



**UNIVERSIDADE FEDERAL FLUMINENSE**  
**FACULDADE DE ADMINISTRAÇÃO E CIÊNCIAS CONTÁBEIS**  
**Programa de Pós-Graduação em Administração**  
**Mestrado Acadêmico em Administração**

**IMPACTO DO BID-ASK SPREAD NA DETERMINAÇÃO DO PREÇO NO MERCADO  
DE AÇÕES**

**LEONARDO MELLO CAMILO DA SILVA**

**Niterói**

**2024**

**LEONARDO MELLO CAMILO DA SILVA**

**IMPACTO DO BID-ASK SPREAD NA DETERMINAÇÃO DO PREÇO NO MERCADO  
DE AÇÕES**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Administração da Faculdade de Administração e Ciências Contábeis da Universidade Federal Fluminense, como requisito para obtenção do título de Mestre em Administração.

Orientador: Prof. Dr. Claudio Henrique da  
Silveira Barbedo

Niterói/RJ

2024

## Resumo

Este estudo investiga o impacto do spread bid-ask na formação de preços de ativos financeiros, com foco no mercado de ações brasileiro. A pesquisa é dividida em duas partes: uma Revisão Narrativa de Literatura (RNL) e um estudo empírico. A RNL explora teorias e abordagens relacionadas ao spread bid-ask, revisando modelos clássicos e contemporâneos.

O estudo empírico aplica métodos reconhecidas para estimar o spread bid-ask no mercado brasileiro, analisando dados de 2017 a 2023. Os resultados indicam que o spread bid-ask é crucial na formação de preços, afetando liquidez e custos de transação. A comparação entre estimadores tem como proposta a identificação da metodologia mais adequada ao mercado brasileiro, oferecendo insights para investidores e reguladores. Este trabalho contribui para a literatura ao preencher lacunas na compreensão do spread bid-ask em mercados emergentes.

**Palavras-chave:** *Bid-ask spread; Microestrutura de mercado; Liquidez de mercado; Custos de transação.*

## SUMÁRIO

1	Introdução . . . . .	8
1.1	Contexto . . . . .	8
1.1.1	Relevância e justificativa . . . . .	8
2	Objetivos . . . . .	10
3	Referencial teórico . . . . .	11
3.1	Hipótese de eficiência de mercado . . . . .	11
3.2	Microestrutura de mercado . . . . .	12
3.3	Spread de oferta e compra (spread bid-ask) . . . . .	13
3.4	Modelos de estimação de Spread bid-ask . . . . .	15
3.5	Revisão Narrativa de Literatura . . . . .	17
4	Metodologia . . . . .	19
4.1	Amostra . . . . .	19
4.2	Modelos econométricos . . . . .	20
4.3	Cálculo do spread bid-ask . . . . .	27
4.4	Métricas de avaliação . . . . .	27
4.4.1	Métricas de erro . . . . .	28
4.4.2	Teste de similaridade estatística - Kolmogorov-Smirnov . . . . .	29
5	Resultados. . . . .	30
5.1	Análise dos estimadores . . . . .	30
5.2	Análise de período de volatilidade . . . . .	35
5.3	Análise comparativa com dados de referência da Bloomberg . . . . .	41
5.3.1	Teste de similaridade estatística . . . . .	41
5.3.2	Métricas de erro . . . . .	42
5.3.3	Interpretação integrada . . . . .	44
5.3.4	Implicações para o mercado brasileiro . . . . .	45
6	Conclusão . . . . .	46
7	Referências . . . . .	48

## LISTA DE FIGURAS

## LISTA DE TABELAS

1	Tabela indicando as principais estatísticas do método EDGE . . . . .	31
2	Tabela indicando as principais estatísticas do método CS . . . . .	31
3	Tabela indicando as principais estatísticas do método AR . . . . .	32
4	Teste t para Diferença de Médias entre Medidas de Liquidez, comparando spreads de grupos de alta e baixa liquidez . . . . .	33
5	Tabela indicando as principais estatísticas do método CS para pré-pandemia, ano da pandemia (2020) e pós-pandemia. . . . .	35
6	Tabela de teste estatístico T para comparação de médias entre os grupos de CS. . . .	36
7	Tabela indicando as principais estatísticas do método EDGE para pré-pandemia, ano da pandemia (2020) e pós-pandemia. . . . .	37
8	Tabela de teste estatístico T para comparação de médias entre os grupos de EDGE . .	37
9	Tabela indicando as principais estatísticas do método AR para pré-pandemia, ano da pandemia (2020) e pós-pandemia. . . . .	38
10	Tabela de teste estatístico T para comparação de médias entre os grupos de AR . . .	39
11	Tabela indicando o resultado do teste de Kolmogorov-Smirnov para cada metodologia de estimativa de spread bid-ask . . . . .	41
12	Painel de métricas de erro RMSE e MAE por grupo de liquidez. . . . .	42
13	Métricas de erro RMSE e MAE por período. . . . .	43

## LISTA DE EQUAÇÕES

1	Estimador CS . . . . .	20
2	Parâmetro $\alpha$ do estimador CS . . . . .	21
3	Parâmetro $\beta$ do estimador CS . . . . .	21
4	Parâmetro $\gamma$ do estimador CS . . . . .	21
5	Parâmetro <i>High</i> de $\beta$ e $\gamma$ do Estimador CS . . . . .	21
6	Parâmetro <i>Low</i> de $\beta$ e $\gamma$ do Estimador CS . . . . .	21
7	Preço médio para cálculo do Estimador AR . . . . .	22
8	Distância quadrática entre o Preço de Fechamento e o Proxy de Preço Eficiente . . . . .	23
9	Estimador AR . . . . .	23
10	Estimador EDGE . . . . .	24
11	Amplitude intradiária . . . . .	25
12	Ajuste para observação discreta . . . . .	25
12	Spread estimado para EDGE . . . . .	25
13	Erro quadrático médio (EQM ou RMSE) . . . . .	28
14	Erro Absoluto Médio (MAE) . . . . .	29
15	Estatística D de Kolmogorov-Smirnov . . . . .	29

# **1 Introdução**

## **1.1 Contexto**

O spread bid-ask é uma medida fundamental na microestrutura de mercado, representando a diferença entre os preços de compra (bid) e venda (ask) de um ativo. Este spread reflete os custos de transação e a liquidez do mercado, sendo influenciado por fatores como a volatilidade do ativo, a frequência de negociação e a assimetria de informações entre os participantes do mercado. A análise do spread bid-ask é crucial para entender a eficiência do mercado e os custos enfrentados pelos investidores.

No contexto dos mercados emergentes, a B3 (Brasil, Bolsa, Balcão) desempenha um papel central. A B3 é um importante centro de negociação de ativos financeiros na América Latina, ela não só facilita a negociação de ações, mas também de derivativos, títulos de renda fixa e outros instrumentos financeiros. A bolsa tem se destacado por sua infraestrutura tecnológica avançada e por iniciativas voltadas à inclusão financeira e ao aumento da liquidez do mercado. No entanto, os mercados emergentes, como o brasileiro, ainda enfrentam desafios significativos, incluindo menor liquidez e maior volatilidade em comparação com mercados desenvolvidos.

A literatura sobre spread bid-ask tem evoluído significativamente ao longo das últimas décadas, com diversos métodos sendo desenvolvidos para sua estimativa. Métodos tradicionais, como o estimador de Roll (1984), foram complementados por abordagens mais recentes, como os estimadores de Corwin e Schultz (2012) e de Abdi e Ranaldo (2017). Apesar desses avanços, há uma lacuna notável na aplicação dessas técnicas a mercados emergentes. A maioria dos estudos concentra-se em mercados desenvolvidos, deixando uma lacuna na compreensão de como essas abordagens se comportam em contextos de menor liquidez e maior volatilidade, como o mercado brasileiro.

### **1.1.1 Relevância e justificativa**

Medir o spread bid-ask é de extrema relevância para diversos stakeholders no mercado financeiro. Corwin (2012) destaca que, para os investidores, o spread bid-ask é um indicador direto dos custos de transação e da liquidez do mercado, influenciando decisões de compra e venda de

ativos. Reguladores utilizam essa medida para monitorar a eficiência do mercado e identificar possíveis práticas de manipulação de preços. Acadêmicos, por sua vez, empregam o spread bid-ask como uma variável chave em estudos sobre microestrutura de mercado, eficiência de preços e assimetria informacional.

Este trabalho busca preencher lacunas importantes na literatura ao realizar uma análise comparativa de métodos de estimativa do spread bid-ask aplicadas ao mercado brasileiro. Ripamonti (2016) observa que a ausência de estudos que comparam diferentes técnicas nesse contexto específico representa uma oportunidade para contribuir com novos insights sobre a eficiência e a liquidez do mercado brasileiro. Os resultados deste estudo podem auxiliar investidores na seleção de ações e na avaliação de riscos de liquidez, fornecendo uma base mais sólida para suas decisões de investimento.

A escolha do mercado brasileiro se justifica por suas características únicas como mercado emergente. A B3, com sua estrutura de mercado order-driven e sua diversidade de ativos e participantes, oferece um ambiente ideal para testar a robustez das metodologias de estimativa do spread bid-ask. Além disso, Ripamonti (2016) sugere que os resultados obtidos podem ser generalizados para outros mercados emergentes com características semelhantes, ampliando o impacto e a relevância do estudo. Dessa forma, este trabalho não só contribui para a literatura acadêmica, mas também oferece implicações práticas para investidores e reguladores no contexto de mercados emergentes.



## 2 Objetivos

Considerando a relevância do spread bid-ask na formação de preços de ativos financeiros e seu impacto na liquidez e nos custos de transação, este estudo tem como objetivo geral comparar, no mercado acionário brasileiro, estimativas de spread bid-ask obtidas pelas abordagens de Corwin & Schultz (2012), Abdi & Ranaldo (2017) e EDGE (Ardia, Guidotti & Kroencke, 2024), tendo como referência a cotação de spread divulgada pela Bloomberg.

A análise comparativa das metodologias selecionadas abrangerá tanto períodos de relativa estabilidade quanto momentos de volatilidade acentuada, com especial atenção ao período da pandemia de COVID-19 em 2020. Esta abordagem permitirá avaliar não apenas a precisão geral dos estimadores, mas também sua robustez e capacidade de capturar as variações nos custos de transação em condições extremas de mercado, fornecendo insights valiosos sobre a microestrutura do mercado brasileiro em diferentes contextos econômicos.

### 3 Referencial teórico

#### 3.1 Hipótese de eficiência de mercado

De acordo com Fama (1970) a Hipótese de eficiência de mercado propõe que os investidores do mercado irão fazer a melhor alocação de ativos possível com todas as informações relevantes que terão acesso, e os preços dos ativos irá refletir todas as informações disponíveis para os investidores, que se basearão na teoria econômica para realizar tais cálculos. Em Finanças, a HEM (Hipótese da Eficiência de Mercado) afirma que os mercados financeiros são “eficientes em relação à informação”. Ou seja, um agente não consegue alcançar consistentemente retornos superiores à média do mercado, considerando as informações publicamente disponíveis no momento em que o investimento é feito (Fama, 1965; Malkiel, 2007; Jensen, 1978).

Existem três versões principais da hipótese: a fraca, a semi-forte e a forte. A hipótese fraca considera que os preços negociados para os bens (por exemplo, ações, obrigações ou propriedade) refletem toda a informação histórica disponível publicamente. Em seu trabalho seminal, Fama (1965) demonstra que os preços em mercados eficientes seguem um passeio aleatório, refletindo todas as informações históricas disponíveis, o que implica que os preços futuros não podem ser previstos com base em padrões passados.

A hipótese semi-forte afirma que os preços refletem todas as informações publicamente disponíveis e que os preços mudam instantaneamente para refletir as novas informações públicas. Esse conceito é extensivamente discutido no trabalho de Fama (1970), onde ele categoriza a eficiência de mercado em formas fraca, semi-forte e forte. A forma semi-forte da HEM sugere que nem mesmo a análise fundamentalista pode proporcionar vantagens, pois todas as informações disponíveis publicamente já estão incorporadas nos preços dos ativos. Estudos empíricos, como os de Malkiel (2007), corroboram essa visão ao demonstrar que as tentativas de prever movimentos de preços com base em análises fundamentais geralmente não resultam em retornos acima da média.

Embora a hipótese de eficiência de mercado proponha que os preços dos ativos refletem todas as informações disponíveis, o spread bid-ask é um fator crucial que pode introduzir ineficiências no mercado. Custos de transação elevados, informação assimétrica, baixa liquidez e problemas de microestrutura do mercado são elementos que podem ampliar o spread e impedir que

os preços se ajustem rapidamente às novas informações. Isso sugere que, na prática, os mercados podem não ser tão eficientes quanto a teoria propõe, e os investidores precisam considerar esses fatores ao tomar decisões de investimento.

### **3.2 Microestrutura de mercado**

A microestrutura de mercado é um campo de estudo que analisa os mecanismos e processos pelos quais os preços dos ativos são formados nos mercados financeiros. Este campo abrange uma variedade de tópicos, incluindo a organização e funcionamento dos mercados, o comportamento dos participantes do mercado, e os custos de transação. Segundo Harris (2003), a microestrutura de mercado examina como a estrutura do mercado e os mecanismos de negociação afetam a formação de preços, a liquidez e a eficiência do mercado.

Compreender a microestrutura é fundamental para analisar como as informações são incorporadas nos preços e como os custos de transação influenciam as decisões de investimento. Os mercados financeiros podem ser organizados de várias maneiras, incluindo mercados de leilão e mercados de dealer. Em mercados de leilão, como as bolsas de valores, os preços são estabelecidos através de leilões competitivos, onde compradores e vendedores submetem ordens que são combinadas para formar o preço de mercado. Nos mercados de dealer, como o mercado de câmbio, os dealers fornecem liquidez ao comprar e vender ativos de seu próprio inventário.

Harris (2003) destaca a importância do livro de ordens, que registra todas as ordens de compra e venda pendentes e desempenha um papel crucial na determinação dos preços de mercado. A liquidez do mercado, definida por como a facilidade com que um ativo pode ser comprado ou vendido sem causar uma mudança significativa em seu preço, é fortemente influenciada pela microestrutura do mercado (Harris, 2003). Chordia, et. al (2000) mostram que a estrutura do mercado e a presença de intermediários financeiros desempenham um papel vital na determinação da liquidez. A presença de market makers, por exemplo, pode aumentar a liquidez ao fornecer cotações firmes de compra e venda e ao absorver temporariamente desequilíbrios entre oferta e demanda. Além disso, a profundidade do livro de ordens e a quantidade de informações disponíveis aos participantes do mercado também afetam a liquidez.

Os custos de transação são um componente crucial da microestrutura de mercado e têm

um impacto significativo na eficiência do mercado. Stoll (1989) e Glosten e Milgrom (1985) identificam três principais componentes dos custos de transação: custos de processamento das ordens, custos de assimetria de informação e custos de inventário. Os custos de processamento referem-se aos custos operacionais incorridos pelos intermediários ao executar ordens de compra e venda. Os custos de assimetria de informação surgem da incerteza sobre a informação que outros participantes do mercado possuem. Já os custos de inventário são os custos associados à manutenção de um estoque de ativos para facilitar a negociação. A eficiência do mercado é afetada por esses custos, pois eles influenciam a capacidade dos preços refletirem todas as informações disponíveis.

As inovações tecnológicas, particularmente o trading eletrônico e os algoritmos de negociação, têm transformado a microestrutura dos mercados financeiros. Hendershott, Jones e Menkveld (2011) investigam o impacto do trading algorítmico na liquidez do mercado e concluem que ele melhora a liquidez ao aumentar a frequência e a velocidade das transações. A introdução de plataformas de negociação eletrônica reduziu os custos de transação e aumentou a transparência do mercado, permitindo uma melhor incorporação de informações nos preços. No entanto, essas inovações também trazem desafios, como a necessidade de regulamentos adequados para mitigar os riscos associados ao trading de alta frequência e para garantir a estabilidade do mercado.

### **3.3 Spread de oferta e compra (spread bid-ask)**

Custos de inventário ou de estoque, segundo De Jong e Rindi (2009), existem apenas em mercados quote-driven, pois nesses mercados os formadores têm a obrigação de fornecer preços de compra e preços de venda ininterruptamente ao mercado. Os custos relacionados à assimetria de informação e de transação ou de processamento de ordens existem em qualquer mercado. O mercado de ações da B3 é order-driven e nesse caso não se identifica o custo de estoque.

Nos mercados quote-driven, também conhecidos como mercados de dealer, um intermediário (dealer) participa de todas as transações. Nesse tipo de mercado, os traders negociam diretamente com o dealer, que fornece cotações dos preços de compra e venda, assumindo o lado oposto de cada transação e, assim, proporcionando liquidez ao mercado. Esses mercados são comuns no comércio de moedas estrangeiras e títulos no mercado de balcão (OTC).

Por outro lado, nos mercados order-driven, os compradores e vendedores negociam

diretamente entre si, sem a intervenção de um dealer, utilizando um conjunto de regras de negociação estabelecidas. Essas regras incluem precedência de ordens, que determina como as ordens de compra e venda são casadas, e regras de precificação de transações. Os mercados order-driven são prevalentes em bolsas de futuros, ações e opções, bem como em redes eletrônicas de comunicação para negociação de títulos e moedas. Em suma, a principal diferença reside no papel do intermediário: enquanto nos mercados quote-driven o dealer facilita todas as transações fornecendo cotações, nos mercados order-driven, as transações são realizadas diretamente entre os traders com base em regras de mercado (Peat, 2009).

No contexto do mercado de ações no Brasil, a B3 adota um modelo de mercado order-driven, mas algumas ações contam com a presença de formadores de mercado. Os formadores de mercado, ou market makers, são contratados pelas empresas para aumentar a liquidez de suas ações. A liquidez refere-se à facilidade com que um ativo pode ser comprado ou vendido sem causar uma mudança significativa em seu preço (Harris, 2003). A presença de um formador de mercado pode reduzir o spread de compra e venda, o que diminui os custos de transação para os investidores e aumenta a atratividade da ação. Empresas contratam formadores de mercado para aumentar a liquidez de suas ações, melhorar a precificação e atrair mais investidores, especialmente em períodos de baixa negociação ou para novos papéis que estão sendo introduzidos no mercado.

A assimetria de informação, por sua vez, ocorre quando diferentes participantes do mercado têm diferentes níveis de acesso à informação relevante para a precificação dos ativos. Nos mercados order-driven, a assimetria de informação pode ser mitigada pela transparência do livro de ordens e pelas regras claras de negociação. No entanto, a assimetria ainda pode existir, especialmente em mercados menos líquidos ou em ações com menor cobertura de analistas. Em mercados quote-driven, a assimetria de informação pode ser mais pronunciada, pois os dealers podem ter acesso a informações privilegiadas que não estão disponíveis para todos os participantes do mercado.

Além disso, os custos de transação referem-se aos custos associados à realização de uma transação econômica, que incluem os custos de negociação e de contratação (Demsetz, 1968). Esses custos incluem taxas pagas à bolsa, comissões de corretagem e custos relacionados ao tempo e esforço necessários para processar as ordens. Em mercados order-driven como a B3, esses custos podem ser reduzidos pela automação e eficiência das plataformas de negociação eletrônica, que

facilitam a correspondência rápida e precisa das ordens de compra e venda.

Em suma, a presença de formadores de mercado em ações na B3 visa principalmente aumentar a liquidez, reduzir o spread de compra e venda e melhorar a eficiência do mercado. Ao contratar formadores de mercado, as empresas buscam criar um ambiente de negociação mais ativo e atraente para os investidores, promovendo assim uma melhor precificação e maior volume de negociações. A compreensão dos diferentes custos associados ao spread de compra e venda, bem como das diferenças entre mercados quote-driven e order-driven, é essencial para avaliar o impacto da microestrutura de mercado na formação de preços dos ativos financeiros.

### **3.4 Modelos de estimação de Spread bid-ask**

A mensuração do spread bid-ask pode ser realizada por meio de dados de alta frequência, que registram de forma contínua as cotações e negociações ao longo do dia, ou através de dados de baixa frequência, os quais utilizam informações agregadas, como os preços de fechamento, máxima e mínima diários. Em contextos onde os dados de alta frequência são escassos ou de difícil acesso, os métodos de estimação baseados em dados agregados tornam-se fundamentais para a avaliação dos custos de transação e da liquidez dos ativos.

Para fundamentar este trabalho, foi realizada uma Revisão Narrativa de Literatura (RNL) durante a etapa de qualificação desta dissertação, a qual permitiu identificar os métodos mais relevantes e inovadoras no campo da estimação do spread bid-ask. A partir dessa revisão, foram selecionadas três técnicas de estimação de baixa frequência que vêm contribuindo de maneira significativa para aprimorar a medição do spread bid-ask e que apresentam potencial aplicabilidade ao mercado brasileiro.

A abordagem de Corwin e Schultz (2012) introduz uma forma simplificada de inferir o spread a partir das oscilações diárias, enquanto a proposta de Abdi e Rinaldo (2017) aprimora esse método ao reduzir incertezas inerentes às estimativas com dados diários. Por fim, a metodologia de Ardia, Guidotti e Kroencke (2024) integra informações adicionais, como o preço de abertura, proporcionando uma estimativa mais robusta e precisa.

O modelo de Corwin e Schultz (2012) foi um dos primeiros a propor uma metodologia baseada apenas em preços diários de máxima e mínima para estimar o spread bid-ask. A ideia

central do modelo é que o intervalo entre os preços mais altos e mais baixos de um dia reflete tanto a volatilidade do ativo quanto a diferença entre os preços de compra e venda. Assim, os autores derivam um estimador baseado na comparação entre a soma dos intervalos de dois dias consecutivos e o intervalo total de dois dias. Esse modelo trouxe uma abordagem simples e aplicável a mercados onde não há disponibilidade de dados de cotações intradiárias

Abdi e Ranaldo (2017) expandiram essa metodologia ao incluir informações do preço de fechamento, além dos preços de máxima e mínima diários. O modelo desenvolvido pelos autores tem como premissa a ideia de que o preço de fechamento contém informações adicionais sobre o comportamento do spread bid-ask. Utilizando uma formulação matemática que combina essas três variáveis, o estimador proposto por Abdi e Ranaldo mostrou-se mais preciso para ativos com menor liquidez, superando modelos anteriores em termos de correlação com medidas de spread obtidas a partir de dados de alta frequência

Mais recentemente, Ardia, Guidotti e Kroencke (2024) propuseram um modelo ainda mais abrangente, que incorpora informações dos preços de abertura, máxima, mínima e fechamento. Diferentemente dos modelos anteriores, que ignoram o preço de abertura, essa nova abordagem busca reduzir vieses na estimação ao aproveitar um conjunto mais completo de informações de preços. Além disso, os autores desenvolveram um estimador eficiente que minimiza a variância da estimativa, tornando-se uma alternativa superior para mensurar spreads em mercados com diferentes níveis de liquidez. O modelo de Ardia et al. (2024) representa um avanço na literatura ao abordar problemas de viés e eficiência estatística na estimação do spread bid-ask

Dessa forma, a evolução dos métodos de estimação do spread bid-ask reflete a necessidade de desenvolver modelos que equilibrem simplicidade e precisão, especialmente em mercados onde dados de alta frequência não estão disponíveis. A aplicação desses modelos ao mercado brasileiro permite uma avaliação mais detalhada da liquidez dos ativos e dos custos de transação, contribuindo para uma melhor compreensão das dinâmicas de precificação no mercado acionário nacional. Além disso, tais estimadores são fundamentais para o desenvolvimento de programas computacionais que viabilizem a estimação contínua do spread bid-ask, permitindo uma análise mais eficiente do comportamento do mercado.

### 3.5 Revisão Narrativa de Literatura

Para fundamentar a compreensão teórica e metodológica deste estudo, foi realizada uma Revisão Narrativa de Literatura (RNL), diferenciando-se da Revisão Sistemática de Literatura (RSL) por permitir maior flexibilidade na seleção e análise dos estudos. A abordagem qualitativa e descritiva da RNL possibilitou explorar e sintetizar os principais achados e teorias relacionadas ao spread bid-ask e sua influência na formação de preços de ativos financeiros. A escolha desta metodologia se justificou pela necessidade de construir um panorama mais amplo sobre o spread bid-ask no contexto brasileiro, onde a literatura ainda é incipiente.

A RNL foi conduzida nas bases de dados Web of Science, Scopus, Periódicos Capes e Google Scholar, com foco na identificação de estudos que respondessem à questão central: “O artigo identificou que o spread bid-ask é importante para a formação de preços?”. Foram selecionados 19 artigos, divididos entre artigos seminais internacionais (11) e estudos específicos sobre o mercado brasileiro (8). Os critérios de inclusão contemplaram estudos que abordassem diretamente a relação entre o spread bid-ask e a formação de preços, utilizando diversas metodologias e tipos de dados. Estudos sem análise empírica ou indisponíveis em texto completo foram excluídos.

A análise dos artigos selecionados revelou que a maioria dos estudos utilizou modelos econométricos, especialmente regressões de Mínimos Quadrados Ordinários (MQO), para investigar a influência do spread bid-ask na formação de preços. Quanto aos tipos de dados, predominaram pesquisas com dados de baixa frequência, como preços diários de abertura, fechamento, máximo e mínimo. Os mercados examinados incluíram tanto o mercado brasileiro quanto mercados internacionais, com destaque para estudos sobre o mercado de ações e debêntures no Brasil (SND) e o mercado de ações dos Estados Unidos (NYSE). Os aspectos mais analisados foram custos de transação, liquidez, assimetria de informação e volatilidade.

Os resultados da RNL confirmaram a relevância do spread bid-ask como componente essencial na microestrutura de mercado, influenciando diretamente a liquidez, os custos de transação e a eficiência dos mercados financeiros. Identificou-se uma clara evolução metodológica na estimativa do spread bid-ask, desde abordagens clássicas como a de Roll (1984) até estimadores contemporâneos como os de Corwin e Schultz (2012), Abdi e Ranaldo (2017) e Ardia, Guidotti e Kroencke (2024). Esta evolução reflete a busca por metodologias mais precisas e adaptáveis a



diferentes contextos de mercado, especialmente em mercados emergentes com características de liquidez heterogêneas. A RNL foi fundamental para identificar as três metodologias utilizadas no presente estudo, selecionadas por sua robustez teórica, ampla aceitação na literatura e capacidade de oferecer insights em diferentes contextos de mercado, especialmente para o caso brasileiro.

## **4 Metodologia**

Este estudo foi desenvolvido para permitir uma análise rigorosa e comparativa das diferentes metodologias de estimativa do spread bid-ask no mercado de ações brasileiro. A seguir, são detalhadas as etapas e procedimentos adotados para a coleta e análise dos dados, bem como a aplicação das metodologias selecionadas.

### **4.1 Amostra**

A base de dados deste estudo foi composta a partir das informações de preços diários das ações listadas na B3, aplicando critérios rigorosos de seleção para garantir a representatividade e a qualidade dos dados utilizados. O critério principal para inclusão das ações na amostra foi o índice de liquidez, calculado conforme o método descrita por Andrade et al. (2009). Foram consideradas apenas as ações com índice de liquidez superior a 0,1, calculado com base em uma janela de seis meses de dados, assegurando a representatividade de ativos com maior frequência de negociação.

A amostra final é composta por 130 ações, distribuídas entre os diferentes segmentos de governança corporativa da B3. Entre elas, 93 pertencem ao Novo Mercado, 16 ao Nível 2, 17 ao Nível 1 e 4 ao segmento Tradicional. Essa segmentação reflete distintos níveis de práticas de governança corporativa, permitindo uma análise abrangente sobre como essas características impactam os spreads bid-ask e a liquidez dos ativos.

A definição da amostra seguiu abordagens amplamente utilizadas na literatura, como a adotada por Andrade et al. (2009), que empregaram critérios de liquidez para selecionar empresas listadas. Empresas do setor financeiro foram excluídas devido às particularidades de suas demonstrações contábeis, que poderiam introduzir vieses na análise. Além disso, foram descartadas companhias com dados incompletos ou inconsistentes, assegurando que a amostra final apresentasse informações confiáveis para a investigação empírica.

O processo de filtragem das ações visou garantir a robustez dos dados utilizados, selecionando ativos que refletissem, de forma mais precisa, o valor de mercado real. A inclusão de ações com maior liquidez não apenas melhora a representatividade da amostra, mas também favorece a precisão das estimativas econométricas realizadas no estudo, reduzindo potenciais

vieses causados por falta de negociação.

## 4.2 Modelos econométricos

Este estudo aplicou três métodos amplamente reconhecidos para a estimativa do spread bid-ask: CS, AR e EDGE utilizando dados obtidos diretamente da B3. Essas abordagens foram selecionadas devido à sua robustez teórica, ampla aceitação na literatura e capacidade de oferecer insights em diferentes contextos de mercado.

Os artigos que fundamentam estes métodos apresentam significativo impacto acadêmico, evidenciado pelo número de citações: o modelo de Corwin e Schultz (2012) possui 1387 citações, seguido pelo método de Abdi e Ranaldo (2017) com 404 citações e o recente EDGE de Ardia, Guidotti e Kroenke (2024) com 18 citações, conforme dados do Google Scholar em abril de 2025. Esses números refletem a relevância e a aceitação das metodologias na comunidade acadêmica, destacando sua importância na análise do spread bid-ask e na microestrutura de mercado.

O estimador de spread bid-ask desenvolvido por Corwin e Schultz (2012) utiliza os preços diários de máxima e mínima das ações para calcular o spread de forma simples e eficiente. A metodologia parte de duas premissas principais. Primeiramente, assume-se que os preços máximos tendem a ser resultados de ordens de compra, enquanto os mínimos são influenciados por ordens de venda. Dessa forma, a relação entre os preços de máxima e mínima reflete tanto a volatilidade fundamental do ativo quanto os custos associados ao spread bid-ask. Em segundo lugar, considera-se que a componente de volatilidade aumenta proporcionalmente com o intervalo de negociação, enquanto a parte do spread permanece constante.

Matematicamente, o estimador é definido pela seguinte expressão:

$$S = \frac{2(e^\alpha - 1)}{1 + e^\alpha}, \quad (1)$$

onde  $S$  representa o spread bid-ask estimado e  $e^\alpha$  denota a função exponencial de  $\alpha$  com base  $e$  (número de Euler). O parâmetro  $\alpha$  é calculado por:

$$\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}}, \quad (2)$$

sendo  $\beta$  e  $\gamma$  definidos a partir dos logaritmos naturais das razões entre os preços máximos, H, e mínimos, L, observados:

$$\beta = E \left[ \sum_{j=0}^1 \left( \ln \frac{H_{t+j}}{L_{t+j}} \right)^2 \right], \quad (3)$$

$$\gamma = E \left[ \left( \ln \frac{H_{t,t+1}}{L_{t,t+1}} \right)^2 \right], \quad (4)$$

onde

$$H_{t,t+1} = \max(H_t, H_{t+1}) \quad (5)$$

e

$$L_{t,t+1} = \min(L_t, L_{t+1}) \quad (6)$$

Nesse contexto, a equação 3 captura as variações diárias entre os preços máximos e mínimos, enquanto a equação 4 reflete a dinâmica entre dois dias consecutivos. Essas variáveis permitem separar a volatilidade fundamental do ativo da componente relacionada ao spread bid-ask.

A partir dessas definições, torna-se evidente que o método permite isolar o impacto do spread bid-ask dos movimentos puramente voláteis do preço. A implementação do estimador é direta, dispensando iterações complexas, o que facilita sua aplicação prática em grandes conjuntos de dados. Corwin e Schultz (2012) validaram o estimador por meio de simulações que replicaram condições realistas de mercado, demonstrando que a correlação entre os valores estimados pelo método e os

spreads reais é elevada, em torno de 0,9. Em comparação com o estimador de covariância de Roll (1984), o desvio padrão das estimativas de Corwin e Schultz é cerca de metade, indicando maior estabilidade nos resultados.

O método foi aplicado a ações de 11 países, incluindo o Brasil (Ripamonti, 2016; Ichimura, Videira, Ripamonti, 2020), evidenciando sua eficácia em medir os custos de transação em diferentes contextos de liquidez e volatilidade. Essa técnica tem sido amplamente utilizada em estudos sobre microestrutura de mercado. Por exemplo, foi empregada para analisar a relação entre a qualidade das informações financeiras e a assimetria informacional, com resultados que destacaram sua eficiência em comparação a outros indicadores. Embora algumas pesquisas apontem limitações no uso do método como medida de liquidez para prever retornos, ele continua sendo uma ferramenta confiável para estimar custos de transação e compreender a dinâmica dos mercados financeiros.

Portanto, o estimador de Corwin e Schultz (2012) oferece uma solução prática e robusta para mensurar o spread bid-ask com base em dados diários, sendo amplamente adotado em estudos de microestrutura, especialmente em mercados onde dados intradiários são limitados. Sua simplicidade e eficácia tornam-no uma ferramenta valiosa para analisar a liquidez e a eficiência do mercado financeiro brasileiro.

Já o estimador de Abdi e Ranaldo (2017) trouxe uma abordagem nova para o cálculo do spread, sendo desenvolvida para estimar o spread bid-ask utilizando preços diários de fechamento, máxima e mínima das ações. A abordagem se baseia na decomposição do spread em componentes de volatilidade e custos de transação, permitindo uma análise detalhada dos fatores que influenciam a formação de preços nos mercados financeiros. Esse estimador é fundamentado em três etapas principais. Primeiro, utiliza-se o preço médio entre a máxima e a mínima (mid-range) como um proxy para o preço eficiente, que é definido como:

$$MR_t = \frac{H_t + L_t}{2} \quad (7)$$

Supondo que o spread bid-ask se cancela nesse cálculo.

Em seguida, calcula-se a distância quadrática entre o preço de fechamento e o proxy de

preço eficiente derivado da média das médias de dois dias consecutivos, com essa distância sendo representada como:

$$D_t = (C_t - MR_t)^2 \quad (8)$$

Por fim, essa distância é ajustada para remover a variância do preço eficiente,  $\sigma_D^2$ , isolando assim o componente relacionado ao spread bid-ask. Essa abordagem, além de simplificar cálculos, não requer ajustes *ad hoc* para períodos sem negociação, como finais de semana e feriados, o que a torna menos suscetível a vieses comuns em outros modelos.

No final, a fórmula do estimador é dada por:

$$S = 2\sqrt{\sigma_D^2}, \quad (9)$$

onde  $S$  é o spread bid-ask estimado, e  $\sigma_D^2$  é a variância do termo  $D_t$ , ajustada com base nos preços de dois dias consecutivos. Essa formulação permite capturar de maneira mais precisa os custos de transação associados às negociações

Adicionalmente, o modelo de Abdi e Rinaldo (2017) se destaca em cenários de baixa liquidez. Ao contrário do estimador de Corwin-Schultz (2012), que apresenta viés descendente em ativos menos líquidos, esse método AR foi dito como mais estável e robusto em condições de mercado desafiadoras, o tornando uma ferramenta valiosa para analisar a eficiência e a liquidez em mercados emergentes e em desenvolvimento, onde a disponibilidade de dados intradiários é limitada. A relevância desse estimador é evidente ao compará-lo com outros modelos, como o de Corwin-Schultz. Enquanto o estimador de Corwin-Schultz depende de uma estrutura não-linear para decompor a volatilidade do preço e o spread bid-ask, o modelo de Abdi e Rinaldo (2017) utiliza uma solução em forma fechada, simplificando o processo de estimação e reduzindo a necessidade de premissas adicionais, como a igualdade de chances de preços de alta e baixa serem iniciados por compradores ou vendedores.

Além disso, o método é amplamente aplicável, tendo sido validado em Abdi e Rinaldo

(2017) em uma extensa amostra de ações nos EUA desde 1926, demonstrando alto grau de correlação com spreads efetivos medidos por dados intradiários (TAQ). Isso confirma sua eficácia como uma ferramenta para mensurar custos de transação e liquidez em longos períodos históricos, contribuindo para a análise de risco de liquidez sistemática e de commonality em liquidez.

O método EDGE, desenvolvido por Ardia, Guidotti e Kroencke (2024), surge como uma proposta robusta para estimar o spread bid-ask de maneira eficiente, utilizando preços diários de abertura, máxima, mínima e fechamento (OHLC). Este método aborda uma limitação crítica de estimadores anteriores, como os de Corwin e Schultz (2012) e Abdi e Ranaldo (2017): o viés introduzido pela suposição de preços observados continuamente, que não reflete a realidade de mercados com baixa frequência de negociações.

O EDGE parte de três premissas centrais, compartilhadas com outros métodos: (1) os retornos fundamentais não são autocorrelacionados; (2) os retornos fundamentais são independentes das flutuações causadas pelo spread bid-ask; e (3) as flutuações do spread não possuem média e são não correlacionadas. No entanto, ele se diferencia ao relaxar a suposição de que preços são continuamente observados. Em mercados reais, a quantidade de negociações por intervalo de tempo é finita, o que resulta em um viés de subestimação dos spreads, especialmente em ativos com baixa liquidez. O EDGE corrige essa limitação introduzindo termos analíticos que ajustam o viés decorrente da observação discreta de preços.

Matematicamente, o spread bid-ask estimado pelo EDGE é definido como:

$$S = f(O_t, H_t, L_t, C_t), \quad (10)$$

onde  $S$  é o spread estimado, e a função  $f$  combina as estatísticas derivadas dos preços OHLC. Em essência, o método utiliza a amplitude intradiária normalizada para capturar a dinâmica entre o preço eficiente e o spread bid-ask. A amplitude é calculada como:

$$A_t = \ln \frac{H_t}{L_t}, \quad (11)$$

e o ajuste para a observação discreta de preços é dado por:

$$D_t = \ln \frac{C_t}{O_t}, \quad (12)$$

onde  $A_t$  reflete a volatilidade e  $D_t$  corrige o viés causado pela não observação contínua dos preços. O spread estimado é então ajustado por um termo de correção baseado na variação observada em múltiplos dias consecutivos, representado como:

$$S = \alpha \cdot A_t + \beta \cdot D_t,$$

em que os coeficientes  $\alpha$  e  $\beta$  são calibrados para garantir que o estimador seja imparcial e apresente a menor variância possível. Esses coeficientes são derivados empiricamente, considerando simulações e validações em diferentes mercados.

De acordo com Ardia, Guidotti e Kroencke (2024), uma das características mais inovadoras do EDGE é a combinação otimizada de estimadores derivados de diferentes combinações de preços (como abertura e fechamento, máxima e mínima). Isso é feito para minimizar a variância de estimação, resultando em um estimador eficiente que se destaca tanto em cenários de spreads baixos quanto elevados. Segundo o artigo, enquanto estimadores como o de Corwin e Schultz (2012) são mais precisos para spreads pequenos, e o de Abdi e Ranaldo (2017) apresenta menor viés para spreads grandes, o EDGE combina o melhor desses cenários, garantindo estimativas com menor variância e viés uniforme.

Comparando com métodos anteriores, como o de Abdi e Ranaldo, o EDGE generaliza sua abordagem ao considerar informações adicionais de preços de abertura e fechamento. Enquanto o estimador de Abdi e Ranaldo utiliza médias de preços máximos, mínimos e de fechamento para isolar o componente de volatilidade do spread, o EDGE aproveita todas as informações de OHLC e introduz ajustes analíticos para corrigir observações discretas. Além disso, o EDGE não requer



ajustes adicionais para períodos sem negociação, como finais de semana e feriados, assim como o estimador de Abdi e Ranaldo.

Sendo assim, o EDGE é uma contribuição relevante para a literatura de microestrutura de mercado, oferecendo uma ferramenta confiável e eficiente para a estimação de custos de transação. Ele não apenas melhora a precisão das estimativas em relação a métodos tradicionais, mas também amplia sua aplicabilidade a diversos contextos de mercado, como períodos históricos e mercados emergentes.

Os três estimadores discutidos – Corwin-Schultz (CS), Abdi-Ranaldo (AR) e EDGE – foram escolhidos para este estudo com base nos achados da Revisão Narrativa de Literatura (RSL), que identificou estes modelos como os mais relevantes e inovadores para a estimação do spread bid-ask no contexto de mercados emergentes. A RSL permitiu mapear a evolução das metodologias de estimação, desde abordagens mais tradicionais até as contribuições mais recentes, destacando estes três estimadores por sua relevância teórica, robustez metodológica e capacidade de oferecer insights em diferentes contextos de mercado. Além disso, suas características complementares permitem uma análise comparativa detalhada de suas performances no mercado brasileiro, que é classificado como emergente.

A aplicação dos métodos será realizada em três períodos distintos: pré-pandemia, durante a pandemia e no pós-pandemia. O período pré-pandemia é caracterizado por uma maior estabilidade econômica e condições de mercado mais previsíveis, oferecendo uma base para avaliar o desempenho dos estimadores em um cenário mais controlado. Durante a pandemia, o mercado foi marcado por alta volatilidade e incertezas, o que representa um contexto de maior desafio para os estimadores, especialmente em ativos menos líquidos. Por fim, o período pós-pandemia reflete um mercado em processo de readaptação, com uma recuperação gradual da liquidez e uma nova estabilidade, permitindo verificar a consistência dos métodos em cenários intermediários.

Com isso, este estudo busca avaliar como cada método se adapta às características específicas do mercado brasileiro e dos diferentes contextos temporais, contribuindo para a literatura de microestrutura ao explorar os pontos fortes e limitações de cada estimador em mercados emergentes e desafiadores.

### 4.3 Cálculo do spread bid-ask

As três metodologias selecionadas foram: EDGE (Ardia, Guidotti e Kroencke, 2024), CS (Corwin e Schultz, 2012) e AR (Abdi e Rinaldo, 2017). O uso de uma janela móvel de 21 dias, representada pelo parâmetro  $width = 21$ , foi adotado para garantir consistência metodológica e facilitar a comparação com estudos anteriores, como os de Ardia, Guidotti e Kroencke (2024), que seguem o mesmo princípio. Essa configuração considera 21 dias como equivalente a aproximadamente um mês de negociações, com base na prática padrão em mercados financeiros, onde a média de dias úteis de negociação em um mês é de cerca de 21.

Além disso, no estudo de Ardia, Guidotti e Kroencke (2024), a aplicação do estimador é baseada em períodos de 21 dias úteis para representar um mês de negociações. Essa abordagem é consistente com os trabalhos de Corwin e Schultz (2012) e Abdi e Rinaldo (2017), que também utilizam essa mesma configuração temporal em suas análises e simulações. O objetivo é criar uma unidade temporal padronizada que permita avaliar variações no spread bid-ask de maneira comparável ao longo de diferentes períodos e contextos de negociação.

Essa escolha também facilita a análise do impacto de diferentes frequências de negociação, uma vez que uma janela de 21 dias captura tanto oscilações diárias quanto tendências mais amplas em cenários de alta e baixa liquidez. Portanto, a adoção desse parâmetro não só mantém a consistência com a literatura, mas também reflete práticas amplamente aceitas no campo de microestrutura de mercado, garantindo a robustez e a validade dos resultados obtidos no contexto do mercado brasileiro.

O cálculo foi realizado sobre as séries de preços de abertura, máxima, mínima, fechamento (OHLC) e volumes diários obtidos diretamente da B3. Os métodos foram aplicados sem modificações adicionais, seguindo rigorosamente os algoritmos descritos nos trabalhos originais, de forma a garantir a comparabilidade com os benchmarks teóricos.

### 4.4 Métricas de avaliação

As estimativas obtidas pelas três abordagens foram comparadas com os valores de referência fornecidos pela Bloomberg, considerados o padrão-ouro para a estimativa do spread bid-ask. Para

realizar essa análise, os dados da Bloomberg foram importados e integrados à base processada, permitindo a construção de comparativos entre as estimativas calculadas e os valores reportados. Essa integração foi necessária para garantir uma avaliação precisa da performance de cada método em relação a uma referência amplamente reconhecida no mercado financeiro. Foram utilizadas métricas estatísticas para medir a proximidade das estimativas com os valores de referência, como o erro quadrático médio (MSE), que avalia a magnitude do erro entre as estimativas, e o coeficiente de correlação de Pearson, que mede a força e a direção da associação entre os valores estimados e os da Bloomberg. Esses indicadores forneceram uma base objetiva para identificar as condições em que os estimadores apresentaram melhor ou pior desempenho, destacando suas capacidades e limitações em diferentes contextos de negociação.

#### 4.4.1 Métricas de erro

Para quantificar a precisão das estimativas geradas por cada abordagem em relação aos valores oficiais da Bloomberg, foram utilizadas as seguintes métricas de erro:

**4.4.1.1 Erro Quadrático Médio (RMSE)** O RMSE é uma medida amplamente utilizada para avaliar a magnitude dos erros de estimação, calculada pela raiz quadrada da média dos quadrados das diferenças entre valores estimados e valores reais:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

onde  $y_i$  representa o valor oficial do spread (Bloomberg) e  $\hat{y}_i$  representa o valor estimado pela metodologia em análise. O RMSE atribui maior peso a erros maiores devido à operação de elevação ao quadrado, tornando-o particularmente sensível a outliers. Esta métrica foi escolhida como principal indicador de proximidade entre as estimativas e os valores reais, sendo que valores menores de RMSE indicam maior precisão do método.

**4.4.1.2 Erro Absoluto Médio (MAE)** Como métrica complementar, foi utilizado o MAE, calculado pela média dos valores absolutos das diferenças entre valores estimados e valores reais:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

O MAE atribui peso igual a todos os erros, independentemente de sua magnitude, oferecendo uma perspectiva alternativa sobre a precisão das estimativas.

#### 4.4.2 Teste de similaridade estatística - Kolmogorov-Smirnov

Para avaliar a similaridade estatística entre as distribuições das estimativas e dos valores oficiais, foi aplicado o teste de Kolmogorov-Smirnov (KS). Este teste não paramétrico avalia a hipótese nula de que duas amostras provêm da mesma distribuição. A estatística D do teste KS é calculada como:

$$D = \sup_x |F_1(x) - F_2(x)| \quad (15)$$

onde  $F_1(x)$  e  $F_2(x)$  são as funções de distribuição acumulada empíricas para as duas amostras sendo comparadas, e  $\sup_x$  representa o supremo (máximo) das diferenças entre as duas distribuições. A interpretação dos resultados do teste KS é baseada no p-valor:

- p-valor < 0,05: Rejeita-se a hipótese nula, indicando diferença estatisticamente significativa entre as distribuições.
- p-valor  $\geq$  0,05: Não se rejeita a hipótese nula, sugerindo que não há evidência estatística suficiente para afirmar que as distribuições são diferentes.

## 5 Resultados

O estimador EDGE tem sua construção baseada na combinação otimizada de diferentes estimadores, minimizando a variância da estimativa final. Essa abordagem faz com que o método seja eficaz tanto em cenários de spreads estreitos quanto amplos, além de se adaptar automaticamente a variações na volatilidade do mercado. Dessa forma, a abordagem EDGE se mostra particularmente superior nos períodos de alta volatilidade, nos quais outras abordagens apresentam desempenho inconsistente.

A seguir, são apresentadas as análises comparativas entre os diferentes estimadores, com o objetivo de avaliar suas performances no contexto do mercado acionário brasileiro e identificar qual metodologia oferece maior precisão na mensuração do spread bid-ask.

### 5.1 Análise dos estimadores

Como visto na Tabela 1, com as estatísticas do EDGE, na Tabela 2, com as estatísticas do CS, e na Tabela 3, com as estatísticas do AR, ao analisar os resultados obtidos, observa-se que o estimador EDGE apresentou valores mínimos de spread bid-ask de  $1,6 \times 10^{-5}$  para ativos mais líquidos,  $2,4 \times 10^{-5}$  para ativos de liquidez intermediária e  $1,8 \times 10^{-5}$  para ativos menos líquidos. Em contraste, o estimador CS retornou valores iguais a zero para todos os grupos de ativos. É importante notar que, conforme o esperado, a média e a mediana do spread bid-ask são consistentemente mais altas para o grupo de menor liquidez (0,008973 e 0,007426, respectivamente) em comparação com os ativos mais líquidos (0,008079 e 0,006728) e os de liquidez intermediária (0,008099 e 0,006832), confirmando o padrão esperado de maior spread para ativos com menor liquidez. Esta tendência se mantém em todos os três estimadores analisados, sendo mais pronunciada no método AR, que apresentou os maiores valores médios entre os três métodos.

Tabela 1: Tabela indicando as principais estatísticas do método EDGE

	EDGE				
	Média	DesvPad	Mediana	Máximo	Mínimo
High Volume (Top 30%)	0.008079	0.006687	0.006728	0.113239	0.000016
Middle Volume (40%)	0.008099	0.006133	0.006832	0.081771	0.000024
Low Volume (Bottom 30%)	0.008973	0.007315	0.007426	0.153692	0.000018

Fonte: Elaborado pelo autor

Tabela 2: Tabela indicando as principais estatísticas do método CS

	CS				
	Média	DesvPad	Mediana	Máximo	Mínimo
High Volume (Top 30%)	0.003897	0.003284	0.003174	0.051537	0.000000
Middle Volume (40%)	0.004129	0.003447	0.003441	0.051915	0.000000
Low Volume (Bottom 30%)	0.004676	0.003926	0.003843	0.066553	0.000000

Fonte: Elaborado pelo autor

Tabela 3: Tabela indicando as principais estatísticas do método AR

	AR				
	Média	DesvPad	Mediana	Máximo	Mínimo
High Volume (Top 30%)	0.010710	0.007551	0.009283	0.108191	0.000000
Middle Volume (40%)	0.010854	0.007644	0.009434	0.090957	0.000000
Low Volume (Bottom 30%)	0.012665	0.009108	0.010901	0.147883	0.000000

Fonte: Elaborado pelo autor

A Tabela 4 confirma estatisticamente a significância dessas diferenças entre as médias dos grupos de liquidez. Os resultados do teste t mostram p-valores extremamente baixos ( $p < 0,05$ ) para todas as comparações entre ativos de baixa e alta liquidez nos três estimadores. As estatísticas t elevadas (23,08 para EDGE, 38,85 para CS e 42,18 para AR) indicam diferenças substanciais e estatisticamente significativas, com as maiores diferenças médias observadas no estimador AR (0,001955), seguido pelo EDGE (0,000894) e CS (0,000779). Esses valores demonstram que o AR é mais sensível às variações de liquidez, capturando de forma mais acentuada as diferenças nos custos de transação entre os grupos de ativos. Esses achados estão em consonância com o exposto por Ardia, Guidotti e Kroencke (2024), que investigaram o desempenho de diferentes estimadores de spread bid-ask.

Tabela 4: Teste t para Diferença de Médias entre Medidas de Liquidez, comparando spreads de grupos de alta e baixa liquidez

Comparação	Estatística t	Valor-p <sup>1</sup>	Diferença Média
Baixa Liquidez vs Alta Liquidez (EDGE)	23.082700	1.214373e-117	0.000894
Baixa Liquidez vs Alta Liquidez (CS)	38.854500	0e+00	0.000779
Baixa Liquidez vs Alta Liquidez (AR)	42.180300	0e+00	0.001955

<sup>1</sup>Valores-p em vermelho indicam diferenças estatisticamente significativas ( $p < 0.05$ ).

Fonte: Elaborado pelo autor

Segundo Ardia, Guidotti e Kroencke (2024), enquanto todos os estimadores avaliados são não viesados em cenários com alta frequência de negociações (por exemplo, 390 negociações por dia), suas performances divergem significativamente à medida que essa frequência diminui. Especificamente, o estimador CS apresenta um viés negativo acentuado em frequências menores, chegando a estimar spreads nulos em situações com menos de dez negociações diárias. Isso ocorre devido à forte dependência do CS na suposição de que os ativos são negociados continuamente, o que não é válido em mercados com baixa liquidez.

Por outro lado, o EDGE mantém estimativas não viesadas independentemente do número de negociações, graças a um termo de correção analítico que considera a negociação infrequente. Nossos resultados corroboram essa evidência, mostrando que o EDGE não retorna valores nulos mesmo em ativos com baixa frequência de negociação, ao contrário do CS. Dessa forma, o EDGE se apresenta como uma técnica mais robusta e apropriada para estimar o spread bid-ask em mercados com ativos menos líquidos.

Ao incluir o estimador de Abdi e Rinaldo (2017) (AR) na análise, conforme apresentado na tabela Tabela 3, observa-se que o AR também fornece resultados significativos na estimativa do spread bid-ask no mercado brasileiro. O estimador AR exibiu valores mínimos de spread bid-ask de 0,0 para todos os grupos de ativos analisados. Embora apresente valores mínimos iguais a zero, o AR tende a estimar spreads consistentemente maiores em comparação com o CS e o EDGE.



Os resultados mostram que o estimador AR possui médias de spread bid-ask de 0,010710 para ativos mais líquidos, 0,010854 para ativos de liquidez intermediária e 0,012665 para ativos menos líquidos, superiores às médias observadas para o EDGE (0,008079, 0,008099 e 0,008973, respectivamente) e significativamente maiores que as do CS (0,003897, 0,004129 e 0,004676). Esse comportamento sugere que o AR captura de maneira mais abrangente os custos de transação, especialmente em mercados com menor liquidez, onde a diferença entre os estimadores se torna ainda mais pronunciada.

Abdi e Ranaldo (2017) destacam que seu estimador é particularmente robusto em condições de baixa liquidez, pois utiliza preços de fechamento, máxima e mínima para estimar o spread bid-ask, sem dependência excessiva da frequência de negociação. Essa característica é essencial no contexto do mercado brasileiro, onde muitos ativos apresentam negociações infrequentes.

Comparando os três estimadores, nota-se que o CS apresenta os menores valores médios em todos os grupos de liquidez, frequentemente retornando valores nulos (mínimo igual a zero) em todos os níveis de volume negociado. O EDGE, ao introduzir termos de correção para a observação discreta de preços, mantém estimativas positivas mesmo em ambientes de baixa liquidez, nunca retornando valores zero. O AR, por sua vez, embora também apresente valores mínimos iguais a zero, tende a estimar spreads consistentemente mais elevados em todos os níveis de liquidez, possivelmente refletindo de forma mais acurada os custos totais enfrentados pelos investidores em mercados emergentes.

Esses achados enfatizam a importância de selecionar metodologias adequadas ao perfil de liquidez dos ativos analisados. Enquanto o CS pode subestimar os spreads em mercados menos líquidos, o EDGE e o AR demonstram vantagens claras ao capturar gradações de liquidez, como evidenciado pelo aumento progressivo nas médias e medianas conforme diminui o volume de negociação dos ativos no mercado brasileiro. O fato de o AR estimar spreads sistematicamente maiores pode fornecer insights adicionais sobre a presença de custos de transação elevados e assimetrias informacionais que não são completamente capturadas pelos outros estimadores.

## 5.2 Análise de período de volatilidade

Para compreender melhor como os diferentes estimadores de spread bid-ask se comportam em condições de mercado distintas, realizamos uma análise comparativa entre três períodos: pré-pandemia (2018-2019), pandemia (2020) e pós-pandemia (2021-2022). Esta segmentação temporal permite avaliar o desempenho dos estimadores em cenários de normalidade relativa, extrema volatilidade e recuperação, respectivamente.

Os dados na Tabela 5 apresentam os resultados do estimador CS nos três períodos analisados. Este método consistentemente mostra os valores médios de spread bid-ask mais baixos em comparação com os outros estimadores. No período pré-pandemia, o spread médio foi de 0,003848, com desvio padrão de 0,002896. Durante a pandemia, observa-se um aumento significativo para 0,005374 (incremento de 39,7%), com desvio padrão de 0,004790, evidenciando tanto um aumento no spread quanto uma maior dispersão nas estimativas. No período pós-pandemia, os valores apresentaram uma redução para 0,004466 (diminuição de 16,9% em relação ao período da pandemia), indicando uma recuperação parcial das condições de mercado, porém ainda 16,1% acima dos níveis pré-pandemia.

Tabela 5: Tabela indicando as principais estatísticas do método CS para pré-pandemia, ano da pandemia (2020) e pós-pandemia

	CS				
	Média	DesvPad	Mediana	Máximo	Mínimo
Pré-pandemia	0.003848	0.002896	0.003300	0.033160	0.000000
Durante pandemia	0.005374	0.004790	0.004274	0.051537	0.000000
Pós-pandemia	0.004466	0.003430	0.003798	0.035070	0.000000

Fonte: Elaborado pelo autor

Os resultados do teste t apresentados na Tabela 6 confirmam que estas variações são estatisticamente significativas, com p-valores muito baixos ( $p < 0,05$ ) para todas as comparações

entre períodos. A diferença média entre os períodos pré-pandemia e pandemia (0,001535) é maior que a diferença entre os períodos pós-pandemia e pré-pandemia (0,000632), corroborando a constatação de que, embora tenha havido recuperação após o período crítico, os spreads não retornaram completamente aos níveis anteriores à crise. O teste estatístico valida, portanto, a hipótese de que a pandemia causou um impacto substantivo e mensurável nos custos de transação do mercado brasileiro, conforme capturado pelo estimador CS.

Tabela 6: Tabela de teste estatístico T para comparação de médias entre os grupos de CS

Comparação	Estatística t	Valor-p <sup>1</sup>	Diferença Média
Pré-pandemia vs Pandemia	-9.299900	5.015175e-17	0.001535
Pandemia vs Pós-pandemia	5.448400	1.623748e-07	0.000903
Pré-pandemia vs Pós-pandemia	-5.635400	3.087069e-08	0.000632

<sup>1</sup>Valores-p em vermelho indicam diferenças estatisticamente significativas ( $p < 0.05$ ).

Fonte: Elaborado pelo autor

Já o estimador EDGE, na Tabela 7, apresenta valores intermediários entre os três métodos analisados, porém com maior sensibilidade às condições de mercado. No período pré-pandemia, o spread médio foi de 0,007435, com desvio padrão de 0,004471. Durante a pandemia, o EDGE registrou um aumento expressivo para 0,012245 (incremento de 64,7%), com desvio padrão mais que duplicado (0,010944), demonstrando sua maior capacidade de capturar a volatilidade e a redução de liquidez desse período. No pós-pandemia, o método estimou um spread médio de 0,008234, representando uma redução de 32,8% em relação ao período crítico, mas ainda 10,7% acima dos níveis pré-pandemia, sugerindo uma normalização incompleta das condições de mercado.

Tabela 7: Tabela indicando as principais estatísticas do método EDGE para pré-pandemia, ano da pandemia (2020) e pós-pandemia

	EDGE				
	Média	DesvPad	Mediana	Máximo	Mínimo
Pré-pandemia	0.007435	0.004471	0.006736	0.051456	0.000016
Durante pandemia	0.012245	0.010944	0.008913	0.113239	0.000018
Pós-pandemia	0.008234	0.004847	0.007415	0.072938	0.000071

Fonte: Elaborado pelo autor

A Tabela 8 corrobora estas observações, apresentando evidências estatísticas robustas da significância destas variações, com p-valores extremamente baixos em todas as comparações entre períodos. A estatística t para a comparação entre os períodos pré-pandemia e pandemia (-14,9366) é consideravelmente maior em valor absoluto do que a observada para o estimador CS (-9,2999). É notável também que, embora o teste t confirme que os valores do pós-pandemia permanecem significativamente diferentes dos níveis pré-pandemia ( $p = 0,00009601$ ), a estatística t é substancialmente menor (-3,9372), sugerindo que o mercado caminha para uma normalização, ainda que incompleta.

Tabela 8: Tabela de teste estatístico T para comparação de médias entre os grupos de EDGE

Comparação	Estatística t	Valor-p <sup>1</sup>	Diferença Média
Pré-pandemia vs Pandemia	-14.936600	1.976012e-33	0.004696
Pandemia vs Pós-pandemia	12.542300	4.148218e-26	0.003877
Pré-pandemia vs Pós-pandemia	-3.937200	9.600929e-05	0.000819

<sup>1</sup>Valores-p em vermelho indicam diferenças estatisticamente significativas ( $p < 0.05$ ).

Fonte: Elaborado pelo autor

O estimador AR, na Tabela 9, consistentemente produziu as maiores estimativas de spread bid-ask nos três períodos analisados. No período pré-pandemia, o spread médio foi de 0,009799, com desvio padrão de 0,005731. Durante a pandemia, o AR registrou o maior valor médio entre todos os métodos, atingindo 0,016082 (aumento de 64,1%), com desvio padrão de 0,013405, refletindo sua alta sensibilidade às condições adversas de mercado.

Tabela 9: Tabela indicando as principais estatísticas do método AR para pré-pandemia, ano da pandemia (2020) e pós-pandemia

	AR				
	Média	DesvPad	Mediana	Máximo	Mínimo
Pré-pandemia	0.009799	0.005731	0.009039	0.072874	0.000000
Durante pandemia	0.016082	0.013405	0.012338	0.112654	0.000000
Pós-pandemia	0.011833	0.006729	0.010759	0.074177	0.000046

Fonte: Elaborado pelo autor

No período pós-pandemia, o spread médio reduziu para 0,011833 (queda de 26,4%), permanecendo 20,8% acima dos níveis pré-pandemia, o que sugere uma percepção mais conservadora deste método quanto à recuperação das condições de liquidez do mercado brasileiro. A Tabela 10 valida estatisticamente estas variações, apresentando p-valores significativos em todas as comparações temporais realizadas. A estatística t para a comparação entre os períodos pré-pandemia e pandemia (-12,6516) demonstra uma sensibilidade similar à do EDGE (-14,9366), porém superior à do CS (-9,2999).

Tabela 10: Tabela de teste estatístico T para comparação de médias entre os grupos de AR

Comparação	Estatística t	Valor-p <sup>1</sup>	Diferença Média
Pré-pandemia vs Pandemia	-12.651600	8.107988e-26	0.006135
Pandemia vs Pós-pandemia	8.321200	2.888963e-14	0.004093
Pré-pandemia vs Pós-pandemia	-6.932300	1.429321e-11	0.002042

<sup>1</sup>Valores-p em vermelho indicam diferenças estatisticamente significativas ( $p < 0.05$ ).

Fonte: Elaborado pelo autor

A análise comparativa dos três estimadores ao longo dos diferentes períodos revela padrões consistentes e significativos sobre seu comportamento e sensibilidade às condições de mercado. Em primeiro lugar, observamos uma hierarquia clara nos valores estimados: o método CS sempre fornece as estimativas mais baixas, o AR as mais altas, e o EDGE ocupa uma posição intermediária. Esta ordenação se manteve inalterada nos três períodos analisados, sugerindo características estruturais inerentes a cada método.

Durante o período pré-pandemia, as diferenças entre os estimadores já eram evidentes, com o AR apresentando um spread médio 154,7% maior que o CS e o EDGE 93,2% maior. Estas diferenças substantivas indicam que, mesmo em condições normais de mercado, a escolha do estimador pode levar a conclusões significativamente distintas sobre os custos de transação no mercado brasileiro.

A pandemia de COVID-19 em 2020 proporcionou um experimento natural para avaliar a sensibilidade dos estimadores às condições extremas de mercado. Todos os métodos registraram aumentos expressivos em suas estimativas, refletindo o choque de liquidez e a alta volatilidade do período. Contudo, é notável que o EDGE e o AR demonstraram incrementos proporcionalmente maiores (64,7% e 64,1%, respectivamente) em comparação ao CS (39,7%). Esta diferença na magnitude da resposta, confirmada pelos valores t mais elevados nas comparações pré-pandemia vs pandemia (-14,9366 para EDGE e -12,6516 para AR contra -9,2999 para CS), sugere que o CS pode subestimar os custos de transação em períodos de estresse de mercado, enquanto o EDGE e o

AR capturam de forma mais completa o aumento dos spreads nessas condições.

Igualmente reveladora foi a análise dos desvios padrão, que apresentaram aumentos substanciais durante a pandemia: 65,4% para o CS, 144,8% para o EDGE e 133,9% para o AR. O aumento mais pronunciado na dispersão das estimativas do EDGE e do AR indica maior sensibilidade destes métodos à heterogeneidade das condições de mercado entre diferentes ativos e momentos, corroborando os resultados dos testes estatísticos que mostraram diferenças médias mais expressivas para estes estimadores.

No período pós-pandemia, todos os estimadores registraram reduções significativas em comparação ao auge da crise, sinalizando a gradual normalização das condições de mercado. No entanto, nenhum retornou completamente aos níveis pré-pandemia, sugerindo efeitos persistentes na liquidez e nos custos de transação. A magnitude da recuperação variou entre os métodos, com o EDGE mostrando a maior reversão (32,8% de redução), seguido pelo AR (26,4%) e pelo CS (16,9%).

Esta análise temporal reforça as conclusões anteriores sobre as características de cada estimador. O CS, com suas premissas de negociação contínua, parece menos sensível às variações nas condições de mercado, o que pode ser vantajoso para análises de longo prazo, mas problemático para capturar choques de liquidez. O EDGE, com seu termo de correção para negociações infrequentes, demonstra maior adaptabilidade a diferentes contextos, enquanto o AR consistentemente produz estimativas mais conservadoras dos custos de transação.

Os resultados sugerem que a escolha do estimador de spread bid-ask deve considerar não apenas o contexto geral do mercado, mas também os objetivos específicos da análise. Para avaliações em períodos de estresse ou em mercados com liquidez heterogênea, como o brasileiro, os estimadores EDGE e AR oferecem maior sensibilidade às variações nas condições de mercado, como demonstrado pelas diferenças médias mais expressivas e estatísticas t de maior magnitude. Em contrapartida, o CS pode ser adequado para análises comparativas de longo prazo em mercados mais estáveis, onde sua menor variância pode ser uma vantagem.

### 5.3 Análise comparativa com dados de referência da Bloomberg

Para avaliar a eficácia dos estimadores estudados em relação a uma referência de mercado amplamente reconhecida, realizou-se uma análise comparativa entre os spreads bid-ask estimados pelos métodos EDGE, CS e AR e os valores reportados pela Bloomberg. Esta comparação é fundamental para determinar qual metodologia oferece estimativas mais precisas no contexto específico do mercado brasileiro.

#### 5.3.1 Teste de similaridade estatística

O teste de Kolmogorov-Smirnov (KS) foi aplicado para verificar se as distribuições dos spreads estimados pelos três métodos são estatisticamente similares à distribuição dos valores de referência da Bloomberg. Este teste não paramétrico avalia a hipótese nula de que duas amostras provêm da mesma distribuição populacional.

Tabela 11: Tabela indicando o resultado do teste de Kolmogorov-Smirnov para cada metodologia de estimativa de spread bid-ask

Metodologia	Estatística D	p-valor	Resultado
Corwin-Schultz	0.982065	0.000000	Distribuições diferentes
Abdi-Ranaldo	0.927737	0.000000	Distribuições diferentes
EDGE	0.947794	0.000000	Distribuições diferentes

Fonte: Elaborado pelo autor

Conforme a Tabela 11, os resultados do teste KS revelaram p-valores iguais a zero e estatísticas D próximas a 1 para todas as três metodologias, indicando uma rejeição contundente da hipótese nula. Isso evidencia que as distribuições dos spreads estimados por CS, AR e EDGE são estatisticamente diferentes da distribuição dos spreads de referência da Bloomberg. Estas diferenças não são meramente aleatórias, mas refletem divergências sistemáticas nas estimativas. Este resultado sugere que os três métodos provavelmente capturam diferentes aspectos dos spreads ou possuem vieses estruturais distintos em relação aos valores de referência. A rejeição da hipótese



nula para todos os estimadores indica a complexidade inerente à mensuração do spread bid-ask em mercados emergentes, onde fatores específicos podem não ser adequadamente capturados pelos modelos desenvolvidos originalmente para mercados mais líquidos e eficientes.

### 5.3.2 Métricas de erro

Para quantificar a precisão das estimativas em relação aos valores de referência, foram calculadas duas métricas de erro: o Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE). Estas métricas permitem uma avaliação objetiva da magnitude das diferenças entre os valores estimados e os valores de referência. Os resultados das métricas de erro revelaram diferenças significativas entre as metodologias:

Tabela 12: Painel de métricas de erro RMSE e MAE por grupo de liquidez.

Grupo de Liquidez	RMSE	MAE
<b>Abdi-Ranaldo</b>		
High Volume (Top 30%)	0.770545	0.207251
Middle Volume (40%)	4.332984	0.606266
Low Volume (Bottom 30%)	1.111480	0.450487
<b>Corwin-Schultz</b>		
High Volume (Top 30%)	0.772264	0.213893
Middle Volume (40%)	4.333463	0.612899
Low Volume (Bottom 30%)	1.115113	0.458440
<b>EDGE</b>		
High Volume (Top 30%)	0.769317	0.209640
Middle Volume (40%)	1.012301	0.313644
Low Volume (Bottom 30%)	1.112719	0.453410

Fonte: Elaborado pelo autor

Tabela 13: Métricas de erro RMSE e MAE por período.

Período	RMSE	MAE
<b>Abdi-Ranaldo</b>		
Pré-pandemia	3.527887	0.580606
Pandemia	1.029691	0.284358
Pós-pandemia	0.600233	0.236721
<b>Corwin-Schultz</b>		
Pré-pandemia	3.528469	0.586528
Pandemia	1.032408	0.294629
Pós-pandemia	0.603152	0.243960
<b>EDGE</b>		
Pré-pandemia	0.549047	0.330544
Pandemia	0.971952	0.279724
Pós-pandemia	0.601579	0.240253

Fonte: Elaborado pelo autor

Como visto nas Tabela 12 e Tabela 13, o estimador EDGE consistentemente apresentou desempenho superior em ambos os painéis, com os menores valores de RMSE e MAE na maioria dos grupos de liquidez e períodos analisados. No Painel A (Tabela 12), observa-se que o EDGE mantém vantagem de precisão especialmente no grupo de liquidez média (“Middle Volume”), onde apresentou RMSE de 1,012301 em comparação com 4,332984 e 4,333463 dos estimadores AR e CS, representando uma redução de aproximadamente 76,7% no erro quadrático médio.

Embora nos grupos de alta e baixa liquidez (“High Volume” e “Low Volume”) os três estimadores tenham apresentado desempenho mais próximo, o EDGE ainda manteve ligeira vantagem. No grupo de maior liquidez, o EDGE registrou RMSE de 0,769317, comparado a 0,770545 do AR e 0,772264 do CS. Para ativos de menor liquidez, o EDGE demonstrou

desempenho intermediário com RMSE de 1,112719, ligeiramente superior ao AR (1,111480), mas inferior ao CS (1,115113).

No Pannel B (Tabela 13), o EDGE demonstrou ser significativamente mais preciso no período pré-pandemia, com RMSE de 0,549047, valor aproximadamente 84,4% menor que os estimadores AR e CS, que apresentaram erros de 3,527887 e 3,528469, respectivamente. Durante o período da pandemia, o EDGE mostrou menor vantagem, mas ainda registrou valores competitivos (RMSE de 0,971952). No período pós-pandemia, os três estimadores apresentaram desempenho mais homogêneo, com o EDGE mantendo-se ligeiramente mais preciso que o CS.

Os valores de MAE reforçam este padrão, com o EDGE apresentando erros absolutos médios geralmente menores ou comparáveis aos outros estimadores. Notavelmente, no segmento de liquidez média, o EDGE apresentou MAE de 0,313644, consideravelmente inferior aos valores de 0,606266 e 0,612899 do AR e CS, respectivamente.

Estes resultados indicam que, embora nenhuma das metodologias reproduza perfeitamente a distribuição dos spreads Bloomberg (conforme indicado pelo teste KS), o estimador EDGE consegue gerar aproximações pontuais mais precisas, especialmente em períodos de menor volatilidade (pré-pandemia) e para ativos de liquidez intermediária. Esta superioridade está alinhada com as observações feitas nas análises temporais, onde o EDGE demonstrou maior adaptabilidade a diferentes contextos de mercado, graças ao seu termo de correção para negociações infrequentes, característica particularmente relevante para o mercado brasileiro.

### **5.3.3 Interpretação integrada**

A análise integrada dos resultados do teste KS e das métricas de erro permite uma compreensão mais completa do desempenho relativo dos estimadores no contexto do mercado brasileiro. Apesar de todas as metodologias produzirem distribuições estatisticamente diferentes dos valores de referência, o EDGE demonstrou capacidade superior para aproximar-se dos valores individuais, o que é evidenciado por seu menor RMSE e MAE. Essa característica é particularmente relevante em aplicações práticas, onde a precisão pontual das estimativas é frequentemente mais importante que a reprodução exata da distribuição completa.

A superioridade do EDGE pode ser atribuída à sua abordagem mais abrangente, que

incorpora informações dos preços de abertura, máxima, mínima e fechamento, conforme descrito por Ardia, Guidotti e Kroencke (2024). Além disso, o EDGE inclui um termo de correção analítico para a negociação infrequente, característica comum em mercados menos líquidos como o brasileiro. Por outro lado, o desempenho similar dos estimadores CS e AR em termos de métricas de erro, apesar de suas diferentes formulações matemáticas, sugere que ambos podem estar sujeitos a limitações semelhantes quando aplicados ao mercado brasileiro. Isso pode estar relacionado a pressupostos subjacentes a esses modelos que não se aplicam plenamente ao contexto de mercados emergentes.

#### **5.3.4 Implicações para o mercado brasileiro**

Os resultados desta análise comparativa têm implicações significativas para a escolha de metodologias de estimativa do spread bid-ask no mercado brasileiro:

O estimador EDGE emerge como a opção mais adequada para aplicações práticas, oferecendo estimativas com erro médio absoluto de aproximadamente 0,33 pontos base em relação aos valores de referência da Bloomberg. O fato de todas as metodologias produzirem distribuições estatisticamente diferentes da Bloomberg sugere que fatores específicos do mercado brasileiro, como a concentração de liquidez em determinados ativos e períodos, podem não estar sendo adequadamente capturados pelos modelos existentes. Estes achados corroboram os achados de Ardia, Guidotti e Kroencke (2024) que sugere que o EDGE é mais robusto em cenários de baixa liquidez e mercados emergentes, como mencionado no referencial teórico.

Em suma, enquanto nenhuma das metodologias analisadas reproduz perfeitamente os valores de referência da Bloomberg, o estimador EDGE oferece uma alternativa mais precisa quando os dados de alta frequência não estão disponíveis. Esta conclusão é particularmente relevante para pesquisadores, investidores e reguladores interessados na microestrutura do mercado brasileiro e em seus custos de transação.

## 6 Conclusão

Este estudo teve como objetivo geral comparar, no mercado acionário brasileiro, estimativas de spread bid-ask obtidas pelas abordagens de Corwin & Schultz (2012), Abdi & Rinaldo (2017) e EDGE (Ardia, Guidotti & Kroencke, 2024), tendo como referência a cotação de spread divulgada pela Bloomberg. A pesquisa desenvolveu-se a partir do reconhecimento da importância fundamental do spread bid-ask como componente da microestrutura de mercado, representando a diferença entre os preços de compra (bid) e venda (ask) de um ativo, refletindo assim os custos de transação e a liquidez do mercado.

A amostra utilizada foi composta por 130 ações listadas na B3, selecionadas com base em critérios de liquidez, onde foram consideradas apenas ações com índice de liquidez superior a 0,1, calculado em uma janela de seis meses. As ações foram distribuídas entre diferentes segmentos de governança corporativa: 93 pertencentes ao Novo Mercado, 16 ao Nível 2, 17 ao Nível 1 e 4 ao segmento Tradicional.

Metodologicamente, o estudo aplicou os três estimadores selecionados aos dados de preços diários (abertura, máxima, mínima, fechamento) e volumes obtidos diretamente da B3, utilizando uma janela móvel de 21 dias para garantir consistência com estudos anteriores. As estimativas foram comparadas com valores de referência da Bloomberg através do teste de Kolmogorov-Smirnov e das métricas de erro RMSE e MAE.

Os resultados do teste de Kolmogorov-Smirnov indicaram que as distribuições dos spreads estimados por todos os três métodos são estatisticamente diferentes da distribuição dos valores da Bloomberg, o que sugere que nenhum dos estimadores reproduz perfeitamente a complexidade do spread bid-ask no mercado brasileiro. Esta constatação aponta para a existência de características específicas deste mercado emergente que podem não estar sendo adequadamente capturadas pelos modelos desenvolvidos originalmente para mercados mais líquidos e eficientes.

A pesquisa se baseou principalmente nos trabalhos seminais dos autores dos próprios estimadores: Corwin e Schultz (2012), que propuseram uma metodologia baseada em preços diários de máxima e mínima; Abdi e Rinaldo (2017), que expandiram essa abordagem incluindo preços de fechamento; e Ardia, Guidotti e Kroencke (2024), que desenvolveram o estimador

EDGE incorporando informações adicionais dos preços de abertura.

Os principais resultados demonstraram que, embora nenhum dos estimadores reproduza perfeitamente a distribuição dos spreads Bloomberg (conforme indicado pelo teste KS), o estimador EDGE apresentou desempenho significativamente superior, com RMSE de 0,97 e MAE de 0,33, valores aproximadamente 65% menores que os dos estimadores CS e AR. Durante o período de alta volatilidade de 2020, todos os estimadores registraram aumento nas médias dos spreads, com EDGE e AR demonstrando maior sensibilidade às condições adversas de mercado.

Esta pesquisa contribui para a literatura ao fornecer evidências sobre a eficácia relativa de diferentes metodologias de estimação do spread bid-ask especificamente no contexto de um mercado emergente como o brasileiro. Para estudos futuros, os resultados oferecem uma base para o desenvolvimento de estimadores mais adaptados às especificidades dos mercados emergentes, que tipicamente apresentam menor liquidez e maior volatilidade. A identificação do EDGE como estimador mais preciso também fornece orientação metodológica para pesquisadores interessados em mensurar custos de transação no mercado brasileiro.

Entre as limitações do estudo, destaca-se que os modelos foram originalmente calibrados e testados em mercados desenvolvidos, com estruturas distintas do brasileiro. Os pressupostos subjacentes a esses modelos, como a hipótese de negociação contínua, podem não se aplicar adequadamente ao contexto brasileiro, especialmente em ativos com menor liquidez. Adicionalmente, particularidades estruturais do mercado brasileiro, como a dinâmica da formação de preços na B3 e a concentração de liquidez em determinados horários, podem não ser completamente capturadas pelos modelos analisados.

## 7 Referências

- ABDI, F.; RANALDO, A. A simple estimation of bid-ask spreads from daily close, high, and low prices. **The review of financial studies**, v. 30, n. 12, p. 4437–4480, 2017.
- ARAÚJO, G. S.; BARBEDO, C. H. DA S.; VICENTE, J. V. M. The adverse selection cost component of the spread of Brazilian stocks. **Emerging markets review**, v. 21, p. 21–41, 2014.
- ARDIA, D.; GUIDOTTI, E.; KROENCKE, T. A. Efficient estimation of bid-ask spreads from open, high, low, and close prices. **SSRN Electronic Journal**, 2023.
- BURTON GORDON MALKIEL. **A random walk down Wall Street : the time- tested strategy for successful investing**. New York: W.W. Norton, 2007.
- CATTIVELLI, L.; PIRINO, D. A SHARP model of bid–ask spread forecasts. **International journal of forecasting**, v. 35, n. 4, p. 1211–1225, 2019.
- CHORDIA, T.; ROLL, R.; SUBRAHMANYAM, A. Commonality in liquidity. **Journal of financial economics**, v. 56, n. 1, p. 3–28, 2000.
- CONROY, R. M.; HARRIS, R. S.; BENET, B. A. The effects of stock splits on bid- ask spreads. **The journal of finance**, v. 45, n. 4, p. 1285–1295, 1990.
- CORREIA, L. F.; AMARAL, H. Determinantes da liquidez de mercado de ações negociadas. **Revista Brasileira de Economia**, v. 68, n. 1, p. 89-110, 2014.
- CORWIN, S. A.; SCHULTZ, P. A simple way to estimate bid-ask spreads from daily high and low prices. **The journal of finance**, v. 67, n. 2, p. 719–760, 2012.
- DE JONG, F.; RINDI, B. **The Microstructure of Financial Markets**. Cambridge: Cambridge University Press, 2009.
- DEMSETZ, H. The cost of transacting. **The Quarterly Journal of Economics**, v. 82, n. 1, p. 33, 1968.
- EASLEY, D.; O’HARA, M. Price, trade size, and information in securities markets. **Journal of financial economics**, v. 19, n. 1, p. 69–90, 1987.

- FAMA, E. F. The behavior of stock-market prices. **The journal of business**, v. 38, n. 1, p. 34, 1965.
- FAMA, E. F. Efficient capital markets: A review of theory and empirical work. **The journal of finance**, v. 25, n. 2, p. 383, 1970.
- FERREIRA, E. et al. Impacto assimétrico do sentimento do investidor na volatilidade do mercado acionário brasileiro. **Revista Brasileira de Gestão de Negócios**, v. 23, n. 2, p. 34-52, 2021.
- GEORGE, T. J.; KAUL, G.; NIMALENDRAN, M. Estimation of the Bid-ask Spreads and its Components: A New Approach. **Review of Financial Studies**, p. 623–656, 1991.
- GLOSTEN, L. R.; MILGROM, P. R. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. **Journal of financial economics**, v. 14, n. 1, p. 71–100, 1985.
- HARRIS, L. **Trading and Exchanges: Market Microstructure for Practitioners**. Oxford: Oxford University Press, 2003.
- HENDERSHOTT, T.; JONES, C. M.; MENKVELD, A. J. Does algorithmic trading improve liquidity? **The journal of finance**, v. 66, n. 1, p. 1–33, 2011.
- HOLDEN, C. W.; JACOBSEN, S. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions: Liquidity measurement problems in fast, competitive markets. **The journal of finance**, v. 69, n. 4, p. 1747–1785, 2014.
- HUANG, R. D.; STOLL, H. R. The components of the bid-ask spread: A general approach. **The review of financial studies**, v. 10, n. 4, p. 995–1034, 1997.
- ICHIMURA, D.; VIDEIRA, R.; RIPAMONTI, A. Asymmetric information and daily stock prices in Brazil. **Estudios gerenciales**, p. 465–472, 2020.
- JENSEN, M. C. Some anomalous evidence regarding market efficiency. **Journal of Financial Economics**, v. 6, n. 2-3, p. 95–101, jun. 1978.
- OLIVEIRA, R. **Como o fator de iliquidez afeta o retorno das ações brasileiras: Estudo Empírico**. São Paulo: Insper, 2018.
- PEAT, M. Market data resources for researchers: The SIRCA data repository. **The Australian**



**economic review**, v. 42, n. 4, p. 490–495, 2009.

PINTO, P. **Análise dos principais componentes do bid-ask spread de opções sobre ações no mercado brasileiro**. São Paulo: Insper, 2018.

RIPAMONTI, A. Corwin-Schultz bid-ask spread estimator in the Brazilian stock market. **Brazilian Administration Review**, v. 13, n. 1, p. 76–97, 2016.

ROLL, R. A simple implicit measure of the effective bid-ask spread in an efficient market. **The journal of finance**, v. 39, n. 4, p. 1127, 1984.

RYAN, Jeffrey; ULRICH, Joshua. Quantmod: Quantitative Financial Modelling Framework. Disponível em: <https://github.com/joshuaulrich/quantmod>. Acesso em: 24 nov. 2024.

STOIKOV, S. The micro-price: a high-frequency estimator of future prices. **Quantitative finance**, v. 18, n. 12, p. 1959–1966, 2018.

STOLL, H. R. Inferring the components of the bid-ask spread: Theory and empirical tests. **The journal of finance**, v. 44, n. 1, p. 115–134, 1989.

TOLEDO, R. **Determinantes da remuneração do spread de certificados de recebíveis do agronegócio no mercado brasileiro**. São Paulo: Fundação Getúlio Vargas, 2016.