

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

**Análise da rentabilidade de investimentos em ofertas  
públicas iniciais de ações entre 2004 a 2023 no Brasil:  
uma abordagem estatística**

**Luiza Neves Graça**

**Trabalho de Conclusão de Curso**



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Análise da rentabilidade de investimentos em ofertas públicas  
iniciais de ações entre 2004 a 2023 no Brasil:  
uma abordagem estatística

**Luiza Neves Graça**

**Orientadora: Teresa Cristina Martins Dias**

Trabalho de Conclusão de Curso apresentado  
como parte dos requisitos para obtenção do  
título de Bacharel em Estatística.

**São Carlos**  
**Março de 2025**



FEDERAL UNIVERSITY OF SÃO CARLOS  
EXACT AND TECHNOLOGY SCIENCES CENTER  
DEPARTMENT OF STATISTICS

Analysis of the profitability of investments in initial public  
offerings of stocks from 2004 to 2023 in Brazil: a statistical  
approach

**Luiza Neves Graça**

**Advisor: Teresa Cristina Martins Dias**

Bachelors dissertation submitted to the Department of Statistics, Federal University of São Carlos - DEs-UFSCar, in partial fulfillment of the requirements for the degree of Bachelor in Statistics.

**São Carlos**  
**February 2025**



Luiza Neves Graça

Análise da rentabilidade de investimentos em ofertas públicas  
iniciais de ações entre 2004 a 2023 no Brasil:  
uma abordagem estatística

Este exemplar corresponde à redação final do trabalho de conclusão de curso devidamente corrigido e defendido por Luiza Neves Graça e aprovado pela banca examinadora.

Aprovado em 5 de fevereiro de 2025

Banca Examinadora:

Teresa Cristina Martins Dias (Orientadora)

Gustavo Henrique de Araujo Pereira

Ricardo Felipe Ferreira



*Aos meus pais por todo apoio e carinho.*



*“Toda vez que um artigo científico apresenta alguns dados, eles vêm acompanhados por uma margem de erro – um lembrete silencioso, mas insistente, de que nenhum conhecimento é completo ou perfeito.”*

*(Carl Sagan, O Mundo Assombrado Pelos Demônios)*



# Resumo

Dentro do sistema financeiro de um país, o mercado de ações representa um importante mecanismo de alocação de recursos entre empresas e investidores. As empresas acessam esse mercado inicialmente por meio da realização de uma oferta pública inicial (*IPO*) e, em troca, o investidor passa a beneficiar-se de eventuais lucros e da possível valorização do ativo no mercado secundário. Com base nisso, neste trabalho, a nossa motivação é investigar, a partir de um banco de dados reais do mercado de ações brasileiro de 2004 a 2023, a performance dos investimentos em *IPOs* frente à rentabilidade do mercado acionário como um todo, representado pelo Ibovespa, durante um período de observação de dois anos. Para alcançar esse objetivo, foram empregados testes não paramétricos para a mediana da rentabilidade dos *IPOs* relativa ao mercado, bem como foi avaliado o comportamento do tempo até a ocorrência de desvalorizações consideráveis desse investimento frente à diferentes variáveis explicativas por meio de métodos de análise de sobrevivência, como os modelos de Cox e de Prentice, Williams e Peterson (PWP). A conclusão foi de que, no curto prazo, o investimento em *IPOs* tende a performar acima do mercado, refletindo um fenômeno documentado como *underpricing*. Contudo, para períodos mais longos, os retornos tornam-se negativos e passam a apresentar uma tendência decrescente. No que diz respeito a esse último fenômeno, pelos modelos de sobrevivência ajustados, desvalorizações maiores tendem a estar associadas a ofertas que foram realizadas em períodos em que o mercado está mais aquecido além de também estarem associadas a empresas que performaram mal desde o primeiro dia.

**Palavras-chave:** *Análise de sobrevivência; Estatística não paramétrica; Mercado acionário brasileiro; Modelo de Cox; Modelo de Prentice, Williams e Peterson; Oferta Pública Inicial.*



# Abstract

Within a country's financial system, the stock market serves as a crucial mechanism for allocating resources between companies and investors. Companies initially access this market by conducting an initial public offering (IPO), and in return, investors benefit from potential profits and the possible appreciation of the asset in the secondary market. Based on this, our motivation in this study is to investigate, using real data from the Brazilian stock market from 2004 to 2023, the performance of IPO investments compared to the overall stock market returns, represented by the Ibovespa, over a two-year observation period. To achieve this objective, we employed nonparametric tests to compare the median return of IPOs relative to the market, as well as survival analysis methods—such as the Cox model and the Prentice, Williams, and Peterson (PWP) model—to examine the time until significant devaluations of these investments in relation to different explanatory variables. The findings indicate that, in the short term, IPO investments tend to outperform the market, reflecting a well-documented phenomenon known as underpricing. However, over longer periods, returns become negative and exhibit a declining trend. Regarding this latter phenomenon, the adjusted survival models suggest that larger devaluations tend to be associated with offerings made during periods in which the market is heated and with companies that performed poorly from the very first day.

**Keywords:** *Brazilian stock market; Cox model; Initial Public Offering; Nonparametric statistics; Prentice, Williams, and Peterson model; Survival analysis.*



# Sumário

<b>1</b>	<b>Introdução</b>	<b>17</b>
<b>2</b>	<b>Metodologia</b>	<b>21</b>
2.1	Estatística não paramétrica . . . . .	21
2.1.1	Teste do sinal para a mediana . . . . .	21
2.2	Análise de sobrevivência . . . . .	23
2.2.1	Estimador de Kaplan-Meier . . . . .	24
2.2.2	Modelo de Cox para riscos proporcionais . . . . .	26
2.2.3	Modelo de Prentice, Williams e Peterson . . . . .	29
2.2.4	Análise de Diagnóstico . . . . .	31
<b>3</b>	<b>Material</b>	<b>33</b>
3.1	Descrição do conjunto de dados . . . . .	33
3.1.1	Inclusão de covariáveis . . . . .	35
<b>4</b>	<b>Resultados</b>	<b>39</b>
4.1	Análise descritiva e exploratória . . . . .	39
4.1.1	Comportamento da variável resposta . . . . .	40
4.1.2	Comportamento da variável resposta na presença de covariáveis . . . . .	44
4.2	Teste de hipótese para a mediana . . . . .	49
4.3	Modelagem . . . . .	52
4.3.1	Modelos de Cox . . . . .	52
4.3.2	Modelo PWP . . . . .	57
<b>5</b>	<b>Conclusão</b>	<b>61</b>
	<b>Referências Bibliográficas</b>	<b>65</b>

<b>A</b>	<b>Construção da base de dados</b>	<b>69</b>
<b>B</b>	<b>Códigos utilizados</b>	<b>73</b>
B.1	Kaplan-Meier . . . . .	73
B.2	Modelo PWP . . . . .	74

# Capítulo 1

## Introdução

No cerne da economia de um país, o mercado de capitais representa um importante braço do sistema financeiro, fomentando o desenvolvimento econômico através da alocação eficiente de recursos de médio e longo prazo entre empresas e investidores.

Nessa linha, o mercado de ações, por sua vez, representa um relevante segmento do mercado de capitais. Enquanto este último engloba diversas formas de instrumentos financeiros para captação de recursos, como títulos de dívida e derivativos, o mercado de ações concentra-se na emissão de ações, as quais simbolizam uma fração da propriedade das companhias de capital aberto, e os recursos advindos dessa emissão viabilizam o crescimento dessas empresas e consequente geração de empregos e renda, contribuindo para o desenvolvimento do país. Para ingressar nesse mercado, a empresa deve passar pelo processo de oferta pública inicial de ações (em inglês *Initial Public Offering (IPO)*).

Uma oferta pública inicial de ações ocorre quando uma empresa já estabelecida busca atrair investidores para financiar novos projetos através de um mecanismo estruturado de acesso aos mercados organizados da bolsa de valores brasileira (B3). Para tanto, o processo consiste em obter as aprovações dos órgãos reguladores, sobretudo da Comissão de Valores Mobiliários (CVM) e da B3, e contactar agentes de mercado, como bancos de investimentos e corretoras, para organizar a distribuição e promoção dessa oferta ao público. Em troca, o investidor passa a ter uma pequena participação na estrutura societária da empresa, podendo beneficiar-se de eventuais lucros no futuro.

Após a oferta inicial, as ações passam a ser negociadas no mercado secundário da B3 e seus preços passam a flutuar conforme a oferta e demanda, o que proporciona liquidez, isto é, a capacidade de transformar seus investimentos em dinheiro, aos investidores que adquiriram suas ações durante a oferta inicial para venderem suas posições conforme

necessário, o que cria também uma oportunidade para lucrarem com a valorização do ativo. Assim, o mercado primário depende do secundário e *vice-versa*.

Ao longo dos anos, no entanto, a atratividade de se investir em *IPOs* tem sido alvo de debates. Nos estudos realizados para avaliar a performance desses investimentos, dois fenômenos recorrentes foram identificados consistentemente em diversos países nos últimos anos, conforme sumarizado em [Gurung \(2024\)](#) e [Martins et al. \(2008\)](#); o primeiro é referente ao fenômeno conhecido como *underpricing*, segundo o qual a performance dos *IPOs* no curto prazo tende a ser positiva, já o segundo é referente ao fenômeno denominado *long-term underperformance*, em que no longo prazo o desempenho das ofertas iniciais tende a ser pior que o mercado no geral.

Por exemplo, em [Ibbotson, Sindelar e Ritter \(1988\)](#), foi coletada uma amostra de 8.668 *IPOs* realizados no período de 1960 a 1987 no mercado estadunidense e os autores concluíram que, em média, estes *IPOs* tiveram um retorno positivo de 16,4% no primeiro dia após o lançamento. Já em [Ritter \(1991\)](#), os autores concluíram, para uma amostra de 1.526 *IPOs* do mercado estadunidense durante o período de 1975 a 1980, que a performance dos *IPOs* três anos após o lançamento, em média, foi em torno de 30% inferior ao mercado. Resultados similares, no curto e longo prazo, foram obtidos por [Siegel \(2008\)](#) com uma amostra de 8.606 *IPOs* estadunidenses lançados entre 1961 a 2001. No contexto de mercados latino-americanos, em [Aggarwal, Leal e Hernandez \(1993\)](#) foram analisados 62 *IPOs* brasileiros, 36 *IPOs* chilenos e 44 *IPOs* mexicanos, todos realizados entre 1980 e 1990 e acompanhados durante três anos após a oferta, sendo também observado o mesmo comportamento. Todavia, vale mencionar que muitos destes estudos limitaram-se a uma análise descritiva dos dados e não englobam dados mais recentes do mercado acionário.

Simultaneamente aos estudos acadêmicos realizados sobre o tema, no debate acerca da atratividade de investir em *IPOs*, algumas vertentes de pensamento se destacam. Por um lado, tem-se a escola fundamentalista, enraizada no *value investing* e idealizada por investidores como Benjamin Graham e David Dodd ([Graham e Dodd, 1934](#); [Graham, 1949](#)). Essa abordagem argumenta que os *IPOs* geralmente apresentam desempenho inferior ao mercado devido a fatores como a escolha do momento ideal pelo vendedor para entrar no mercado, a disparidade na disponibilidade de informações entre os proprietários originais e os compradores, e o incentivo dos bancos de investimento e proprietários das empresas para promover uma visão excessivamente otimista das perspectivas futuras, visando maximizar o preço inicial e os seus ganhos pessoais.

Em contrapartida, o investimento em *IPOs* é amplamente recomendado por corretoras, analistas de mercado e personalidades em mídias sociais, que exaltam as oportunidades, o potencial e as qualidades das empresas recém-ingressantes no mercado.

Paralelamente, como uma terceira frente, tem-se a teoria de mercados eficientes, preconizada pelo ganhador do prêmio Nobel de economia Eugene Fama ([Fama, 1970](#)), que postula que os preços dos ativos tendem a refletir todas as informações disponíveis e que os participantes do mercado são racionais e, portanto, não seria plausível assumir que o investimento em *IPOs* possa ser consistentemente pior ou melhor que o mercado de ações em geral.

Em síntese, com base nas teorias e estudos elencados, nossa motivação é verificar, no mercado brasileiro, se os investimentos em *IPOs* têm performance superior, inferior ou equivalente ao mercado acionário no geral em diferentes horizontes de tempo, do curto ao médio e longo prazo, e como essa rentabilidade se comporta sob o efeito de diferentes variáveis, selecionadas de acordo com algumas das principais teorias acerca das distorções que podem ocorrer no processo de precificação de novas ofertas.

Para alcançar esse objetivo, foram levantados dados acerca da rentabilidade relativa ao Ibovespa para 220 *IPOs* brasileiros realizados entre 2004 e 2023, acompanhando seus desempenhos por dois anos após suas respectivas estreias. Nesse contexto, a metodologia que será empregada a esse conjunto de dados abordará duas técnicas principais.

A primeira é referente ao emprego de testes não paramétricos para avaliar se os retornos dos *IPOs* são significativamente superiores ou inferiores aos retornos do Ibovespa. Para tanto, os retornos relativos ao Ibovespa de cada *IPO* foram observados em horizontes de tempo equivalente, em dias corridos, a um dia, uma semana, um mês, seis meses, um ano e dois anos, sendo os três primeiros referentes a prazos de curta duração e os três últimos referentes a prazos de média a longa duração. Assim, é possível avaliar o comportamento do conjunto de dados sem ter que analisar todas as cotações diárias.

A segunda abordagem engloba a aplicação de modelos de sobrevivência ao problema. Tal método já foi aplicado em estudos anteriores que avaliam o tempo até que uma nova oferta deixe de ser negociada na bolsa, seja por falência, fusões, aquisições ou outros motivos ([Espanlaub et al., 2016](#); [Baluja, 2017](#)). No entanto, neste estudo, o evento de interesse não é a retirada do registro da empresa, mas sim o momento em que determinado nível de perda relativo ao Ibovespa for alcançado. Desse modo, foram estabelecidos três patamares de perda relativa ao Ibovespa de acordo com a análise descritiva, que atuam como

o critério de determinação da “mortalidade” das observações para o ajuste de modelos de sobrevivência. Optamos pelos modelos de sobrevivência por sua capacidade não apenas de examinar a ocorrência de retornos negativos, mas também de investigar o tempo até essa ocorrência e avaliar quais variáveis podem acelerar ou influenciar o evento de interesse.

Dentro da análise de sobrevivência, estimamos a função de sobrevivência pelo método não paramétrico de Kaplan-Meier ([Kaplan e Meier, 1958](#)), o que auxilia a compreensão do comportamento da curva de sobrevivência para diferentes patamares de perda. Em seguida, frente à inclusão das covariáveis, foram abordadas duas técnicas de modelagem. A primeira se trata do ajuste de múltiplos modelos de Cox para riscos proporcionais ([Colosimo, 2024](#); [Cox, 1972](#)) para diferentes níveis de perdas relativas. Já a segunda abordagem engloba o ajuste de modelos de sobrevivência para eventos múltiplos ordenados estruturados ([Carvalho \*et al.\*, 2011](#)), mais especificamente do modelo proposto por [Prentice \*et al.\* \(1981\)](#).

Assim, neste estudo, com base nos objetivos estabelecidos e nas metodologias brevemente abordadas acima, visamos contribuir para melhorar a transparência do mercado acionário, através da compreensão de possíveis tendências históricas e potenciais distorções na precificação dos ativos.

Este trabalho está organizado da seguinte forma: no Capítulo 2, encontra-se uma revisão da teoria subjacente aos métodos que foram empregados ao problema proposto; no Capítulo 3, tem-se uma descrição da base de dados utilizada e de sua construção, bem como os critérios de seleção de covariáveis e como estas foram obtidas; no Capítulo 4, são analisados os resultados obtidos a partir do emprego dos métodos propostos aos dados; e, por fim, no Capítulo 5, são elencadas as conclusões obtidas a partir das etapas anteriores.

# Capítulo 2

## Metodologia

Neste capítulo, são abordados os métodos utilizados para a realização deste trabalho, que incluem testes não paramétricos, na Seção 2.1, e técnicas de análise de sobrevivência, mais especificamente a curva de Kaplan-Meier e modelos de sobrevivência, na Seção 2.2.

### 2.1 Estatística não paramétrica

O campo de estudo da estatística não paramétrica pode ser definido como a coleção de métodos estatísticos que, ao contrário da sua contraparte paramétrica, não supõe que os dados provêm de uma distribuição específica, como a distribuição normal (M. G. Kendall, 1954). À vista disso, estes métodos podem ser utilizados em substituição de técnicas paramétricas em diversas situações, por exemplo: nas situações em que as suposições do método a ser utilizado não se verificam, seja porque as observações em estudo não possuem normalidade ou homogeneidade da variância; nos casos em que é difícil de se obter estimativas para os parâmetros populacionais ou as variáveis são qualitativas.

Os métodos não paramétricos, portanto, são menos exigentes e mais flexíveis em suas suposições, mais adequados a pequenas amostras e menos sensíveis a erros de medida. No caso deste estudo, a escolha de métodos não paramétricos foi feita em função da não normalidade dos dados, a qual ficará evidente mais à frente.

#### 2.1.1 Teste do sinal para a mediana

O teste do sinal para a mediana (Conover, 1999; John W. Pratt, 1981) é um teste não paramétrico e parte de duas suposições; a de que a unidade de medida dos dados é no

mínimo ordinal e que cada unidade amostral é independente das demais. Assim, trata-se de uma ferramenta flexível e abrangente a uma variedade de problemas e dados.

A princípio, considere  $X_1, \dots, X_n$  variáveis aleatórias, independentes e identicamente distribuídas retiradas como amostra da distribuição da variável aleatória de interesse  $X$ , cuja escala de medida é no mínimo ordinal e com mediana populacional desconhecida  $M$ . Neste caso, o teste do sinal nos ajudará a verificar se há evidências, com base na amostra, de que a mediana  $M$  é diferente, maior ou menor que determinado valor  $M_0$ .

Logo, três possíveis hipóteses podem ser formuladas e testadas – a hipótese bilateral, a unilateral à esquerda e a unilateral à direita –, dessa forma, considerando  $i = 1, \dots, n$ , estas podem ser detalhadas da seguinte forma:

$$\left\{ \begin{array}{l} H_0 : M = M_0 \Leftrightarrow (X_i < M_0) = 1/2 \\ H_1 : M \neq M_0 \Leftrightarrow (X_i < M_0) \neq 1/2 \end{array} \right. \quad (\text{hipótese bilateral})$$

$$\left\{ \begin{array}{l} H_0 : M = M_0 \Leftrightarrow (X_i < M_0) = 1/2 \\ H_1 : M < M_0 \Leftrightarrow (X_i < M_0) > 1/2 \end{array} \right. \quad (\text{hipótese unilateral à esquerda})$$

$$\left\{ \begin{array}{l} H_0 : M = M_0 \Leftrightarrow (X_i < M_0) = 1/2 \\ H_1 : M > M_0 \Leftrightarrow (X_i < M_0) < 1/2. \end{array} \right. \quad (\text{hipótese unilateral à direita.})$$

Em seguida, a fim de verificar essas hipóteses, as unidades amostrais são transformadas em valores binários, ou sinais, que indicam se aquela observação se encontra acima ou abaixo de  $M_0$ . Usualmente, zero é atribuído a observações abaixo de  $M_0$ , enquanto um é atribuído a observações acima, já no caso em que o valor é igual a  $M_0$  esta observação é desconsiderada. Caso a hipótese nula seja verdadeira, isto é,  $M = M_0$ , é esperado que o número de unidades amostrais acima de  $M_0$  seja igual ao número de observações abaixo. Por conseguinte, a estatística teste é definida como a soma de todos os valores acima de  $M_0$ , e, sob  $H_0$ , esta segue uma distribuição binomial para  $n$  ensaios de Bernoulli, cada um com probabilidade de 0,5 de ser um sucesso. As regras de decisão, considerando um nível de significância  $\alpha \in (0, 1)$ , são:

Para a hipótese bilateral, rejeitamos  $H_0$  caso o valor amostral para a estatística teste seja maior do que o quantil  $1 - \alpha/2$  da distribuição binomial( $n, 1/2$ ) ou caso seja menor do que o quantil  $\alpha/2$  da mesma distribuição.

Para a hipótese unilateral à esquerda, rejeitamos  $H_0$  caso o valor obtido para a estatística teste seja menor do que o quantil da distribuição binomial( $n, 1/2$ ).

Por fim, para a hipótese unilateral à direita, rejeitamos  $H_0$  caso o valor obtido para a estatística teste seja maior do que o quantil 1 da distribuição binomial( $n, 1/2$ ).

Vale mencionar que o nível de significância em um teste determina a probabilidade de cometermos um erro do tipo I – rejeitar a hipótese nula no caso em que esta é verdadeira – e, em geral, assume-se níveis significância de 0,01, 0,05 ou 0,1. Neste estudo, é utilizado 0,05 como nível de significância.

## 2.2 Análise de sobrevivência

O campo de estudo da análise de sobrevivência engloba uma variedade de métodos estatísticos utilizados para descrever, analisar e modelar, como variável resposta, o tempo até a ocorrência de determinado evento (Colosimo, 2024). De início, é importante destacar que uma das principais características dos dados de sobrevivência é a censura, a qual representa uma observação parcial ou incompleta. Apesar de existirem diversas formas de censura, neste estudo, é abordada apenas a censura à direita do tipo I, que é determinada quando, até o final do período de acompanhamento, o evento não ocorreu para a unidade amostral em questão.

Com base nessas considerações, em uma amostra de tamanho  $n$ , definimos os dados de sobrevivência para uma observação  $i$ , tal que  $i = 1, \dots, n$ , como um vetor  $(t_i, \delta_i)$ , em que  $t_i$  representa o tempo de ocorrência do evento ou da censura e  $\delta_i$  é uma variável indicadora que assume valor zero, se a observação foi censurada, e valor um, caso o evento tenha ocorrido. Quanto à inclusão de covariáveis, a cada observação  $i$  corresponde um vetor  $\mathbf{x}_i$  de comprimento igual ao número de covariáveis e com seus respectivos valores para a observação em questão, de forma que os dados passam a ser representados pelas variáveis  $t_i$  e  $\delta_i$  e pelo vetor  $\mathbf{x}_i$ .

Em seguida, na Subseção 2.2.1, é apresentado o estimador não paramétrico de Kaplan-Meier. Nas Subseções 2.2.2 e 2.2.3, são apresentados métodos de modelagem para dados e sobrevivência. Por fim, na Subseção 2.2.4, são elencadas técnicas de análise de diagnóstico.

### 2.2.1 Estimador de Kaplan-Meier

Em linha com o panorama do campo de estudo da análise de sobrevivência brevemente delineado, tem-se que a definição do tempo de sobrevivência é uma peça central para que possamos nos aprofundar nesse tópico. Dessa forma, considere o tempo de sobrevivência como sendo uma variável aleatória  $T$  não negativa que representa o tempo em que o evento sob estudo ocorreu para determinada unidade amostral. Partindo desse princípio, considera-se que esta variável  $T$  é analisada, sobretudo, por meio da função de sobrevivência.

Em primeiro lugar, definimos a função de sobrevivência, denotada por  $S(t)$ , como a probabilidade do evento não acontecer até determinado tempo  $t$ , o que é expresso matematicamente pela equação (2.1), em que  $F(t)$  representa a função de distribuição (acumulada) de  $t$ .

$$S(t) = (T > t) = 1 - F(t). \quad (2.1)$$

A função de sobrevivência empírica, na ausência de censuras, é dada por:

$$\hat{S}(t) = \frac{\text{n}^\circ \text{ de observações que não sofreram o evento até o tempo } t}{\text{n}^\circ \text{ total de observações no estudo}}.$$

Neste contexto, o estimador de Kaplan-Meier (Kaplan e Meier, 1958) é uma forma não paramétrica de estimar a função de sobrevivência em casos em que o conjunto de dados apresenta censuras à direita. Assim, para  $j = 1, \dots, k$ ,  $t_1 \leq t_2 \leq \dots \leq t_k$  como sendo  $k$  tempos distintos e ordenados, considere  $n_j$  como o número de unidades amostrais sob risco de sofrerem o evento no tempo  $t_j$ , e  $d_j$  o número de ocorrências do evento registradas em  $t_j$ . O estimador de Kaplan-Meier, portanto, é calculado pela equação (2.2),

$$\hat{S}(t) = \prod_{j:j \leq k} \left( \frac{n_j - d_j}{n_j} \right). \quad (2.2)$$

Ao se tratar do estimador em (2.2), algumas considerações são importantes. Em relação às suas propriedades, por ser considerado um estimador de máxima verossimilhança de  $S(t)$ , tem-se que este é consistente e, portanto, não-viciado para grandes amostras, além de possuir normalidade assintótica, uma vez estabelecido que as observações são independentes e identicamente distribuídas.

Com base na propriedade de convergência para uma distribuição normal, é possível

definir um intervalo de  $100\gamma\%$  de confiança para o estimador, em que  $\gamma$  representa o nível de confiança, isto é,  $1 - \alpha$ , o nível de significância. Este intervalo é calculado conforme a equação (2.3):

$$IC(S(t), \gamma) = \hat{S}(t) \pm z_{(1-\alpha)/2} \sqrt{\widehat{\text{Var}}(\hat{S}(t))}, \quad (2.3)$$

em que  $\hat{S}(t)$  é dado pela Equação (2.2),  $z_{(1-\alpha)/2}$  é o quantil  $(1 - \alpha)/2$  da distribuição normal padrão. A variância do estimador, por sua vez, é estimada pela fórmula (2.4),

$$\widehat{\text{Var}}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}. \quad (2.4)$$

É importante destacar que, embora o estimador de Kaplan-Meier não incorpore covariáveis diretamente, é possível observar o comportamento da função de sobrevivência estimada para diferentes níveis de variáveis qualitativas ou de variáveis quantitativas categorizadas.

Em síntese, o estimador de Kaplan-Meier é uma importante ferramenta descritiva para analisar o comportamento do tempo de sobrevivência amostral, dado que nos ajuda a observar a evolução temporal da probabilidade de ocorrência do evento.

### Teste de *logrank*

Um aspecto importante ao ser observado nas curvas de Kaplan-Meier é se, para diferentes categorias de determinada covariável sob estudo, as curvas podem ser consideradas suficientemente diferentes de forma que a inclusão desta covariável no modelo seja útil.

Para determinar se a diferença entre duas curvas é significativa, podemos utilizar o teste de *logrank* (Mantel, 1966), no qual a estatística teste é dada pela diferença entre o número de falhas obtidas em cada grupo e o número esperado sob a hipótese nula de que todas as curvas de sobrevivência são iguais.

Para uma melhor definição deste teste, considere que temos  $r$  curvas de sobrevivência que gostaríamos de comparar a fim de determinar se as curvas são iguais ou distintas. Neste caso, os dados podem ser organizados em uma tabela de contingência com duas linhas e  $r$  colunas, de modo que para a coluna  $i$ , na primeira linha, teríamos o número de falhas  $d_{ij}$  no tempo  $j$  e, na segunda linha, teríamos o número de sobreviventes  $n_{ij} - d_{ij}$ . Assim, a distribuição conjunta de todos os tempos de falha em determinado tempo  $j$  é dada por uma distribuição hipergeométrica multivariada, cuja função de probabilidade é

dada por:

$$\frac{\prod_{i=1}^r \binom{n_{ij}}{d_{ij}}}{\binom{n_j}{d_j}}.$$

A estatística teste, denotada por  $W$ , sob essas circunstâncias é dada pela equação (2.5):

$$W = \mathbf{v}' V^{-1} \mathbf{v}, \quad (2.5)$$

cuja distribuição, sob a hipótese nula de que as curvas são iguais, é qui-quadrado com  $r-1$  graus de liberdade e em que o vetor  $\mathbf{v}$  é dado pela seguinte equação:

$$\mathbf{v} = \sum_j^k \mathbf{v}_j,$$

tal que  $\mathbf{v}'_j = (d_{2j} \quad w_{2j}, \dots, d_{rj} \quad w_{rj})$ , dado que  $w_{ij} = n_{ij} d_{ij} n_j^{-1}$  é a média de  $d_{ij}$ . Por sua vez, a matriz  $V$ , representa a variância da estatística  $\mathbf{v}$ .

Vale ainda mencionar que, caso a hipótese nula de que as curvas de sobrevivência são iguais for rejeitada, é necessário fazer comparações múltiplas a fim de identificar quais grupos diferem e, para isso, utiliza-se algum método para controlar o erro tipo I, em geral, o método de Bonferroni (Bonferroni, 1936).

### 2.2.2 Modelo de Cox para riscos proporcionais

No centro das metodologias que juntas constituem a análise de sobrevivência, um papel central é atribuído à modelagem do problema e à interpretação deste a partir da inclusão de covariáveis, ou variáveis explicativas. Entre os modelos de sobrevivência que podem ser utilizados, o modelo de regressão semi-paramétrico de Cox assume uma posição de destaque.

Também conhecido como modelo de riscos proporcionais de Cox, este é utilizado para modelar a função de risco, denotada por  $h(t)$ , e, naturalmente, admite covariáveis (Cox, 1972). A função de risco, ou taxa de falha, representa o risco instantâneo de uma unidade amostral sofrer o evento de interesse em determinado tempo  $t$  e é condicional à sobrevivência até o tempo  $t$ . Em linhas gerais, a função de risco é dada pela fórmula (2.6),

caso a variável aleatória para o tempo de sobrevivência  $T$  seja contínua, e pela fórmula (2.7), caso  $T$  seja discreta, considerando  $S(t)$  como sendo a função de sobrevivência. As duas fórmulas são dadas por:

$$h(t) = \lim_{t \rightarrow 0} \frac{[t \leq T \leq t + t, T \geq t]}{tS(t)}, \quad (2.6)$$

e

$$h(t) = [T = t | T \geq t] = \frac{[T = t]}{S(t) - 1}. \quad (2.7)$$

Para definirmos o modelo de Cox, considere uma amostra de tamanho  $n$  e um conjunto de  $p$  covariáveis, de modo que a  $i$ -ésima unidade amostral esteja associada a um vetor de covariáveis  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ . Além disso, considere também o vetor de coeficientes  $\boldsymbol{\beta}$  de dimensão  $p$ , e  $h_0(t)$  como sendo a função de taxa de falha basal comum a todos as unidades amostrais, isto é,  $h_0(t)$  é igual a  $h(t|x)$  quando  $x = 0$ . Assim, o modelo de Cox é dado pela fórmula (2.8),

$$h(t|\mathbf{x}_i) = h_0(t)\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}. \quad (2.8)$$

De modo geral, o modelo de Cox é formado por duas partes: uma parte paramétrica em que os parâmetros associados ao modelo são estimados, e uma parte não-paramétrica em que temos a taxa de risco de base não especificada. Para aplicar o modelo de Cox a um problema, é importante que algumas suposições sejam satisfeitas. Em primeiro lugar, é necessário que as taxas de falha de dois indivíduos diferentes sejam proporcionais para quaisquer pares de unidades amostrais sob estudo e, em segundo lugar, é necessário que as observações sejam independentes. É importante ressaltar que esse modelo aborda somente casos com censura à direita e não incorpora demais tipos de censura.

Em relação ao processo de estimação dos parâmetros do modelo, utiliza-se comumente o método da máxima verossimilhança parcial, no qual estamos interessados em encontrar o valor mais provável, ou mais verossímil, para os parâmetros dado um conjunto de observações. No entanto, neste caso, o método é considerado parcial dado que não levamos em consideração o componente não paramétrico, isto é, a taxa de risco de base, para a construção do método de estimação dos parâmetros. Resumidamente, a função de verossimilhança parcial é dada por:

$$\begin{aligned}
L(\boldsymbol{\beta}|\mathbf{x}) &= \prod_{i=1} \frac{\left( \text{u.a. com } \mathbf{x}_i \text{ sofrer o evento em determinado tempo } t_j \right)}{\left( \text{uma ocorrência em } t_j \right)} \Big)^{\delta_i} \\
&= \prod_{i=1} \left( \frac{h_i(t_j|\mathbf{x}_i)}{\sum_{k \in R(t_j)} h_k(t_j|\mathbf{x}_k)} \right)^{\delta_i} \\
&= \prod_{i=1} \left( \frac{h_0(t_j) \exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}{\sum_{k \in R(t_j)} h_0(t_j) \exp\{\mathbf{x}'_k \boldsymbol{\beta}\}} \right)^{\delta_i} \\
&= \prod_{i=1} \left( \frac{\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}{\sum_{k \in R(t_j)} \exp\{\mathbf{x}'_k \boldsymbol{\beta}\}} \right)^{\delta_i}.
\end{aligned}$$

Em seguida, para achar o ponto de máximo do logaritmo da função de verossimilhança parcial, ou seja, resolvemos a equação (2.9):

$$\frac{\partial \ln(L(\boldsymbol{\beta}|\mathbf{x}))}{\partial \boldsymbol{\beta}} = 0. \quad (2.9)$$

Ao todo, deverá ser solucionado um sistema com  $p$  equações, uma para cada parâmetro do modelo, o que é comumente feito pelo método iterativo de Newton-Raphson (Akram e Ann, 2015). Note que a informação sobre os tempos está embutida no grupo de risco, de modo que eles não entram diretamente na estimação dos parâmetros, sendo considerados explicitamente apenas na componente não paramétrica do modelo.

Já para estimarmos a parte não paramétrica do modelo, ou seja, a função de risco de base, utilizamos a estimativa proposta por Breslow (1972) e dada pela equação (2.10):

$$\hat{h}_0(\mathbf{t}) = \sum_{j:t_j < t} \left( \frac{d_j}{\sum_{k \in R_k} \exp\{\mathbf{x}'_k \hat{\boldsymbol{\beta}}\}} \right), \quad (2.10)$$

em que  $d_j$  representa o número de ocorrências em  $t_j$ .

Por fim, vale destacar que uma das principais vantagens de se utilizar o modelo de Cox é a sua fácil interpretabilidade a partir da razão das taxas de falha. Suponha-se, por exemplo, duas observações,  $g$  e  $v$ , que compartilham os mesmos valores para todas as covariáveis, exceto por uma covariável binária  $x_f$ . Para essas duas unidades amostrais, a razão entre as taxas de falha é dada por:

$$\frac{\hat{h}(t|\mathbf{x}_g)}{\hat{h}(t|\mathbf{x}_v)} = \exp\{\hat{\beta}_f\},$$

ou seja, a taxa de risco para observações que assumem o valor  $x_f = 1$  é multiplicada

por  $\exp\{\hat{\beta}_f\}$  vezes em relação às unidades amostrais cujo valor para  $x_f$  é 0, mantendo-se as demais variáveis constantes. De modo análogo, essa interpretação se estende a outras variáveis, independentemente de serem categóricas com múltiplos níveis ou contínuas.

### 2.2.3 Modelo de Prentice, Williams e Peterson

A escolha das técnicas de modelagem para dados de sobrevivência depende exclusivamente da natureza dos dados à nossa disposição e, em grande parte dos problemas com dados de sobrevivência, tem-se que o interesse não é abordar um único evento que só pode acontecer uma vez a cada unidade amostral, mas sim abordar diferentes tipos de eventos relacionados a um único fator sob estudo. Este é o caso de eventos múltiplos ordenados e estruturados, em que são selecionados alguns eventos de interesse e assumimos que a unidade amostral está em risco de sofrer o  $i$ -ésimo evento quando o evento anterior já ocorreu, e considera-se também que o risco de uma observação sofrer um evento difere do risco de sofrer outro evento (Carvalho *et al.*, 2011). Exemplos de eventos deste tipo são a ocorrência de infartos do miocárdio, a recorrência de internações hospitalares, a incidência de lesões a determinado ligamento em esportistas, e, como é o caso deste trabalho, a ocorrência de perdas de diferentes magnitudes em determinado investimento (Cabete, 2012).

Entre os modelos de sobrevivência para eventos múltiplos, destacam-se as seguintes classes, melhor descritas nos livros Colosimo (2024) e Carvalho *et al.* (2011): os modelos de riscos competitivos, nos quais a ocorrência de um evento exclui a possibilidade de ocorrência de outros eventos; os modelos de fragilidade, em que é incluída uma variável, chamada de variável de fragilidade, para captar efeitos não mensuráveis nas unidades amostrais, sejam estas individuais ou para dados agrupados; e os modelos marginais, que se baseiam na expansão do modelo de Cox e são utilizados em uma ampla variedade de casos em que os eventos são recorrentes. Com base na natureza dos dados sob estudo, esta última classe de modelos aparenta ser a mais adequada.

Em relação aos modelos marginais para dados ordenados e estruturados, o modelo de Prentice, Williams e Peterson (Prentice *et al.*, 1981), também chamado de modelo PWP, é o que mais se destaca. Neste modelo, assumimos que os eventos seguem determinada ordem de ocorrência e que seus riscos diferem entre si. As diferenças entre o modelo de Cox e o modelo PWP considerando o estudo de  $k$  eventos sequenciais são mais bem ilustradas na Figura 2.1.

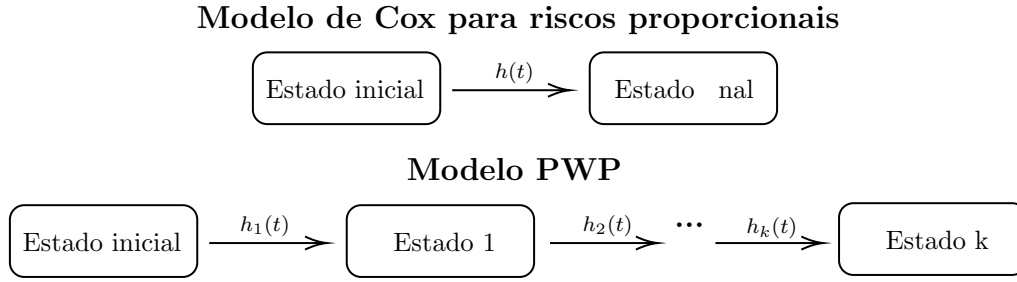


Figura 2.1: Representações ilustrativas dos modelos PWP e Cox para riscos proporcionais. Elaboração própria baseada em [Carvalho \*et al.\* \(2011\)](#).

Desse modo, para definirmos em termos matemáticos, considere  $j$  como sendo um dos  $k$  eventos selecionados para serem estudados, e  $\mathbf{x}_i$  como sendo o vetor de covariáveis para a  $i$ -ésima unidade amostral. O modelo de riscos proporcionais adaptado para este caso assume a forma:

$$h_{ij}(t|\mathbf{x}_i) = Y_{ij}(t)h_{0j}(t)\exp\{\mathbf{x}_i'\boldsymbol{\beta}\}, \quad (2.11)$$

em que  $Y_{ij}(t)$  é uma variável indicadora que assume o valor zero até que o evento  $j$  ocorra e, a partir do momento em que este ocorre, a variável assume o valor 1. Além disso,  $h_{0j}$  representa o risco basal para o evento  $j$ .

Vale citar que há duas variações deste modelo: uma em que apenas o risco basal se altera entre eventos de forma que o vetor de parâmetros  $\boldsymbol{\beta}$  permanece o mesmo para todos os eventos, e uma segunda variação em que ambos se alteram entre eventos.

No que diz respeito à estimação dos parâmetros, assim como foi realizado no modelo de Cox, no modelo PWP, faremos uso da função de verossimilhança parcial para estimar a parte paramétrica do modelo. Neste caso, a função é dada pela equação 2.12, tal que  $n$  representa o número de unidades amostrais e  $s$  representa um estado dos  $k$  estados pertencentes ao espaço  $S$ .

$$L(\boldsymbol{\beta}|\mathbf{x}) = \prod_{i=1}^n \prod_{s \in S} \left( \frac{\exp\{\mathbf{x}_i'\boldsymbol{\beta}_s\}}{\sum_{l=1}^n Y_{sl} \exp\{\mathbf{x}_l'\boldsymbol{\beta}_s\}} \right) \quad (2.12)$$

De forma similar ao que é realizado na estimação do modelo de Cox, a maximização da função de verossimilhança parcial pode ser feita por meio do método de Newton-Raphson. Já a estimação da parte não paramétrica do modelo é feita nos mesmos moldes do modelo de Cox.

## 2.2.4 Análise de Diagnóstico

Após o ajuste dos modelos, a análise de diagnóstico se mostra uma etapa fundamental, pois é nesta que podemos avaliar a qualidade do modelo no que tange a adequação às suposições e a sua capacidade de se moldar aos dados.

Neste trabalho, quanto às técnicas de análise de diagnóstico, foram utilizados o resíduo de Schoenfeld e o resíduo martingal, os quais estão detalhados brevemente abaixo.

### Resíduo de Schoenfeld

O resíduo de Schoenfeld é uma das principais ferramentas utilizadas para avaliar a suposição de proporcionalidade das curvas de sobrevivência presente no modelo de Cox, tanto a nível global quanto para cada covariável (Schoenfeld, 1982).

Considere que, para determinado indivíduo  $i$ , atribuímos um vetor de covariáveis  $\mathbf{x}_i$  com dimensão  $p$ . Para este indivíduo é atribuído um vetor de resíduos de Schoenfeld,  $\mathbf{r}_i$ , também de dimensão  $p$ , de forma que para determinada covariável  $q$  seja atribuído um valor para o resíduo dado por:

$$r_{iq} = x_{iq} \frac{\sum_l^n x_{lq} \exp\{\mathbf{x}'_l \hat{\boldsymbol{\beta}}\}}{\sum_l^n \exp\{\mathbf{x}'_l \hat{\boldsymbol{\beta}}\}},$$

tal que os resíduos sejam definidos apenas para falhas, e não para censuras. Contudo, a forma mais utilizada dos resíduos de Schoenfeld é a padronizada, que é definida por:

$$\mathbf{r}_i^* = [I(\hat{\boldsymbol{\beta}})]^{-1},$$

sendo  $I(\hat{\boldsymbol{\beta}})$  a matriz de informação empírica.

Uma técnica gráfica para avaliar a proporcionalidade das curvas de sobrevivência por meio deste resíduo, proposta por Terry M. Therneau (2000), consiste no gráfico de  $r_{iq}^* + \hat{\beta}_q$  versus o tempo. Dado que, quando a suposição de proporcionalidade é satisfeita, o valor do parâmetro  $\beta_q$  deve permanecer invariante ao longo do tempo, um gráfico com inclinação zero mostra indícios a favor da suposição de proporcionalidade.

### Resíduo martingal

Por fim, na análise de diagnóstico, o resíduo martingal se destaca por sua capacidade de identificar *outliers* e avaliar qual é a melhor forma funcional de determinada variável,

por exemplo, se esta deve ser utilizada em sua forma quadrática, logarítmica ou até mesmo se deve ser categorizada, além de outras formas possíveis. O cálculo deste resíduo, para determinado indivíduo  $i$ , é dado pela equação (2.13):

$$\hat{m}_i = \delta_i \hat{e}_i, \quad (2.13)$$

em que  $\delta_i$  é uma variável que indica se a observação é uma censura e  $\hat{e}_i$  representa o resíduo de Cox e Snell (1968), dado por:

$$\hat{e}_i = \hat{\Lambda}_0(t_i) \exp \left\{ \sum_q^p x_{iq} \hat{\beta}_q \right\},$$

de forma que  $\hat{\Lambda}_0$  representa a função de taxa de falha basal acumulada, por sua vez calculada como:

$$\hat{\Lambda}_0(t) = \int_0^t \lambda_0(u) du.$$

Ao observar o gráfico de pontos formados determinados pelo vetor  $(x_{iq}, \hat{m}_i)$  para cada indivíduo  $i$  e covariável  $q$ , uma nuvem de pontos sem tendência indica que nenhuma transformação é necessária à forma da variável, enquanto que pontos distantes do aglomerado de pontos podem indicar valores atípicos.

# Capítulo 3

## Material

Neste capítulo, encontra-se, na Seção 3.1, uma breve descrição de como o conjunto de dados foi elaborado, calculado e compilado, e no Apêndice A podem ser encontrados detalhes adicionais sobre as considerações feitas durante a construção da base de dados. Para acessar a base de dados completa, acesse <https://github.com/luizaneves/tcc>.

### 3.1 Descrição do conjunto de dados

A construção da base de dados foi feita em linha com o objetivo do estudo de avaliar se os investimentos em *IPOs* têm rentabilidade superior, inferior ou equivalente ao mercado acionário em geral. Assim, durante o estudo o valor a ser mensurado e analisado é a rentabilidade de cada *IPO* em relação a um referencial de mercado, neste caso, o Ibovespa, durante um horizonte de tempo determinado.

Para tanto, foram observadas as cotações diárias (B3, 2024a) dos *IPOs* ocorridos desde 2004 a 2023 em um período de até dois anos de sua estreia. Paralelamente, também foi acompanhada a cotação do Ibovespa correspondente ao período de acompanhamento de cada empresa. O índice Bovespa foi escolhido por ser a referência mais utilizada para avaliar o desempenho do mercado acionário brasileiro e por ser constituído pelas empresas que correspondem por 80% do número de negócios e do volume negociado no mercado, assim, é possível considerá-lo um bom representante para o desempenho do mercado acionário como um todo (B3, 2024). Vale mencionar que foram consideradas a distribuição de proventos e a ocorrência de desdobramentos e agrupamentos de ações no cálculo da rentabilidade, conforme melhor elucidado no apêndice A.

Em suma, foram simulados dois investimentos com valores iniciais iguais simultanea-

mente; um deles em um *IPO* e o outro no Ibovespa. Ao final, o interesse é observar a diferença da performance de um em relação ao outro.

Quanto ao cálculo da rentabilidade relativa de cada oferta pública inicial, partiremos do pressuposto que na data da oferta foi investida a mesma quantia tanto no *IPO* quanto no índice Bovespa, de forma que no período de  $w$  dias depois do *IPO* a rentabilidade relativa é dada por:

$$\text{Rent. relativa no período } w = \frac{(\text{Rent. do investimento no } IPO \text{ no período } w) + 1}{(\text{Rent. do investimento no Ibovespa no período } w) + 1} \quad 1,$$

de modo que ambas as rentabilidades, do *IPO* e do Ibovespa, sejam calculadas da seguinte forma:

$$\text{Rentabilidade} = \frac{\text{Valor final do investimento}}{\text{Valor inicial investido}} \quad 1.$$

Note que como consideramos os valores iniciais dos investimentos como sendo iguais em ambos, estes se cancelam na fórmula da rentabilidade relativa. Logo, esta equação pode ser simplificada por:

$$\text{Rentabilidade relativa no período } w = \frac{\text{Valor do investimento no } IPO \text{ na data } w}{\text{Valor do investimento no Ibovespa na data } w} \quad 1.$$

Adicionalmente, o período de até dois anos foi escolhido por ser representativo tanto do curto prazo – por exemplo, de um dia, uma semana e um mês – quanto do médio e longo prazo – como seis meses, um ano e dois anos –, de forma que estes sejam razoavelmente longos a ponto de dar à empresa tempo para divulgar resultados após o *IPO* e para o mercado ajustar suas expectativas estabelecidas na data da oferta, ao passo que não é um tempo tão longo para que as considerações feitas na oferta inicial já não tenham tanto peso na avaliação da ação pelos investidores. Vale mencionar que durante o estudo utilizaremos o conceito de dias úteis para evitar distorções provocadas por feriados existentes e demais dias em que não se tem negociações nos períodos de dias corridos. Assim, adotamos o padrão existente no mercado financeiro que utiliza o conceito de 5 dias úteis como representativo de uma semana, 21 dias úteis para um mês e 252 dias úteis para um ano.

Dessa forma, a base de dados é composta pelas rentabilidades relativas ao Ibovespa de 220 *IPOs*<sup>1</sup>, registradas em cada um dos primeiros 504 dias desde sua oferta, o que corresponde a dois anos. As empresas incluídas no estudo estão distribuídas por ano de

---

<sup>1</sup>Em alguns casos, o acompanhamento das cotações dos *POs* durante todo o período foi impossibilitado, seja por fusões, falências ou outros motivos. Para mais informações, consulte o Apêndice A.

estreia conforme ilustrado na Tabela 3.1.

Tabela 3.1: Distribuição dos *IPOs* por ano de lançamento.

Anos	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Número de <i>IPOs</i>	7	8	23	55	3	4	10	11	2	9
Anos	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Número de <i>IPOs</i>	1	1	1	8	3	5	27	42	0	0

Ademais, para aplicarmos as técnicas inferenciais propostas, os dados foram interpretados como dados de sobrevivência. Desse modo, as cotações diárias de cada ação foram vistas sob a perspectiva de um problema de dados longitudinais, segundo a qual cada empresa foi analisada considerando o tempo até que ocorresse determinado evento de interesse, demarcado por uma perda de rentabilidade do investimento relativa ao Ibovespa, por exemplo um prejuízo relativo de 20%. Portanto, sob o ótica de dados de sobrevivência, a base de dados é composta por duas colunas; a primeira é um vetor com os tempos de falha ou de censura, enquanto a segunda é composta por uma variável indicadora que assume valor 0 quando a observação foi censurada e valor 1 quando o evento ocorreu. Essa abordagem permite explorar o risco de ocorrerem determinados níveis de perda na rentabilidade das ações ao longo do tempo.

Informações mais detalhadas acerca da construção da base de dados bem como um recorte dos dados utilizados podem ser encontradas no Apêndice A.

### 3.1.1 Inclusão de covariáveis

Além de analisar o comportamento dos retornos em investimentos em ofertas públicas iniciais por si só, buscamos compreender quais covariáveis estariam relacionadas a diferentes comportamentos desta rentabilidade, sobretudo no longo prazo. Quanto à seleção das covariáveis que foram incluídas no estudo, foram consideradas as teorias formuladas acerca do fenômeno de baixa rentabilidade no longo prazo citadas em [Ibbotson e Ritter \(1995\)](#), [Ritter \(1991\)](#) e [Shiller \(1990\)](#).

A primeira teoria, proposta por [Shiller \(1990\)](#), sugere que os investimentos em *IPOs* são influenciados por um fenômeno de modismo. De acordo com essa hipótese, os retornos iniciais tendem a ser excessivamente altos, espelhando expectativas e um otimismo elevados por parte dos investidores. Contudo, à medida que as empresas começam a divulgar

seus resultados e essas expectativas não se concretizam, os retornos tendem a se reverter, levando a desempenhos negativos no longo prazo. Para testar essa teoria, propomos avaliar se as empresas com os maiores retornos no curto prazo, em específico após um dia da oferta, também enfrentam os maiores riscos de sofrerem retornos negativos posteriormente. Nesse sentido, incluímos uma covariável categórica ordinal que indica o nível de retorno de curto prazo para determinada ação e pode assumir três valores: baixo, para retornos menores que -1,07%, moderado, para retornos entre -1,07% e 6,83%, e alto, para retornos acima de 6,83%. Vale destacar que as classes foram selecionadas de modo que cada categoria tenha um número igual de observações, garantindo classes balanceadas, o que contribui para estimativas mais confiáveis durante a aplicação de modelos estatísticos.

Uma segunda vertente de pensamento, proposta em [Ritter \(1991\)](#), parte do pressuposto de que os mercados passam por ciclos de euforia e depressão. Assim, as empresas que querem arrecadar fundos por meio de um *IPO* escolhem fazer suas ofertas iniciais durante os períodos, ou janelas de oportunidade, em que o mercado está mais aquecido e quando o apetite por risco por parte dos investidores está maior, sobretudo, como uma estratégia para fazer ofertas a preços excessivamente altos. Eventualmente, o ciclo de mercado é revertido e o comportamento oposto ao que foi observado no curto prazo é observado em períodos mais longos. Dessa forma, incluímos uma covariável binária que indica se o *IPO* foi emitido em um período de euforia do mercado ou não, como realizado em estudos como [Barihas et al. \(2024\)](#). Para definir esta variável foi observado o volume de *IPOs* ocorridos, ilustrado na Tabela 3.1, de forma que os períodos com maior volume, isto é, os anos de 2006, 2007, 2020 e 2021, foram classificados como períodos dentro da janela de oportunidade, enquanto os demais períodos foram classificados como não pertencentes a nenhuma janela de oportunidade.

Ademais, outras covariáveis foram investigadas, como o setor de atuação e o tamanho da oferta. Embora elas não estejam diretamente relacionadas a teorias existentes, essas variáveis foram facilmente acessíveis e, portanto, incluídas com o objetivo de explorar possíveis efeitos no retorno dos *IPOs* ao longo do tempo. O tamanho da oferta, em milhões de reais, foi corrigido pelo IPCA ([IBGE, 2025](#)), tendo como referência os valores de dezembro de 2023 e, além disso, vale mencionar que trata-se de uma variável contínua, enquanto a variável setor é categórica com 10 categorias. Na Tabela 3.2, é possível observar quais são os setores e como as empresas sob estudo estão distribuídas entre eles.

Tabela 3.2: Distribuição dos *IPOs* por setor.

<b>Setor</b>	<b>Número de <i>POs</i></b>
Bens Industriais	24
Comunicações	5
Consumo Cíclico	66
Consumo não Cíclico	18
Financeiro	37
Materiais Básicos	5
Petróleo, Gás e Biocombustíveis	11
Saúde	24
Tecnologia da Informação	17
Utilidade Pública	13



# Capítulo 4

## Resultados

Neste capítulo, estão documentados os resultados obtidos a partir do emprego das metodologias descritas no Capítulo 2 ao conjunto de dados. Iniciamos o estudo por meio de uma análise descritiva e exploratória dos dados, na Seção 4.1, a qual encontra-se dividida em duas partes: na primeira parte, foi realizada a análise exclusivamente da variável resposta, na Subseção 4.1.1; enquanto que na segunda parte, na Subseção 4.1.2, foi feita a análise descritiva incluindo covariáveis. Em seguida, na Seção 4.2 aplicamos testes de hipóteses não-paramétricos a fim de elucidar algumas das questões centrais acerca do comportamento da variável resposta. Por fim, na Seção 4.3, encontram-se os resultados obtidos por meio da aplicação dos modelos de sobrevivência aos dados.

Os resultados apresentados neste capítulo foram desenvolvidos por meio do *software* R (R Core Team, 2023). Mais detalhes sobre a aplicação e códigos utilizados podem ser encontrados no Apêndice B.

### 4.1 Análise descritiva e exploratória

A análise descritiva e exploratória dos dados é uma etapa fundamental para podermos compreender o comportamento dos dados, como a presença de padrões, tendências e anomalias, o que nos ajuda a determinar quais técnicas estatísticas são mais apropriadas para uma análise mais aprofundada posteriormente. Primeiramente, na Subseção 4.1.1, foi realizada somente a análise da variável resposta. Já na Subseção 4.1.2, foram incorporadas as variáveis explicativas à análise.

### 4.1.1 Comportamento da variável resposta

De início, foi analisado o comportamento da rentabilidade dos *IPOs* relativa ao Ibovespa, em seis horizontes de tempo desde a oferta inicial; nos períodos de um dia, uma semana, um mês, seis meses, um ano e dois anos. Começamos a análise por meio dos *boxplots* encontrados na Figura 4.1, em que temos todos os seis períodos no gráfico (A) e os três primeiros períodos separados dos três últimos, nos gráficos (B) e (C), para facilitar a visualização. Nos *boxplots*, observamos que a variabilidade, representada no gráfico pela amplitude das caixas, aumenta conforme aumentamos o tempo desde a oferta, o que é esperado, pois intervalos de tempo maiores proporcionam mais oportunidades para que sejam alcançados níveis mais extremos de rentabilidades, tanto positivas quanto negativas.

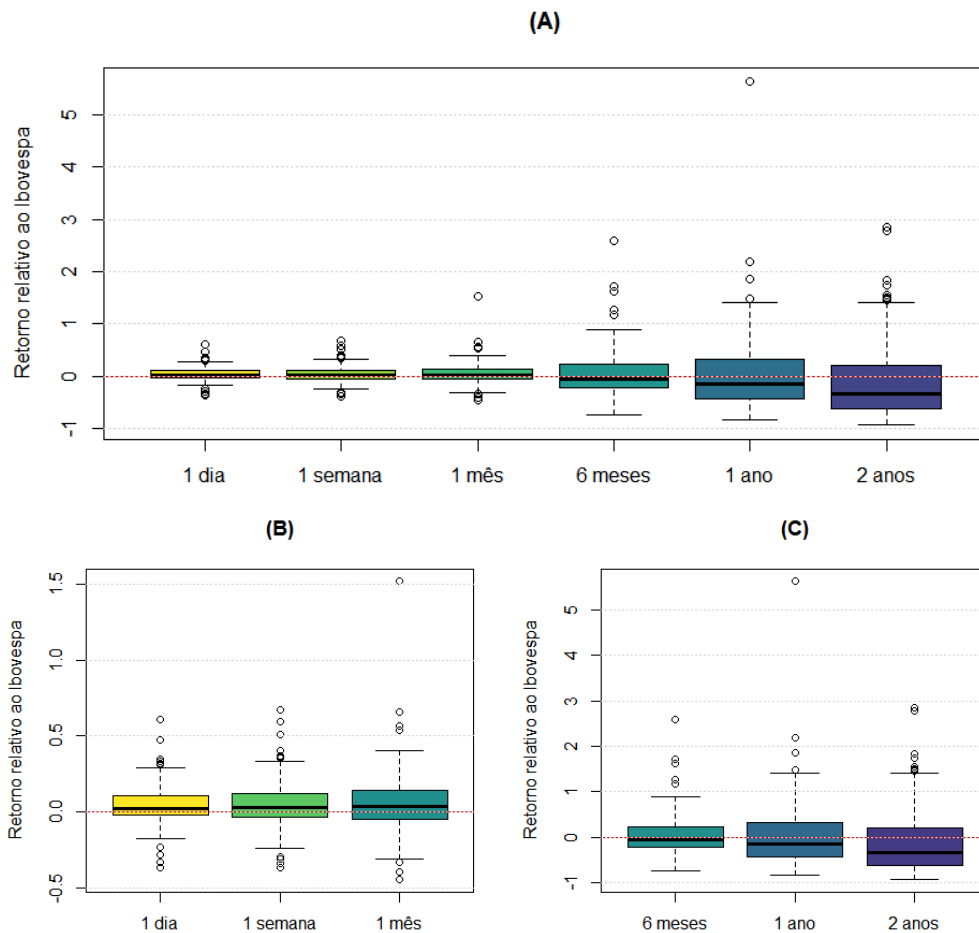


Figura 4.1: Comportamento dos retornos relativos ao Ibovespa dos investimentos em *IPOs*, em números absolutos, para todos os períodos (gráfico (A)), para os três primeiros períodos isoladamente (gráfico (B)), e para os três últimos períodos (gráfico (C)).

Adicionalmente, também é possível observar similaridades entre os períodos; em todos os casos, os dados possuem uma assimetria positiva, o que confere também um número

maior de *outliers* positivos em todas as instâncias. Essa assimetria pode estar presente em decorrência do suporte truncado da variável resposta, a qual pertence ao intervalo de -1 a infinito.

Quando aos *outliers*, ao todo, foram detectados 10, 11, 7, 5, 4 e 9 *outliers* respectivamente em cada período e, destes, apenas os *outliers* dos quatro últimos períodos mais se destacam devido à sua magnitude. Apesar disso, era esperado que houvessem alguns retornos extremos em uma amostra de mais de 200 ações.

De maneira complementar, também é possível observar, tanto pelos *boxplots* da Figura 4.1 quanto pela Tabela 4.1, que nos três primeiros períodos há uma tendência levemente crescente na média e na mediana, enquanto nos últimos três períodos há uma reversão da tendência anterior e passamos a observar uma diminuição considerável tanto da média, que vai de 4,4% em um dia para -10% em dois anos, quanto da mediana, que parte de 2,1% em um dia para -32,3% em dois anos.

Tabela 4.1: Medidas descritivas de tendência central e de dispersão.

	1 dia	1 semana	1 mês	6 meses	1 ano	2 anos
<b>Mínimo</b>	-36,51%	-37,06%	-44,95%	-72,54%	-82,47%	-93,03%
<b>1 Quartil</b>	-2,29%	-3,75%	-4,70%	-22,22%	-42,21%	-62,02%
<b>Mediana</b>	2,07%	3,13%	3,17%	-4,73%	-14,06%	-32,33%
<b>Média</b>	4,42%	4,68%	5,39%	2,84%	0,12%	-10,02%
<b>3 Quartil</b>	10,46%	12,01%	14,10%	22,85%	33,50%	20,92%
<b>Máximo</b>	60,61%	67,20%	151,95%	259,30%	561,90%	284,86%
<b>Desvio Padrão</b>	11,79%	13,84%	19,34%	41,76%	64,82%	69,29%

Ainda em relação à Tabela 4.1, é possível observar como os períodos entre um mês e seis meses após o *IPO* constituem um marco de transição, de forma que podemos observar uma diferença marcante entre todas as medidas descritivas dos três primeiros períodos em relação aos três últimos.

Na Figura 4.2, é possível observar por uma outra perspectiva como a distribuição da variável resposta se altera ao longo do tempo. Nota-se como a assimetria tende a ficar mais acentuada conforme o tempo passa.

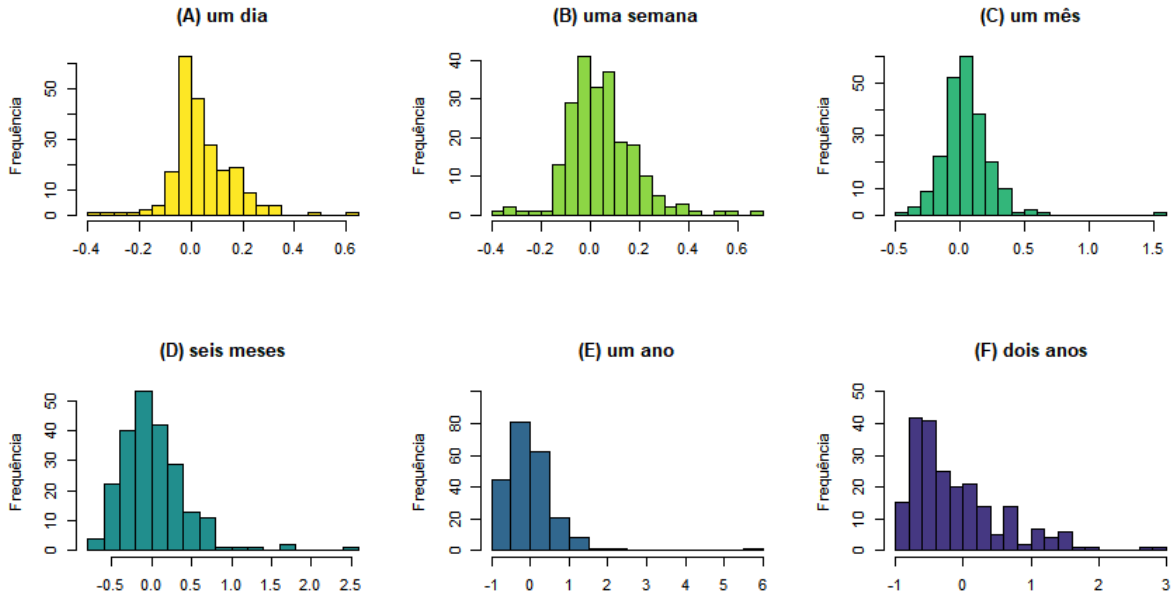


Figura 4.2: Comportamento da rentabilidade dos *IPOs* relativa ao Ibovespa para cada período.

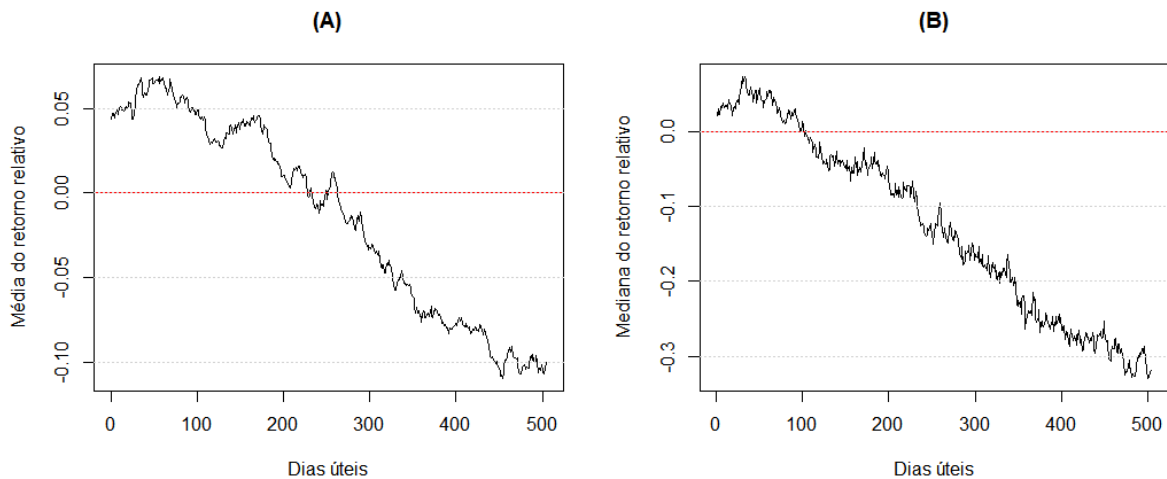


Figura 4.3: Evolução diária da média, no gráfico (A), e da mediana, no gráfico (B), do retorno dos investimentos em *IPOs* relativo ao Ibovespa.

Já nos dois gráficos apresentados na Figura 4.3, é possível observar com maior nuance como a média e a mediana dos retornos relativos ao Ibovespa se comportam para cada dia adicional após a oferta inicial. Com este e os demais resultados apresentados, é possível notar que o comportamento observado é condizente com os padrões documentados em outros estudos, como em [Aggarwal, Leal e Hernandez \(1993\)](#); [Ritter \(1991\)](#); [Ibbotson, Sindelar e Ritter \(1988\)](#) citados no Capítulo 1, uma vez que, para os períodos de curto prazo, a tendência de aumento da rentabilidade é indicativa do fenômeno de *underpricing*, enquanto, para o longo prazo, a tendência de perda de rentabilidade é indicativa do

fenômeno de baixa performance.

Ainda com base na Figura 4.3, é possível elencar possíveis patamares de perda relativa ao Ibovespa para servirem como os nossos eventos de interesse durante a aplicação de técnicas de análise de sobrevivência. Ao observar a mediana no gráfico, torna-se nítido que a definição de níveis de perda muito superiores a 30% resultaria em um número excessivo de censuras, comprometendo a eficácia da análise.

No contexto de análise de sobrevivência, na Figura 4.4, é possível observar o comportamento do estimador de Kaplan-Meier conforme o tempo passa, considerando a definição do evento de interesse como sendo ter atingido determinado nível de perda. Neste caso, foram selecionados os níveis de perda de 10%, 15%, 20%, 30%, 40% e 60% para avaliar como o estimador é sensível às mudanças na definição do evento de interesse e determinar os níveis de perda mais interessantes para o ajuste de modelos de sobrevivência posteriormente.

Ao observar o gráfico da Figura 4.4, notamos que após o nível 40% de perda o número de censuras ultrapassa 50%, o que não é ideal para análises subsequentes. Consequentemente, acreditamos que os níveis de perda de 20%, 30% e 40% representam patamares adequados para serem utilizados como eventos de interesse na análise de sobrevivência, dado que não resultam em uma quantidade muito grande de censuras e, mesmo assim, podem ser consideradas perdas substanciais.

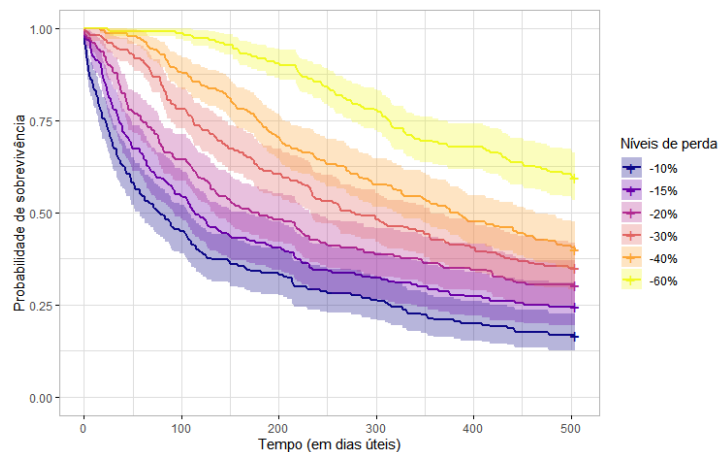


Figura 4.4: Curvas de Kaplan-Meier para diferentes definições do evento de interesse.

Em suma, esta análise inicial nos possibilitou identificar as características da amostra coletada e como esta está distribuída. É possível dizer que a amostra aparenta ser proveniente de uma distribuição não normal com uma considerável assimetria positiva, o que inviabiliza a realização de testes para a média, que dependem de distribuições normais

ou pelo menos simétricas, mas atende as suposições necessárias para o teste do sinal para a mediana. Ademais, foi observado um comportamento de aumento da rentabilidade relativa ao Ibovespa no curto prazo – até um mês – seguida por uma relevante perda de rentabilidade no mais longo prazo – de seis meses até dois anos. Por último, nesta etapa do estudo estabelecemos os níveis de perda de 20%, 30% e 40% como os eventos de interesse para o ajuste dos modelos de sobrevivência.

### 4.1.2 Comportamento da variável resposta na presença de covariáveis

Agora, incorporando as covariáveis à análise descritiva, podemos observar como a rentabilidade dos *IPOs* se comporta para cada nível das variáveis explicativas.

Em primeiro lugar, na Figura 4.5, observamos como as séries temporais da média e da mediana dos retornos relativos ao Ibovespa se comportam para *IPOs* realizados dentro e fora de períodos demarcados como janelas de oportunidade. Ao todo 73 empresas ficaram na categoria “fora da janela de oportunidade”, enquanto que as demais 147 foram consideradas pertencentes a períodos em que o mercado está mais aquecido.

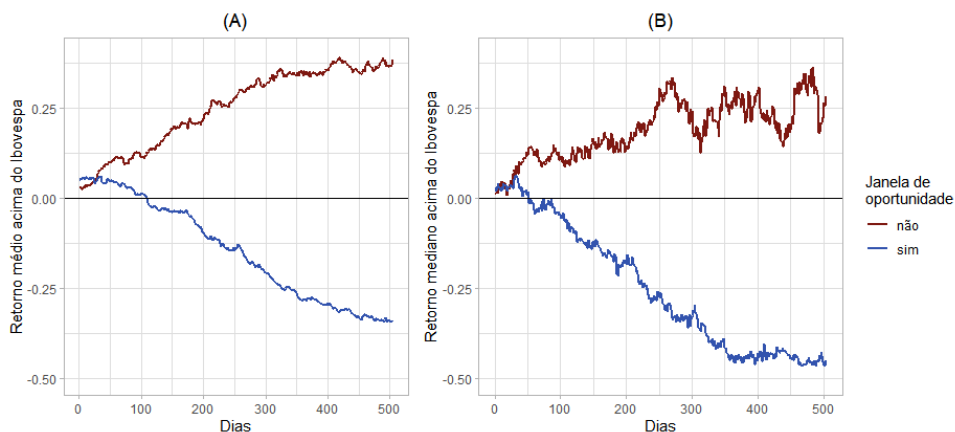


Figura 4.5: Média (A) e mediana (B) dos retornos ao longo do tempo para *IPOs* realizados dentro e fora das janelas de oportunidade.

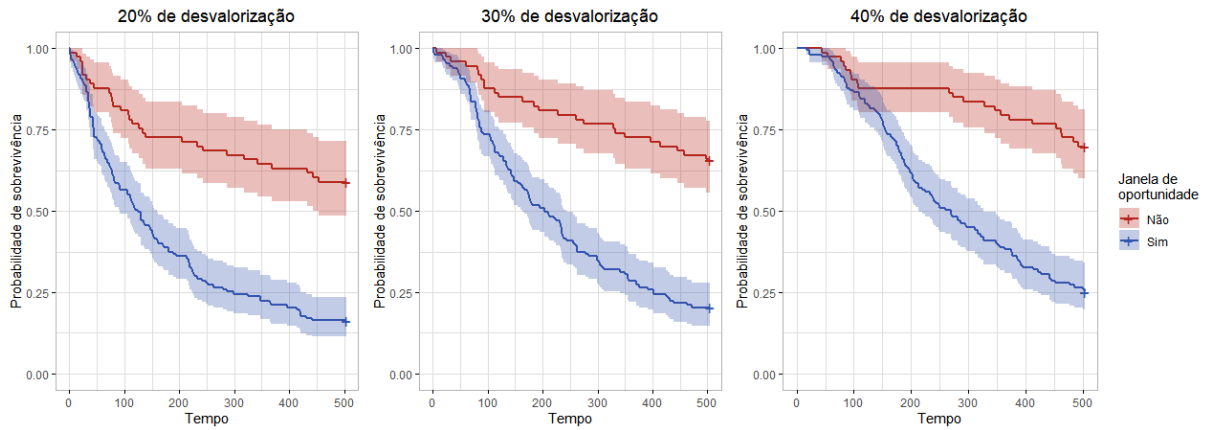


Figura 4.6: Curvas de Kaplan-Meier para cada evento de interesse e cada classificação dos *IPOs* como dentro e fora das janelas de oportunidade.

Nota-se, tanto na Figura 4.5 quanto na Figura 4.6, que a diferença entre os dois grupos é expressiva. No curto prazo, o retorno dos *IPOs* realizados dentro da janela de oportunidade foram maiores, em média, e equivalentes em mediana aos demais *IPOs*. Já no longo prazo, a performance das ofertas realizadas em períodos em que o mercado estava mais inerte supera em muitas vezes o retorno do outro grupo. Estes resultados estão em linha com a teoria postulada por Ritter (1991). Além dessas observações, na Figura 4.6, dado que as curvas de Kaplan-Meier não se cruzam, temos um indicativo de que a suposição de proporcionalidade, necessária para o ajuste de modelos de Cox, está sendo atendida.

Em seguida, também são observados comportamentos muito distintos da variável resposta para cada nível de retorno após um dia da oferta. Vale lembrar que as três categorias estão definidas como: baixo, para retornos menores que -1,07%; moderado, para retornos entre -1,07% e 6,83%; e alto, para retornos acima de 6,83% no primeiro dia.

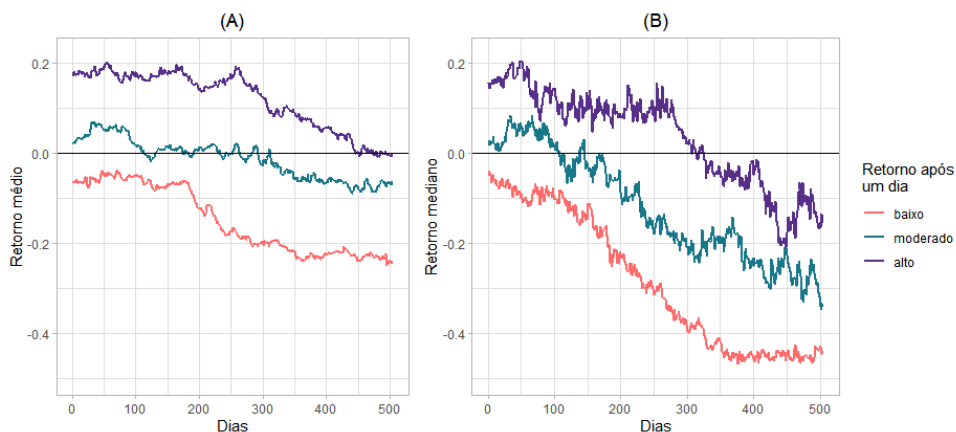


Figura 4.7: Média (A) e mediana (B) dos retornos ao longo do tempo para cada nível de valorização no primeiro dia após a oferta.

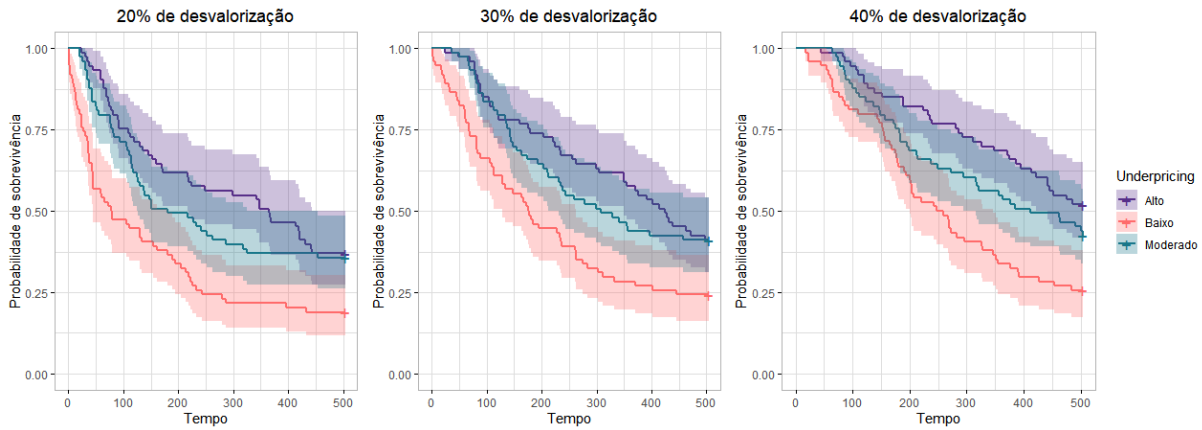


Figura 4.8: Curvas de Kaplan-Meier para cada evento de interesse e para cada nível de valorização no primeiro dia após a oferta.

Tanto na Figura 4.7 quanto na Figura 4.8, observa-se que valorizações mais expressivas no primeiro dia estão associadas a valorizações mais elevadas nos dias subsequentes, até mesmo dois anos após a oferta. Por outro lado, valorizações iniciais menores também estão relacionadas a valorizações superiores nos períodos seguintes. Contudo, é importante destacar que, em todos os casos, a tendência dos retornos de longo prazo é decrescente, indicando que, à medida que o tempo avança, os retornos médios e medianos tendem a diminuir. Além disso, ao observar as curvas de Kaplan-Meier da Figura 4.8, podemos dizer que há indicativos de que a suposição de proporcionalidade está sendo atendida.

Esses resultados divergem das teorias apresentadas por Shiller (1990) e Ibbotson e Ritter (1995), que sugerem que empresas com maiores retornos no curto prazo são frequentemente associadas a expectativas de crescimento mais altas, as quais têm maior probabilidade de não se concretizarem, resultando em desempenhos inferiores para os investidores no longo prazo. Ainda assim, a covariável analisada aparenta ser útil para explicar o comportamento da variável resposta.

Em relação à variável setor de atuação, nas Figuras 4.9 e 4.10, é possível observar como a média e a mediana dos retornos de ofertas públicas evoluem ao longo do tempo para diferentes setores.

Em relação à média, a maioria dos setores, com exceção dos setores financeiro, de saúde e de utilidade pública, apresentaram tendência negativa, enquanto os três restantes não apresentaram tendência. Já considerando a mediana, quase todas as categorias apresentaram tendência negativa, com exceção do setor de utilidade pública, que não apresentou tendência.

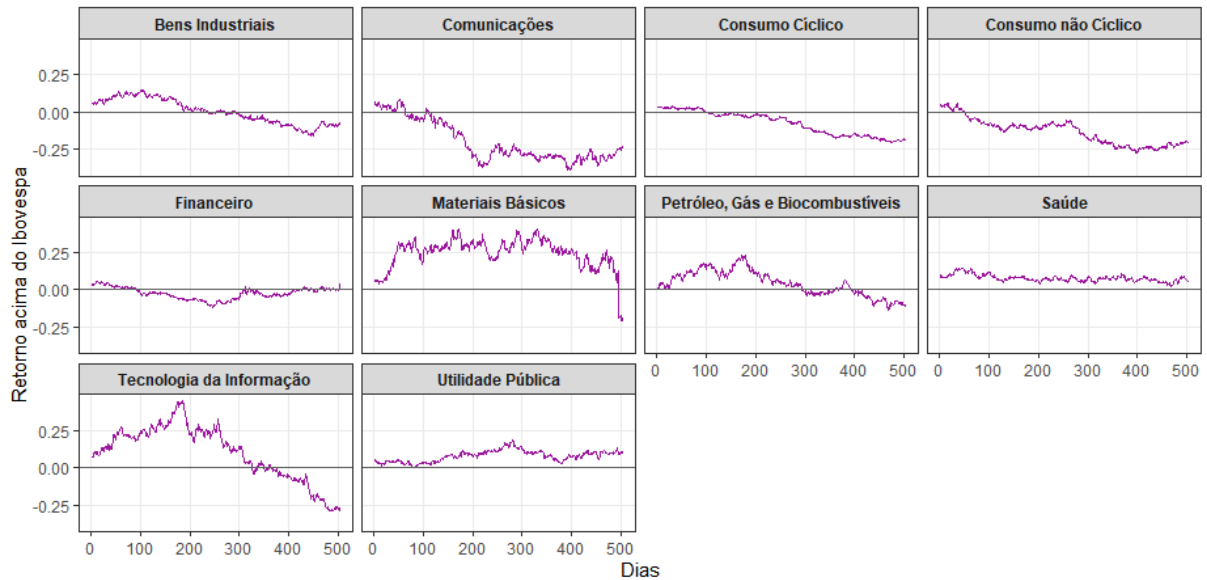


Figura 4.9: Retorno médio diário dos *IPOs* por setor.

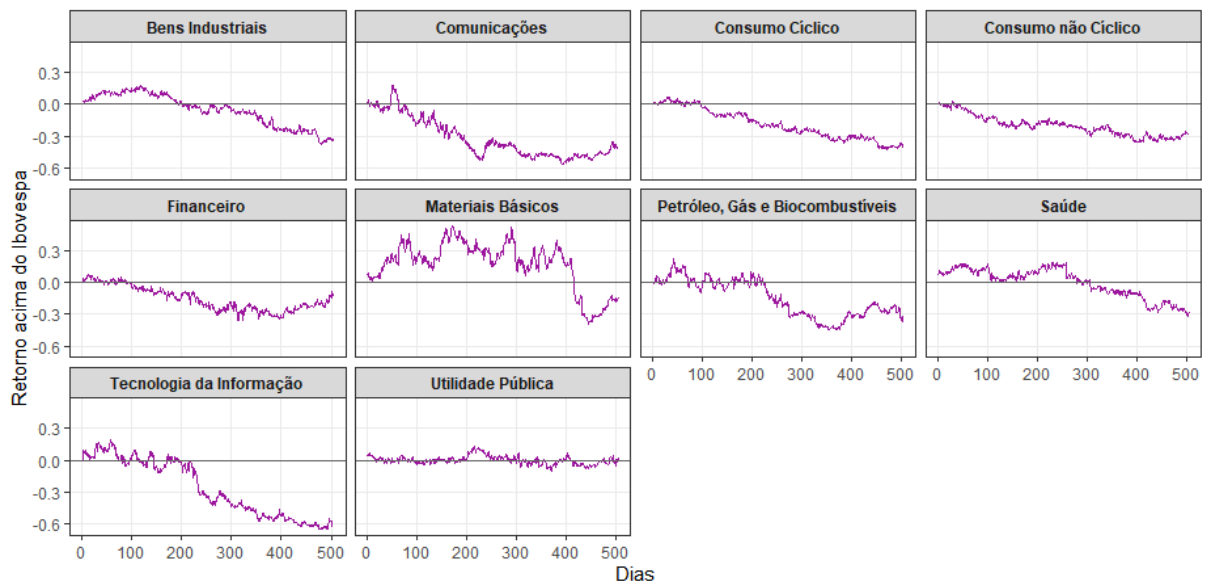


Figura 4.10: Retorno mediano diário dos *IPOs* por setor.

Com base nestes gráficos, é plausível pensar que o setor poderia ser uma variável útil para explicar a rentabilidade das ofertas públicas, no entanto, dado que as classes encontram-se excessivamente desbalanceadas, conforme podemos notar na Tabela 3.2, é possível que um modelo com esta covariável não tenha um ajuste adequado.

Nas Figuras 4.11 e 4.12, acompanhamos como está distribuída a nuvem de pontos entre o tamanho do *IPO* e o seu retorno relativo ao Ibovespa após um dia e dois anos da oferta. No geral, não é possível observar tendência alguma e, portanto, esta covariável foi desconsiderada nas análises posteriores.

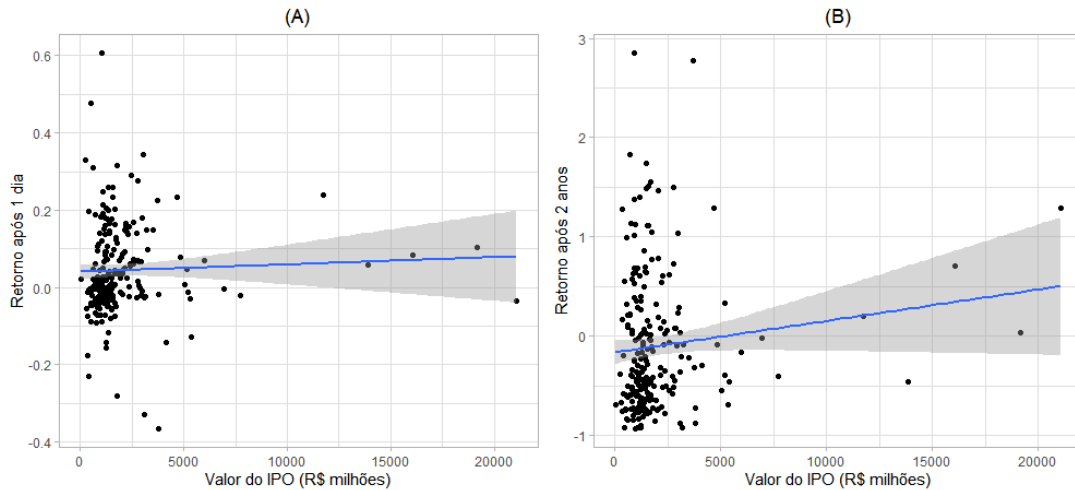


Figura 4.11: Retorno dos *IPOs* após 1 dia (A) e dois anos (B) *vis-à-vis* o tamanho da oferta, em milhões de reais.

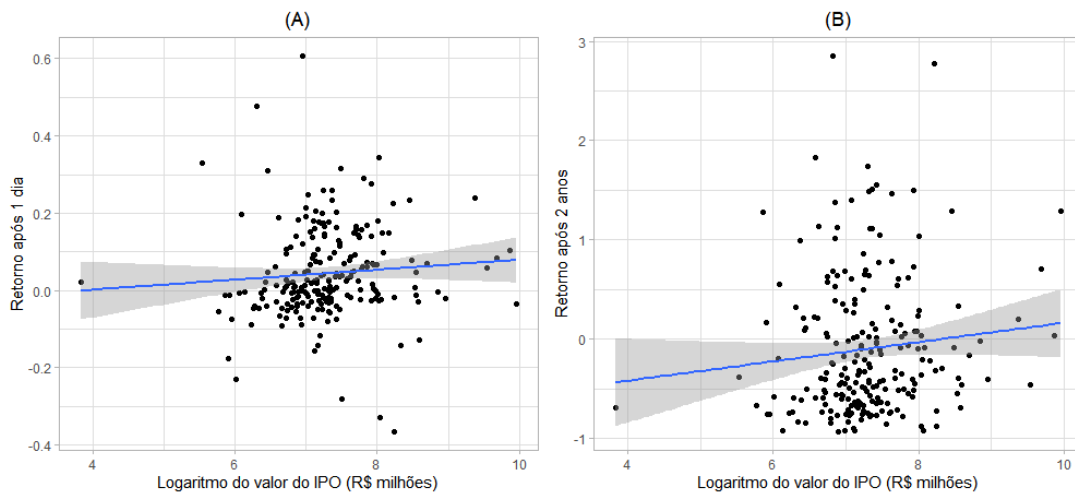


Figura 4.12: Retorno dos *IPOs* após 1 dia (A) e dois anos (B) *vis-à-vis* o logaritmo do tamanho da oferta, em milhões de reais.

Tabela 4.2: Valores-p para o teste de *logrank* para cada evento e covariável.

	Nível de desvalorização		
	20%	30%	40%
<b>Janela de oportunidade</b>	$1 \times 10^{-9}$	$2 \times 10^{-1}$	$8 \times 10^{-1}$
<b>Valorização no primeiro dia</b>	$2 \times 10^{-4}$	0,002	$5 \times 10^{-4}$
<b>Setor</b>	0,08	0,2	0,09

Por fim, na Tabela 4.2, podemos observar quais variáveis foram consideradas significativas segundo o teste de *logrank*, de forma que valores pequenos para o p-valor indicam

que as curvas de Kaplan-Meier são diferentes entre si. Nesta tabela, observa-se que as primeiras duas covariáveis são significativas a um nível de significância de 5%, enquanto que não há evidências de que as curvas de Kaplan-Meier associadas à covariável setor são diferentes entre si.

Agora, para determinar quais curvas diferem entre si, é necessário a realização de comparações múltiplas. Na Tabela 4.3, é possível observar os p-valores obtidos para as comparações múltiplas pelo método de Bonferroni. Nela, nota-se que não há evidências de diferença entre os níveis alto e moderado, enquanto que as maiores diferenças são observadas entre o nível baixo e os demais.

Tabela 4.3: Valores-p para o teste de *logrank* com comparações múltiplas dos níveis da variável valorização no primeiro dia, para os três modelos de Cox.

	Níveis de desvalorização		
	-20%	-30%	-40%
<b>baixo versus alto</b>	0,0004	0,0042	0,0004
<b>baixo versus moderado</b>	0,0104	0,0240	0,0565
<b>moderado versus alto</b>	1	1	0,4846

Por se tratar de uma variável binária, não foi necessário realizar as comparações múltiplas para a variável janela de oportunidade. Já para a variável setor, nenhuma categoria foi significativamente diferente das demais. Fora isso, é importante destacar que o teste de *logrank* não foi aplicado à variável tamanho da oferta, uma vez que, a partir da análise do gráfico de dispersão, não foi identificada qualquer relação aparente entre essa variável e a variável resposta, o que nos levou a optar por não categorizá-la para tornar a aplicação do teste possível.

## 4.2 Teste de hipótese para a mediana

Com base na análise descritiva realizada anteriormente, temos indicativos favoráveis à presença dos fenômenos de *underpricing* e de baixa performance a longo prazo no mercado de *IPOs* brasileiros nos últimos anos. Diante disso, estes comportamentos estão presentes, torna-se necessária a realização de testes. Ao todo, foram realizados seis testes do sinal para verificar se a mediana populacional é maior, menor ou igual a zero para os seis períodos elencados anteriormente, dado que todas as suposições para esta aplicação

foram atendidas. Vale ressaltar que testes para a mediana, neste caso, são preferíveis em relação a testes para a média, pois a mediana é robusta a *outliers*, permitindo uma avaliação mais precisa da tendência geral dos investimentos sem a distorção causada por valores extremos.

Dessa forma, para os três primeiros períodos – de um dia, uma semana e um mês – foram feitos testes para verificar se a mediana é maior do que zero, ou seja, se a rentabilidade dos *IPOs* é significativamente maior do que a rentabilidade do Ibovespa em mediana, dado que queremos avaliar a presença do fenômeno de *underpricing* nos períodos mais curtos, que é caracterizado por retornos relativos positivos no curto prazo. Já para os três últimos períodos – de seis meses, um ano e dois anos – foram realizados testes para verificar se a mediana é significativamente menor do que zero, o que indicaria que há evidências de que a performance dos *IPOs* no longo prazo é menor do que a do Ibovespa em mediana.

Logo, as estatísticas de teste foram calculadas atribuindo o valor um às unidades amostrais cuja rentabilidade relativa ao Ibovespa excedeu zero e o valor zero àquelas com rentabilidade inferior a zero, em conformidade com a teoria subjacente ao teste, detalhada no Capítulo 2.

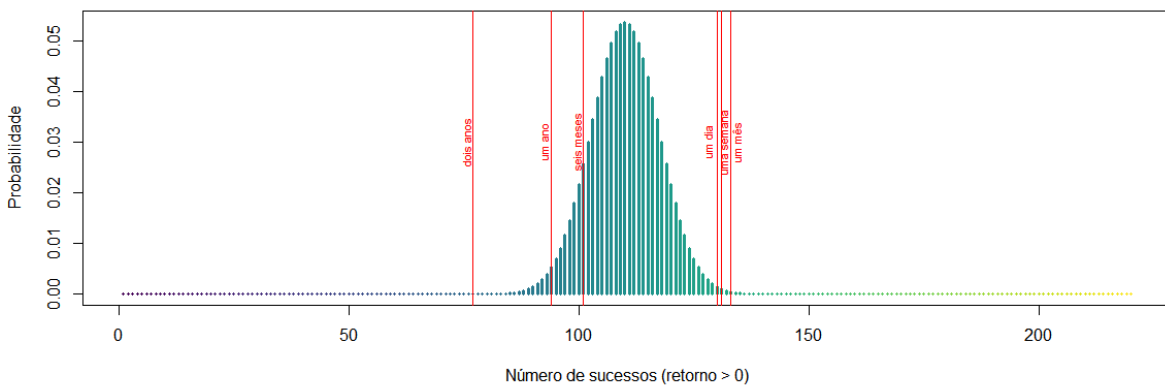


Figura 4.13: Distribuição binomial para  $n = 220$  e  $p = 0,5$  com os valores obtidos para a estatística do teste em cada período analisado.

Na Figura 4.13, podemos observar como cada um dos valores obtidos para as estatísticas do teste se encontram na distribuição binomial sob a hipótese nula, de modo que os três primeiros períodos se concentram na cauda direita da distribuição, sugerindo um número maior de instâncias em que a rentabilidade superou zero do que o esperado sob a hipótese nula de uma mediana igual a zero, já os últimos três períodos, tiveram o comportamento oposto, agrupando-se no lado esquerdo da distribuição, o que indica

maior incidência de valores negativos do que os esperados pela hipótese nula.

Com base na Tabela 4.4, em que estão apresentados os p-valores obtidos em cada um dos testes, concluímos para um nível de significância de 5% que há evidências, com base na amostra, de que a mediana dos três primeiros períodos são maiores que zero, enquanto as medianas dos dois últimos períodos são menores que zero. Já para o período de seis meses não foram encontradas evidências de que a mediana é diferente de zero.

Tabela 4.4: Teste do sinal para a mediana de cada período e os respectivos p-valores obtidos.

	p-valor	Hipótese alternativa
<b>1 dia</b>	0,0042	Mediana >0
<b>1 semana</b>	0,0028	Mediana >0
<b>1 mês</b>	0,0012	Mediana >0
<b>6 meses</b>	0,1258	Mediana <0
<b>1 ano</b>	0,0182	Mediana <0
<b>2 anos</b>	$5,085 \times 10^{-6}$	Mediana <0

Assim, no contexto do problema sob estudo, concluímos que há evidências no curto prazo, de um dia a um mês, de que a rentabilidade dos *IPOs* é maior do que a do Ibovespa em mediana, enquanto que para o longo prazo, de um a dois anos, há evidências de que a rentabilidade dos *IPOs* possui uma performance menor do que a do Ibovespa em mediana. Já para o período de seis meses, não foram encontradas evidências de que a mediana é negativa. Tal resultado era esperado de acordo com a análise descritiva e também está em linha com estudos anteriores que avaliaram a rentabilidade do investimento em ofertas públicas iniciais.

## 4.3 Modelagem

Nesta seção, apresentamos os resultados obtidos a partir do ajuste dos modelos de sobrevivência ao problema em análise. Inicialmente, foram testados três modelos de Cox, cada um ajustado para um evento específico, conforme descrito na Subseção 4.3.1. Em sequência, avaliou-se o modelo PWP com um vetor de parâmetros compartilhado entre os eventos, na Subseção 4.3.2.

Para o ajuste dos modelos, foi ajustado inicialmente o modelo completo com interações duas a duas, incluindo as covariáveis setor, retorno no primeiro dia e a variável que indica se o *IPO* foi feito em um período otimista ou não (dentro ou fora da janela de oportunidade). A partir deste ajuste inicial, foram removidas as covariáveis não significativas e o modelo foi reajustado até obtermos um modelo final. Apenas os resultados do modelo final foram incluídos.

### 4.3.1 Modelos de Cox

De início, os primeiros modelos ajustados foram três modelos de Cox distintos, um para cada definição do evento de interesse, isto é, para quando o retorno atinge os níveis de perda de 20%, 30% e 40%. Na Tabela 4.5, encontram-se os resultados obtidos nos modelos finais ajustados; vale mencionar que tanto as interações quanto a variável setor não foram consideradas significativas pelo teste  $t$  a um nível de significância de 5%.

Tabela 4.5: Valores estimados para os parâmetros dos modelos de Cox.

Eventos	<b>v1</b>	<b>v2</b>	<b>v3</b>	<b>exp(v1)</b>	<b>exp(v2)</b>	<b>exp(v3)</b>
<b>-20%</b>	1,5100	1,0733	0,5543	4,5268	2,9249	1,7407
<b>-30%</b>	1,4777	0,9492	0,3254	4,3828	2,5837	1,3845
<b>-40%</b>	1,368	1,035	0,351	3,928	2,816	1,421

Já na Tabela 4.6, temos os intervalos de 95% de confiança para os exponenciais dos parâmetros, o que nos ajuda a avaliar a estimativa dos efeitos das variáveis explicativas sobre o risco. Quando o intervalo não contém o valor 1, isso sugere que a variável tem um efeito estatisticamente significativo sobre a taxa de falha. Além disso, a amplitude do intervalo reflete a precisão da estimativa, com intervalos mais estreitos indicando maior precisão.

Tabela 4.6: Intervalos de 95% de confiança para os exponenciais cada parâmetro dos modelos de Cox.

	<b>exp(v1)</b>		<b>exp(v2)</b>		<b>exp(v3)</b>	
	limite inferior	limite superior	limite inferior	limite superior	limite inferior	limite superior
<b>-20%</b>	2,597	5,940	1,896	4,184	0,943	2,140
<b>-30%</b>	2,8107	6,834	1,7180	3,886	0,9036	2,121
<b>-40%</b>	2,834	7,230	1,895	4,514	1,107	2,736

As variáveis v1, v2 e v3 foram definidas conforme especificado nas equações (4.1), (4.2) e (4.3).

$$v1 = \begin{cases} 0, & \text{se a empresa não pertence a nenhuma janela de oportunidade} \\ 1, & \text{caso contrário} \end{cases} \quad (4.1)$$

$$v2 = \begin{cases} 1, & \text{se o retorno no primeiro dia foi baixo} \\ 0, & \text{caso contrário} \end{cases} \quad (4.2)$$

$$v3 = \begin{cases} 1, & \text{se o retorno no primeiro dia foi moderado} \\ 0, & \text{caso contrário} \end{cases} \quad (4.3)$$

A partir do exponencial das estimativas dos parâmetros de cada covariável, podemos obter a interpretação do modelo, uma vez que  $\exp\{\hat{\beta}_q\}$  representa em quantas vezes a taxa de risco de base é acelerada ou desacelerada conforme os valores que a covariável q assume. Para este modelo em específico, a interpretação é como se segue:

**exp(v1):** caso a empresa seja categorizada como pertencente a um período de grande volume de *IPOs*, ou seja, caso a oferta pertença a uma janela de oportunidade, estima-se que seu risco de sofrer perdas de -20%, -30% e -40% relativo ao Ibovespa aumenta em 4,5, 4,4 e 3,9 vezes respectivamente;

**exp(v2):** caso a empresa apresente um retorno classificado como baixo no primeiro dia, estima-se que seu risco de sofrer perdas de -20%, -30% e -40% é multiplicado por 2,92, 2,58 e 2,82 vezes respectivamente, quando comparado ao caso em que a empresa possui um retorno alto no primeiro dia, conforme a definição de cada classificação no Capítulo 3;

$\exp(\mathbf{v3})$ : por último, caso a empresa possua um retorno moderado no primeiro dia, comparativamente a um retorno alto, estima-se que seu risco de sofrer perdas de -20%, -30% e -40% é multiplicado por 1,74, 1,38 e 1,42 vezes, respectivamente.

Além dessas interpretações, é possível observar que as estimativas de cada um dos parâmetros entre os três modelos são similares entre si, o que indica que cada covariável afeta a função de risco de base de cada modelo de maneira semelhante.

Já na Figura 4.14, é possível observar como cada risco de base progride em relação ao tempo e para cada evento definido. Naturalmente, o risco aumenta conforme o tempo passa e valores mais altos de perda têm riscos menores de ocorrerem quando comparados a valores menores. Vale mencionar que o gráfico também pode ser interpretado como sendo representante do risco de se sofrer cada um dos eventos dada uma unidade amostral cujo vetor de covariáveis é nulo, ou seja, tratam-se das curvas de risco para um investimento em uma empresa cujo *IPO* não foi feito durante uma janela de oportunidade e que obteve um retorno considerado alto no primeiro dia.

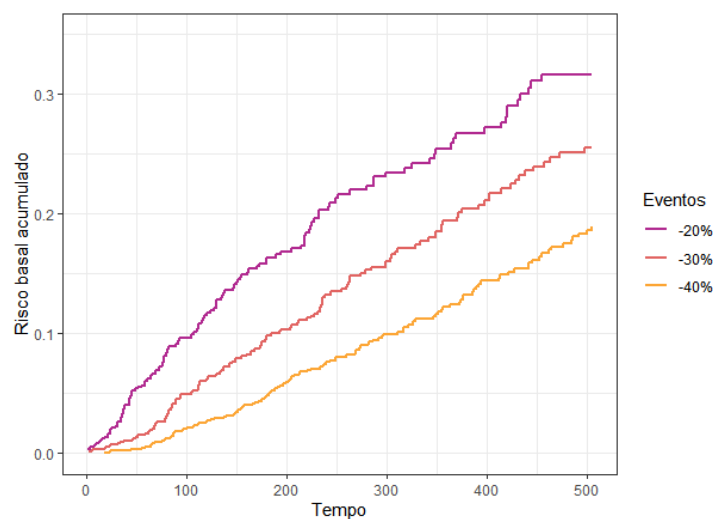


Figura 4.14: Funções de risco basal para cada um dos modelos e Cox.

## Análise de Diagnóstico

No que diz respeito à qualidade do ajuste, em primeiro lugar, foi avaliada se a suposição de proporcionalidade está sendo satisfeita. Ao observar os gráficos da Figura 4.15, notamos que, em todos os modelos e variáveis, a nuvem de pontos não apresenta tendência e também se encontra centrada no zero permanecendo dentro das bandas de confiança, o que indica que a suposição de proporcionalidade está sendo atendida.

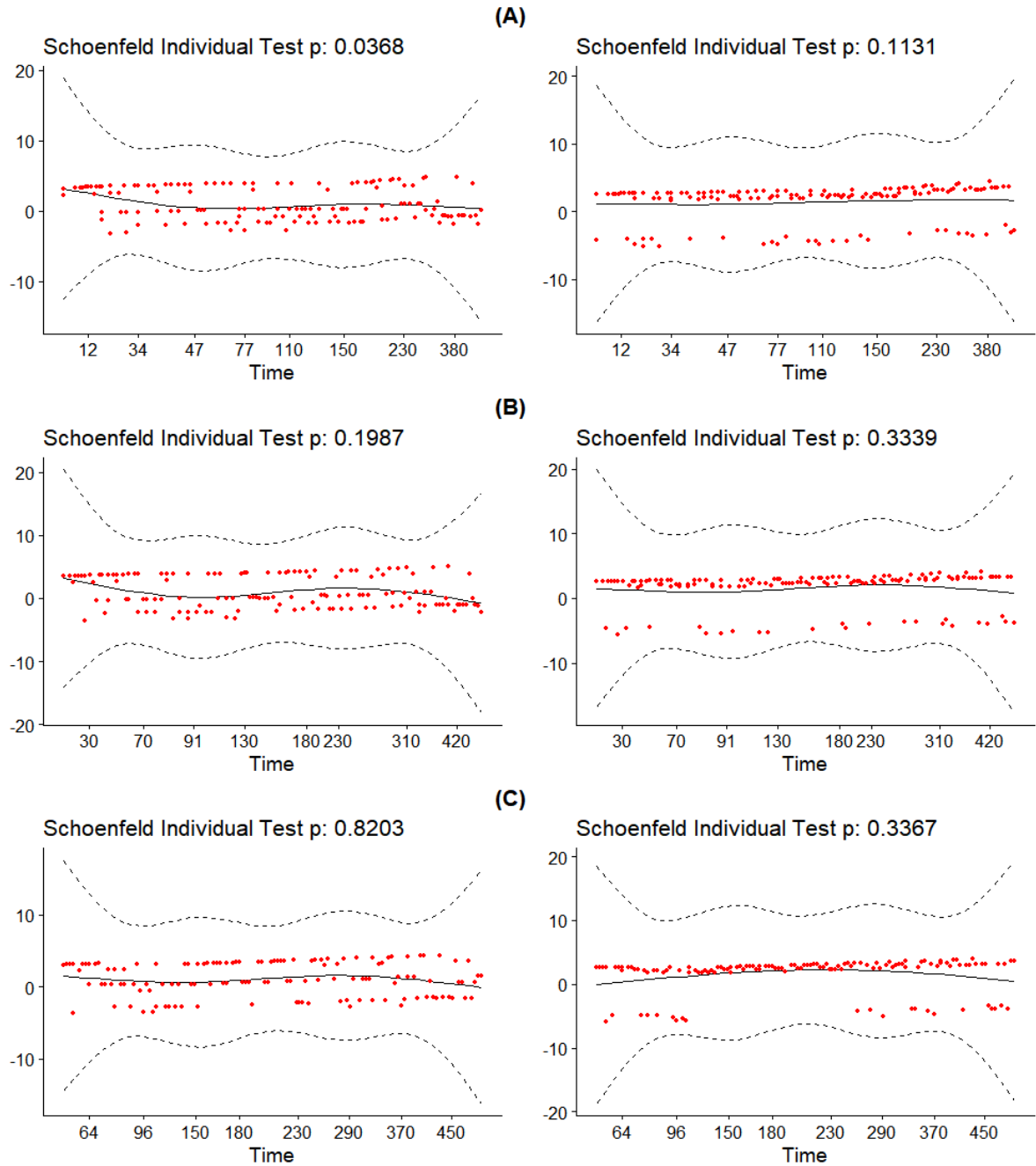


Figura 4.15: Resíduos de Schoenfeld por covariável e para os modelos para o níveis de desvalorização 20% (gráficos (A)), 30% (gráficos (B)) e 40% (gráficos (C)); em todos, à esquerda encontra-se o gráfico para a covariável retorno no dia 1, enquanto que à direita encontra-se valores para a covariável janela de oportunidade.

Tabela 4.7: Valores-p para o teste de Schoenfeld de proporcionalidade.

Valores-p para cada modelo			
	-20%	-30%	-40%
<b>Janela de oportunidade</b>	0,037	0,20	0,82
<b>Retorno no dia 1</b>	0,113	0,33	0,34
<b>Global</b>	0,051	0,30	0,75

Na Tabela 4.7, pode-se observar os valores-p para o teste de proporcionalidade de Schoenfeld, cuja hipótese nula é de que os riscos são proporcionais. Apesar de ser possível observar valores-p menores no modelo ajustado para a desvalorização de -20% como evento de interesse, consideramos que a suposição de proporcionalidade está sendo atingida, dado a qualidade dos gráficos para os resíduos.

Por último, analisando pela ótica dos resíduos martingal, na Figura 4.16, notamos que os resíduos, de modo geral, estão centrados no zero, e não apresentam tendência aparente, apesar de que, caso houvesse tendência, é possível que seria difícil de identificá-la por tratarem-se de covariáveis categóricas. Além disso, é possível observar alguns pontos atípicos, mas que, por serem poucos, não indicam problemas com o ajuste do modelo. Logo, concluímos que o ajuste aparenta ser adequado.

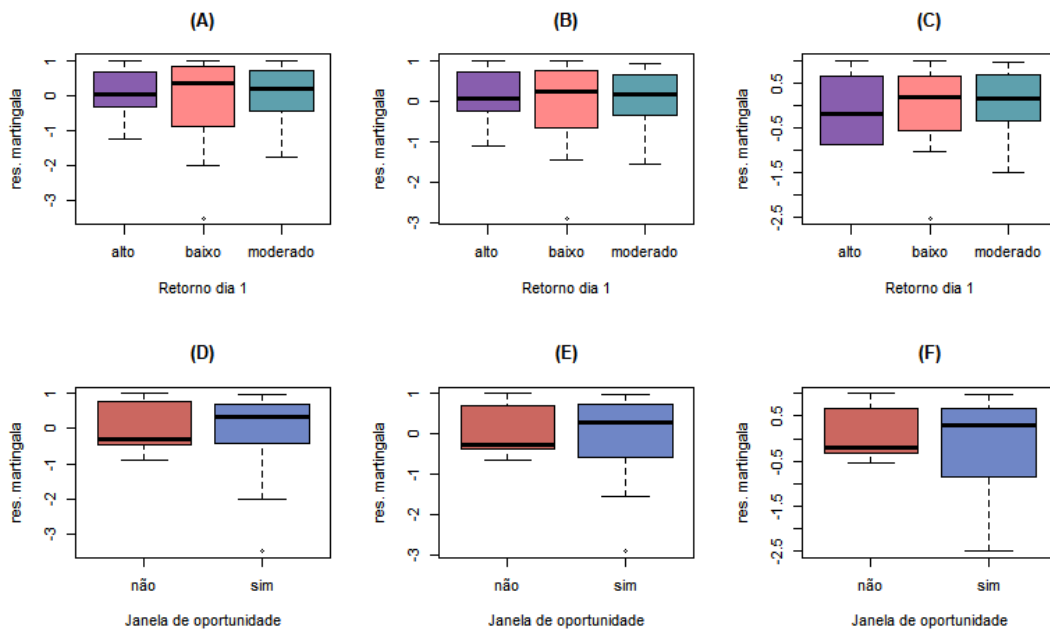


Figura 4.16: Resíduos martingal por covariável e para os modelos para o níveis de desvalorização 20% (gráficos (A) e (D)), 30% (gráficos (B) e (E)) e 40% (gráficos (C) e (F)).

### 4.3.2 Modelo PWP

Por fim, dado que as estimativas para os parâmetros dos três modelos de Cox foram similares para a mesma covariável, é razoável supor que também seja adequado ajustar o modelo PWP com parâmetros invariantes, isto é, com um único vetor  $\beta$  para todos os eventos. Para o ajuste do modelo, os dados tiveram que ser reformatados de modo que seu conteúdo fosse compatível com os argumentos da função do *software* R utilizada para o ajuste do modelo<sup>1</sup>.

Quanto ao processo de seleção de modelos, foi ajustado o modelo completo primeiro e foram removidas, uma a uma, as covariáveis não significativas, chegando a um modelo final, cujas estimativas para o vetor de parâmetros pode ser encontrada na Tabela 4.8.

Tabela 4.8: Estimativas dos parâmetros do modelo PWP com vetor de parâmetros igual para todos os eventos.

Covariáveis		$\exp( )$
<b>v1</b>	1,4446	4,2401
<b>v2</b>	1,0159	2,7619
<b>v3</b>	0,4029	1,4962

Já na Tabela 4.9, tem-se os intervalos de 95% de confiança estimados para cada parâmetro. Note que as amplitudes destes intervalos permaneceram similares às encontradas nos modelos de Cox.

Tabela 4.9: Intervalos de 95% de confiança para os exponenciais dos parâmetros do modelo PWP.

Covariáveis	limite inferior	limite superior
<b>exp(v1)</b>	2.7341	6.576
<b>exp(v2)</b>	1.9056	4.003
<b>exp(v3)</b>	0.9985	2.242

Vale mencionar que as covariáveis v1, v2 e v3, são as mesmas que as do ajuste de múltiplos modelos de Cox, dado que ambos identificaram as mesmas covariáveis como sendo significativas. A interpretação do modelo é como segue:

<sup>1</sup>Mais detalhes podem ser encontrados no Apêndice B.

**exp(v1):** caso a oferta seja realizada em um período com grande volume de ofertas, é estimado que o risco de ocorrerem perdas, de qualquer nível, é multiplicado em 4,24 vezes quando comparado ao caso contrário;

**exp(v2):** caso o *IPO* tenha um retorno, no primeiro dia, considerado baixo, isto é, que pertença aos retornos do tercil inferior, estima-se que o risco de ocorrerem perdas ao longo dos dois anos após a oferta é multiplicado em 2,76 vezes quando comparado ao caso em que o retorno no primeiro dia é considerado alto;

**exp(v3):** por fim, caso o investimento tenha um retorno no primeiro dia considerado moderado, estima-se que o risco de ocorrerem perdas, de qualquer nível, nos dois anos seguintes é multiplicado por 1,49 quando comparado também ao caso em que o retorno no primeiro dia é considerado alto.

Na Figura 4.17, é possível observar a evolução, ao longo do tempo, da função de risco de base estimada. Novamente, ela deve ter comportamento estritamente crescente ou com tendência nula, com riscos maiores para perdas menores e riscos menores para perdas maiores.

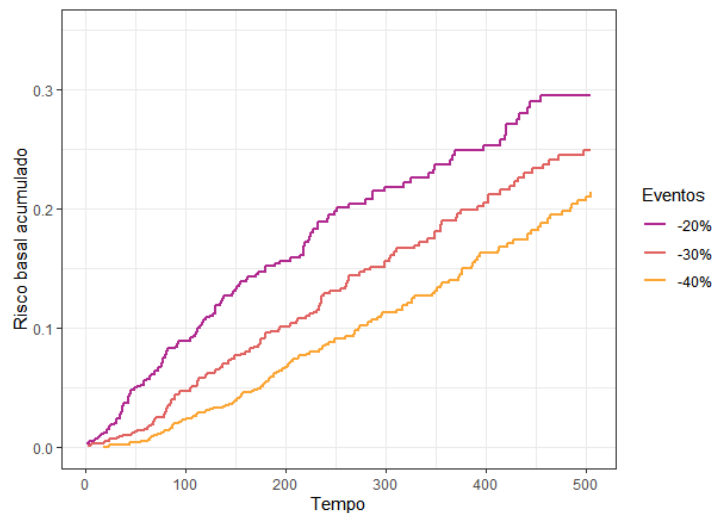


Figura 4.17: Funções de risco basal para o modelo PWP com vetor de parâmetros invariante.

## Análise de Diagnóstico

No que diz respeito à qualidade do ajuste, a começar pela adequação à suposição de proporcionalidade, pelos gráficos da Figura 4.18, temos que ambas as nuvens de pontos aparentam não ter tendência, além de estarem centradas ao redor do zero, não ultrapassando os limites das bandas de confiança, o que indica que a suposição de proporcionalidade dos riscos é razoável.

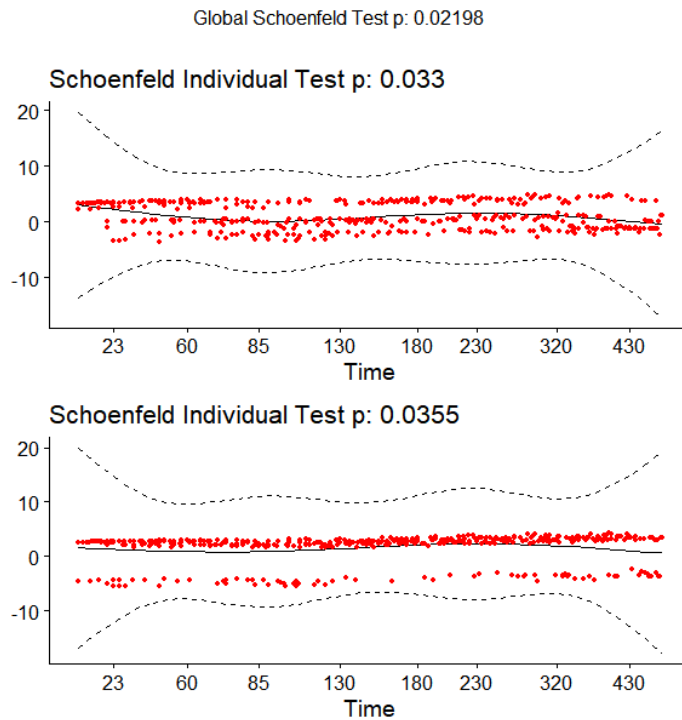


Figura 4.18: Resíduos de Schoenfeld para o modelo PWP para as covariáveis retorno no primeiro dia (gráfico de cima) e janela de oportunidade (gráfico de baixo).

No entanto, ao analisar os resultados obtidos no teste de Schoenfeld, que podem ser observados na Tabela 4.10, temos que a suposição de proporcionalidade não está sendo satisfeita, apesar de todos os indicativos contrários ao analisarmos a nuvem de pontos dos resíduos e as curvas de Kaplan-Meier. Por esses motivos, é razoável supor que a proporcionalidade está sendo atendida, uma vez que os demais resultados se mostraram bastante favoráveis.

Tabela 4.10: Valores-p para o teste de Schoenfeld.

	Valor-p
<b>Janela de oportunidade</b>	0,035
<b>Retorno no dia 1</b>	0,033
<b>Global</b>	0,022

Por fim, ao analisar os resíduos martingal, na Figura 4.19, notamos a ausência de tendência, o que indica que não há necessidade realizar transformações nas covariáveis, além de também notarmos a presença de alguns poucos pontos discrepantes. De modo geral, pela análise de diagnóstico realizada, temos que o ajuste está adequado.

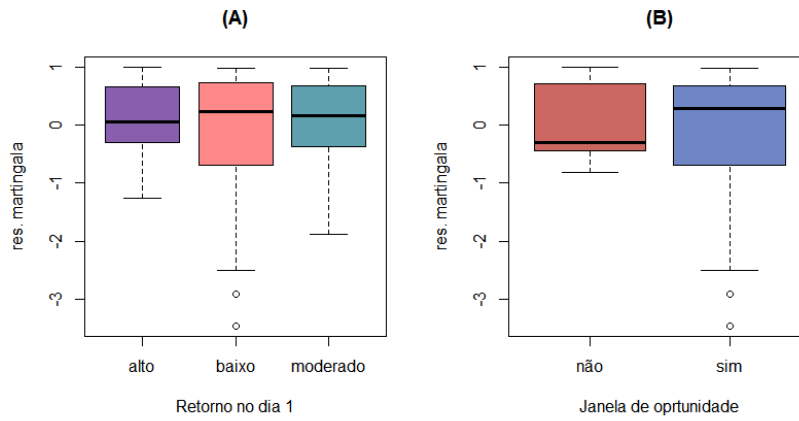


Figura 4.19: Resíduos martingal para o modelo PWP para as covariáveis retorno no primeiro dia (gráfico (A)) e janela de oportunidade (gráfico (B)).

# Capítulo 5

## Conclusão

Em síntese, este estudo teve como objetivo explorar o comportamento dos retornos de investimentos em ofertas públicas iniciais de ações no mercado brasileiro dos últimos 20 anos. Em linha com pesquisas anteriores acerca deste tema, o foco desta análise recaiu sobre dois fenômenos principais: o *underpricing*, caracterizado por retornos positivos no primeiro dia após a oferta; e a baixa performance do investimento em prazos maiores.

A fim de atender esse objetivo, foram coletados dados de 220 ofertas realizadas entre os períodos de 2004 a 2023, abordando apenas fontes primárias. A variável resposta, portanto, foi definida como o retorno do investimento em determinado *IPO* relativo ao Ibovespa, o qual representa o desempenho do mercado acionário brasileiro de modo geral. No que tange os métodos utilizados para analisar os dados, em primeiro lugar, foi feita uma análise descritiva e exploratória, em seguida, foram realizados testes para verificar se há, de fato, a presença dos fenômenos mencionados anteriormente e, por último, foram ajustados modelos de sobrevivência a fim de avaliar como os retornos a longo prazo são afetados pela presença de diferentes variáveis explicativas.

Nesse contexto, a inclusão de covariáveis foi pautada pelas teorias subjacentes acerca da baixa rentabilidade a longo prazo documentada em outros estudos. A primeira teoria, proposta por [Shiller \(1990\)](#), sugere que *IPOs* possuem retornos iniciais tendem a ser excessivamente altos, espelhando expectativas e um otimismo elevados por parte dos investidores, o que é seguido por prejuízos em períodos mais longos, ao passo que tais expectativas não são concretizadas. Desse modo, foi adicionada uma variável indicadora do nível de retorno no primeiro dia após o *IPO*, com seguintes categorias: baixo, para retornos menores que -1,07%, moderado, para retornos entre -1,07% e 6,83%, e alto, para retornos acima de 6,83%. Por outro lado, a segunda teoria, elaborada por [Ritter \(1991\)](#),

sugere que *IPOs* ocorrem em ciclos de euforia, quando o mercado está aquecido e os preços estão inflacionados, os quais são seguidos por reversões no longo prazo. Para testar essa teoria, incluímos uma covariável binária indicando se o *IPO* foi emitido em períodos de alta atividade (2006, 2007, 2020, 2021) ou fora de janelas de oportunidade. Além disso, também foram incluídas as covariáveis setor de atuação e tamanho da oferta, em milhões de reais.

Quanto aos resultados obtidos, na análise descritiva, para os retornos dos *IPOs* foi verificada uma tendência positiva, no curto prazo, marcada por uma média de 4,4% de valorização no primeiro dia. No médio e longo prazos, esta tendência se transforma em uma tendência negativa e o prazo de dois anos de observação é marcado por um retorno mediano de -32,33%. Ao realizar testes do sinal para mediana, foi concluído que há evidências, com base na amostra, de que para os períodos de um dia, uma semana e um mês, os retornos dos *IPOs* tendem a ser maiores que os do Ibovespa em mediana. Já para os prazos de um ano e dois anos obtivemos evidências de que os retornos dos investimentos em ofertas públicas iniciais são significativamente menores do que os retornos para o Ibovespa no mesmo período, em mediana.

Quando incluímos variáveis explicativas na análise, encontramos que ofertas realizadas em períodos em que o mercado está mais aquecido possuem padrões mais acentuados de retornos positivos no curto prazo e negativos em períodos maiores, enquanto que para os demais *IPOs* o fenômeno de *underpricing* aparenta estar presente em menor escala e não houve tendência negativa para os retornos, em média ou em mediana, no longo prazo, em linha com Ritter (1991). Já para a variável indicadora do nível de retorno no primeiro dia após a oferta, na análise descritiva, foi possível observar que ofertas com retornos maiores no primeiro dia tendem a ter retornos maiores nos dias subsequentes, enquanto que ofertas com retornos classificados como baixos no primeiro dia também obtiveram os retornos mais baixos nos períodos seguintes, ao contrário de Shiller (1990). Por fim, não houveram evidências de que as covariáveis setor de atuação e tamanho da oferta são significativas para a análise.

Por último, para melhor compreendermos os efeitos de cada covariável na variável resposta, os dados foram interpretados como dados de sobrevivência, em que, ao ser atingido determinado nível de perda em um investimento, é considerado que ocorreu uma “mortalidade”. Considerando os níveis de perda de 20%, 30 % e 40%, primeiro, foram ajustados três modelos de Cox, um para cada evento, e, em seguida, foi ajustado um modelo PWP

com vetor de parâmetros igual para cada evento. Ambos os modelos obtiveram ajustes satisfatórios, no entanto vale mencionar que o modelo PWP se mostrou mais parcimonioso ao trabalhar com os eventos de forma conjunta e, portanto, a interpretação de seus parâmetros também foi facilitada. A partir destes ajustes, obtivemos que os efeitos de cada variável ocorre de forma similar em todos os eventos de modo que, caso a oferta seja realizada durante um período de mercado aquecido, estima-se que o risco de se sofrerem perdas aumenta em 4,2 vezes. Caso a oferta tenha tido um retorno considerado baixo ou moderado no primeiro dia, o risco de ocorrerem perdas estimado é aumentado em 2,76 e 1,49 vezes, respectivamente, quando comparado ao risco caso o retorno tenha sido alto no primeiro dia.

Em conclusão, este estudo contribuiu para investigar e elucidar algumas questões acerca de possíveis ineficiências na precificação de ofertas públicas iniciais de ações. Conjuntamente, foi possível a aplicação de técnicas estatísticas, sobretudo no que diz respeito à análise de sobrevivência, em um domínio pouco explorado por essas metodologias, como o é o caso da área de finanças. Também foram explorados métodos que levam em consideração eventos múltiplos e ordenados, como o modelo PWP, que geralmente não fazem parte do escopo de cursos de graduação, ampliando a compreensão sobre suas aplicações. Como parte do desafio, a implementação computacional desses métodos no *software* R demandou esforço adicional, visto o entendimento do código aplicado à base de dados.



# Referências Bibliográficas

- Aggarwal, R., Leal, R. e Hernandez, L. (1993). The aftermarket performance of initial public offerings in latin america. *Financial Management*, **22**(1), 42–53.
- Akram, S. e Ann, Q. U. (2015). Newton-Raphson method. *International Journal of Scientific & Engineering Research*, **6**(7), 1748–1752.
- B3 (2024a). Séries históricas dos preços dos ativos listados na B3. Disponível em: [https://www.b3.com.br/pt\\_br/market\\_data\\_e\\_indices/servicos\\_de\\_dados/market\\_data/historico/mercado\\_a\\_vista/series\\_historicas/](https://www.b3.com.br/pt_br/market_data_e_indices/servicos_de_dados/market_data/historico/mercado_a_vista/series_historicas/). Último acesso em 12 de junho de 2024.
- B3 (2024b). Dados sobre ofertas públicas iniciais realizadas. Disponível em: [https://www.b3.com.br/pt\\_br/produtos\\_e\\_servicos/solucoes\\_para\\_emissores/ofertas\\_publicas/estatisticas/](https://www.b3.com.br/pt_br/produtos_e_servicos/solucoes_para_emissores/ofertas_publicas/estatisticas/). Último acesso em 12 de junho de 2024.
- B3 (2024). Índice Bovespa. Disponível em: [https://www.b3.com.br/pt\\_br/market\\_data\\_e\\_indices/indices/indices\\_amplos/ibovespa.htm](https://www.b3.com.br/pt_br/market_data_e_indices/indices/indices_amplos/ibovespa.htm). Último acesso em 12 de junho de 2024.
- Baluja, G. (2017). Survival of the fittest: an empirical analysis of IPOs in the post-SEBI era. *Business Analyst*, **2**, 79–106.
- Barihas, R., Weber, P. e Zimmermann, H. (2024). Ipo pricing in germany and switzerland in the new millennium: Tests on underpricing and long-term performance. *Journal of Risk and Financial Management*, **17**(11).
- Bonferroni, C. (1936). *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R. Istituto superiore di scienze economiche e commerciali di Firenze. Seeber.
- Breslow, N. (1972). Discussion on Professor Cox's Paper. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**(2), 202–220.

- Cabete, A. B. d. A. e. B. (2012). *Análise de sobrevivência com acontecimentos múltiplos: aplicação ao estudo do tempo até à ocorrência de enfarte do miocárdio*. Dissertação de Mestrado em Bioestatística, Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa.
- Carvalho, M. S., Andreozzi, V. L., Codeço, C. T., Campos, D. P., Barbosa, M. T. S. e Shimakura, S. E. (2011). *Análise de sobrevivência: teoria e aplicações em saúde*. SciELO – Editora FIOCRUZ. ISBN 9788575412169.
- Colosimo, E. A.; Giolo, S. R. (2024). *Análise de Sobrevivência Aplicada*. Ed. Edgard Blucher, second edition.
- Conover, W. (1999). *Practical Nonparametric Statistics*. Wiley, third edition.
- Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2), 187–220.
- Cox, D. R. e Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, **30**(2), 248–265.
- CVM (2024). Consulta de documentos de empresas listadas. Disponível em: <https://www.rad.cvm.gov.br/ENET/frmConsultaExternaCVM.aspx>. Último acesso em 19 de abril de 2024.
- Espenlaub, S., Khurshed, A., Mohamed, A. e Saadouni, B. (2016). Committed anchor investment and IPO survival – The roles of cornerstone and strategic investors. *Journal of Corporate Finance*, **41**, 139–155.
- Fama, E. (1970). Efficient Capital Markets: a Review of Theory and Empirical Work. *Journal of Finance*, **25**, 383–417.
- Graham, B. (1949). *The intelligent investor*. Harper & Brothers.
- Graham, B. e Dodd, D. (1934). *Security Analysis*. McGraw-Hill.
- Gurung, J. B. (2024). Long-run price performance phenomenon of initial public offerings: A review of theoretical and empirical perspectives. *Contemporary Research: An Interdisciplinary Academic Journal*, **7**(2), 114–137.

- Ibbotson, R. G. e Ritter, J. R. (1995). Initial public offerings. Em *Finance*, volume 9 of *Handbooks in Operations Research and Management Science*, páginas 993–1016. Elsevier.
- Ibbotson, R. G., Sindelar, J. L. e Ritter, J. R. (1988). Initial Public Offerings. *Journal of Applied Corporate Finance*, **1**(2), 37–45.
- IBGE (2025). Índice Nacional de Preços ao Consumidor Amplo (IPCA) - Séries Históricas. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/precos-e-custos/9256-indice-nacional-de-precos-ao-consumidor-amplo>. Acessado em: 24 fev. 2025.
- John W. Pratt, J. D. G. (1981). *Concepts of Nonparametric Theory*. Springer New York, NY.
- Kaplan, E. L. e Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association*, **53**(282), 457–481.
- M. G. Kendall, A. S. (1954). *The advanced theory of statistics*. Charles Griffin, London.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports*, **50**(3), 163–170.
- Martins, G. J. T., Pai, C. D. e Pereira, M. F. (2008). Análise da Performance de Longo Prazo dos IPOs no mercado de ações Brasileiro. Em *XXXII Encontro da ANPAD*, Rio de Janeiro.
- Microsoft Corporation (2019). *Microsoft Excel*.
- Prentice, R. L., Williams, B. J. e Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**(2), 373–379.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ritter, J. R. (1991). The long term performance of Initial Public Offerings. *Journal of Finance*, **46**(1), 3–27.
- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, **69**(1), 239–241.

Shiller, R. J. (1990). Speculative prices and popular models. *The Journal of Economic Perspectives*, 4(2), 55–65.

Siegel, J. (2008). *Investindo em ações no longo prazo*. McGraw-Hill e Elsevier, fourth edition.

Terry M. Therneau, P. M. G. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer New York, NY.

# Apêndice A

## Construção da base de dados

Neste apêndice, o escopo escolhido para os dados, assim como considerações feitas durante a construção da base de dados, serão abordados em maiores detalhes. Ao final, encontra-se um recorte do conjunto de dados.

Em primeiro lugar, em relação às categorias de ações incluídas no estudo, foram selecionados apenas *IPOs* de empresas listadas nos segmentos de mercado Nível 1 (N1), Nível 2 (N2) e Novo Mercado (NM). Os segmentos de mercado são classes definidas pela bolsa de valores brasileira para empresas de capital aberto de acordo com seus níveis de governança corporativa, a qual simboliza a transparência e o comprometimento das empresas em organizar seus processos internos e prestar contas ao mercado; o maior nível de governança corporativa é encontrado no segmento NM, o segmento N2 possui um nível de governança corporativa intermediário e o segmento N1 é o mais flexível dos três.

Com base no escopo estabelecido, informações sobre a data e preço por ação na oferta inicial bem como a cotação do Ibovespa e as cotações diárias das ações no mercado secundário por dois anos após a oferta foram obtidas a partir das bases de dados abertos da B3 (B3, 2024a,b).

Ademais, para calcular a rentabilidade dos investimentos em *IPOs* foi necessário incorporar a distribuição de proventos e eventuais desdobramentos ou agrupamentos que uma empresa pode fazer.

Estes três eventos societários podem ser descritos da seguinte forma: desdobramentos correspondem a um aumento no número de ações em circulação, reduzindo o valor unitário da ação mas não influenciando o valor do investimento; agrupamentos, pelo contrário, reduzem o número de ações em circulação, aumentando o preço da ação, mas também não alteram o valor do investimento; já a distribuição de proventos é caracterizada por paga-

mentos feitos pelas empresas aos acionistas, como juros sobre capital próprio, dividendos e bonificações, e resultam em uma queda no preço da ação a partir da data de negociação em que novos acionistas não terão direitos a esses proventos, conhecida como data “*ex*”. Todas estas informações foram obtidas através da consulta dos documentos históricos de cada companhia na CVM (CVM, 2024).

Quanto à realização de desdobramentos e agrupamentos, é feita uma correção na quantidade de ações na carteira hipotética. Já em relação à distribuição de proventos, após a data *ex*, seu valor em dinheiro é reinvestido na sua respectiva ação pelo preço vigente na data.

É importante destacar que todos estes eventos societários geram uma alteração na quantidade das ações obtidas inicialmente, e o tratamento efetuado é equivalente ao que é feito pela B3 ao calcular o índice Bovespa quando estes eventos ocorrem nas ações que constituem o índice, dado que se trata de um índice de retorno total. Dessa forma, foi garantida a equivalência de tratamento entre os dois investimentos teóricos; no *IPO* e no Ibovespa.

Por fim, um último ponto a ser considerado na construção da base de dados são três casos especiais que foram recorrentes durante a compilação dos dados: (i) algumas empresas deixaram de ser negociadas durante os dois anos de acompanhamento; (ii) algumas ações na amostra são de baixa liquidez, isto é, de baixo volume de negociação e não foram negociadas por alguns dias durante o acompanhamento; e (iii) algumas das datas das ofertas públicas iniciais registradas na base de dados da B3 coincidiram com fins de semana e feriados, quando não se tem cotações do Ibovespa.

Quanto ao primeiro caso, optamos por excluir da análise ações que não puderam ser observadas até o final do estudo, a fim de simplificar os tipos de censura com os quais teríamos que trabalhar durante a análise de sobrevivência. Ao todo, foi interrompido o acompanhamento de nove ofertas primárias – Submarino, BMF, Bovespa Holding, Satipel, Raia S.A., Mosaico Tecnologia do Consumidor, Focus Energia Holding, Tivit, Vivax –, destas, as seis primeiras foram fundidas com ou adquiridas por outras empresas, já as três últimas empresas tiveram seu capital fechado durante os primeiros dois anos desde sua estreia.

Em relação ao segundo caso, em dias em que porventura o volume de negociação de determinada empresa fosse nulo, a cotação do último dia em que houve alguma negociação era utilizada para substituir a cotação faltante. Essas ocorrências foram raras e o número

de cotações ausentes não ultrapassou a marca de cinco dias consecutivos em nenhuma das empresas.

Já em relação ao último caso mencionado, foi utilizada a última cotação do Ibovespa registrada antes da data do *IPO* para ser utilizada no cálculo do valor inicial investido no índice Bovespa.

Com base nessas considerações, o número final de *IPOs* incluídos na amostra foi de 220, cada um acompanhado diariamente por 504 dias úteis, o que é equivalente a dois anos. O processo de compilação e manuseio dos dados foi feito tanto por meio do *software* estatístico R (R Core Team, 2023) quanto do *software* Excel (Microsoft Corporation, 2019); o primeiro, por facilitar o manejo de grandes bases de dados, foi utilizado para compilar as cotações diárias de cada empresa, enquanto o segundo foi utilizado para compilar os dados referentes à ocorrência de eventos societários e para calcular a rentabilidade relativa de cada *IPO*.

Na Tabela A.1 temos um recorte da base de dados com as rentabilidades relativas de cada *IPO* nos recortes temporais de um dia, uma semana, um mês, seis meses, um ano e dois anos. Na tabela, o *ticker* representa o código de negociação e identificação das ações no mercado da bolsa de valores.

Tabela A.1.: Rentabilidade das cinco primeiras e cinco últimas ações em relação ao Ibovespa nos primeiros e os últimos cinco dias após o IPO.

Dias úteis após PO	1	2	3	4	5	...	216	217	218	219	220
	<i>Ticker</i>										
<b>1</b>	AALR3	ABCB4	ABNB3	ABRE11	ABYA3	...	VITT3	VIVA3	VNET3	VVEO3	WEST3
<b>2</b>	-4,09%	-5,87%	7,64%	-6,52%	-0,80%	19,58%	-0,10%	10,40%	9,25%	-15,62%	
<b>3</b>	-10,64%	-6,58%	3,06%	-7,97%	-2,11%	14,45%	-3,58%	12,38%	12,50%	-17,90%	
<b>4</b>	-9,20%	-6,43%	1,42%	-6,53%	-4,26%	15,43%	-4,91%	18,73%	16,18%	-20,64%	
<b>5</b>	-6,41%	-7,52%	-1,25%	-4,68%	-4,64%	19,97%	-4,92%	21,22%	13,95%	-20,51%	
...	-7,43%	-7,17%	-2,54%	-3,39%	-3,82%	16,13%	-3,56%	20,22%	22,09%	-15,64%	
<b>500</b>	-56,89%	-14,89%	-43,93%	117,17%	-7,59%	46,49%	12,58%	-2,47%	7,90%	-92,56%	
<b>501</b>	-56,97%	-11,30%	-43,52%	117,42%	-9,62%	44,41%	12,15%	-1,88%	7,44%	-92,46%	
<b>502</b>	-55,65%	-8,63%	-43,96%	116,60%	-11,07%	43,72%	8,96%	0,88%	7,49%	-92,71%	
<b>503</b>	-55,62%	-12,21%	-44,55%	114,87%	-18,68%	47,39%	6,97%	2,34%	11,02%	-92,28%	
<b>504</b>	-55,56%	-13,39%	-44,17%	113,74%	-19,76%	55,69%	7,74%	3,25%	8,36%	-92,15%	

# Apêndice B

## Códigos utilizados

Neste apêndice, encontram-se os códigos utilizados para a estimação das curvas de Kaplan-Meier, na Seção [B.1](#), e a aplicação do modelo PWP, na Seção [B.2](#).

### B.1 Kaplan-Meier

De início, para a aplicação dos métodos de sobrevivência, os dados tiveram que ser transformados do seu formato de séries temporais para um formato compatível com o formato requisitado para aplicação destes métodos.

```
1 survival_df <- function(taxa, dados){
2   survlista <- list()
3
4   for (acao in colnames(dados[,])) {
5     # seleciona os retornos de uma ação da lista
6     retornos <- dados[[acao]]
7     # retorna a 1 linha cujo valor é menor que a taxa
8     dia <- which(retornos <= taxa)[1]
9     if (is.na(dia)) {
10      # censura
11      survlista[[acao]] <- list(time = nrow(dados), status = 0)
12    } else {
13      # morte
14      survlista[[acao]] <- list(time = dia, status = 1)
15    }
16  }
```

```

16   }
17
18   # lista -> df
19   survdf <- do.call(rbind, lapply(survlista, data.frame,
20                                stringsAsFactors=FALSE))
21   survdf <- as.data.frame(survdF)
22   names(survdF) <- c("tempo", "status")
23
24   return(survdF)
25 }

```

Abaixo segue como a função acima seria aplicada para transformar o *dataframe* contendo as 220 séries temporais, conforme representado pela Tabela A.1, em dados de sobrevivência.

```

1 library(survival)
2 library(survminer)
3 library(ggplot2)
4
5 # curva de Kaplan-Meier considerando que o evento de interesse é a
6   ocorrência de uma desvalorização de 20%
7
8 survdf_20 <- survival_df(-0.2, temporal)
9 surv_obj_20 <- Surv(time = survdf_20$tempo, event = survdf_20$status)
10 km_fit_20 <- survfit(surv_obj_20 ~ 1, data = surv_obj_20)
11 ggsurvplot(km_fit_20, conf.int = TRUE, risk.table = TRUE)

```

## B.2 Modelo PWP

Para aplicação do modelo PWP, novamente, a base de dados foi transformada de forma que a base de dados utilizada para fazer as curvas de Kaplan-Meier incorporasse as covariáveis. No entanto, na formulação desta base de dados, foi necessário garantir que um evento, por exemplo a ocorrência de uma perda de 30% no valor do investimento, só pudesse ocorrer a partir do momento que o evento anterior já ocorreu, neste caso uma

perda de 20%, de forma a satisfazer as exigências desse modelo.

```

1 survival_df_pwp <- function(taxa1, taxa2, taxa3, dados) {
2   # armazenar os resultados em uma lista
3   surv_list <- list()
4   taxas <- c(taxa1, taxa2, taxa3)
5
6   for (acao in colnames(dados)) {
7     # extrair os retornos de uma ação específica
8     retornos <- dados[[acao]]
9
10    # um dataframe p colocar os detalhes de cada evento para esta ação
11    df_eventos <- data.frame(
12      ticker = rep(acao, length(taxas)), # 3 linhas por ação
13      tempo = NA,
14      status = 0,
15      obs.evento = 1:length(taxas)
16    )
17
18    dia1 <- which(retornos <= taxas[1])[1]
19    retornos[dia1] <- -retornos[dia1]
20    dia2 <- which(retornos <= taxas[2])[1]
21    retornos[dia2] <- -retornos[dia2]
22    dia3 <- which(retornos <= taxas[3])[1]
23    dias <- c(dia1, dia2, dia3)
24
25    for (i in 1:3){
26      if (!is.na(dias[i])) {
27        # registra o evento que ocorreu
28        df_eventos$tempo[i] <- dias[i]
29        df_eventos$status[i] <- 1
30      } else {
31        # se não houver eventos, censurar
32        df_eventos$tempo[i] <- nrow(dados)
33      }

```

```

34     }
35
36     # armazenar o resultado desta ação na lista
37     surv_list[[acao]] <- df_eventos
38 }
39
40 # combinar tudo em um único dataframe
41 surv_df <- do.call(rbind, surv_list)
42 rownames(surv_df) <- NULL
43
44 return(surv_df)
45 }
46
47 df_pwp <- survival_df_pwp(-0.2, -0.3, -0.4, temporal)
48
49 # adicionar as covariáveis
50 df <- merge(df_pwp, covariaveis, by = ticker )

```

Dessa forma a base de dados após estes ajustes, denotada por “df”, possui seis colunas: a primeira é uma coluna contendo o *ticker* de cada empresa, a segunda é uma coluna com o tempo de ocorrência do evento ou censura, a terceira é uma coluna com uma variável indicadora que assume valor um se o evento ocorreu e zero se ocorreu uma censura, a quarta coluna indica qual evento se trata (1 para 20%, 2 para 30% e 3 para 40%), enquanto as demais colunas são as colunas de covariáveis, conforme exemplificado na Tabela B.1. A partir de um conjunto de dados com o esse formato, é possível ajustar um modelo PWP com o mesmo vetor de parâmetros para cada evento conforme ilustrado no código seguinte.

```

1 library(survival)
2
3 modeloPWP <- coxph(Surv(tempo, status) ~ as.factor(underpricing) +
4                   as.factor(janela) +
5                   strata(obs.evento), cluster = ticker, data = df,
6                   method = breslow )

```

Tabela B.1: Algumas observações da base de dados utilizada na aplicação dos códigos do modelo PWP.

<b>ticker</b>	<b>tempo</b>	<b>status</b>	<b>obs.evento</b>	<b>underpricing</b>	<b>janela</b>
AALR3	24	1	1	baixo	não
AALR3	61	1	2	baixo	não
AALR3	86	1	3	baixo	não
ABCB4	45	1	1	baixo	sim
ABCB4	82	1	2	baixo	sim
ABCB4	151	1	3	baixo	sim
ABNB3	347	1	1	alto	sim
...	...	...	...	...	...