e dois movimentos	(-) = (--) + (-+)
	(+) = (+-) + (++)
Dependências entre dois e três movimentos anteriores	(--) = (---) + (--+)
	(-+) = (-+-) + (-++)
	(+-) = (+--)
	(++) = (++) + (+++)
Dependências entre três e quatro movimentos anteriores	(---) = (----) + (---+)
	(--+)
	(-+-)
	(-++)
	(+--)
	(++-)
	(+++)
	(+++)

## 7.1 MODELO DE REGRESSÃO LOGÍSTICA UTILIZADO

Modelou-se, para cada dia de negociação e para cada período, a probabilidade de alta em função dos movimentos anteriores por meio do seguinte modelo logístico:

$$y = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (20)$$

Em que  $y$  é uma regressão linear que representa o *log-odd* da probabilidade  $p$  de um movimento para cima,  $\beta_0$  é o parâmetro do intercepto, relacionado a todos os movimentos anteriores negativos, e  $\beta_1$  a  $\beta_n$  são os coeficientes das variáveis  $x_1$  a  $x_n$ , que representam todos os demais movimentos anteriores.

Adotou-se o nível de 5% de significância para a verificação da confiabilidade de cada um dos coeficientes encontrados. Nota-se que  $x_i$ , com  $i$  variando de 1 a  $n$ , assumirá somente os valores "0" ou "1", quando a variável respectiva for ou não observada. O *odd* da probabilidade  $p$  será dado pela seguinte equação:

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}$$

Assim, a probabilidade de um movimento para cima pode ser encontrada por:

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}$$

Ou ainda

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

É importante salientar que, devido às características dos dados em estudo, nenhuma das variáveis independentes pode ocorrer concomitantemente com alguma outra.

Assim, a regressão linear  $y$  sempre será da forma  $y = \beta_0$ , quando todos os movimentos anteriores forem negativos, ou  $y = \beta_0 + \beta_i$ , caso contrário, em que  $i$  representa o índice das demais variáveis observadas. Por conseguinte, o *odd* sempre será igual a  $e^{\beta_0}$  ou  $e^{\beta_0 + \beta_i}$ , respectivamente.

A ocorrência de cada uma das variáveis  $x_1$  a  $x_n$  será responsável pela multiplicação do *odd* do intercepto ( $e^{\beta_0}$ ) por  $e^{\beta_i}$ . Esse termo é conhecido como *odds-ratio* e seu valor é importante para saber a influência de cada variável sobre a probabilidade  $p$ .

Salienta-se que  $y$  pode assumir qualquer valor na reta dos números reais. Entretanto, independentemente do valor assumido, a probabilidade  $p$  sempre estará discriminada no intervalo 0 e 1, sendo menor do que 1/2, para  $y > 0$  e maior do que 1/2, para  $y < 0$ . A Tabela 5 e a Figura 14 mostram a relação existente entre os valores de  $y$ , *odd* e  $p$ .

Tabela 5 - Relação entre  $y$ , *odd* e  $p$

$y$	$odd = \frac{p}{1-p}$	$p$
$-\infty$	0	0
$< 0$	$< 1$	$< 1/2$
$= 0$	$= 1$	$= 1/2$
$> 0$	$> 1$	$p > 1/2$
$+\infty$	$+\infty$	1

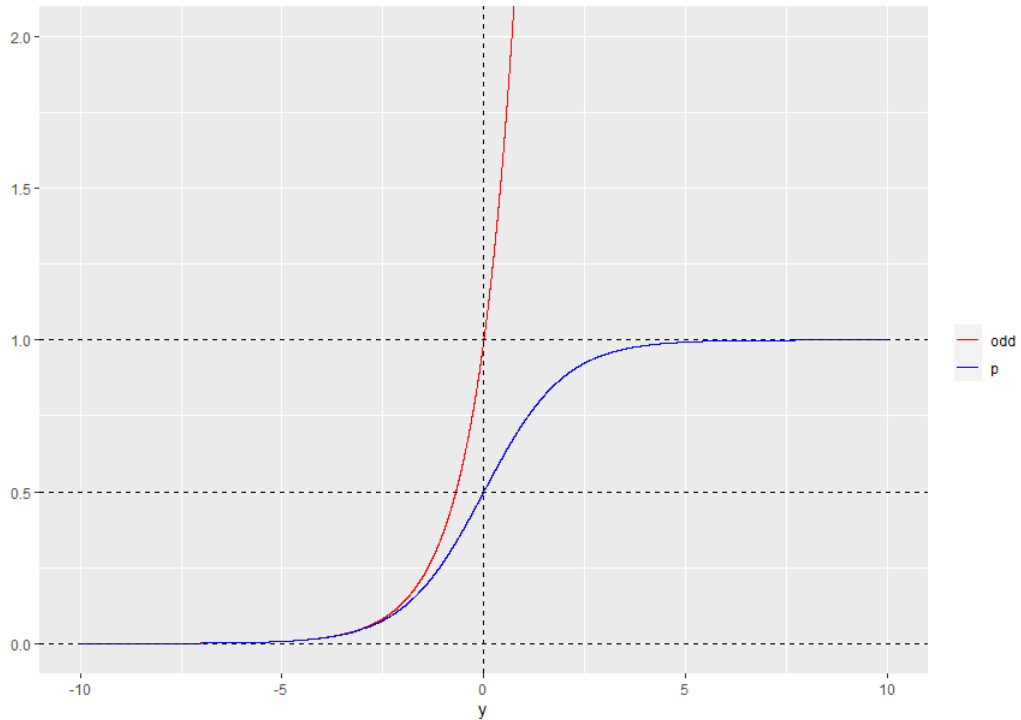


Figura 14 - Relação entre  $y$ ,  $odd$  e  $p$ .

## 7.2 REGRESSÃO LOGÍSTICA PARA 1 (UM) MOVIMENTO ANTERIOR

Para 1 (um) movimento anterior, foi utilizado o seguinte modelo de regressão logística:

$$y = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1$$

Nesse caso,  $y$  representa o *log-odd* da probabilidade  $p$  de um movimento para cima,  $\beta_0$  é o coeficiente relacionado à variável (-) e  $\beta_1$  é o coeficiente da variável (+), representada por  $x_1$ .

O *odd* e a probabilidade  $p$  serão dados pelas seguintes equações:

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x_1}$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1)}}$$

## 7.2.1 Coeficientes e Probabilidades

A aplicação do modelo logístico nos dados em cada dia de negociação gerou coeficientes e  $p$ -valores (significâncias) mostrados na Figura 15. As linhas vermelhas indicam os níveis 0 e 5% nos gráficos dos coeficientes e significâncias, respectivamente.

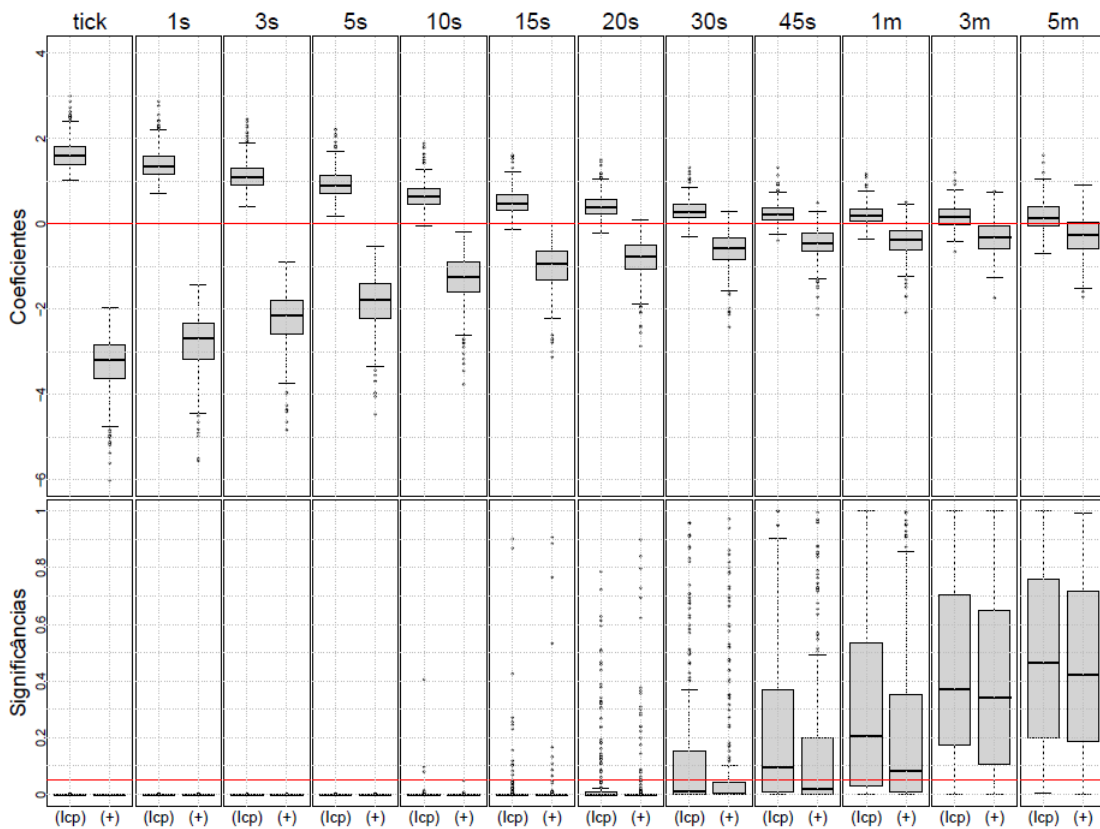


Figura 15 - Coeficientes e significâncias do modelo logístico para 1 (um) movimento anterior.

Na maioria dos modelos, somente os dados *tick a tick* e de 1 a 30 segundos apresentaram os coeficientes do intercepto (-) e da variável (+) significativos ao nível de 5%. Para o período de 45 segundos, somente  $\beta_1$  foi em geral significativo e, para os dados de 1 a 5 minutos, nenhum dos dois coeficientes foram significativos na maior parte dos modelos.

Vê-se ainda que, em todos os casos,  $\beta_0$  apresentou valores positivos e  $\beta_1$  negativos. Assim, quando a variável (-) for verificada, como  $\beta_0$  foi positivo, o  $odd e^{\beta_0}$  será maior do 1 e a probabilidade de alta  $p$  será maior do que 1/2 (Tabela 5). Esse

resultado confirma a tendência de reversão verificada na seção 6. Ou seja, movimentos anteriores negativos tendem a ser seguidos por um movimento positivo.

Por sua vez, quando a variável (+) for observada, o *odd* da variável (-) será multiplicado por  $e^{\beta_1}$ . Assim, o *odd* resultante será igual a  $e^{\beta_0 + \beta_1}$ , um valor menor do que o do intercepto ( $e^{\beta_0}$ ), dado que  $\beta_1$  foi negativo.

Graficamente, nota-se que, em geral, os valores de  $\beta_1$  foram, em módulo, maiores do que  $\beta_0$ . Assim, na ocorrência dessa variável,  $y = \beta_0 + \beta_1$  será negativo. Por conseguinte, conforme a Tabela 5,  $p$  será menor do que 1/2, e a probabilidade de baixa  $1 - p$  será maior do que 1/2, corroborando também o que foi visto na seção 6, que movimentos positivos tendem a ser seguidos por um movimento em sentido contrário.

A Figura 16 mostra a relação entre a probabilidade de alta  $p$  e os movimentos anteriores. A linha em vermelho indica o nível de 50% em cada gráfico.

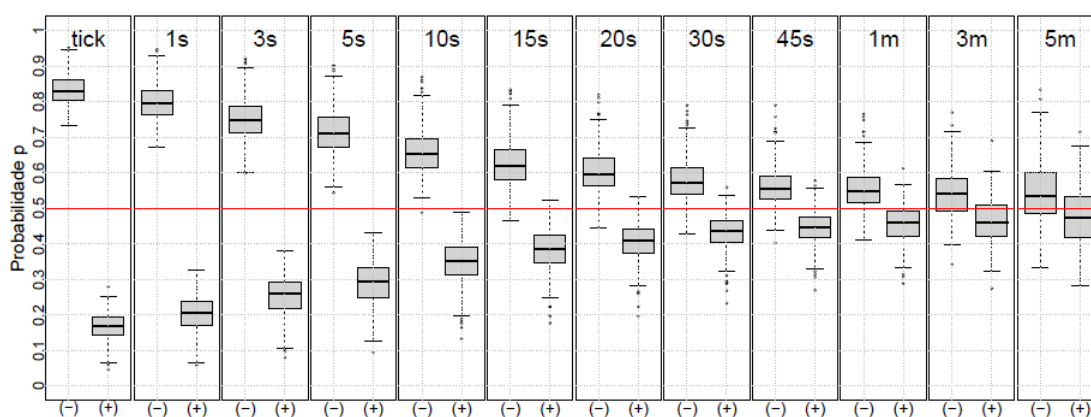


Figura 16 - Probabilidades do modelo logístico para 1 (um) movimento anterior.

Para todos os dados, a probabilidade  $p$  foi, na maioria dos modelos, maior do que 1/2 para a variável (-) e menor que do que 1/2 para a variável (+). Como esperado, os modelos indicaram uma tendência de reversão do último movimento observado.

### 7.3 REGRESSÃO LOGÍSTICA PARA 2 (DOIS) MOVIMENTOS ANTERIORES

Para 2 (dois) movimentos anteriores, foi utilizado o seguinte modelo de regressão logística:

$$y = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

Nesse caso,  $y$  representa o *log-odd* da probabilidade  $p$  de um movimento para cima,  $\beta_0$  é o coeficiente relacionado à variável (--) e  $\beta_1$  a  $\beta_3$  são os coeficientes das variáveis (+), (+-), (++) , representadas por  $x_1$  a  $x_3$ , respectivamente.

O *odd* e a probabilidade  $p$  serão dados pelas seguintes equações:

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)}}$$

### 7.3.1 Coeficientes e Probabilidades

A aplicação do modelo logístico nos dados em cada dia de negociação gerou coeficientes e *p-valores* (significâncias) mostrados nas Figuras 17 e 18. As linhas vermelhas indicam os níveis 0 e 5% nos gráficos dos coeficientes e significâncias, respectivamente.

Para os dados *tick a tick* e para períodos de 1 a 20 segundos, os coeficientes do intercepto (--) e das variáveis (+) e (++) foram, na maioria dos modelos, muito significativos, com valores positivos para o intercepto e negativos para as outras duas variáveis.

Assim, quando a variável (--) for verificada, como  $\beta_0$  foi positivo, o *odd*  $e^{\beta_0}$  será maior do 1 e a probabilidade de alta  $p$  será maior do que 1/2 (Tabela 5). Esse resultado confirma a tendência de reversão verificada na seção 6. Ou seja, movimentos anteriores negativos tendem a ser seguidos por um movimento positivo.

Ainda nos dados *tick a tick* e para períodos de 1 a 20 segundos, quando as variáveis (+) ou (++) forem observadas, o *odd* da variável (--) será multiplicado por  $e^{\beta_i}$ , com  $i$  igual a 2 ou 3. Assim, o *odd* resultante será igual a  $e^{\beta_0 + \beta_i}$ , um valor menor do que o do intercepto ( $e^{\beta_0}$ ), dado que  $\beta_i$  foi negativo.



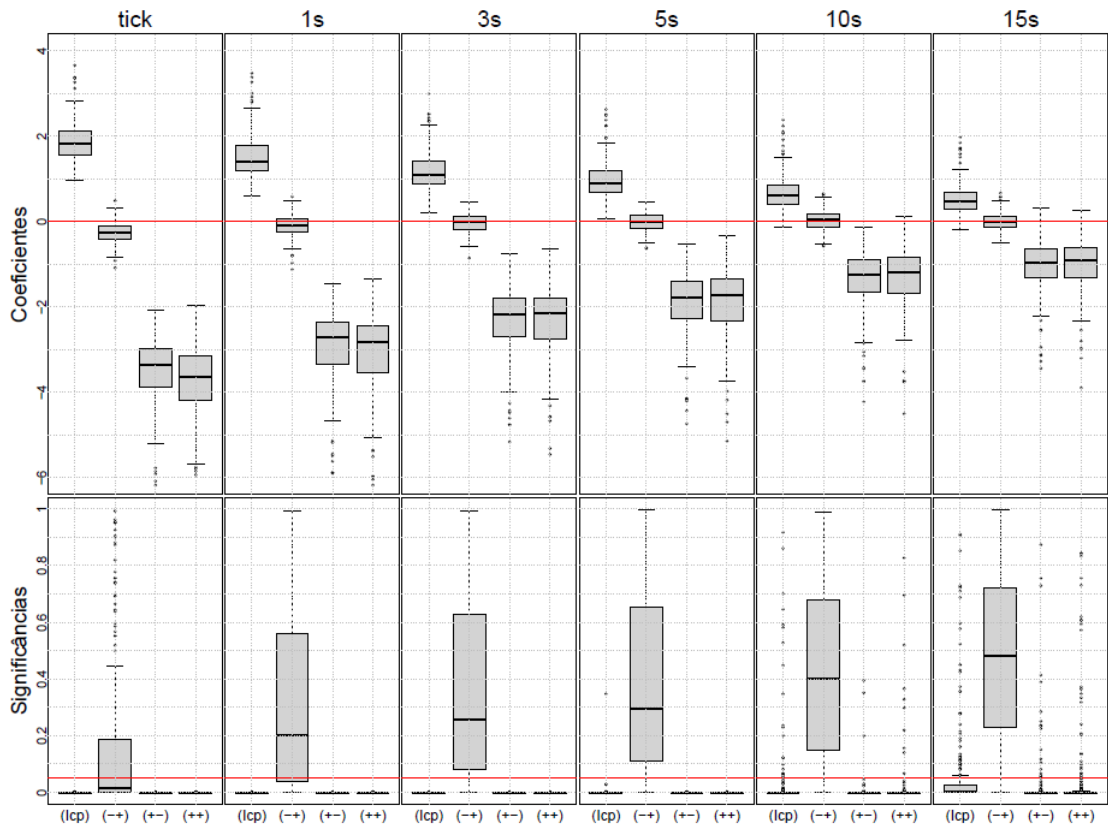


Figura 17 - Coeficientes e significâncias do modelo logístico para 2 (dois) movimentos anteriores. Dados até 15 segundos

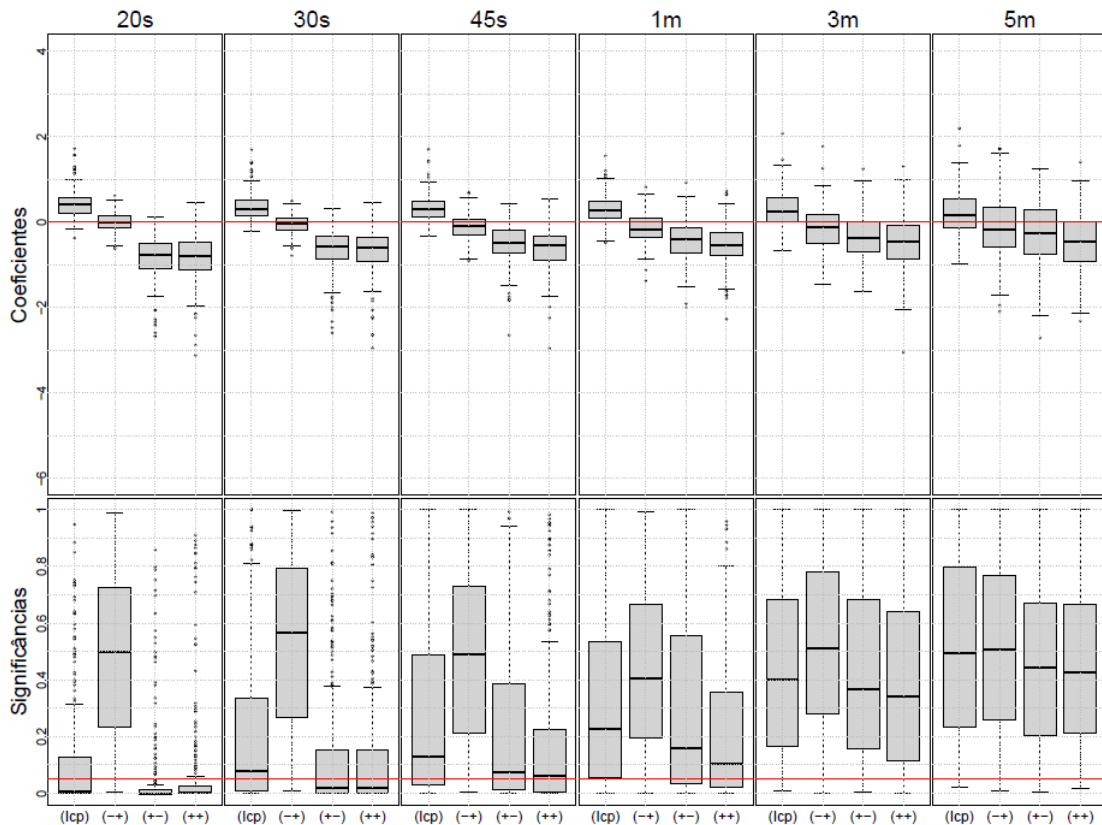


Figura 18 - Coeficientes e significâncias do modelo logístico para 2 (dois) movimentos anteriores. Dados de 20 segundos a 5 minutos

Graficamente, em geral,  $\beta_2$  e  $\beta_3$  foram, em módulo, maiores do que  $\beta_0$ . Assim, na ocorrência de qualquer dessas variáveis,  $y = \beta_0 + \beta_i$  será negativo. Por conseguinte, a probabilidade  $p$  será menor do que  $1/2$ , corroborando também o que foi visto na seção 6, que movimentos positivos tendem a ser seguidos por um movimento em sentido contrário.

Com relação aos coeficientes da variável (-+), ele só foi significativo, na maioria dos modelos, para os dados *tick a tick*. No entanto, para os dados até 45 segundos, sua mediana, em módulo, sempre foi menor que a mediana de  $\beta_0$ , de forma que  $y = \beta_0 + \beta_1 > 0$ . Portanto, nesses dados, a probabilidade  $p$  será maior do que  $1/2$ , prevalecendo a tendência de reversão do movimento anterior.

Para os dados de 30 segundos, o coeficiente do intercepto (--) não foi, em geral, significativo. No entanto, a mediana de  $\beta_0$  foi levemente positiva, indicando uma probabilidade de alta maior do que  $1/2$  para dois movimentos anteriores negativos. A mediana de  $\beta_1$ , por sua vez, foi próxima de zero, de forma que  $y = \beta_0 + \beta_1$  também

será positivo e  $p$  permanecerá maior do que  $1/2$  para a variável (-+). Nos dois casos, a tendência foi de reversão do movimento anterior.

Ainda nesse período, os coeficientes das variáveis (+-) e (++) foram, em geral, significativos e menores do que zero. No entanto, em módulo, eles foram maiores do que  $\beta_0$ , de forma que  $y = \beta_0 + \beta_i$ ,  $i$  igual a 2 ou 3, será negativo e a probabilidade de alta  $p$  será menor do que  $1/2$ , confirmando a tendência de reversão do movimento anterior.

Para os dados de 45 segundos até 5 minutos, os coeficientes não foram significativos na maioria dos modelos, no entanto, em geral eles confirmaram a tendência reversiva. Em todos os casos,  $\beta_0$  foi positivo e  $\beta_1$ ,  $\beta_2$  e  $\beta_3$  foram negativos.

Nota-se que  $\beta_2$  e  $\beta_3$  foram sempre maiores, em módulo, do que  $\beta_0$ , acarretando  $y = \beta_0 + \beta_i$  positivo e  $p$  maior do que  $1/2$  para as variáveis (+-) e (++) . Com relação a  $\beta_1$ , geralmente, o seu módulo foi menor do que  $\beta_0$ , de forma que  $y = \beta_0 + \beta_i > 0$  e  $p > 1/2$  para o movimento (-+), confirmando a tendência reversiva.

A Figura 19 mostra a relação entre a probabilidade de alta  $p$  e os movimentos anteriores. A linha em vermelho indica o nível de 50% em cada gráfico.

Em todos os casos, a probabilidade  $p$  foi, na maioria dos modelos, maior do que  $1/2$  para as variáveis (--) e (-+), e menor que do que  $1/2$  para as variáveis (+-) e (++) , indicando uma tendência reversiva do último movimento observado. A única exceção foi com relação à variável (+-) para os dados de 5 minutos, cuja mediana da probabilidade foi próxima a  $1/2$ .

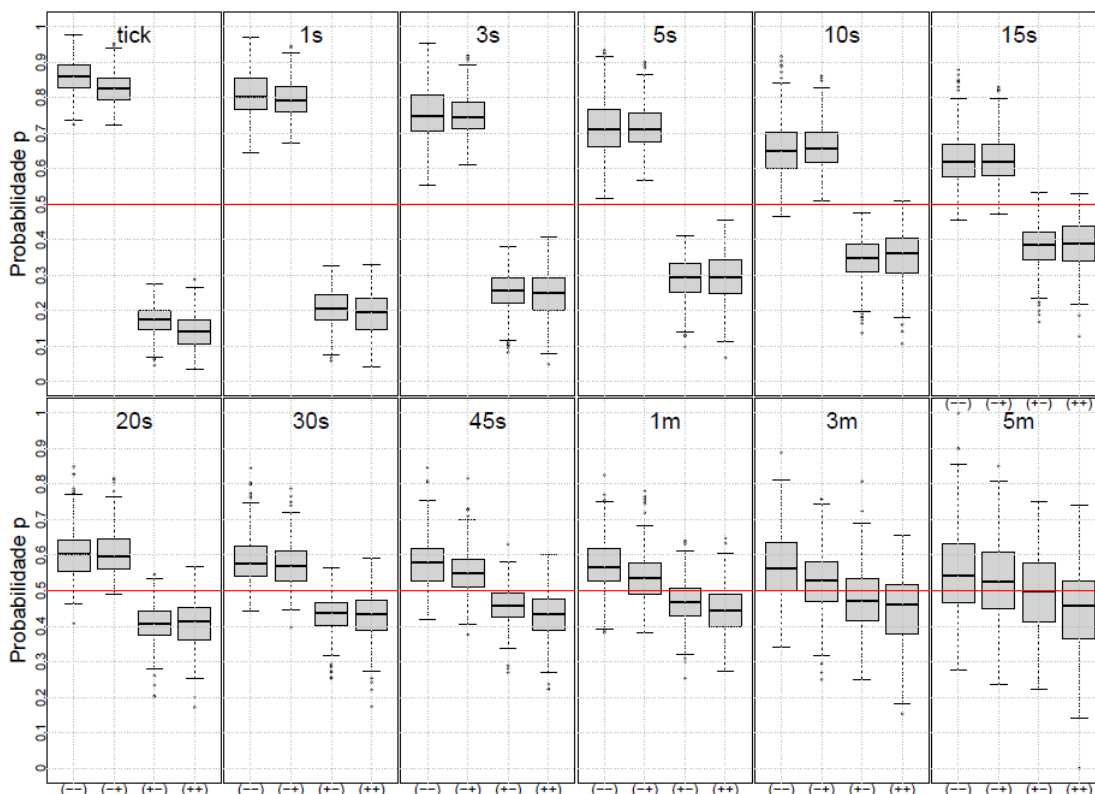


Figura 19 - Probabilidades do modelo logístico para 2 (dois) movimentos anteriores.

## 7.4 REGRESSÃO LOGÍSTICA PARA 3 (TRÊS) MOVIMENTOS ANTERIORES

Para 3 (três) movimentos anteriores, foi utilizado o seguinte modelo de regressão logística:

$$y = \log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$$

Nesse caso,  $y$  representa o  $\log$ -*odd* da probabilidade  $p$  de um movimento para cima,  $\beta_0$  é o coeficiente relacionado à variável ( $---$ ) e  $\beta_1$  a  $\beta_7$  são os coeficientes das variáveis ( $--+$ ), ( $-+-$ ), ( $-++$ ), ( $+--$ ), ( $+++$ ), ( $++-$ ) e ( $+++$ ), representadas por  $x_1$  a  $x_7$ , respectivamente.

O *odd* e a probabilidade  $p$  serão dados pelas seguintes equações:

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7}$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7)}}$$

### 7.4.1 Coeficientes e Probabilidades

A aplicação do modelo logístico nos dados em cada dia de negociação gerou coeficientes e *p-valores* (significâncias) mostrados nas Figuras 20 a 22. As linhas vermelhas indicam os níveis 0 e 5% nos gráficos dos coeficientes e significâncias, respectivamente.

Para os dados *tick a tick* e para períodos de 1 a 10 segundos, os coeficientes do intercepto (---) e das variáveis (---), (---), (---) e (---) foram, na maioria dos modelos, muito significativos, com valores positivos para o intercepto e negativos para as outras quatro variáveis.

Assim, quando a variável (---) for verificada, como  $\beta_0$  foi positivo, o *odd*  $e^{\beta_0}$  será maior que 1 e a probabilidade de alta  $p$  será maior do que 1/2 (Tabela 5). Esse resultado confirma a tendência de reversão verificada na Seção 6. Ou seja, movimentos anteriores negativos tendem a ser seguidos por um movimento positivo.

Ainda nesses dados, quando as variáveis (---), (---), (---) ou (---) forem observadas, o *odd* da variável (---) será multiplicado por  $e^{\beta_i}$ , com  $i$  variando de 4 a 7. Assim, o *odd* resultante será igual a  $e^{\beta_0 + \beta_i}$ , um valor menor do que o do intercepto ( $e^{\beta_0}$ ), dado que  $\beta_i$  foi negativo.

Graficamente, nota-se que, em geral,  $\beta_4$ ,  $\beta_5$ ,  $\beta_6$  e  $\beta_7$  foram, em módulo, maiores do que  $\beta_0$ . Assim, na ocorrência de qualquer dessas variáveis,  $y = \beta_0 + \beta_i$  será negativo. Por conseguinte, conforme a Tabela 5, a probabilidade  $p$  será menor do que 1/2, corroborando a tendência de movimentos positivos serem seguidos por um movimento em sentido contrário.

Com relação aos coeficientes das variáveis (---), (---) e (---), mesmo não sendo, na maioria dos casos, significativos ao nível de 5%, aplicando o mesmo raciocínio para a mediana de cada um deles (mostrada graficamente), pode-se notar que a probabilidade  $p$  será maior do que 1/2, prevalecendo a tendência de reversão do movimento anterior.

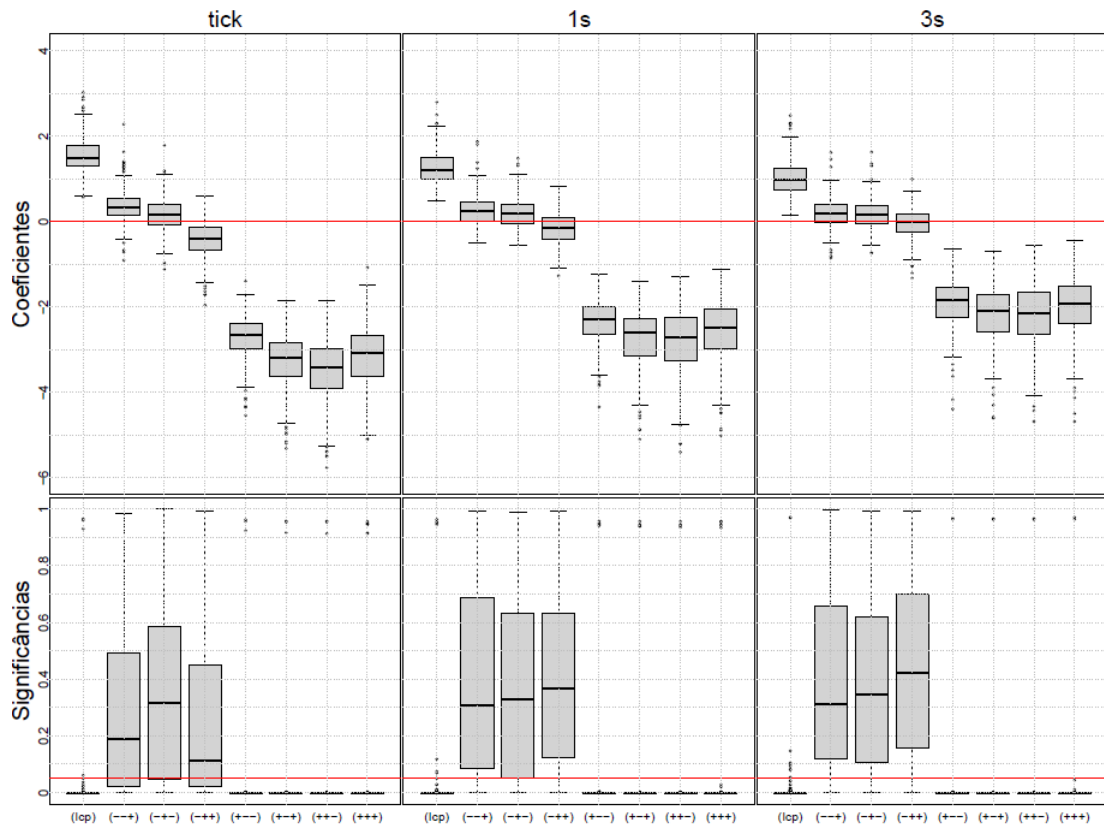


Figura 20 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados tick a tick a 3 segundos

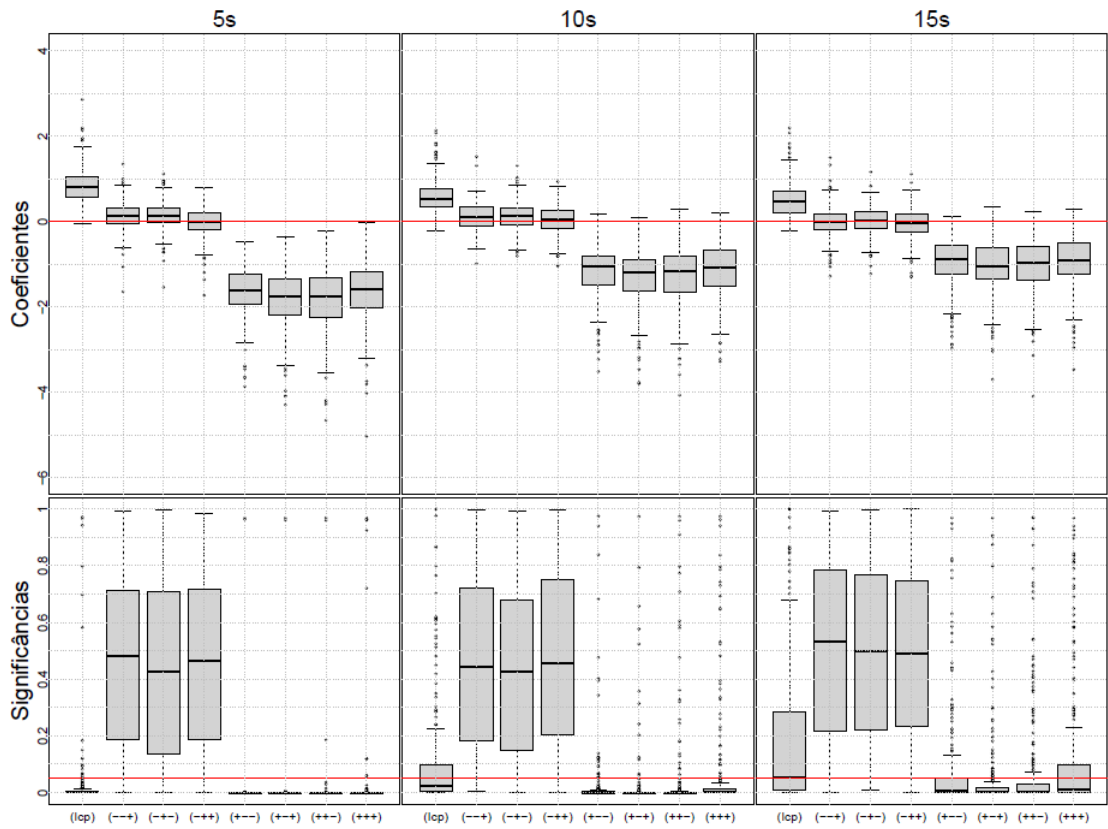


Figura 21 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados 5 a 15 segundos

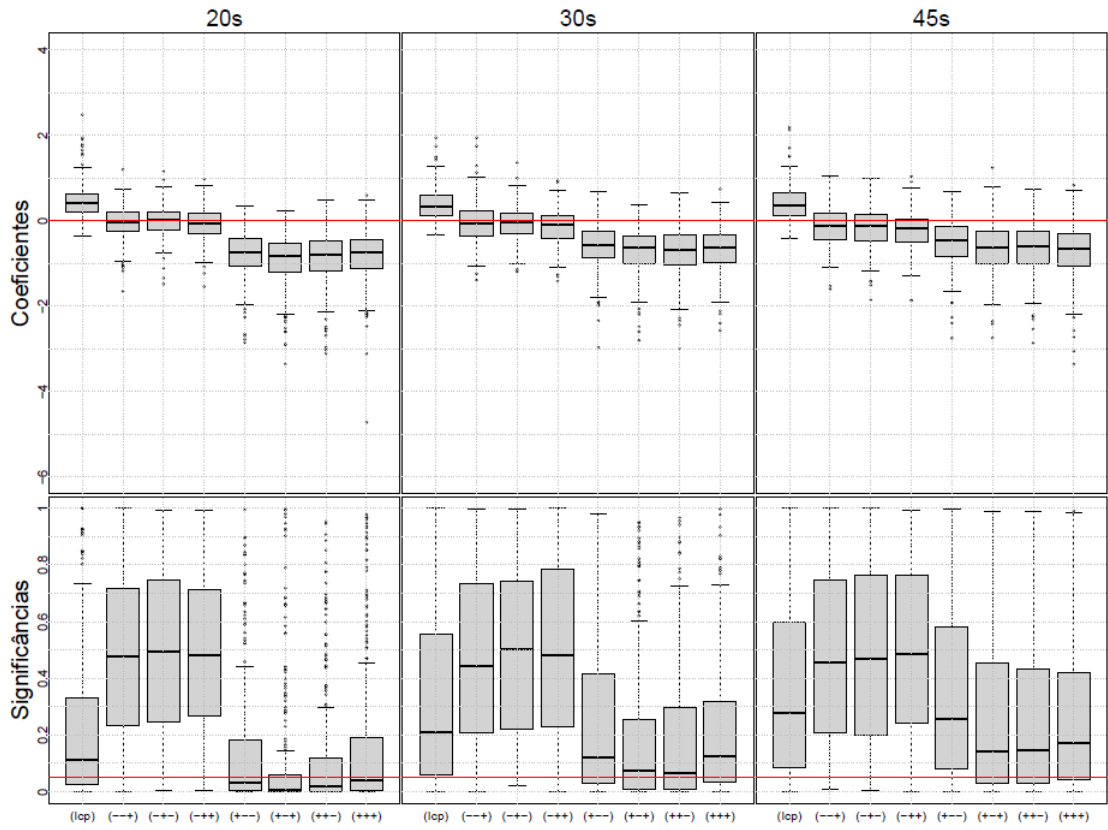


Figura 22 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados 20 a 45 segundos



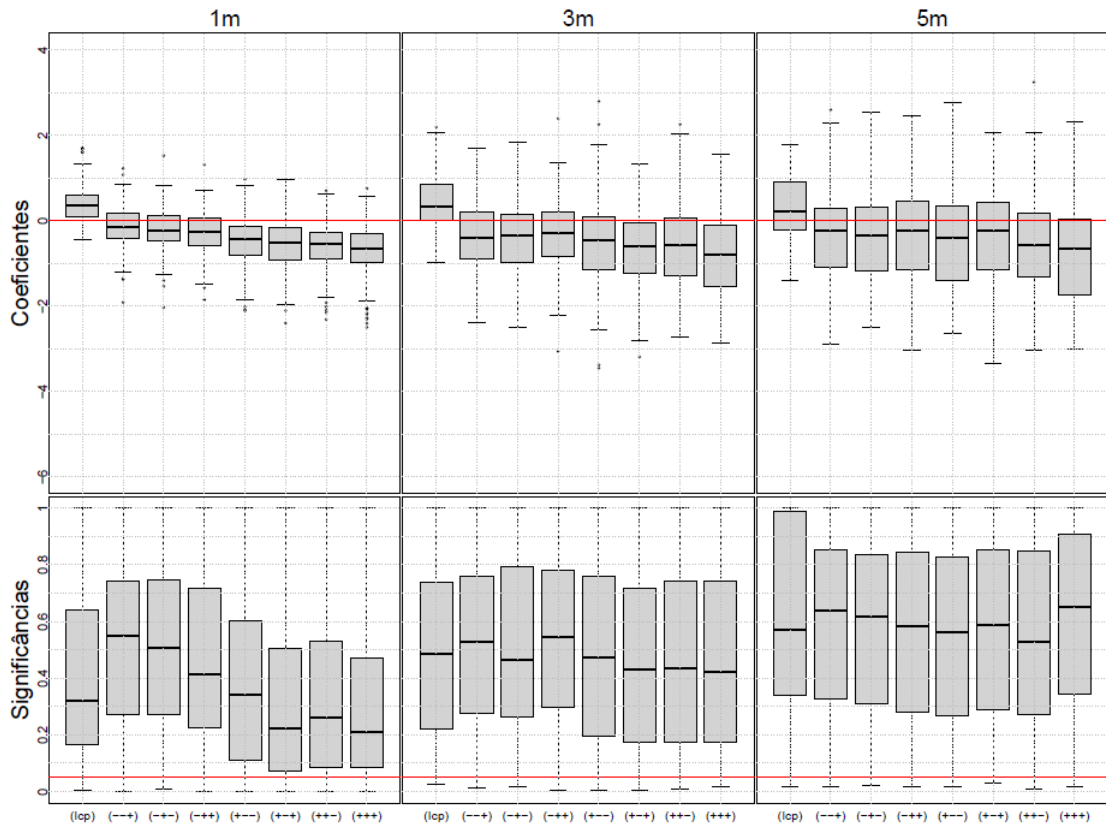


Figura 23 - Coeficientes e significâncias do modelo logístico para 3 (três) movimentos anteriores. Dados 1 a 5 minutos

Para os dados de 15 e 20 segundos, somente os coeficientes das variáveis (+-), (-++), (++-) e (+++) foram, em geral, significativos, enquanto que, nos demais dados, nenhuma das variáveis apresentaram coeficientes significativos na maioria dos modelos. Em todos eles, entretanto, a análise da mediana nos mostra uma tendência de reversão do movimento anterior.

Por exemplo, para os dados de 45 segundos, a mediana do coeficiente do intercepto, variável (--+), foi positiva e das demais variáveis foram negativas. Com relação às variáveis (--+), (-+-) e (-+-), os módulos de seus coeficientes foram em geral menores do que  $\beta_0$ , assim  $y = \beta_0 + \beta_i$  será positivo e  $p$  será maior do que 1/2 (Tabela 5), confirmando a tendência reversiva.

Para as variáveis (+--), (-++), (++-) e (+++), por sua vez, os módulos das medianas dos coeficientes foram maiores do que o coeficiente  $\beta_0$ . Assim,  $y = \beta_0 + \beta_i$  será negativo e  $p$  será menor do que 1/2, confirmando também essa tendência.

As Figuras 24 e 25 mostram a relação entre a probabilidade de alta  $p$  e os movimentos anteriores. A linha em vermelho indica o nível de 50% em cada gráfico.

Em todos os dados, a probabilidade  $p$  foi, na maioria dos modelos, maior do que  $1/2$  para as variáveis  $(---)$ ,  $(--\rightarrow)$ ,  $(-\rightarrow-)$  e  $(-\rightarrow\rightarrow)$ , e menor que do que  $1/2$  para as variáveis  $(\rightarrow--)$ ,  $(\rightarrow\rightarrow-)$ ,  $(\rightarrow\rightarrow\rightarrow)$  e  $(\rightarrow\rightarrow\rightarrow)$ . Como esperado, os modelos indicaram uma tendência de reversão do último movimento.

Pode-se notar ainda que alguns modelos forneceram probabilidades próximas de  $1/2$  para algumas variáveis. Isso foi verificado com relação ao movimento  $(-\rightarrow-)$  e  $(\rightarrow\rightarrow-)$ , para os dados de 5 minutos.

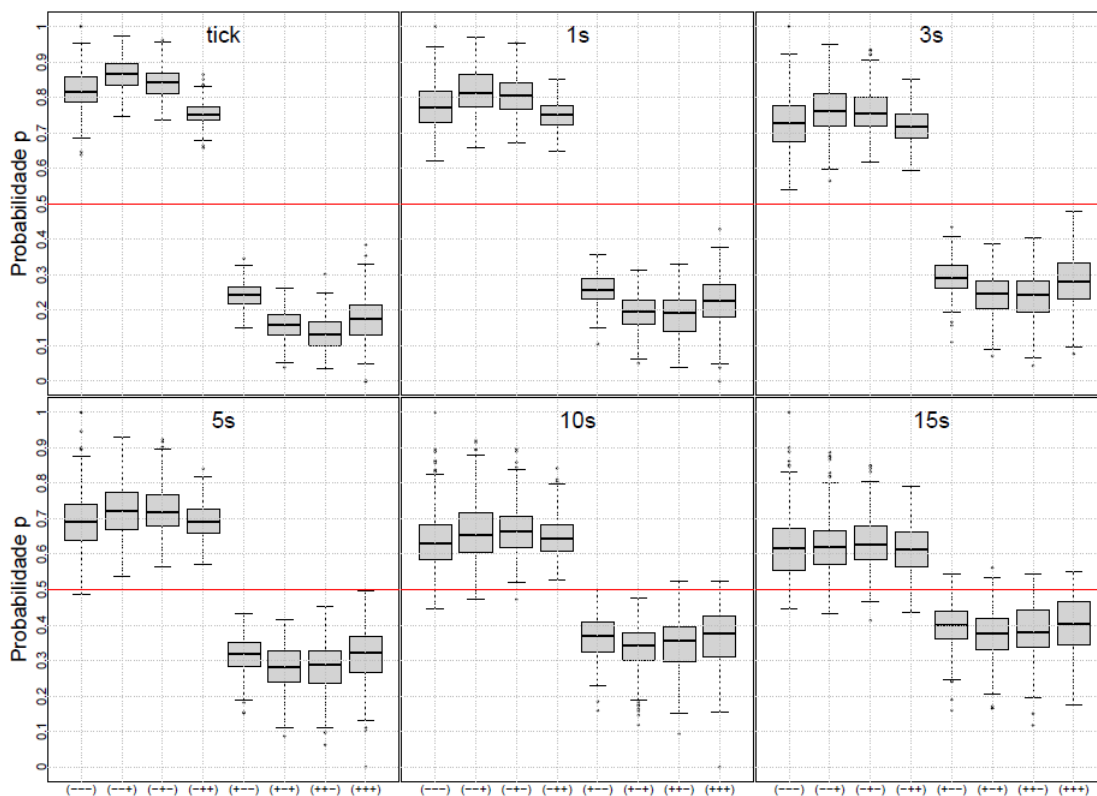


Figura 24 - Probabilidades do modelo logístico para 3 (três) movimentos anteriores. Dados tick a tick até 15 segundos

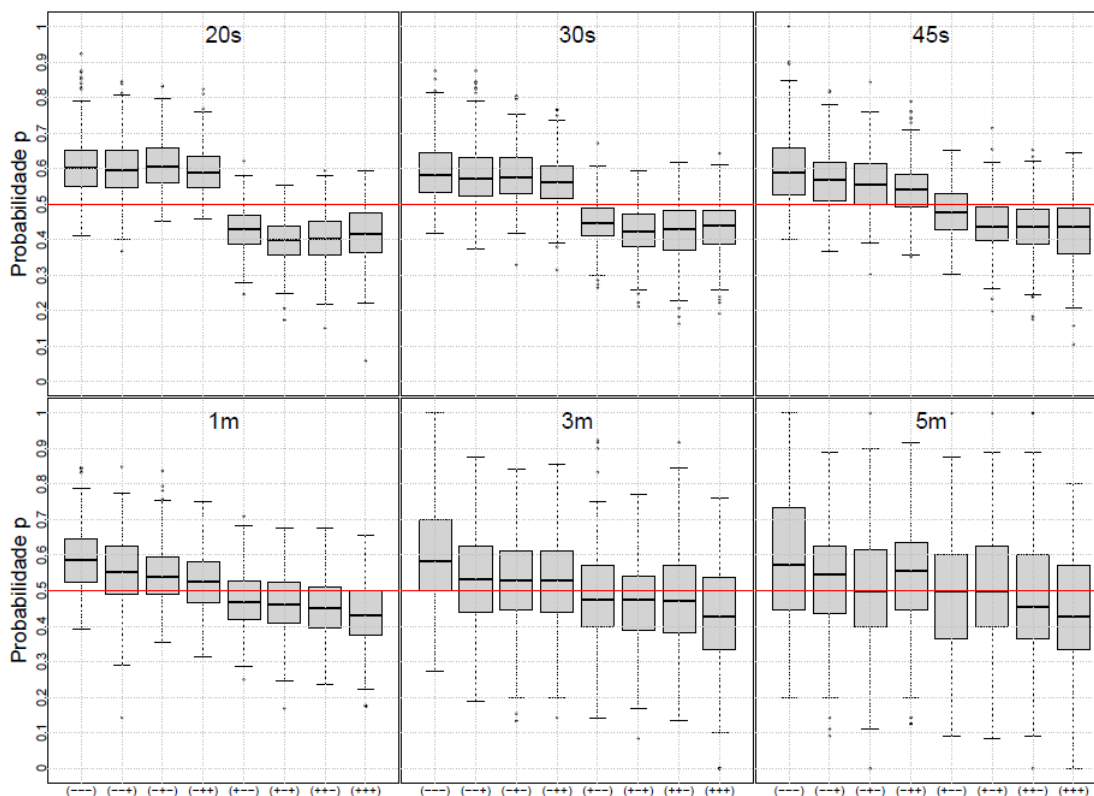


Figura 25- Probabilidades do modelo logístico para 3 (três) movimentos anteriores. Dados de 20 segundos até 5 minutos

## 7.5 CAPACIDADE PREDITIVA DOS MODELOS

Para a análise da capacidade preditiva dos modelos, cada um deles foi utilizado para se prever os movimentos do ativo no dia seguinte. Assim, os coeficientes do primeiro dia de negociação foram utilizados para prever a movimentação do segundo dia; os coeficientes do segundo dia foram utilizados para prever a movimentação do terceiro; e assim sucessivamente.

Apenas os coeficientes do último dia de negociação não foram utilizados para fazer algum tipo de previsão. Portanto, foram calculados os percentuais de acerto somente dos primeiros 244 modelos logísticos.

Em cada dia, o limiar da probabilidade  $p$  utilizado para inferir o sentido da próxima movimentação foi igual a  $1/2$ . Dessa forma, caso o valor de  $p$ , calculado a partir das variáveis representativas dos últimos movimentos, for menor ou igual a  $1/2$ , foi considerado que o próximo movimento seria negativo, caso contrário, foi considerado que seria positivo. A Tabela 6 ilustra essa relação.

Tabela 6 - Limiar Preditivo da probabilidade  $p$  e valor predito

Valor de $p$	Estimativa do Próximo Movimento	Valor Predito
$\leq 1/2$	Negativo	0
$> 1/2$	Positivo	1

Os valores preditos de cada movimento foram então comparados com os valores efetivamente observados, sendo enfim calculadas as probabilidades de acertos para cada dia. A Figura 26 mostra os resultados encontrados. A linha vermelha indica o nível de 50% de acerto.

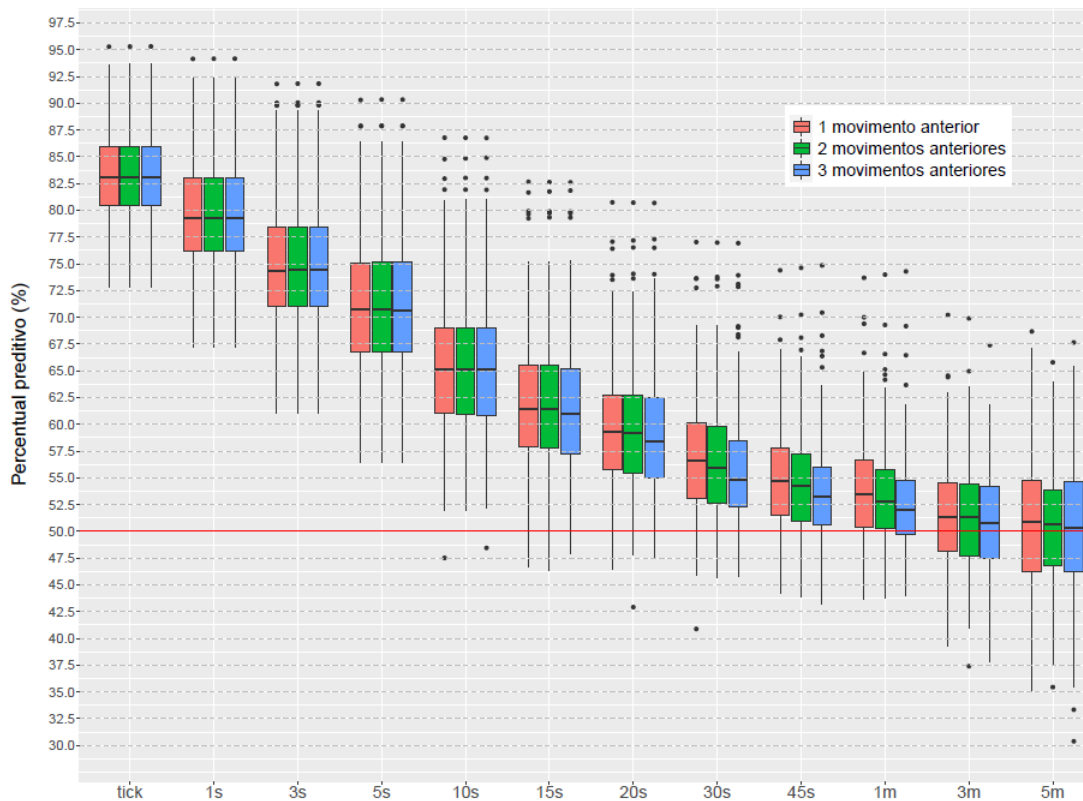


Figura 26 - Percentuais de acerto para cada modelo logístico.

Para os dados *tick a tick* e de 1 a 10 segundos, os modelos apresentaram percentuais de acerto muito semelhantes, variando em torno de 83.1%, 79.3%, 74.4%, 70.7% e 65.2%, respectivamente.

Isso aconteceu pois, independentemente da quantidade de movimentos anteriores, os modelos forneceram valores preditos sempre na mesma direção, tendo em vista a maior influência do último movimento para fins de determinação da

probabilidade preditiva. Nesses casos, portanto, o modelo somente para um movimento anterior foi suficiente para se prever adequadamente as movimentações futuras do ativo.

Nos demais casos, em geral, os modelos para 1 movimento anterior apresentaram melhores resultados que os de 2 e 3 movimentos anteriores. Nota-se ainda que os percentuais preditivos foram diminuindo gradativamente, tornando-se cada vez mais próximos dos 50%.

A Tabela 7 mostra as medianas e as médias das probabilidades de acerto observadas para cada modelo. Nota-se que essas medidas são relativamente próximas em todos os intervalos considerados.

*Tabela 7 - Acurácia dos Modelos*

<i>Período</i>	<i>Movimentos Anteriores</i>	<i>Mediana</i>	<i>Média</i>
<i>tick</i>	1	83,06%	83,38%
<i>tick</i>	2	83,06%	83,39%
<i>tick</i>	3	83,06%	83,39%
<i>1s</i>	1	79,29%	79,77%
<i>1s</i>	2	79,29%	79,77%
<i>1s</i>	3	79,29%	79,77%
<i>3s</i>	1	74,38%	74,82%
<i>3s</i>	2	74,39%	74,83%
<i>3s</i>	3	74,39%	74,83%
<i>5s</i>	1	70,71%	71,21%
<i>5s</i>	2	70,70%	71,22%
<i>5s</i>	3	70,68%	71,22%
<i>10s</i>	1	65,15%	65,41%
<i>10s</i>	2	65,13%	65,38%
<i>10s</i>	3	65,12%	65,34%
<i>15s</i>	1	61,38%	61,91%
<i>15s</i>	2	61,42%	61,80%
<i>15s</i>	3	61,00%	61,42%
<i>20s</i>	1	59,27%	59,50%
<i>20s</i>	2	59,22%	59,42%
<i>20s</i>	3	58,36%	58,92%
<i>30s</i>	1	56,63%	56,85%
<i>30s</i>	2	55,89%	56,44%
<i>30s</i>	3	54,84%	55,90%
<i>45s</i>	1	54,74%	55,03%
<i>45s</i>	2	54,25%	54,55%
<i>45s</i>	3	53,26%	53,71%
<i>1m</i>	1	53,41%	53,77%
<i>1m</i>	2	52,83%	53,15%
<i>1m</i>	3	51,99%	52,55%
<i>3m</i>	1	51,29%	51,51%
<i>3m</i>	2	51,34%	51,38%
<i>3m</i>	3	50,73%	50,94%
<i>5m</i>	1	50,92%	50,63%
<i>5m</i>	2	50,62%	50,49%
<i>5m</i>	3	50,29%	50,31%

## 7.6 ESTRATÉGIA DE OPERAÇÕES PARA APROVEITAMENTO DA CAPACIDADE PREDITIVA DOS MODELOS

Tendo sido verificada a capacidade dos modelos logísticos em prever, para a maioria dos dias, a movimentação do ativo com uma probabilidade de sucesso acima dos 50%, será proposta uma estratégia de operação em alta frequência que busque obter lucros financeiros a partir dos dados desses modelos.

A análise da viabilidade operacional da estratégia será realizada somente para os modelos com apenas um movimento anterior, tendo em vista sua capacidade preditiva ter sido, na maioria dos casos, superior aos modelos com dois ou três movimentos anteriores, sobretudo para períodos acima de 3 segundos. Da mesma forma que na seção 6, os coeficientes de cada modelo serão utilizados para prever os movimentos do ativo no dia seguinte.

A estratégia proposta é extremamente simples e consiste em operar comprado quando a probabilidade  $p$  fornecida pelo modelo for maior do que  $1/2$  e vendido à descoberto quando  $p$  for menor ou igual a  $1/2$ , sem nunca aumentar a posição de ambas as operações. Os preços de compra ou venda serão considerados como sendo o valor de fechamento do ativo para cada intervalo, conforme o modelo logístico adotado.

A Figura 27 mostra o funcionamento da estratégia de operação. Inicialmente, se a probabilidade de alta  $p$  for maior do que  $1/2$ , será realizada uma compra, caso contrário será realizado uma venda a descoberto.

Se a compra tiver sido realizada, independentemente da direção tomada pelo mercado, avalia-se novamente a probabilidade  $p$ . Caso ela permaneça maior do que  $1/2$ , será mantida a posição comprada, sem aumentar a quantidade de ativos em carteira. Do contrário, será realizada uma venda em dobro dos ativos, de forma a assumir uma posição vendida.

Por outro lado, se inicialmente tenha sido realizada uma venda a descoberto, independentemente da direção tomada pelo mercado, também se avalia novamente o valor de  $p$ . Só que agora o raciocínio será o inverso, ou seja, se ele permanecer menor ou igual a  $1/2$ , será mantida a posição vendida, sem aumentar a quantidade de

vendas. Caso contrário, será realizada uma compra dobrada, assumindo assim uma posição comprada.

As operações continuam seguindo esse padrão até o fechamento do pregão do dia, no qual será realizada uma operação de compra ou venda simples, para zerar a posição. Uma suposição da estratégia é que o investidor possui recursos suficientes para continuar executando as operações, independentemente da lucratividade de cada uma delas.

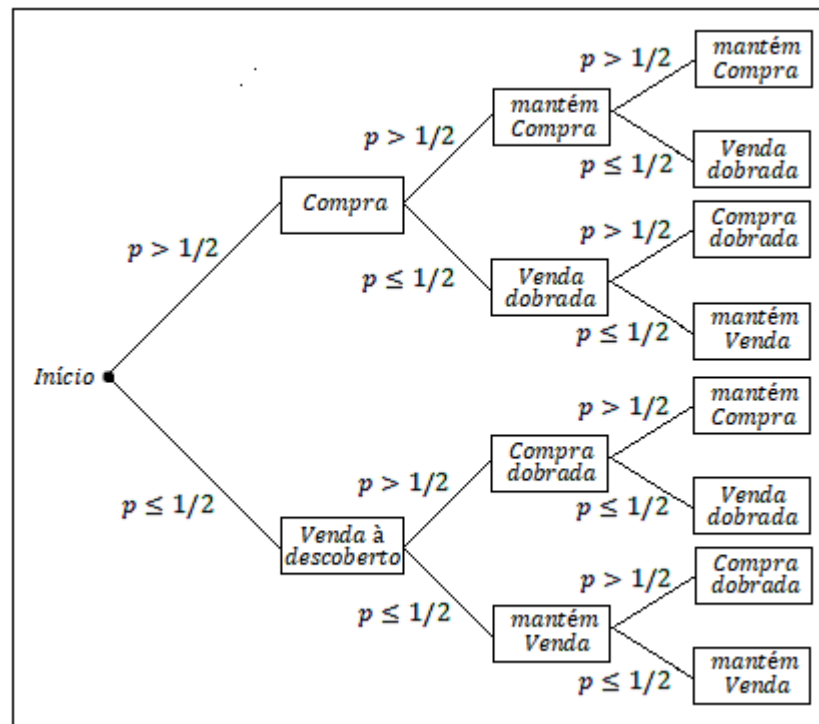


Figura 27 - Estratégia de operação.

As Figuras 28 e 29 mostram, respectivamente, as estratégias de operação quando os preços estão sempre caminhando na direção indicada pelas probabilidades dos modelos ou quando estão caminhando na direção oposta. As movimentações positivas dos preços são representadas pelas linhas ascendentes, enquanto as negativas pelas descendentes.

De forma bastante simplificada, se a diferença entre os preços sucessivos de mercado sempre for um valor fixo, ganha-se essa diferença em cada movimento na direção indicada por  $p$  ou perde-se em cada movimento na direção contrária. Assim, certamente obtém-se lucro na maioria dos dias, tendo em vista que os modelos

conseguiram prever a movimentação do ativo com uma probabilidade de acerto acima dos 50%.

Ocorre, entretanto, que não se sabe de antemão qual será a diferença dos preços em cada movimento, sendo necessária a avaliação individualizada de cada uma delas. A princípio, os custos operacionais para a verificação da viabilidade dos modelos serão desconsiderados.

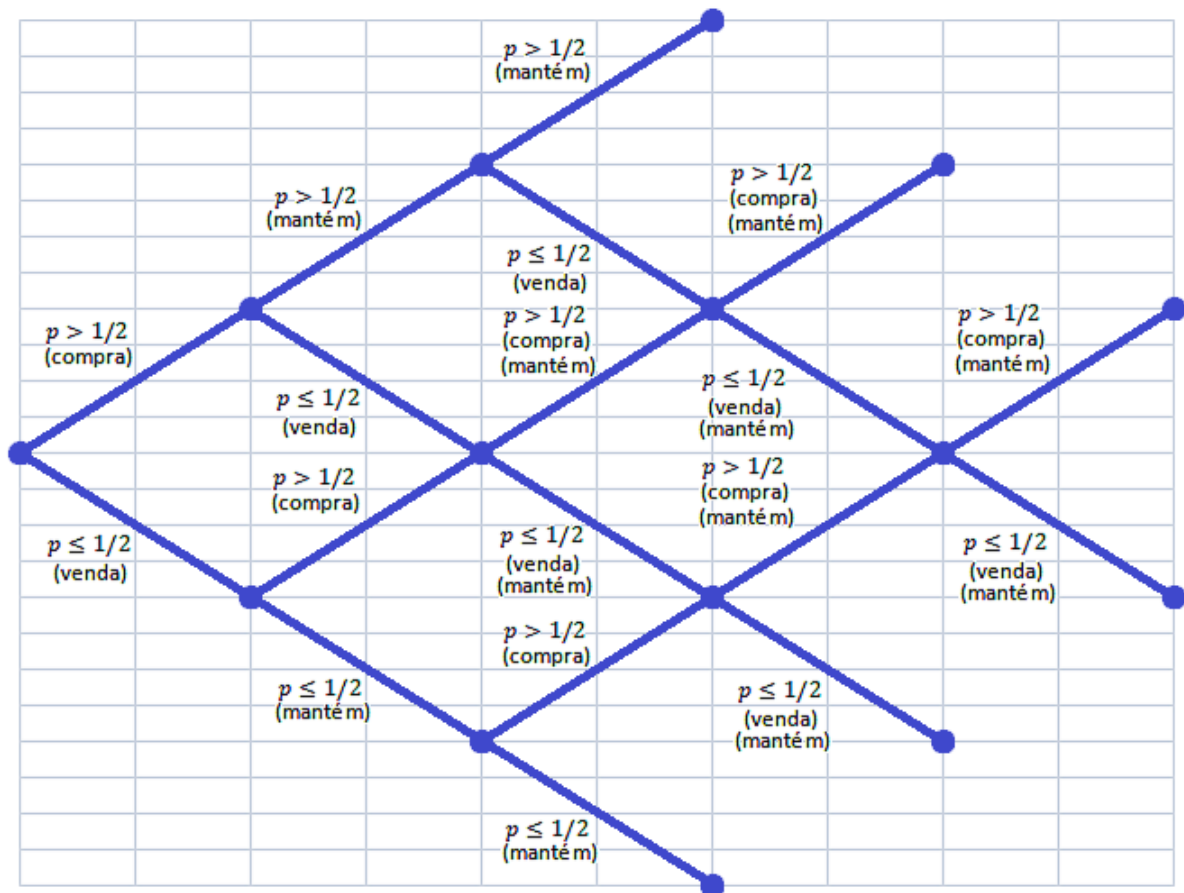


Figura 28 - Estratégias de operação bem-sucedidas



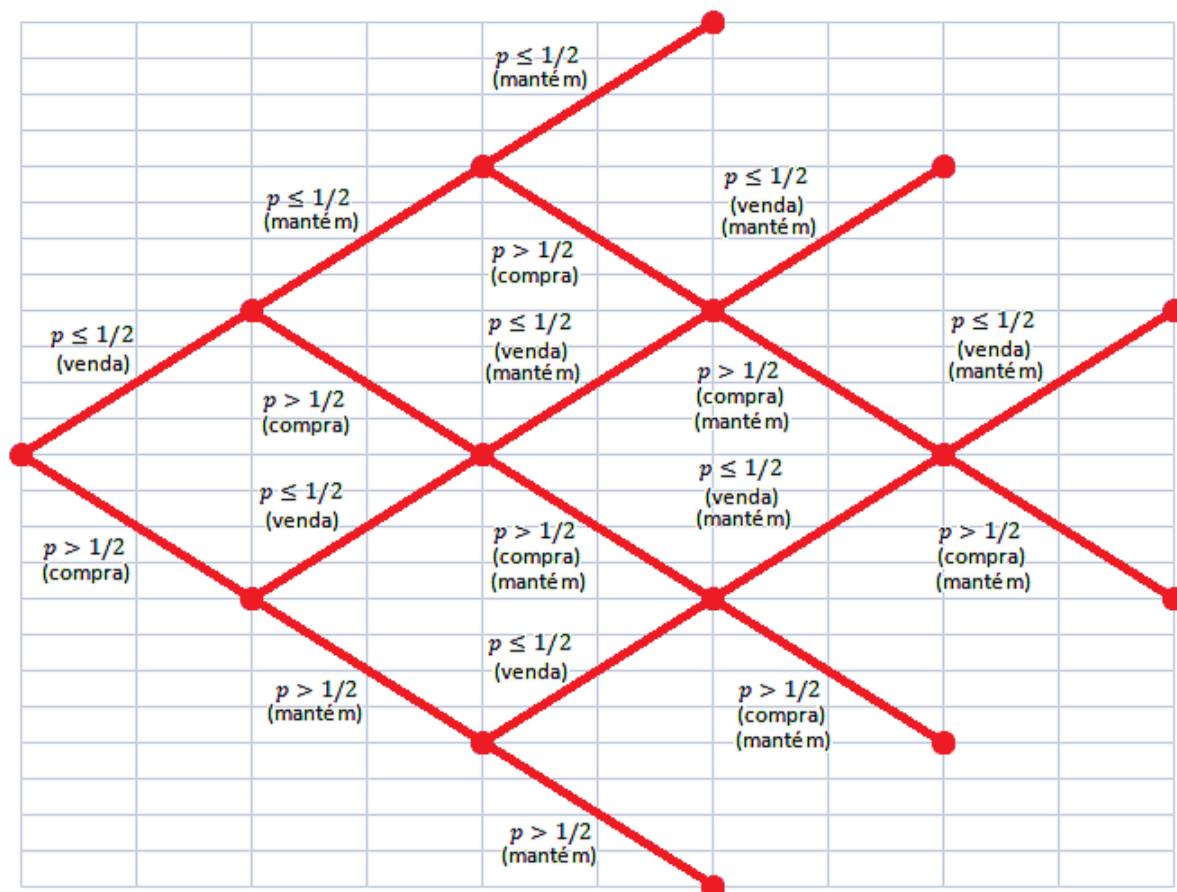


Figura 29 - Estratégias de operação malsucedidas

## 7.7 ESTRATÉGIA DE OPERAÇÃO DESCONSIDERANDO CUSTOS OPERACIONAIS

Para análise da viabilidade da estratégia operacional, as probabilidades de alta  $p$ , em cada passo, serão denotadas por  $p_0, p_1, \dots, p_n$  e os preços sucessivos do ativo por  $v_0, v_1, \dots, v_n$ . Será criada uma sequência  $s_i$ , com  $i$  variando de 1 a  $n$ , que assumirá os seguintes valores:

$$1, \quad \text{se} \begin{cases} p_{i-1} > \frac{1}{2} & \text{e } v_i > v_{i-1} \\ & \text{ou} \\ p_{i-1} \leq \frac{1}{2} & \text{e } v_i < v_{i-1} \end{cases}$$

$$-1, \quad \text{se} \begin{cases} p_{i-1} > \frac{1}{2} & \text{e } v_i < v_{i-1} \\ & \text{ou} \\ p_{i-1} \leq \frac{1}{2} & \text{e } v_i > v_{i-1} \end{cases}$$

Assim, o lucro percentual após todas as operações do dia, desconsiderando qualquer custo operacional, será dado por:

$$L = \frac{\sum_{i=1}^n s_i |v_i - v_{i-1}|}{v_0}$$

Ou seja, o módulo da diferença dos preços sucessivos será somado quando o modelo prever corretamente a movimentação do ativo e subtraído quando a previsão do modelo for incorreta. Ao final, esse resultado será dividido pelo preço inicial de compra ou venda, para verificar o lucro percentual das operações do dia.

As Figuras 30 e 31 mostram os lucros percentuais encontrados para cada intervalo nos 244 dias de aplicação da estratégia de negociação. A linha em vermelho indica o nível de 0% em cada gráfico. Nota-se que, em praticamente todos os períodos considerados, os lucros percentuais diários foram, em grande parte dos dias, positivos.

Para os dados *tick a tick*, os lucros diários foram extremamente elevados, variando em torno de 400%. Eles foram diminuindo gradativamente, ficando em torno de 183%, 90%, 57%, 27% e 15%, para os dados de 1 a 15 segundos, respectivamente.

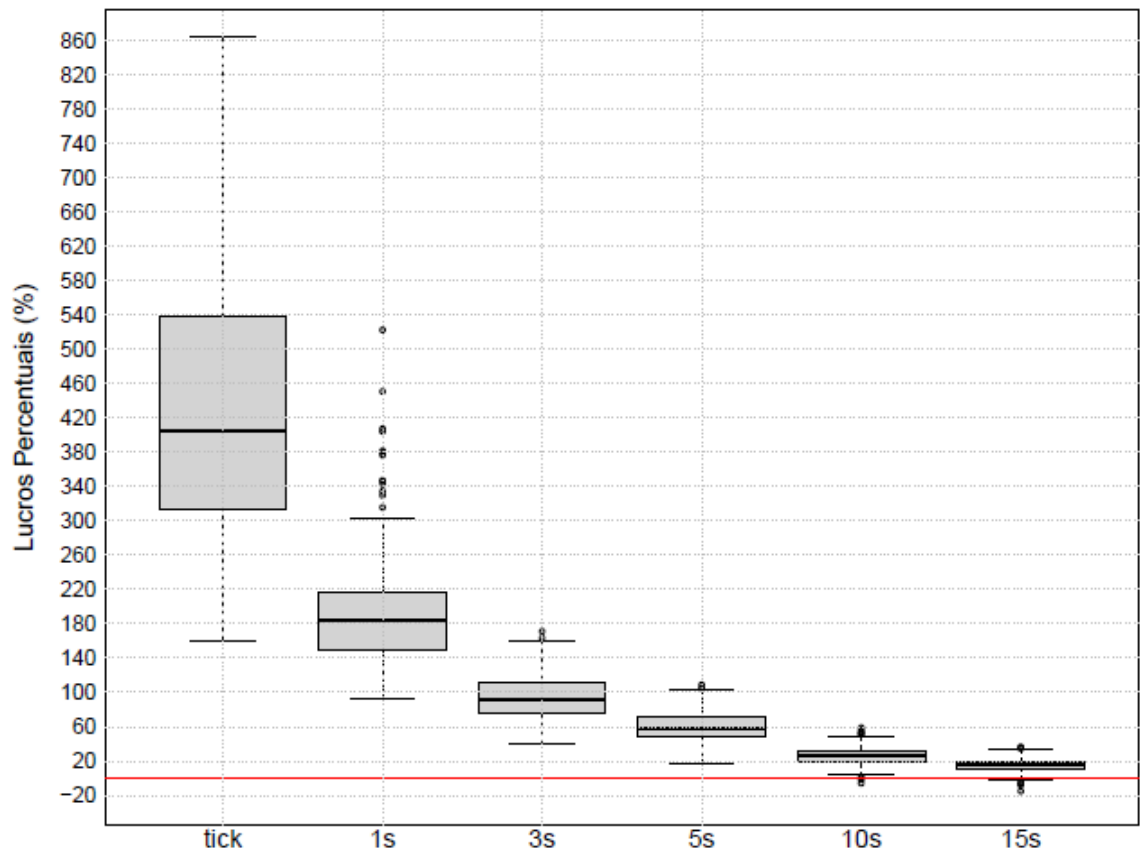


Figura 30 - Lucros Percentuais para os dados tick a tick e de até 15 segundos

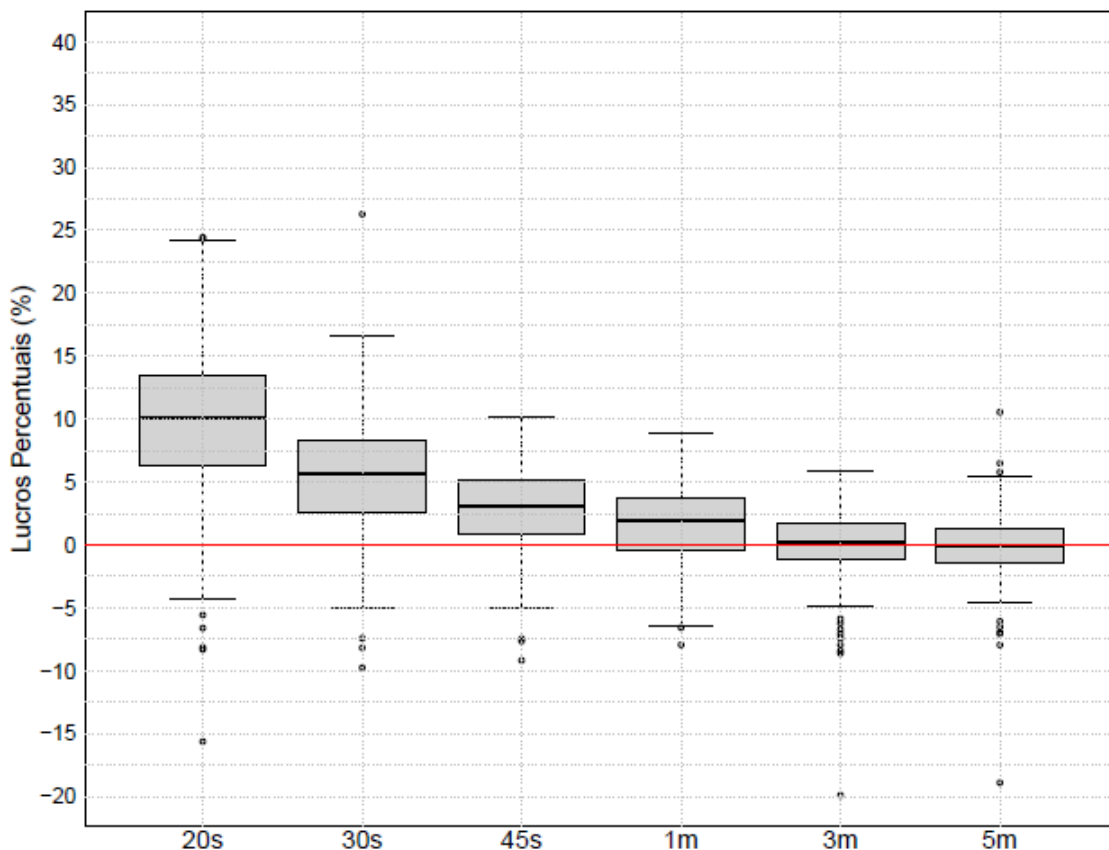


Figura 31 - Lucros Percentuais para os dados de 20 segundos até 5 minutos

Para os dados de 20 segundos a 1 minuto, os lucros percentuais ficaram entre 10% e 2%, enquanto, para os dados de 3 e 5 minutos, ficaram próximos a 0%. Nota-se ainda que a variância dos lucros percentuais se tornou cada vez menor, à medida que os períodos foram sendo aumentados.

A Tabela 8 mostra as medianas e as médias encontradas. Nota-se, com exceção dos dados *tick a tick*, uma aproximação entre essas duas medidas.

Tabela 8 – Medianas e Médias dos Lucros Diários

<i>Período</i>	<i>Mediana</i>	<i>Média</i>
<i>tick</i>	404,43%	489,25%
1s	183,28%	192,01%
3s	90,85%	95,04%
5s	56,96%	59,27%
10s	26,99%	26,50%
15s	15,64%	15,13%
20s	10,11%	9,58%
30s	5,68%	5,44%
45s	3,09%	2,73%
1m	1,86%	1,46%
3m	0,22%	0,00%
5m	-0,09%	-0,16%

## 7.8 ESTRATÉGIA DE OPERAÇÃO CONSIDERANDO CUSTOS OPERACIONAIS

Para a execução das operações de mercado, existem diversos custos envolvidos que são cobrados pela B3, pelas corretoras responsáveis pela intermediação dos negócios e pelo Estado.

Com relação às negociações à vista de ações, objeto do presente estudo, os custos cobrados pela B3 em cada negócio são denominados taxa de negociação (ou emolumento) e taxa de liquidação. Esses custos são obrigatórios e podem variar ao longo do ano, sendo utilizados para remunerar os serviços prestados pela bolsa.

Por outro lado, em cada operação, as corretoras cobram, em geral, a chamada taxa de corretagem, que pode ter um valor fixo ou variável. Caso seja cobrada, sobre esse valor também é devido o Imposto sobre Serviço (ISS), que varia conforme o município e pode chegar a até 5% do valor da corretagem.

Ocorre que atualmente já existem corretoras que não cobram taxa alguma dos seus clientes para as execuções operações. Por isso, os únicos custos que serão considerados para a análise da viabilidade da estratégia operacional são os da própria B3, discriminados na Tabela 9, para as operações que se iniciem e encerrem num mesmo dia (*day-trade*).

Tabela 9 - Custos Operacionais B3 - Day Trade. Fonte B3 [19]

Volume daytrade (R\$ milhões)		Negociação	Liquidação	Total
Pessoa física	Pessoa jurídica			
Ate 4 (inclusive)	Até 20 (inclusive)	0,003089%	0,0200%	0,023089%
De 4 até 12,5 (inclusive)	De 20 até 50 (inclusive)	0,0030%	0,0200%	0,0230%
De 12,5 até 25 (inclusive)	De 50 até 250 (inclusive)	0,0005%	0,0195%	0,0200%
De 25 até 50 (inclusive)	De 250 até 500 (inclusive)	0,0005%	0,0175%	0,0180%
Mais de 50	Mais de 500	0,0005%	0,0155%	0,0160%

Os custos para um volume financeiro de até 4 milhões de reais é de 0,023089% por operação. Esse será o valor cobrado em cada compra ou venda executada no decorrer do dia, independentemente da lucratividade da negociação.

É importante observar que, sendo o emolumento um custo proporcional, o lucro percentual obtido ao considerá-lo nas operações diárias será o mesmo, independentemente da quantidade de ações negociadas.

No nosso modelo de negociação, inicialmente, será realizada uma compra ou venda, caso valor da probabilidade  $p$  seja  $> 1/2$  ou  $\leq 1/2$ , respectivamente. Após

essa primeira operação, as demais compras ou vendas serão dobradas, a depender da mudança do valor de  $p$ . Somente a última operação voltará a ser uma compra ou venda simples, que irá zerar a posição existente.

Por exemplo, se inicialmente foi realizada uma compra, apenas se realizará uma venda dobrada quando  $p$  se tornar menor ou igual a  $1/2$ . Caso contrário, a compra será mantida, sem incorrer em novos custos operacionais.

Por outro lado, se inicialmente foi realizada uma venda, apenas será feita uma compra dobrada quando  $p$  se tornar maior do que  $1/2$ , sendo mantida a posição vendida se  $p$  permanecer menor ou igual a  $1/2$ , sem novos custos operacionais.

### 7.8.1 Diferença entre os preços e a lucratividade do modelo

Uma ponderação que deve ser feita é com relação à diferença mínima entre os preços de compra e venda para que operações bem-sucedidas venham a incorrer em lucro. Em cada operação, denominando a taxa operacional por  $T$ , o lucro obtido após o seu fechamento será dado por:

$$\text{lucro fechamento} = |v_2 - v_1| - T(v_1 + v_2)$$

Em que  $v_1$  e  $v_2$  são, respectivamente, o preço de entrada e saída da operação. Nota-se que a taxa  $T$  é aplicada tanto na operação de entrada, quando na de saída. Assim, para que  $\text{lucro fechamento} \geq 0$ , deve-se ter:

$$|v_2 - v_1| \geq T(v_1 + v_2) \leftrightarrow (v_1 + v_2) \leq \frac{|v_2 - v_1|}{T}$$

Ou seja, para ativos com preços elevados, a diferença entre os preços de entrada e saídas deverão ser cada vez maiores para que o modelo seja lucrativo. E isso se torna um problema para papéis com fortes tendências reversivas no curto prazo, pois, nesses casos, as diferenças entre os preços são geralmente muito pequenas.

Por exemplo, considere uma compra e venda bem sucedida de um ativo hipotético com  $v_1 = R\$ 25,00$  e  $v_2 > v_1$ . Substituindo na desigualdade, tem-se:

$$|v_2 - 25| \geq T(25 + v_2) \leftrightarrow v_2 \geq \frac{30(T + 1)}{1 - T}$$

Para  $T = 0,023089\%$ , tem-se que  $v_2 \geq R\$ 25,012$  e  $v_2 - v_1 \geq R\$0,012$ , que é um valor maior do que 1 (um) centavo. Ocorre que ativos muito líquidos possuem fortes tendências reversivas no curto prazo e costumam variar em torno de 1 (um)

centavo para os dados *tick* e *tick*. Assim, certamente essa estratégia operacional teria prejuízos nesses casos.

Pode-se ainda verificar qual o limiar máximo do preço para que operações bem sucedidas com diferenças da ordem de 1 (um) centavo sejam lucrativas. Considerando uma compra e venda em que  $v_2 = v_1 + 0,01$  e substituindo na desigualdade, tem-se:

$$|v_1 + v_1 + 0,01| \leq \frac{|v_1 + 0,01 - v_1|}{T} \leftrightarrow v_1 \leq \frac{1 - T}{200T}$$

Para  $T = 0,023089\%$ , tem-se que  $v_1 \leq R\$ 21,65$ , que é um valor abaixo dos preços de grande parte dos ativos mais líquidos negociados na B3.

### 7.8.2 Análise dos Resultados encontrados

A partir da formulação discriminada na Seção 7.7, será criada uma nova sequência denominada  $r_i$ , com  $i$  variando de 1 a  $n-1$ , que assumirá os seguintes valores:

$$1, \quad \text{se} \begin{cases} p_i > \frac{1}{2} \text{ e } p_{i-1} \leq \frac{1}{2} \\ \text{ou} \\ p_i \leq \frac{1}{2} \text{ e } p_{i-1} > \frac{1}{2} \end{cases}$$

$$0, \quad \text{se} \begin{cases} p_i > \frac{1}{2} \text{ e } p_{i-1} > \frac{1}{2} \\ \text{ou} \\ p_i \leq \frac{1}{2} \text{ e } p_{i-1} \leq \frac{1}{2} \end{cases}$$

Denotando a taxa operacional por  $T$ , o prejuízo percentual decorrente somente dos custos operacionais será dado por:

$$P = \frac{Tv_0 + 2T \sum_{i=1}^{n-1} r_i v_i + Tv_n}{v_0} = T \left( \frac{v_0 + v_n + 2 \sum_{i=1}^{n-1} r_i v_i}{v_0} \right)$$

Ou seja, será realizada uma operação inicial de compra ou venda, acarretando num custo  $Tv_0$ . Após, serão feitas somente operações de compra ou venda dobradas sempre que o modelo logístico indicar mudança no sentido do movimento do ativo. Essa indicação será feita por  $r_i = 1$ , implicando num aumento dos custos operacionais em  $2Tv_i$ . Apenas na última operação, volta-se a realizar um compra ou venda simples, para zerar a posição, acarretando num custo final  $Tv_n$ .

Assim, o resultado percentual  $R$  de todas as operações do dia, com a inclusão dos custos operacionais, será dado por:

$$R = L - P = \frac{\sum_{i=1}^n s_i |v_i - v_{i-1}|}{v_0} - T \left( \frac{v_0 + v_n + 2 \sum_{i=1}^{n-1} r_i v_i}{v_0} \right) \rightarrow$$

$$R = \frac{\sum_{i=1}^n s_i |v_i - v_{i-1}| - T(v_0 + v_n + 2 \sum_{i=1}^{n-1} r_i v_i)}{v_0}$$

As Figuras 32 e 33 mostram os resultados percentuais encontrados para cada período nos 244 dias de aplicação da estratégia de negociação. A linha em vermelho indica o nível de 0% em cada gráfico.

Nota-se que, em todos os períodos considerados, os resultados percentuais diários foram, na maioria dos dias, negativos, o que mostra uma influência muito alta dos custos operacionais para a lucratividade da estratégia.

Para os dados *tick a tick* de 1 segundo, os resultados variaram em torno de -27,5% e -20,5%. Eles foram aumentando gradativamente, ficando em torno de -1,4% para os dados de 5 minutos. Vê-se, ainda, uma diminuição da variância dos lucros percentuais com o aumento da duração dos períodos.

A Tabela 10 mostra as medianas e as médias dos resultados diários. Observa-se uma aproximação gradual entre essas medidas.



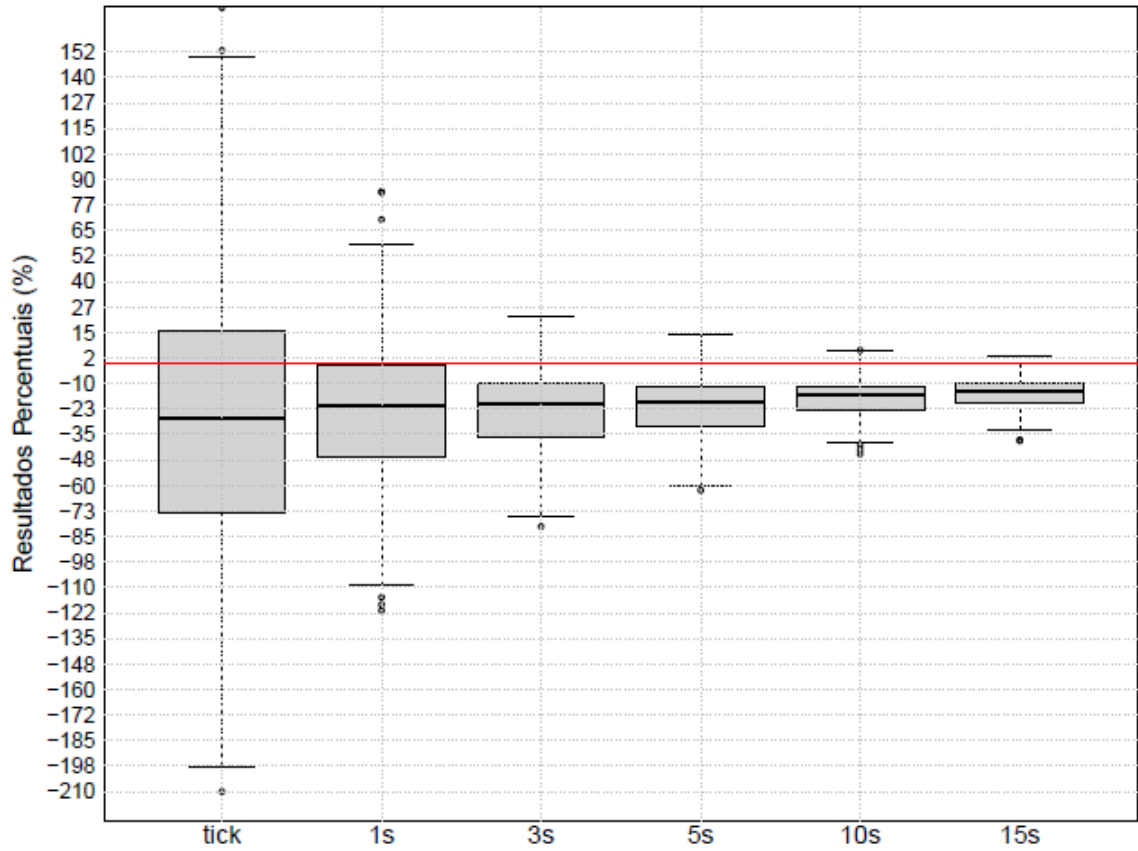


Figura 32 - Resultados Percentuais para os dados tick a tick e de até 15 segundos

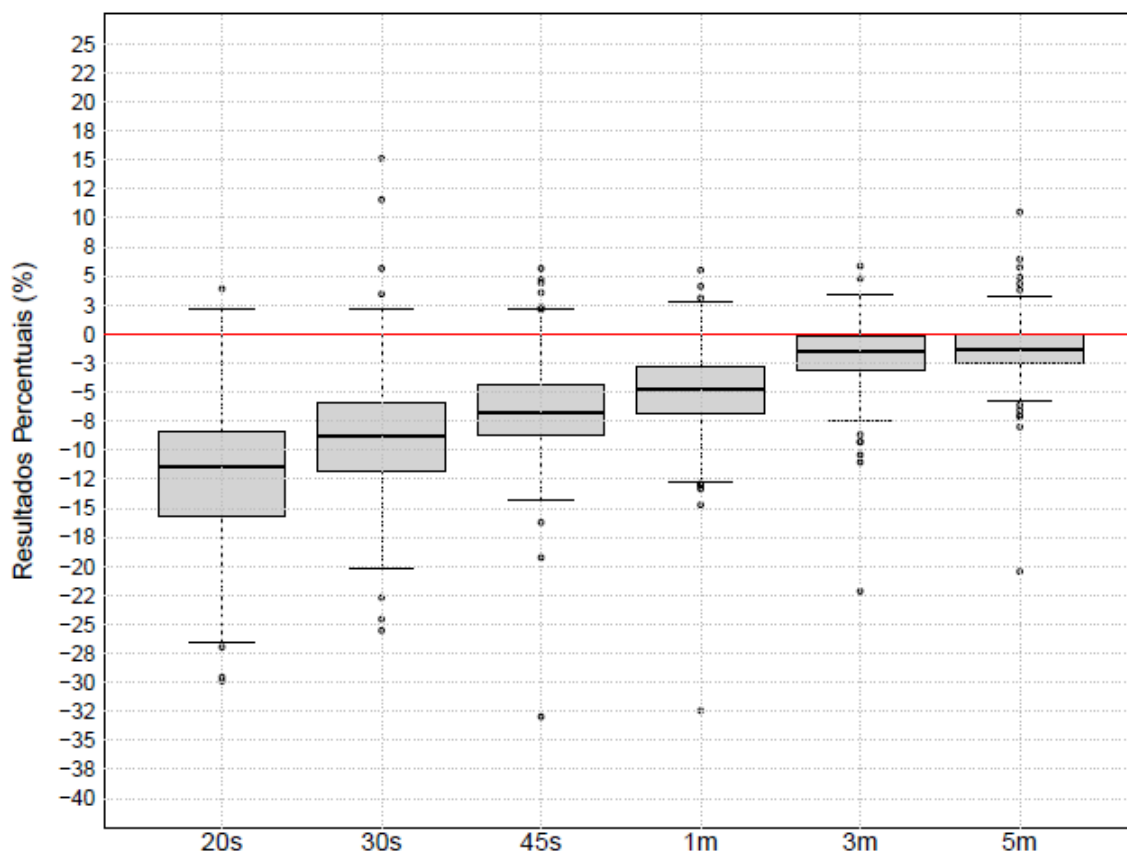


Figura 33 - Resultados Percentuais para os dados de 20 segundos até 5 minutos

Tabela 10 – Medianas e Médias dos Resultados Diários

Período	Mediana	Média
tick	-27,46%	-22,59%
1s	-20,52%	-23,94%
3s	-19,60%	-23,16%
5s	-18,83%	-20,99%
10s	-15,47%	-17,12%
15s	-13,54%	-14,51%
20s	-11,42%	-12,10%
30s	-8,83%	-8,69%
45s	-6,72%	-6,42%
1m	-4,74%	-4,75%
3m	-1,56%	-1,74%
5m	-1,39%	-1,24%

Portanto, a inclusão dos custos operacionais no modelo de negociação eliminou praticamente todos os lucros que seriam obtidos sem eles, apesar de alguns dias mostrarem resultados bastante elevados.

Isso ocorreu não apenas em função da necessidade de se ter uma diferença mínima entre os preços de entrada e saída para a lucratividade dos negócios bem

sucedidos, mas também pelo fato dos custos operacionais incidirem, inclusive, sobre as operações que tiveram prejuízo (malsucedidas).

### 7.8.3 Análise dos Custos operacionais

As Figuras 34 e 35 mostram os custos operacionais calculados para cada período nos 244 dias de aplicação da estratégia. Para os dados *tick a tick*, os custos foram extremamente elevados, variando em torno de 435%. Eles diminuíram gradativamente com o aumento dos períodos, variando em torno de 1,6% para os dados de 5 minutos.

Nota-se também uma diminuição gradual da variância, com exceção dos dados de 1 a 5 minutos, para os quais houve o aumento dessa medida em relação ao período de 45 segundos. A Tabela 11 mostra as medianas e as médias dos custos diários.

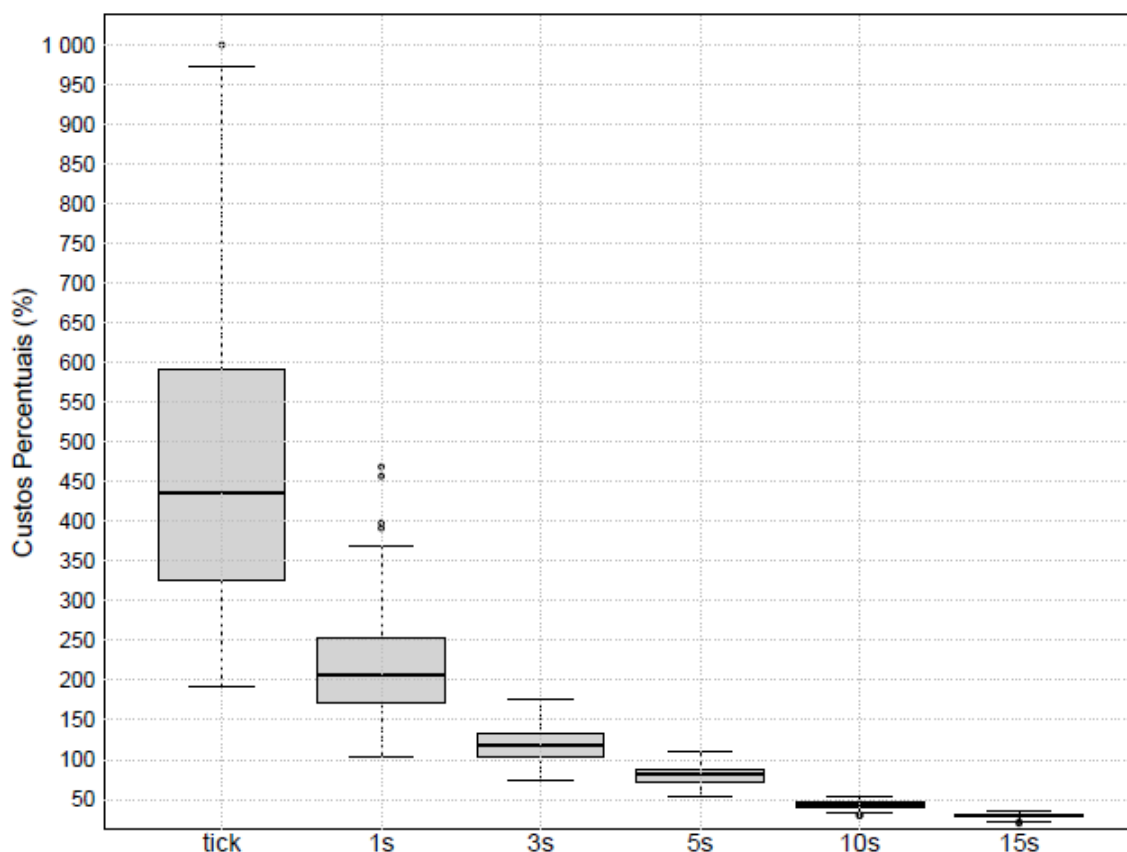


Figura 34 - Custos Percentuais para os dados *tick a tick* e de até 15 segundos

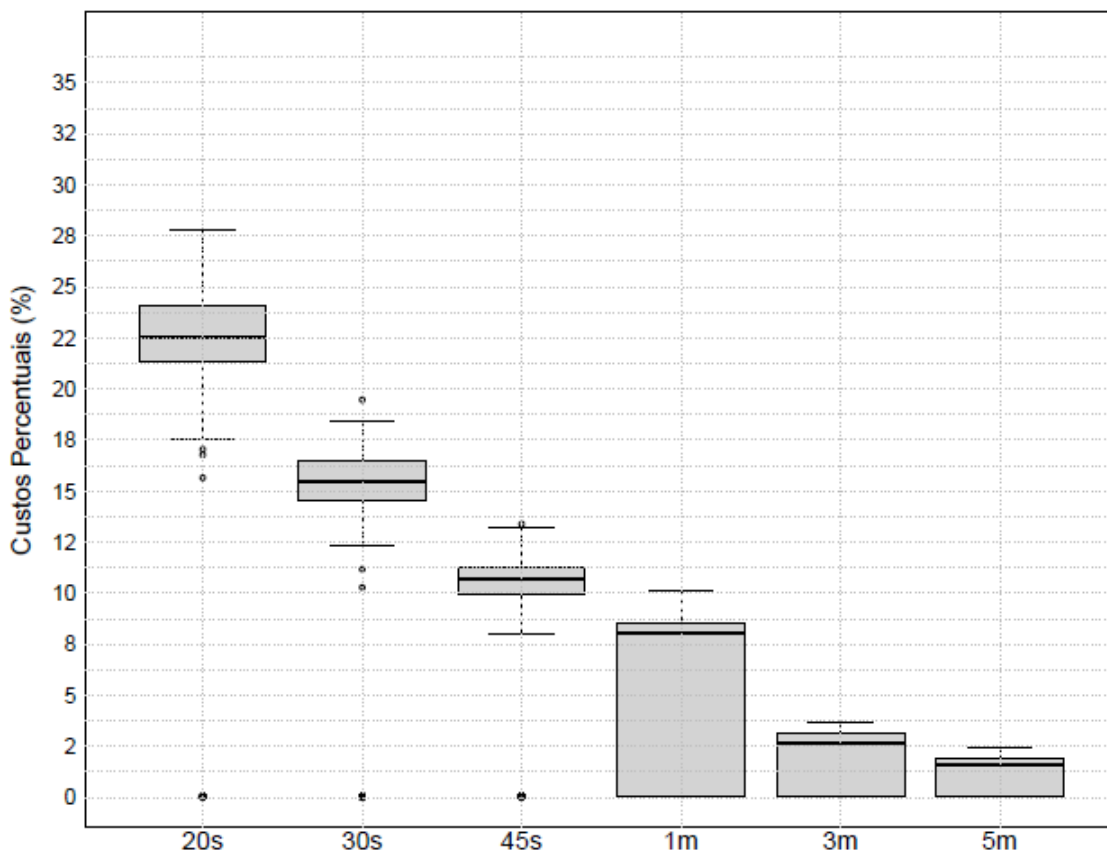


Figura 35 - Custos Percentuais para os dados de 20 segundos até 5 minutos

Tabela 11 – Medianas e Médias dos Custos Diários

Período	Mediana	Média
tick	435,16%	511,84%
1s	205,98%	215,95%
3s	117,95%	118,20%
5s	81,16%	80,27%
10s	43,74%	43,62%
15s	30,12%	29,64%
20s	22,54%	21,68%
30s	15,51%	14,12%
45s	10,68%	9,14%
1m	8,06%	6,22%
3m	2,66%	1,74%
5m	1,63%	1,08%

A Tabela 12 mostra, de forma consolidada, as médias diárias de cada intervalo dos seguintes dados: resultados sem qualquer custo operacional; custos operacionais; e resultados com a inserção dos custos operacionais.

Observa-se, de forma clara, os custos operacionais superarem os resultados dos modelos sem qualquer custo envolvido. Assim, considerando os níveis atuais dos

custos operacionais cobrados pela B3, a estratégia operacional proposta não traria resultados favoráveis.

Tabela 12 - Médias diárias dos Lucros, Custos operacionais e dos Resultados

Período	Lucros	Custos Operacionais	Resultados
<i>tick</i>	489,25%	511,84%	-22,59%
1s	192,01%	215,95%	-23,94%
3s	95,04%	118,20%	-23,16%
5s	59,27%	80,27%	-20,99%
10s	26,50%	43,62%	-17,12%
15s	15,13%	29,64%	-14,51%
20s	9,58%	21,68%	-12,10%
30s	5,44%	14,12%	-8,69%
45s	2,73%	9,14%	-6,42%
1m	1,46%	6,22%	-4,75%
3m	0,00%	1,74%	-1,74%
5m	-0,16%	1,08%	-1,24%

As Figuras 36 a 38 mostram os percentuais acumulados aritmeticamente dos lucros, custos operacionais e resultados, nas cores azul, vermelho e verde, respectivamente. A linha em preto indica o nível de 0%.

Para os dados *tick a tick*, os lucros e custos acumulados foram próximos entre si. No entanto, conforme visto, em média, os custos foram 22,29% maior do que os lucros, evidenciando, ao final dos dias, um resultado percentual extremamente negativo.

Para os demais intervalos, os custos acumulados foram claramente superiores aos respectivos lucros. Vê-se ainda um distanciamento cada vez maior entre essas medidas no decorrer dos dias.

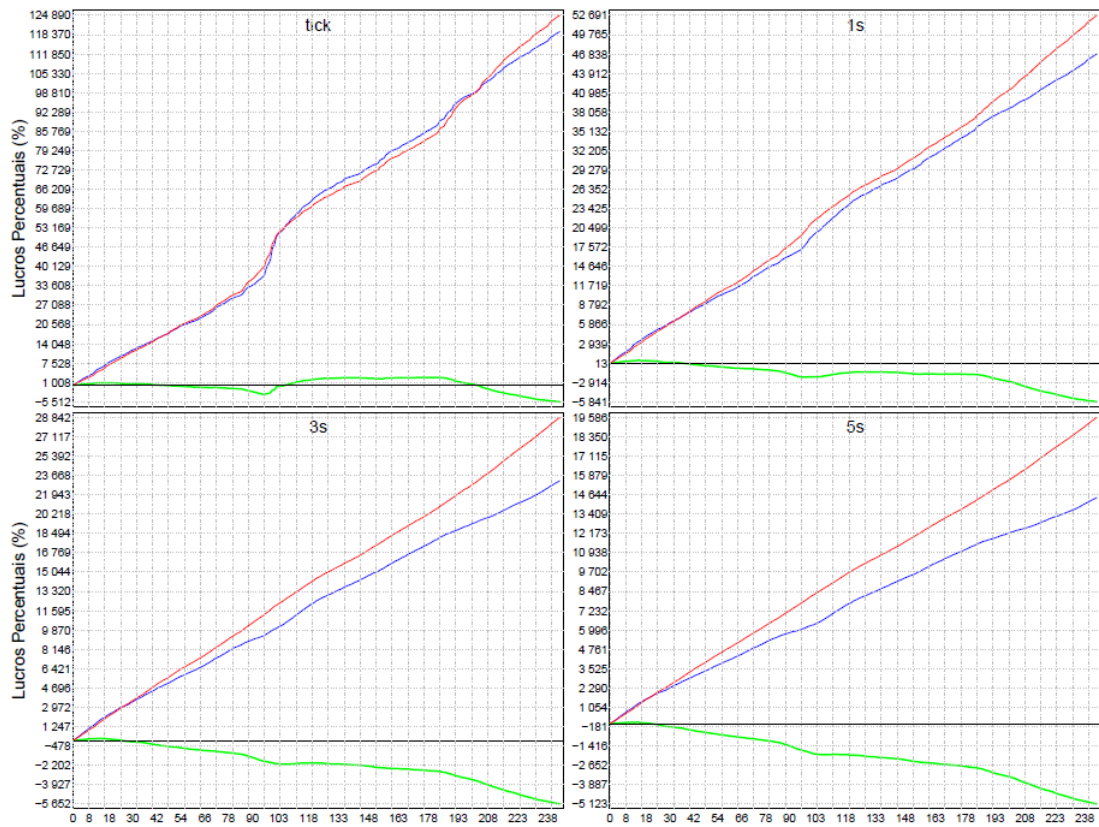


Figura 36 - Lucros, Custos e Resultados Percentuais Acumulados Aritmeticamente para os dados tick a tick até 5 segundos.

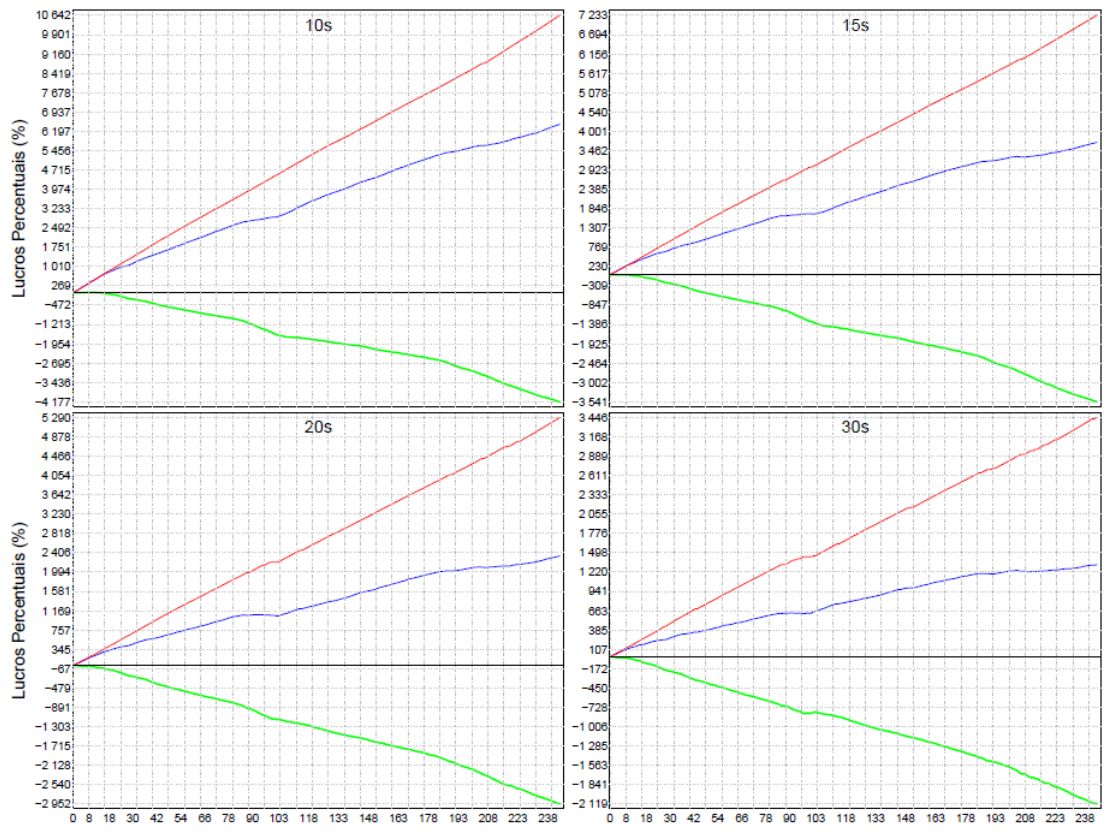


Figura 37 - Lucros, Custos e Resultados Percentuais Acumulados Aritmeticamente para os dados de 10 a 45 segundos.

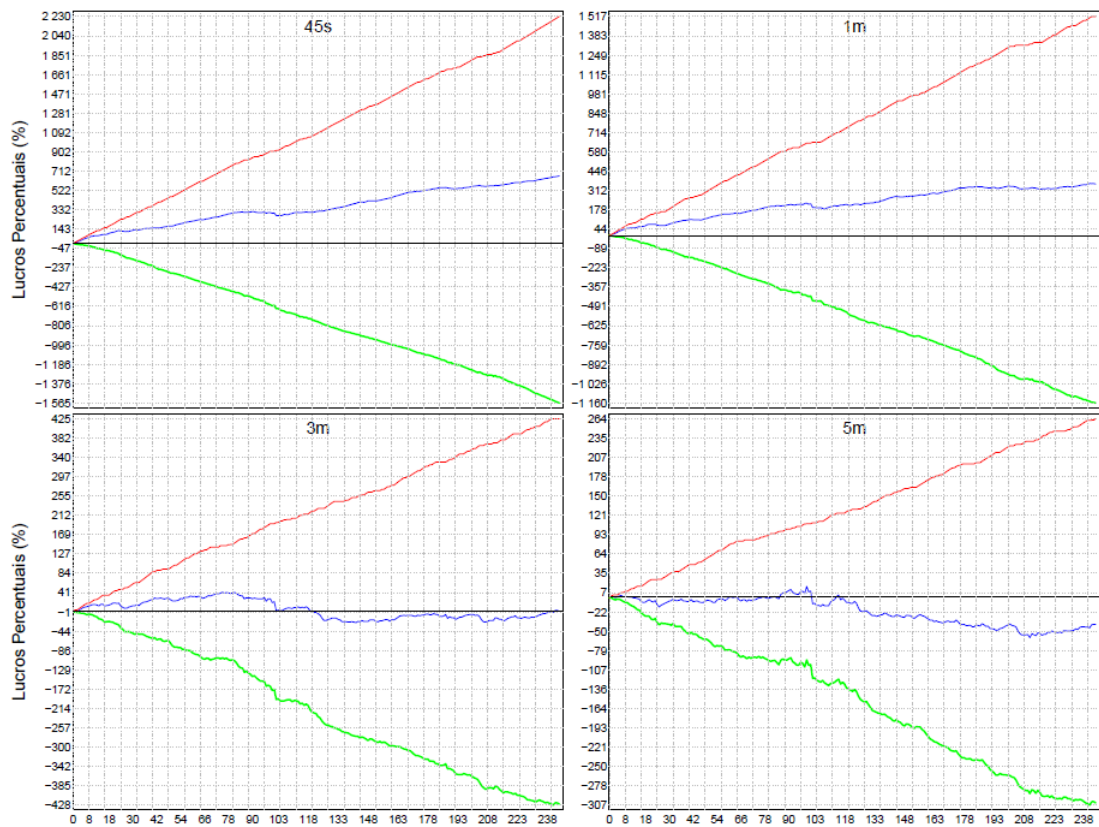


Figura 38 - Lucros, Custos e Resultados Percentuais Acumulados Aritmeticamente para os dados de 1 a 5 minutos.



## 8. MODELO PARA A MOVIMENTAÇÃO DOS PREÇOS DOS ATIVOS E PRECIFICAÇÃO DE OPÇÕES

Na seção 6, concluiu-se que a movimentação dos preços dos ativos financeiros é potencialmente ineficiente no curto prazo e que informações a priori impactam o comportamento dos preços durante certo período. Chegou-se ainda à conclusão de que, no curto prazo, a hipótese do passeio aleatório não seria válida.

Na seção 7, foram propostos modelos logísticos de predição da variação dos preços a partir dos últimos movimentos. Os modelos se mostraram bastante satisfatórios, sendo capazes de prever o sentido da movimentação do mercado com percentuais de acerto acima dos 50%, chegando a superar a barreira dos 60% para os dados *tick a tick* e de 1 a 15 segundos.

Nesta seção, será proposto um modelo de movimentação dos ativos financeiros dentro de um mesmo dia que busque absorver os resultados da presente pesquisa.

Viu-se, na seção 3, que o movimento Browniano geométrico é comumente aceito e utilizado para descrever o comportamento dos ativos financeiros. No entanto, esse movimento está alicerçado sobre a teoria do passeio aleatório, não sendo, portanto, um bom modelo para a movimentação dos ativos no curto prazo.

Conforme foi abordado, a versão de tempo discreto para o movimento Browniano geométrico pode ser representada por:

$$S_i = S_{i-1} \exp(a\Delta t + \sigma Y \sqrt{\Delta t}), \quad i = 1, 2, \dots$$

Em que  $Y \sim N(0,1)$ . Constatou-se também que, em cada etapa, a probabilidade de alta do ativo será a probabilidade de  $E_i = a\Delta t + \sigma Y \sqrt{\Delta t} \sim N(a\Delta t, \sigma^2 \Delta t)$  ser maior do que 0.

O modelo proposto utilizará como base a versão em tempo discreto do movimento Browniano geométrico. No entanto, a variável estocástica  $Y$  será substituída por outras duas com médias diferentes de zero, de tal forma que a probabilidade de alta do ativo será  $P(+|-)$ , quando o último movimento tiver sido negativo, ou  $P(+|+)$ , quando o último movimento tiver sido positivo ou nulo.

Na proposta, tanto  $P(+|-)$  quanto  $P(+|+)$  serão indicadas pelas médias da respectiva probabilidade  $p$  fornecida pelos modelos logísticos aplicados às movimentações dos  $r=20$  dias anteriores.

Viu-se que, para 1 movimento anterior, o valor de  $p$  será dado por:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1)}}$$

Em que  $\beta_0$  é o coeficiente do intercepto, relacionado ao último movimento negativo, variável (-), e  $\beta_1$  é o coeficiente relacionado ao último movimento positivo, variável (+). Assim, para um dia  $j$  qualquer,  $P(+|-)$  e  $P(++)$  serão estimados por:

$$P(+|-) = \frac{1}{r} \sum_{d=j-r}^{j-1} p_d, \quad \text{para } p_d = \frac{1}{1 + e^{-(\beta_{0d})}}$$

$$P(++ ) = \frac{1}{r} \sum_{d=j-r}^{j-1} p_d, \quad \text{para } p_d = \frac{1}{1 + e^{-(\beta_{0d} + \beta_{1d})}}$$

Em que  $\beta_{0d}$  e  $\beta_{1d}$  são os coeficientes  $\beta_0$  e  $\beta_1$  estimados para cada um dos  $d$  dias. O primeiro movimento do dia será o mesmo fornecido pela versão em tempo discreto do movimento Browniano geométrico, ou seja:

$$S_1 = S_0 \exp(a\Delta t + \sigma Y \sqrt{\Delta t}), \quad Y \sim N(0,1).$$

A partir daí, os demais movimentos serão dados pelas seguintes equações:

$$S_i = \begin{cases} S_{i-1} \exp(a\Delta t + \sigma D \sqrt{\Delta t}), & \text{se } S_{i-1} < S_{i-2} \\ S_{i-1} \exp(a\Delta t + \sigma U \sqrt{\Delta t}), & \text{se } S_{i-1} \geq S_{i-2} \end{cases}, \quad i = 2, 3, \dots$$

Em que  $D$  e  $U$  são variáveis aleatórias normais com médias  $\mu_d$  e  $\mu_u$ , respectivamente, e variâncias 1, ou seja  $D \sim N(\mu_d, 1)$  e  $U \sim N(\mu_u, 1)$ . Os parâmetros  $\mu_d$  e  $\mu_u$  serão definidos de forma que as probabilidades de  $E_D = a\Delta t + \sigma D \sqrt{\Delta t} \sim N(a\Delta t + \sigma \mu_d \sqrt{\Delta t}, \sigma^2 \Delta t)$  e  $E_U = a\Delta t + \sigma U \sqrt{\Delta t} \sim N(a\Delta t + \sigma \mu_u \sqrt{\Delta t}, \sigma^2 \Delta t)$  serem maior do que zero sejam iguais a  $P(+|-)$  e  $P(++ )$ .

Deste modo, a partir das funções de distribuições de probabilidade de  $E_D$  e  $E_U$ , indicadas por  $f_D$  e  $f_U$ , deseja-se que:

$$P(+|-) = \int_0^{\infty} f_D(x) dx \quad \text{e} \quad P(++ ) = \int_0^{\infty} f_U(x) dx$$

Nota-se que  $\mu_d$  e  $\mu_u$  podem ser obtidos a partir de uma simples transformação nos quantis  $Q_{P(+|-)}$  e  $Q_{P(++ )}$  de uma distribuição  $N(0, \sigma^2 \Delta t)$ . Assim, a partir desses valores, pode-se encontrar:

$$\mu_d = \frac{Q_{P(+|-)} - a\Delta t}{\sigma \sqrt{\Delta t}} \quad \text{e} \quad \mu_u = \frac{Q_{P(++ )} - a\Delta t}{\sigma \sqrt{\Delta t}}$$

Por fim, da mesma forma que no movimento Browniano geométrico, considera-se  $a = \mu - \frac{1}{2}\sigma^2$ , em que  $\mu$  é a taxa de crescimento esperada. A seguir, estão resumidas as equações do movimento Browniano geométrico e do novo modelo.

Movimento Browniano Geométrico:

$$S_i = S_{i-1} \exp(a\Delta t + \sigma Y \sqrt{\Delta t}), \quad Y \sim N(0,1), \quad i = 1,2,3, \dots$$

Novo Modelo:

$$S_i = \begin{cases} S_{i-1} \exp(a\Delta t + \sigma Y \sqrt{\Delta t}), & Y \sim N(0,1), \quad i = 1 \\ S_{i-1} \exp(a\Delta t + \sigma D \sqrt{\Delta t}), & \text{se } S_{i-1} < S_{i-2}, \quad D \sim N(\mu_d, 1), \quad i > 1 \\ S_{i-1} \exp(a\Delta t + \sigma U \sqrt{\Delta t}), & \text{se } S_{i-1} \geq S_{i-2}, \quad U \sim N(\mu_u, 1), \quad i > 1 \end{cases}$$

Observou-se na seção 7 que, à medida que se aumenta o intervalo considerado,  $P(+|-)$  e  $P(+|+)$  ficam cada vez mais próximos de  $P(+)$ , ou seja, o mercado se comporta de forma mais eficiente e o movimento Browniano geométrico se torna um bom modelo para a movimentação dos ativos financeiros.

É importante observar que o modelo proposto incorpora essa característica do mercado, pois, à medida que  $P(+|-)$  e  $P(+|+)$  se aproximam de  $P(+)$ ,  $\mu_d$  e  $\mu_u$  se aproximam de zero e o modelo proposto se aproxima do movimento Browniano geométrico.

### 8.1 UM EXEMPLO HIPOTÉTICO

Para ilustrar o modelo, gerou-se computacionalmente alguns dados para um dia hipotético com  $S(0) = 10$ ,  $\mu = 0,08$  e  $\sigma = 0,4$ . Para esses parâmetros,  $a = \mu - \frac{1}{2}\sigma^2 = 0$  e, no movimento Browniano geométrico,  $P(+)$  = 50%. Considerou-se que os modelos logísticos forneceram  $P(+|-)$  = 80% e  $P(+|+)$  = 20%, para um  $\Delta t = 10^{-3}$ .

As médias das variáveis  $D$  e  $U$  utilizadas foram  $\mu_d = 0,8416$  e  $\mu_u = -0,8416$ . Com essa atribuição, foi possível garantir que as probabilidades de  $E_D = a\Delta t + \sigma D \sqrt{\Delta t}$  e  $E_U = a\Delta t + \sigma U \sqrt{\Delta t}$  serem maior do que zero fossem iguais a  $P(+|-)$  e  $P(+|+)$ , respectivamente.

A Figura 39 mostra o caminho percorrido pelo ativo durante um dia para 20 simulações diferentes. Em cada uma delas foram gerados um total de  $n = 1/\Delta t = 1.000$  preços diários. As médias das probabilidades encontradas foram  $P(+)$  =

49,89% ,  $P(+|-) = 79,79\%$  e  $P(++ ) = 19,84\%$  , valores muito próximos dos esperados.

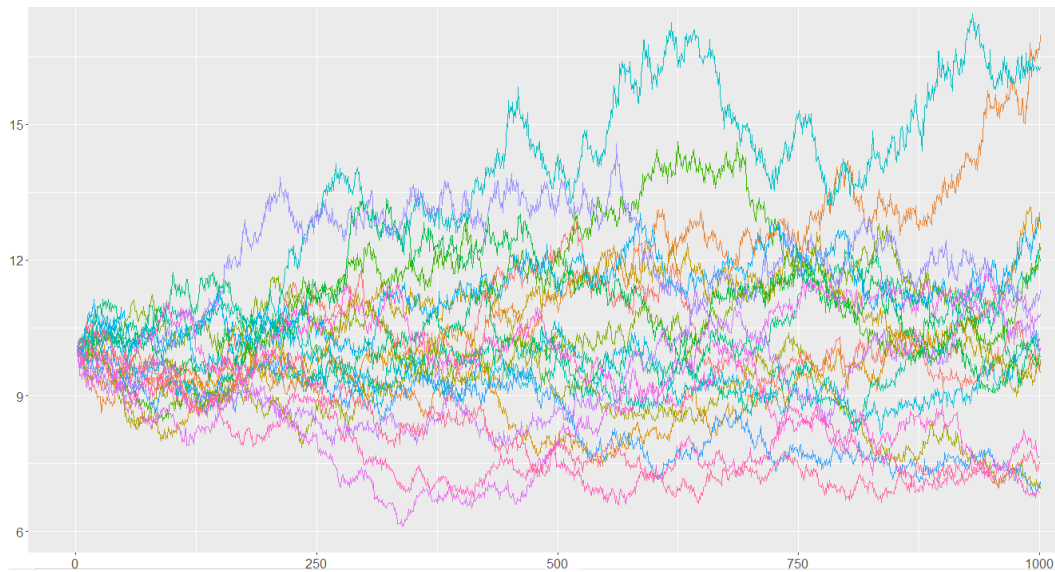


Figura 39 - Movimentação diária para 20 simulações diferentes.

Para a análise da distribuição dos preços finais ( $S_n$ ), foram realizadas 100.000 simulações diárias. Desta vez, as médias das probabilidades encontradas foram  $P(+ ) = 50,00\%$  ,  $P(+|-) = 80,01\%$  e  $P(++ ) = 19,98\%$  . A Figura 40 mostra o histograma e a densidade aproximada desses preços. A distribuição dos preços finais é aproximadamente *log-normal*.

Verificou-se ainda que o logaritmo natural de  $S_n$  possui uma média de 2,3 e variância de 0,09452. A Figura 41 mostra a distribuição aproximada de  $\log(S_n)$ , em azul, e a densidade exata de uma variável normal com esses mesmos parâmetros, em preto. As curvas são extremamente parecidas, corroborando a *log-normalidade* de  $S_n$ .

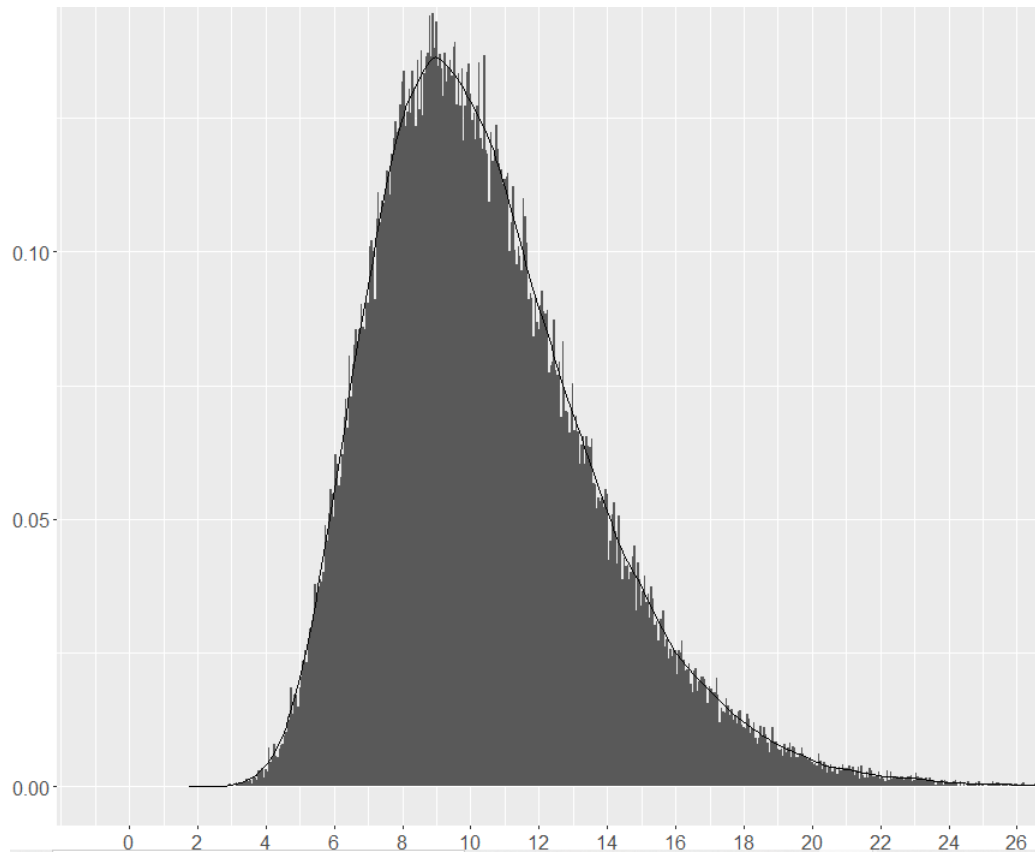


Figura 40 - Histograma e densidade aproximada de 100.000 simulações diárias.

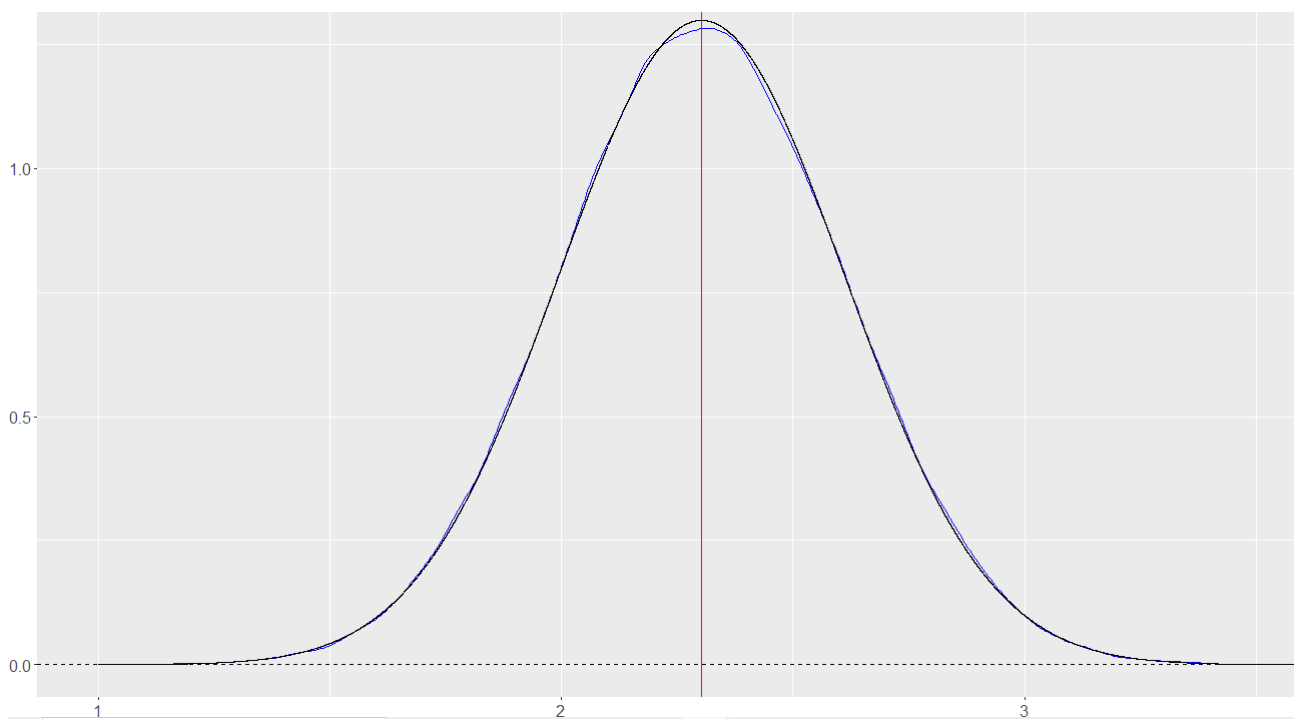


Figura 41 - Densidade de  $\log(S_n)$  e de uma distribuição  $N(2,3; 0,09452)$

Na seção 3, observou-se que o movimento Browniano geométrico possui uma distribuição *log-normal* da forma  $\ln S(t) \sim N(\ln S(0) + at, \sigma^2 t)$ . A Figura 42 mostra a curva *log-normal* de  $S(t)$ , em  $t = 1$ , em preto, e a distribuição aproximada obtida pelo modelo proposto, em azul. A linha vertical indica a média de  $S_n$ , no valor de 10,49.

Viu-se também que a média de  $S(t)$  é dada por  $S(0)e^{\mu t}$ . Assim, em  $t = 1$ ,  $E(S(t)) = 10,83$ , maior do que a média do novo modelo.

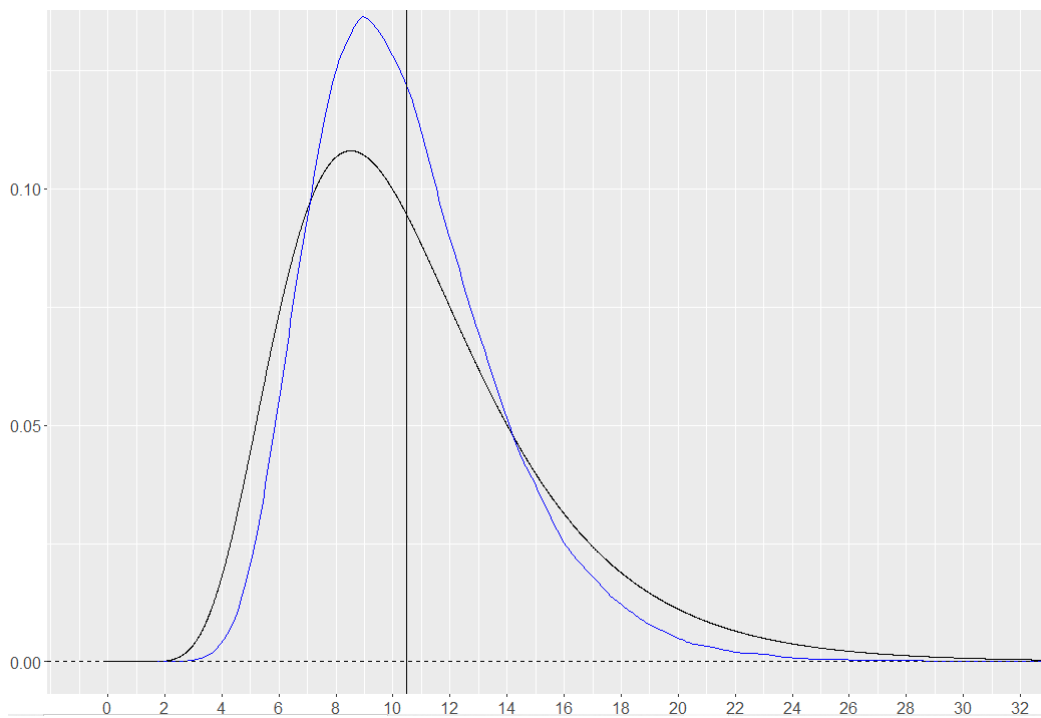


Figura 42 - Densidade de  $S_n$  e a distribuição *log-normal* do movimento Browniano geométrico.

A distribuição gerada pelo novo modelo é menos achatada que a do movimento Browniano geométrico. Esse resultado é intuitivo, tendo em vista uma maior probabilidade de reversão dos movimentos anteriores dada por  $P(+|-) = 80\%$  e  $P(+|+) = 20\%$ .

Essa característica pode ser observada de forma mais clara por meio de suas correspondentes distribuições normais. A Figura 43 mostra a distribuição normal que ajustou  $\log(S_n)$ , em azul, e a distribuição de  $\ln S(t) \sim N(\ln S(0) + at, \sigma^2 t)$ , em preto. A linha vertical indica as suas médias, ambas iguais a  $\ln S(0) + at = 2,3$ .

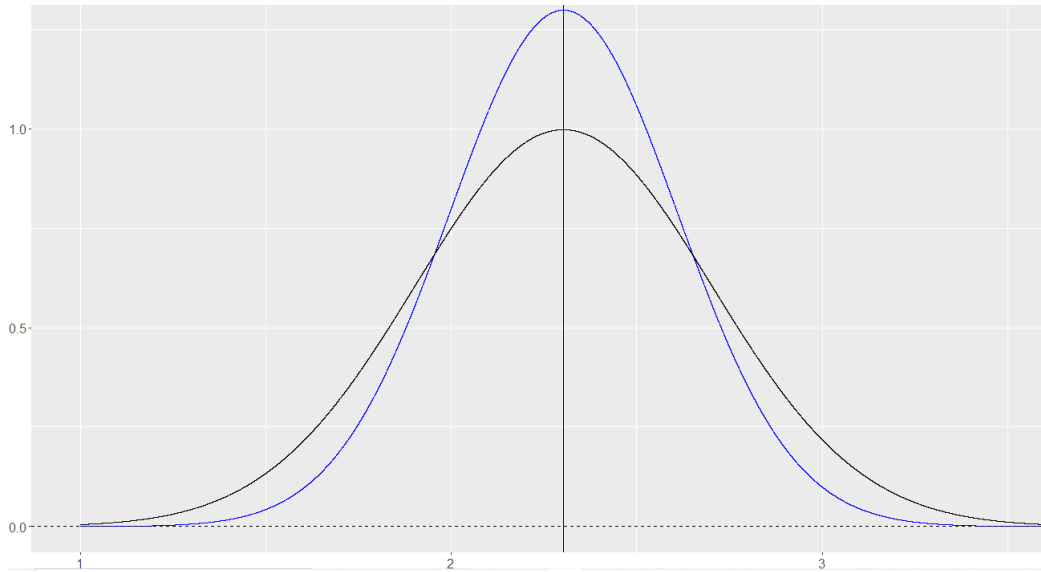


Figura 43 - Distribuição normal que ajustou  $\log(S_n)$ ,  $N(2,3; 0,09452)$ , e a distribuição de  $\ln S(t) \sim N(\ln S(0) + at, \sigma^2 t)$

A fim de verificar a aproximação do modelo proposto com movimento Browniano geométrico, à medida que  $P(+|-)$  e  $P(+|+)$  se aproximam de  $P(+)$ , foram realizadas 100.000 simulações para diferentes probabilidades condicionais. A Tabela 13 especifica as probabilidades utilizadas e os respectivos valores de  $\mu_d$  e  $\mu_u$  necessários em cada caso.

Tabela 13 -  $P(+|-), P(+|+), \mu_d$  e  $\mu_u$  utilizados nas simulações

$P(+ -)$	$P(+ +)$	$\mu_d$	$\mu_u$
0,80	0,20	0,84162	-0,84162
0,68	0,32	0,4677	-0,4677
0,65	0,35	0,38532	-0,38532
0,63	0,37	0,33185	-0,33185
0,60	0,40	0,25335	-0,25335
0,57	0,43	0,17637	-0,17637
0,55	0,45	0,12566	-0,12566
0,53	0,47	0,07527	-0,07527

A Figura 44 mostra as distribuições *log-normais* que ajustaram  $S_n$ , para cada simulação, e a distribuição *log-normal* de  $S_t$ , em  $t=1$ , fornecida pelo movimento Browniano geométrico. A linha vertical indica a média de  $S_t$ , dada por  $S(0)e^{\mu t} = 10,83$ ,  $t=1$ . Essa média foi maior do que a dos preços finais simulados de  $S_n$ .

Como esperado, as distribuições que ajustaram  $S_n$  são menos achatadas que a do movimento Browniano geométrico, em razão das maiores probabilidades de reversão dadas por  $P(+|-) > 1/2$  e  $P(+|+) < 1/2$  em todos os casos.

Percebe-se ainda uma aproximação das distribuições em relação ao movimento Browniano geométrico, à medida que  $P(+|-)$  e  $P(+|+)$  se aproximam de  $P(+)=1/2$ .

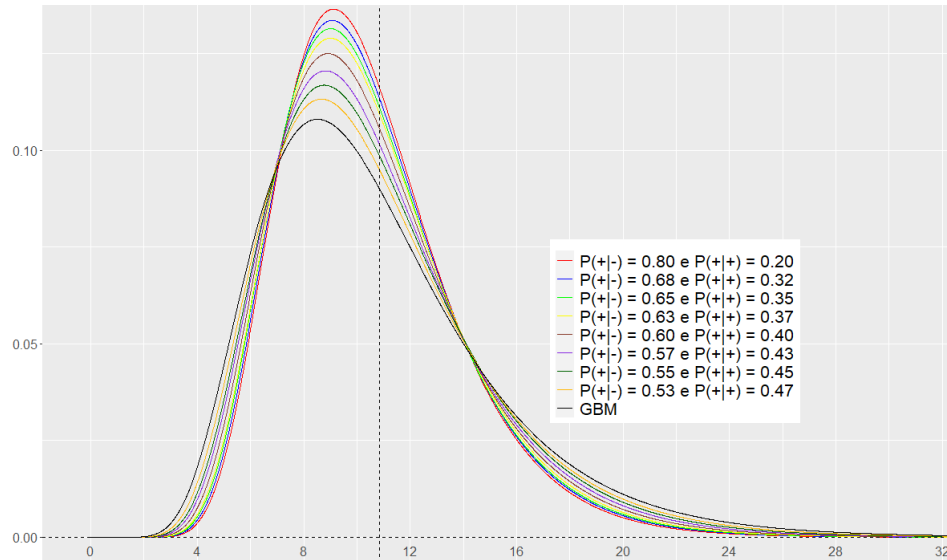


Figura 44 - Distribuições log-normais que ajustam  $S_n$  em cada simulação e a distribuição de  $S(t)$

A Figura 45 mostra as distribuições normais que ajustam  $\log(S_n)$  e a distribuição de  $\ln S(t) \sim N(\ln S(0) + at, \sigma^2 t)$ , em  $t=1$ . A linha vertical indica as médias, que foram todas iguais à fornecida pelo movimento Browniano geométrico,  $\ln S(0) + at = 2,3$ .

Mais uma vez, à medida que  $P(+|-)$  e  $P(+|+)$  se aproximam de  $P(+)$ , nota-se uma aproximação gradual das distribuições do novo modelo em relação ao movimento Browniano geométrico.



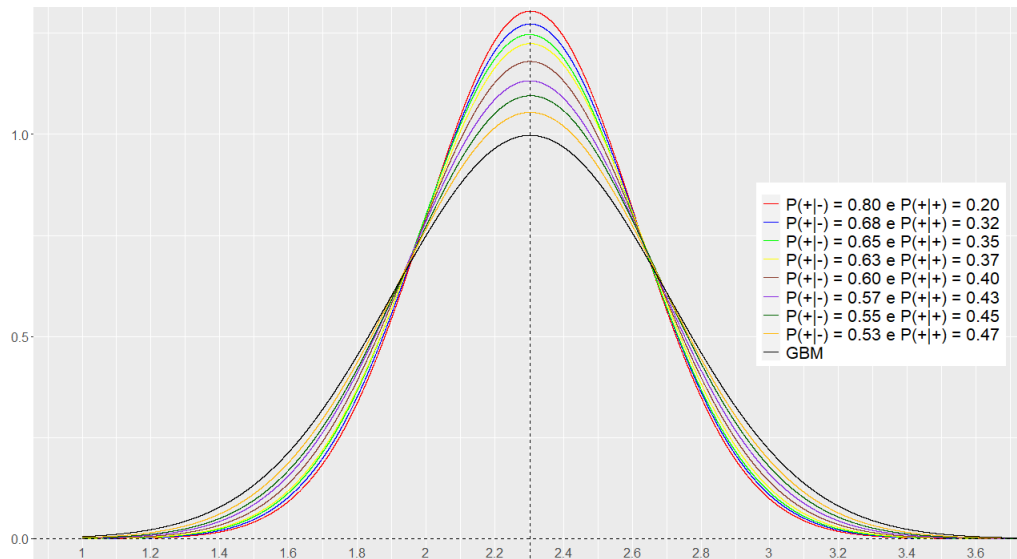


Figura 45 - Distribuições normais que ajustaram  $\log(S_n)$  em cada simulação e a distribuição de  $\ln S(t)$

## 8.2 ESTIMATIVA DA VOLATILIDADE $\sigma$ PARA OS DADOS DIÁRIOS

Na seção 3, observou-se que a volatilidade  $\sigma$  pode ser definida como o desvio padrão do retorno oferecido pela ação em 1 ano quando o retorno é expresso utilizando-se capitalização contínua.

No entanto, considerando o interesse na variação dos preços intradiários, é razoável, no âmbito da presente pesquisa, que a volatilidade  $\sigma$  seja expressa como o desvio padrão do retorno oferecido dentro de um dia somente.

Frisa-se que a volatilidade diária poderia ser obtida a partir da anual por meio da relação  $\sigma_D = \sigma_A \sqrt{1/252}$ , em que  $\sigma_D$  e  $\sigma_A$  são as respectivas volatilidades diária e anual e 252 é o valor considerado de dias úteis no ano. A partir daí, seria utilizado procedimento semelhante para calcular as volatilidades de cada período dentro do mesmo dia.

Entretanto, essa metodologia não traria resultados tão próximos da realidade. Por exemplo, a volatilidade para o período de 1 segundo seria, em última instância, aproximada a partir da volatilidade anual, não refletindo as movimentações mais recentes do ativo.

Para estimar a volatilidade diária, observaram-se os preços de fechamento do ativo a cada 5 minutos nos  $t=20$  dias anteriores. Considerando que, em cada dia, as

operações foram efetuadas das 10h às 17h, tem-se um total de observações diárias dado por  $n = 7.60/5 = 84$ .

Assim, definindo  $S_0$  como o valor da primeira negociação do dia e  $S_i$ ,  $i = 1, 2, \dots, n$ , como os preços médios a cada 5 minutos, calculou-se o *log-retorno* em cada etapa por:

$$u_i = \ln (S_i/S_{i-1})$$

Assim, a estimativa normal da volatilidade de cada dia  $d$  será dada por:

$$\hat{\sigma}_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (u_{id} - \bar{u}_d)^2}, \text{ em que } \bar{u}_d \text{ é a média de } u_{id}.$$

Finalmente, a volatilidade diária para um dia  $j$  qualquer, será estimada pela média de  $\hat{\sigma}_d$  nos  $r=20$  anteriores, ou seja:

$$\hat{\sigma} = \frac{1}{r} \sum_{d=j-r}^{j-1} \hat{\sigma}_d$$

### 8.3 APLICAÇÃO DO MODELO PARA OS DADOS REAIS

A aplicabilidade do modelo aos dados reais será demonstrada pela estimação da movimentação dos preços, considerando o intervalo de 5 segundos, para o 21º dia do pregão do ano de 2018, dia 31/01/2018.

Primeiramente, serão realizadas as estimações das probabilidades condicionais,  $P(+|-)$  e  $P(+|+)$ , e da volatilidade  $\hat{\sigma}$  do dia  $j=21$ , a partir dos dados dos  $r=20$  primeiros dias de negociação. Utilizou-se as seguintes formulações, já abordadas no decorrer desta seção 8:

$$P(+|-) = \frac{1}{r} \sum_{d=j-r}^{j-1} \frac{1}{1 + e^{-(\beta_{0d})}}$$

$$P(+|+) = \frac{1}{r} \sum_{d=j-r}^{j-1} \frac{1}{1 + e^{-(\beta_{0d} + \beta_{1d})}}$$

$$\hat{\sigma} = \frac{1}{r} \sum_{d=j-r}^{j-1} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (u_{id} - \bar{u}_d)^2}$$

Encontraram-se os seguintes resultados:  $P(+|-) = 82,83\%$ ,  $P(+|+) = 18,65\%$  e  $\hat{\sigma} = 0,001249$ . Complementarmente, foi calculado a média das proporções de movimentos positivos nos  $r=20$  primeiros dias de negociação, chegando-se a um resultado de  $P(+) = 50,46\%$ .

Tem-se ainda que  $\Delta t = 1/n$ , em que  $n = \text{total de segundos diários} / \text{duração do período}$ . Para um intervalo de 5 segundos,  $n = 7 * 60 * 60 / 5 = 5.040$  e  $\Delta t = 1,98 * 10^{-4}$ .

O valor de  $\mu$  é definido a partir da taxa de juros livre de risco capitalizada continuamente [18]. Estimou-se esse parâmetro por meio da taxa SELIC vigente em 31/01/2018, que era de 7% ao ano [21]. Portanto, o valor de  $\mu$ , para o 21º dia, foi estimado por  $\mu = 0,07/252 = 2,78 * 10^{-4}$ , considerando o ano com 252 dias úteis. Assim,  $a = \mu - \frac{1}{2}\sigma^2 = 2,78 * 10^{-4}$ .

Nota-se que, com esses parâmetros, a probabilidade de alta no movimento Browniano geométrico, dada pela a probabilidade de  $E_i \sim N(a\Delta t, \sigma^2 \Delta t)$  ser maior do que 0, será de  $P(+) = 50,12\%$ .

Para que  $P(E_D > 0) = P(+|-)$  e  $P(E_U > 0) = P(+|+)$ , os parâmetros  $\mu_d$  e  $\mu_u$  estimados foram de 0,9442 e  $-0,8941$ , respectivamente. E finalmente  $S_0$  foi considerado como sendo o valor da primeira negociação do dia 31/01/2018,  $S_0 = 19,74$ .

O modelo de precificação do ativo para o dia 31/01/2018 será então definido por:

$$S_i = \begin{cases} S_{i-1} \exp(a\Delta t + \hat{\sigma}Y\sqrt{\Delta t}), & Y \sim N(0,1), \quad i = 1 \\ S_{i-1} \exp(a\Delta t + \hat{\sigma}D\sqrt{\Delta t}), & \text{se } S_{i-1} < S_{i-2}, \quad D \sim N(\mu_d, 1), \quad i > 1 \\ S_{i-1} \exp(a\Delta t + \hat{\sigma}U\sqrt{\Delta t}), & \text{se } S_{i-1} \geq S_{i-2}, \quad U \sim N(\mu_u, 1), \quad i > 1 \end{cases}$$

Em que  $S_0 = 19,74$ ,  $a = 2,78 * 10^{-4}$ ,  $\mu_d = 0,9442$ ,  $\mu_u = -0,8941$ ,  $\hat{\sigma} = 0,001249$  e  $\Delta t = 1,98 * 10^{-4}$ .

Realizou-se um total de 40.000 simulações desse modelo. As médias das probabilidades condicionais encontradas foram  $P(+|-) = 82,83\%$  e  $P(+|+) = 18,65\%$ , exatamente iguais às esperadas.

A proporção de movimentos positivos gerados pelo novo modelo foi de  $P(+) = 50,45\%$ , valor também muito próximo dessa medida verificada nos  $r=20$  primeiros dias de negociação. Portanto, além de refletir bem as probabilidades condicionais, o

modelo fornece uma probabilidade de alta condizente com a verificada nos últimos dias de negociação.

Nota-se que essa probabilidade foi maior do que a do movimento Browniano geométrico, dada por  $P(+)=50,12\%$ . Verifica-se ainda que  $P(+)$  foi maior do que  $1/2$  justamente em função da assimetria existente entre  $P(+|-)$  e  $P(+|+)$ . Diferentemente das probabilidades propostas nos exemplos hipotéticos, essas medidas não são complementares entre si, ou seja,  $P(+|-) > 1 - P(+|+) = P(-|+)$  e  $P(+|+) > 1 - P(+|-) = P(-|-)$ .

A Figura 46 mostra o histograma e a densidade aproximada dos preços ao final do dia ( $S_n$ ). O logaritmo natural de  $S_n$  possui uma média de  $2,9844$  e variância de  $9,31 * 10^{-7}$ . A Figura 47 mostra a distribuição aproximada de  $\log(S_n)$ , em azul, e a densidade exata de uma variável normal com esses mesmos parâmetros, em preto. A proximidade entre as curvas evidencia a *log-normalidade* de  $S_n$ .

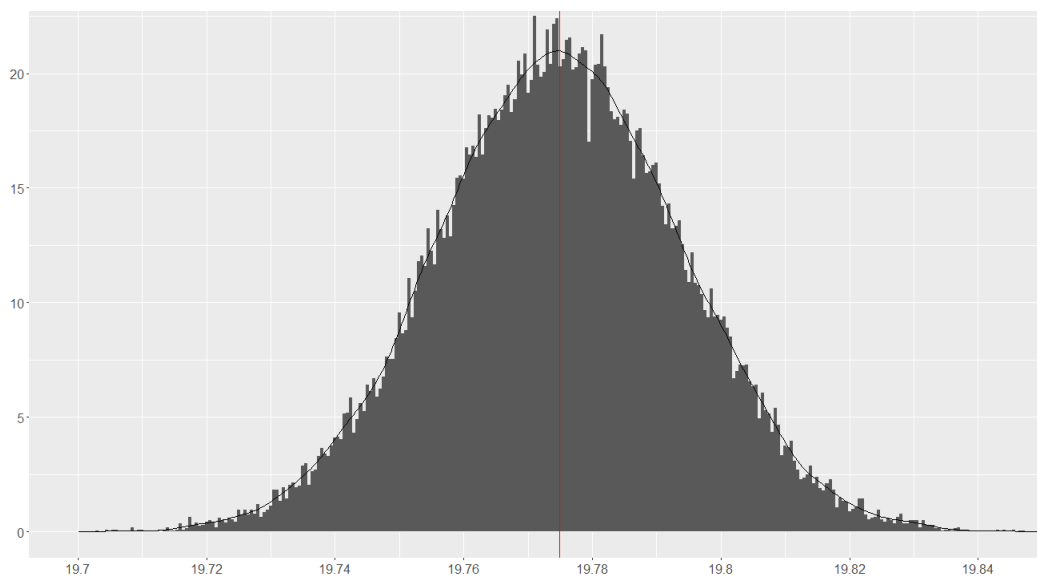


Figura 46 - Histograma e densidade aproximada de 40.000 simulações para o dia 31/01/2018.

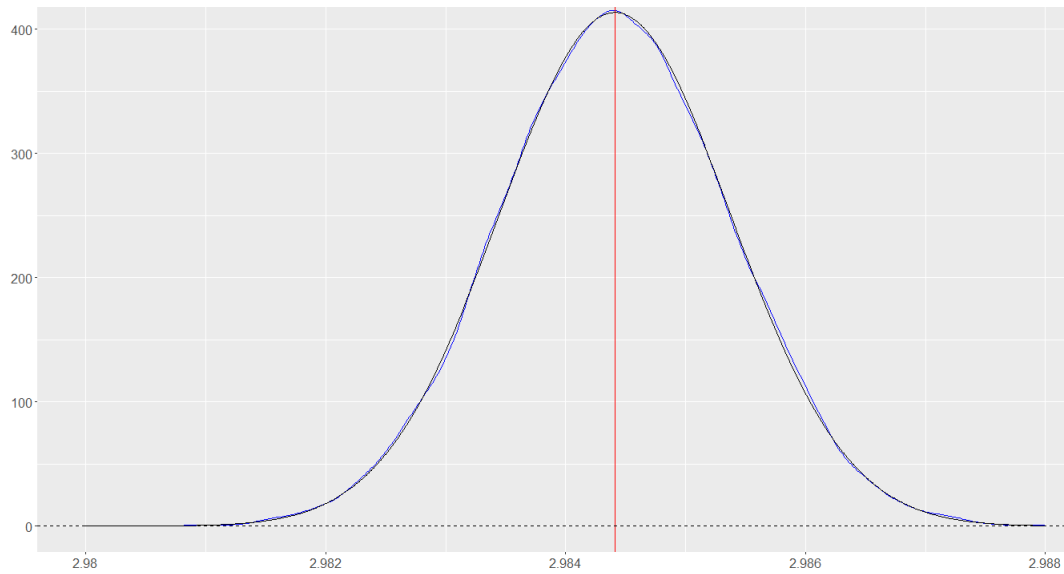


Figura 47 - Densidade de  $\log(S_n)$  e da curva normal ajustada pelos parâmetros da média a variância de  $\log(S_n)$

A Figura 48 ilustra, para os mesmos parâmetros, a densidade *log-normal* de um movimento Browniano geométrico em  $t = 1$  (curva em preto) e a densidade aproximada dos preços para o novo modelo (curva em vermelho). As linhas verticais indicam as respectivas médias. Vê-se que a nova densidade é menos achatada, em razão da maior probabilidade de reversão dos movimentos anteriores.

A média dos preços finais no novo modelo foi igual a 19,775, superior a do movimento Browniano geométrico, em  $t = 1$ , dada por  $S(0)e^{\mu t} = 19,745$ . Esse resultado foi condizente com o fato da probabilidade de alta ter sido maior do que a do movimento Browniano geométrico.

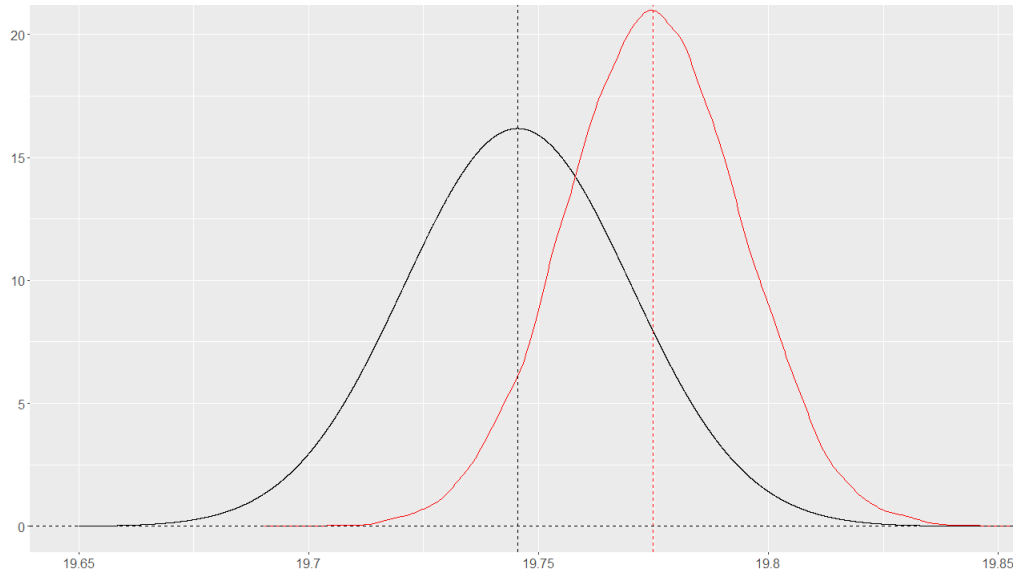


Figura 48 - Densidade de  $S_n$  e a distribuição *log-normal* do movimento Browniano geométrico.

Por último, a Figura 49 mostra a distribuição *log-normal* que ajustou  $S_n$ , em azul, e a distribuição *log-normal* de  $S_t$ , em  $t=1$ , em preto, fornecida pelo movimento Browniano geométrico. Ela é equivalente à Figura 48, substituindo a distribuição dos preços finais pela distribuição *log-normal* exata, ajustada pela média e variância de  $\log(S_n)$ .

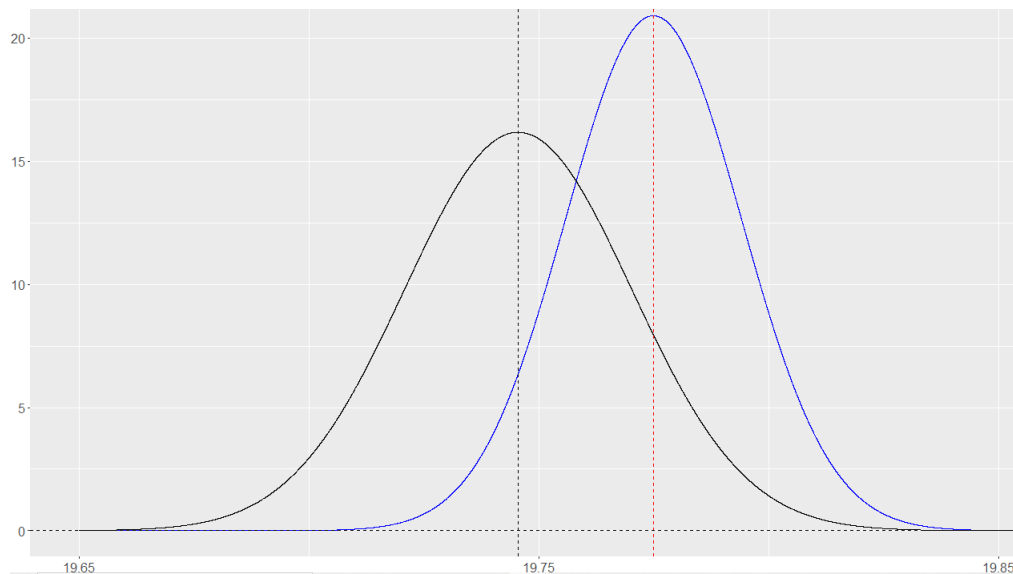


Figura 49 - Distribuição *log-normal* ajustada de  $S_n$  e a distribuição do movimento Browniano geométrico.

Foi verificado o comportamento do modelo em função de diferentes de intervalos analisados. A Tabela 14 mostra, para os intervalos de 10, 20, 30 e 45 segundos e para o intervalo de 3 minutos, as estimações das probabilidades condicionais do dia  $j=21$ , a partir dos  $r=20$  primeiros dias de negociação. Também estão sendo mostrados o valor de  $n = \text{total de segundos diários} / \text{duração do período}$ , o  $\Delta t$  utilizado e os respectivos valores de  $\mu_d$  e  $\mu_u$  necessários em cada caso.

Tabela 14 - Intervalo,  $n$ ,  $P(+|+)$ ,  $P(+|-)$ ,  $\mu_d$  e  $\mu_u$  utilizados nas simulações

intervalo	$n$	$\Delta t = 1/n$	$P(+ +)$	$P(+ -)$	$\mu_d$	$\mu_u$
10s	2520	0,0003968	0,24	0,78	-0,7076	0,7625
20s	1260	0,0007937	0,32	0,71	-0,4786	0,5451
30s	840	0,0011905	0,35	0,68	-0,3808	0,4498
45s	560	0,0017857	0,39	0,64	-0,2792	0,3477
3m	140	0,0071429	0,45	0,60	-0,1505	0,2261

Foram realizadas, desta vez, 100.000 simulações de cada modelo. A Figura 50 mostra as distribuições *log-normais* que ajustaram  $S_n$ , para cada intervalo, e a distribuição *log-normal* de  $S_t$ , em  $t=1$ , fornecida pelo movimento Browniano geométrico. A linha vertical indica a média de  $S_t$ , dada por  $S(0)e^{\mu t} = 19,745$ ,  $t=1$ .

Como esperado, as distribuições que ajustaram  $S_n$  são menos achatadas que a do movimento Browniano geométrico, em razão das maiores probabilidades de reversão dadas por  $P(+|-) > 1/2$  e  $P(+|+) < 1/2$  em todos os casos.

No entanto, diferentemente do que se verificou no exemplo hipotético, em todos os casos, as médias dos preços finais simulados de  $S_n$  foram maiores do que a do movimento Browniano geométrico ( $S(0)e^{\mu t} = 19,745$ ).

Isso ocorreu, pois as probabilidades condicionais dos modelos refletiram as proporções de movimentos positivos nos  $r=20$  primeiros dias de negociação, que foram maiores que a probabilidade de alta do movimento Browniano geométrico.

A Tabela 15 mostra as proporções de movimentos positivos observadas nos  $r=20$  primeiros dias de negociação e a probabilidade de alta no movimento Browniano geométrico, dada pela a probabilidade de  $E_i \sim N(a\Delta t, \sigma^2 \Delta t)$  se maior do que zero.

Tabela 15 -  $P(+)$  - GBM e  $P(+)$  - 20 dias anteriores

intervalo	$P(+)$ - GBM	$P(+)$ - 20 dias anteriores
10s	0,5018	0,5065
20s	0,5025	0,5100
30s	0,5031	0,5119
45s	0,5037	0,5130
3m	0,5075	0,5209

Vê-se ainda uma aproximação gradual das distribuições em relação à densidade do movimento Browniano geométrico, à medida que  $P(+|-)$  e  $P(+|+)$  se aproximam do  $P(+)$  calculado para cada intervalo desse movimento.

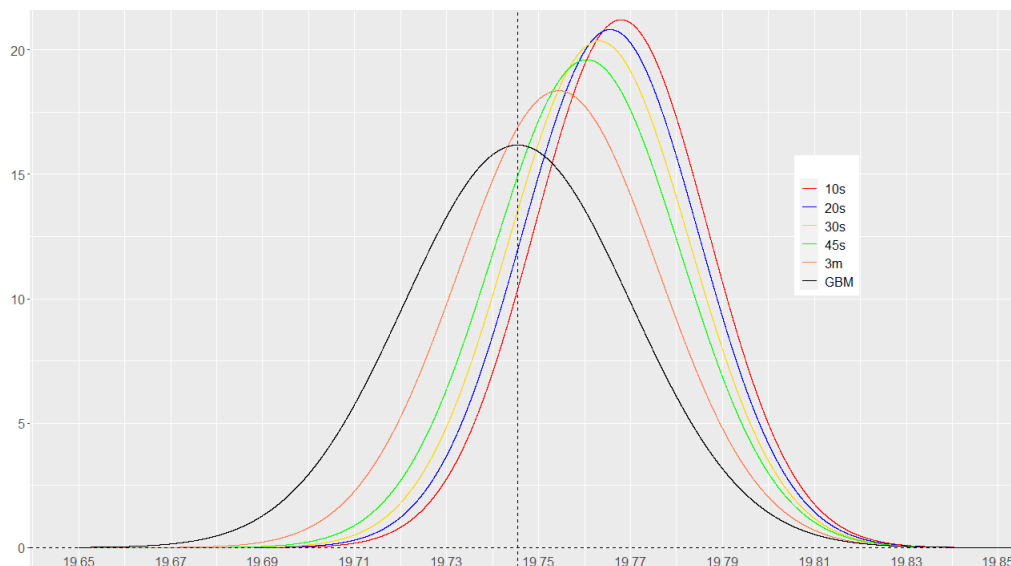


Figura 50 - Distribuições log-normais que ajustaram  $S_n$  em cada simulação e a distribuição de  $S(t)$

#### 8.4 UTILIZAÇÃO DO NOVO MODELO PARA A PRECIFICAÇÃO DE OPÇÕES

Viu-se, na seção 4, que as opções podem ser adequadamente precificadas a partir da distribuição dos preços dos ativos subjacentes na data de vencimento desses derivativos.

Neste tópico, a distribuição *log-normal* que ajustou a movimentação do ativo no intervalo de 5 segundos, denominada  $f(S_n)$ , será utilizada para obter uma estimativa mais precisa dos preços dessas opções no final do dia.

Em toda análise, serão consideradas hipoteticamente duas opções, *call* e *put*, com preços de exercício ao final do dia de  $K_{call} = 19,78$  e  $K_{put} = 19,74$ .



Conforme abordado na seção 4, a probabilidade de execução de uma *call* ou *put* será dada pela probabilidade do preço do ativo, no exercício, ser maior ou menor do que os *strikes* respectivos.

A Figura 51 ilustra essas probabilidades por meio das áreas sob as densidades da nova distribuição e do movimento Browniano geométrico. A área em azul indica a probabilidade de execução da opção de compra, enquanto a em vermelho indica a probabilidade de execução da opção de venda.

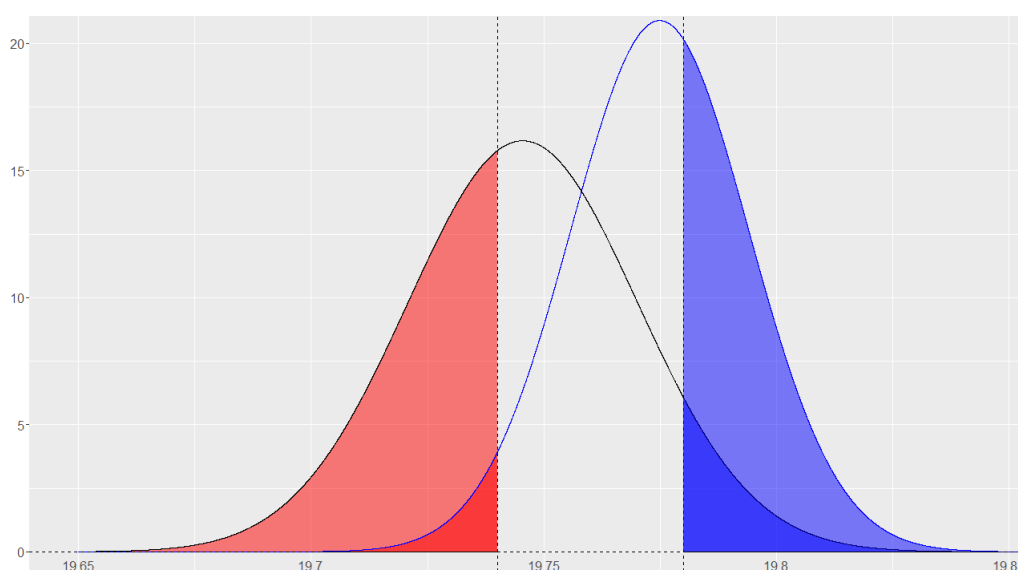


Figura 51 - Densidade de  $S_n$  e a distribuição log-normal do movimento Browniano geométrico.

É possível ver que as probabilidades de execução da *call* e *put* no novo modelo são maior e menor, respectivamente, do que as fornecidas pelo movimento Browniano geométrico.

A partir da curva de  $f(S_n)$ , calculam-se essas probabilidades por:

$$P(S_n > K) = \int_K^{\infty} f(s)ds \quad \text{e} \quad P(S_n < K) = \int_0^K f(s)ds$$

Encontraram-se  $P(S_n > K_{call}) = 39,45\%$  e  $P(S_n < K_{put}) = 3,36\%$ , enquanto, para o movimento Browniano geométrico, essas probabilidades, obtidas diretamente da distribuição log-normal de  $S(t)$ , em  $t=1$ , foram 8,09% e 41,22%, respectivamente.

Viu-se também que os *payoffs ponderados pela probabilidade* das opções são calculados a partir do produto dos *payoffs* pela densidade dos preços no vencimento. A Figura 52 mostra, de forma consolidada, a densidade do novo modelo, a densidade

do movimento Browniano geométrico, os *payoffs* e os *payoffs ponderados pela probabilidade* de cada opção.

Nota-se que os *payoffs ponderados* da nova distribuição, em relação aos do movimento Browniano geométrico, foram superiores para a *call* e inferiores para a *put*.

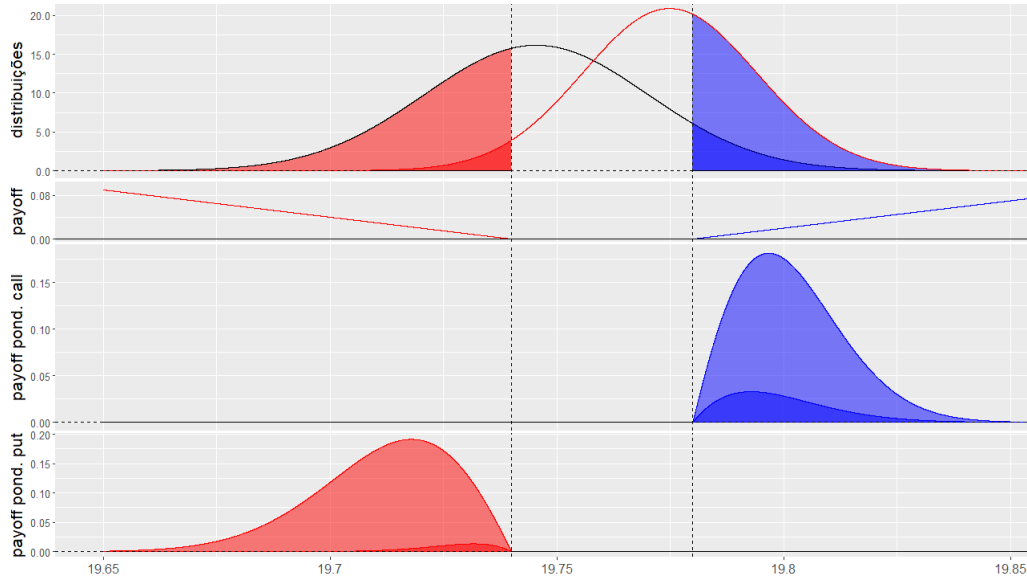


Figura 52 - Densidades, *payoffs* e os *payoffs ponderados pela probabilidade* da *call* e da *put*.

Os valores esperados da *call* e da *put*,  $V_{call}$  e  $V_{put}$ , serão calculados pelos valores esperados dos respectivos *payoffs*. Graficamente, eles serão as áreas sob os *payoffs ponderados pela probabilidade*.

Pela Figura 52, vê-se que, em relação aos obtidos pelo movimento Browniano geométrico,  $V_{call}$  e  $V_{put}$  serão superiores para a *call* e inferiores para a *put*. Eles serão calculados por:

$$V_{call} = \int_{-\infty}^{\infty} g_{call}(s) \cdot f(s) ds = \int_{K_{call}}^{\infty} w_{call}(s) ds$$

$$V_{put} = \int_{-\infty}^{\infty} g_{put}(s) \cdot f(s) ds = \int_0^{K_{put}} w_{put}(s) ds$$

Finalmente, descontando  $V_{call}$  e  $V_{put}$  pelo retorno esperado  $\mu$ , capitalizado continuamente, os preços das opções  $P_{call}$  e  $P_{put}$ , em  $t = 0$ , serão dados por:

$$P_{call} = e^{-\mu n \Delta t} \cdot V_{call} \quad \text{e} \quad P_{put} = e^{-\mu n \Delta t} \cdot V_{put}$$

Para o novo modelo, foram encontrados  $P_{call} = 5,33 \cdot 10^{-3}$  e  $P_{put} = 2,52 \cdot 10^{-4}$ .

Para o movimento Browniano geométrico, os preços das opções podem ser obtidos de forma semelhante, integrando os *payoffs ponderados pela probabilidade* e descontando o resultado pelo retorno esperado. Entretanto, conforme visto na seção 4, esses preços podem ser calculados diretamente pelas fórmulas de Black - Scholes - Merton.

A aplicação dessas fórmulas forneceu preços para a *call* e *put* de  $0,9084 * 10^{-3}$  e  $73,36.10^{-4}$ , inferior e superior, respectivamente, aos do novo modelo. Portanto, o modelo proposto, além de refletir melhor as movimentações mais recentes do ativo, forneceu estimativas diferentes para os preços das opções de mercado.

Uma observação importante a ser feita é que, a depender do intervalo utilizado, os preços obtidos para as opções de mercado serão diferentes entre si. Entretanto, a despeito disso, os preços da *call* e da *put* serão sempre superiores e inferiores, respectivamente, aos valores calculados a partir do movimento Browniano geométrico.

O aumento gradual dos intervalos aproxima cada vez mais os preços da *call* e da *put* daqueles calculados pelas fórmulas de Black - Scholes - Merton.

## 9. CONCLUSÃO

O presente trabalho estudou a teoria do passeio aleatório e a eficiência do mercado para dados intradiários de ações negociadas na B3. Viu-se que, se a movimentação dos preços realmente seguisse um passeio aleatório, a probabilidade de um movimento para cima deveria ser semelhante à probabilidade de um movimento para cima, dado as informações a priori.

Em todos os intervalos estudados, a estimativa de  $P(+)$  sempre foi próxima dos 50%. Entretanto, verificou-se que  $P(+|-)$ , foi, em geral, maior do que 50%, sobretudo para os menores intervalos. Para os dados *tick a tick*, por exemplo, a estimativa da probabilidade de alta, dado um último movimento negativo ficou acima dos 80%, aproximando-se gradativamente dos 50% a medida que os intervalos foram sendo incrementados.

Por sua vez, a estimativa da probabilidade de alta, dado um último movimento positivo,  $P(+|+)$ , foi normalmente inferior aos 50%. Para os dados *tick a tick*, essa estimativa ficou abaixo dos 20%, aproximando-se gradativamente dos 50% a medida que os intervalos foram sendo incrementados.

Foram então realizados testes estatísticos Z, para comparação das proporções de  $P(+)$  com as de  $P(+|-)$  e  $P(+|+)$ . Para os intervalos *tick a tick* e de até 30 segundos, os resultados rejeitaram a igualdade dessas proporções na maioria dos dias. Repetiu-se os testes para todos os dias conjuntamente, concluindo-se pela rejeição da independência dos movimentos do ativo em todos os intervalos.

Concluiu-se que o mercado é potencialmente ineficiente no curto prazo e que informações a priori impactam a movimentação dos preços para intervalos de até 5 minutos. Deduziu-se ainda que, até esse intervalo, o mercado apresenta tendências reversivas e a hipótese do passeio aleatório não é satisfeita, não podendo ser utilizada para descrever de forma fidedigna os movimentos do ativo.

A partir dessa verificação, foram propostos modelos de regressão logística para estimar a probabilidade de alta/baixa de cada movimento em função dos movimentos passados. Em todos os modelos, confirmou-se a tendência de reversão do sentido do último movimento observado.

Verificou-se ainda a capacidade preditiva dos modelos regressivos. Para isso, cada um deles foi utilizado para se prever os movimentos do ativo no dia seguinte. Os

resultados obtidos foram muitos satisfatórios, pois eles foram capazes de antecipar o sentido de cada movimentação com uma probabilidade de sucesso acima de 50%.

Propôs-se então uma estratégia de operação em alta frequência que buscasse obter lucros financeiros a partir dos modelos para um movimento anterior somente. Considerou-se que as operações foram realizadas pelos preços de fechamento de cada intervalo. A estratégia mostrou-se muito eficaz ao se desconsiderar os custos operacionais, chegando, por exemplo, a obter lucros diários acima dos 100% para o intervalo de 1 segundo.

Entretanto, ao inserir os custos operacionais, os lucros verificados foram completamente absorvidos por eles, de forma que os resultados diários se tornaram negativos na maioria dos dias de negociação. Nota-se que os únicos custos considerados foram os cobrados obrigatoriamente pela B3, utilizados para remunerar os serviços prestados por ela.

Concluiu-se que esse resultado decorreu da necessidade de se ter uma diferença mínima entre os preços de entrada e saída para a lucratividade dos negócios e pelo fato dos custos operacionais incidirem, inclusive, sobre as operações que tiveram prejuízo. Assim, apesar da movimentação dos preços não seguir um processo do tipo passeio aleatório, a estratégia proposta não foi suficiente para superar a barreira do menor custo operacional existente atualmente.

Após, apresentou-se um modelo de precificação de ativos que buscasse incorporar os resultados da pesquisa. Para tanto, propôs-se uma modificação na versão em tempo discreto do movimento Browniano geométrico, de forma que ele pudesse refletir as probabilidades condicionais verificadas nos 20 dias anteriores de negociação.

O novo modelo forneceu resultados computacionais bastantes coerentes, incorporando, concomitantemente, a aleatoriedade observada no movimento Browniano geométrico, as probabilidades condicionais fornecidas pelos modelos logísticos de regressão, a proporção de movimentos positivos verificadas nos 20 dias anteriores de negociação e a aproximação gradual a um processo do tipo passeio aleatório a medida que os intervalos foram sendo incrementados (e portanto o mercado foi se tornando cada vez mais eficiente).

A partir dele, foram realizadas estimações mais precisas dos preços de opções de compra e venda europeias, *call* e *put*, dentro de um mesmo dia. Viu-se que, a

dependem do intervalo utilizado, os preços obtidos para as opções de mercado serão diferentes entre si. Entretanto, foi concluído que, em todos os intervalos considerados, os preços da *call* e *put* serão sempre superiores e inferiores, respectivamente, a esses valores calculados a partir do movimento Browniano geométrico.

Concluiu-se ainda que o aumento gradual dos intervalos aproxima cada vez mais os preços das opções àqueles calculados pelas fórmulas de Black - Scholes – Merton, que têm como pressuposto o comportamento eficiente do mercado de ações.

## REFERÊNCIAS

- [1] Dimson, E. and Mussavian, M. (1998), A brief history of market efficiency. *European Financial Management*, 4: 91-103.
- [2] Eugene F. Fama (1995), Random Walks in Stock Market Prices. *Financial Analysts Journal*, 51.
- [3] Brogaard, Jonathan; Carrion, Allen; Moyaert, Thibaut; Riordan, Ryan; Shkilko, Andriy; Sokolov, Konstantin (2018), High frequency trading and extreme price movements. *Journal of Financial Economics* 1–13
- [4] Jain, Pankaj; Jain, Pawan; McInish, Thomas H. (2016), Does high-frequency trading increase systemic risk? *Journal of Financial Markets*, 4: 91-103.
- [5] Menkveld, Albert J. (2013), High frequency trading and the new market makers. *Journal of Financial Markets*, 16: 712--740.
- [6] Conrad, Jennifer; Wahal, Sunil; Xiang, Jin. (2015), High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics*, 116: 271-291.
- [7] Rechenthin, Michael; Street, W. Nick. (2013) Using conditional probability to identify trends in intra-day high-frequency equity pricing. *Physica A: Statistical Mechanics and its Applications*, 392: 6169-6188.
- [8] Toru Ohira, Naoya Sazuka, Kouhei Marumo, Tokiko Shimizu, Misako Takayasu, Hideki Takayasu(2002), Predictability of currency market exchange, *Physica A: Statistical Mechanics and its Applications* 308: 368–374.
- [9] M. Tanaka-Yamawaki(2003), Stability of Markovian structure observed in high frequency foreign exchange data, *Annals of the Institute of Statistical Mathematics* 55: 437–446.
- [10] Mark H.A. Davis and Alison Etheridge, "Louis Bachelier's Theory of Speculation: the Origins of Modern Finance", Princeton University Press 2006
- [11] Ross, Sheldon M. (2014). *Introduction to Probability Models* (11th ed.). Amsterdam: Elsevier. pp. 612–14.
- [12] M.D Godfrey, C.W.J. Granger, and O. Morgenstern, "The Random Walk Hypothesis of Stock Market Behavior" *Kyklos*, vol 17 (Janeiro 1964):1-30.
- [13] Alexander, "Price Movements in Speculative Markets: Trends or Random Walks, Number 2."

- [14] Mikosch, Thomas , "Elementary Stochastic Calculus with Finance in View", World Scientific, 1998.
- [15] Révész, Pál, "Random Walk in Randon and Non-Random Environments", World Scientific, 1990.
- [16] Gregory, Lawer; Limic, Vlada, "Random Walk: A modern introduction", Cambridge studies in advanced mathematics, 2010.
- [17] Lo, Andrew; Mackinlay, A. Craig, "A non-randow walk Down Wall Street", Princeton University Press, 1999.
- [18] Hull, John C, "Opções, Futuros e Outros Derivativos", Bookman, 2016.
- [19] [http://www.b3.com.br/pt\\_br/produtos-e-servicos/tarifas](http://www.b3.com.br/pt_br/produtos-e-servicos/tarifas)
- [20] <https://www.up2dataondemand.com.br/>
- [21] <https://www.bcb.gov.br/controleinflacao/historicotaxasjuros>



## APÊNDICE

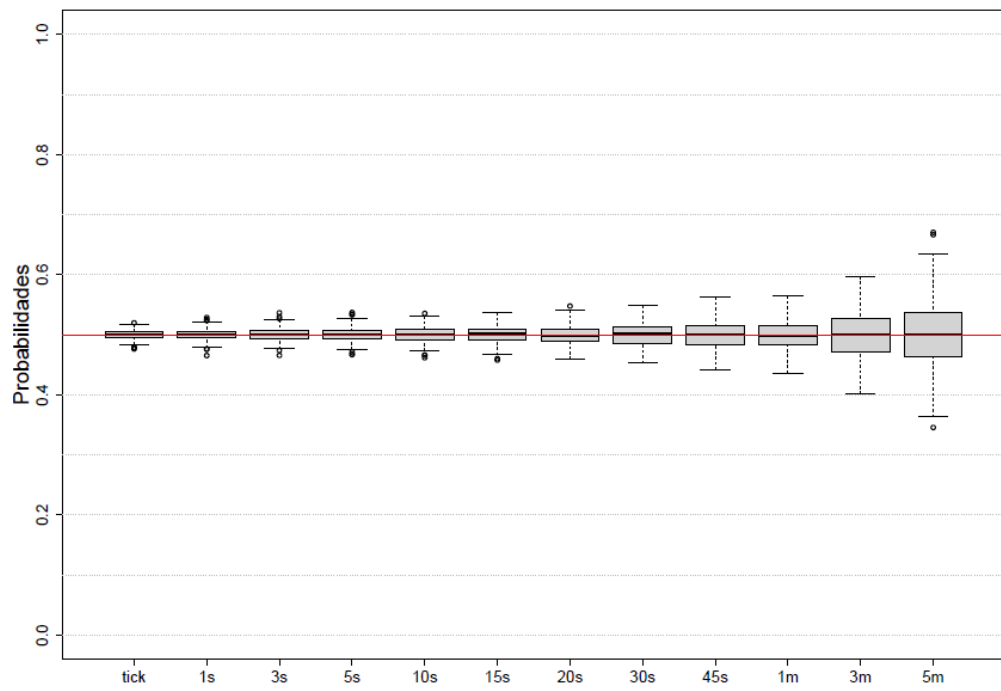


Figura A 1 - Boxplot de  $P(+)$  para diferentes períodos (ABEV3).

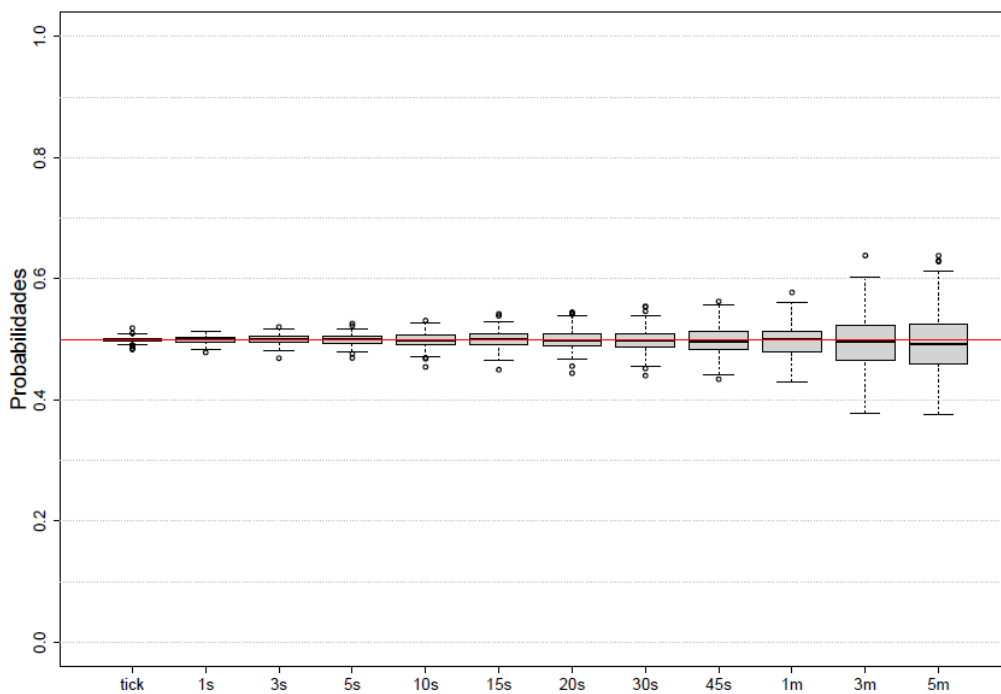


Figura A 2 - Boxplot de  $P(+)$  para diferentes períodos (BBDC4).

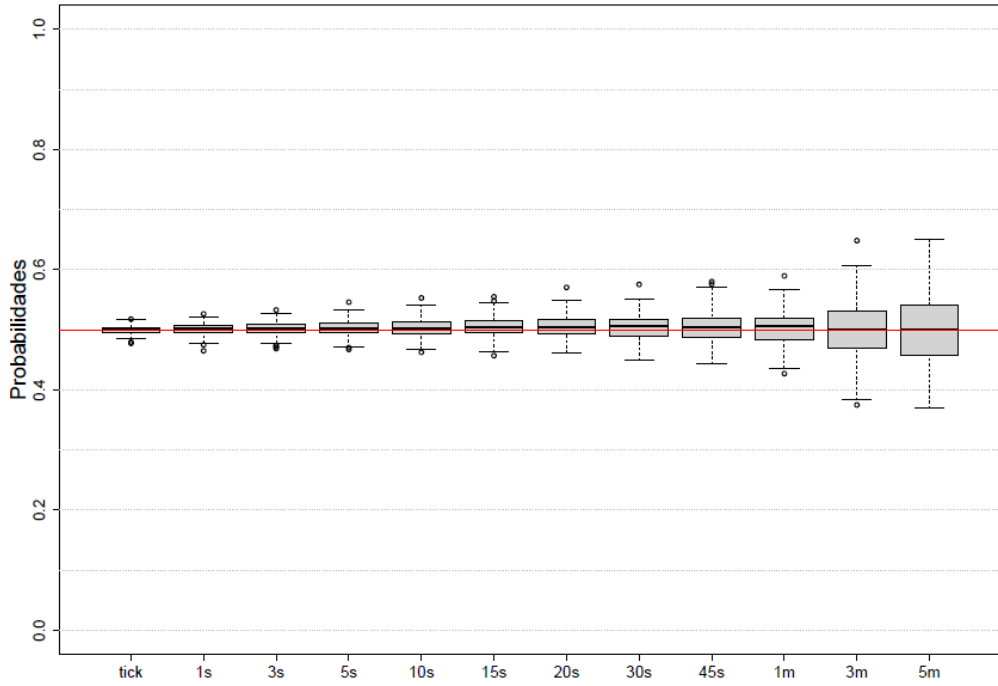


Figura A 3 - Boxplot de  $P(+)$  para diferentes períodos (VALE3).

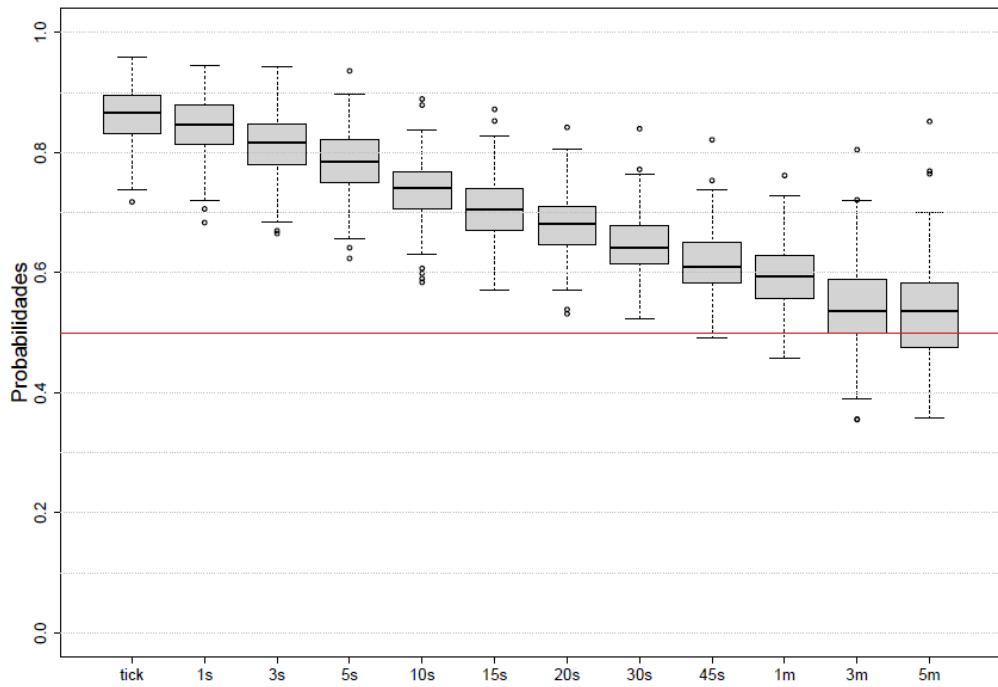


Figura A 4 - Boxplot de  $P(+/-)$  para diferentes períodos (ABEV3)

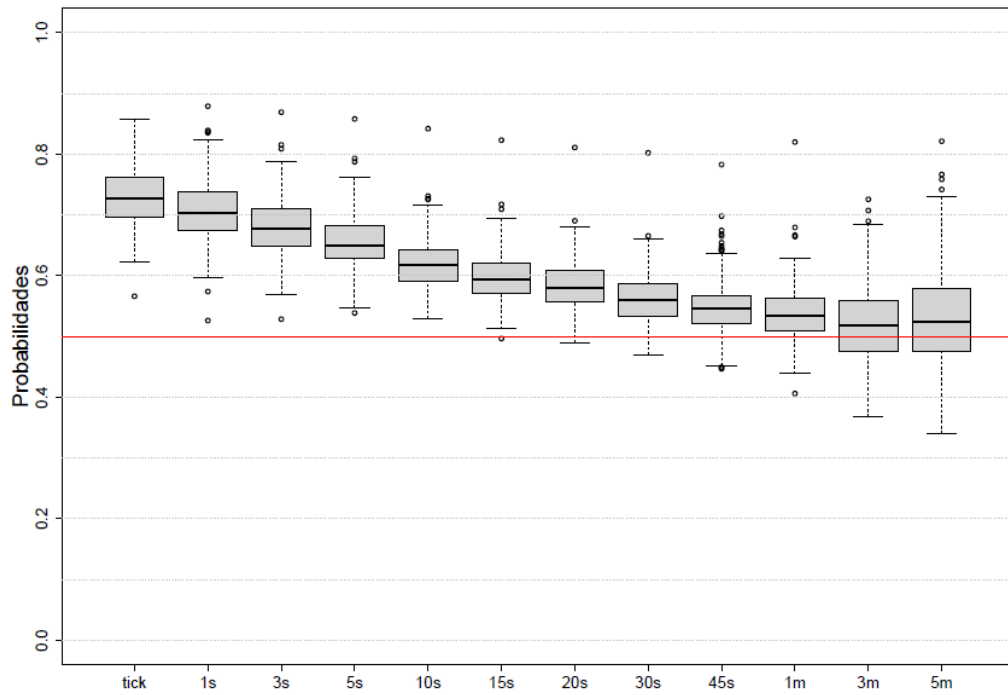


Figura A 5 - Boxplot de  $P(+/-)$  para diferentes períodos (BBDC4)

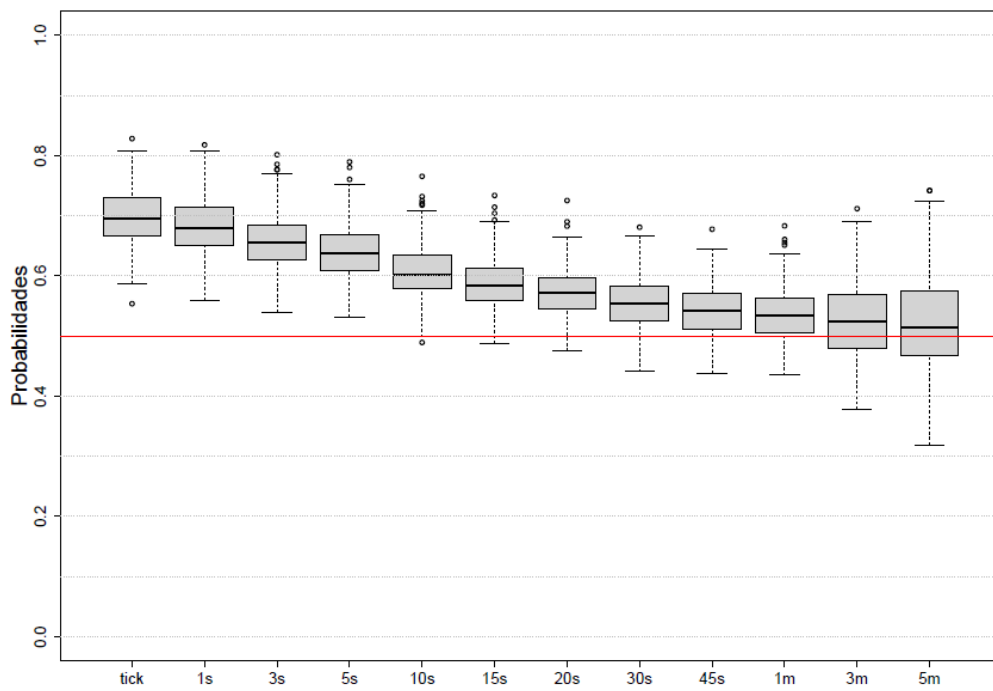


Figura A 6 - Boxplot de  $P(+/-)$  para diferentes períodos (VALE3)

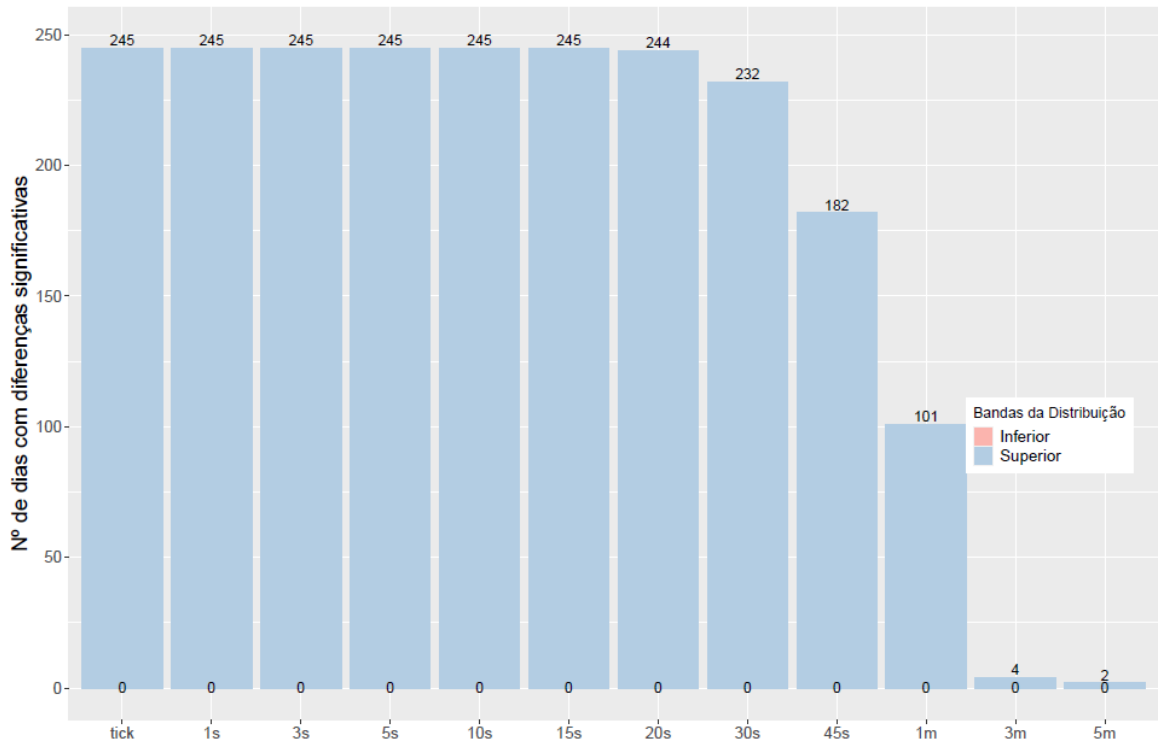


Figura A 7 - Número de dias estatisticamente significativos (Teste 1) - ABEV3

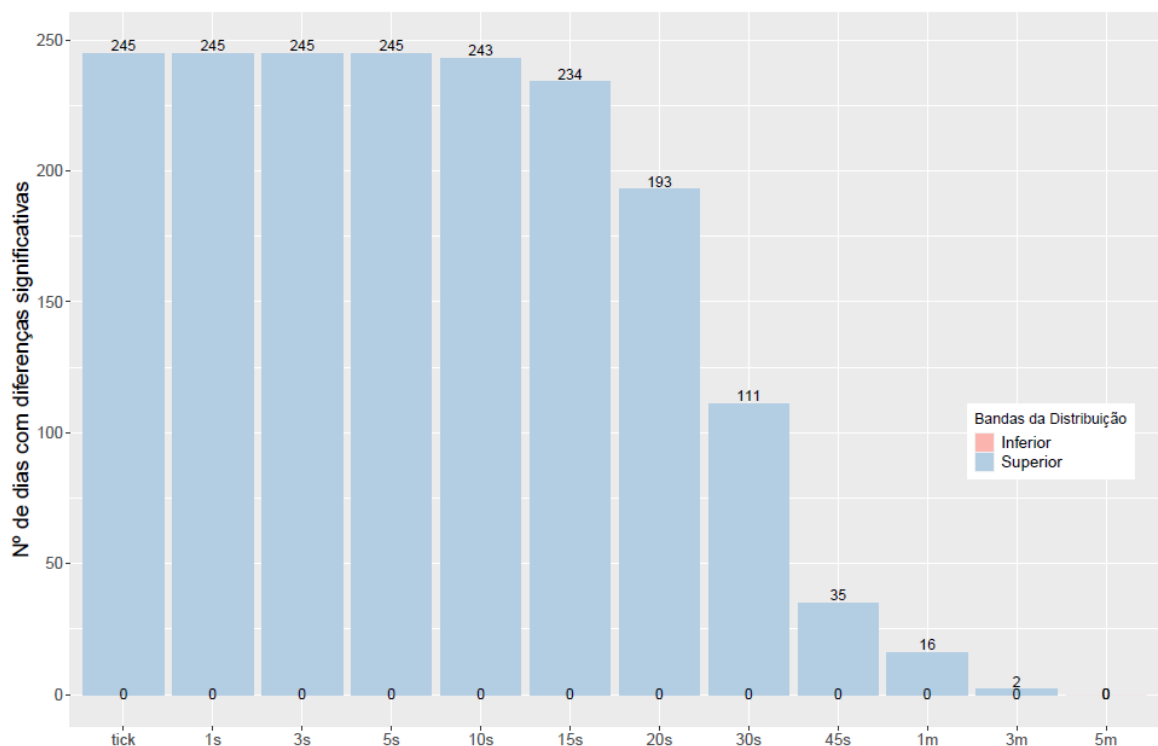


Figura A 8 - Número de dias estatisticamente significativos (Teste 1) - BBDC4

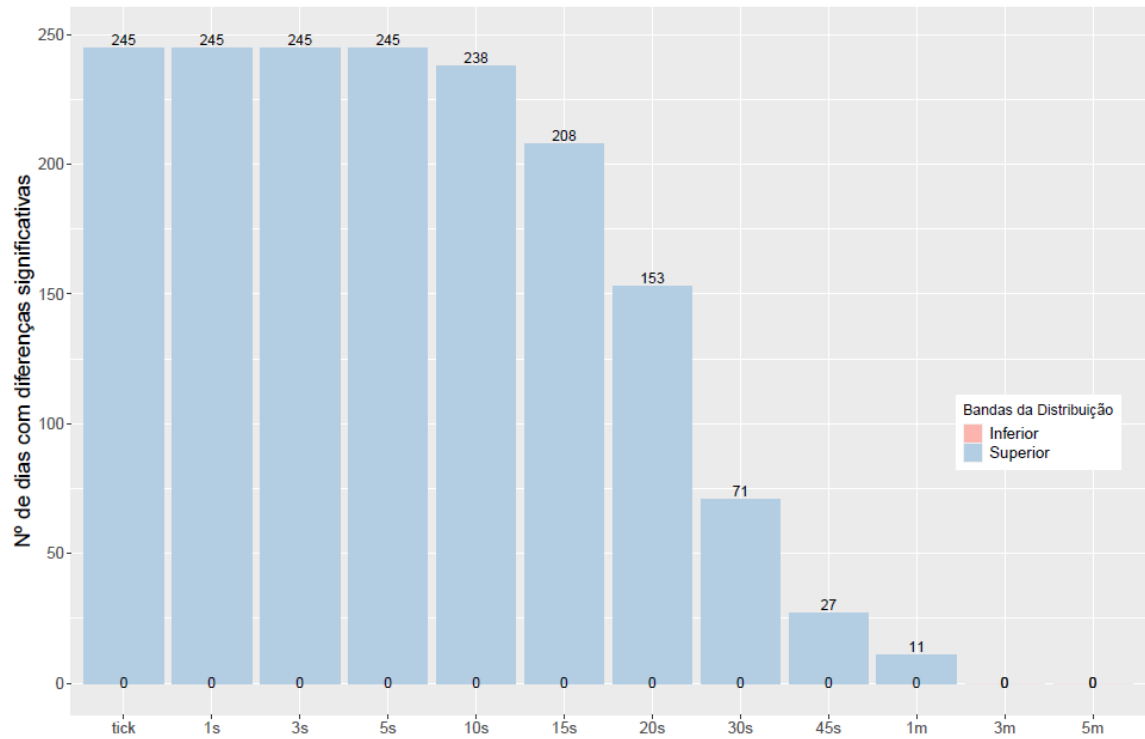


Figura A 9 - Número de dias estatisticamente significativos (Teste 1) - VALE3

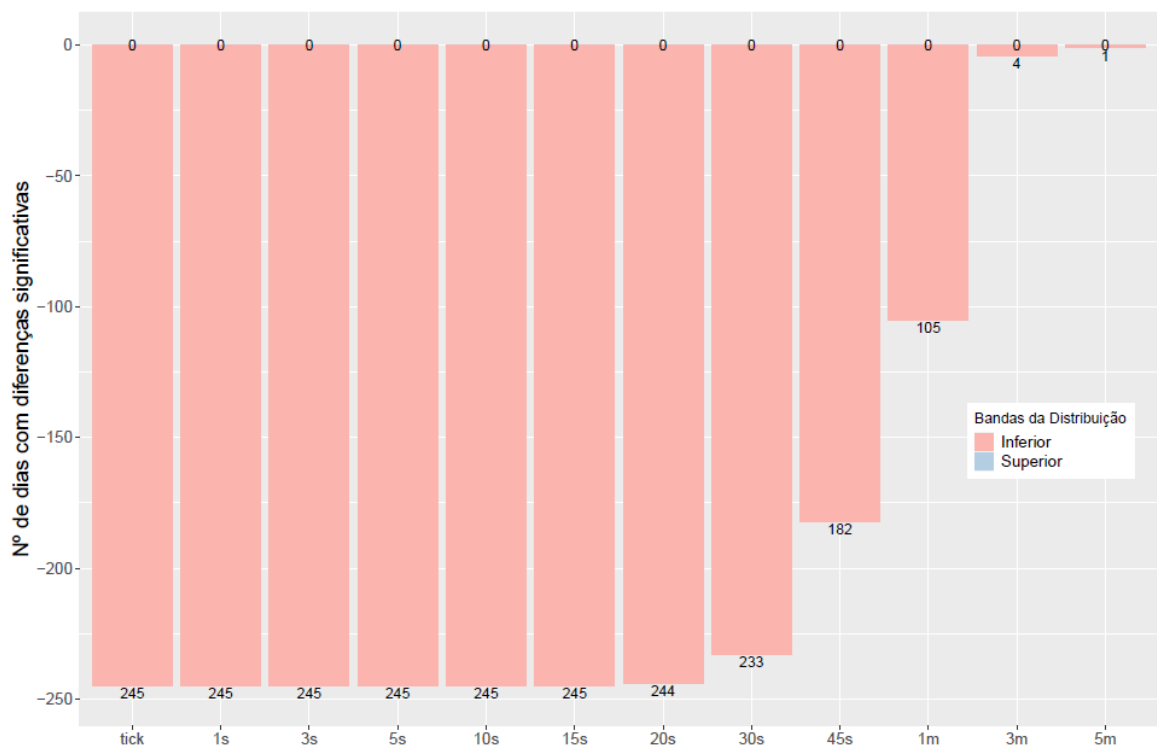


Figura A 10 - Número de dias estatisticamente significativos. (Teste 2) - ABEV3

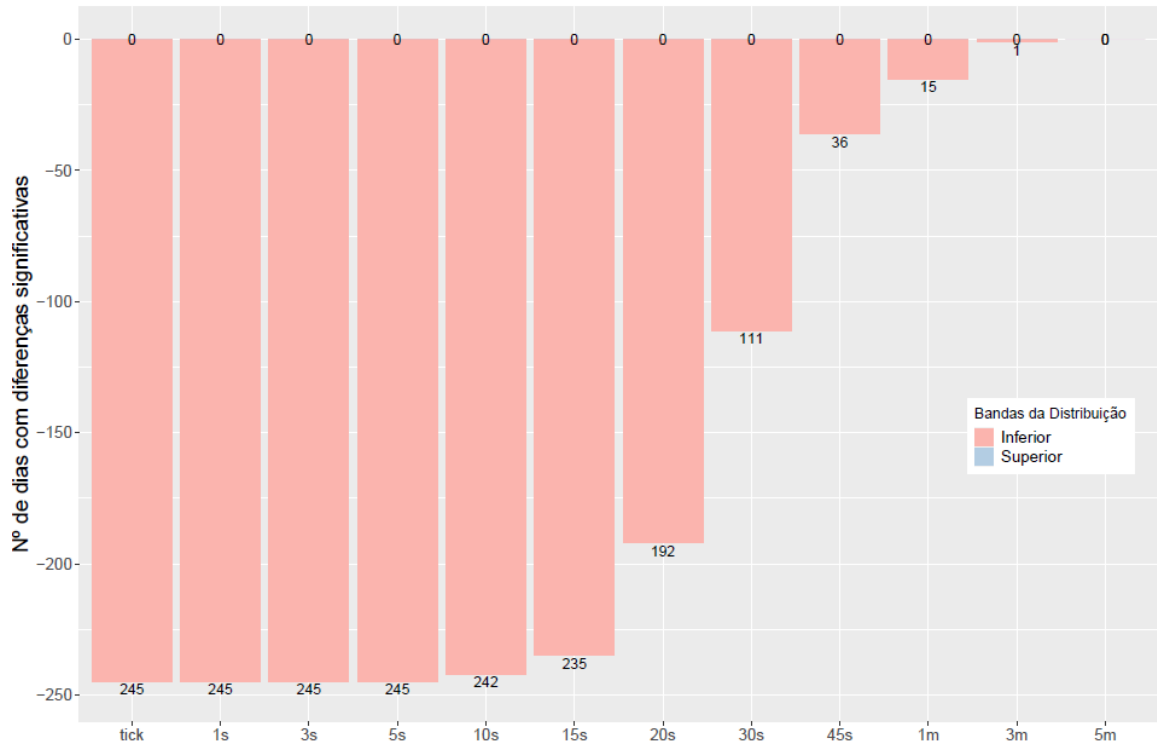


Figura A 11 - Número de dias estatisticamente significativos. (Teste 2) - BBDC4

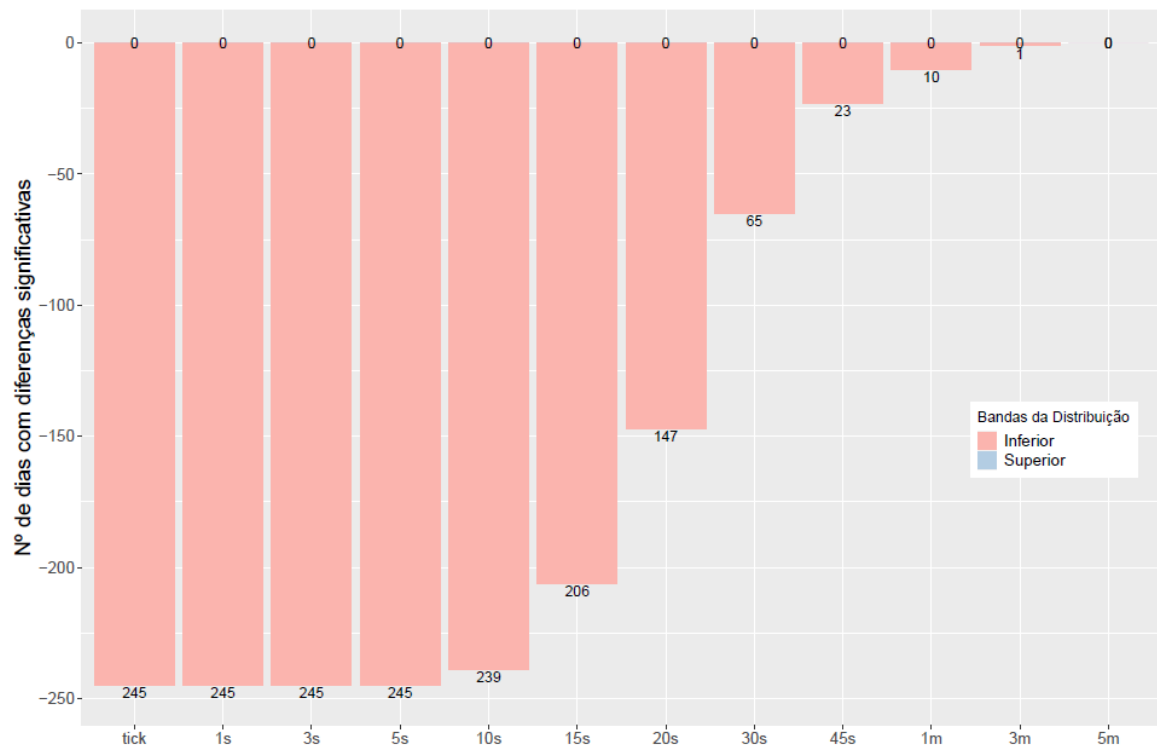


Figura A 12 - Número de dias estatisticamente significativos. (Teste 2) - VALE3

Tabela A 1 - Resultado do Teste 1 para todos os dias agrupados (ABEV3)

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
tick	557,81	0
1s	386,89	0
3s	280,72	0
5s	220,98	0
10s	145,66	0
15s	107,75	0
20s	84,43	0
30s	58,17	0
45s	39,26	0
1m	28,54	0
3m	8,29	0
5m	5,34	4,68E-08

Tabela A 2 - Resultado do Teste 1 para todos os dias agrupados (BBDC4)

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
Período	387,91	0
1s	252,31	0
3s	175,93	0
5s	135,16	0
10s	83,65	0
15s	58,30	0
20s	43,94	0
30s	27,78	0
45s	17,74	0
1m	11,96	0
3m	4,25	1,06E-05
5m	3,70	1,07E-04

Tabela A 3 - Resultado do Teste 1 para todos os dias agrupados (VALE3)

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
tick	376,99	0
1s	241,71	0
3s	167,48	0
5s	127,53	0
10s	75,60	0
15s	51,28	0
20s	36,75	0
30s	23,15	0
45s	14,05	0
1m	10,28	0
3m	4,17	1,50E-05
5m	2,20	1,39E-02

Tabela A 4 - Resultado do Teste 2 para todos os dias agrupado (ABEV3).

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
<i>tick</i>	-558,37	0
1s	-387,81	0
3s	-281,20	0
5s	-221,49	0
10s	-146,05	0
15s	-108,03	0
20s	-84,64	0
30s	-58,53	0
45s	-39,51	0
1m	-28,81	8,76E-183
3m	-8,49	1,06E-17
5m	-5,53	1,58E-08

Tabela A 5 - Resultado do Teste 2 para todos os dias agrupado (BBDC4).

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
<i>tick</i>	-387,56	0
1s	-252,16	0
3s	-175,86	0
5s	-135,13	0
10s	-83,71	0
15s	-58,26	0
20s	-43,93	0
30s	-27,81	1,5E-170
45s	-17,80	3,31E-71
1m	-12,02	1,42E-33
3m	-4,49	3,64E-06
5m	-3,83	6,40E-05

Tabela A 6 - Resultado do Teste 2 para todos os dias agrupado (VALE3).

<b>Período</b>	<b>z-valor</b>	<b>p-valor</b>
<i>tick</i>	-376,84	0
1s	-240,69	0
3s	-166,41	0
5s	-126,43	0
10s	-74,86	0
15s	-50,69	0
20s	-36,29	1,3E-288
30s	-22,93	1,07E-116
45s	-13,95	1,56E-44
1m	-10,22	7,75E-25
3m	-4,38	6,07E-06
5m	-2,37	8,80E-03



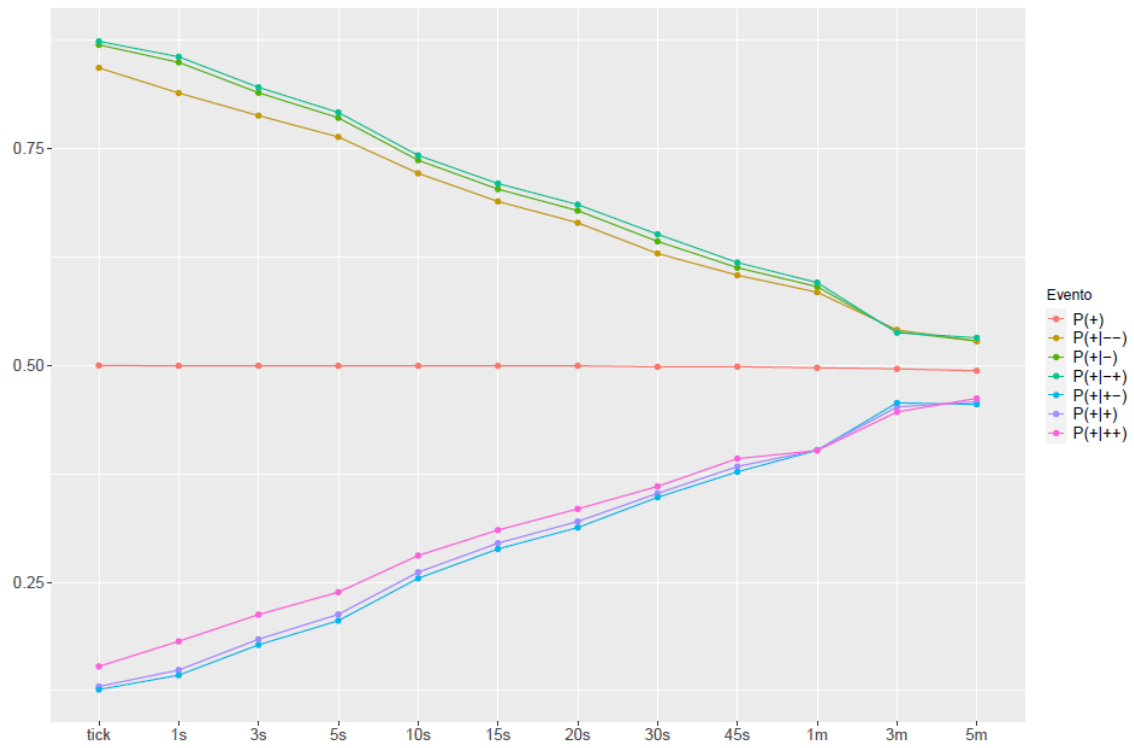


Figura A 13 - Probabilidade Condicional para até dois movimentos à priori (ABEV3).

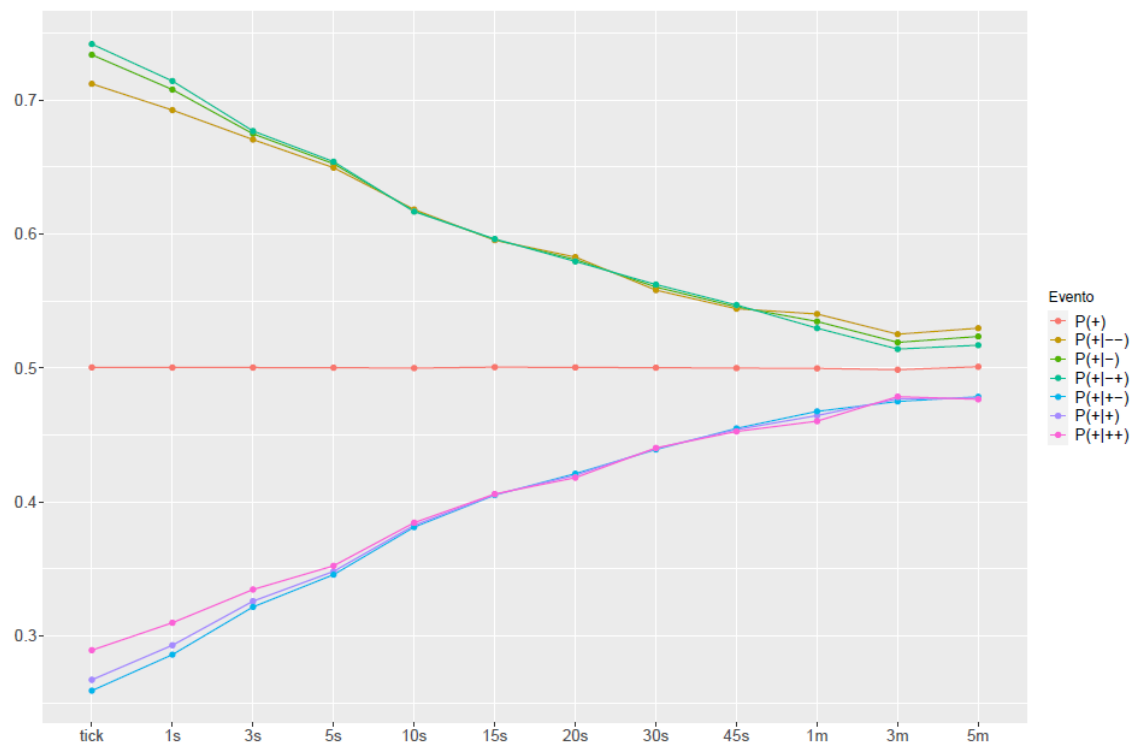


Figura A 14 - Probabilidade Condicional para até dois movimentos à priori (BBDC4).

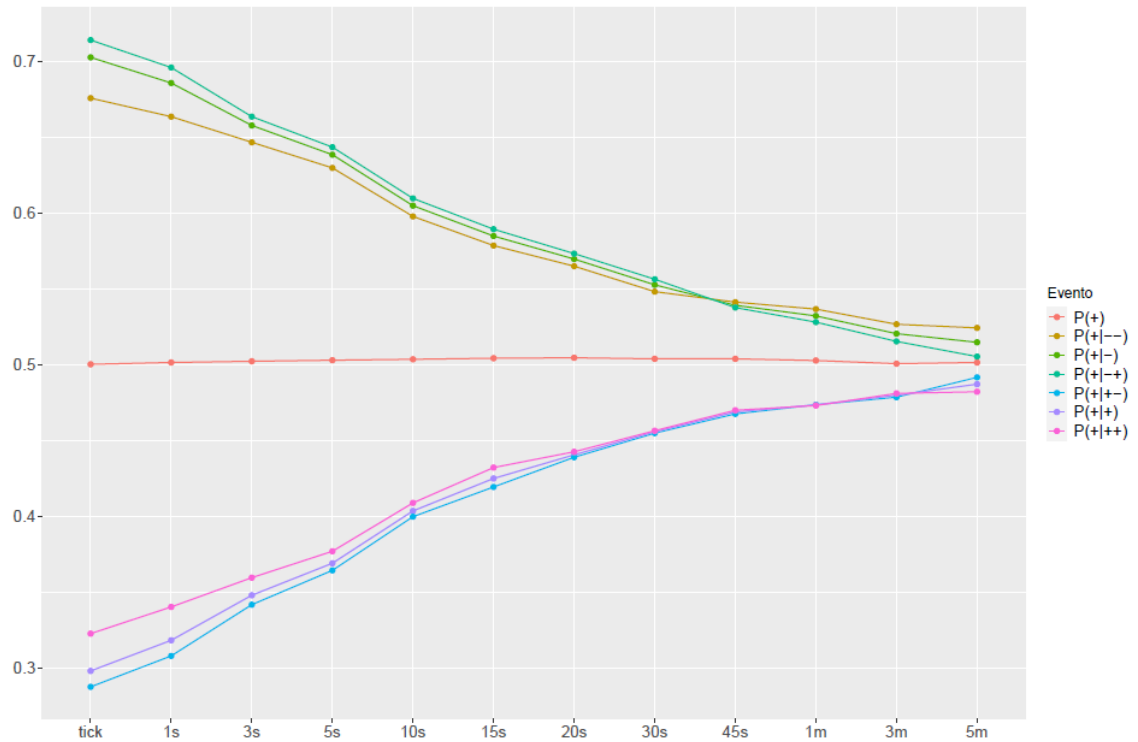


Figura A 15 - Probabilidade Condicional para até dois movimentos à priori (VALE3).

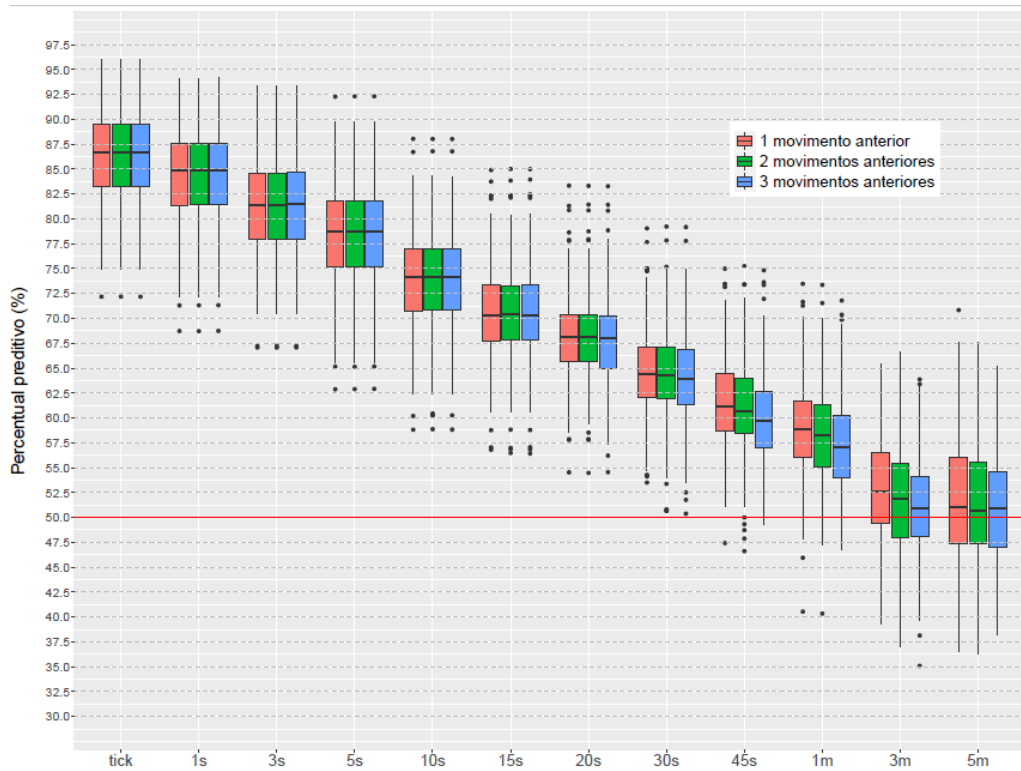


Figura A 16 - Percentuais de acerto para cada modelo logístico (ABEV3).

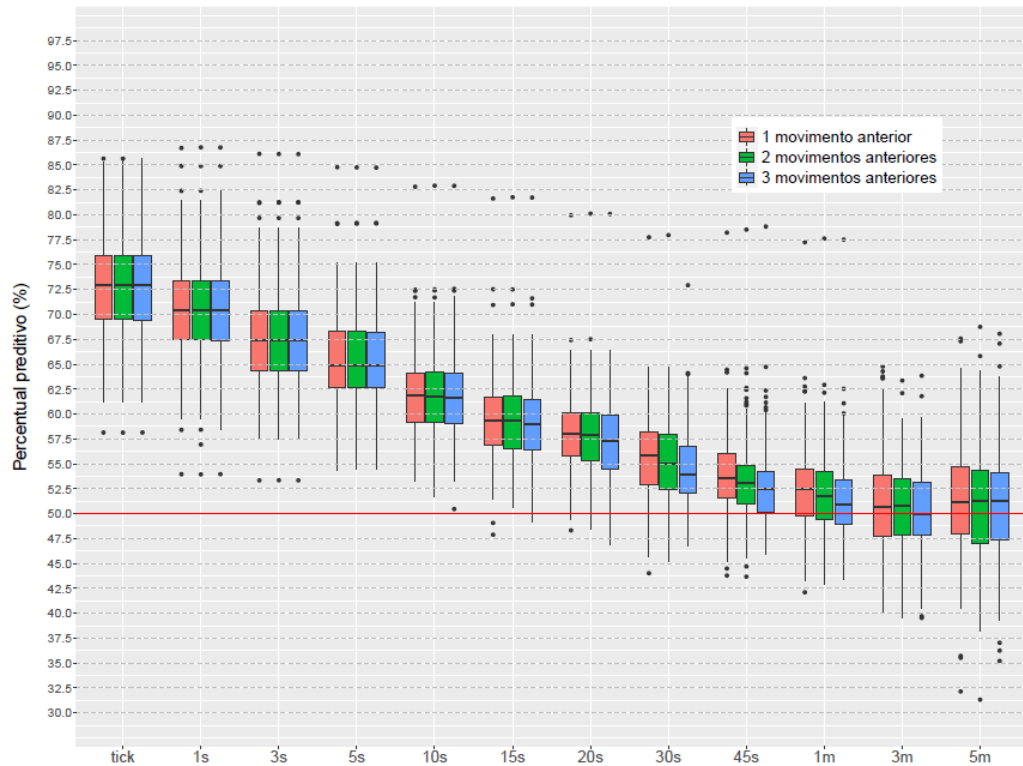


Figura A 17 - Percentuais de acerto para cada modelo logístico (BBDC4).

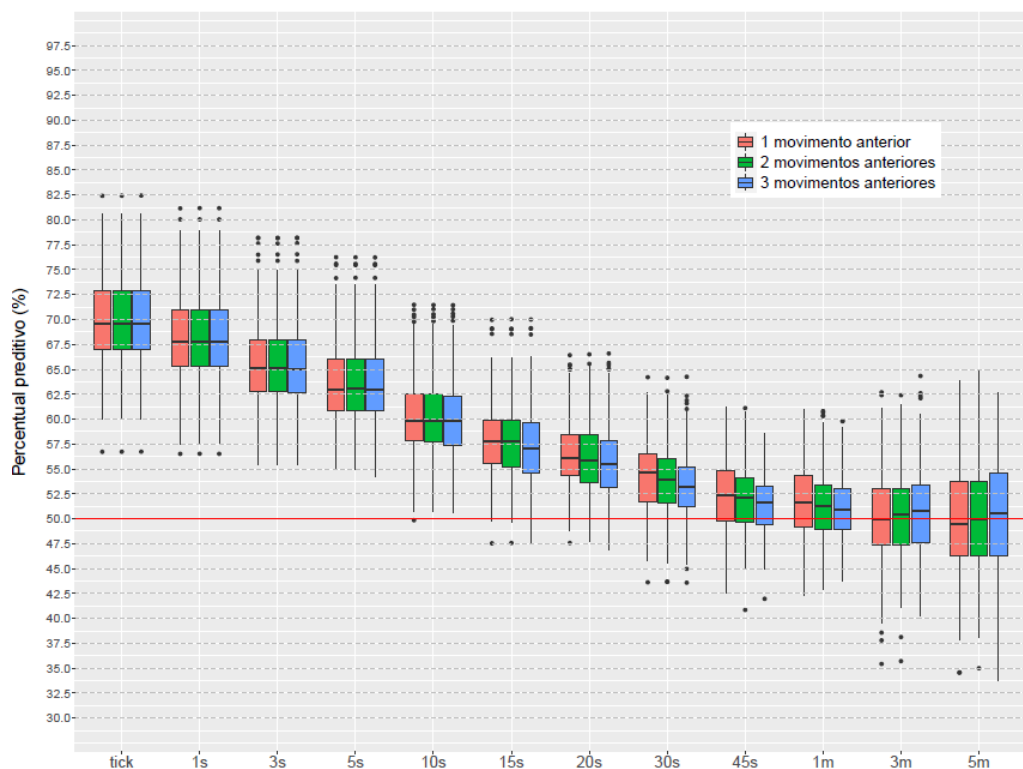


Figura A 18 - Percentuais de acerto para cada modelo logístico (VALE3).

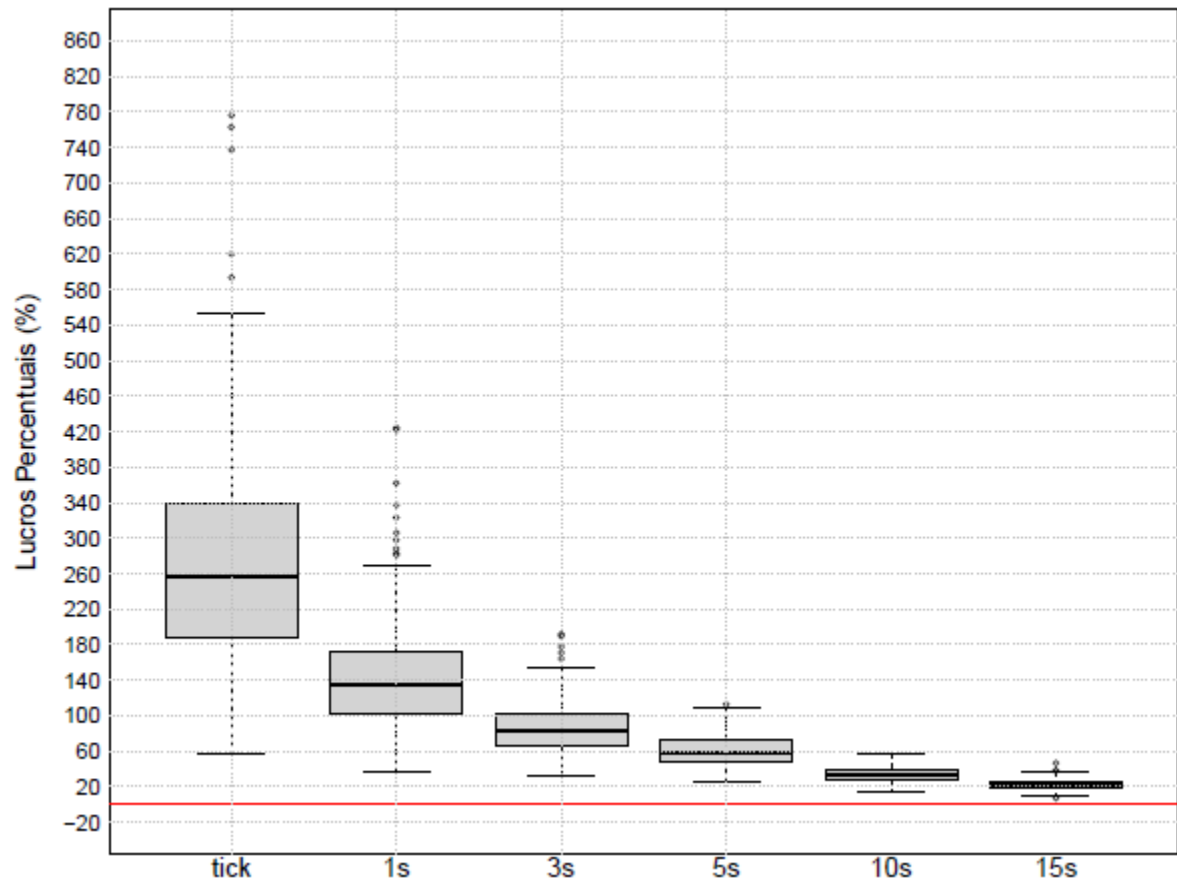


Figura A 19 - Lucros Percentuais para os dados tick a tick e de até 15 segundos (ABEV3)

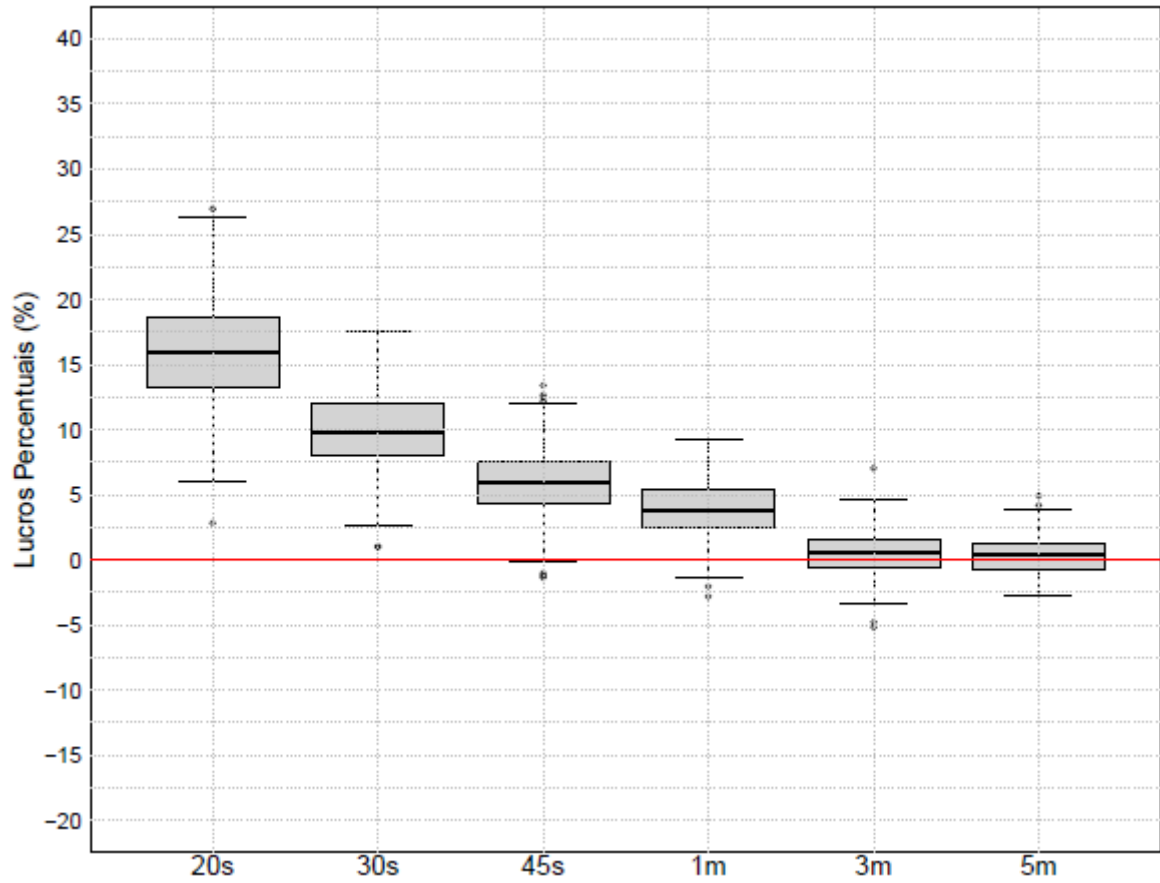


Figura A 20 - Lucros Percentuais para os dados de 20 segundos até 5 minutos (ABEV3).

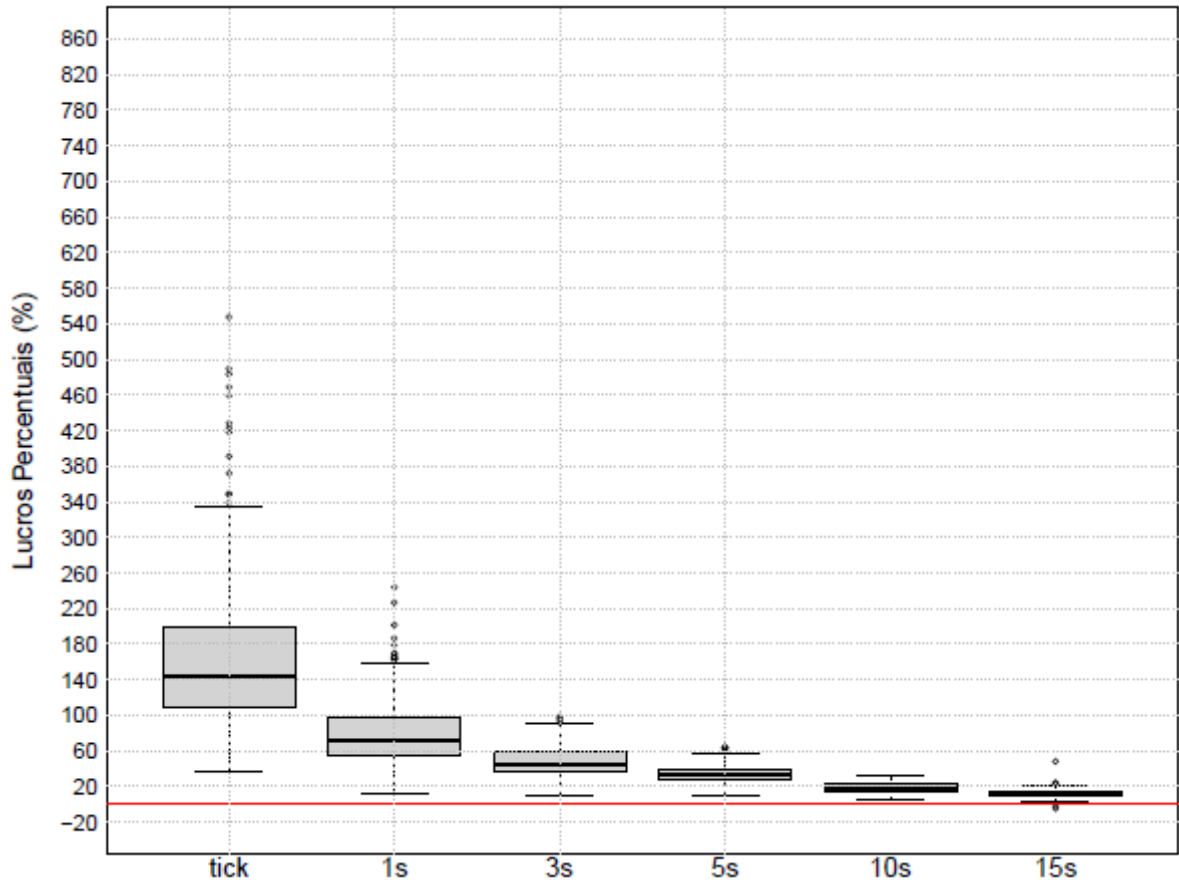


Figura A 21 - Lucros Percentuais para os dados tick a tick e de até 15 segundos (BBDC4)

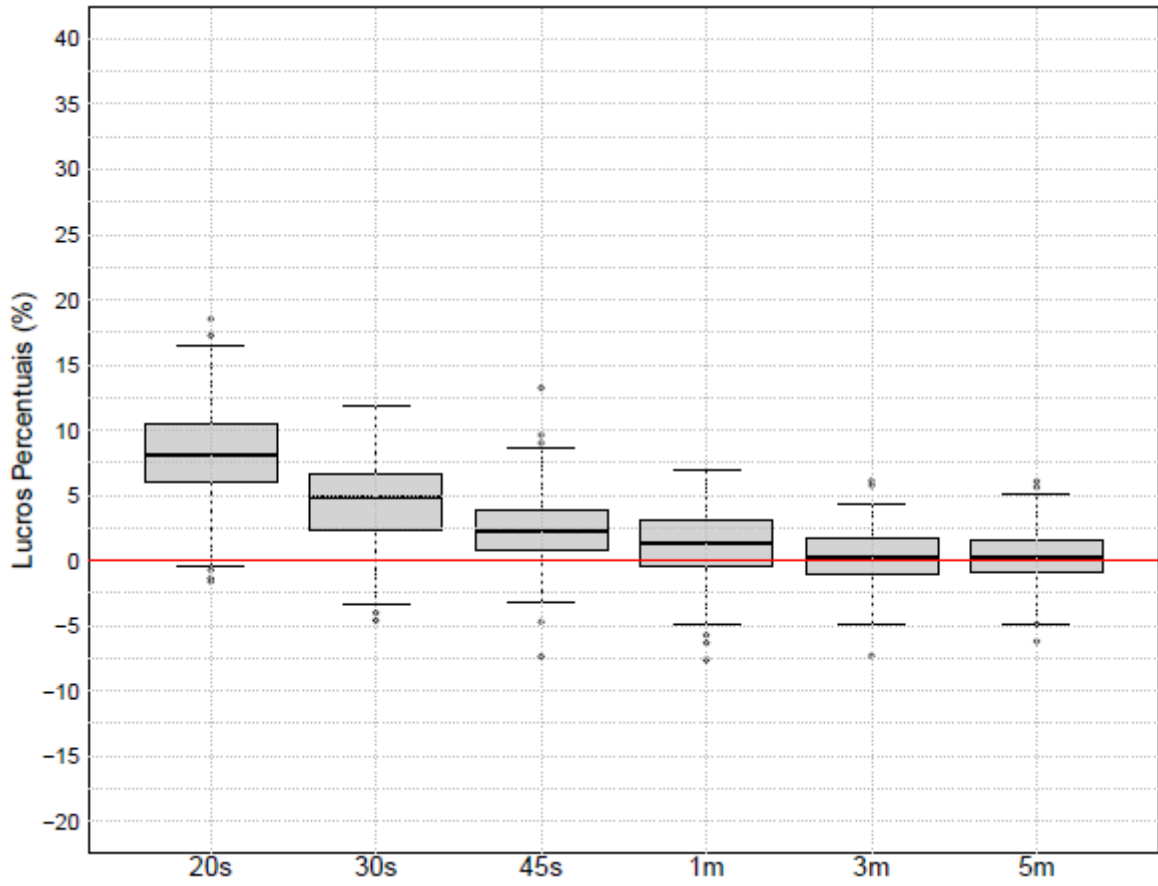


Figura A 22 - Lucros Percentuais para os dados de 20 segundos até 5 minutos (BBDC4).

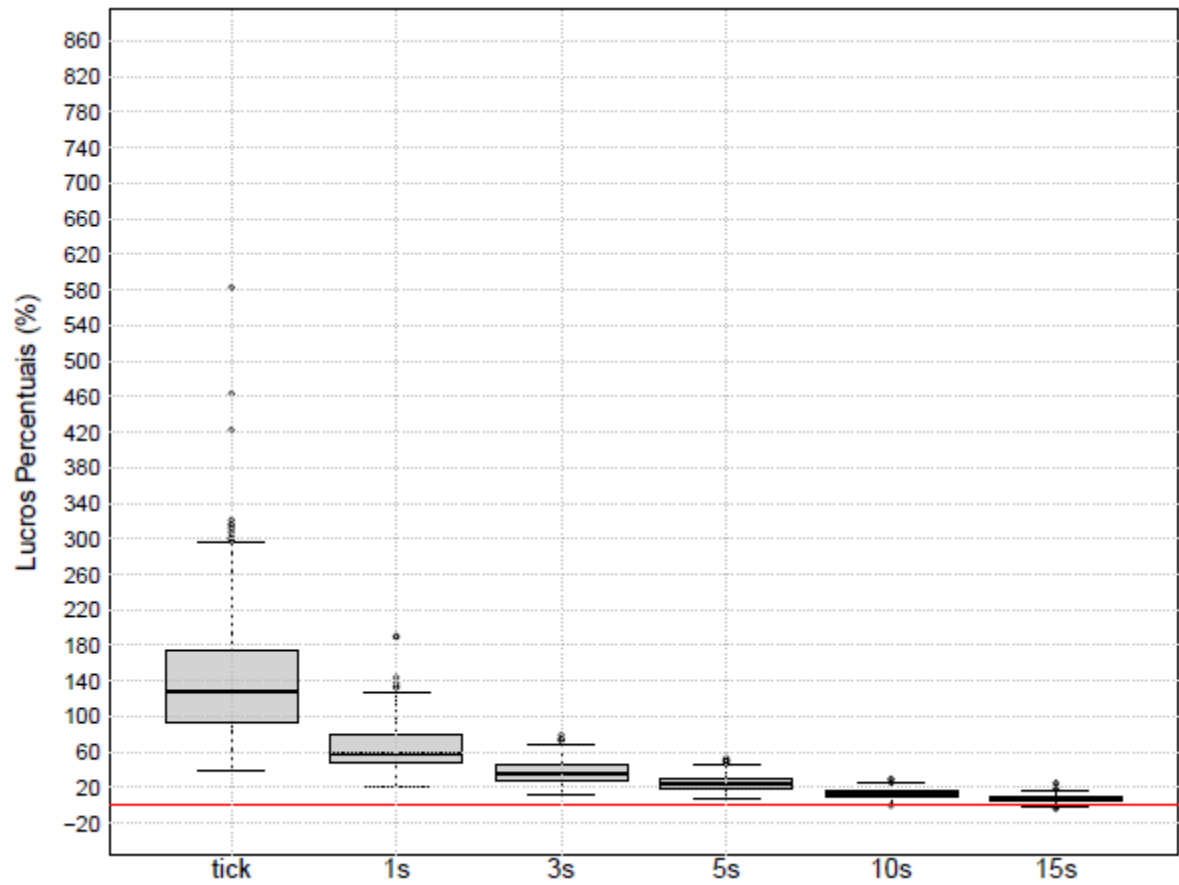


Figura A 23 - Lucros Percentuais para os dados tick a tick e de até 15 segundos (VALE3)



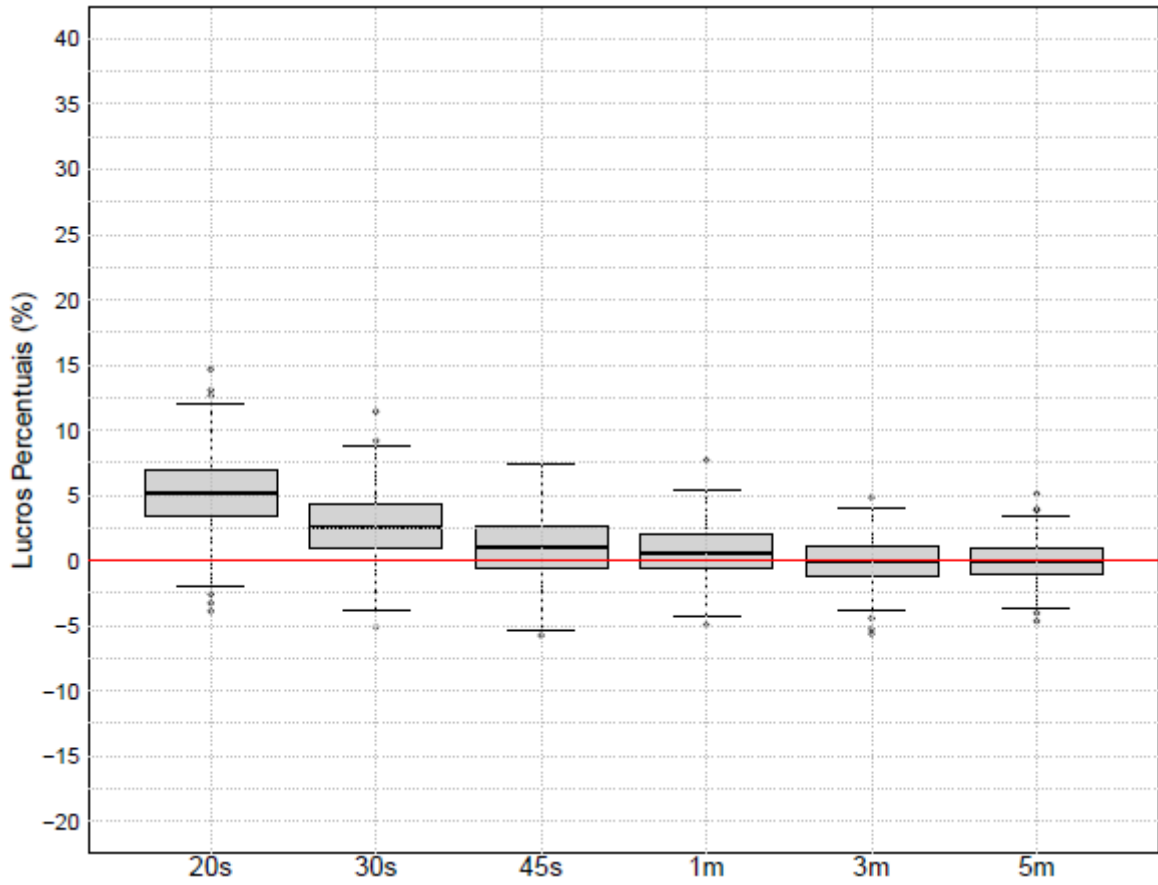


Figura A 24 - Lucros Percentuais para os dados de 20 segundos até 5 minutos (VALE3).

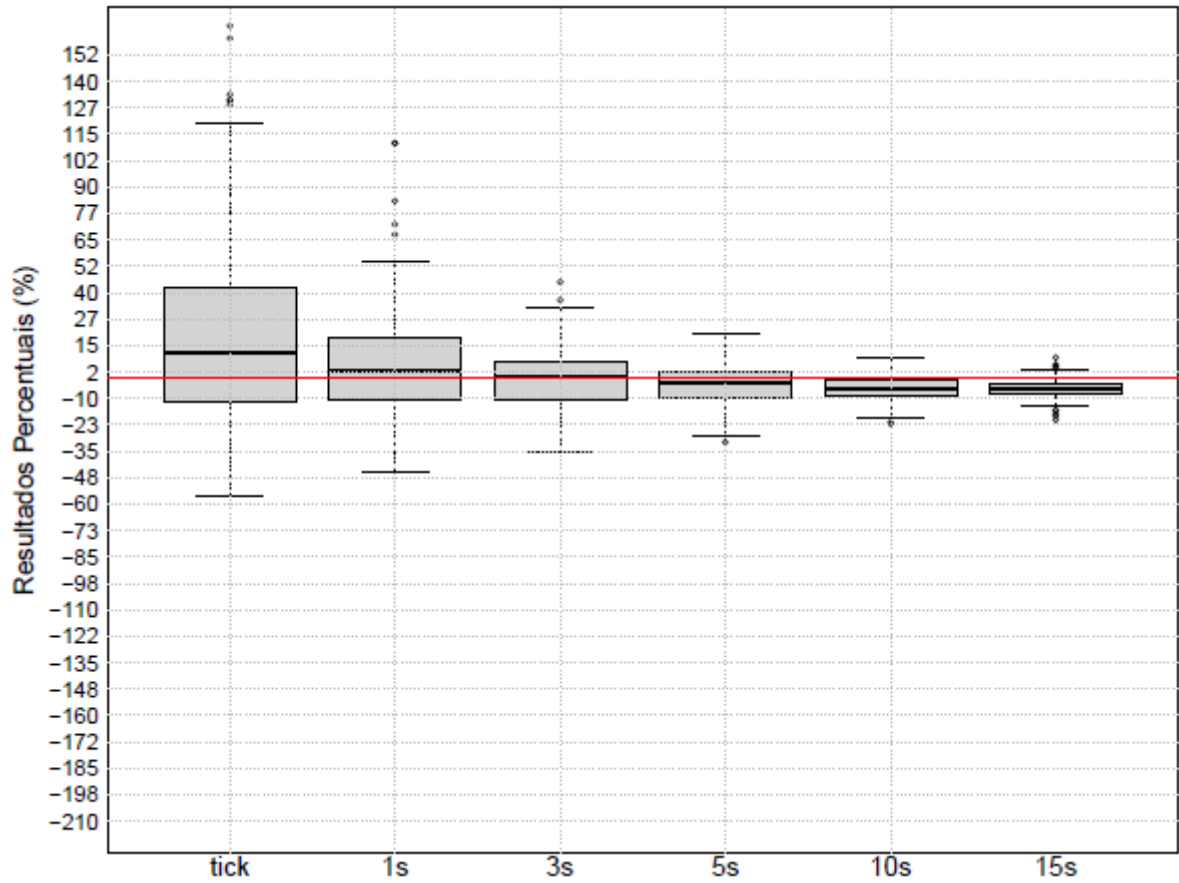


Figura A 25 - Resultados Percentuais para os dados tick a tick e de até 15 segundos (ABEV3)

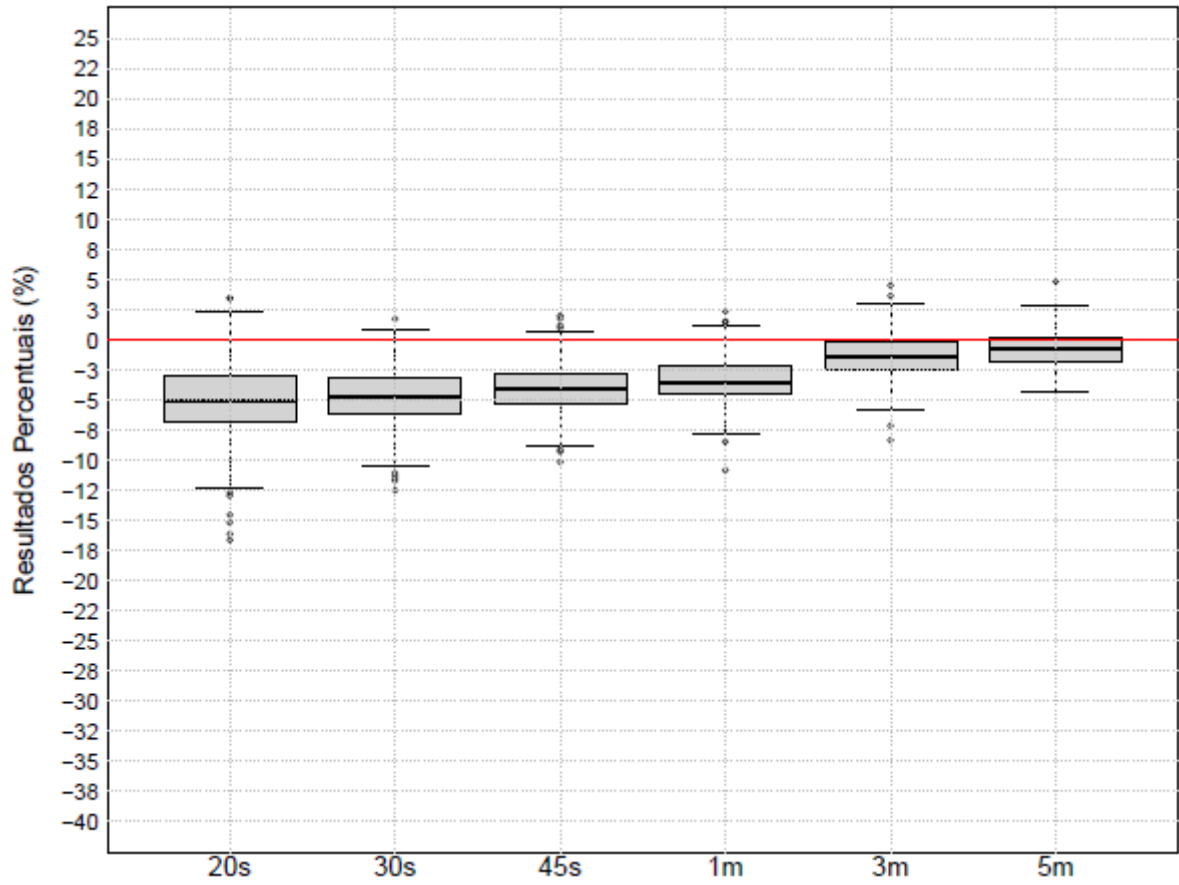


Figura A 26 - Resultados Percentuais para os dados de 20 segundos até 5 minutos (ABEV3)

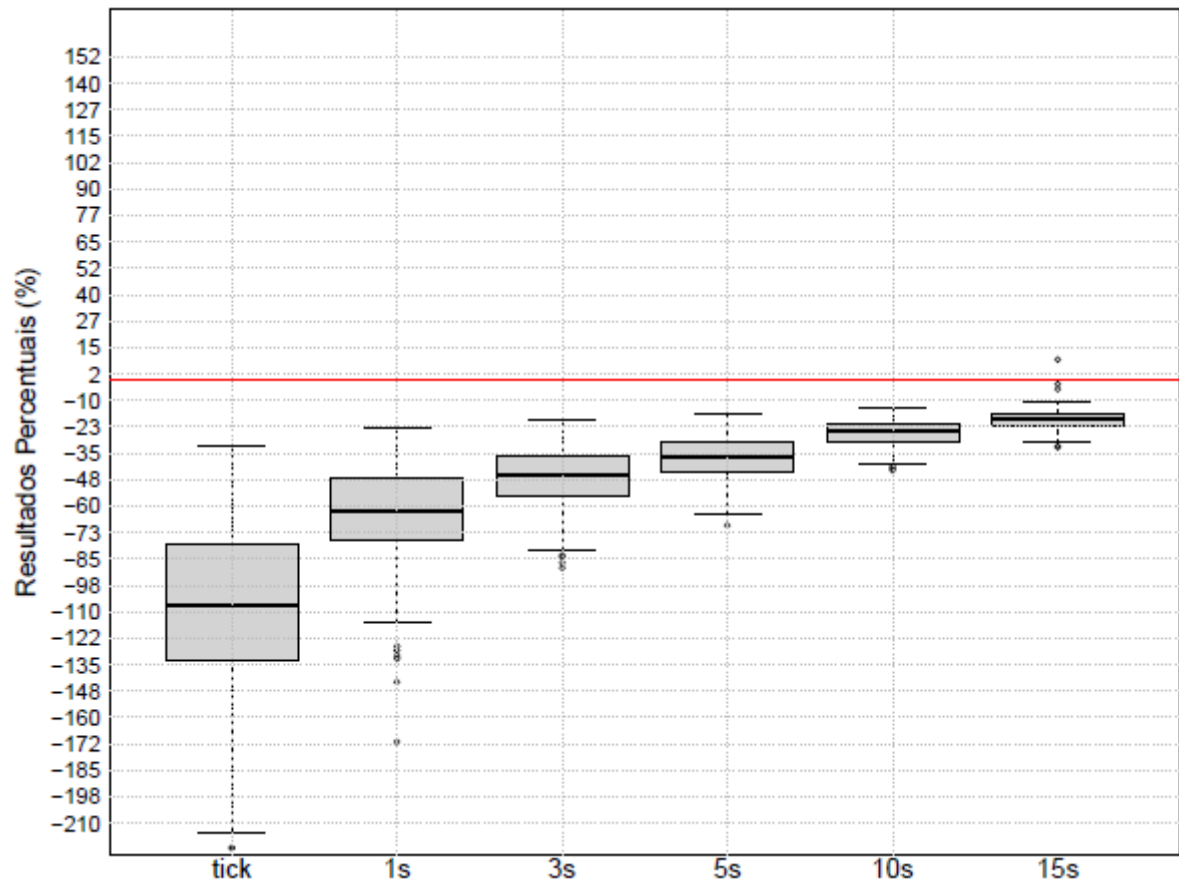


Figura A 27 - Resultados Percentuais para os dados tick a tick e de até 15 segundos (BBDC4)

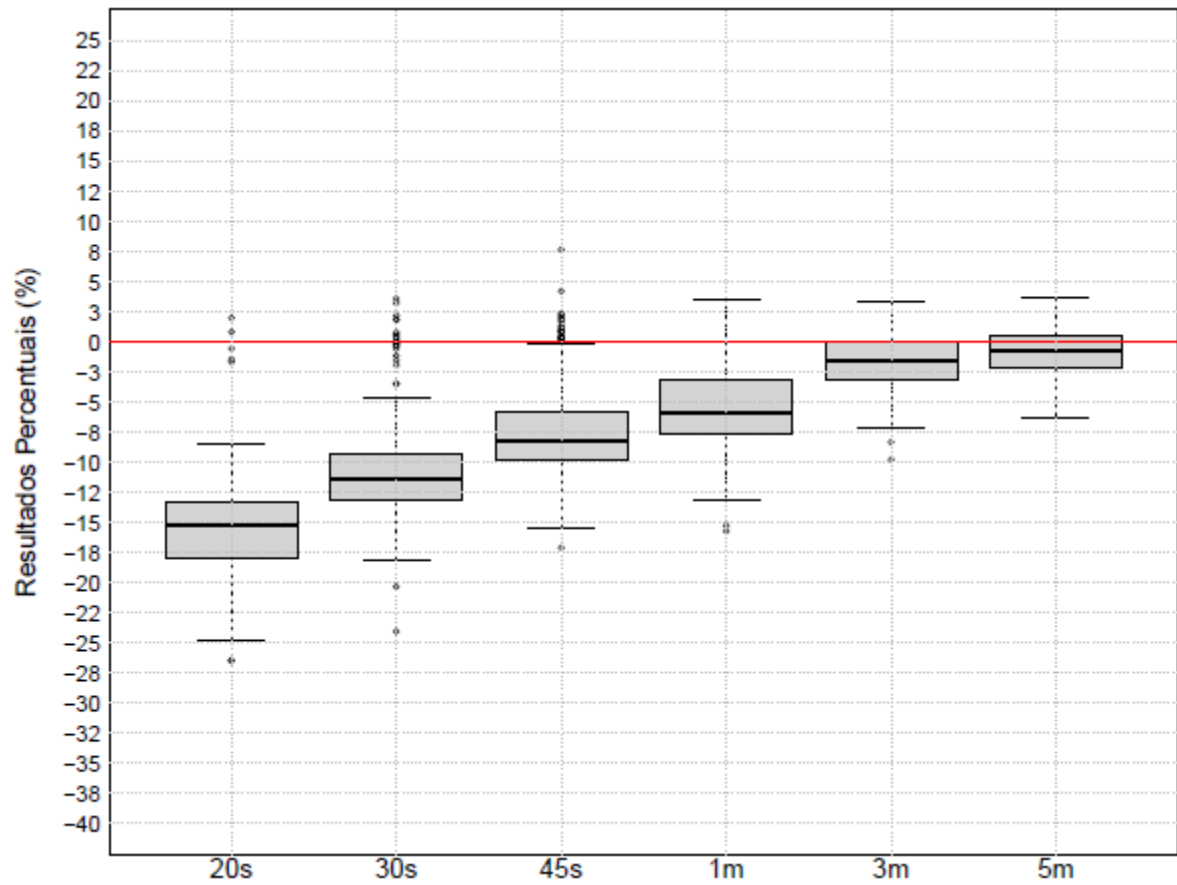


Figura A 28 - Resultados Percentuais para os dados de 20 segundos até 5 minutos (BBDC4)

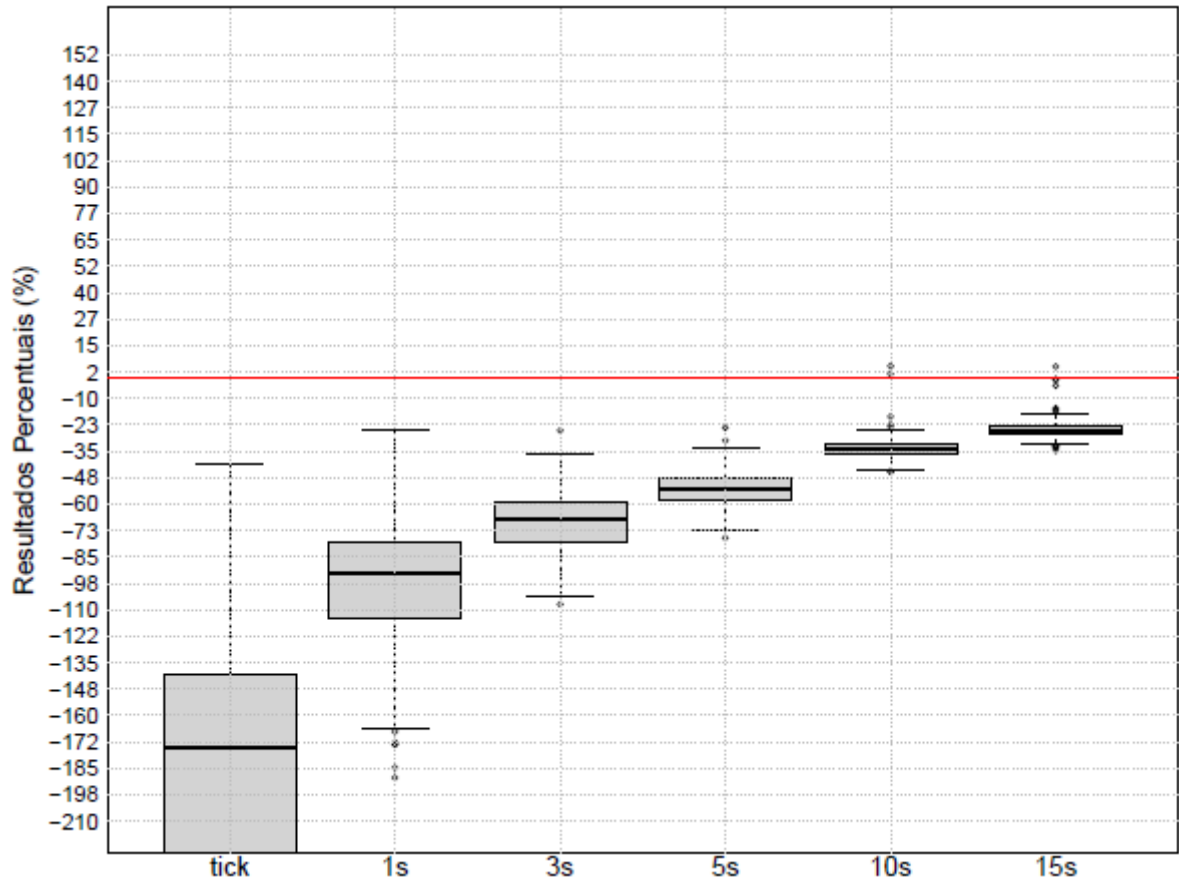


Figura A 29 - Resultados Percentuais para os dados tick a tick e de até 15 segundos (VALE3)

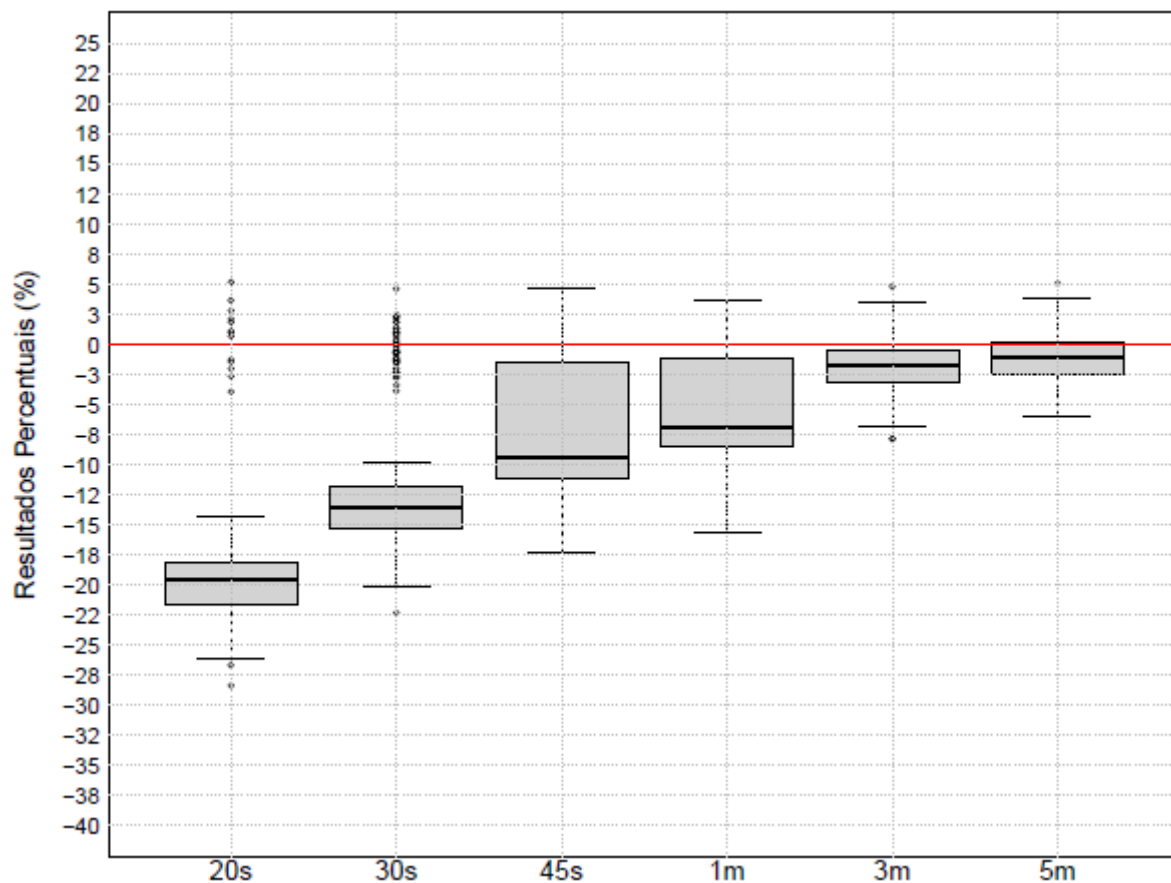


Figura A 30 - Resultados Percentuais para os dados de 20 segundos até 5 minutos (VALE3)