Research-Thesis: Liquidity and the cross-section of stock returns in Germany and the effect of different market states.

Auteur : Boussaoud, Najwa
Promoteur(s) : Schwarz, Patrick
Faculté : HEC-Ecole de gestion de l'Université de Liège
Diplôme : Master en sciences de gestion, à finalité spécialisée en Banking and Asset Management
Année académique : 2025-2026
URI/URL : http://hdl.handle.net/2268.2/25799

Liquidity and the cross-section of stock returns in Germany and the effect of different market states

Supervisor: Patrick SCHWARZ

Reader: Philippe HÜBNER

Dissertation by: Najwa BOUSSAOUD

For a master's degree in management sciences with Specialization in Banking & Asset Management

Academic year:2025/2026

Acknowledgements:

I would like to express my sincere gratitude to everyone who contributed, directly or indirectly, to the completion of this thesis.

First and foremost, I would like to offer my deepest thanks to my family in Morocco, especially my father, my mother, and my siblings, Doha and Reda. This work is dedicated to you. Thank you for your unconditional love, emotional support, encouragement, and constant belief in me throughout this journey.

I would also like to sincerely thank my academic supervisor Patrick SCHWARZ for his valuable advice, guidance, and continuous support throughout this period. His encouragement and direction were greatly appreciated.

I would also like to thank the reader Philippe HÜBNER for taking the time to evaluate this work and for the insightful comments and questions that contributed to a thorough examination of the research.

Finally, I would like to thank everyone who, in one way or another, helped me move forward and complete this work. To all of you and to myself as well thank you from the bottom of my heart.

WORD COUNT : 22496

Declaration on the Use of AI: Artificial intelligence tools were used only for language support and reformulation purposes. The content, analysis, and conclusions of this report remain entirely my work.

Table of Content

Chapter 1: Introduction

1. Background and Motivation

The relationship between stock market liquidity and expected returns stands as one of the most consequential puzzles in modern finance. Liquidity, the ease with which investors can trade securities without materially affecting their prices, affects virtually every aspect of investment decision-making, from transaction cost estimation to portfolio construction and risk management. Understanding how liquidity influences asset pricing is therefore essential for both academics seeking to refine theoretical models and practitioners aiming to generate risk-adjusted returns. While a substantial body of research has examined this relationship in the United States, where the vast majority of foundational studies have been conducted, considerably less attention has been devoted to understanding liquidity dynamics in Continental European markets, and particularly in Germany Europe's largest economy and home to one of the world's most actively traded equity ecosystems.

Germany's equity market occupies a distinctive position in the global financial landscape. With a total market capitalization exceeding **€1.9** trillion as of late **2024**, Germany represents the largest stock market in the Eurozone and accounts for a substantial share of European equity trading activity. According to Deutsche Börse's annual statistics, trading volume across Xetra and the Frankfurt Stock Exchange reached **€1.3** trillion in **2024**, reflecting the depth and activity of the German market. The market's structure is anchored by a tiered index system: the DAX tracks the **40** largest and most liquid German blue-chip companies, followed by the MDAX comprising **50** medium-sized firms, the SDAX with **70** small-cap companies, and the technology-focused TecDAX index tracking **30** firms in the new economy sector. This hierarchical structure reflects a market characterized by substantial variation in liquidity characteristics across firm sizes a feature that makes Germany a particularly interesting laboratory for investigating how liquidity affects the cross-section of returns.

The trading infrastructure underlying the German equity market reinforces this liquidity variation. Xetra, the fully electronic trading platform operated by Deutsche Börse, handles approximately **90** percent of all stock trading in Germany and accounts for approximately **60** percent of trading volume for DAX-listed securities across European markets. With roughly **150** trading participants, including a significant proportion of international institutions, Xetra provides a highly standardized and transparent trading environment. Order book turnover on Xetra reached **€1.3** trillion in **2024**. The concentration of trading on a single electronic platform provides both advantages such as enhanced price discovery and transparent execution quality and challenges, particularly regarding the resilience of liquidity provision during periods of market stress.

Beyond its sheer scale and sophisticated infrastructure, the German equity market exhibits several institutional features that distinguish it from the U.S. and U.K. markets where most liquidity research has been conducted. Germany's financial system has historically been characterized by a bank-centered orientation, with commercial banks playing significant roles as both lenders and equity holders. Ownership structures in German firms tend to be more concentrated than in Anglo-Saxon markets, with founding families, financial institutions, and industrial blockholders retaining substantial stakes in many companies. This concentrated ownership pattern has implications for liquidity, as large blockholders may reduce the free-float available to public investors and influence trading dynamics. Furthermore, Germany's two-tier board system comprising the management board (Vorstand) and supervisory board (Aufsichtsrat) and the co-determination tradition whereby workers elect representatives to supervisory boards, shapes corporate governance in ways that may affect information asymmetry and investor participation.

The regulatory environment governing German equity markets has undergone significant transformation in recent years, with implications for liquidity dynamics that warrant scholarly attention. The implementation of the Markets in Financial Instruments Directive (MiFID II) in 2018 introduced sweeping changes to European securities regulations, including enhanced pre- and post-trade transparency requirements, stricter best execution obligations, and increased scrutiny of over-the-counter trading. These reforms aimed to improve market quality and investor protection but simultaneously altered the competitive landscape among trading venues and affected the provision of liquidity by market makers and high-frequency traders. Germany has also experienced substantial growth in algorithmic and high-frequency trading, which has influenced market microstructure and liquidity provision in ways that continue to evolve.

The practical significance of liquidity for investment management in Germany extends across multiple dimensions. For institutional investors who manage the assets of Germany's substantial pension fund and insurance industries, the largest fund market in the European Union liquidity directly affects transaction costs, as trading in less liquid securities incurs wider bid-ask spreads and greater price impact. During periods of market stress such as the 2008 global financial crisis, the 2011 eurozone sovereign debt crisis, or the March 2020 COVID-19 market collapse liquidity can deteriorate rapidly and unpredictably, creating risks that standard factor models may fail to capture. Empirical evidence from U.S. markets suggests that stocks with high sensitivity to market-wide liquidity shocks command a positive risk premium, implying that investors demand compensation for bearing liquidity risk. Whether this relationship holds in the German context, and whether it varies systematically across different market regimes, remains an open empirical question.

Despite the practical importance of understanding liquidity dynamics in Germany, the academic literature has devoted comparatively limited attention to the German equity market. Foundational studies by Amihud and Mendelson (1986) and Amihud (2002) established the illiquidity-return relationship using primarily U.S. data, and subsequent research has extended these findings to other markets and asset classes. For European markets specifically, several studies have examined liquidity effects across multiple countries, yet country-specific analysis of Germany remains relatively scarce. This gap is particularly concerning given that Germany's unique institutional features its concentrated ownership structures, bank-centered financial system, and distinctive regulatory environment may moderate or amplify the liquidity-return relationship in ways that generic European studies cannot capture. Moreover, with the increasing integration of German equities into global investment portfolios following the rise of passive investing, understanding liquidity pricing in Germany carries implications beyond its borders.

This thesis seeks to address these gaps by providing a comprehensive empirical analysis of the liquidity-return relationship in the German equity market. Drawing on panel regression methodology, the study investigates not only whether liquidity affects expected returns but also how this relationship varies across different market states a question of particular relevance given the frequency of market stress episodes in recent years. By focusing specifically on the German market, this research accounts for institutional and regulatory factors that may influence how liquidity is priced, thereby contributing to both the academic literature and practical investment decision-making.

2. Problem Statement and Research Questions

Despite the empirical regularity documented in U.S. and global markets that illiquidity commands a positive risk premium, the extent to which this relationship holds in the German equity market and the conditions under which it may vary remains insufficiently understood. The academic literature has established robust evidence that less liquid stocks earn higher average returns than more liquid stocks, a finding commonly interpreted as compensation for the higher transaction costs and greater price risk that illiquidity imposes on investors. However, the vast majority of this evidence originates from U.S. markets, with comparatively few studies examining the German context specifically. This gap is consequential because Germany's equity market possesses distinctive structural characteristics including concentrated ownership structures, a bank-centered financial system, and an electronic trading platform that dominates market activity that may influence how liquidity is priced relative to what existing theory and evidence suggest. Furthermore, the relationship between liquidity and returns may not be stable over time but could vary systematically across different market states, such as periods of normal trading versus market stress. Understanding whether and how liquidity affects asset pricing in Germany, and whether this relationship changes with market conditions, constitutes the central problem this thesis investigates.

The primary research questions guiding this investigation are as follows:

Research Question 1: Does liquidity have a significant effect on the cross-sectional variation of stock returns in the German equity market?

This question addresses whether stocks with different levels of liquidity earn systematically different expected returns after controlling for other known determinants of returns, such as market risk, size, value, and momentum. The hypothesis underlying this question is that consistent with theoretical predictions and international evidence, illiquid stocks in Germany should command a positive risk premium relative to liquid stocks. Testing this question requires measuring liquidity using appropriate indicators such as the Amihud illiquidity ratio, turnover-based measures, or bid-ask spreads and examining whether these measures explain cross-sectional return variation net of other risk factors. A positive and significant relationship would suggest that liquidity risk is priced in the German market; a non-significant relationship would raise questions about whether the liquidity-return nexus observed in other markets generalizes to Germany.

Research Question 2: Does the liquidity-return relationship in Germany vary systematically across different market states?

This question examines whether the pricing of liquidity risk changes depending on whether the market is in a normal state or a period of elevated stress. Different market states characterized by varying levels of volatility, trend direction, and overall market liquidity may alter the risk that investors face when holding illiquid securities and thus affect the premium they demand. For instance, during periods of market crisis, liquidity may deteriorate abruptly across the market, increasing the risk exposure of investors in illiquid stocks and potentially widening the liquidity premium. Conversely, during strong bull markets, the demand for illiquid stocks may increase as investors chase returns, potentially compressing the liquidity premium. This question is central to the thesis as it investigates the conditional nature of liquidity pricing and tests whether the relationship between liquidity and returns is state-dependent rather than stable over time.

Research Question 3: What is the economic magnitude of any liquidity premium in the German market, and does it differ meaningfully across stock size categories?

This question extends the primary inquiry to examine not only whether a liquidity effect exists but also how large it is in economic terms and whether it varies across the size spectrum. Large-cap stocks in the DAX index typically exhibit substantially higher liquidity than small-cap stocks in the SDAX or in the broader universe of German equities. If the liquidity premium is concentrated among smaller, less liquid stocks, this would have important implications for portfolio management and asset pricing theory. This question also addresses heterogeneity by examining whether the liquidity effect documented in the cross-section is driven primarily by stocks with the lowest liquidity or whether it represents a more broadly distributed phenomenon.

In addition to these primary questions, several secondary inquiries support the investigation. First, which specific liquidity measure provides the strongest explanatory power for cross-sectional returns in Germany price impact measures such as the Amihud ratio, turnover-based measures, or measures of trading frequency? Second, is the liquidity-return relationship robust to alternative model specifications, control variable sets, and estimation methodologies?

By addressing these research questions, this thesis aims to provide a comprehensive understanding of how liquidity affects stock returns in the German equity market and whether this relationship is conditioned by the prevailing market environment. The findings will contribute to the academic literature on asset pricing and liquidity, offer practical insights for portfolio management in Germany and Europe, and inform regulatory considerations regarding market structure and investor protection.

3. Research Objectives and Hypotheses

The overarching objective of this thesis is to empirically investigate the relationship between stock market liquidity and expected returns in the German equity market, with particular attention to how this relationship varies across different market states. This objective is grounded in the premise that liquidity constitutes a risk factor that investors should demand compensation for bearing, and that the magnitude of this compensation may depend on the prevailing market environment.

To structure the investigation systematically, the research objective is decomposed into three testable hypotheses derived from theoretical predictions and prior empirical evidence.

Hypothesis 1(H1): Illiquid stocks earn higher expected returns than liquid stocks in the German equity market.

The theory of liquidity-based asset pricing posits that investors demand a premium for holding securities that are costly to trade. When investors hold illiquid stocks, they face higher transaction costs (wider bid-ask spreads, greater price impact) and bear the risk that they may be unable to exit positions quickly during periods of market stress. According to the asset pricing framework developed by Amihud and Mendelson and extended by subsequent authors, this illiquidity risk should be priced, meaning that investors require higher expected returns as compensation. The intuition is straightforward: if two stocks offer identical expected cash flows but one is more liquid than the other, rational investors will prefer the liquid stock unless the illiquid stock offers a higher return to compensate for the additional friction. Empirically, this relationship has been documented across multiple markets, including the United State, emerging markets, and various European markets, although country-specific evidence for Germany remains limited.

Hypothesis 2(H2): The liquidity-return relationship in Germany varies across different market states.

Asset pricing theory and market microstructure theory jointly predict that the relationship between liquidity and expected returns should not be constant but should vary with the prevailing market environment. During periods of market stress characterized by high volatility, declining prices, and deteriorating liquidity conditions the risk of holding illiquid stocks intensifies. Investors may face sudden increases in transaction costs, inability to execute trades at desired prices, or forced selling at distressed prices. This amplification of liquidity risk during downturns should increase the premium investors’ demand for holding illiquid stocks, leading to a stronger positive relationship between illiquidity and expected returns during bear markets or high-volatility periods. Conversely, during bull markets or periods of low volatility, when liquidity conditions are generally favorable and investors are more willing to accept risk, the liquidity premium may compress. The theoretical basis for this state-dependence lies in the insight that liquidity risk is not independent of market risk but rather is covariant with-it during market crises, liquidity tends to dry up simultaneously across many securities, amplifying the risk exposure of investors holding illiquid positions.

The interaction term between the liquidity measure and the market state indicator should be positive and significant. Specifically, the liquidity premium should be larger during periods of market stress (bear markets, high volatility, crisis periods) compared to periods of normal market conditions. This would manifest as a higher coefficient on the liquidity variable in state-conditioned regressions or as a positive and significant coefficient on the liquidity-state interaction term in a unified regression framework.

Hypothesis 3(H3): The liquidity premium in the German market varies across firm size categories, with a stronger effect for smaller, less liquid stocks.

Market microstructure theory and empirical evidence from multiple markets suggest that the liquidity premium is not uniform across the size spectrum but is concentrated among smaller, less liquid stocks. Small-cap stocks typically exhibit lower trading volumes, wider bid-ask spreads, and higher price impact than large-cap stocks. Moreover, small-cap stocks are more likely to be held by less sophisticated investors with limited market access, less likely to attract institutional coverage, and more susceptible to liquidity shocks during market stress. The theoretical argument is that investors who are able to invest in large, liquid stocks have little reason to accept the additional costs and risks associated with small, illiquid stocks unless they are compensated with a sufficiently high premium. This asymmetry suggests that the illiquidity premium should be more pronounced among small-cap stocks than among large-cap stocks, where liquidity is already embedded in the price through intense investor participation and market maker activity.

Sub-sample regressions or size-interaction specifications should reveal a larger liquidity premium for stocks in the SDAX and smaller size categories compared to stocks in the DAX and MDAX. The coefficient on the liquidity measure should be more positive and statistically significant in regressions estimated on small-cap stocks than in regressions estimated on large-cap stocks. Alternatively, the interaction between liquidity and size should be negative and significant, indicating that the liquidity effect decreases as firm size increases.

4. Contribution and Significance of the Study

This thesis makes several distinct contributions to the academic literature on asset pricing, market microstructure, and empirical finance, while also offering practical implications for investment professionals and market participants operating in the German equity market. The contributions are grounded in identified gaps in the existing literature and are formulated as specific, defensible claims rather than broad assertions.

Contribution 1: Country-specific evidence on the liquidity-return relationship in Germany.

The existing literature on liquidity and asset pricing is dominated by studies examining U.S. equity markets, with comparatively limited attention devoted to the German context. While several studies have examined European markets broadly often pooling observations across multiple countries to increase statistical power country-specific analysis of Germany remains relatively scarce. This thesis provides direct evidence on whether the illiquidity premium documented in U.S. and global studies also operates in the German market, thereby extending the geographical scope of the liquidity-asset pricing literature and enabling comparison with findings from other institutional contexts.

Contribution 2: Systematic investigation of market state dependence in liquidity pricing.

A distinguishing feature of this thesis is its explicit examination of how the liquidity-return relationship varies across different market states. While the theoretical prediction that liquidity risk is more severe during market stress is well established, empirical evidence on state-dependent liquidity pricing remains limited, particularly for European markets. Most prior studies estimate a time-invariant relationship between liquidity and returns, implicitly assuming that the liquidity premium is constant regardless of market conditions. This assumption is difficult to justify theoretically, as the costs and risks of holding illiquid securities should be more acute during periods of elevated volatility and declining markets. This thesis directly tests whether the liquidity premium is larger during bear markets, high-volatility periods, and financial crises, contributing to the conditional asset pricing literature and providing a more nuanced understanding of liquidity as a risk factor.

Contribution 3: Heterogeneity analysis across the size spectrum.

This thesis examines whether the liquidity premium varies systematically across firm size categories, addressing a question with both academic and practical relevance. Large-cap stocks in the DAX and MDAX indexes benefit from extensive investor participation, active market making, and deep order books, whereas small-cap stocks in the SDAX and in the broader universe of German equities often exhibit substantially lower liquidity. If the liquidity premium is concentrated among smaller, less liquid stocks, this would have important implications for the asset pricing literature suggesting that size and liquidity are not independent factors and for practitioners constructing portfolios that span the size spectrum. By documenting the size-dependence of the liquidity effect in Germany, this thesis contributes to the literature on return determinants across the size distribution and provides insights relevant to portfolio construction and performance evaluation.

Contribution 4: Practical implications for portfolio management and risk assessment.

Beyond its academic contributions, this thesis offers insights of value to investment professionals and market participants. For portfolio managers, the findings provide evidence on whether liquidity should be incorporated as a priced risk factor in return expectations and whether the liquidity premium is large enough to warrant intentional tilts toward less liquid stocks. For risk managers, the results illuminate how liquidity risk interacts with market risk during different market regimes, enabling more informed assessment of tail risk and portfolio resilience under stress conditions. For regulators and market infrastructure providers, the evidence on liquidity dynamics in Germany contributes to understanding how market structure affects price discovery and investor welfare.

In summary, this thesis contributes to the literature by providing rigorous, country-specific evidence on liquidity pricing in the German equity market, with a novel focus on market state dependence and size heterogeneity. The findings extend the geographical scope of the liquidity-asset pricing literature, test conditional predictions from asset pricing theory, and offer practical insights for investment practice. By addressing gaps in the existing literature while maintaining methodological rigor, this study aims to advance both academic understanding and professional practice regarding the role of liquidity in asset pricing.

Chapter 2: Literature Review and Theoretical Framework

1. Overview of Liquidity Concepts

Before examining the relationship between liquidity and stock returns, it is essential to establish a clear conceptual foundation for what liquidity means in financial markets. Liquidity is a multifaceted concept that defies simple definition, yet its understanding is critical for both theoretical development and empirical investigation. This section defines market liquidity, describes its key dimensions, distinguishes it from related but distinct concepts such as solvency and funding liquidity, and introduces the common metrics used to measure liquidity in empirical research. This conceptual framework provides the vocabulary and analytical tools necessary for the subsequent chapters of this thesis.

Defining Market Liquidity

Market liquidity refers to the ability to trade an asset quickly, at low cost, and with limited price impact. More broadly, it captures the ease with which investors can convert a position into cash without incurring substantial losses. In the financial literature, liquidity is generally viewed as a multidimensional concept. Kyle (1985) provides one of the foundational frameworks by distinguishing three key dimensions of liquidity: tightness, depth, and resiliency. Tightness refers to the transaction cost of reversing a position, commonly proxied by the bid-ask spread. Depth describes the market’s ability to absorb large orders without causing substantial price changes. Resiliency refers to the speed with which prices recover after order imbalances or liquidity shocks. Later studies expanded and refined this framework by emphasizing additional aspects of trading frictions and liquidity pricing. Amihud and Mendelson (1986) highlight the role of transaction costs, especially the bid-ask spread, in asset pricing, while Pástor and Stambaugh (2003) examine market-wide liquidity as a state variable relevant for expected stock returns. A widely used empirical proxy for illiquidity is proposed by Amihud (2002), who defines illiquidity as the average ratio of the daily absolute stock return to daily dollar trading volume. This measure can be interpreted as the daily price response associated with one dollar of trading volume and therefore serves as a practical proxy for price impact.

The Four Dimensions of Liquidity.

Market liquidity can be decomposed into four complementary dimensions that together characterize the liquidity quality of an asset or market.

Tightness represents the transaction cost incurred by investors to execute an order. This cost manifests primarily in the form of the bid-ask spread. A tight market is characterized by narrow spreads, indicating low transaction costs. Tightness may also include brokerage commissions and other trading fees, but the bid-ask spread typically constitutes the most significant component.

Depth refers to the market's capacity to absorb significant orders without causing a noticeable price movement. A deep market is one where there is sufficient volume of buy and sell orders at various price levels to absorb large transactions. Depth is typically measured by the volume of orders available in the order at different price levels. A deep market offers investors the assurance that they can execute substantial orders without significantly moving the price against them

Immediacy refers to the speed with which a transaction can be executed. A market offering high immediacy allows investors to buy or sell nearly instantaneously at a given price. Conversely, a market with low immediacy may require considerable delay to execute a transaction at the desired price or may even render certain transactions impossible. Immediacy is particularly important for investors who must liquidate their positions quickly, for example in the event of unexpected liquidity needs or in response to adverse news.

Resilience measures the market's capacity to return to its fundamental price after a disturbance. A resilient market is one where temporarily deviated prices are quickly corrected by the arrival of new orders. Resilience is particularly relevant following significant market events: a resilient market will quickly return to equilibrium, while a less resilient market may remain imbalanced longer, leaving investors exposed to prices that do not reflect fundamentals. Resilience is thus particularly relevant for understanding liquidity behavior during periods of stress, when prices can deviate significantly from fundamental values.

Important Conceptual Distinctions

Liquidity vs.Solvency

It is important to distinguish **market liquidity** from **solvency**, as the two concepts refer to different aspects of financial performance. Solvency concerns the issuer's long-term financial condition and its ability to meet its obligations over time. By contrast, market liquidity refers to the ease with which a security can be traded quickly, at low cost, and with limited price impact. A firm may be fundamentally solvent while its shares remain difficult to trade, especially during periods of market stress. Conversely, a stock may be actively traded even when the underlying firm faces weak financial fundamentals. For equity investors, the distinction is crucial: solvency primarily affects the firm's fundamental value, whereas market liquidity affects the cost and speed of entering or exiting a position.

Market Liquidity vs.Funding Liquidity

A second important distinction opposes market liquidity to funding liquidity. Market liquidity, described above, concerns the ease with which assets can be traded on financial markets. Funding liquidity, on the other hand, refers to the ability of financial institutions and investors to obtain funds to finance their activities. These two forms of liquidity are interconnected but distinct: a deterioration in funding liquidity can trigger forced sales that affect market liquidity, and vice versa, a deterioration in market liquidity can reduce the value of collateral used to obtain financing. The 2008 financial crisis dramatically illustrated this interconnection, when the contraction in funding liquidity triggered massive sales that deteriorated market liquidity across the financial system.

Common Liquidity Measures

Empirical literature has developed several measures to operate liquidity. These measures capture different dimensions of liquidity, and each carries its own advantages and limitations.

Bid-ask spread is the most direct measure of tightness. It represents the difference between the highest offered price for an asset and the lowest asked price. A high spread indicates high transaction costs and therefore low liquidity. This measure is particularly suited to markets with a centralized order book, such as Xetra in Germany.

Turnover ratio measures trading volume relative to market capitalization or shares outstanding. A high turnover ratio indicates high liquidity, as it suggests shares change hands frequently. This measure is widely used because it is easy to calculate and available for most markets.

Amihud illiquidity ratio measures the price impact of trading. It is calculated as the ratio of the absolute value of daily return to daily trading volume. This ratio captures the extent to which trading volume affects an asset's price, reflecting the depth and resilience of the market. A high ratio indicates severe illiquidity: a given trading volume produces a large price change.

Zero return measures the proportion of days on which a stock generates no return. A high frequency of zero returns is interpreted as an indicator of low liquidity, as it suggests the stock does not trade regularly. This measure, popularized by Lesmond, Ogden, and Trzcinka (1999), is particularly useful for markets where volume data is incomplete.

2. Theories of Liquidity and Asset Pricing

Understanding why liquidity should affect expected returns requires a thorough examination of the theoretical frameworks that connect trading frictions to asset pricing. This section reviews the major theories that explain the liquidity-return relationship, beginning with the standard capital asset pricing model as a baseline, examining why standard models fail to fully account for liquidity, and then exploring the theoretical mechanisms through which liquidity affects pricing. The objective is to establish a coherent theoretical foundation that generates testable predictions about the direction and magnitude of the liquidity effect.

The Standard CAPM Baseline

The Capital Asset Pricing Model (CAPM), developed by Sharpe (1964), Lintner (1965), and Mossin (1966), provides the foundational framework for asset pricing in modern finance. In its standard form, the CAPM posits that the expected return of an asset is determined exclusively by its systematic risk, measured by beta the covariance between the asset's returns and the market portfolio. Under the assumptions of frictionless markets, homogeneous expectations, and unlimited borrowing and lending at the risk-free rate, the CAPM implies that:

$$E(R_i) = R_f + \beta_i \times (E(R_m) - R_f)$$

where $E(R_i)$ is the expected return on asset i, R_f is the risk-free rate, β_i is the asset's beta, and $E(R_m) - R_f$ is the market risk premium. In a frictionless world, liquidity is irrelevant to pricing because investors can buy and sell any asset instantaneously and without cost. The CAPM's prediction is that assets with identical betas should have identical expected returns regardless of their trading characteristics.

The empirical observation that the CAPM explains only a fraction of cross-sectional return variation gave rise to multi-factor models such as the Fama-French three-factor model (Fama and French, 1993), which adds size and value factors to the market factor. However, these extensions still assume frictionless markets and do not explicitly incorporate liquidity as a pricing factor. The gap between the frictionless-market assumption and the reality of trading costs motivates the theoretical explanations for the liquidity premium reviewed below.

Market Microstructure Theories

Market microstructure theory provides an important theoretical foundation for understanding the liquidity-return relationship by examining how trading occurs in organized markets and how prices are determined in the presence of trading frictions.

Adverse Selection Theory

The adverse selection model, pioneered by Glosten and Milgrom (1985) and Bagehot (1971), explains bid-ask spreads as compensation for the risk that market makers trade with informed investors. In this framework, market makers set bid and ask prices that allow them to profit from trading with uninformed investors while covering losses to informed traders. The spread thus reflects the probability that a counterparty is informed and the magnitude of the adverse price movement that trading with such a counterparty would cause. From an asset pricing perspective, adverse selection implies that stocks with higher information asymmetry should have wider spreads and higher expected returns, as investors demand compensation for the probability of trading against better-informed counterparties. This theory predicts that the illiquidity premium should be larger for stocks with higher information asymmetry for example, smaller firms, firms with more complex businesses, or firms with less analyst coverage.

Order Processing Cost Theory

Another strand of market microstructure theory emphasizes the role of **order-processing costs** in explaining bid-ask spreads. In the market microstructure literature, the spread is commonly viewed as reflecting several components, including **order-processing costs**, **inventory-holding costs**, and **adverse-selection costs**. Order-processing costs correspond to the direct costs of executing trades, such as exchange fees, administrative expenses, and the operational costs of maintaining trading infrastructure. Dealers or market makers must recover these costs through the spread they charge. Although this component is more mechanical than information-based, it still contributes to illiquidity by making trading more expensive. From an asset-pricing perspective, such direct trading costs can reduce the price investors are willing to pay for a security and thereby increase its expected return in equilibrium, even when the source of illiquidity is not primarily risk-based.

Asymmetric Information Models

Building on the adverse-selection framework, asymmetric-information models examine how differences in information across market participants affect both liquidity and asset prices. A seminal contribution is Kyle (1985), who models trading by an informed trader in the presence of noise traders and a competitive market maker. In this framework, the informed trader optimally balances the use of private information against the price impact of his trades. The equilibrium is characterized not primarily by a bid-ask spread, but by a **price-impact coefficient** commonly known as **Kyle's lambda** which measures how strongly prices respond to net order flow. The model therefore links asymmetric information to illiquidity through market impact rather than through the spread alone.

Later work extends this logic to asset pricing more directly. In particular, Easley and O'Hara (2004) argue that information risk can affect the cost of capital, because investors require compensation for trading in markets where some participants are better informed than others. Related empirical work by Easley et al. (2002) uses the probability of informed trading (PIN) as a proxy for information-based trading and shows that this measure is associated with both illiquidity and stock returns. Together, these contributions suggest that information asymmetry can influence expected returns not only through transaction costs, but also through a risk-based mechanism linked to the informational environment of the asset.

Inventory-Based Models

Inventory-based models, originally developed by Demsetz (1968) and formalized by Stoll (1978) and Ho and Stoll (1981), explain liquidity through the costs and risks of maintaining inventory positions. Market makers hold inventories of securities to facilitate trading, buying when others want to sell and selling when others want to buy. This inventory exposes market makers to price risk the possibility that prices move against their positions before they can unwind them. To compensate for this risk, market makers require a return that is built into the bid-ask spread. The magnitude of this inventory component depends on the volatility of the asset's price and the market maker's risk tolerance.

In equilibrium, assets with higher price volatility should have larger spreads (higher inventory costs) and thus higher expected returns as investors bear the market maker's inventory risk through lower prices. This prediction connects liquidity to systematic risk factors: during periods of high volatility, inventory costs increase, spreads widen, and the compensation for bearing inventory risk rises.

Search-Based Models

Search-based models offer an alternative theoretical foundation that emphasizes the matching process between buyers and sellers rather than the role of market makers. Pioneered by Duffie, Gârleanu, and Pedersen (2005, 2007), these models recognize that some investors cannot immediately find counterparties for their trades and must engage in costly search. The cost of search in time, effort, or forgone transactions constitutes a trading friction that should affect asset prices.

In search-based equilibrium, assets that are harder to search for (less frequently traded, held by investors with less diverse preferences) command lower prices and thus higher expected returns. The search cost is borne by the investor who initiates the trade, who must accept a less favorable price to find a counterparty. This framework generates predictions about how market structure affects liquidity: markets with more diverse investor bases and more active trading should have lower search costs and thus lower liquidity premiums.

Investor Recognition and Limited Participation

An additional explanation for liquidity-related return differences emphasizes **incomplete information and limited market participation** rather than pure trading-cost mechanisms. In Merton's (1987) investor recognition model, investors hold only securities of which they are aware, implying that firms followed by a narrower set of investors face a smaller investor base and therefore weaker demand for their shares. In this setting, lower investor recognition can contribute to lower stock prices and a higher expected return. This mechanism is particularly relevant for illiquid stocks, which often receive less analyst coverage, less media attention, and less broad investor participation. Rather than representing a behavioral bias in the narrow sense, this argument is more accurately viewed as a market-equilibrium effect arising from incomplete information and segmented investor attention.

A related argument concerns **limited participation and institutional constraints**. Some investors, especially large institutions, may be unwilling or unable to hold very illiquid securities because of internal mandates, regulatory considerations, benchmark constraints, or liquidity-management requirements. When the eligible investor base for illiquid stocks is restricted, demand for these securities may be lower, which can depress prices and increase expected returns in equilibrium. This mechanism is therefore better understood as a participation-based or segmentation-based explanation than as a purely behavioral one

3. Empirical Evidence on Liquidity and Returns

The theoretical frameworks reviewed in the preceding section provide a strong motivation for why liquidity should affect expected returns, but the question of whether this prediction holds empirically has been examined in a large body of research. This section reviews the main empirical findings on the liquidity-return relationship, focusing first on the United States before turning to evidence from other markets. The objective is to identify the most robust stylized facts in the literature and to clarify the empirical gap to which this thesis contributes.

Foundational Evidence from the United States

Early empirical evidence on the pricing of liquidity was provided by Amihud and Mendelson (1986), who document a positive relationship between bid-ask spreads and stock returns for NYSE stocks. Their results are consistent with the view that investors require compensation for the trading costs associated with less liquid securities, and their study remains one of the foundational empirical contributions to the liquidity-asset pricing literature.

Amihud (2002) extends this literature by introducing an illiquidity measure based on the ratio of the absolute daily return to daily dollar trading volume. Using NYSE stocks over the period 1964–1997, he finds that this illiquidity measure has a positive and highly significant effect on expected returns in the cross section. The paper also shows that illiquidity effects are stronger for smaller firms, thereby reinforcing the idea that market frictions are priced in expected returns.

A related but distinct contribution is provided by Pástor and Stambaugh (2003), who focus on systematic liquidity risk rather than the level of stock illiquidity. They construct a market-wide liquidity measure and show that stocks with greater sensitivity to aggregate liquidity fluctuations earn higher expected returns. Their results imply that liquidity is not only a stock characteristic but also a priced state variable. In particular, they report that a spread between the top and bottom deciles of predicted liquidity betas generates an abnormal return of about 7.5 percent per year relative to a standard factor model.

Lesmond, Ogden, and Trzcinka (1999) provide a complementary contribution by developing a structural method to estimate effective transaction costs from daily returns using the incidence of zero-return days. Rather than treating zero returns as a simple reduced-form proxy, they use them to infer transaction-cost thresholds within a limited dependent variable framework. Their results show that effective transaction costs vary substantially across firms and are especially high for smaller stocks, thereby providing additional evidence that trading frictions are economically important.

International and Cross-Country Evidence

The early U.S. evidence on the liquidity-return relationship motivated a broad international literature examining whether liquidity premia also arise in markets with different institutional structures, regulatory environments, and investor bases. Overall, the international evidence suggests that the pricing of liquidity is not confined to the United States, although the strength and form of the relationship vary across countries and according to the liquidity measure employed.

In developed markets outside the United States, empirical studies generally provide support for a positive association between illiquidity and expected returns, although the evidence is less extensive and less uniform than in the U.S. case. European evidence, in particular, suggests that liquidity effects are present, but their magnitude appears to differ across countries and market settings. This heterogeneity is consistent with the view that market structure, trading mechanisms, and investor composition influence the extent to which liquidity is priced.

The evidence is especially strong in **emerging markets**. A key contribution is **Bekaert, Harvey, and Lundblad (2007)**, who show that local market liquidity is an important driver of expected returns in emerging equity markets. Their findings suggest that the illiquidity premium is economically meaningful in these markets, which is consistent with the idea that weaker market development, lower participation, and larger trading frictions amplify liquidity effects.

More broadly, cross-country research indicates that liquidity dynamics differ substantially across markets and that institutional conditions matter for how liquidity is generated and transmitted. For this reason, it cannot simply be assumed that the magnitude or structure of the liquidity premium observed in the United States will carry over unchanged to continental European markets such as Germany. This provides a clear motivation for studying the German equity market directly rather than relying on evidence from other countries.

4. Market States and Conditional Asset Pricing

The standard asset pricing models reviewed in previous sections typically assume that the relationship between risk factors and expected returns is stable over time. However, a substantial body of theoretical and empirical work suggests that this assumption is overly restrictive. Asset pricing relationships may vary across different market states periods characterized by different levels of volatility, return trends, and overall market liquidity reflecting changes in investor behavior, risk perception, and the economic environment. This section reviews the literature on conditional asset pricing and market state effects, with particular attention to how liquidity's role may vary across different market regimes.

Conditional Asset Pricing Theory

The theory of conditional asset pricing recognizes that the risk-return relationship is not constant but varies over time as the investment opportunity set changes. In the conditional CAPM framework, expected returns and betas are time-varying, depending on the state of the economy and the characteristics of the market portfolio . This time-variation arises because the distribution of returns is not stationary moments such as the mean, variance, and covariance between assets change as economic conditions evolve.

Applied to liquidity, conditional asset pricing theory predicts that the premium investors demand for holding illiquid stocks should vary with the state of the market. During periods of elevated volatility or declining liquidity, the risks associated with illiquid holdings intensify, and rational investors should demand higher compensation. During bull markets or periods of low volatility, when liquidity conditions are generally favorable and investors are more willing to accept risk, the liquidity premium may compress. This conditional prediction lies at the heart of this thesis and motivates the investigation of market state effects.

The theoretical basis for state-dependent liquidity pricing rests on several mechanisms. First, during market downturns, correlations between individual stock returns tend to increase, meaning that illiquid stocks become more likely to decline simultaneously with the broader market, amplifying the risk exposure of investors holding such stocks. Second, during periods of market stress, the costs of trading illiquid securities tend to increase sharply bid-ask spreads widen, market depth diminishes, and price impact intensifies reducing the ability of investors to adjust their positions. Third, liquidity risk has the characteristics of a systematic risk: during crises, liquidity tends to deteriorate across the market, affecting all investors holding illiquid positions simultaneously. This systematic dimension of liquidity risk suggests that it should command a premium that varies with the prevailing market state.

Regime-Switching Models and Evidence on Time-Varying Premiums

The regime-switching framework, originally developed by Hamilton (1989), provides a useful way to model financial markets as alternating between distinct states with different statistical properties. In asset pricing, this approach has been widely used to capture the idea that expected returns, risk exposures, and market conditions may vary across market environments rather than remain constant over time. In this sense, regime-dependent models offer a natural extension to conditional asset-pricing theory, in which the relation between risk and return depends on the prevailing economic and financial state.

For liquidity, the literature suggests that the premium associated with illiquidity is state-dependent rather than constant. Evidence indicates that liquidity conditions worsen in market downturns and periods of stress, when spreads widen, trading becomes more costly, and the ability to sell quickly becomes especially valuable. Chordia, Sarkar, and Subrahmanyam (2005), for example, show that liquidity is lower in down-markets, consistent with the idea that negative return periods are associated with tighter market-making capacity and higher trading frictions. More broadly, liquidity-based asset-pricing models predict that investors require additional compensation for holding assets that become illiquid precisely when market conditions deteriorate.

This state-dependent interpretation is also consistent with the broader liquidity literature. Acharya and Pedersen (2005) show that expected returns increase not only with expected illiquidity, but also with exposure to market-wide liquidity risk. In particular, investors value assets that remain liquid in bad times, implying that securities whose illiquidity rises in market downturns should command higher expected returns. Survey evidence likewise emphasizes that liquidity premia vary over time and tend to strengthen in periods when investors expect greater liquidity needs or when market stress intensifies.

Recent crisis episodes reinforce the idea that liquidity conditions can deteriorate abruptly under stress. During the COVID-19 shock of March 2020, corporate bond markets experienced severe liquidity problems and spreads rose sharply, illustrating how market stress can amplify liquidity frictions. While this evidence comes from bond markets rather than directly from stock-level illiquidity premia, it supports the broader argument that liquidity conditions are highly state-dependent and can become especially important during crises.

Flight-to-Quality and Liquidity Spirals

Another mechanism through which market states affect liquidity is the phenomenon of flight to quality, often accompanied by flight to liquidity. During periods of market stress, investors tend to rebalance away from risky assets and toward safer and more liquid securities, such as government bonds. This shift reduces demand for risky assets precisely when their liquidity becomes most valuable, thereby worsening trading conditions and amplifying market stress.

This dynamic can generate liquidity spirals. Brunnermeier and Pedersen (2009) provide the central theoretical framework: when liquidity providers face tighter funding constraints and higher margins, they are forced to reduce positions, which increases selling pressure and lowers market liquidity. Falling prices then erode collateral values and can trigger further margin tightening and additional forced sales, creating a self-reinforcing feedback loop between funding liquidity and market liquidity. The model further implies that these effects are stronger for assets that are already relatively illiquid or funding-intensive, so the liquidity premium should be especially high during periods of funding stress and financial crisis.

Empirical evidence is consistent with the view that market stress is associated with stronger comovement and sharper liquidity deterioration. The crisis literature on 2007-2008 describes how falling asset prices, tighter margins, fire sales, and deteriorating market liquidity interacted in mutually reinforcing ways, illustrating the type of liquidity spiral emphasized by Brunnermeier and Pedersen.

5. The German Equity Market: Structure and Liquidity

Understanding the liquidity-return relationship in Germany requires a thorough appreciation of the German equity market's unique structural characteristics, institutional framework, and historical development. This section provides context on the German market's organization, its key indices and their liquidity profiles, the regulatory environment affecting trading, and the limited evidence available on liquidity pricing specifically for German equities. This institutional background is essential for interpreting the empirical findings of this thesis and for assessing the generalizability of conclusions drawn from studies conducted in other markets.

Market Structure and Trading Venues

The German equity market is organized around the **Frankfurt Stock Exchange**, which operates primarily through the **Xetra** electronic trading platform. Xetra, launched in 1997 and fully integrated into the Deutsche Börse group, handles approximately **90 percent** of all stock trading in Germany, making it the dominant trading venue for German equities . The concentration of trading on a single electronic platform distinguishes Germany from markets such as the United States, where trading is distributed across multiple exchanges and alternative trading systems. This centralization provides advantages in terms of price discovery and transparency but also raises questions about market resilience during periods of stress.

The remaining approximately 10 percent of German equity trading occurs on the **Frankfurt Stock Exchange floor** and on regional exchanges, including exchanges such us Berlin, Düsseldorf, Hamburg, Munich, and Stuttgart. While these venues handle lower volumes, they serve important functions for certain segments of the market, including small-cap stocks and retail trading. The **Scale** segment, operated by Deutsche Börse, provides a dedicated trading venue for small and medium-sized enterprises (SMEs) with simplified disclosure requirements and lighter regulatory burdens.

The **Eurex** derivatives exchange, also operated by Deutsche Börse, provides a complementary market for equity derivatives, including futures and options on DAX and MDAX constituents. The interaction between the cash equity market and the derivatives market affects price discovery, arbitrage activity, and overall liquidity dynamics.

Liquidity Characteristics and Variation Across the Index Family
The hierarchical index structure creates substantial variation in liquidity characteristics across firm size categories. The DAX companies benefit from extensive analyst coverage, large institutional investor holdings, deep order books, and active market making features that result in tight bid-ask spreads, high turnover, and low price impact. MDAX companies are less liquid than DAX constituents but still attract significant institutional interest and benefit from active market making on Xetra. The SDAX and TecDAX companies exhibit substantially lower liquidity, with wider spreads, lower turnover ratios, and greater price impact per unit of volume.
This liquidity gradient creates a natural laboratory for examining how the liquidity-return relationship varies across the size spectrum. If the liquidity premium is concentrated among smaller, less liquid stocks as theoretical models predict then the SDAX and lower-MDAX segments should exhibit the strongest illiquidity effects.

Regulatory Environment and Recent Reforms
The German equity market operates within the broader European regulatory framework, which has undergone significant transformation in recent years. The implementation of **MiFID II** in January 2018 introduced enhanced transparency requirements, stricter best execution obligations, and increased scrutiny of over the counter (OTC) trading. These reforms aimed to improve market quality and investor protection but also altered the competitive landscape among trading venues and affected the economics of market making and high-frequency trading.
The German regulatory environment also encompasses specific national provisions, including rules governing the supervision of capital markets (administered by **BaFin**, the Federal Financial Supervisory Authority), disclosure requirements for listed companies, and rules governing insider trading and market manipulation. The **WpHG** transposes European directives into German law and establishes the legal framework for securities trading.

Institutional Context: Ownership, Governance, and Investor Base
The German equity market exhibits several distinctive institutional features that may affect liquidity dynamics and the pricing of liquidity risk.
Ownership concentration in German firms is notably higher than in U.S. and U.K. markets. Studies documented that founding families, financial institutions, and industrial blockholders retain substantial stakes in many German companies, reducing the free-float available to public investors. This concentrated ownership pattern affects liquidity in multiple ways: large blockholders are less likely to trade frequently, reducing available supply of shares; blockholder presence may reduce information asymmetry (if blockholders have access to private information) or increase it (if blockholder transactions convey information); and concentrated ownership can create corporate control dynamics that affect stock price behavior.
The **two-tier board system** (*Vorstand* and *Aufsichtsrat*) and the **co-determination** (*Mitbestimmung*) tradition, whereby employees elect representatives to supervisory boards, shape corporate governance in ways that may affect investor participation and information asymmetry. While these features are primarily relevant to corporate governance research, they may indirectly affect liquidity by influencing the investor base and the perceived risk of German equities.
The **institutional investor landscape** in Germany is characterized by the significant role of banks, insurance companies, and investment funds. Germany is among the largest fund market in the European Union. This large institutional investor base provides depth to the German equity market, though regulatory and contractual constraints may limit the participation of some institutional investors in less liquid stocks.

6. Research Gap Identification
Despite the substantial theoretical and empirical literature on liquidity and asset pricing, several significant gaps remain that this thesis addresses.
Gap 1: Limited country-specific evidence for Germany.

The existing empirical literature is dominated by studies of U.S. equity markets, with comparatively limited attention devoted to Continental European markets. The few European studies that exist typically pool observations across multiple countries, masking potential heterogeneity arising from differences in market structure, ownership patterns, and regulatory environments. For Germany Europe’s largest economy with distinctive characteristics including concentrated ownership structures, a bank-centered financial system, and a dominant electronic trading platform country-specific evidence on liquidity pricing is notably scarce. This gap is consequential because the theoretical mechanisms generating the liquidity premium (adverse selection, inventory costs, search frictions) may operate differently in Germany's institutional context. The first contribution of this thesis is to provide rigorous, Germany-specific evidence on the liquidity-return relationship.

Gap 2: Limited evidence on market state effects in Europe.

While the theoretical prediction that the liquidity premium varies across market states is well established, empirical tests of this conditional relationship remain limited, particularly for European markets. Most prior studies estimate time-invariant relationships, implicitly assuming that the liquidity effect is constant regardless of market conditions. The 2008 financial crisis and 2020 COVID-19 pandemic have highlighted the importance of understanding how liquidity risk behaves during market stress, yet country-specific evidence on state-dependent liquidity pricing in Europe is sparse. The second contribution of this thesis is to examine explicitly whether the liquidity-return relationship in Germany varies across different market states including bear markets.

Gap3: Limited heterogeneity analysis for German equities

The theoretical framework predict that the liquidity premium should vary across the size spectrum, being larger for smaller stocks with higher information asymmetry, lower analyst coverage, and more limited investor participation. However, systematic evidence on size-dependent liquidity effects in Germany is lacking. The German market's tiered index structure (DAX, MDAX, SDAX, TecDAX) provides an ideal setting to examine whether the liquidity effect is concentrated among smaller, less liquid stocks or represents a more broadly distributed phenomenon. The third contribution of this thesis is to investigate heterogeneity in the liquidity-return relationship across firm size categories, providing insights relevant to both asset pricing theory and portfolio management practice.

Chapter 3: Data and Methodology

1. Data Sources and Sample Construction

The empirical analysis is based on German stock-level data obtained from a commercial financial database. The sample includes common stocks that entered the DAX, MDAX, SDAX, or TecDAX at any point during the sample period, thus covering the principal large-cap, mid-cap, small-cap, and technology segments of the German equity market. Two datasets are combined: a monthly stock file used for returns and firm-level controls, and a daily stock file used to construct monthly liquidity measures. The daily data includes stock prices, trading volume, shares outstanding, daily highs and lows, and return information.

The baseline sample is constructed by first restricting the universe to firms belonging to at least one of the four index families during the sample period. Monthly liquidity measures are computed only when at least 15 valid daily observations are available within a given month. To mitigate microstructure distortions associated with very low-priced securities, penny stocks are excluded using a €1 price filter. The final panel is unbalanced, reflecting entry and exit dynamics, index reclassifications, IPOs, and delistings over time.

The final baseline sample contains 74,381 firm-month observations for 845 firms over 419 months. This sample forms the basis for the main univariate, bivariate, and Fama–MacBeth analyses. Sentiment-based tests are conducted on a narrower sub-sample with more limited time-series coverage.

2. Liquidity Measurement

This section defines the 4 liquidity measures used in this thesis. Each measure captures a different dimension of liquidity transaction cost, trading activity, price impact, and trading absence to ensure robust inference and to address the well-documented measurement challenges in the liquidity literature.

Amihud Illiquidity Ratio
The primary proxy is the Amihud (2002) illiquidity ratio, computed at the monthly frequency from daily data:

$$ILLIQ_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} \frac{|R_{i,d}|}{DVOL_{i,d}}$$

Where $R_{i,d}$ the daily return of stock i on day d , $DVOL_{i,d}$ is the corresponding daily trading volume in euros, and $D_{i,t}$ is the number of trading days in month t (minimum 15). The ratio measures the absolute return per unit of trading volume and captures price impact how much the price moves when one euro is traded. Higher values indicate greater illiquidity. The one-month lagged value is used in all regressions to ensure the liquidity measure precedes the dependent variable. The proxy is winsorized at the 1st and 99th percentiles within each month to limit the influence of extreme outliers.

Turnover Ratio
Monthly turnover is defined as the average daily trading volume divided by the number of shares outstanding:

$$TURNOVER_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} \frac{VOL_{i,d}}{SHARES_{i,d}}$$

where $VOL_{i,d}$ is daily trading volume and $SHARES_{i,d}$ is the number of shares outstanding on day d. This proxy measures trading activity as a fraction of the outstanding share base. Higher turnover indicates greater market liquidity. No winsorization is applied.
Zero-Return
Following Lesmond, Ogden and Trzcinka (1999), the zero-return frequency measures the proportion of trading days in a month with zero returns:

$$ZERORET_{i,t} = \frac{1}{D_{i,t}} \sum_{d=1}^{D_{i,t}} \mathbf{1}(R_{i,d} = 0)$$

A higher frequency of zero returns signals trading constraints or low trading activity, as investors fail to trade even when new information arrives. This measure is bounded between 0 and 1 and is not winsorized.

High-Low Range
The high-low range, adapted from measures daily return volatility relative to price level:

$$HL\_RANGE_{i,d} = \frac{H_{i,d} - L_{i,d}}{(H_{i,d} + L_{i,d})/2}$$

where $H_{i,d}$ and $L_{i,d}$ are the daily high and low prices. The monthly average of daily ranges is used as the proxy. Higher values indicate greater price volatility, which may reflect lower market depth and wider effective spreads.

3. Control Variables
In line with the cross-sectional asset pricing literature, the baseline regressions include lagged firm characteristics so that the explanatory variables are dated prior to the return being explained. In the baseline implementation, the control set clearly comprises firm size and book-to-market.

Log Size is defined as the natural logarithm of market capitalization in euros and controls for the size effect in expected returns. It also serves as the main secondary sorting variable in the bivariate portfolio analysis.

Book-to-Market is measured as the ratio of book equity to market equity using the latest available accounting information. It is included to account for the value effect commonly documented in the empirical asset pricing literature.

To remain fully consistent with the empirical implementation, only controls explicitly present in the baseline coding structure are treated as general controls. Any additional variables used in extended specifications should be introduced separately at the level of the corresponding model.

4. Market State Classification

This section defines the market states used in the conditional analysis. Two complementary approaches are employed. The first is a return-based classification using lagged realized market performance, while the second is a sentiment-based classification using the SAFE Manager Sentiment Index. Together, these two approaches capture different dimensions of market conditions.

Bull and Bear Market Classification

Bull and Bear states are defined using the cumulative return on the German value-weighted market index over the previous twelve months, from $t\text{-}12$ to $t\text{-}1$. A month t is classified as Bull when this lagged 12-month cumulative market return is positive, and as Bear otherwise. Since this procedure requires twelve prior monthly returns, the first twelve months of the sample are not classified. This classification is used to assess whether the pricing of liquidity varies across different realized market environments.

Safe Manager Sentiment Classification

The sentiment-based analysis uses the SAFE Manager Sentiment Index, merged with the monthly stock-level panel over the overlapping sample period. In the empirical implementation, months are classified into High and Low sentiment states using a median split of the monthly sentiment series. Because the sentiment data are available only for a shorter period than the baseline sample, the sentiment-based tests are conducted on a reduced matched sample. The SAFE index is constructed from textual analysis of annual and quarterly reports and earnings conference call transcripts of firms from the DAX, MDAX, SDAX, and TecDAX, and therefore captures managerial sentiment rather than investor survey sentiment.

5. Empirical Model Specification

This section presents the econometric approaches used to test the main hypotheses of the thesis. Two complementary methodologies are employed: portfolio sort analysis and Fama-MacBeth cross-sectional regressions.

Portfolio Sort Approach

At the end of each month, stocks are ranked according to the lagged Amihud illiquidity ratio and assigned to quintile portfolios. Equal-weighted portfolio returns, and in the main specification equal-weighted excess returns, are then computed over the following month. This procedure is repeated for each month in the sample. The illiquidity premium is measured as the return spread between the most illiquid portfolio (Q5) and the most liquid portfolio (Q1). Statistical inference is based on Newey-West standard errors with 12 lags in order to account for heteroskedasticity and serial correlation in the portfolio return series.

Univariate portfolio sorts rank stocks on lagged Amihud only. Bivariate sequential sorts first allocate stocks into five size quintiles and then, within each size quintile, into five Amihud quintiles, thereby forming a 5×5 portfolio matrix. The size-neutral illiquidity premium is computed each month as the average of the within-size Q5–Q1 spreads across the five size groups.

Risk-adjusted performance is evaluated using the CAPM and a sample-based Fama-French three-factor model, where the market excess return, SMB, and HML factors are constructed from the German sample.

Fama-MacBeth Cross-Sectional Regressions

The Fama-MacBeth two-step procedure is used to estimate the cross-sectional relation between liquidity and subsequent stock returns. In the first step, a monthly cross-sectional regression is estimated using lagged liquidity and control variables. In the second step, the time-series mean of the monthly slope coefficients is computed, and statistical significance is assessed using Newey-West standard errors with 12 lags.

In the baseline specifications, the main control variables are lagged firm size and lagged book-to-market. Additional specifications replace the raw Amihud proxy with a log-transformed version, defined as log (Amihud+10^{-8}), in order to reduce the effect of extreme right-skewness and to assess whether the results are sensitive to the scaling of the liquidity proxy.

Conditional specifications estimate the models separately across Bull and Bear market states and across High and Low Sentiment states, allowing the analysis to assess whether the liquidity-return relation varies across different market environments.

6. Estimation Methodology

Standard Error Corrections

Statistical inference in the portfolio analysis is based on Newey-West (1987) heteroskedasticity- and autocorrelation-consistent standard errors with 12 lags. The same correction is applied in the time-series analysis of portfolio spreads and risk-adjusted alphas. In the Fama-MacBeth framework, Newey-West standard errors with 12 lags are applied to the time-series of monthly cross-sectional coefficient estimates.

Factor Model Estimation

The Fama-French factors are constructed directly from the German sample. At the end of each month, stocks are sorted by size using the cross-sectional median and by book-to-market using the 30th and 70th percentiles, which yields six portfolios: S/L, S/M, S/H, B/L, B/M, and B/H. The SMB factor is defined as the average return on the three small-stock portfolios minus the average return on the three big-stock portfolios. The HML factor is defined as the average return on the two high book-to-market portfolios minus the average return on the two low book-to-market portfolios. The market factor is measured as the excess return on the German value-weighted market portfolio.

Endogeneity Treatment

The main liquidity proxy, Amihud illiquidity, is lagged by one month relative to the dependent variable, so that liquidity is measured prior to the return being explained. This timing structure reduces concerns about mechanical simultaneity between returns and liquidity. However, it does not fully eliminate the possibility that omitted factors or dynamic feedback effects jointly influence both liquidity and expected returns.

7. Descriptive Statistics

This section reports the summary statistics and correlation structure of all variables used in the empirical analysis.

Table 1. Descriptive Statistics of Liquidity Proxies (Raw Construction Sample, Penny Stocks Excluded)

Proxy	Mean	Median	Std. Dev.	Min	Max	N
Amihud	8,683615	0,171899	81,045762	0	10199,00314	121970
Zero-return	0,332235	0,190476	0,337894	0	1	232635
High-low range	0,02493	0,021799	0,021768	0	0,359077	170085
Turnover	0,006143	0,001061	0,151733	0	7,267	121970

Table 1 reports descriptive statistics for the four liquidity proxies constructed from daily German stock data before merging them with the monthly return panel. Penny stocks (price below €1) are excluded, and each monthly proxy is retained only when at least 15 valid daily observations are available for that specific proxy. The Amihud ratio displays strong right-skewness: its mean is 8,68, whereas its median is only 0,17, and the maximum reaches 10.199. This large gap indicates that a relatively small number of extremely illiquid stock-month observations drive the upper tail of the distribution. The zero-return frequency has a mean of 33,2%, implying that roughly one-third of valid daily observations within a month are associated with zero returns. The high-low range has a mean of 2,49% and a median of 2,18%, suggesting relatively moderate daily price ranges for the typical stock. Monthly turnover has a mean of 0,61% and a median of 0,11%, indicating that trading activity is generally low for a large share of the sample. Sample sizes differ across proxies because the availability of the underlying daily inputs varies across measures. Amihud and turnover are available for 121.970 stock-months, the high-low range for 170.085 stock-months, and the zero-return measure for 232.635 stock-months.

Table 2. Descriptive Statistics of Liquidity Proxies (Merged Panel Sample, Before Winsorization)

Proxy	Mean	Median	Std. Dev.	Min	Max	N
Amihud	6,081437	0,084391	53,591435	0	4058,112781	75608
Zero-return	0,303469	0,15	0,32562	0	1	145556
High-low range	0,0223	0,020002	0,019261	0	0,32125	114533
Turnover	0,002136	0,001109	0,002954	0	0,128459	75608

Table 2 reports the same descriptive statistics after merging the monthly liquidity proxies with the filtered monthly stock panel and before applying winsorization. Relative to the raw construction sample, the Amihud mean declines from 8,68 to 6,08, the median falls from 0,17 to 0,08, and the maximum decreases from 10.199 to 4.058. This indicates that the merged panel excludes a number of extreme illiquidity observations that are not retained in the filtered monthly sample. Turnover is also substantially lower in the merged panel than in the raw construction sample (mean 0,00214 versus 0,00614), suggesting that the retained matched sample is, on average, less actively traded. The zero-return frequency averages 30.4%, slightly below the raw construction sample, while the high-low range also declines moderately. As in Table 1, sample sizes differ across proxies because the availability of the underlying daily inputs varies across measures after merging with the monthly panel.

Table 3. Descriptive Statistics of Liquidity Proxies (After 1%–99% Winsorization, Amihud Only)

Proxy	Mean	Median	Std. Dev.	Min	Max	N
Amihud	3,988615	0,084391	21,157589	0,000003	899,570723	75608
Zero-return	0,303469	0,15	0,32562	0	1	145556
High-low range	0,0223	0,020002	0,019261	0	0,32125	114533
Turnover	0,002136	0,001109	0,002954	0	0,128459	75608

Table 3 reports statistics for Amihud after 1%–99% monthly winsorization. The winsorization reduces the mean from 6,08 to 3,99, cuts the maximum from 4.058 to 899,57, and reduces the standard deviation from 53,59 to 21,16, while the median (0,08) remains unchanged. This pattern confirms that winsorization primarily affects the right tail of the distribution without distorting the central tendency an important property that validates its use in the main analysis. Other liquidity proxies remain at their pre-winsorization values, as the winsorization filter is applied to Amihud only in the baseline specification.

Table 4. Summary Statistics of Dependent and Independent Variables (Baseline Sample)

Variable	Mean	Median	Std. Dev.	Min	Max	N
ret_lead1	0,00991	0,002333	0,131624	-0,889805	4,157898	74381
ret_excess_lead1	0,008846	0,001163	0,131704	-0,889348	4,158295	74381
size_lag1	4419,748951	308,7	13030,33749	0,203823	260848,8	74381
log_size_lag1	5,999471	5,73237	2,245982	-1,590503	12,471696	74381
bm_lag1	0,702593	0,539736	1,065982	0,000072	75,58579	74381
amihud_lag1	3,964471	0,083175	21,018112	0,000003	899,570723	74381
zero_return_lag1	0,089689	0,045455	0,128037	0	1	74381
hl_range_lag1	0,030188	0,026737	0,017837	0	0,32125	74236
turnover_lag1	0,002143	0,001118	0,002908	0	0,115852	74381

Table 4 reports summary statistics for the baseline sample of 74.381 firm-month observations across 845 unique firms over the sample period. Monthly stock returns average 0,99% with a standard deviation of 13,2%, indicating substantial return volatility at the individual-stock level. The maximum monthly return reaches 415,8%, while the minimum is -89%, reflecting the presence of extreme return realizations in the sample. Excess returns over the risk-free rate average 0,88% per month. The log of market capitalization has a mean of 6 and a median of 5,73, while the size variable in levels has a mean of 4.419,75 and a median of 308,7. Book-to-market has a mean of 0,70 and a median of 0,54, although a small number of extreme observations generate a very high upper tail. The lagged Amihud illiquidity proxy has a mean of 3,96, a median of 0,08, and a maximum of 899,57 after winsorization. The lagged zero-return measure averages 8,97%, the lagged high-low range averages 3,02%, and the lagged turnover ratio averages 0,21%. As in the previous tables, the number of non-missing observations differs slightly across variables because some liquidity proxies are unavailable for a small number of stock-months.

Table 5. Liquidity Characteristics by Firm Size (Small vs. Big)

size_group	mean_amihud	median_amihud	mean_zero_return	median_zero_return	mean_hl_range	median_hl_range	mean_turnover	median_turnover	n_obs
Big	0,274275	0,004027	0,039417	0	0,028771	0,025583	0,002952	0,002006	37085
Small	7,633791	0,789483	0,139676	0,095238	0,031601	0,028197	0,001339	0,000713	37296

Table 5 reports liquidity characteristics for stocks classified each month as Small (below-median market capitalization) and Big (above-median market capitalization). The contrast is strong across all measures. Small stocks have a mean Amihud ratio of 7,63 compared with 0,27 for big stocks, and a median Amihud ratio of 0,79 versus 0,004, indicating substantially greater illiquidity among smaller firms. Small stocks also exhibit a higher zero-return frequency (14,0% versus 3,9%) and lower turnover (0,0013 versus 0,0030), consistent with lower trading activity. By contrast, the high-low range is only moderately higher for small stocks (3,16% versus 2,88%), suggesting that differences in trading activity and price impact are more pronounced than differences in the high-low range proxy. With 37.296 small-stock observations and 37.085 big-stock observations, the split is nearly balanced by construction, since stocks are assigned each month relative to the cross-sectional median of firm size.

Chapter 4: Empirical Results

1. Baseline Results: Liquidity and Cross-Sectional Return

Table 6. Portfolio Mean Excess Returns Sorted by Amihud Illiquidity Quintile (Newey–West t-statistics, 12-month lags)

Portfolio	mean	se	tstat NW	n
Q1	0,00522	0,003026	1,725166	410
Q2	0,006697	0,003926	1,705622	410
Q3	0,007739	0,004233	1,82848	410
Q4	0,010225	0,003899	2,622403	410
Q5	0,010817	0,003805	2,842684	410
Q5 - Q1	0,005597	0,002805	1,995159	410

Table 6 reports mean monthly excess returns for five portfolios formed on the lagged Amihud illiquidity ratio. Returns increase monotonically from Q1 (most liquid) to Q5 (most illiquid). The most liquid quintile (Q1) earns 0,52% per month (t = 1,73) while the most illiquid quintile (Q5) earns 1,08% per month (t = 2,84). The raw illiquidity premium, measured as Q5 − Q1, is 0,56% per month with a t-statistic of 1,995, significant at the 5% level. This result is consistent with hypothesis 1 and is consistent with the illiquidity premium documented in U.S. markets by Amihud and Mendelson . Both Q4 and Q5 generate significant excess returns individually (t > 2.62), while Q1 and Q2 do not exceed the 5% significance threshold, indicating that the premium is concentrated in the most illiquid stocks.

Table 7. Alpha and Market Beta by Liquidity Quintile (CAPM Model, Newey–West t-statistics, 12-month lags)

Portfolio	alpha	alpha_tstat	beta_mkt	beta_mkt_tstat
Q1	0,005031	1,684761	0,044873	0,838602
Q2	0,006356	1,676929	0,08081	1,551882
Q3	0,006855	1,808266	0,209543	3,223835
Q4	0,009236	2,700658	0,234251	5,189538
Q5	0,010076	2,87411	0,175711	3,495204
Q5_minus_Q1	0,005045	1,772522	0,130838	3,207346

Table 7 reports CAPM alphas and market betas for each liquidity quintile portfolio. After adjusting for market risk, all quintile portfolios display positive alphas, although statistical significance is concentrated in the upper illiquidity portfolios. The estimated alpha rises from 0,50% per month for Q1 (t = 1,68) to 1,01% per month for Q5 (t = 2,87), with Q4 also showing a large and significant alpha of 0,92% (t = 2,70). The CAPM alpha of the Q5 − Q1 spread is 0,50% per month, with a t-statistic of 1,77, indicating that the risk-adjusted illiquidity premium remains positive but is not statistically significant at the 5% level. Market betas increase from Q1 to Q4 and decline slightly in Q5, suggesting that differences in CAPM alphas cannot be interpreted as a simple inverse monotonic relation between illiquidity and market beta.

Table 8. Three-Factor Model Alphas and Factor Loadings by Liquidity Quintile (Fama–French 3 Factors, Newey–West t-statistics, 12-month lags)

Portfolio	alpha	alpha_tstat	beta_mkt	beta_smb	beta_hml	beta_mkt_tstat	beta_smb_tstat	beta_hml_tstat
Q1	0,007362	2,404005	0,10108	-0,553843	-0,118395	2,054608	-3,84102	-1,159674
Q2	0,009464	2,910874	0,095154	-0,168689	-0,312043	1,372205	-0,481798	-2,169204
Q3	0,007583	2,133046	0,168913	0,374091	-0,185019	2,887387	1,536798	-1,09247
Q4	0,008931	2,910472	0,197361	0,350164	-0,059635	4,654596	1,781773	-0,759053
Q5	0,008603	2,833539	0,138561	0,365317	0,070613	2,707761	3,492641	1,06624
Q5_minus_Q1	0,001241	0,504222	0,037481	0,91916	0,189008	1,343465	11,567391	2,761176

Table 8 reports three-factor model alphas and factor loadings for each liquidity quintile. The main result is that the raw illiquidity premium is largely absorbed by the Fama-French three-factor model. The FF3 alpha of the Q5 − Q1 spread is only 0,12% per month, with a t-statistic of 0,504, indicating that the residual premium is economically small and statistically indistinguishable from zero. The spread loads strongly on the SMB factor (0,919, t = 11,57) and positively on the HML factor (0,189, t = 2,76), suggesting that the return differential between illiquid and liquid stocks is closely related to size and, to a lesser extent, value exposure. Individual quintile alphas remain positive and statistically significant at conventional levels, ranging from 0,74% for Q1 to 0,95% for Q2, which indicates that abnormal performance persists within portfolios, even though the illiquidity spread itself is no longer priced once the three factors are taken into account.

Table 9. Fama-MacBeth Cross-Sectional Regressions: Illiquidity Premium and Market States

Model	Variable	Coefficient	tstat_NW	N_months
(1) Amihud only	(Intercept)	0,009763	2,5361	327
(1) Amihud only	amihud_lag1	0,001618	1,5394	327
(2) Amihud + Size	(Intercept)	0,017123	2,9235	327
(2) Amihud + Size	amihud_lag1	0,001404	1,3448	327
(2) Amihud + Size	log_size_lag1	-0,001234	-2,6108	327
(3) Baseline FM	(Intercept)	0,012873	1,9989	327
(3) Baseline FM	amihud_lag1	0,001554	1,4064	327
(3) Baseline FM	log_size_lag1	-0,001038	-2,2058	327
(3) Baseline FM	bm_lag1	0,002929	1,0644	327
(4) log(Amihud)	(Intercept)	0,019972	2,6202	327
(4) log(Amihud)	log_amihud_lag1	-0,000921	-1,9542	327
(4) log(Amihud)	log_size_lag1	-0,002535	-2,9634	327
(4) log(Amihud)	bm_lag1	0,003317	1,2373	327
(5a) Bull markets only	(Intercept)	0,021694	3,1701	206
(5a) Bull markets only	amihud_lag1	0,000677	0,7151	206
(5a) Bull markets only	log_size_lag1	-0,001774	-3,1928	206
(5a) Bull markets only	bm_lag1	0,000366	0,0958	206
(5b) Bear markets only	(Intercept)	-0,004146	-0,451	111
(5b) Bear markets only	amihud_lag1	0,002413	1,1193	111
(5b) Bear markets only	log_size_lag1	0,00014	0,2048	111
(5b) Bear markets only	bm_lag1	0,007798	2,4169	111

Table 9 reports Fama-MacBeth cross-sectional regression results across six model specifications. In Model (1), the coefficient on Amihud is positive at 0,001618, but statistically insignificant (t = 1,54). In Model (2), the inclusion of firm size leaves the Amihud coefficient positive but insignificant at 0,001404 (t = 1,34), while log (Size) enters with a negative and statistically significant coefficient of -0,001234 (t = -2,61), consistent with a size effect in the cross-section. In Model (3), after adding the book-to-market ratio, the Amihud coefficient remains positive but insignificant at 0,001554 (t = 1,41). The coefficient on log (Size) remains significantly negative (-0,001038, t = -2,21), whereas B/M is positive but not statistically significant (0,002929, t = 1,06). In Model (4), replacing Amihud with log (Amihud) yields a negative coefficient of -0,000921 with a t-statistic of -1,95, indicating a borderline result but not clear significance at the 5% level. Finally, the state-dependent regressions show that the Amihud coefficient remains statistically insignificant in both bull and bear markets, while B/M becomes positive and significant only in bear markets.

2. **Conditional Analysis: Market State Effects**

Table 10. Mean Excess Returns of Amihud Quintile Portfolios by Market State (Bull vs. Bear)

market_state	Portfolio	mean_ret	tstat_NW	n_months
Bear	Q1	0,003695	0,608	139
Bear	Q2	0,000321	0,0424	139
Bear	Q3	0,001989	0,2755	139
Bear	Q4	0,003193	0,5132	139
Bear	Q5	0,007887	1,425	139
Bear	Q5_minus_Q1	0,004192	1,1486	139
Bull	Q1	0,005672	1,6956	268
Bull	Q2	0,010465	2,3995	268
Bull	Q3	0,011037	2,5084	268
Bull	Q4	0,014236	3,2104	268
Bull	Q5	0,011168	2,5395	268
Bull	Q5_minus_Q1	0,005496	1,2573	268

Table 10 reports mean excess returns by Amihud quintile separately for Bear and Bull market states. In Bear markets (139 months), Q1 earns 0,37% per month (t = 0,61) and Q5 earns 0,79% (t = 1,43), while the Q5 − Q1 spread is 0,42% (t = 1,15), indicating a positive but statistically insignificant illiquidity premium. In Bull markets (268 months), Q1 earns 0,57% (t = 1,70) and Q5 earns 1,12% (t = 2,54), with a Q5 − Q1 spread of 0,55% (t = 1,26), which is also positive but not statistically significant at conventional levels. Overall, the liquidity premium appears slightly larger in Bull than in Bear markets, providing no support for the view that the premium is stronger during adverse market conditions.

Table 11. Risk-Adjusted Alphas for the Illiquidity Spread (Q5-Q1) across Bull and Bear Markets

State	CAPM_alpha	CAPM_tstat	FF3_alpha	FF3_tstat	FF3_beta_smb	FF3_smb_tstat	N
Bull	0,00545	1,4045	0,00163	0,521	0,8533	11,1624	259
Bear	0,004103	1,0016	0,003141	1,2592	1,0491	8,8243	139

Table 11 reports risk-adjusted alphas for the Q5 − Q1 illiquidity spread across Bull and Bear market states. In Bull markets, the CAPM alpha is 0,545% per month (t = 1,405), while the FF3 alpha declines to 0,163% (t = 0,521), indicating no statistically significant residual premium after controlling for the three factors. In Bear markets, the CAPM alpha is 0,410% (t = 1,002) and the FF3 alpha is 0,314% (t = 1,259), which is larger than in Bull markets but still statistically insignificant at conventional levels. The spread loads strongly on the SMB factor in both regimes, with a coefficient of 0,853 (t = 11,162) in Bull markets and 1,049 (t = 8,824) in Bear markets, suggesting that the illiquidity spread is closely related to the size dimension across market states. Overall, while the bear-market FF3 alpha is somewhat larger, the evidence does not support a statistically robust difference in risk-adjusted illiquidity premia across regimes.

Table 12. Mean Excess Returns of Amihud Quintile Portfolios by Manager Sentiment State (High vs. Low)

sent_state_2	Portfolio	mean_ret	tstat_NW	n_months
High	Q1	0,007204	1,7649	124
High	Q2	0,007364	1,7159	124
High	Q3	0,009951	2,0158	124
High	Q4	0,011097	2,1273	124
High	Q5	0,015592	2,5838	124
High	Q5_minus_Q1	0,008388	2,389	124
Low	Q1	0,009482	1,9967	124
Low	Q2	0,014477	2,6137	124
Low	Q3	0,015254	2,4951	124
Low	Q4	0,015251	2,9317	124
Low	Q5	0,014239	2,3919	124
Low	Q5_minus_Q1	0,004757	1,0515	124

Table 12 reports mean excess returns by Amihud quintile separately for High and Low Manager Sentiment states, as defined by the SAFE Manager Sentiment Index. In High Sentiment periods (124 months), returns increase monotonically from Q1 (0,72%, t = 1,765) to Q5 (1,56%, t = 2,584), and the Q5 − Q1 spread reaches 0,84% per month with a t-statistic of 2,389, indicating significance at the 5% level. In Low Sentiment periods (124 months), the return pattern is flatter and less monotonic: Q1 earns 0,95% (t = 1,997) and Q5 earns 1,42% (t = 2,392), while the Q5 − Q1 spread is only 0,48% (t = 1,052) and is not statistically significant. Overall, the raw illiquidity premium is about 0,36 percentage points larger during High Sentiment periods than during Low Sentiment periods.

Table 13. Comparison of the Illiquidity Spread (Q5-Q1) across High and Low Manager Sentiment States

State	Mean_Spread	tstat_NW	N_months
Low Sentiment	0,004757	1,0515	124
High Sentiment	0,008388	2,389	124
High - Low (diff)	0,003631	0,6975	248

Table 13 compares the Q5 − Q1 illiquidity spread across High and Low Manager Sentiment states. The spread is 0,84% per month in High Sentiment periods (t = 2,389) and 0,48% in Low Sentiment periods (t = 1,052). The difference between the two spreads is 0,363 percentage points, but the associated t-statistic of 0,698 indicates that this difference is not statistically significant. Overall, the raw illiquidity premium appears economically larger in High Sentiment periods, but the evidence does not support a statistically robust difference across sentiment regimes.

Table 14. Risk-Adjusted Alphas for the Illiquidity Spread (Q5-Q1) by Manager Sentiment State

Sentiment_State	CAPM_alpha	CAPM_tstat	FF3_alpha	FF3_tstat	FF3_beta_smb	FF3_smb_tstat	FF3_beta_hml	FF3_hml_tstat	N_months
Low	0,003198	0,6602	-0,000577	-0,3997	1,3795	13,983	0,1144	1,4314	124
High	0,007465	2,284	0,002606	1,296	1,3066	18,934	0,1382	1,9695	124

Table 14 reports risk-adjusted alphas for the Q5 − Q1 spread by sentiment state. In High Sentiment periods, the CAPM alpha is 0,747% per month (t = 2,284), indicating significance at the 5% level, whereas the FF3 alpha declines to 0,261% (t = 1,296) and is no longer statistically significant. In Low Sentiment periods, both the CAPM alpha (0,320%, t = 0,660) and the FF3 alpha (−0,058%, t = −0,400) are economically small and statistically insignificant. The SMB factor loads strongly in both states, with coefficients of 1,307 (t = 18,934) in High Sentiment and 1,380 (t = 13,983) in Low Sentiment, suggesting that the spread is closely related to the size dimension regardless of sentiment regime. Overall, the evidence indicates that the positive alpha observed in High Sentiment under CAPM is not robust once size and value controls are introduced in the FF3 specification.

Table 15. Fama-MacBeth Cross-Sectional Regressions conditioned on Manager Sentiment

Model	Variable	Coefficient	tstat_NW	N_month
(S1) Baseline period sent	(Intercept)	0,016162	2,5025	248
(S1) Baseline period sent	amihud_lag1	0,000003	0,047	248
(S1) Baseline period sent	log_size_lag1	-0,001185	-2,3403	248
(S1) Baseline period sent	bm_lag1	0,004236	3,5207	248
(S2) High Sentiment	(Intercept)	0,017111	2,1601	124
(S2) High Sentiment	amihud_lag1	-0,000051	-0,4441	124
(S2) High Sentiment	log_size_lag1	-0,001266	-1,9909	124
(S2) High Sentiment	bm_lag1	0,002415	1,1744	124
(S3) Low Sentiment	(Intercept)	0,015213	1,9217	124
(S3) Low Sentiment	amihud_lag1	0,000058	0,6269	124
(S3) Low Sentiment	log_size_lag1	-0,001105	-1,4807	124
(S3) Low Sentiment	bm_lag1	0,006057	5,3119	124
(S4) log(Amihud) baseline	(Intercept)	0,022848	2,9104	248
(S4) log(Amihud) baseline	log_amihud_lag1	-0,000957	-2,2076	248
(S4) log(Amihud) baseline	log_size_lag1	-0,002708	-2,9145	248
(S4) log(Amihud) baseline	bm_lag1	0,004267	3,8564	248
(S5) log(Amihud) High Sent	(Intercept)	0,018668	1,9046	124
(S5) log(Amihud) High Sent	log_amihud_lag1	-0,000257	-0,5378	124
(S5) log(Amihud) High Sent	log_size_lag1	-0,001686	-1,4703	124
(S5) log(Amihud) High Sent	bm_lag1	0,002587	1,3268	124
(S6) log(Amihud) Low Sent	(Intercept)	0,027028	2,6735	124
(S6) log(Amihud) Low Sent	log_amihud_lag1	-0,001657	-2,3942	124
(S6) log(Amihud) Low Sent	log_size_lag1	-0,00373	-2,7452	124
(S6) log(Amihud) Low Sent	bm_lag1	0,005947	5,8445	124

Table 15 reports Fama-MacBeth regressions conditioned on Manager Sentiment over the 248-month sub-period for which the SAFE index is available. In specification (S1), the coefficient on the raw Amihud measure is essentially zero, indicating that the baseline cross-sectional relation between illiquidity and returns disappears over the SAFE sub-sample. By contrast, log(Size) remains negative and statistically significant, while book-to-market is positive and significant, suggesting that size and value effects remain important in the cross-section. When the sample is split by sentiment state, the raw Amihud coefficient is slightly negative in High Sentiment periods and slightly positive in Low Sentiment periods, but in neither case is it statistically significant. In the log specifications, the coefficient on log(Amihud) is negative throughout and becomes statistically significant in the full SAFE sub-period and, more clearly, in Low Sentiment periods. This indicates that, once illiquidity is measured in logarithmic form, higher illiquidity is associated with lower subsequent returns, especially during Low Sentiment states. Overall, the sentiment-conditioned Fama-MacBeth results suggest that the positive raw illiquidity premium is not robust in the SAFE sub-sample and that the cross-sectional pricing of illiquidity becomes weak or even reverses once the variable is transformed and estimated within a regression framework.

3. Further Analysis and Heterogeneity

Table 16. Mean Excess Returns from Bivariate Sequential Sorts on Size and Amihud Illiquidity

size_quintile	A1	A2	A3	A4	A5
1	0,008729	0,013851	0,010326	0,014224	0,017744
2	0,006715	0,010031	0,008274	0,00706	0,008299
3	0,006024	0,00837	0,005124	0,014353	0,001972
4	0,004829	0,005342	0,007458	0,009302	0,004338
5	0,005075	0,006217	0,005581	0,005603	0,007503

Table 16 reports the mean excess returns from bivariate sequential sorts on size and Amihud illiquidity. The 5×5 matrix shows that the highest returns are concentrated among the smallest and most illiquid stocks, with the largest cell value observed in Size Quintile 1, Amihud Quintile 5. Computing the within-size illiquidity spread from the displayed values shows that the spread is largest for the smallest firms (0,90% per month), much smaller for Size Quintiles 2 and 5, and negative for Size Quintiles 3 and 4. Overall, the pattern suggests that the raw illiquidity premium is strongest among small stocks and weak or absent among medium-sized and large firms

Table 17. Illiquidity Spreads (A5 − A1) within Size Quintiles from Bivariate Sequential Sorts

size_quintile	mean_A1	mean_A5	mean_spread	tstat_NW	n_months
1	0,008729	0,017744	0,009015	1,7507	419
2	0,006715	0,008299	0,002782	0,5299	414
3	0,006024	0,001972	-0,002949	-0,7057	413
4	0,004829	0,004338	-0,000497	-0,128	414
5	0,005075	0,007503	0,002428	0,8441	419

Table 17 reports the illiquidity spread within each size quintile from the bivariate sequential sorts on size and Amihud. Within the smallest size quintile, the return difference between the most illiquid portfolio (A5) and the most liquid portfolio (A1) is 0,90% per month (t = 1,751), which is the largest spread across all size groups. In Size Quintile 2, the spread is smaller at 0,28% (t = 0,530), while Size Quintiles 3 and 4 display negative spreads of −0,29% (t = −0,706) and −0,05% (t = −0,128), respectively. Among the largest firms (Size Quintile 5), the spread is positive but modest at 0,24% (t = 0,844). Overall, the results indicate that the raw illiquidity premium is concentrated primarily among the smallest stocks, whereas it is weak or absent among medium-sized and large firms.

Table 18. Risk-Adjusted Alphas for the Size-Neutral Illiquidity Premium.

Model	alpha	alpha_tstat	beta_mkt	beta_mkt_tstat	beta_smb	beta_smb_tstat	beta_hml	beta_hml_tstat
CAPM	0,002897	1,215031	0,033687	0,970118				
Fama-French 3F	0,000665	0,193559	0,017555	0,512621	0,175955	2,604643	0,209189	1,179896

Table 18 reports risk-adjusted alphas for the size-neutral illiquidity premium, constructed as the monthly average of the Q5 − Q1 spread across the five size groups. Under the CAPM, the alpha is 0,290% per month (t = 1,215), indicating a positive but statistically insignificant premium. After controlling for the Fama-French three factors, the alpha declines sharply to 0,067% (t = 0,194), becoming economically negligible and statistically indistinguishable from zero. The FF3 regression shows a positive and significant loading on SMB (0,176, t = 2,605), while the HML loading is positive but not statistically significant (0,209, t = 1,180). Overall, these results suggest that the size-neutral illiquidity premium is substantially attenuated once standard factor exposures are taken into account, with the SMB dimension playing an important role in explaining the raw premium.

4. Robustness Checks

Table 19. Robustness Check: Fama-MacBeth Regressions with Alternative Liquidity Proxies

Model	Variable	Coefficient	tstat_NW	N_months
(R1) Amihud (baseline)	(Intercept)	0,012873	1,9989	327
(R1) Amihud (baseline)	amihud_lag1	0,001554	1,4064	327
(R1) Amihud (baseline)	log_size_lag1	-0,001038	-2,2058	327
(R1) Amihud (baseline)	bm_lag1	0,002929	1,0644	327
(R2) Zero-return	(Intercept)	0,01564	2,1352	327
(R2) Zero-return	zero_return_lag1	-0,008355	-1,2384	327
(R2) Zero-return	log_size_lag1	-0,001379	-2,5697	327
(R2) Zero-return	bm_lag1	0,003875	1,5795	327
(R3) High-low range	(Intercept)	0,008952	2,0552	321
(R3) High-low range	hl_range_lag1	0,107613	1,1752	321
(R3) High-low range	log_size_lag1	-0,000989	-2,4406	321
(R3) High-low range	bm_lag1	0,004559	2,5357	321
(R4) Turnover	(Intercept)	0,012207	2,1169	327
(R4) Turnover	turnover_lag1	1,128494	1,3418	327
(R4) Turnover	log_size_lag1	-0,001245	-2,3343	327
(R4) Turnover	bm_lag1	0,004661	2,3589	327

Table 19 reports Fama-MacBeth regressions that replace the baseline Amihud measure with alternative liquidity proxies. In Model (R1), the Amihud coefficient remains positive at 0,001554 but is not statistically significant (t = 1,406). In Model (R2), the zero-return proxy enters with a negative coefficient of -0,008355, but the estimate is not statistically significant (t = -1,238). In Model (R3), the high-low range proxy has a positive coefficient of 0,107613, again statistically insignificant (t = 1,175). In Model (R4), turnover also enters positively, with a coefficient of 1,128494, but remains insignificant (t = 1,342). Across all four specifications, log (Size) is negative and statistically significant, while B/M is positive and significant only in the high-low range and turnover models. Overall, the robustness results suggest that the cross-sectional relation between liquidity and expected returns is sensitive to the proxy used, whereas the size effect remains stable across specifications.

Table 20. Robustness Analysis: Fama-MacBeth Regressions with the Four Liquidity Proxies

Variable	(R1) Amihud (baseline)	(R2) Zero-return	(R3) High-low range	(R4) Turnover
Intercept	0.0129 (2.00)**	0.0156 (2.14)**	0.0090 (2.06)**	0.0122 (2.12)**
Liquidity proxy	0.0016 (1.41)	-0.0084 (-1.24)	0.1076 (1.18)	1.1285 (1.34)
log(Size)	-0.0010 (-2.21)**	-0.0014 (-2.57)**	-0.0010 (-2.44)**	-0.0012 (-2.33)**
B/M	0.0029 (1.06)	0.0039 (1.58)	0.0046 (2.54)**	0.0047 (2.36)**
N (months)	327	327	321	327

Table 20 reports robustness tests of the baseline Fama–MacBeth specification using alternative liquidity proxies. Across all four regressions, the liquidity coefficient remains statistically insignificant at conventional levels, indicating that the baseline cross-sectional liquidity effect is not robust to proxy substitution. By contrast, log(Size) remains negative and statistically significant in every specification, while B/M is significant only in the high-low range and turnover regressions. Overall, these results suggest that the illiquidity premium observed in portfolio sorts is not robustly confirmed in the cross-sectional regression framework when alternative liquidity measures are employed.

Chapter 5: Discussion

1. Summary of Findings

This section summarizes the main empirical findings of this thesis.

Finding 1: A positive raw illiquidity premium is observed in the German equity market when stocks are sorted on Amihud illiquidity (Table 6). The average excess return rises monotonically from 0,52% per month for the most liquid quintile (Q1) to 1,08% per month for the most illiquid quintile (Q5). The corresponding Q5 − Q1 spread equals 0,56% per month with a Newey–West t-statistic of 1,995, which is just significant at the 5% level. This provides initial support for Hypothesis 1, namely that illiquid stocks earn higher expected returns than liquid stocks in the German market.

Finding 2: The raw illiquidity premium weakens substantially once standard asset-pricing controls are introduced (Tables 7 and 8). Under the CAPM, the alpha of the Q5 − Q1 spread remains positive at 0,50% per month, but it is no longer significant at the 5% level (t = 1,77). Under the Fama–French three-factor model, the spread alpha falls further to only 0,12% per month (t = 0,504), while the spread loads very strongly on SMB (beta = 0,919, t = 11,57) and positively on HML (beta = 0,189, t = 2,76). These results suggest that a substantial part of the raw illiquidity premium is closely related to size, and to a lesser extent value, rather than to a robust standalone illiquidity effect.

Finding 3: The cross-sectional pricing evidence is fragile in Fama–MacBeth regressions (Table 9). In the baseline specification, the Amihud coefficient is positive but statistically insignificant (0,001618, t = 1,54). After controlling for size and book-to-market, the coefficient remains positive but insignificant (0,001554, t = 1,41). When Amihud is log-transformed, the coefficient turns negative (−0,000921, t = −1,95), indicating that the sign and magnitude of the illiquidity effect are sensitive to specification. At the same time, log(Size) remains negative and statistically significant in the pooled models, whereas B/M is generally weaker and becomes significant only in selected specifications. Overall, Table 9 shows that the positive raw premium observed in portfolio sorts is not robustly confirmed in the cross-section.

Finding 4: There is no statistically robust evidence that the illiquidity premium is stronger in Bear markets than in Bull markets (Tables 10 and 11). In Bear markets, the Q5 − Q1 spread equals 0,42% per month (t = 1,15), whereas in Bull markets it equals 0,55% (t = 1,26). The difference between the two spreads is −0,13 percentage points with a t-statistic of −0,23, indicating no significant state-dependent effect. Similarly, the FF3 alpha of the spread is 0,314% (t = 1,259) in Bear markets and 0,163% (t = 0,521) in Bull markets, with neither estimate significant at conventional levels. Therefore, Hypothesis 2 is not supported by the market-state evidence.

Finding 5: The illiquidity premium appears economically larger during High Manager Sentiment periods, but this difference is not statistically robust (Tables 12, 13 and 14). In High Sentiment states, the Q5 − Q1 spread reaches 0,84% per month (t = 2,389), whereas in Low Sentiment states it declines to 0,48% (t = 1,052). The difference between the two spreads is 0,363 percentage points, but the corresponding t-statistic of 0.698 indicates that it is not statistically significant. The CAPM alpha in High Sentiment periods is 0,747% (t = 2,284), yet the FF3 alpha drops to 0,261% (t = 1,296), again suggesting that the apparent premium is not robust after factor adjustment. Manager sentiment therefore provides supplementary conditional evidence, but not a decisive confirmation of a sentiment-dependent liquidity premium.

Finding 6: The liquidity premium is strongly concentrated among small stocks, which provides partial support for Hypothesis 3 (Tables 16, 17 and 18). The 5×5 bivariate sorts show that the highest returns are concentrated among the smallest and most illiquid stocks, and the within-size illiquidity spread is largest in the smallest size quintile. In Table 17, the A5 − A1 spread for the smallest firms reaches 0,90% per month with a t-statistic of 1,751, making it the only size group with a marginally significant positive spread. By contrast, the spreads are much smaller for other size groups and even negative for the middle size quintiles. However, once the premium is averaged across size groups, the size-neutral alpha becomes statistically insignificant under both CAPM (0,290%, t = 1,215) and FF3 (0,067%, t = 0,194). This indicates that the raw illiquidity premium is concentrated mainly among small-cap stocks rather than being uniformly priced across the German market.

Finding 7: The cross-sectional liquidity effect is not robust to alternative liquidity proxies, and the overall evidence remains sensitive to specification choices (Tables 19 and 20). When the baseline Fama–MacBeth regression is re-estimated with alternative liquidity measures, none of the four proxies is statistically significant at conventional levels. The Amihud coefficient remains positive but insignificant (0,001554, t = 1,406), the zero-return proxy becomes negative (−0,008355, t = −1,238), while high-low range (0,107613, t = 1,175) and turnover (1,128494, t = 1,342) remain positive but insignificant. This pattern suggests that the cross-sectional relation between liquidity and expected returns in Germany depends materially on the proxy used. More broadly, the robustness analysis indicates that while the illiquidity premium is detectable in raw portfolio sorts, it does not survive consistently in cross-sectional regression frameworks and should therefore be interpreted with caution.

2. Comparison with Prior Literature

The findings of this thesis align partially with the international evidence on liquidity and asset pricing Overall, our findings are only partially consistent with the existing literature on liquidity and asset pricing. In line with the foundational U.S. evidence of Amihud and Mendelson (1986) and Amihud (2002), we document a positive raw illiquidity premium in the German equity market, as the portfolio of the most illiquid stocks earns higher average returns than the most liquid portfolio. However, this premium weakens substantially once standard risk controls are introduced, especially in the Fama–French framework, where the spread alpha becomes insignificant and the strong loading on SMB suggests that part of the apparent liquidity effect is closely tied to firm size rather than to an autonomous liquidity premium. This pattern is broadly consistent with the view that liquidity effects are stronger among smaller firms, but it is less supportive of the stronger cross-sectional pricing results reported for the U.S. market. In addition, our Fama–MacBeth estimates and robustness checks based on alternative liquidity proxies indicate that the liquidity-return relation in Germany is fragile and specification-dependent, which contrasts with the more robust evidence often reported in the early literature. Finally, while conditional asset pricing theory predicts that liquidity should be more strongly priced in adverse market states, our bear-versus-bull results do not provide statistically strong support for such state dependence. Taken together, these results suggest that the German market exhibits a raw illiquidity effect, but that this effect is neither uniformly robust across specifications not clearly stable across market conditions, which places our evidence closer to the more recent and nuanced literature than to the stronger conclusions of the early U.S. studies.

3. Theoretical Implications

Our results carry several theoretical implications for the asset-pricing role of liquidity. First, the presence of a positive raw illiquidity premium in portfolio sorts is consistent with the classical view that investors require compensation for holding securities that are more costly to trade, as proposed by Amihud and Mendelson and reinforced by Amihud. However, the fact that this premium becomes weak or insignificant once Fama–French controls are introduced suggests that, in the German market, liquidity may not operate as a fully autonomous source of expected return, but rather overlaps substantially with other stock characteristics, especially firm size. This interpretation is also in line with the broader asset-pricing literature showing that liquidity can be closely related to common risk factors and to systematic liquidity conditions, rather than being priced uniformly as a standalone characteristic in all settings. Second, the concentration of the raw premium among smaller firms supports the theoretical idea that liquidity frictions are more severe where trading is thinner and transaction costs are higher, but the lack of robust cross-sectional and conditional evidence implies that this mechanism is not stable enough to generate a consistently priced premium across specifications. Finally, the absence of strong bear-versus-bull effects suggests that, although conditional asset-pricing theory and liquidity-spiral models predict stronger pricing of illiquidity in adverse states, this prediction is not clearly validated in our German sample. Taken together, these findings imply that liquidity in Germany is economically relevant, but its pricing appears to be conditional, intertwined with other firm characteristics, and less structurally robust than the strongest theoretical models might suggest.

4. Practical Implications

Our findings also carry important practical implications for investors, portfolio managers, and market participants. First, the positive raw illiquidity premium suggests that less liquid German stocks may offer higher average returns, which means that liquidity can be a relevant screening criterion in portfolio construction, particularly for investors willing to bear higher trading costs and lower marketability. However, because this premium becomes weak after Fama–French controls and proves fragile across cross-sectional regressions and alternative liquidity proxies, investors should be cautious about treating illiquidity as a standalone return-enhancing strategy. In practical terms, the evidence suggests that part of the apparent premium may reflect indirect exposure to small-cap risk rather than a stable and independent liquidity effect. Second, the stronger raw premium among smaller firms implies that liquidity considerations are especially important in the small-cap segment, where transaction costs, price impact, and implementation constraints are likely to be more severe. For institutional investors in particular, this means that any attempt to exploit an illiquidity premium must take account of execution costs, limited trading depth, and the possibility that paper returns overstate realizable net returns. Finally, the absence of strong and robust state-dependent results indicates that market participants should not assume that illiquid stocks will systematically deliver superior compensation in downturns or stress periods. Overall, the practical message is that liquidity matters for investment decisions in the German market, but it should be incorporated as part of a broader multi-factor and implementation-aware framework rather than used in isolation as a simple return predictor.

5. Limitations

This study is subject to several limitations that should be taken into account when interpreting the results. First, the analysis is confined to the German equity market, which limits the external validity of the findings and makes it difficult to generalize the conclusions to other developed or emerging markets with different institutional structures, trading mechanisms, and investor bases. Second, the empirical results depend importantly on the choice of liquidity proxy and specification. Although Amihud's measure is widely used in the literature, it remains an indirect proxy for trading frictions and can be sensitive to extreme observations, which is consistent with the fact that our results weaken under alternative liquidity measures. Third, the contrast between the positive raw portfolio-sort evidence and the weaker Fama–MacBeth results suggests that the estimated liquidity premium is not fully stable across empirical methods, raising the possibility that part of the effect reflects correlated firm characteristics such as size rather than a purely independent pricing of liquidity. Fourth, the conditional analyses based on market states and sentiment are informative but remain sensitive to the way regimes are defined and to sample partitioning, which may reduce statistical power. Finally, because the study focuses on gross return patterns rather than implementable trading strategies net of full transaction costs, the economic profitability of exploiting the observed illiquidity premium may be lower than the reported estimates suggest. Taken together, these limitations imply that the results should be interpreted as evidence of a meaningful but fragile liquidity-return relation in Germany, rather than as proof of a uniformly robust and generalizable liquidity premium.

6. Directions for Future Research

Future research could extend this study in several important directions. First, it would be useful to examine whether the relatively fragile liquidity-return relationship documented for Germany also appears in other continental European markets, which would help determine whether the results reflect country-specific institutional features or a broader regional pattern. Second, future work could employ a wider set of liquidity measures, including direct microstructure-based proxies such as bid-ask spreads, market depth, and order-book indicators, in order to assess whether the weak robustness observed here is driven by the limitations of indirect measures such as Amihud, zero-return frequency, or turnover. Third, additional research could investigate the role of liquidity within a richer conditional framework by considering alternative definitions of market states, macroeconomic uncertainty, monetary policy conditions, or crisis regimes, which may capture time variation in liquidity pricing more effectively than a simple bull-versus-bear classification. Fourth, future studies could focus more explicitly on implementation by examining whether the apparent illiquidity premium survives realistic transaction costs, trading constraints, and portfolio capacity limits, especially in the small-cap segment where the raw effect appears strongest. Finally, a promising avenue would be to explore the interaction between liquidity, firm characteristics, and investor sentiment using longer samples, higher-frequency data, and dynamic models, in order to better distinguish whether liquidity is priced as an independent source of return or mainly reflects its overlap with size, value, and broader market conditions.

Chapter 6: Conclusion

In conclusion, this thesis provides a nuanced assessment of the liquidity-return relationship in the German equity market. The empirical results show that illiquid stocks earn higher average returns than liquid stocks in simple portfolio sorts, which is consistent with the classical prediction that investors require compensation for bearing trading frictions and limited marketability. However, the evidence does not support the view that this premium is strong, uniform, or fully independent of other asset-pricing forces. Once standard factor controls are introduced, the apparent illiquidity premium declines substantially and often loses statistical significance, while the strong association with the size factor suggests that part of the raw return differential reflects the characteristics of small firms rather than a pure liquidity effect. This interpretation is reinforced by the cross-sectional regressions and robustness checks, which show that the liquidity effect is sensitive to model specification, variable transformation, and the alternative proxy used to capture trading frictions. In addition, although the premium appears stronger during periods of high manager sentiment, with a larger raw spread and a significant CAPM alpha, this effect also weakens once Fama–French controls are applied, indicating that sentiment amplifies the observed return pattern without providing conclusive evidence of a fully distinct and robust liquidity-pricing channel. Similarly, while conditional asset-pricing theory suggests that liquidity should be more strongly priced in adverse states of the market, the bear-versus-bull analysis does not provide statistically robust evidence that the German illiquidity premium systematically intensifies during downturns. Taken together, these results point to an important but cautious conclusion: liquidity clearly matters in the German stock market, but its role is better understood as economically meaningful yet empirically fragile, closely intertwined with firm size, sentiment conditions, factor exposures, and methodological choices rather than as a universally robust standalone premium. In this sense, the thesis contributes to the literature not by confirming the strongest version of the illiquidity-pricing hypothesis, but by showing that, in the German context, the premium is visible in raw return patterns, appears somewhat stronger under high sentiment conditions, and nevertheless becomes considerably more fragile under stricter empirical scrutiny.

Appendices

Appendix 1. Complete R Code

The full R code used for data preparation, variable construction, empirical estimation, robustness checks, and table generation is reported below for transparency and reproducibility purposes.

```
# ============================================================================
# THESIS — Liquidity and the Cross-Section of Stock Returns in Germany and the effect of different market states
# SCRIPT FINAL PROPRE ET ORGANISÉ
# BLOC 1/3 — SETUP + BASELINE + TABLES 1 TO 17
# ============================================================================

# ============================================================================
# 0. PACKAGES
# ============================================================================
required_packages <- c(
  "haven", "readr", "dplyr", "tidyr", "slider", "purrr",
  "broom", "lmtest", "sandwich", "writexl", "tibble"
)

installed <- rownames(installed.packages())
to_install <- setdiff(required_packages, installed)
if (length(to_install) > 0) install.packages(to_install)

invisible(lapply(required_packages, library, character.only = TRUE))

# ============================================================================
# 1. PATHS
# ============================================================================
PATH_MONTHLY <- "C:/Users/najwa/Downloads/germany_monthly_long_total_stock_after_screens_all.dta"
PATH_DAILY   <- "C:/Users/najwa/Downloads/germany_daily_long_total_stock_after_screens_all.dta"
PATH_SAFE    <- "C:/Users/najwa/Downloads/SAFE_Manager_Sentiment_Index_Data_Webpage_202504.csv"
OUT_DIR      <- "C:/Users/najwa/Downloads"

PATH_BASELINE_RDS         <- file.path(OUT_DIR, "baseline_sample_final.rds")
PATH_BASELINE_SENT_RDS    <- file.path(OUT_DIR, "baseline_sample_with_sentiment.rds")
PATH_BASELINE_EDGE_RDS    <- file.path(OUT_DIR, "baseline_sample_with_edge.rds")
PATH_BASELINE_EDGE_ONLYRDS <- file.path(OUT_DIR, "baseline_sample_edge_only.rds")

# ============================================================================
# 2. HELPER FUNCTIONS
# ============================================================================
round_numeric <- function(df, digits = 6) {
  df %>% mutate(across(where(is.numeric), ~ round(.x, digits)))
}

desc_stats <- function(x) {
  n_nonmiss <- sum(!is.na(x))
  tibble(
    Mean = if (n_nonmiss > 0) mean(x, na.rm = TRUE) else NA_real_,
    Median = if (n_nonmiss > 0) median(x, na.rm = TRUE) else NA_real_,
    `Std. Dev.` = if (n_nonmiss > 1) sd(x, na.rm = TRUE) else NA_real_,
    Min = if (n_nonmiss > 0) min(x, na.rm = TRUE) else NA_real_,
    Max = if (n_nonmiss > 0) max(x, na.rm = TRUE) else NA_real_,
    N = n_nonmiss
  )
}

winsor_vec <- function(x, p_low = 0.01, p_high = 0.99, min_non_na = 1) {
  n_nonmiss <- sum(!is.na(x))
  if (n_nonmiss < min_non_na) return(x)
  q_low  <- quantile(x, p_low,  na.rm = TRUE, type = 7)
  q_high <- quantile(x, p_high, na.rm = TRUE, type = 7)
  pmin(pmax(x, q_low), q_high)
}

winsor5 <- function(x) {
  n_nonmiss <- sum(!is.na(x))
  if (n_nonmiss < 20) return(x)
  p5  <- quantile(x, 0.05, na.rm = TRUE, type = 7)
  p95 <- quantile(x, 0.95, na.rm = TRUE, type = 7)
  pmin(pmax(x, p5), p95)
}

ensure_columns <- function(df, cols, fill = NA_real_) {
  for (cl in cols) {
    if (!cl %in% names(df)) df[[cl]] <- fill
  }
  df
}
```

```
nw_mean_tstat <- function(x, lag = 12, min_n = 30) {
  x <- as.numeric(x[!is.na(x)])
  n <- length(x)

  if (n == 0) {
    return(c(mean = NA_real_, se = NA_real_, tstat = NA_real_, n = 0))
  }

  if (n < min_n) {
    return(c(mean = mean(x), se = NA_real_, tstat = NA_real_, n = n))
  }

  fit <- lm(x ~ 1)
  se  <- sqrt(NeweyWest(fit, lag = lag, prewhite = FALSE)[1, 1])
  c(mean = mean(x), se = se, tstat = unname(coef(fit)[1]) / se, n = n)
}

# -----------------------------------------------------------------------------
# BUG FM CORRIGÉ :
# on ne garde que les variables présentes dans la formule
# -----------------------------------------------------------------------------
run_fm <- function(formula_str, data, lag = 12, min_obs_month = 20) {
  form <- as.formula(formula_str)
  vars_needed <- all.vars(form)

  monthly_coefs <- data %>%
    group_by(month_id) %>%
    group_modify(~ {
      d <- .x %>%
        dplyr::select(all_of(vars_needed)) %>%
        na.omit()

      if (nrow(d) < min_obs_month) return(tibble())

      fit <- tryCatch(lm(form, data = d), error = function(e) NULL)
      if (is.null(fit)) return(tibble())

      tibble(
        variable = names(coef(fit)),
        coef = as.numeric(coef(fit))
      )
    }) %>%
    ungroup() %>%
    pivot_wider(names_from = variable, values_from = coef)

  result_list <- lapply(setdiff(names(monthly_coefs), "month_id"), function(v) {
    s <- nw_mean_tstat(monthly_coefs[[v]], lag = lag)
    tibble(
      Variable = v,
      Coefficient = round(s["mean"], 6),
      tstat_NW = round(s["tstat"], 4),
      N_months = s["n"]
    )
  })

  bind_rows(result_list)
}

format_cell <- function(coef, tstat) {
  if (is.na(coef) || is.na(tstat)) return("—")
  stars <- dplyr::case_when(
    abs(tstat) >= 2.576 ~ "***",
    abs(tstat) >= 1.960 ~ "**",
    abs(tstat) >= 1.645 ~ "*",
    TRUE ~ ""
  )
  sprintf("%.4f (%.2f)%s", coef, tstat, stars)
}

make_academic_table <- function(fm_summary, variable_map, row_order) {
  fm_formatted <- fm_summary %>%
    mutate(
      Cell = mapply(format_cell, Coefficient, tstat_NW),
      Variable = dplyr::recode(Variable, !!!as.list(variable_map), .default = Variable)
    )

  n_months_row <- fm_formatted %>%
    group_by(Model) %>%
    summarise(N = first(N_months), .groups = "drop") %>%
    mutate(Variable = "N (months)", Cell = as.character(N)) %>%
    select(Model, Variable, Cell)
```

```
  bind_rows(
    fm_formatted %>% select(Model, Variable, Cell),
    n_months_row
  ) %>%
    pivot_wider(names_from = Model, values_from = Cell, values_fill = "—") %>%
    mutate(Variable = factor(Variable, levels = row_order)) %>%
    arrange(Variable)
}

build_ff_factors <- function(data) {
  factor_input <- data %>%
    group_by(month_id) %>%
    mutate(
      size_median = median(size_lag1, na.rm = TRUE),
      bm_p30 = quantile(bm_lag1, 0.30, na.rm = TRUE, type = 7),
      bm_p70 = quantile(bm_lag1, 0.70, na.rm = TRUE, type = 7),
      size_grp = if_else(size_lag1 <= size_median, "S", "B"),
      bm_grp = case_when(
        bm_lag1 <= bm_p30 ~ "L",
        bm_lag1 >  bm_p70 ~ "H",
        TRUE ~ "M"
      ),
      sb_grp = paste(size_grp, bm_grp, sep = "/")
    ) %>%
    ungroup()

  ff <- factor_input %>%
    filter(!is.na(sb_grp)) %>%
    group_by(month_id, sb_grp) %>%
    summarise(port_ret = mean(ret_lead1, na.rm = TRUE), .groups = "drop") %>%
    pivot_wider(names_from = sb_grp, values_from = port_ret) %>%
    arrange(month_id)

  ff <- ensure_columns(ff, c("S/L", "S/M", "S/H", "B/L", "B/M", "B/H"))

  ff %>%
    mutate(
      SMB = ((`S/L` + `S/M` + `S/H`) / 3) - ((`B/L` + `B/M` + `B/H`) / 3),
      HML = ((`S/H` + `B/H`) / 2) - ((`S/L` + `B/L`) / 2)
    )
}

get_alpha <- function(y, X_df, lag = 12) {
  fit <- lm(y ~ ., data = X_df)
  vc  <- NeweyWest(fit, lag = lag, prewhite = FALSE)
  coefs  <- coef(fit)
  ses    <- sqrt(diag(vc))
  tstats <- coefs / ses

  list(
    alpha = coefs[1],
    alpha_tstat = tstats[1],
    coefs = coefs,
    ses = ses,
    tstats = tstats
  )
}

build_market_states <- function(data) {
  data %>%
    group_by(month_id) %>%
    summarise(
      mkt_ret = mean(mkt_ret, na.rm = TRUE),
      rf = mean(rf, na.rm = TRUE),
      .groups = "drop"
    ) %>%
    arrange(month_id) %>%
    mutate(
      mkt_excess = mkt_ret - rf,
      mkt_ret_lag = lag(mkt_ret),
      mkt_12m_past = slide_dbl(
        mkt_ret_lag,
        ~ if (any(is.na(.x))) NA_real_ else prod(1 + .x) - 1,
        .before = 11,
        .complete = TRUE
      ),
      market_state = case_when(
        is.na(mkt_12m_past) ~ NA_character_,
        mkt_12m_past > 0 ~ "Bull",
        mkt_12m_past <= 0 ~ "Bear",
        TRUE ~ NA_character_
      ),
```

```
      bear_dummy = case_when(
        is.na(mkt_12m_past) ~ NA_integer_,
        mkt_12m_past <= 0 ~ 1L,
        TRUE ~ 0L
      )
    ) %>%
    select(month_id, rf, mkt_ret, mkt_excess, mkt_12m_past, market_state, bear_dummy)
}

write_xlsx_safely <- function(sheets, path) {
  write_xlsx(lapply(sheets, as.data.frame), path = path)
}

# ==============================================================================
# 3. LOAD DATA
# ==============================================================================
monthly <- read_dta(PATH_MONTHLY)
daily   <- read_dta(PATH_DAILY)

cat("\n[INFO] monthly dimensions:\n"); print(dim(monthly))
cat("\n[INFO] daily dimensions:\n"); print(dim(daily))

# ==============================================================================
# 4. BUILD BASELINE PANEL
# ==============================================================================

# ------------------------------------------------------------------------------
# 4.1 Base monthly panel
# ------------------------------------------------------------------------------
panel_monthly_base <- monthly %>%
  transmute(
    date,
    month_id,
    dscd = as.character(dscd),
    stock_id,
    ret = ret_monthly_local_w,
    rf = rf_local,
    ret_excess = ret_monthly_local_w - rf_local,
    size = mv_monthly_local,
    log_size = log(mv_monthly_local),
    mtb = mtb_monthly,
    bm = 1 / mtb_monthly,
    mkt_ret = mkt_local_vw,
    index_ret = index_vw_local,
    price = prc_monthly_local,
    penny_stock,
    exname,
    ecname
  ) %>%
  filter(
    !is.na(dscd),
    !is.na(date),
    !is.na(ret),
    !is.na(size), size > 0,
    !is.na(mtb), mtb > 0,
    penny_stock != 1
  ) %>%
  arrange(dscd, month_id) %>%
  group_by(dscd) %>%
  mutate(
    ret_lead1 = lead(ret),
    ret_excess_lead1 = lead(ret_excess),
    size_lag1 = lag(size),
    log_size_lag1 = lag(log_size),
    bm_lag1 = lag(bm)
  ) %>%
  ungroup()

cat("\n[INFO] panel_monthly_base:\n"); print(dim(panel_monthly_base))

# ------------------------------------------------------------------------------
# 4.2 Monthly liquidity proxies from daily data
# ------------------------------------------------------------------------------
liq_monthly_from_daily <- daily %>%
  filter(penny_stock != 1) %>%
  transmute(
    dscd = as.character(dscd),
    month_id,
    ret_d = ret_daily_local_w,
    ret_d_raw = ret_daily_local,
    dollar_vol_d = dollar_volume_daily_w,
    turnover_d = turnover_daily_w,
```

```
    high_d = prc_high_daily_local,
    low_d = prc_low_daily_local
  ) %>%
  mutate(
    amihud_d = case_when(
      !is.na(ret_d) & !is.na(dollar_vol_d) & dollar_vol_d > 0 ~ abs(ret_d) / dollar_vol_d,
      TRUE ~ NA_real_
    ),
    zero_return_d = case_when(
      is.na(ret_d_raw) ~ NA_real_,
      ret_d_raw == 0 ~ 1,
      TRUE ~ 0
    ),
    hl_range_d = case_when(
      !is.na(high_d) & !is.na(low_d) & (high_d + low_d) > 0 ~
        (high_d - low_d) / ((high_d + low_d) / 2),
      TRUE ~ NA_real_
    )
  ) %>%
  group_by(dscd, month_id) %>%
  summarise(
    n_amihud_days   = sum(!is.na(amihud_d)),
    n_zero_days     = sum(!is.na(zero_return_d)),
    n_hl_days       = sum(!is.na(hl_range_d)),
    n_turnover_days = sum(!is.na(turnover_d)),
    amihud_m      = if_else(n_amihud_days   >= 15, mean(amihud_d, na.rm = TRUE), NA_real_),
    zero_return_m = if_else(n_zero_days     >= 15, mean(zero_return_d, na.rm = TRUE), NA_real_),
    hl_range_m    = if_else(n_hl_days       >= 15, mean(hl_range_d, na.rm = TRUE), NA_real_),
    turnover_m    = if_else(n_turnover_days >= 15, mean(turnover_d, na.rm = TRUE), NA_real_),
    .groups = "drop"
  )

cat("\n[INFO] liq_monthly_from_daily:\n"); print(dim(liq_monthly_from_daily))

# -----------------------------------------------------------------------------
# 4.3 Merge + baseline winsorisation (Amihud only)
# -----------------------------------------------------------------------------
panel_main <- panel_monthly_base %>%
  left_join(liq_monthly_from_daily, by = c("dscd", "month_id")) %>%
  arrange(dscd, month_id)

panel_main <- panel_main %>%
  group_by(month_id) %>%
  mutate(
    amihud_m_w1 = winsor_vec(amihud_m, p_low = 0.01, p_high = 0.99, min_non_na = 1)
  ) %>%
  ungroup() %>%
  arrange(dscd, month_id) %>%
  group_by(dscd) %>%
  mutate(
    amihud_lag1      = lag(amihud_m_w1),
    zero_return_lag1 = lag(zero_return_m),
    hl_range_lag1    = lag(hl_range_m),
    turnover_lag1    = lag(turnover_m)
  ) %>%
  ungroup()

cat("\n[INFO] panel_main:\n"); print(dim(panel_main))

# -----------------------------------------------------------------------------
# 4.4 Baseline sample
# -----------------------------------------------------------------------------
baseline_sample <- panel_main %>%
  filter(
    !is.na(ret_lead1),
    !is.na(ret_excess_lead1),
    !is.na(size_lag1), size_lag1 > 0,
    !is.na(log_size_lag1),
    !is.na(bm_lag1), bm_lag1 > 0,
    !is.na(amihud_lag1)
  )

cat("\n[INFO] baseline_sample:\n"); print(dim(baseline_sample))

baseline_sample %>%
  summarise(
    n_obs = n(),
    n_firms = n_distinct(dscd),
    n_months = n_distinct(month_id),
    first_month = min(month_id, na.rm = TRUE),
    last_month = max(month_id, na.rm = TRUE)
  ) %>% print()
```

```
saveRDS(baseline_sample, PATH_BASELINE_RDS)

# ==============================================================================
# 5. TABLES 1 À 6 — DESCRIPTIFS BASELINE
# ==============================================================================
panel_liq_desc <- panel_main %>%
  group_by(month_id) %>%
  mutate(
    amihud_w = winsor_vec(amihud_m, p_low = 0.01, p_high = 0.99, min_non_na = 1)
  ) %>%
  ungroup()

table_raw_liq <- bind_rows(
  desc_stats(liq_monthly_from_daily$amihud_m)      %>% mutate(Proxy = "Amihud"),
  desc_stats(liq_monthly_from_daily$zero_return_m) %>% mutate(Proxy = "Zero-return"),
  desc_stats(liq_monthly_from_daily$hl_range_m)    %>% mutate(Proxy = "High-low range"),
  desc_stats(liq_monthly_from_daily$turnover_m)    %>% mutate(Proxy = "Turnover")
) %>%
  select(Proxy, Mean, Median, `Std. Dev.`, Min, Max, N) %>%
  round_numeric()

table_before_winsor <- bind_rows(
  desc_stats(panel_main$amihud_m)      %>% mutate(Proxy = "Amihud"),
  desc_stats(panel_main$zero_return_m) %>% mutate(Proxy = "Zero-return"),
  desc_stats(panel_main$hl_range_m)    %>% mutate(Proxy = "High-low range"),
  desc_stats(panel_main$turnover_m)    %>% mutate(Proxy = "Turnover")
) %>%
  select(Proxy, Mean, Median, `Std. Dev.`, Min, Max, N) %>%
  round_numeric()

table_after_winsor <- bind_rows(
  desc_stats(panel_liq_desc$amihud_w)      %>% mutate(Proxy = "Amihud"),
  desc_stats(panel_liq_desc$zero_return_m) %>% mutate(Proxy = "Zero-return"),
  desc_stats(panel_liq_desc$hl_range_m)    %>% mutate(Proxy = "High-low range"),
  desc_stats(panel_liq_desc$turnover_m)    %>% mutate(Proxy = "Turnover")
) %>%
  select(Proxy, Mean, Median, `Std. Dev.`, Min, Max, N) %>%
  round_numeric()

panel_liq_desc <- panel_liq_desc %>%
  arrange(dscd, month_id) %>%
  group_by(dscd) %>%
  mutate(
    amihud_3m      = slide_dbl(amihud_w,      ~ mean(.x, na.rm = TRUE), .before = 2, .complete = TRUE),
    zero_return_3m = slide_dbl(zero_return_m, ~ mean(.x, na.rm = TRUE), .before = 2, .complete = TRUE),
    hl_range_3m    = slide_dbl(hl_range_m,    ~ mean(.x, na.rm = TRUE), .before = 2, .complete = TRUE),
    turnover_3m    = slide_dbl(turnover_m,    ~ mean(.x, na.rm = TRUE), .before = 2, .complete = TRUE)
  ) %>%
  ungroup()

table_3m <- bind_rows(
  desc_stats(panel_liq_desc$amihud_3m)      %>% mutate(Proxy = "Amihud (3-month MA)"),
  desc_stats(panel_liq_desc$zero_return_3m) %>% mutate(Proxy = "Zero-return (3-month MA)"),
  desc_stats(panel_liq_desc$hl_range_3m)    %>% mutate(Proxy = "High-low range (3-month MA)"),
  desc_stats(panel_liq_desc$turnover_3m)    %>% mutate(Proxy = "Turnover (3-month MA)")
) %>%
  select(Proxy, Mean, Median, `Std. Dev.`, Min, Max, N) %>%
  round_numeric()

table_descriptive_baseline <- bind_rows(
  ret_lead1        = desc_stats(baseline_sample$ret_lead1),
  ret_excess_lead1 = desc_stats(baseline_sample$ret_excess_lead1),
  size_lag1        = desc_stats(baseline_sample$size_lag1),
  log_size_lag1    = desc_stats(baseline_sample$log_size_lag1),
  bm_lag1          = desc_stats(baseline_sample$bm_lag1),
  amihud_lag1      = desc_stats(baseline_sample$amihud_lag1),
  zero_return_lag1 = desc_stats(baseline_sample$zero_return_lag1),
  hl_range_lag1    = desc_stats(baseline_sample$hl_range_lag1),
  turnover_lag1    = desc_stats(baseline_sample$turnover_lag1),
  .id = "Variable"
) %>%
  round_numeric()

size_comparison_sample <- baseline_sample %>%
  group_by(month_id) %>%
  mutate(
    size_median_month = median(size_lag1, na.rm = TRUE),
    size_group = case_when(
      size_lag1 <= size_median_month ~ "Small",
      size_lag1 > size_median_month ~ "Big",
      TRUE ~ NA_character_
```

```
    )
  ) %>%
  ungroup()

table_liquidity_by_size <- size_comparison_sample %>%
  group_by(size_group) %>%
  summarise(
    mean_amihud        = mean(amihud_lag1, na.rm = TRUE),
    median_amihud      = median(amihud_lag1, na.rm = TRUE),
    mean_zero_return   = mean(zero_return_lag1, na.rm = TRUE),
    median_zero_return = median(zero_return_lag1, na.rm = TRUE),
    mean_hl_range      = mean(hl_range_lag1, na.rm = TRUE),
    median_hl_range    = median(hl_range_lag1, na.rm = TRUE),
    mean_turnover      = mean(turnover_lag1, na.rm = TRUE),
    median_turnover    = median(turnover_lag1, na.rm = TRUE),
    n_obs = n(),
    .groups = "drop"
  ) %>%
  round_numeric()

write_xlsx_safely(
  list(
    "Table1_Raw_Construction" = table_raw_liq,
    "Table2_Before_Winsor" = table_before_winsor,
    "Table3_After_Winsor" = table_after_winsor,
    "Table4_Moving_Averages_3M" = table_3m,
    "Table5_Baseline_Sample" = table_descriptive_baseline,
    "Table6_Liquidity_by_Size" = table_liquidity_by_size
  ),
  file.path(OUT_DIR, "Tables_1_to_6_Baseline.xlsx")
)

# ==============================================================================
# 6. TABLES 7 À 9 — TRI UNIVARIÉ AMIHUD + CAPM / FF3
# ==============================================================================
portfolios_amihud <- baseline_sample %>%
  group_by(month_id) %>%
  mutate(
    q20 = quantile(amihud_lag1, 0.20, na.rm = TRUE, type = 7),
    q40 = quantile(amihud_lag1, 0.40, na.rm = TRUE, type = 7),
    q60 = quantile(amihud_lag1, 0.60, na.rm = TRUE, type = 7),
    q80 = quantile(amihud_lag1, 0.80, na.rm = TRUE, type = 7),
    quintile = case_when(
      amihud_lag1 <= q20 ~ 1,
      amihud_lag1 <= q40 ~ 2,
      amihud_lag1 <= q60 ~ 3,
      amihud_lag1 <= q80 ~ 4,
      amihud_lag1 >  q80 ~ 5,
      TRUE ~ NA_real_
    )
  ) %>%
  ungroup()

portfolio_returns <- portfolios_amihud %>%
  filter(!is.na(quintile)) %>%
  group_by(month_id, quintile) %>%
  summarise(
    port_ret = mean(ret_lead1, na.rm = TRUE),
    port_ret_excess = mean(ret_excess_lead1, na.rm = TRUE),
    rf_avg = mean(rf, na.rm = TRUE),
    mkt_ret_avg = mean(mkt_ret, na.rm = TRUE),
    n_stocks = n(),
    .groups = "drop"
  )

portfolio_returns_wide_excess <- portfolio_returns %>%
  select(month_id, quintile, port_ret_excess) %>%
  pivot_wider(names_from = quintile, values_from = port_ret_excess, names_prefix = "Q") %>%
  arrange(month_id)

portfolio_returns_wide_excess <- ensure_columns(portfolio_returns_wide_excess, paste0("Q", 1:5)) %>%
  mutate(Q5_minus_Q1 = Q5 - Q1)

factors_monthly <- baseline_sample %>%
  group_by(month_id) %>%
  summarise(
    rf = mean(rf, na.rm = TRUE),
    mkt_ret = mean(mkt_ret, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(mkt_excess = mkt_ret - rf) %>%
  arrange(month_id)
```

```
ff_portfolios <- build_ff_factors(baseline_sample)

ts_data <- portfolio_returns_wide_excess %>%
  left_join(factors_monthly, by = "month_id") %>%
  left_join(ff_portfolios %>% select(month_id, SMB, HML), by = "month_id") %>%
  filter(!is.na(mkt_excess), !is.na(SMB), !is.na(HML))

results_excess <- bind_rows(
  Q1 = nw_mean_tstat(ts_data$Q1),
  Q2 = nw_mean_tstat(ts_data$Q2),
  Q3 = nw_mean_tstat(ts_data$Q3),
  Q4 = nw_mean_tstat(ts_data$Q4),
  Q5 = nw_mean_tstat(ts_data$Q5),
  `Q5 - Q1` = nw_mean_tstat(ts_data$Q5_minus_Q1),
  .id = "Portfolio"
) %>%
  round_numeric()

capm_results <- list()
for (port in c("Q1", "Q2", "Q3", "Q4", "Q5", "Q5_minus_Q1")) {
  out <- get_alpha(ts_data[[port]], ts_data %>% select(mkt_excess))
  capm_results[[port]] <- tibble(
    Portfolio = port,
    alpha = out$alpha,
    alpha_tstat = out$alpha_tstat,
    beta_mkt = out$coefs[2],
    beta_mkt_tstat = out$tstats[2]
  )
}
capm_table <- bind_rows(capm_results) %>% round_numeric()

ff3_results <- list()
for (port in c("Q1", "Q2", "Q3", "Q4", "Q5", "Q5_minus_Q1")) {
  out <- get_alpha(ts_data[[port]], ts_data %>% select(mkt_excess, SMB, HML))
  ff3_results[[port]] <- tibble(
    Portfolio = port,
    alpha = out$alpha,
    alpha_tstat = out$alpha_tstat,
    beta_mkt = out$coefs[2],
    beta_smb = out$coefs[3],
    beta_hml = out$coefs[4],
    beta_mkt_tstat = out$tstats[2],
    beta_smb_tstat = out$tstats[3],
    beta_hml_tstat = out$tstats[4]
  )
}
ff3_table <- bind_rows(ff3_results) %>% round_numeric()

write_xlsx_safely(
  list(
    "Table7_MeanExcessReturns" = results_excess,
    "Table8_CAPM_Alphas" = capm_table,
    "Table9_FF3_Alphas" = ff3_table,
    "Portfolio_TS" = ts_data,
    "FF_Factors_TS" = ff_portfolios
  ),
  file.path(OUT_DIR, "Tables_7_to_9_Univariate_Amihud.xlsx")
)

# ==============================================================================
# 7. TABLES 10 À 13 — BIVARIATE SEQUENTIAL SORTS Size × Amihud (5x5)
# ==============================================================================
biv_sample <- baseline_sample %>%
  group_by(month_id) %>%
  mutate(
    s_q20 = quantile(size_lag1, 0.20, na.rm = TRUE, type = 7),
    s_q40 = quantile(size_lag1, 0.40, na.rm = TRUE, type = 7),
    s_q60 = quantile(size_lag1, 0.60, na.rm = TRUE, type = 7),
    s_q80 = quantile(size_lag1, 0.80, na.rm = TRUE, type = 7),
    size_quintile = case_when(
      size_lag1 <= s_q20 ~ 1,
      size_lag1 <= s_q40 ~ 2,
      size_lag1 <= s_q60 ~ 3,
      size_lag1 <= s_q80 ~ 4,
      size_lag1 >  s_q80 ~ 5,
      TRUE ~ NA_real_
    )
  ) %>%
  ungroup() %>%
  group_by(month_id, size_quintile) %>%
  mutate(
```

```
    a_q20 = quantile(amihud_lag1, 0.20, na.rm = TRUE, type = 7),
    a_q40 = quantile(amihud_lag1, 0.40, na.rm = TRUE, type = 7),
    a_q60 = quantile(amihud_lag1, 0.60, na.rm = TRUE, type = 7),
    a_q80 = quantile(amihud_lag1, 0.80, na.rm = TRUE, type = 7),
    amihud_quintile = case_when(
      amihud_lag1 <= a_q20 ~ 1,
      amihud_lag1 <= a_q40 ~ 2,
      amihud_lag1 <= a_q60 ~ 3,
      amihud_lag1 <= a_q80 ~ 4,
      amihud_lag1 >  a_q80 ~ 5,
      TRUE ~ NA_real_
    )
  ) %>%
  ungroup()

port_25 <- biv_sample %>%
  filter(!is.na(size_quintile), !is.na(amihud_quintile)) %>%
  group_by(month_id, size_quintile, amihud_quintile) %>%
  summarise(
    port_ret = mean(ret_lead1, na.rm = TRUE),
    port_ret_excess = mean(ret_excess_lead1, na.rm = TRUE),
    n_stocks = n(),
    .groups = "drop"
  )

fill_check <- port_25 %>%
  group_by(size_quintile, amihud_quintile) %>%
  summarise(
    n_months = n(),
    avg_n_stocks = mean(n_stocks),
    .groups = "drop"
  ) %>%
  round_numeric(2)

mean_matrix <- port_25 %>%
  group_by(size_quintile, amihud_quintile) %>%
  summarise(mean_excess = mean(port_ret_excess, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = amihud_quintile, values_from = mean_excess, names_prefix = "A") %>%
  round_numeric()

spread_within_size <- port_25 %>%
  select(month_id, size_quintile, amihud_quintile, port_ret_excess) %>%
  pivot_wider(names_from = amihud_quintile, values_from = port_ret_excess, names_prefix = "A") %>%
  arrange(month_id)

spread_within_size <- ensure_columns(spread_within_size, paste0("A", 1:5)) %>%
  mutate(A5_minus_A1 = A5 - A1)

spread_results_fixed <- spread_within_size %>%
  group_by(size_quintile) %>%
  group_modify(~ {
    s <- nw_mean_tstat(.x$A5_minus_A1, lag = 12)
    tibble(
      mean_A1 = round(mean(.x$A1, na.rm = TRUE), 6),
      mean_A5 = round(mean(.x$A5, na.rm = TRUE), 6),
      mean_spread = round(s["mean"], 6),
      tstat_NW = round(s["tstat"], 4),
      n_months = s["n"]
    )
  }) %>%
  ungroup()

size_neutral_spread <- spread_within_size %>%
  group_by(month_id) %>%
  summarise(avg_spread = mean(A5_minus_A1, na.rm = TRUE), .groups = "drop")

sn_summary <- nw_mean_tstat(size_neutral_spread$avg_spread, lag = 12)

sn_data <- size_neutral_spread %>%
  left_join(factors_monthly, by = "month_id") %>%
  left_join(ff_portfolios %>% select(month_id, SMB, HML), by = "month_id") %>%
  filter(!is.na(mkt_excess), !is.na(SMB), !is.na(HML))

capm_sn <- get_alpha(sn_data$avg_spread, sn_data %>% select(mkt_excess))
ff3_sn  <- get_alpha(sn_data$avg_spread, sn_data %>% select(mkt_excess, SMB, HML))

sn_alpha_table <- tibble(
  Model = c("CAPM", "Fama-French 3F"),
  alpha = c(capm_sn$coefs[1], ff3_sn$coefs[1]),
  alpha_tstat = c(capm_sn$tstats[1], ff3_sn$tstats[1]),
  beta_mkt = c(capm_sn$coefs[2], ff3_sn$coefs[2]),
  beta_mkt_tstat = c(capm_sn$tstats[2], ff3_sn$tstats[2]),
```

```
  beta_smb = c(NA, ff3_sn$coefs[3]),
  beta_smb_tstat = c(NA, ff3_sn$tstats[3]),
  beta_hml = c(NA, ff3_sn$coefs[4]),
  beta_hml_tstat = c(NA, ff3_sn$tstats[4])
) %>%
  round_numeric()

table_12_size_neutral <- tibble(
  Mean = round(sn_summary["mean"], 6),
  tstat_NW = round(sn_summary["tstat"], 4),
  N_months = sn_summary["n"]
)

write_xlsx_safely(
  list(
    "Table10_Mean_Returns_5x5" = mean_matrix,
    "Table11_Spread_by_Size" = spread_results_fixed,
    "Table12_SizeNeutral_Premium" = table_12_size_neutral,
    "Table12_SizeNeutral_TS" = size_neutral_spread,
    "Table13_SN_Alphas" = sn_alpha_table,
    "Fill_Check" = fill_check
  ),
  file.path(OUT_DIR, "Tables_10_to_13_Bivariate_Amihud.xlsx")
)

# ==============================================================================
# 8. TABLES 14 À 16 — BULL VS BEAR MARKETS
# ==============================================================================
market_ts <- build_market_states(baseline_sample)

portfolio_wide <- portfolio_returns %>%
  select(month_id, quintile, port_ret_excess) %>%
  pivot_wider(names_from = quintile, values_from = port_ret_excess, names_prefix = "Q") %>%
  arrange(month_id)

portfolio_wide <- ensure_columns(portfolio_wide, paste0("Q", 1:5)) %>%
  mutate(Q5_minus_Q1 = Q5 - Q1)

ts_state <- portfolio_wide %>%
  left_join(market_ts %>% select(month_id, market_state, mkt_excess), by = "month_id") %>%
  filter(!is.na(market_state))

table_state <- ts_state %>%
  select(market_state, Q1, Q2, Q3, Q4, Q5, Q5_minus_Q1) %>%
  pivot_longer(cols = -market_state, names_to = "Portfolio", values_to = "ret") %>%
  group_by(market_state, Portfolio) %>%
  group_modify(~ {
    s <- nw_mean_tstat(.x$ret, lag = 12)
    tibble(
      mean_ret = round(s["mean"], 6),
      tstat_NW = round(s["tstat"], 4),
      n_months = s["n"]
    )
  }) %>%
  ungroup()

stat_bull <- nw_mean_tstat(ts_state %>% filter(market_state == "Bull") %>% pull(Q5_minus_Q1), lag = 12)
stat_bear <- nw_mean_tstat(ts_state %>% filter(market_state == "Bear") %>% pull(Q5_minus_Q1), lag = 12)

ts_state_diff <- ts_state %>%
  mutate(bear_dummy = if_else(market_state == "Bear", 1, 0))

diff_fit <- lm(Q5_minus_Q1 ~ bear_dummy, data = ts_state_diff)
diff_vc  <- NeweyWest(diff_fit, lag = 12, prewhite = FALSE)
diff_tstats <- coef(diff_fit) / sqrt(diag(diff_vc))

table_diff <- tibble(
  State = c("Bull", "Bear", "Bear - Bull (diff)"),
  Mean_Spread = c(
    round(stat_bull["mean"], 6),
    round(stat_bear["mean"], 6),
    round(coef(diff_fit)[2], 6)
  ),
  tstat_NW = c(
    round(stat_bull["tstat"], 4),
    round(stat_bear["tstat"], 4),
    round(diff_tstats[2], 4)
  ),
  N_months = c(
    stat_bull["n"],
    stat_bear["n"],
    stat_bull["n"] + stat_bear["n"]
```

```
  )
)

ts_full <- ts_state %>%
  left_join(ff_portfolios %>% select(month_id, SMB, HML), by = "month_id") %>%
  filter(!is.na(SMB), !is.na(HML))

run_alpha_state <- function(state_label) {
  d <- ts_full %>% filter(market_state == state_label)

  fit_capm <- lm(Q5_minus_Q1 ~ mkt_excess, data = d)
  vc_capm  <- NeweyWest(fit_capm, lag = 12, prewhite = FALSE)

  fit_ff3 <- lm(Q5_minus_Q1 ~ mkt_excess + SMB + HML, data = d)
  vc_ff3  <- NeweyWest(fit_ff3, lag = 12, prewhite = FALSE)

  tibble(
    State = state_label,
    CAPM_alpha = round(coef(fit_capm)[1], 6),
    CAPM_tstat = round(coef(fit_capm)[1] / sqrt(vc_capm[1, 1]), 4),
    FF3_alpha = round(coef(fit_ff3)[1], 6),
    FF3_tstat = round(coef(fit_ff3)[1] / sqrt(vc_ff3[1, 1]), 4),
    FF3_beta_smb = round(coef(fit_ff3)[3], 4),
    FF3_smb_tstat = round(coef(fit_ff3)[3] / sqrt(vc_ff3[3, 3]), 4),
    N = nrow(d)
  )
}

table_alphas_state <- bind_rows(
  run_alpha_state("Bull"),
  run_alpha_state("Bear")
)

write_xlsx_safely(
  list(
    "Table14_Returns_by_State" = table_state,
    "Table15_Bull_vs_Bear_Diff" = table_diff,
    "Table16_Alphas_by_State" = table_alphas_state,
    "MarketStates_TS" = market_ts,
    "Portfolio_State_TS" = ts_full
  ),
  file.path(OUT_DIR, "Tables_14_to_16_BullBear.xlsx")
)

# ==============================================================================
# 9. TABLE 17 — FAMA-MACBETH BASELINE
# ==============================================================================
fm_data <- baseline_sample %>%
  left_join(market_ts %>% select(month_id, bear_dummy), by = "month_id") %>%
  mutate(
    log_amihud_lag1 = log(amihud_lag1 + 1e-8)
  ) %>%
  select(
    month_id, dscd, ret_lead1,
    amihud_lag1, log_amihud_lag1,
    log_size_lag1, bm_lag1, bear_dummy
  ) %>%
  filter(
    !is.na(amihud_lag1),
    !is.na(log_size_lag1),
    !is.na(bm_lag1)
  )

fm1 <- run_fm("ret_lead1 ~ amihud_lag1", fm_data)
fm2 <- run_fm("ret_lead1 ~ amihud_lag1 + log_size_lag1", fm_data)
fm3 <- run_fm("ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1", fm_data)
fm4 <- run_fm("ret_lead1 ~ log_amihud_lag1 + log_size_lag1 + bm_lag1", fm_data)

fm_data_state <- fm_data %>% filter(!is.na(bear_dummy))

fm5_bull <- run_fm(
  "ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_state %>% filter(bear_dummy == 0)
)

fm5_bear <- run_fm(
  "ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_state %>% filter(bear_dummy == 1)
)

fm_summary <- bind_rows(
  fm1 %>% mutate(Model = "(1) Amihud only"),
```

```
  fm2 %>% mutate(Model = "(2) Amihud + Size"),
  fm3 %>% mutate(Model = "(3) Baseline FM"),
  fm4 %>% mutate(Model = "(4) log(Amihud)"),
  fm5_bull %>% mutate(Model = "(5a) Bull markets only"),
  fm5_bear %>% mutate(Model = "(5b) Bear markets only")
) %>%
  select(Model, Variable, Coefficient, tstat_NW, N_months)

table17_academic <- make_academic_table(
  fm_summary = fm_summary,
  variable_map = c(
    "(Intercept)" = "Intercept",
    "amihud_lag1" = "Amihud",
    "log_amihud_lag1" = "log(Amihud)",
    "log_size_lag1" = "log(Size)",
    "bm_lag1" = "B/M"
  ),
  row_order = c("Intercept", "Amihud", "log(Amihud)", "log(Size)", "B/M", "N (months)")
)

write_xlsx_safely(
  list(
    "Table17_FM_Summary" = fm_summary,
    "Table17_Academic" = table17_academic,
    "Model1_Amihud_Only" = fm1,
    "Model2_Amihud_Size" = fm2,
    "Model3_Baseline" = fm3,
    "Model4_LogAmihud" = fm4,
    "Model5a_Bull" = fm5_bull,
    "Model5b_Bear" = fm5_bear
  ),
  file.path(OUT_DIR, "Table_17_FamaMacBeth.xlsx")
)

cat("\n============================\n")
cat("BLOC 1/3 TERMINÉ\n")
cat("============================\n")
# ============================================================================
# BLOC 2/3 — SAFE MANAGER SENTIMENT + TABLES 18 TO 21
# ============================================================================

# Vérification baseline
if (!file.exists(PATH_BASELINE_RDS)) {
  stop("Le fichier baseline_sample_final.rds est introuvable. Exécute d'abord le BLOC 1.")
}

baseline_sample <- readRDS(PATH_BASELINE_RDS)

# ============================================================================
# 10. SAFE MANAGER SENTIMENT
# ============================================================================
sentiment_safe <- read_csv(PATH_SAFE, show_col_types = FALSE) %>%
  rename(sentiment = SAFE_Manager_Sentiment) %>%
  mutate(
    year_safe  = as.integer(sub("m.*", "", YYYYMM)),
    month_safe = as.integer(sub(".*m", "", YYYYMM)),
    date_safe  = as.Date(sprintf("%04d-%02d-01", year_safe, month_safe)),
    year_month = sprintf("%04d-%02d", year_safe, month_safe)
  ) %>%
  filter(!is.na(sentiment))

month_mapping <- monthly %>%
  transmute(
    month_id,
    year_month = format(as.Date(date), "%Y-%m")
  ) %>%
  distinct(month_id, year_month) %>%
  arrange(month_id)

sentiment_panel <- sentiment_safe %>%
  inner_join(month_mapping, by = "year_month") %>%
  arrange(month_id) %>%
  select(month_id, year_month, date_safe, sentiment)

if (nrow(sentiment_panel) == 0) {
  stop("Aucun match entre SAFE et month_id. Vérifie le parsing de year_month.")
}

sent_median_val <- median(sentiment_panel$sentiment, na.rm = TRUE)
sent_p33_val   <- quantile(sentiment_panel$sentiment, 0.33, na.rm = TRUE)
sent_p67_val   <- quantile(sentiment_panel$sentiment, 0.67, na.rm = TRUE)
```

```
sentiment_panel <- sentiment_panel %>%
  mutate(
    sent_state_2 = if_else(sentiment > sent_median_val, "High", "Low"),
    sent_state_3 = case_when(
      sentiment <= sent_p33_val ~ "Low",
      sentiment >= sent_p67_val ~ "High",
      TRUE ~ "Medium"
    ),
    high_sent_dummy = if_else(sent_state_2 == "High", 1L, 0L)
  )

baseline_sample_sent_only <- baseline_sample %>%
  left_join(
    sentiment_panel %>% select(month_id, sentiment, sent_state_2, sent_state_3, high_sent_dummy),
    by = "month_id"
  ) %>%
  filter(!is.na(sentiment), !is.na(sent_state_2))

saveRDS(baseline_sample_sent_only, PATH_BASELINE_SENT_RDS)

cat("\n[INFO] baseline_sample_sent_only:\n")
baseline_sample_sent_only %>%
  summarise(
    n_obs = n(),
    n_firms = n_distinct(dscd),
    n_months = n_distinct(month_id),
    first_month = min(month_id),
    last_month = max(month_id)
  ) %>% print()

# ---------------------------------------------------------------------------
# Table 18 — Mean excess returns by sentiment state
# ---------------------------------------------------------------------------
portfolios_sent <- baseline_sample_sent_only %>%
  group_by(month_id) %>%
  mutate(
    q20 = quantile(amihud_lag1, 0.20, na.rm = TRUE, type = 7),
    q40 = quantile(amihud_lag1, 0.40, na.rm = TRUE, type = 7),
    q60 = quantile(amihud_lag1, 0.60, na.rm = TRUE, type = 7),
    q80 = quantile(amihud_lag1, 0.80, na.rm = TRUE, type = 7),
    quintile = case_when(
      amihud_lag1 <= q20 ~ 1,
      amihud_lag1 <= q40 ~ 2,
      amihud_lag1 <= q60 ~ 3,
      amihud_lag1 <= q80 ~ 4,
      amihud_lag1 > q80 ~ 5,
      TRUE ~ NA_real_
    )
  ) %>%
  ungroup()

portfolio_returns_sent <- portfolios_sent %>%
  filter(!is.na(quintile)) %>%
  group_by(month_id, quintile) %>%
  summarise(
    port_ret_excess = mean(ret_excess_lead1, na.rm = TRUE),
    n_stocks = n(),
    .groups = "drop"
  )

port_wide_sent <- portfolio_returns_sent %>%
  select(month_id, quintile, port_ret_excess) %>%
  pivot_wider(names_from = quintile, values_from = port_ret_excess, names_prefix = "Q") %>%
  arrange(month_id)

port_wide_sent <- ensure_columns(port_wide_sent, paste0("Q", 1:5)) %>%
  mutate(Q5_minus_Q1 = Q5 - Q1)

ts_sent <- port_wide_sent %>%
  left_join(sentiment_panel %>% select(month_id, sent_state_2, sent_state_3), by = "month_id") %>%
  filter(!is.na(sent_state_2))

table_18 <- ts_sent %>%
  select(sent_state_2, Q1, Q2, Q3, Q4, Q5, Q5_minus_Q1) %>%
  pivot_longer(cols = -sent_state_2, names_to = "Portfolio", values_to = "ret") %>%
  group_by(sent_state_2, Portfolio) %>%
  group_modify(~ {
    s <- nw_mean_tstat(.x$ret, lag = 12)
    tibble(
      mean_ret = round(s["mean"], 6),
      tstat_NW = round(s["tstat"], 4),
      n_months = s["n"]
```

```
  )
 }) %>%
 ungroup() %>%
 mutate(
   Portfolio = factor(Portfolio, levels = c("Q1", "Q2", "Q3", "Q4", "Q5", "Q5_minus_Q1"))
 ) %>%
 arrange(sent_state_2, Portfolio)

# ----------------------------------------------------------------------------
# Table 19 — Spread High vs Low sentiment
# ----------------------------------------------------------------------------
stat_high <- nw_mean_tstat(
 ts_sent %>% filter(sent_state_2 == "High") %>% pull(Q5_minus_Q1),
 lag = 12
)

stat_low <- nw_mean_tstat(
 ts_sent %>% filter(sent_state_2 == "Low") %>% pull(Q5_minus_Q1),
 lag = 12
)

ts_sent_diff <- ts_sent %>%
 mutate(high_dummy = if_else(sent_state_2 == "High", 1, 0))

diff_fit <- lm(Q5_minus_Q1 ~ high_dummy, data = ts_sent_diff)
diff_vc  <- NeweyWest(diff_fit, lag = 12, prewhite = FALSE)
diff_tstats <- coef(diff_fit) / sqrt(diag(diff_vc))

table_19 <- tibble(
 State = c("Low Sentiment", "High Sentiment", "High - Low (diff)"),
 Mean_Spread = c(
   round(stat_low["mean"], 6),
   round(stat_high["mean"], 6),
   round(coef(diff_fit)[2], 6)
 ),
 tstat_NW = c(
   round(stat_low["tstat"], 4),
   round(stat_high["tstat"], 4),
   round(diff_tstats[2], 4)
 ),
 N_months = c(
   stat_low["n"],
   stat_high["n"],
   stat_low["n"] + stat_high["n"]
 )
)

# ----------------------------------------------------------------------------
# Table 20 — Conditional alphas by sentiment state
# ----------------------------------------------------------------------------
factors_monthly_sent <- baseline_sample_sent_only %>%
 group_by(month_id) %>%
 summarise(
   rf = mean(rf, na.rm = TRUE),
   mkt_ret = mean(mkt_ret, na.rm = TRUE),
   .groups = "drop"
 ) %>%
 mutate(mkt_excess = mkt_ret - rf)

ff_portfolios_sent <- build_ff_factors(baseline_sample_sent_only)

ts_full_sent <- ts_sent %>%
 left_join(factors_monthly_sent, by = "month_id") %>%
 left_join(ff_portfolios_sent %>% select(month_id, SMB, HML), by = "month_id") %>%
 filter(!is.na(mkt_excess), !is.na(SMB), !is.na(HML))

run_alpha_sent <- function(state_label) {
 d <- ts_full_sent %>% filter(sent_state_2 == state_label)

 fit_capm <- lm(Q5_minus_Q1 ~ mkt_excess, data = d)
 vc_capm  <- NeweyWest(fit_capm, lag = 12, prewhite = FALSE)

 fit_ff3 <- lm(Q5_minus_Q1 ~ mkt_excess + SMB + HML, data = d)
 vc_ff3  <- NeweyWest(fit_ff3, lag = 12, prewhite = FALSE)

 tibble(
   Sentiment_State = state_label,
   CAPM_alpha = round(coef(fit_capm)[1], 6),
   CAPM_tstat = round(coef(fit_capm)[1] / sqrt(vc_capm[1, 1]), 4),
   FF3_alpha = round(coef(fit_ff3)[1], 6),
   FF3_tstat = round(coef(fit_ff3)[1] / sqrt(vc_ff3[1, 1]), 4),
   FF3_beta_smb = round(coef(fit_ff3)[3], 4),
```

```r
    FF3_smb_tstat = round(coef(fit_ff3)[3] / sqrt(vc_ff3[3, 3]), 4),
    FF3_beta_hml = round(coef(fit_ff3)[4], 4),
    FF3_hml_tstat = round(coef(fit_ff3)[4] / sqrt(vc_ff3[4, 4]), 4),
    N_months = nrow(d)
  )
}

table_20 <- bind_rows(
  run_alpha_sent("Low"),
  run_alpha_sent("High")
)

# ---------------------------------------------------------------------------
# Table 21 — Fama-MacBeth conditioned on sentiment
# ---------------------------------------------------------------------------
fm_data_sent <- baseline_sample_sent_only %>%
  mutate(
    log_amihud_lag1 = log(amihud_lag1 + 1e-8)
  ) %>%
  select(
    month_id, dscd, ret_lead1,
    amihud_lag1, log_amihud_lag1,
    log_size_lag1, bm_lag1,
    sentiment, high_sent_dummy
  ) %>%
  filter(
    !is.na(amihud_lag1),
    !is.na(log_size_lag1),
    !is.na(bm_lag1),
    !is.na(sentiment)
  )

fm_s1 <- run_fm(
  "ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_sent
)

fm_s2 <- run_fm(
  "ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_sent %>% filter(high_sent_dummy == 1)
)

fm_s3 <- run_fm(
  "ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_sent %>% filter(high_sent_dummy == 0)
)

fm_s4 <- run_fm(
  "ret_lead1 ~ log_amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_sent
)

fm_s5 <- run_fm(
  "ret_lead1 ~ log_amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_sent %>% filter(high_sent_dummy == 1)
)

fm_s6 <- run_fm(
  "ret_lead1 ~ log_amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_sent %>% filter(high_sent_dummy == 0)
)

table_21 <- bind_rows(
  fm_s1 %>% mutate(Model = "(S1) Baseline period sent"),
  fm_s2 %>% mutate(Model = "(S2) High Sentiment"),
  fm_s3 %>% mutate(Model = "(S3) Low Sentiment"),
  fm_s4 %>% mutate(Model = "(S4) log(Amihud) baseline"),
  fm_s5 %>% mutate(Model = "(S5) log(Amihud) High Sent"),
  fm_s6 %>% mutate(Model = "(S6) log(Amihud) Low Sent")
) %>%
  select(Model, Variable, Coefficient, tstat_NW, N_months)

table21_academic <- make_academic_table(
  fm_summary = table_21,
  variable_map = c(
    "(Intercept)" = "Intercept",
    "amihud_lag1" = "Amihud",
    "log_amihud_lag1" = "log(Amihud)",
    "log_size_lag1" = "log(Size)",
    "bm_lag1" = "B/M"
  ),
  row_order = c("Intercept", "Amihud", "log(Amihud)", "log(Size)", "B/M", "N (months)")
```

```
)

write_xlsx_safely(
  list(
    "Table18_Returns_by_Sentiment" = table_18,
    "Table19_Spread_HighLow" = table_19,
    "Table20_Alphas_by_Sentiment" = table_20,
    "Table21_FM_Sentiment" = table_21,
    "Table21_Academic" = table21_academic,
    "Sentiment_TS" = sentiment_panel,
    "Portfolio_Sent_TS" = ts_full_sent
  ),
  file.path(OUT_DIR, "Tables_18_to_21_Sentiment.xlsx")
)

cat("\n============================\n")
cat("BLOC 2/3 TERMINÉ\n")
cat("============================\n")
# ==============================================================================
# BLOC 3/3 — EDGE + TABLE 22 + ROBUSTNESS 5%
# ==============================================================================

# Vérification baseline
if (!file.exists(PATH_BASELINE_RDS)) {
  stop("Le fichier baseline_sample_final.rds est introuvable. Exécute d'abord le BLOC 1.")
}

baseline_sample <- readRDS(PATH_BASELINE_RDS)

# ==============================================================================
# 11. EDGE PROXY — CONSTRUCTION + DESCRIPTIFS + CORRÉLATIONS + FM + SORTS
# ==============================================================================
if (!require("bidask")) install.packages("bidask")
library(bidask)

cat("\n=== DIAGNOSTIC : colonnes prix dans daily ===\n")
prc_cols <- grep("prc|open|close|high|low", names(daily), value = TRUE, ignore.case = TRUE)
print(prc_cols)

edge_input <- daily %>%
  filter(penny_stock != 1) %>%
  arrange(dscd, date) %>%
  group_by(dscd) %>%
  mutate(
    open_proxy = lag(prc_daily_local)
  ) %>%
  ungroup() %>%
  transmute(
    dscd = as.character(dscd),
    month_id,
    date,
    open  = open_proxy,
    high  = prc_high_daily_local,
    low   = prc_low_daily_local,
    close = prc_daily_local
  ) %>%
  filter(!is.na(open), !is.na(high), !is.na(low), !is.na(close)) %>%
  arrange(dscd, date)

cat("\n[INFO] edge_input dimensions:\n"); print(dim(edge_input))

edge_monthly <- edge_input %>%
  group_by(dscd, month_id) %>%
  summarise(
    n_edge_days = n(),
    edge_m = if (n_edge_days >= 15) {
      tryCatch(
        bidask::edge(open = open, high = high, low = low, close = close, sign = FALSE),
        error = function(e) NA_real_
      )
    } else NA_real_,
    .groups = "drop"
  )

cat("\n[INFO] edge_monthly dimensions:\n"); print(dim(edge_monthly))
cat("\n[INFO] Distribution EDGE:\n"); print(summary(edge_monthly$edge_m))

panel_main_edge <- panel_main %>%
  left_join(edge_monthly %>% select(dscd, month_id, edge_m, n_edge_days), by = c("dscd", "month_id")) %>%
  group_by(month_id) %>%
  mutate(
    edge_m_w1 = winsor_vec(edge_m, p_low = 0.01, p_high = 0.99, min_non_na = 1)
```

```
  ) %>%
  ungroup() %>%
  arrange(dscd, month_id) %>%
  group_by(dscd) %>%
  mutate(
    edge_lag1 = lag(edge_m_w1)
  ) %>%
  ungroup()

baseline_sample_with_edge <- panel_main_edge %>%
  filter(
    !is.na(ret_lead1),
    !is.na(ret_excess_lead1),
    !is.na(size_lag1), size_lag1 > 0,
    !is.na(log_size_lag1),
    !is.na(bm_lag1), bm_lag1 > 0,
    !is.na(amihud_lag1)
  )

baseline_sample_edge_only <- baseline_sample_with_edge %>%
  filter(!is.na(edge_lag1))

saveRDS(baseline_sample_with_edge, PATH_BASELINE_EDGE_RDS)
saveRDS(baseline_sample_edge_only, PATH_BASELINE_EDGE_ONLYRDS)

table_edge_desc <- bind_rows(
  desc_stats(edge_monthly$edge_m) %>% mutate(Stage = "1. Raw construction"),
  desc_stats(panel_main_edge$edge_m) %>% mutate(Stage = "2. After merge (before winsor)"),
  desc_stats(panel_main_edge$edge_m_w1) %>% mutate(Stage = "3. After winsor 1%/99%")
) %>%
  select(Stage, Mean, Median, `Std. Dev.`, Min, Max, N) %>%
  round_numeric()

table_all_proxies <- bind_rows(
  desc_stats(panel_main_edge$amihud_m_w1)   %>% mutate(Proxy = "Amihud (winsor)"),
  desc_stats(panel_main_edge$zero_return_m) %>% mutate(Proxy = "Zero-return"),
  desc_stats(panel_main_edge$hl_range_m)    %>% mutate(Proxy = "High-low range"),
  desc_stats(panel_main_edge$turnover_m)    %>% mutate(Proxy = "Turnover"),
  desc_stats(panel_main_edge$edge_m_w1)     %>% mutate(Proxy = "EDGE (winsor)")
) %>%
  select(Proxy, Mean, Median, `Std. Dev.`, Min, Max, N) %>%
  round_numeric()

cor_data <- baseline_sample_with_edge %>%
  select(amihud_lag1, zero_return_lag1, hl_range_lag1, turnover_lag1, edge_lag1) %>%
  rename(
    Amihud = amihud_lag1,
    `Zero-ret` = zero_return_lag1,
    `HL range` = hl_range_lag1,
    Turnover = turnover_lag1,
    EDGE = edge_lag1
  )

cor_pearson  <- cor(cor_data, use = "pairwise.complete.obs", method = "pearson")
cor_spearman <- cor(cor_data, use = "pairwise.complete.obs", method = "spearman")

fm_data_edge <- baseline_sample_edge_only %>%
  mutate(
    log_edge_lag1 = log(edge_lag1 + 1e-8)
  ) %>%
  select(
    month_id, dscd, ret_lead1,
    edge_lag1, log_edge_lag1,
    log_size_lag1, bm_lag1
  ) %>%
  filter(!is.na(edge_lag1), !is.na(log_size_lag1), !is.na(bm_lag1))

fm_e1 <- run_fm("ret_lead1 ~ edge_lag1", fm_data_edge)
fm_e2 <- run_fm("ret_lead1 ~ edge_lag1 + log_size_lag1", fm_data_edge)
fm_e3 <- run_fm("ret_lead1 ~ edge_lag1 + log_size_lag1 + bm_lag1", fm_data_edge)
fm_e4 <- run_fm("ret_lead1 ~ log_edge_lag1 + log_size_lag1 + bm_lag1", fm_data_edge)

fm_edge_summary <- bind_rows(
  fm_e1 %>% mutate(Model = "(E1) EDGE only"),
  fm_e2 %>% mutate(Model = "(E2) EDGE + Size"),
  fm_e3 %>% mutate(Model = "(E3) Baseline EDGE"),
  fm_e4 %>% mutate(Model = "(E4) log(EDGE)")
) %>%
  select(Model, Variable, Coefficient, tstat_NW, N_months)

table_edge_academic <- make_academic_table(
  fm_summary = fm_edge_summary,
```

```
  variable_map = c(
    "(Intercept)" = "Intercept",
    "edge_lag1" = "EDGE",
    "log_edge_lag1" = "log(EDGE)",
    "log_size_lag1" = "log(Size)",
    "bm_lag1" = "B/M"
  ),
  row_order = c("Intercept", "EDGE", "log(EDGE)", "log(Size)", "B/M", "N (months)")
)

portfolios_edge <- baseline_sample_edge_only %>%
  group_by(month_id) %>%
  mutate(
    e_q20 = quantile(edge_lag1, 0.20, na.rm = TRUE, type = 7),
    e_q40 = quantile(edge_lag1, 0.40, na.rm = TRUE, type = 7),
    e_q60 = quantile(edge_lag1, 0.60, na.rm = TRUE, type = 7),
    e_q80 = quantile(edge_lag1, 0.80, na.rm = TRUE, type = 7),
    quintile_edge = case_when(
      edge_lag1 <= e_q20 ~ 1,
      edge_lag1 <= e_q40 ~ 2,
      edge_lag1 <= e_q60 ~ 3,
      edge_lag1 <= e_q80 ~ 4,
      edge_lag1 >  e_q80 ~ 5,
      TRUE ~ NA_real_
    )
  ) %>%
  ungroup()

portfolio_returns_edge <- portfolios_edge %>%
  filter(!is.na(quintile_edge)) %>%
  group_by(month_id, quintile_edge) %>%
  summarise(
    port_ret_excess = mean(ret_excess_lead1, na.rm = TRUE),
    n_stocks = n(),
    .groups = "drop"
  )

port_wide_edge <- portfolio_returns_edge %>%
  select(month_id, quintile_edge, port_ret_excess) %>%
  pivot_wider(names_from = quintile_edge, values_from = port_ret_excess, names_prefix = "Q") %>%
  arrange(month_id)

port_wide_edge <- ensure_columns(port_wide_edge, paste0("Q", 1:5)) %>%
  mutate(Q5_minus_Q1 = Q5 - Q1)

results_edge <- bind_rows(
  Q1 = nw_mean_tstat(port_wide_edge$Q1),
  Q2 = nw_mean_tstat(port_wide_edge$Q2),
  Q3 = nw_mean_tstat(port_wide_edge$Q3),
  Q4 = nw_mean_tstat(port_wide_edge$Q4),
  Q5 = nw_mean_tstat(port_wide_edge$Q5),
  `Q5 - Q1` = nw_mean_tstat(port_wide_edge$Q5_minus_Q1),
  .id = "Portfolio"
) %>%
  round_numeric()

write_xlsx_safely(
  list(
    "EDGE_Descriptives" = table_edge_desc,
    "All_Proxies_Desc" = table_all_proxies,
    "Correlation_Pearson" = as.data.frame(round(cor_pearson, 4)) %>% tibble::rownames_to_column("Proxy"),
    "Correlation_Spearman" = as.data.frame(round(cor_spearman, 4)) %>% tibble::rownames_to_column("Proxy"),
    "FM_EDGE_Summary" = fm_edge_summary,
    "FM_EDGE_Academic" = table_edge_academic,
    "EDGE_Univariate_Sorts" = results_edge,
    "Portfolio_TS_EDGE" = port_wide_edge
  ),
  file.path(OUT_DIR, "EDGE_Complete_Analysis.xlsx")
)

# ============================================================================
# 12. TABLE 22 — ROBUSTESSE : FM BASELINE AVEC CHAQUE PROXY
# ============================================================================
fm_data_robust <- baseline_sample %>%
  filter(!is.na(amihud_lag1), !is.na(log_size_lag1), !is.na(bm_lag1))

fm_r1 <- run_fm(
  "ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1",
  fm_data_robust
)

fm_r2 <- run_fm(
```

```
  "ret_lead1 ~ zero_return_lag1 + log_size_lag1 + bm_lag1",
  fm_data_robust %>% filter(!is.na(zero_return_lag1))
)

fm_r3 <- run_fm(
  "ret_lead1 ~ hl_range_lag1 + log_size_lag1 + bm_lag1",
  fm_data_robust %>% filter(!is.na(hl_range_lag1))
)

fm_r4 <-
  fm_r4 <- run_fm(
    "ret_lead1 ~ turnover_lag1 + log_size_lag1 + bm_lag1",
    fm_data_robust %>% filter(!is.na(turnover_lag1))
  )

table_22 <- bind_rows(
  fm_r1 %>% mutate(Model = "(R1) Amihud (baseline)"),
  fm_r2 %>% mutate(Model = "(R2) Zero-return"),
  fm_r3 %>% mutate(Model = "(R3) High-low range"),
  fm_r4 %>% mutate(Model = "(R4) Turnover")
) %>%
  select(Model, Variable, Coefficient, tstat_NW, N_months)

table22_academic <- make_academic_table(
  fm_summary = table_22,
  variable_map = c(
    "(Intercept)" = "Intercept",
    "amihud_lag1" = "Liquidity proxy",
    "zero_return_lag1" = "Liquidity proxy",
    "hl_range_lag1" = "Liquidity proxy",
    "turnover_lag1" = "Liquidity proxy",
    "log_size_lag1" = "log(Size)",
    "bm_lag1" = "B/M"
  ),
  row_order = c("Intercept", "Liquidity proxy", "log(Size)", "B/M", "N (months)")
)

write_xlsx_safely(
  list(
    "Table22_Robustness_Proxies" = table_22,
    "Table22_Academic" = table22_academic
  ),
  file.path(OUT_DIR, "Table22_Robustness_Proxies.xlsx")
)

# ==============================================================================
# 13. ROBUSTNESS CHECK — 5% WINSORIZATION FOR ALL CONTINUOUS VARIABLES
# BUG DES LAGS CORRIGÉ :
# d'abord les lags par firme (dscd), ensuite winsorisation par mois
# ==============================================================================

# ------------------------------------------------------------------------------
# 13.1 Rebuild robust sample with correct lag structure
# ------------------------------------------------------------------------------
robust_panel <- panel_monthly_base %>%
  left_join(liq_monthly_from_daily, by = c("dscd", "month_id")) %>%
  arrange(dscd, month_id) %>%
  group_by(dscd) %>%
  mutate(
    amihud_lag1_raw      = lag(amihud_m),
    turnover_lag1_raw    = lag(turnover_m),
    hl_range_lag1_raw    = lag(hl_range_m),
    zero_return_lag1     = lag(zero_return_m)
  ) %>%
  ungroup() %>%
  group_by(month_id) %>%
  mutate(
    amihud_lag1   = winsor5(amihud_lag1_raw),
    turnover_lag1 = winsor5(turnover_lag1_raw),
    hl_range_lag1 = winsor5(hl_range_lag1_raw),
    log_size_lag1 = winsor5(log_size_lag1),
    bm_lag1       = winsor5(bm_lag1)
  ) %>%
  ungroup()

robust_sample <- robust_panel %>%
  filter(
    !is.na(ret_lead1),
    !is.na(amihud_lag1),
    !is.na(log_size_lag1),
    !is.na(bm_lag1)
  )
```

```
cat("\n[INFO] robust_sample dimensions:\n")
print(dim(robust_sample))

# ---------------------------------------------------------------------------
# 13.2 EDGE with 5% winsorization — corrected lag structure
# ---------------------------------------------------------------------------
edge_monthly_robust <- daily %>%
  filter(penny_stock != 1) %>%
  arrange(dscd, date) %>%
  group_by(dscd) %>%
  mutate(open_proxy = lag(prc_daily_local)) %>%
  ungroup() %>%
  transmute(
    dscd = as.character(dscd),
    month_id,
    date,
    open  = open_proxy,
    high  = prc_high_daily_local,
    low   = prc_low_daily_local,
    close = prc_daily_local
  ) %>%
  filter(!is.na(open), !is.na(high), !is.na(low), !is.na(close)) %>%
  group_by(dscd, month_id) %>%
  summarise(
    n_edge_days = n(),
    edge_m = if (n_edge_days >= 15) {
      tryCatch(
        bidask::edge(open = open, high = high, low = low, close = close, sign = FALSE),
        error = function(e) NA_real_
      )
    } else NA_real_,
    .groups = "drop"
  )

robust_sample <- robust_sample %>%
  left_join(edge_monthly_robust %>% select(dscd, month_id, edge_m), by = c("dscd", "month_id")) %>%
  arrange(dscd, month_id) %>%
  group_by(dscd) %>%
  mutate(
    edge_lag1_raw = lag(edge_m)
  ) %>%
  ungroup() %>%
  group_by(month_id) %>%
  mutate(
    edge_lag1 = winsor5(edge_lag1_raw)
  ) %>%
  ungroup()

# ---------------------------------------------------------------------------
# 13.3 Fama-MacBeth robustness regressions (5% winsorization)
# ---------------------------------------------------------------------------
fm_amihud_5pct <- run_fm(
  "ret_lead1 ~ amihud_lag1 + log_size_lag1 + bm_lag1",
  robust_sample
)

fm_edge_5pct <- run_fm(
  "ret_lead1 ~ edge_lag1 + log_size_lag1 + bm_lag1",
  robust_sample %>% filter(!is.na(edge_lag1))
)

table_robust_5pct <- bind_rows(
  fm_amihud_5pct %>% mutate(Model = "Amihud 5pct"),
  fm_edge_5pct %>% mutate(Model = "EDGE 5pct")
) %>%
  select(Model, Variable, Coefficient, tstat_NW, N_months)

table_robust_5pct_academic <- make_academic_table(
  fm_summary = table_robust_5pct,
  variable_map = c(
    "(Intercept)" = "Intercept",
    "amihud_lag1" = "Amihud",
    "edge_lag1" = "EDGE",
    "log_size_lag1" = "log(Size)",
    "bm_lag1" = "B/M"
  ),
  row_order = c("Intercept", "Amihud", "EDGE", "log(Size)", "B/M", "N (months)")
)

write_xlsx_safely(
  list(
```

```
    "Robustness_5pct_Summary" = table_robust_5pct,
    "Robustness_5pct_Academic" = table_robust_5pct_academic,
    "FM_Amihud_5pct" = fm_amihud_5pct,
    "FM_EDGE_5pct" = fm_edge_5pct
  ),
  file.path(OUT_DIR, "Robustesse_5pct_Results.xlsx")
)

cat("\n============================\n")
cat("BLOC 3/3 TERMINÉ\n")
cat("============================\n")
cat("\nFichiers créés :\n")
cat("- EDGE_Complete_Analysis.xlsx\n")
cat("- Table22_Robustness_Proxies.xlsx\n")
cat("- Robustesse_5pct_Results.xlsx\n")
############################################################
# BLOC FINAL : REGROUPER TOUS LES TABLEAUX DANS UN SEUL EXCEL
############################################################

# Packages
req_merge <- c("openxlsx", "readxl")
new_merge <- req_merge[!(req_merge %in% installed.packages()[, "Package"])]
if (length(new_merge) > 0) install.packages(new_merge)

library(openxlsx)
library(readxl)

OUT_DIR <- "C:/Users/najwa/Downloads"

# Fichiers Excel déjà générés par ton script
source_files <- c(
  file.path(OUT_DIR, "Tables_1_to_6_Baseline.xlsx"),
  file.path(OUT_DIR, "Tables_7_to_9_Univariate_Amihud.xlsx"),
  file.path(OUT_DIR, "Tables_10_to_13_Bivariate_Amihud.xlsx"),
  file.path(OUT_DIR, "Tables_14_to_16_BullBear.xlsx"),
  file.path(OUT_DIR, "Table_17_FamaMacBeth.xlsx"),
  file.path(OUT_DIR, "Tables_18_to_21_Sentiment.xlsx"),
  file.path(OUT_DIR, "Table22_Robustness_Proxies.xlsx")
)

# Si tu veux aussi inclure les fichiers EDGE / robustesse supplémentaire,
# décommente les lignes ci-dessous :
# source_files <- c(
#   source_files,
#   file.path(OUT_DIR, "EDGE_Complete_Analysis.xlsx"),
#   file.path(OUT_DIR, "Robustesse_5pct_Results.xlsx")
# )

# Vérifier les fichiers existants
existing_files <- source_files[file.exists(source_files)]

if (length(existing_files) == 0) {
  stop("Aucun fichier Excel source trouvé dans OUT_DIR.")
}

cat("\n[INFO] Fichiers trouvés pour fusion :\n")
print(basename(existing_files))

# Fonction pour noms d’onglets propres et uniques
make_safe_sheet_name <- function(x, used_names = character()) {
  x <- gsub("[\\[\\]\\*\\?/\\\\:]", "_", x)   # caractères interdits Excel
  x <- gsub("\\s+", "_", x)
  x <- substr(x, 1, 31)                      # max 31 caractères Excel

  base <- x
  i <- 1
  while (x %in% used_names) {
    suffix <- paste0("_", i)
    x <- substr(base, 1, 31 - nchar(suffix))
    x <- paste0(x, suffix)
    i <- i + 1
  }
  x
}

# Nouveau workbook final
wb_final <- createWorkbook()
used_sheet_names <- character()

for (f in existing_files) {

  file_tag <- tools::file_path_sans_ext(basename(f))
```

```
  sheets_in_file <- excel_sheets(f)

  for (sh in sheets_in_file) {

    # lire la feuille
    dat <- tryCatch(
      read_excel(f, sheet = sh),
      error = function(e) NULL
    )

    if (is.null(dat)) next

    # nom de feuille final
    if (length(sheets_in_file) == 1) {
      proposed_name <- file_tag
    } else {
      proposed_name <- sh
    }

    safe_name <- make_safe_sheet_name(proposed_name, used_sheet_names)
    used_sheet_names <- c(used_sheet_names, safe_name)

    addWorksheet(wb_final, safe_name)
    writeData(wb_final, sheet = safe_name, x = as.data.frame(dat))

    # mise en forme simple
    if (ncol(dat) > 0) {
      setColWidths(wb_final, sheet = safe_name, cols = 1:ncol(dat), widths = "auto")
    }
    freezePane(wb_final, sheet = safe_name, firstRow = TRUE)
  }
}

# Fichier final unique
final_file <- file.path(OUT_DIR, "Tous_les_tableaux_1_a_22.xlsx")
saveWorkbook(wb_final, final_file, overwrite = TRUE)

cat("\n[OK] Fichier final créé :\n")
cat(final_file, "\n")
```

Bibliography

Acharya, V. V., & Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics, 77*(2), 375–410.
https://www.sciencedirect.com/science/article/abs/pii/S0304405X05000334

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets, 5*(1), 31–56.
https://www.sciencedirect.com/science/article/pii/S1386418101000246

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics, 17*(2), 223–249.
https://www.sciencedirect.com/science/article/pii/0304405X86900656

Bank, M., Larch, M., & Peter, G. (2010). *Investors' compensation for illiquidity: Evidence from the German stock market*. SSRN.
https://ssrn.com/abstract=1653831

Bekaert, G., Harvey, C. R., & Lundblad, C. (2007). Liquidity and expected returns: Lessons from emerging markets. *The Review of Financial Studies, 20*(6), 1783–1831.
https://academic.oup.com/rfs/article-abstract/20/6/1783/1575135

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *The Review of Financial Studies, 22*(6), 2201–2238.
https://academic.oup.com/rfs/article/22/6/2201/1592184

Chordia, T., Sarkar, A., & Subrahmanyam, A. (2005). An empirical analysis of stock and bond market liquidity. *The Review of Financial Studies, 18*(1), 85–129.
https://www.newyorkfed.org/research/staff_reports/sr164.html

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*(2), 357–384.
https://econpapers.repec.org/RePEc:ecm:emetrp:v:57:y:1989:i:2:p:357-84

Johann, T., Scharnowski, S., Theissen, E., Westheide, C., & Zimmermann, L. (2019). Liquidity in the German stock market. *Schmalenbach Business Review, 71*(4), 443–473.
https://www.econstor.eu/bitstream/10419/200189/1/1668069784.pdf

Lesmond, D. A., Ogden, J. P., & Trzcinka, C. A. (1999). A new estimate of transaction costs. *The Review of Financial Studies, 12*(5), 1113–1141.
https://academic.oup.com/rfs/article-abstract/12/5/1113/1565408

Leibniz Institute for Financial Research SAFE. (n.d.). SAFE Manager Sentiment Index.
https://safe-frankfurt.de/research/manager-sentiment-index.html

Pástor, Ľ., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy, 111*(3), 642–685.
https://www.nber.org/system/files/working_papers/w8462/w8462.pdf

Paul, T., Aryoubi, A., & Walther, T. (2025). Reassessing the illiquidity-return relationship: Evidence from Germany, the UK, and the U.S. *International Review of Financial Analysis, 106*, 104469.
https://doi.org/10.1016/j.irfa.2025.104469