A Work Project, presented as part of the requirements for the Award of a master's degree in Finance from the Nova School of Business and Economics.

Price Dynamics in OTC Bond Market: The Role of Market Microstructure

Can Enis Keskin

Work project carried out under the supervision of:

Professor Tim Eisert

20/01/2024

Abstract

This thesis investigates the drivers of daily price returns in over-the-counter corporate bond markets by comparing market microstructure elements with factors like liquidity, market conditions, and bond-specific characteristics. Using a dynamic Instrumental Variable Two-Stage Least Squares (IV-2SLS) model, the study identifies changes in the 10-year Treasury yield and yield spreads as primary determinants of bond returns. Additionally, market microstructure factors significantly influence returns as secondary mechanisms refining pricing based on market conditions, liquidity, and bond characteristics. These findings underscore the dominant role of macroeconomic indicators and the context-dependent role of microstructure factors, providing insights to enhance market efficiency and stability.

Keywords: Corporate Bonds, Market Microstructure, OTC Markets, Financial Markets, Price Dynamics

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

Over-the-counter (OTC) corporate bond markets, characterized by decentralized trading and a network of dealers negotiating prices bilaterally, present a unique environment for price discovery and price dynamics. Unlike equities trading on centralized exchanges, the OTC bond market's fragmentation and opacity can lead to substantial heterogeneity in how prices are formed and transmitted, making the roles of trading frictions, dealer behavior, and intermediation more complex (Saunders et al. 2002; Edwards et al. 2007; Ederington et al. 2015). This complexity is not merely of academic interest: corporate bonds constitute a large share of global capital markets, and their price stability, resilience, and overall efficiency have significant implications for investors, issuers, and regulators alike. According to the 2024 SIFMA Capital Markets Fact Book, by the end of 2023, outstanding U.S. corporate bonds reached approximately USD 10.76 trillion, reflecting a 3% increase from 2022. Additionally, corporate bond issuance in 2023 totaled USD 1.4 trillion, highlighting their essential role in financing business expansion, mergers, and other strategic initiatives. Corporate bonds are the second-largest category in the fixed income market, following U.S. Treasury securities.

This study addresses a pivotal research question: How does market microstructure influence daily price dynamics in OTC corporate bond markets compared to traditional factors such as liquidity conditions, market-wide shocks, and bond-specific characteristics? Daily returns are used to measure price dynamics, capturing both the direction and magnitude of price changes. The analysis examines the comparative and combined effects of market microstructure variables—including dealer trading behavior, transaction costs, and trading platforms—against conventional determinants like liquidity measures, broad market conditions, and fundamental bond attributes (Bessembinder et al. 2006; Duffee 2011; Elton et al. 1999).

The primary contribution of this thesis lies in its comparative approach. Instead of analyzing market microstructure variables in isolation, it contrasts them with other explanatory factors, including liquidity proxies (e.g., price spreads, trading frequency), market conditions (e.g., volatility, yield curve changes), and bond-specific characteristics (e.g., credit rating, maturity, coupon). While previous studies typically examine one dimension at a time, this research systematically evaluates a diverse set of factors to determine the relative impact of market microstructure on OTC bond price dynamics.

Furthermore, the study investigates the interactions between market microstructure and other variables by incorporating interaction terms and utilizing robust panel data econometric techniques (Baltagi 2020; Petersen 2009; Newey and West 1987; Andrews 1991) This methodology reveals complex relationships and deepens the understanding of price formation in OTC bond markets. By integrating industry best practices and conducting a comparative analysis across multiple themes, the research ranks the significance of various determinants in bond market price setting, thereby enhancing the existing literature.

To address the research question, this thesis formulates several testable hypotheses. Hypothesis 1 (H1) examines whether market microstructure variables have a greater impact on daily price dynamics than external market conditions, liquidity, and bond characteristics. The analysis reveals that fluctuations in market conditions—particularly the 10-year Treasury yield—and bond characteristics, especially the yield spread, are the primary drivers of short-term price movements. In contrast, market microstructure variables and liquidity measures play a secondary role in refining the pricing process. These results are consistent across different subsets and estimation methods.

Hypothesis 2 (H2) investigates whether market microstructure effects on short-term price dynamics are influenced by market conditions and liquidity. Regression analysis with interaction terms reveals that the impact of market microstructure variables on daily bond returns varies with market conditions and credit grades. Specifically, higher liquidity tightens prices by altering dealers' compensation mechanisms through commissions. The interaction between Alternative Trading System (ATS) usage and commissions shows that efficiency gains from electronic trading platforms offset transaction costs. Additionally, market volatility enables dealers to leverage compensation for providing liquidity. These effects, however, depend on bond characteristics.

By disentangling these effects, the study helps to determine that microstructure considerations merit a central place in models of corporate bond price formation, complementing well-known liquidity-based explanations and responses to macro-financial shocks.

2. Literature Review

While a substantial body of literature investigates factors influencing bond pricing—particularly liquidity, credit risk, and macroeconomic conditions—relatively fewer studies integrate market microstructure variables, such as dealer behavior, trading protocols, and transaction cost components, into a unified empirical framework (Bessembinder et al. 2006; Edwards et al. 2007; Saunders et al. 2002). The existing research, though extensive, often examines these factors in isolation, leaving a gap in understanding their relative importance in driving price dynamics, making it unclear which variables are primary drivers and which are secondary.

Early seminal works emphasized liquidity as a key driver of bond price behavior. Elton et al. (1999) highlight the role of credit risk and tax-adjusted spreads, but their focus primarily rests on broad yield differentials rather than the intricacies of day-to-day returns or dealer intermediation. Chen et al. (2007) delve deeper into liquidity proxies—such as bid-ask spreads and zero-trading days—demonstrating that illiquidity can cause significant deviations in bond prices from their fundamental values. Similarly, Dick-Nielsen et al. (2012) analyze liquidity during the financial crisis and show that liquidity shortages profoundly influence corporate bond pricing, particularly in stressed markets. These studies establish liquidity as a fundamental determinant, which is often treated as a exogenous feature of the market environment. However, they primarily measure liquidity's impact on yield spreads or long-term pricing rather than short-run price dynamics.

Beyond liquidity, macroeconomic shocks and broader market conditions, have long been recognized as key determinants of bond prices. Duffee (2011) and others have shown how interest rate volatility, equity market turbulence, and credit spread fluctuations affect bond values. More nuanced perspectives incorporate both macro factors and trading characteristics. Ederington et al. (2015) , for instance, decompose dealer spreads into distinct cost drivers—market-making risks, agent-related costs, and variability in underlying market conditions—to explain transaction-level bond price formation. Their work signals that macro conditions do not act in isolation; rather, they interact with micro-level features like dealer activity and volatility-sensitive markups.

In parallel, research on corporate bond fundamentals established that default risk, taxation, and issuer characteristics influence long-term bond pricing. Elton et al. (1999) and related work confirm that rating classes and other issuer-level factors shape yield spreads. More recently, Acharya et al. (2013) show that liquidity risk interacts with credit conditions and that the importance of liquidity can shift dramatically under stressed regimes, implying that bond fundamentals alone cannot fully explain short-term price fluctuations. While these studies highlight fundamental determinants, they remain anchored to either static yield levels or scenario-specific spreads, rather than continuous daily price adjustments.

The literature on the microstructure of OTC corporate bond markets was pioneered by works such as Edwards et al. (2007) and Saunders et al. (2002), which laid the groundwork for understanding how dealer markups, trade size, and other basic transaction characteristics influence corporate bond trading costs. Bessembinder et al. (2006) further emphasized the critical role of transparency and trading protocols in shaping corporate bond prices. More recent research has deepened this analysis by considering how evolving market structures and technologies alter bond price dynamics. For example, Hendershott and Madhavan (2015) highlight how the introduction and adoption of electronic trading platforms and auction mechanisms enhance market access and liquidity provision, while Friewald et al. (2012) show that liquidity and credit risk are closely intertwined, especially during periods of financial turmoil, when illiquidity premiums sharply increase.

Despite these advances, existing studies often focus on selective aspects of microstructure or treat key factors in isolation. Hendershott and Madhavan (2015), for instance, analyze the implications of electronic trading for liquidity but do not simultaneously control for the interplay of changing credit conditions, heterogeneous bond attributes, and macroeconomic volatility. Conversely, while Friewald et al. (2012) document how varying market conditions affect the relative importance of liquidity vis-à-vis credit risk, their study does not explicitly dissect how dealer behaviors, inter-dealer competition, or inventory constraints shape this relationship.

Thus, there is a gap in research that treats microstructure as one component in a richer ecosystem of determinants affecting daily returns. The present thesis aims to fill this gap. Building on the present literature this thesis advances the field in several key ways. Firstly, it moves beyond liquidity alone to incorporate dealer activities, transaction costs, and trading platforms, exploring how these elements interact with liquidity conditions and market shocks to shape daily bond returns. By shifting the focus from isolated trade-level spreads to aggregated daily returns, the study simultaneously examines the influence of evolving market conditions and intra-day trading microstructures on short-horizon bond performance, capturing more immediate responses to market shocks. Secondly, the research integrates bond-specific attributes—such as credit ratings, maturity, and coupon structures—with a comprehensive set of market microstructure and liquidity variables. This connection bridges the established fundamentals literature with the dynamic evolution of market prices, offering a more granular understanding of how fundamental attributes interact with dealer strategies and trading conditions to influence daily returns. Furthermore, the study employs interaction terms and robust panel data econometric techniques (Baltagi 2020; Petersen 2009; Newey and West 1987; Andrews 1991) to assess whether the importance of microstructure varies under different liquidity regimes or market volatility states. This methodological approach allows for the disentanglement of the relative importance of each factor and reveals how their interactions drive daily price movements. By placing market microstructure variables alongside traditional liquidity measures and bond characteristics, this study not only disentangles the relative importance of each factor but also elucidates their combined effects on daily price movements. Ultimately, this thesis bridges the gap between microstructure and broader market determinants, providing valuable insights for investors, issuers, and regulators.

3. Theory

3.1 Market Microstructure

Understanding the institutional and microstructural aspects of the U.S. OTC corporate bond market is essential for interpreting the empirical results of this study. Unlike the transparent, centralized frameworks of equity and futures exchanges, the OTC corporate bond market is decentralized, fragmented, and less transparent. Prices are determined through bilateral negotiations rather than a centralized order book, making trading frictions, search costs, and dealer intermediation crucial in shaping price discovery and dynamics (Saunders et al. 2002; Edwards et al. 2007; Ederington et al. 2015).

In the U.S., corporate bonds are traded over-the-counter among dealers and institutional investors, including broker-dealers, asset managers, pension funds, insurance companies, and hedge funds. Dealers act as intermediaries, providing both agency services by matching buyers and sellers, and market-making services by maintaining inventories to offer immediate liquidity. Institutional investors often rely on dealers for trading less liquid bonds or large block trades (Saunders et al. 2002). Despite the rise of electronic trading platforms and alternative trading systems (ATS), the market remains largely relationship-driven with significant variation in trade execution and pricing (Bessembinder et al. 2006; Hendershott and Madhavan 2015).

The OTC market employs a request-for-quote (RFQ) mechanism, where investors contact multiple dealers for price quotes. This process is time-consuming and incurs search costs due to limited pre-trade transparency, as there is no consolidated order book (Bergault and Guéant 2023). Additionally, corporate bonds are heterogeneous in terms of credit quality, maturity, coupon features, and liquidity, complicating price discovery and increasing reliance on dealers' expertise and networks (Edwards et al. 2007).

The introduction of the Trade Reporting and Compliance Engine (TRACE) by the Financial Industry Regulatory Authority (FINRA) has enhanced post-trade transparency by mandating reporting of bond transactions. However, pre-trade transparency remains limited, requiring investors to use TRACE data, dealer quotes, and proprietary models to assess fair prices (Schneider 2018; Edwards et al. 2007).

Key frictions in the OTC corporate bond market include pre-trade opacity, search costs, inventory-carrying costs, and information asymmetry. Dealers face costs and risks from holding inventories and from trading with informed investors, leading to wider bid-ask spreads and cautious pricing (Duffie et al. 2005; Hendershott and Madhavan 2015; Ederington et al. 2015). Illiquid bonds incur higher transaction costs due to the difficulty in finding counterparties and longer inventory holding periods (Bessembinder et al. 2006; Dick-Nielsen et al. 2012).

Understanding these microstructural elements is crucial for analyzing how market microstructure variables, along with liquidity, market conditions, and bond characteristics, influence daily price dynamics. This research leverages TRACE transaction data and robust econometric techniques to assess the role of market microstructure alongside other factors, enhancing the comprehension of price formation in the U.S. corporate bond market. Find the discussion of corporate bond characteristics in U.S. bond markets in the **Appendix 1: Corporate Bonds.**

3.3 Empirical Model: Dynamic IV-2SLS

This research utilizes a dynamic Instrumental Variable (IV) Two-Stage Least Squares (2SLS) model to estimate the effects of market microstructure, liquidity, bond characteristics, and market conditions on daily bond price returns. Unlike previous studies such as Edwards et al., (2007) that assume exogeneity, this approach directly addresses potential endogeneity—particularly of the lagged dependent variable and the volatility of daily bond price returns — while also preserving key variables and allowing for a comprehensive, causal interpretation of the determinants of bond returns (Chen et al. 2007).

To capture short-term trending or mean reverting behavior of bond returns, a lagged dependent variable is included. However, this introduces endogeneity concerns, as past returns and current volatility can simultaneously influence each. Following Dick-Nielsen et al. (2012), a Durbin-Wu-Hausman test confirmed that endogeneity is present. Consequently, the IV-2SLS framework is employed. The lagged dependent variable and daily price volatility are instrumented using their second lags. Comprehensive testing of various instrument combinations identified the second lag of the endogenous variable as the most stable configuration. First-stage tests evaluated this approach, confirming it as a robust and valid instrument selection (see **Appendix 2: Statistical Tests** for details).

Prior influential studies (e.g., Edwards et al. 2007; Ederington et al. 2015) primarily rely on Weighted Least Squares (WLS) or differences-in-differences approaches without explicit endogeneity corrections. However, estimation methods such as the Generalized Method of Moments (GMM) are less suitable in environments characterized by both a large cross-sectional dimension (N) and a long temporal dimension (T), as GMM tends to overfit when faced with an excessive number of instruments in large-T settings (Moon and Weidner 2017). In contrast, the simpler IV-2SLS approach effectively avoids this issue. With a cross section of up to approximately 4500 bonds (N) across 322 days (T) the IV-2SLS framework enables a more reliable disentangling of the relative importance of market microstructure, liquidity, bond characteristics, and market conditions (see **Appendix 3.1: Data**). Critically, this approach supports the inclusion of time- or entity-invariant variables—essential for comparing the relative importance of all variable groups—while still adequately addressing unobserved heterogeneity. See **Appendix 3.2: Regression Model** for the detailed model specifications.

Applying fixed effects as outlined by Chen et al. (2007) would result in perfect collinearity for time-invariant variables (e.g., bond characteristics) and entity-invariant variables (e.g., market conditions). To overcome this issue, this study adopts a refined transformation strategy aligned with advanced panel techniques (Bai 2009; Moon and Weidner 2017). Specifically, time- and entity-variant variables, such as market microstructure and liquidity, are both demeaned and cross-sectionally averaged to eliminate unobserved heterogeneity at both the entity and time levels. Meanwhile, entity-invariant variables like market conditions are only demeaned, preserving their necessary variation without completely removing their influence. Time-invariant variables, such as bond characteristics, are solely cross-sectionally averaged to ensure their effects remain identifiable. This selective transformation approach prevents key regressors from being absorbed due to perfect collinearity, effectively addressing unobserved heterogeneity across both dimensions while retaining critical explanatory variables, including bond characteristics and market shocks. See **Appendix 6: Detailed Data Pre-Processing Pipeline and Key Variable Transformations** for further information.

Beyond addressing endogeneity and heterogeneity, model diagnostics reveal the presence of both heteroskedasticity and autocorrelation in the error structure, see **Appendix 2.4: Model Diagnostics (Residual analysis).** To account for these issues, the literature typically employs two-way clustered standard errors (e.g., Dick-Nielsen et al. 2012) or heteroskedasticity- and autocorrelation-consistent (HAC) robust standard errors (e.g., Bessembinder et al. 2006). Two-way clustered standard errors effectively capture correlations across two dimensions, such as entity and time, providing a straightforward and conservative inference framework. However, they may not fully address more complex serial dependence patterns. Conversely, HAC robust standard errors offer greater flexibility by accommodating non-constant variance and intricate autocorrelation structures, which is beneficial for dynamic relationships, though they require careful selection of bandwidth and kernel parameters (Petersen 2009; Newey and West 1987). To ensure robust and reliable inferences, this study employs both two-way clustered and HAC robust standard errors. Standard errors are clustered at the bond and trade-date levels, while HAC standard errors use automatic bandwidth allocation to optimize the estimation without imposing restrictive assumptions. By comparing results from both methods, the analysis enhances the robustness of the findings and strengthens confidence in the validity of the conclusions (Andrews 1991).

The modeling strategy presented does not merely replicate one existing approach. Instead, it strategically combines insights and techniques from various strands of literature—dynamic panel modeling, IV-2SLS estimation, advanced panel transformations and covariance estimations —to suit the unique requirements of this study. By doing so, this model addresses the challenges posed by large, complex panel data, preserves essential variables for comparative analysis, and provides consistent and unbiased estimates.

4. Empirical Analysis

4.1 Model Results

The regression results reveal that while market microstructure variables are consistently statistically significant, they often play a secondary role relative to key macroeconomic and bond-level factors. This finding speaks directly to the first hypothesis (H1), which posited that market microstructure variables would have a more pronounced impact on daily price dynamics than external conditions, liquidity measures, and bond characteristics.

The estimated model suggests that macroeconomic and bond-level characteristics primarily govern daily bond price dynamics. Among these, deviations in the 10-year Treasury yield stand out as the most influential factor, with even a one basis point departure from its mean exerting a discernible downward effect on bond returns. This aligns with fundamental fixed-income theory: as benchmark yields rise, bond prices typically fall. Additionally, non-linear shifts in yield curve curvature, captured by a cubic transformation, reveal secondary yet meaningful nuances. Significant deviations in curvature—whether reflecting bull or bear steepening—negatively affect bond returns. In particular, the interaction of 10-year yield changes with yield curvature deviations deepens the negative return impact in steepening scenarios. Thus, while the level of benchmark rates is the paramount driver, the shape of the curve also informs the market's pricing response and risk assessment.

Beyond macro-level factors, fundamental bond characteristics further shape returns. Changes in yield spreads, time-to-maturity, and coupon structures carry substantial weight. Yield spreads, for instance, show a clear inverse relationship with returns, confirming their status as a core proxy for credit conditions and investor risk sentiment. Non-traditional bond features, such as zero-coupon or "other" coupon structures, also stand out, often introducing heightened return volatility and sensitivity to interest rate movements. Together, these macroeconomic and bond-level determinants dominate the explanatory hierarchy, underscoring their primary role in daily price formation.

Market microstructure variables, while secondary, still matter. Joint significance tests, via a Wald, test confirm that the set of microstructure factors is collectively nontrivial and statistically relevant (see **Appendix 2: Statistical Tests** for test results). Within this group, Customer Selling Pressure (CSP) emerges as the most consistently influential component. Elevated selling pressure—characterized by a larger proportion of sell trades—reduces returns, capturing how dealers adjust prices downward to manage inventory risk under selling imbalances. Other microstructure measures, such as commissions, ATS usage, or customer flow, have subtler direct effects, yet remain statistically significant, suggesting that trading frictions, infrastructure, and participant composition fine-tune the price formation process. For instance, increases in the share of commission-bearing trades slightly bolster returns. Meanwhile, an increased reliance on Alternative Trading Systems corresponds to modest return improvements, hinting at enhanced price discovery and reduced transaction costs in more transparent and technology-driven environments. Lastly, customer flow presents an additional insightful dimension to the analysis as greater customer involvement may signal stronger demand from end investors, driving bond prices higher, while reduced inter-dealer activity might reflect less competitive pressure and lose pricing. Higher customer participation rates also indicate more frequent investor trading, potentially signaling broader market confidence or evolving trading strategies.

Market liquidity metrics, on the other hand, play a more moderate yet still meaningful role. Changes in daily trading volume, inactivity (time between trades), and trading frequency each display statistically significant but comparatively modest effects on returns. For instance, higher daily trading volume, reflecting market depth (Ranaldo 2002), marginally lowers returns—likely due to increased competition and tighter pricing. Conversely, longer waiting times between trades, a proxy for inactivity, appear to slightly increase returns, perhaps reflecting moments of less "noisy" trading and more stable price discovery. These liquidity dimensions act as secondary refinements: they neither overshadow macroeconomic nor bond-level factors, nor do they outstrip the leading microstructure variables. Instead, they shape the conditions under which both macro-level fundamentals and microstructure forces manifest, influencing the smoothness, stability, and incremental pricing adjustments.

In sum, H1 is not supported. However, while market conditions and bond attributes clearly take precedence in explaining daily return variations, market microstructure variables, confirmed jointly significant by the Wald test, and liquidity measures do provide valuable secondary layers of explanation. The key takeaway is that microstructure and liquidity factors become consequential in refining the price response, even if they do not surpass the dominant role played by broader yield conditions and inherent bond characteristics.

See the **Appendix 4: Detailed Regression Analysis** for detailed individual interpretation of the model parameters.

Full Model	Coefficients		P-value		Individual Rank Full Model		Group Rank Full Model	
			Bond-Date Two-way clustered robust standard errors model	HAC robust standard errors mode	Bond-Date Two-way clustered robust standard errors model	HAC robust standard errors mode	Bond-Date Two-way clustered robust standard errors model	HAC robust standard errors mode
R-squared	0,0549	0,0549						
Adj. R-squared	0,0549	0,0549						
F-statistic	1216,1	8526,5						
P-value (F-stat)	0,0000	0,0000						
Distribution	chi2(29)	chi2(29)						
Estimator	IV-2SLS	IV-2SLS						
No. Observations	1.238.547	1.238.547						
Cov. Estimator	clustered (6082, 322)	kernel (bandwith 70)						
Δ 10Y Treasury Yield	-0,0017		0,0026	0,0000	3	8	1	2
Δ Market Volatility	-0,0007		0,4747	0,0880	14	24		
Δ Treasury Yield Curvature (TYC)	-0,0006		0,0329	0,0000	5	11		
Commission	4,47E-05		0,0097	0,0004	11	20	2	5
ATS Usage	8,35E-05		0,0119	0,0007	10	19		
Customer Flow	1,13E-05		0,1690	0,0111	14	22		
Customer Selling Pressure (CSP)	-0,0002		0,0000	0,0000	7	15		
Δ Daily Trading Volume	-0,0001		0,0160	0,0002	8	16	5	4
Δ Average Trading Inactivity	9,49E-05		0,0006	0,0000	9	17		
Δ Daily Trading Frequency (DTF)	-0,0004		0,0736	0,0155	14	12		
Δ Price Spread	4,85E-05		0,6044	0,3276	14	24		
Δ Intraday Return Variance	-0,0002		0,1397	0,1418	14	24		
ATS - DTF Interaction	-3,87E-05		0,1640	0,1076	14	24	4	3
Commission - DTF Interaction	-0,0004		0,0282	0,0075	6	12		
Commission - ATS Interaction	6,66E-06		0,0089	0,0002	13	23		
CSP - Daily Trading Volume Interaction	3,73E-05		0,0023	0,0000	12	21		
Customer Flow - Price Spread Interaction	-9,05E-05		0,0553	0,0089	14	18		
Commission - Market Volatiltiy Interaction	-0,0015		0,0852	0,0003	14	9		
Δ 10Y Treasury Yield - TYC Interaction	-0,0010		0,0245	0,0000	4	10		
Δ 10Y Treasury Yield - TTM Interaction	-0,0003		0,1557	0,0000	14	14		
Time-to-maturity (TTM)	-0,0262		0,1018	0,0000	14	5	3	1
Coupon Rate	0,0066		0,3164	0,0000	14	7		
Credit Grade	-0,0090		0,0759	0,0000	14	6		
Δ Yield Spread	-0,1324		0,0053	0,0000	1	3		
Zero Coupon Bond Flag	-0,6445		0,1502	0,0032	14	1		
Floating Coupon Bond Flag	0,1041		0,0322	0,0000	2	4		
Other Coupon Bond Flag	-0,2625		0,3293	0,0126	14	2		
Lag 1 Daily Price Returns	-0,3624		0,3659	0,1933	14	24		
Rolling 5-day Bond Price Volatility	0,0006		0,2290	0,1435	14	24		

Table 1: IV-2SLS Model Output and Ranking

The second hypothesis posited that the relevance of market microstructure factors is context-dependent, particularly in interaction with market conditions and other thematic groups of variables. The results substantiate this point. Interactions between commissions and trading frequency, for example, reveal that returns are more adversely affected when both transaction costs and trading intensity rise concurrently. This could imply that frequent trading in these scenarios involves bonds with lower-than-average embedded commissions, potentially leading to less inflated bond prices. It could also reflect the compensation mechanisms as with higher liquidity there is less need for embedding commissions in the price as dealers need less compensation for providing liquidity when liquidity is ample.

Similarly, the interplay between commissions and ATS usage reveals a notable dynamic. When both commission-bearing trades and ATS activity rise above average levels, there is a combined positive impact on daily bond returns. This suggests that the efficiency gains from electronic trading platforms can help mitigate the return-reducing effects of higher commissions, highlighting the complementary role of technology in enhancing market performance.

These findings portray market microstructure elements as part of a dynamic ecosystem, where their impact amplifies, mitigates, or reshapes itself depending on concurrent trading conditions, market stability, or investor behavior.

The conditional significance of these interactions is further highlighted by how microstructure variables respond differently to changes in liquidity and volatility regimes. The presence of higher market volatility, when paired with elevated commissions, compounds negative return impacts. In contrast, robust trading volume can somewhat cushion the downward pressure exerted by heavy customer selling. Thus, microstructure factors are not static; their influence evolves based on broader conditions. While they may seldom lead price dynamics independently, their effect is material when layered atop shifts in benchmark yields, adjustments in yield curve shape, changes in credit conditions, and evolving liquidity environments, strongly supporting H2.

In this study, regressors were carefully chosen to address multicollinearity and ensure sufficient variation across clusters. The process involved an iterative assessment of Variance Inflation Factors (VIF), correlation analysis, and the use of low variance filters both within and between clusters.

To mitigate multicollinearity, variables exceeding a VIF threshold of 10 were scrutinized and either combined or excluded. Pairwise correlation analysis further identified variables with high redundancy (e.g., correlations above 0.8), leading to targeted adjustments. Additionally, variables with low variability—especially those with near-constant values within clusters—were transformed (e.g., through binning) to enhance variance. If no improvement was achieved, they were excluded to ensure greater model stability and interpretability

Key adjustments included combining market volatility indicators. Principal Component Analysis (PCA) was applied to merge the VIX and MOVE indices into a single composite variable, effectively capturing market volatility shocks while reducing collinearity. Further, the embedded and explicit commission variables were combined into a single Total Commission Flag variable to address high collinearity and low variation. This approach balances comprehensiveness with interpretability, as it retains all relevant information without excluding a parameter. While this trade-off reduces the ability to distinguish between embedded and explicit commissions, it was deemed acceptable to preserve the overall integrity of the model.

Moreover, the market platform variable was excluded due to its low variability and exceptionally high VIF, which introduced multicollinearity concerns. While this variable distinguishes between primary and secondary market trades, its near-constant nature—caused by the very limited occurrence of primary market trades—rendered it unsuitable for inclusion in the model.

While some variables slightly exceeded the VIF threshold, exceptions were made for key market microstructure variables—such as ATS Usage and its interaction with commissions—due to their critical role in addressing the research question. Similarly, Daily Trading Frequency and the composite Market Volatility variable were retained for their theoretical and empirical importance in capturing trading intensity and market shocks. These variables were maintained as they are central to their respective groups, with the slight VIF breaches considered acceptable given their significance to the analysis. See the VIF overview in the **Appendix 2.3: Variance Inflation Factor (VIF).**

The final variable set reflects a balance between multicollinearity management and comprehensive variable coverage, incorporating macroeconomic factors, bond-specific attributes, and market microstructure insights. This rigorous selection process enhances the model's robustness, ensuring its reliability in analyzing bond return dynamics while preserving theoretical underpinnings.

4.2 Robustness Checks

4.2.1 Covariance Estimator

To ensure the reliability and consistency of the regression results, robustness checks were performed using two different covariance estimators: two-way clustered standard errors, as proposed by Dick-Nielsen et al. (2012), and HAC robust standard errors, following Bessembinder et al. (2006). This approach tests whether the key findings remain robust under varying error structure assumptions, thereby strengthening the credibility of the inferences (Petersen 2009).

The results confirm that key macroeconomic and bond-level variables—such as deviations in the 10-year Treasury yield, yield curve curvature, and yield spreads—are consistently significant across both estimators. This reinforces their central role in driving daily bond price dynamics. Similarly, market microstructure variables, including CSP, commissions, ATS usage, trading volume, and trading inactivity, maintain robust significance, underscoring their stable but secondary contribution to price formation.

However, certain variables and interaction terms, such as customer trade flow, trading frequency, time-to-maturity, coupon rate, and credit grade, exhibit conditional robustness. These factors remain significant only under HAC robust standard errors, indicating their sensitivity to heteroskedasticity or autocorrelation. Interaction terms like Customer Flow–Price

Spread and Commission–Market Volatility also require cautious interpretation due to their varying significance.

Overall, the robustness checks validate the core findings, emphasizing the reliable influence of macroeconomic and market microstructure variables on bond returns. The conditional robustness of some factors highlights the complexity of bond price dynamics and the need for careful inference. Future research could explore alternative models or additional robustness measures to further clarify these nuanced relationships and enhance the understanding of bond return drivers.

4.2.2 Credit Grade Subset

Further robustness checks were conducted by segmenting the data based on credit grades. The results indicate that while macroeconomic factors consistently influence bond returns across all credit grades, the relevance of market microstructure variables, liquidity measures, and certain bond characteristics varies. Market conditions, such as Treasury yields, and bond characteristics, like yield spreads, appear to be slightly more significant for high-yield bonds, likely reflecting their greater sensitivity to economic fluctuations.

Conversely, market microstructure factors, such as commissions and ATS usage, are more relevant for high-yield bonds, highlighting their role in trading efficiency and liquidity within lower credit quality segments. This likely reflects higher trading frictions and liquidity constraints in these markets. Meanwhile, CSP remains consistently significant with similar effect sizes across both bond segments, underscoring its pervasive influence on bond returns, regardless of credit quality.

Liquidity variables, such as daily trading volume, exhibit consistent yet nuanced effects across subsets, suggesting that greater market depth generally increases competition and tightens pricing. However, other liquidity measures vary by credit grade, reflecting distinct underlying liquidity conditions of each segment. Trading time shows conditional significance in the High-

Yield segment, while tightness becomes conditionally significant in the Investment-Grade segment. Additionally, resilience, measured by intraday return variance (Ranaldo 2002), is only significant for Investment-Grade bonds, likely due to higher trading activity, which enables a more reliable calculation of intraday return variance compared to the less frequently traded High-Yield bonds.

Bond-specific attributes, particularly coupon features, exhibit greater significance in high-yield markets. This suggests that in the high-yield segment, higher coupon rates are interpreted as signals of elevated risk, whereas in the investment-grade segment, they are perceived as indicators of greater attractiveness. Therefore, the influence of bond characteristics, such as coupon rates, interacts with credit quality, leading to differing effects on returns across market segments.

The robustness checks support the primary model's findings for H1. Macroeconomic factors, such as changes in Treasury yields and yield spreads, remain the dominant drivers of bond returns across both high-yield and investment-grade subsets, confirming their central role. While market microstructure variables retain significance, their effect sizes and importance vary across credit grades, consistent with the primary model’s findings that these factors play a secondary yet reliable role relative to macroeconomic and bond-level determinants. The stability of macroeconomic factors across subsets further underscores their foundational importance in explaining daily bond price dynamics. Additionally, the Wald test confirms that market microstructure variables are jointly significant in both subsets, highlighting their relevance within the model.

The analysis also robustly supports H2, which posited that the relevance of market microstructure factors is context-dependent, particularly in their interactions with market conditions and other variable groups. The interaction terms between commissions and trading frequency, as well as commissions and ATS usage, demonstrate that the impact of microstructure variables on returns is contingent upon other market conditions. These conditional effects highlight the dynamic interplay between different factors, reinforcing H2's claim that the influence of market microstructure is not static but varies with the broader trading environment and economic conditions. The subset analysis confirms that interaction effects vary in magnitude and significance across High Yield and Investment Grade bonds, highlighting their context-dependent nature. For instance, in the Investment Grade segment, ample liquidity likely keeps compensation mechanisms tightly optimized, leaving little room for dealers to adjust commissions to influence returns. In contrast, in the High Yield segment, where liquidity is more limited, the efficiency gains from electronic trading platforms help offset the return-reducing effects of higher commissions, as transaction costs are more pronounced.

However, caution is warranted when interpreting these subset results. The model explains a larger portion of the variance in Investment Grade bonds ($R^2 = 0.1036$) compared to High Yield bonds ($R^2 = 0.0741$), suggesting it is more effective in capturing the price dynamics of the former. While the F-statistics remain significant across subsets, their robustness depends on the covariance estimator, with only the HAC estimator yielding reliable coefficients in the High Yield segment. Additionally, certain interactions and variables lose significance in the subset models, indicating that the model's explanatory power may be limited by differences in liquidity and credit quality across segments. See **Appendix 4.2. Subset Analysis** for the detailed results.

While the regression model provides robust insights into the determinants of bond returns, it has potential limitations related to parameter selection, endogeneity, and variable transformations. These criticisms are thoroughly examined and addressed in detail in the **Appendix 5: Potential Criticism**.

5. Conclusion

This study advances the understanding of price dynamics in OTC corporate bond markets by evaluating the relative importance of market microstructure variables alongside traditional determinants, such as liquidity, macroeconomic conditions, and bond-specific characteristics. Using a dynamic IV-2SLS model and robust econometric techniques, the research reveals the complex interactions that shape daily bond returns in a decentralized and opaque trading environment. The key contribution lies in its integrative approach, assessing the joint and relative impacts of these factors within a unified framework. Findings highlight that macroeconomic factors, particularly deviations in the 10-year Treasury yield and yield spreads, are the primary drivers of daily bond price movements. Market microstructure and liquidity variables, while secondary, serve as crucial adjustment mechanisms. For practitioners, the results emphasize the need to monitor macroeconomic indicators and understand market microstructure dynamics for improved price prediction and risk management. Regulators can leverage these insights to enhance market transparency, reduce trading frictions, and promote overall market efficiency. This study builds on prior work emphasizing liquidity and macroeconomic conditions (e.g., Elton et al. 1999; Chen et al. 2007) but diverges by quantifying the substantial, context-dependent role of market microstructure variables. Extending earlier contributions by Edwards et al. (2007) and Ederington et al. (2015), this research quantitatively compares dealer influences with other key determinants and uncovers interaction effects between microstructure variables and market conditions. These findings underscore the dynamic and nuanced nature of short-term price formation, providing a more granular, day-to-day perspective. Future research could explore alternative identification strategies, refine parameter selection methods, and incorporate order book data to further enhance the generalizability and depth of the analysis.

References

Acharya, Viral, Yakov Amihud, and Sreedhar T. Bharath. 2013. "Liquidity Risk of Corporate Bond Returns: Conditional Approach." Journal of Financial Economics 110 (2): 358–386.

Annaert, Jan, Anouk G.P. Claes, Marc J.K. De Ceuster, and Hairui Zhang. 2024. Estimating the Yield Curve Using the Nelson-Siegel Model: A Ridge Regression Approach. Antwerpen: Universiteit Antwerpen.

Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." Econometrica 59 (3): 817–858. https://doi.org/10.2307/2938229

Baltagi, B.H., and L. Liu. 2020. "Forecasting with Unbalanced Panel Data." Journal of Forecasting 39: 709–724. https://doi.org/10.1002/for.2646.

Bessembinder, Hendrik, William Maxwell, and Krishna Venkataraman. 2006. "Market Transparency, Liquidity Externalities, and Institutional Trading Costs in Corporate Bonds." Journal of Financial Economics 82 (2): 251–288. https://doi.org/10.1016/j.jfineco.2005.10.002.

Bergault, Pierre, and Olivier Guéant. 2023. "Modeling Liquidity in Corporate Bond Markets: Applications to Price Adjustments." arXiv preprint arXiv:2309.04216. https://arxiv.org/pdf/2309.04216.

Bai, Jushan. 2009. "Panel Data Models with Interactive Fixed Effects." Econometrica 77: 1229–1279. https://doi.org/10.3982/ECTA6135.

Chen, Li, David A. Lesmond, and Jiwei Wei. 2007. "Corporate Yield Spreads and Bond Liquidity." The Journal of Finance 62: 119–149. https://doi.org/10.1111/j.1540-6261.2007.01203.x.

Cleveland, R.B., and W.S. Cleveland. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess." Journal of Official Statistics 6: 3–33.

Dick-Nielsen, Jens. 2009. "Liquidity Biases in TRACE." Journal of Fixed Income 19 (2). https://doi.org/10.3905/jfi.2009.19.2.043. Available at SSRN: https://ssrn.com/abstract=1424870 or http://dx.doi.org/10.2139/ssrn.1424870.

Dick-Nielsen, Jens, Peter Feldhütter, and David Lando. 2012. "Corporate Bond Liquidity before and after the Onset of the Subprime Crisis." *Journal of Financial Economics* 103 (3): 471–492.

Dick-Nielsen, Jens. 2014. "How to Clean Enhanced TRACE Data." Available at SSRN: https://ssrn.com/abstract=2337908 or http://dx.doi.org/10.2139/ssrn.2337908.

Duffee, Gregory R. 2011. "Information in (and Not in) the Term Structure." The Review of Financial Studies 24 (9): 2895–2934. https://doi.org/10.1093/rfs/hhr033.

Duffie, Darrell, Nicola Gârleanu, and Lasse Heje Pedersen. 2005. "Over-the-Counter Markets." Econometrica 73: 1815–1847. https://doi.org/10.1111/j.1468-0262.2005.00639.x.

Edwards, A.K., L.E. Harris, and M.S. Piwowar. 2007. "Corporate Bond Market Transaction Costs and Transparency." The Journal of Finance 62: 1421–1451. https://doi.org/10.1111/j.1540-6261.2007.01240.x.

Ederington, Louis, Wei Guan, and Pradeep K. Yadav. 2015. "Dealer Spreads in the Corporate Bond Market: Agent vs. Market-Making Roles." CFR Working Papers 15-11. University of Cologne, Centre for Financial Research (CFR). https://ideas.repec.org/p/zbw/cfrwps/1511.html.

Elton, Edwin J. and Gruber, Martin J. and Agrawal, Deepak and Mann, Christopher, Explaining the Rate Spread on Corporate Bonds (September 1999). NYU Working Paper No. FIN-99-082, Available at SSRN: https://ssrn.com/abstract=1299452

Friewald, Nils, Rainer Jankowitsch, and Marti G. Subrahmanyam. 2012. "Illiquidity or Credit Deterioration: A Study of Liquidity in the US Corporate Bond Market During Financial Crises." Journal of Financial Economics 105 (1): 18–36. https://doi.org/10.1016/j.jfineco.2012.02.001.

Hendershott, Tom, and Avanidhar Madhavan. 2015. "Click or Call? Auction versus Search in the Over-the-Counter Market." The Journal of Finance 70: 419–447. https://doi.org/10.1111/jofi.12164.

Moon, Hyungsik, and Martin Weidner. 2017. "Dynamic Linear Panel Regression Models with Interactive Fixed Effects." Econometric Theory 33 (1): 158–195.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." Econometrica 55 (3): 703–708. https://doi.org/10.2307/1913610.

Petersen, Mitchell A. 2009. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." The Review of Financial Studies 22 (1): 435–480. https://doi.org/10.1093/rfs/hhn053.

Ranaldo, Angelo. 2001. "Intraday Market Liquidity on the Swiss Stock Exchange." Financial Markets & Portfolio Management 15 (3). Posted March 11, 2002.

Saunders, Anthony, Anand Srinivasan, and Ingo Walter. 2002. "Price Formation in the OTC Corporate Bond Markets: A Field Study of the Inter-Dealer Market." Journal of Economics and Business 54: 95–113. https://doi.org/10.1016/S0148-6195(01)00055-8.

Schneider, Michael T. 2018. "Market Microstructure, Price Impact, and Liquidity in Fixed Income Markets." Scuola Normale Superiore. http://ricerca.sns.it/bitstream/11384/85739/1/thesis_schneider.pdf.

Securities Industry and Financial Markets Association. 2024. Capital Markets Fact Book July 2024. Accessed December 16, 2024. https://www.sifma.org/resources/research/statistics/fact-book/.

Young Ho Eom, Jean Helwege, and Jing-Zhi Huang. 2004. "Structural Models of Corporate Bond Pricing: An Empirical Analysis." The Review of Financial Studies 17 (2): 499–544. https://doi.org/10.1093/rfs/hhg053.

Appendix 1: Corporate Bonds

U.S. corporate bonds, as delineated by the Securities Industry and Financial Markets Association (SIFMA, 2024), are debt securities issued by both public and private corporations to raise capital for investment initiatives and expansion plans. These bonds are inherently riskier than U.S. Treasury securities, prompting credit rating agencies to assign ratings that reflect the issuers' creditworthiness and the probability of debt repayment (SIFMA, 2024). Corporate bonds are broadly categorized into several types based on their issuance and credit quality. Publicly traded bonds are registered with the Securities and Exchange Commission (SEC) and are traded on public exchanges, providing greater transparency and liquidity. 144A securities refer to bonds issued under Securities Act Rule 144A, allowing the sale of unregistered bonds to qualified institutional buyers under specific conditions, thereby facilitating more flexible financing options (SIFMA, 2024).

Further classification divides corporate bonds into high yield and investment grade categories. High yield bonds are rated below BBB by credit rating agencies, indicating a higher risk of default but offering higher interest rates to compensate investors for the increased risk. In contrast, investment-grade bonds receive ratings of BBB or higher, signifying lower default risk and typically offering lower yields compared to high yield counterparts (SIFMA, 2024).

In addition to these classifications, corporate bonds exhibit various structural features that cater to different investor preferences and issuer needs. Fixed-rate bonds provide consistent interest payments throughout their term, offering predictability to investors. Floating-rate bonds have interest rates that adjust periodically based on benchmark rates such as the Federal Funds Rate or the Secured Overnight Financing Rate (SOFR), thereby mitigating interest rate risk for both issuers and investors (SIFMA, 2024). Callable bonds grant issuers the option to redeem the bonds before their maturity date, typically at a premium, allowing issuers to refinance debt if interest rates decline. Conversely, non-callable bonds do not offer this redemption feature, providing greater security to investors regarding the bond's lifespan (SIFMA, 2024). Additionally, convertible bonds offer bondholders the option to convert their bonds into a predetermined number of the issuer's equity shares, blending features of debt and equity and providing potential upside benefits if the issuer's stock performs well (SIFMA, 2024).

These diverse characteristics make U.S. corporate bonds a complex yet essential component of the fixed income market. They cater to a wide range of investment strategies and risk appetites, from conservative investors seeking stable returns through investment-grade bonds to more aggressive investors aiming for higher yields via high yield bonds or potential equity upside through convertible bonds. Understanding these distinct features is crucial for analyzing price dynamics and the role of market microstructure within the OTC corporate bond market.

Appendix 2: Statistical Tests

Appendix 2.1: Overview Test Statistics

Test statistics	Full model	Grade subsets	
		High Yield	Investment Grade
Durbin-Wu-Hausman Test			
Test Stat.	6619	10178	3406
P-value	0,0000	0,0000	0,0000
First Stage F-Stat.			
L1_log_daily_return	915	148	2036
log_volatility_squared_dev	30791	7783	18976
First Stage R-squared	0,4190	0,418	0,363
Breuasch-Pagan Test			
LM Statistic	207854	51281	160171
LM p-value	0,0000	0,0000	0,0000
F-Statistic	8612	2204	6904
F-Statistic (p-value)	0,0000	0,0000	0,0000
Ljung-Box Test			
Q-Statistic	4,17E+10	1,69E+09	1,99E+10
p-value	0,0000	0,0000	0,0000
Wald test			
H0: Linear equality contraint is valid			
Statistic	2793	941	5460
P-value	0,0000	0,0000	0,0000
Distributed	chi2(4)	chi2(4)	chi2(4)

Table 2: Test Statistics Overview

Since the model is directly identified, with the number of instruments exactly matching the number of endogenous variables, it is not overidentified. Sargan Tests and Hansen J Tests are designed to evaluate the validity of instruments in overidentified models where there are more instruments than necessary. These tests rely on having additional instruments to test for potential overidentifying restrictions. In a directly identified model, there are no extra instruments available to perform such tests, making Sargan and Hansen J Tests inapplicable. Therefore, these tests cannot be used in this context.

Appendix 2.2: ACF/PACF

Full Model

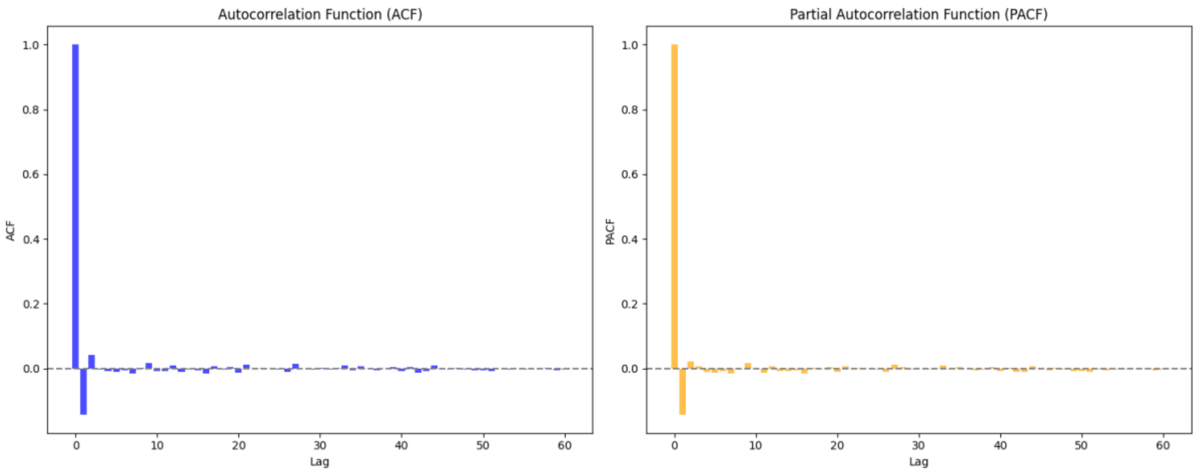

Figure 1: ACF/PACF Full Model

Grade Subset: High Yield

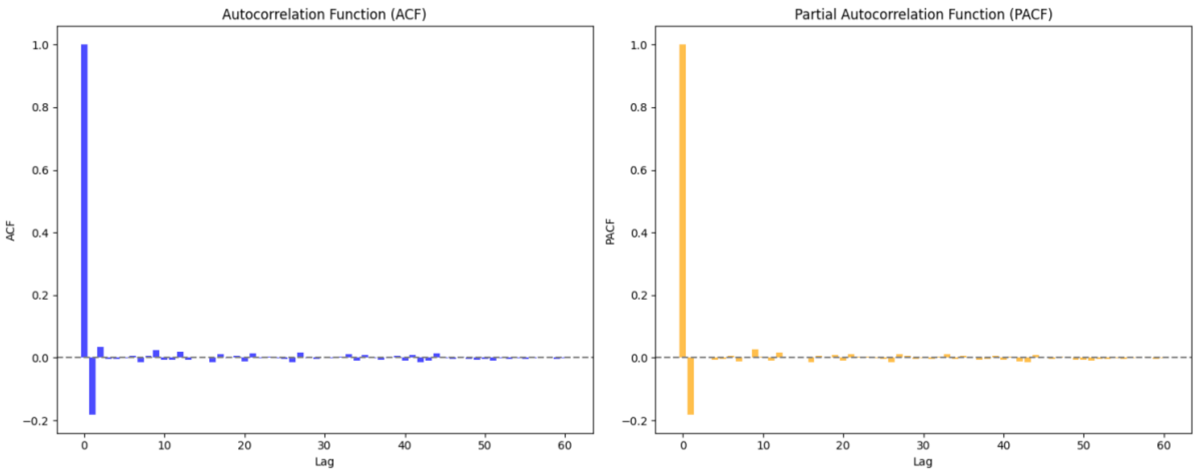

Figure 2:ACF/PACF Grade Subset - High Yield

Grade Subset: Investment Grade

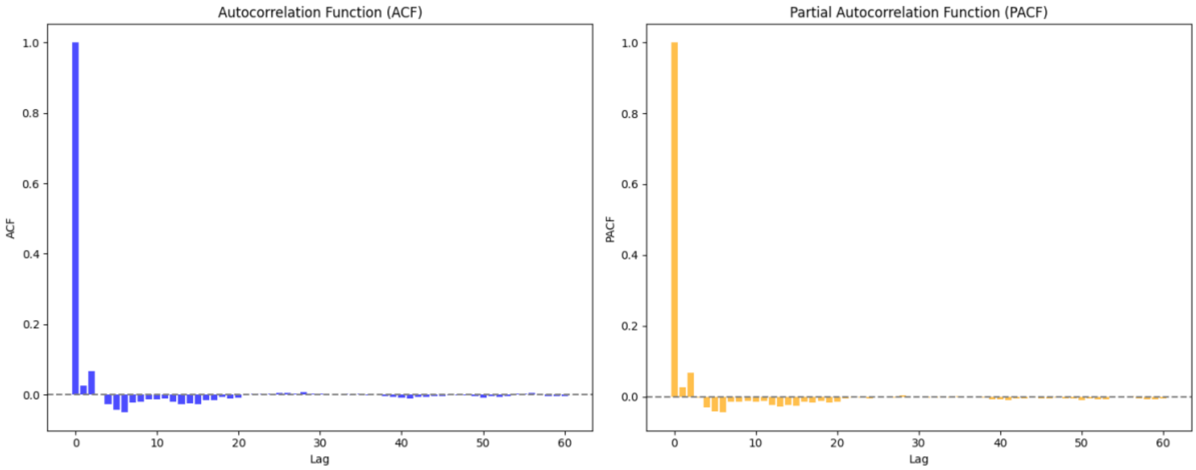

Figure 3: ACF/PACF Grade Subset - Investment Grade

Appendix 2.3: Variance Inflation Factor (VIF)

Index	Variable	VIF
0	Commission	1.809066
1	ATS Usage	21.214063
2	Customer Flow	1.367469
3	Customer Selling Pressure	1.055912
4	Δ 10Y Treasury Yield	1.023771
5	Δ Market Volatility	12.007103
6	Δ Treasury Yield Curvature (TYC)	1.028965
7	Δ Time-to-maturity (TTM)	1.389776
8	Coupon Rate	1.436800
9	Credit Grade	1.326280
10	Δ Yield Spread	1.000516
11	Zero Coupon Bond Flag	1.019472
12	Floating Coupon Bond Flag	1.295905
13	Other Coupon Bond Flag	1.019155
14	Δ Daily Trading Volume	1.279214

15	Δ Average Trading Inactivity	1.096491
16	Δ Daily Trading Frequency (DTF)	12.842232
17	Δ Price Spread	1.378139
18	Δ Intraday Return Variance	1.000528
19	Rolling 5-day Bond Price Volatility	1.004386
20	Δ 10Y Treasury Yield – TYC Interaction	1.112488
21	Δ 10Y treasury Yield – TTM Interaction	1.000095
22	Commission – Market Volatility Interaction	11.924136
23	ATS – DTF Interaction	2.261720
24	Commission – DTF Interaction	12.392693
25	Customer Flow – Price Spread Interaction	1.253019
26	Commission – ATS Interaction	23.155679
27	CSP – Daily Trading Volume Interaction	1.045236

Table 3: VIF of Model Paramters

Appendix 2.4: Model Diagnostics (Residual analysis)

Full Model

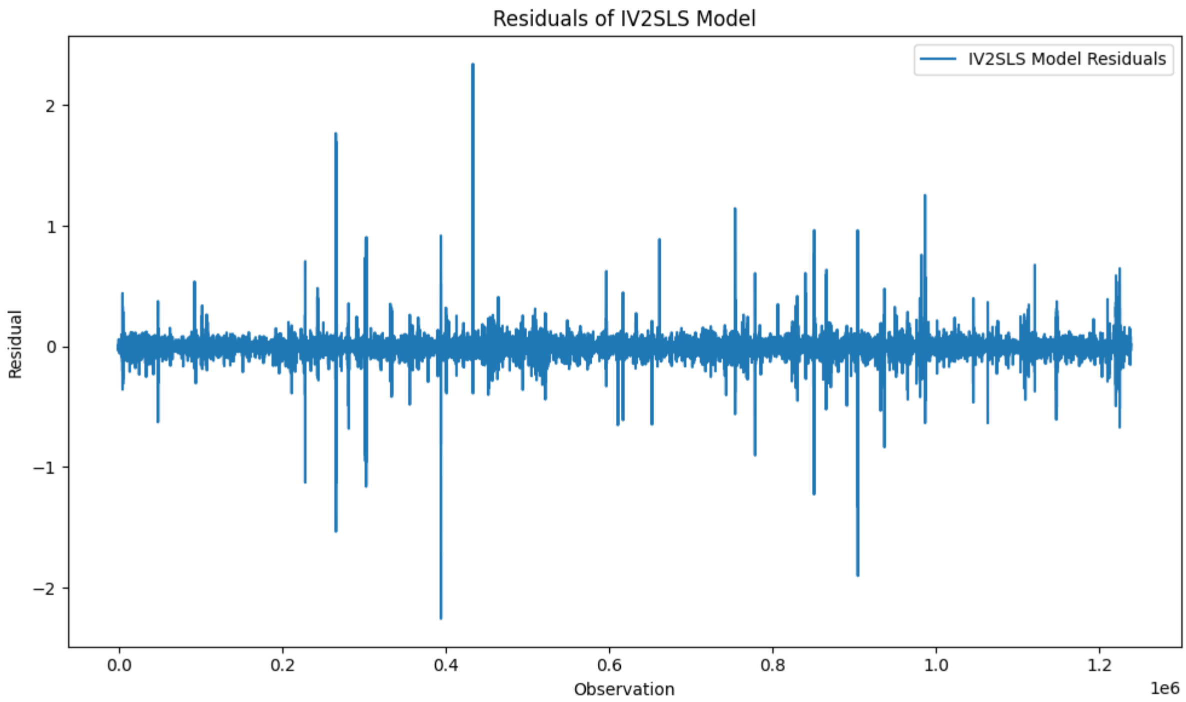

Figure 4: Model Diagnostics - Time Series Residual Plot - Full Model

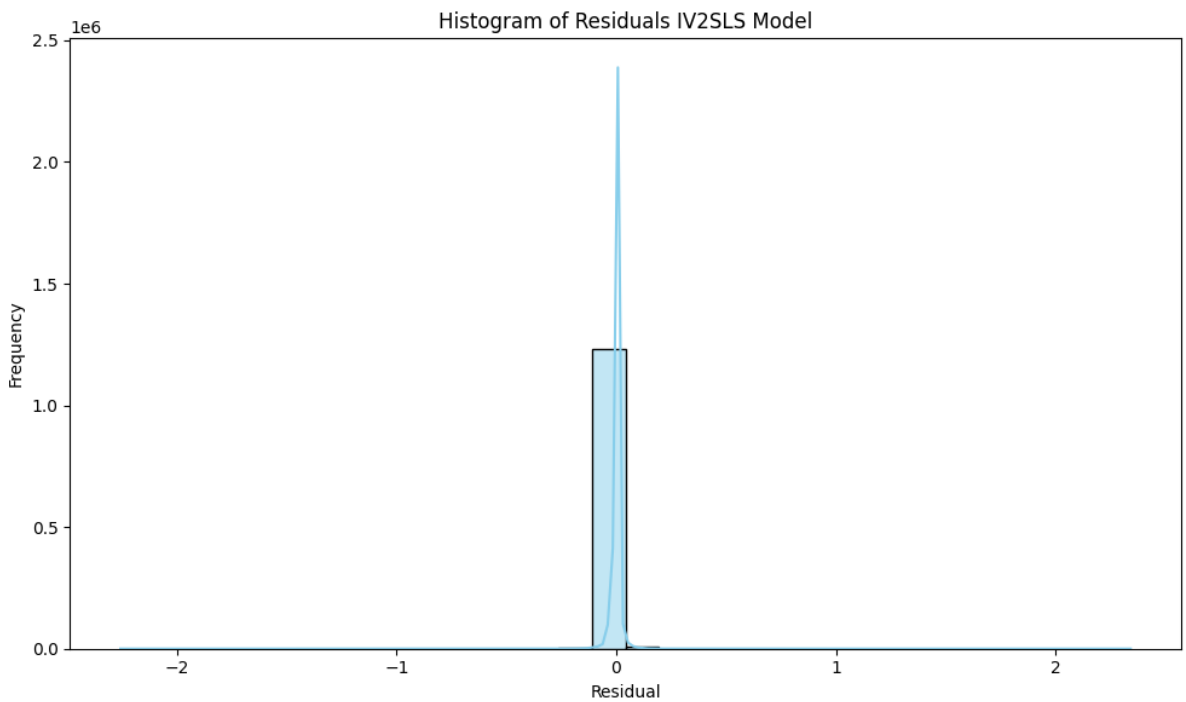

Figure 5: Model Diagnostics - Residual Histogram - Full Model

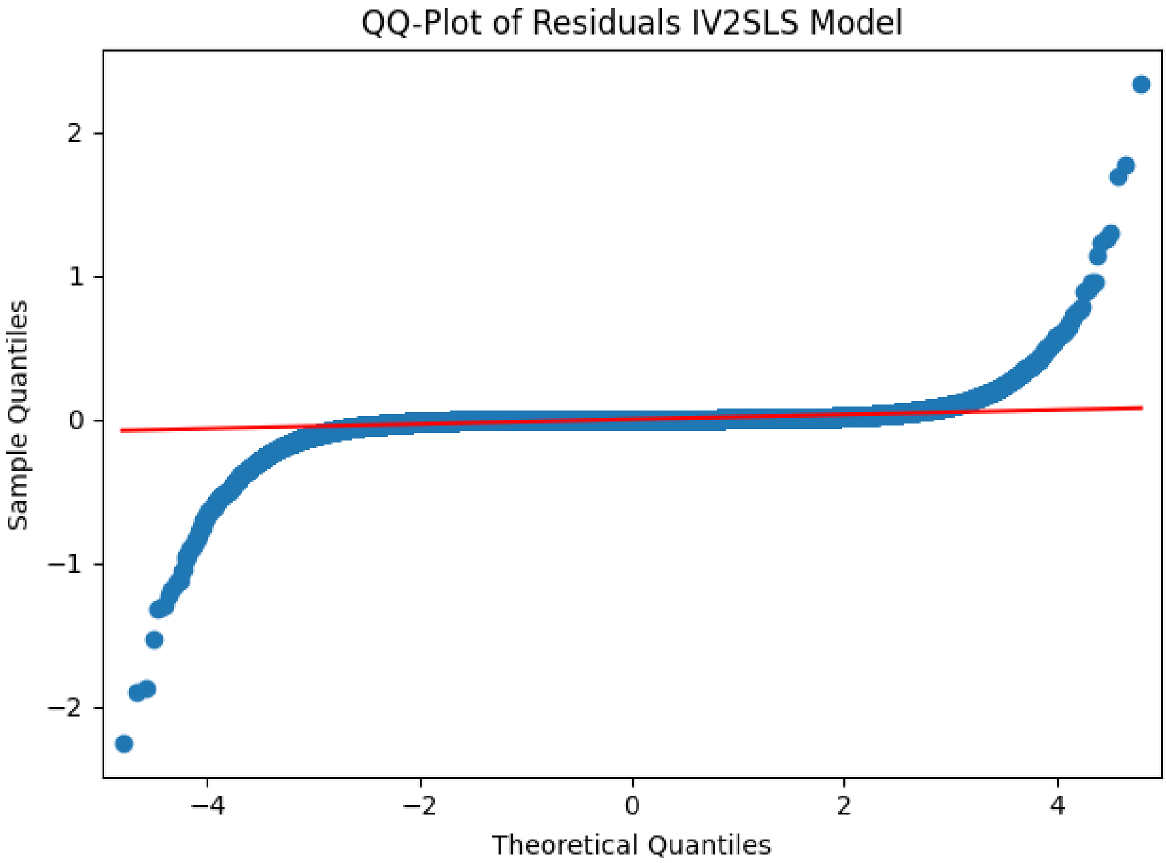

Figure 6: Model Diagnostics - Residuals QQ-Plot - Full Model

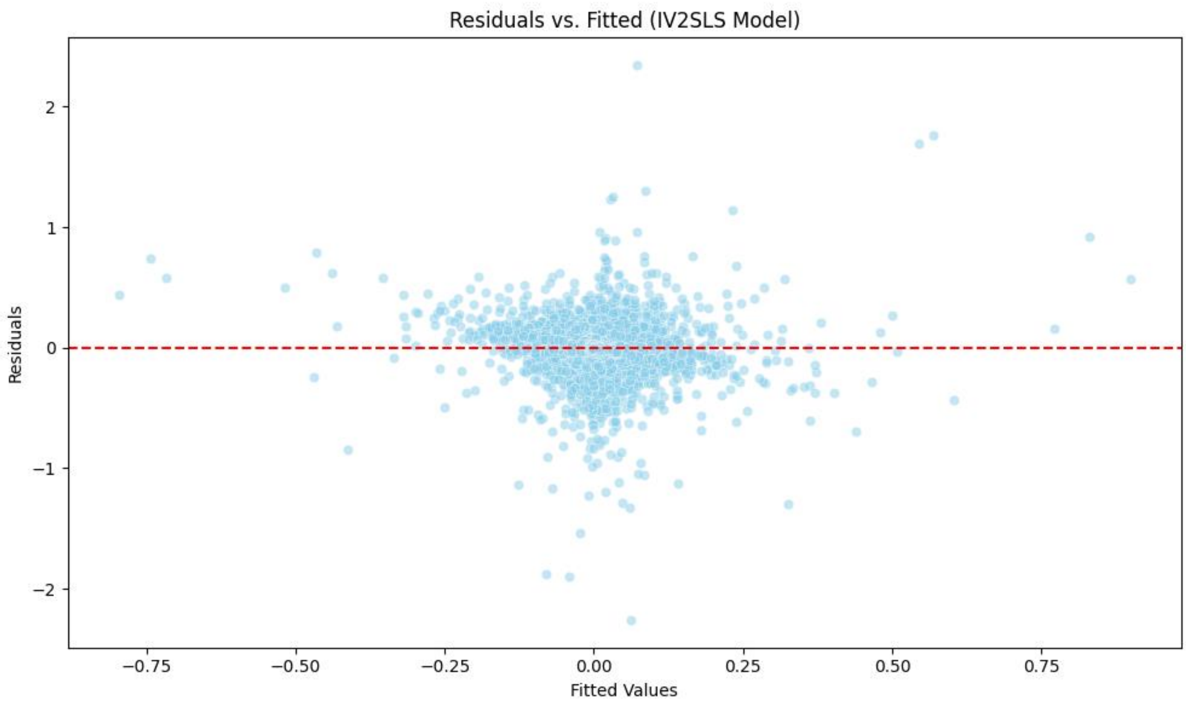

Figure 7: Model Diagnostics - Residuals vs. Fitted Values - Full Model

Grade Subset: High Yield

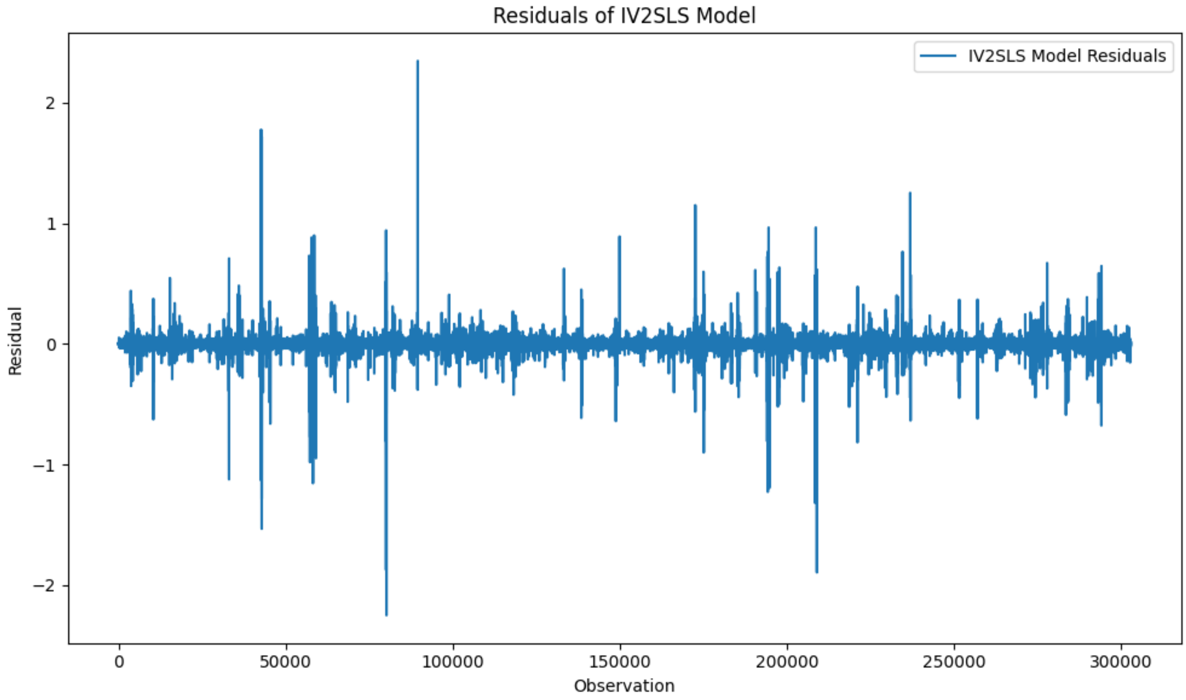

Figure 8: Model Diagnostics - Time Series Residual Plot - Grade Subset (High Yield)

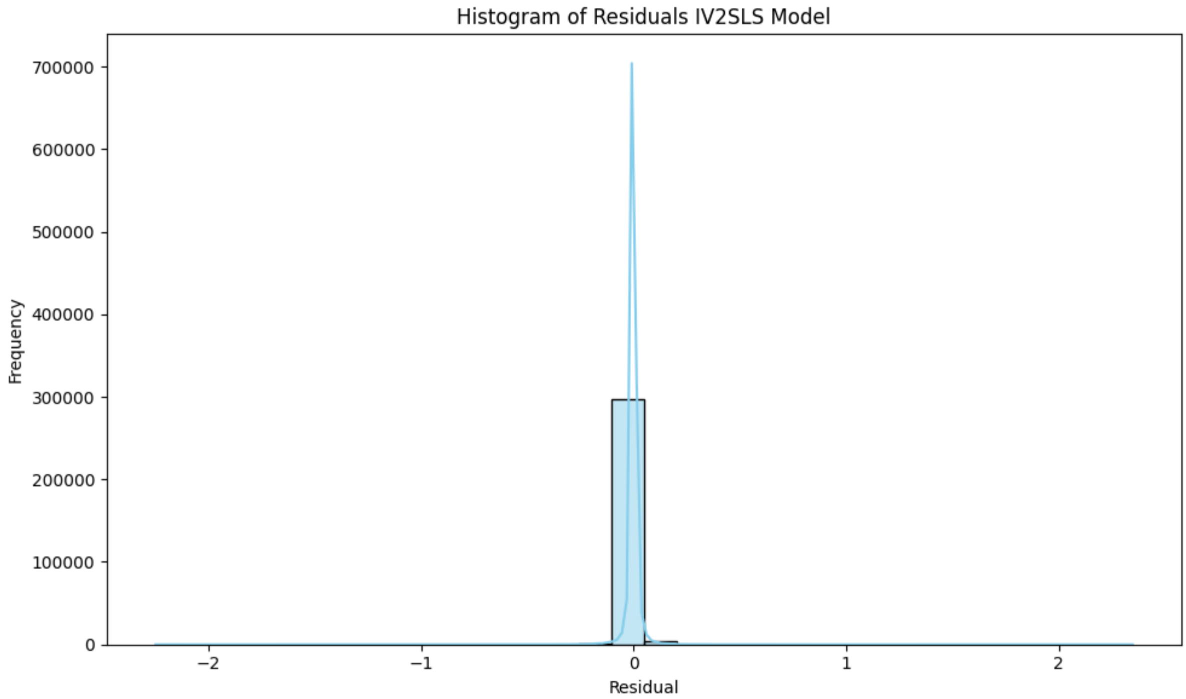

Figure 9: Model Diagnostics - Residual Histogram - Grade Subset (High Yield)

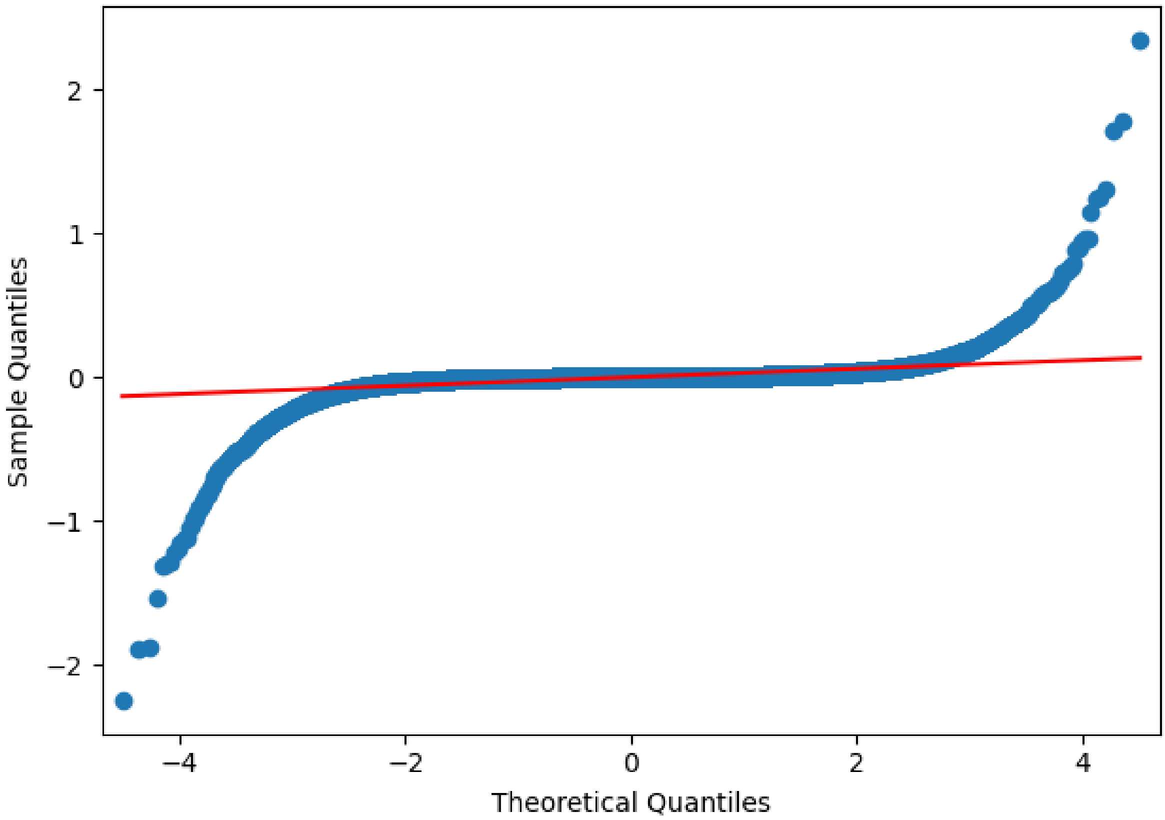

Figure 10: Model Diagnostics - Residuals QQ-Plot - Grade Subset (High Yield)

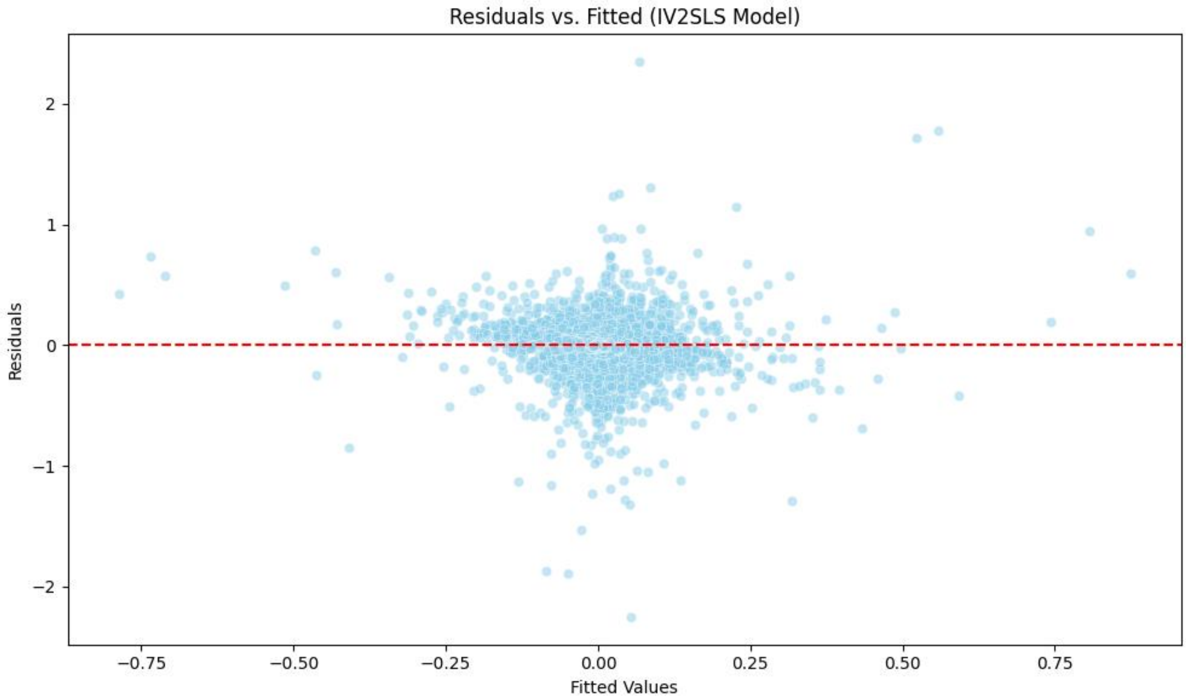

Figure 11: Model Diagnostics - Residuals vs. Fitted Values - Grade Subset (High Yield)

Grade Subset: Investment Grade

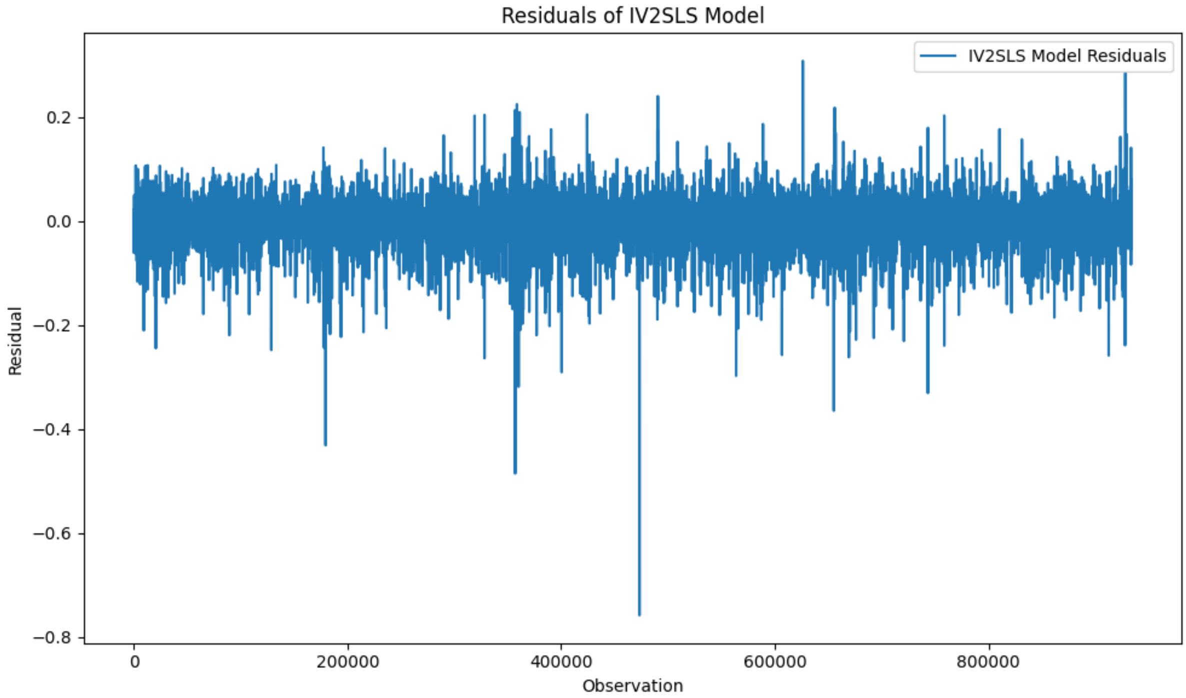

Figure 12: Model Diagnostics - Time Series Residual Plot - Grade Subset (Investment Grade)

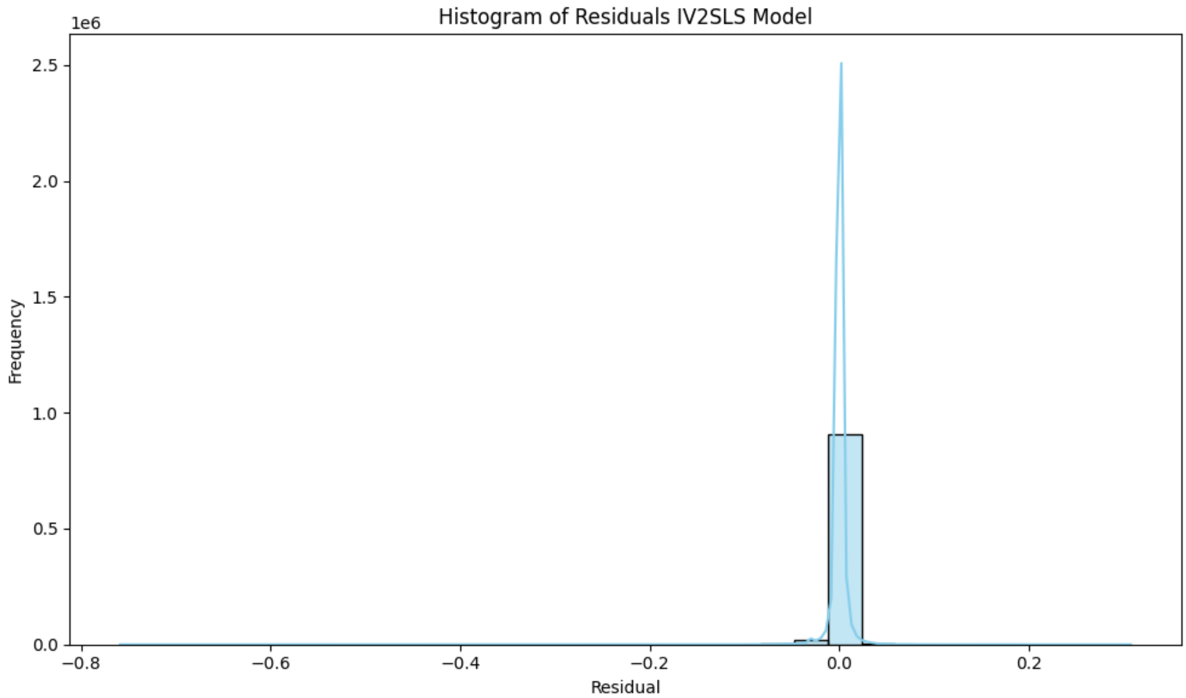

Figure 13: Model Diagnostics - Residual Histogram - Grade Subset (Investment Grade)

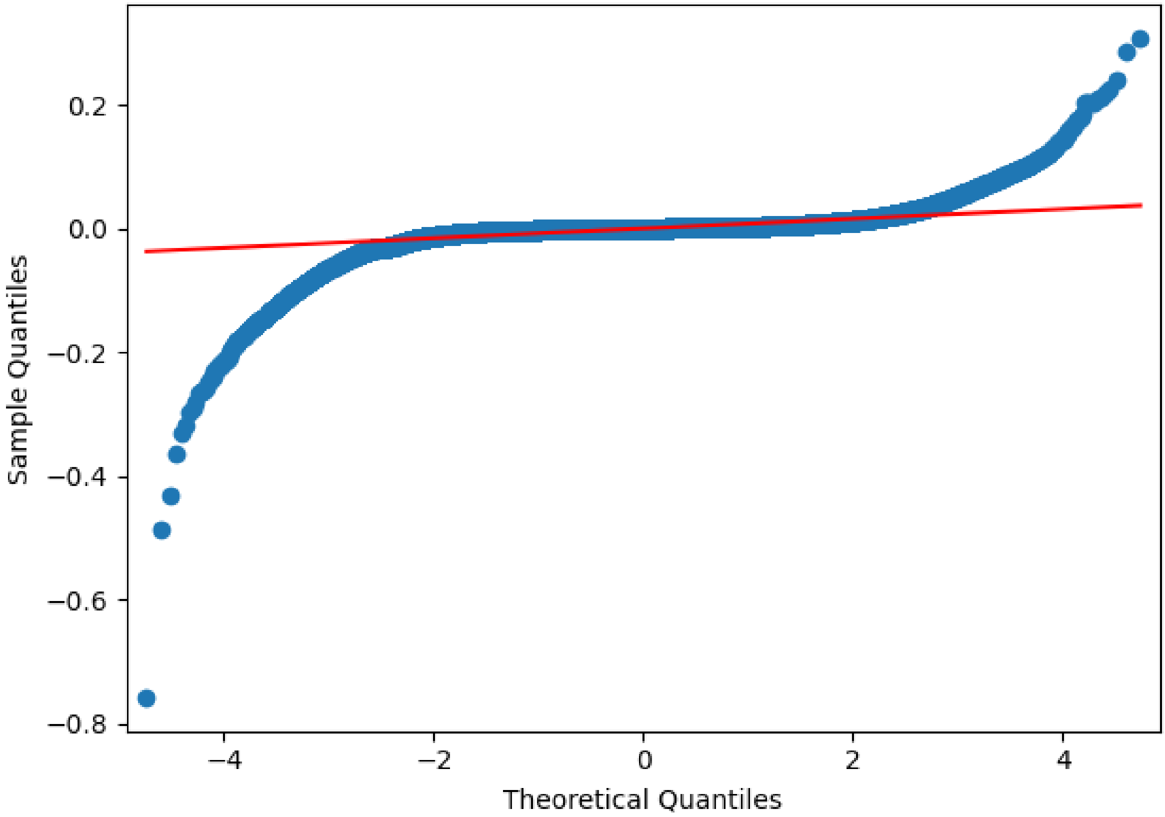

Figure 14: Model Diagnostics - Residuals QQ-Plot - Grade Subset (Investment Grade)

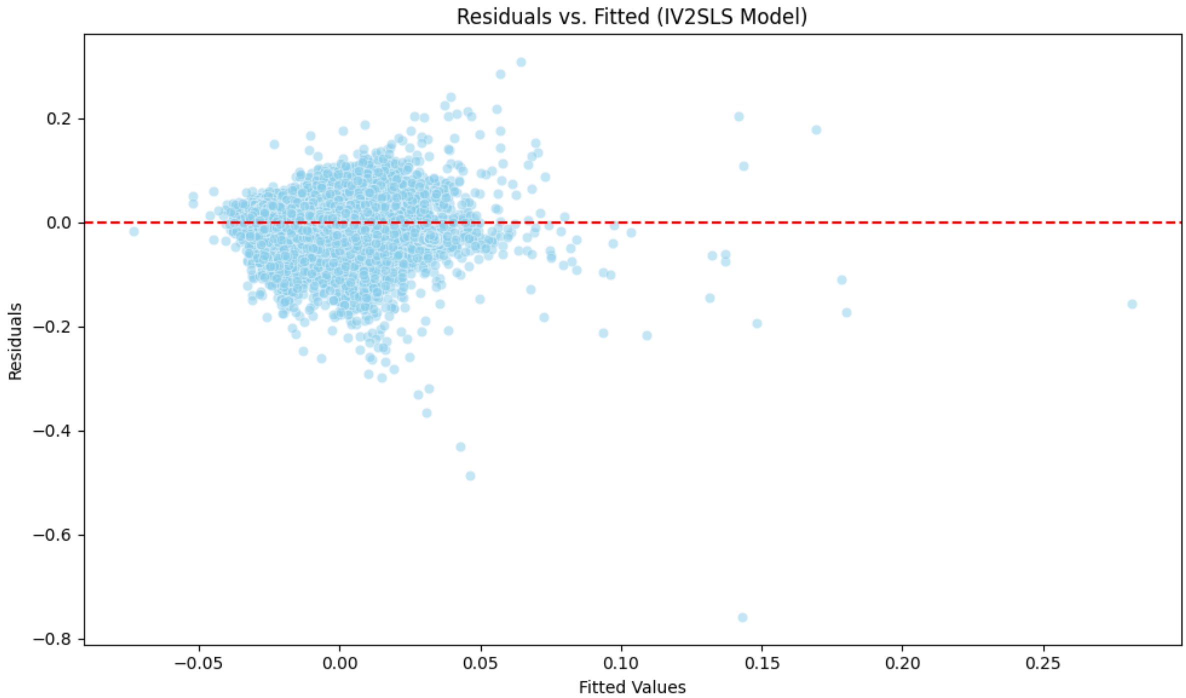

Figure 15: Model Diagnostics - Residuals vs. Fitted Values - Grade Subset (Investment Grade)

Appendix 3: Model Overview

Appendix 3.1: Data

Filtering Steps

- Post-aggregation: Unfiltered: 1.626.016 observations
- After filtering out bonds (cusip_id) with less than 100 observations to ensure appropriate and meaningful lagged variables and demeaning: 1.257.088 observations
- After filtering out trade days (trade_date) with less than 1000 bonds in their cross sections to ensure meaningful cross sectional averaging: 1.257.796 observations

Cross-Sectional Size

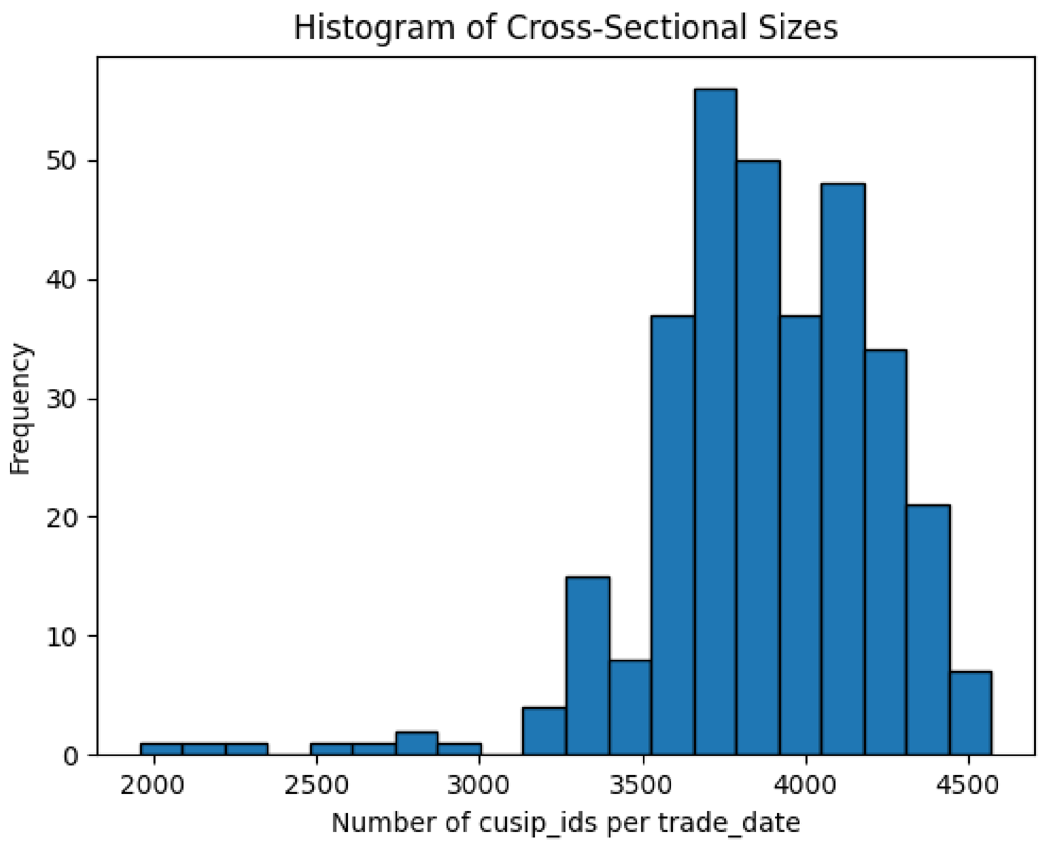

Figure 16: Histogram of Cross-Sectional Sizes - Daily Number of distinct Bonds

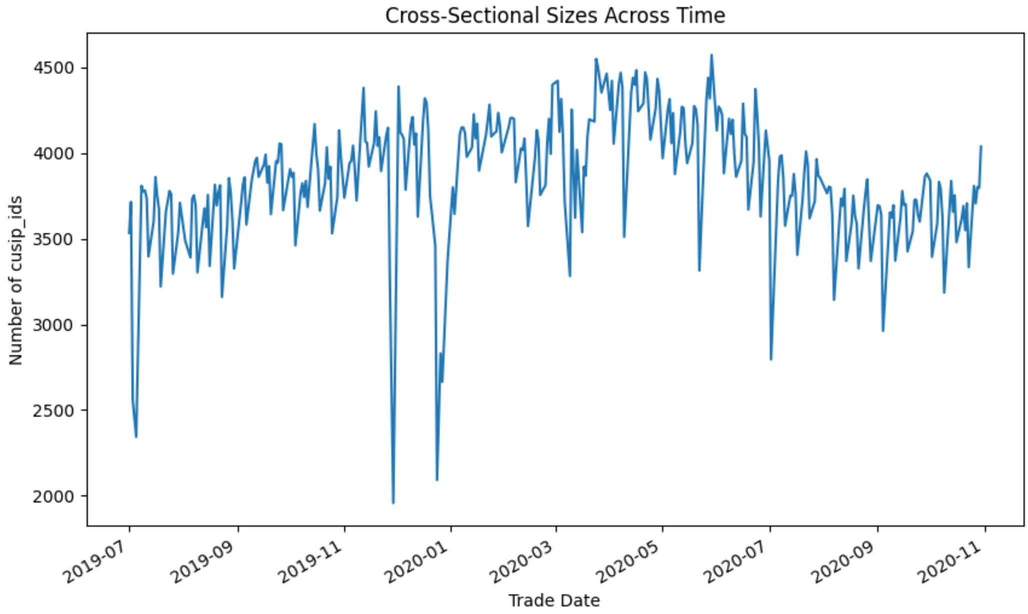

Figure 17: Time series plot of Cross-Sectional Sizes - Daily Number of distinct Bonds

Appendix 3.2: Regression Model

Regression Model Formula

$$
\begin{aligned}
R_{i,t} = \alpha + \beta_1(R_{i,t-1}) + \beta_2[D(Vol_{i,t}) - C(Vol_{i,t})] \\
&+ \sum_m \gamma_m [D(Microstructure_{m,i,t}) - C(Microstructure_{m,i,t})] \\
&+ \sum_l \delta_l [D(Liquidity_{l,i,t}) - C(Liquidity_{l,i,t})] + \sum_b \mu_b C(BondChar_{b,i,t}) \\
&+ \sum_s \nu_s D(MarketShock_{s,i,t}) + \beta_3 [D(ATS_{i,t} * COMMISSION_{i,t}) \\
&- C(ATS_{i,t} * COMMISSION_{i,t})] \\
&+ \sum_{m,l} \lambda_{m,l} [D(Microstructure_{m,i,t} * Liquidity_{l,i,t}) \\
&- C(Microstructure_{m,i,t} * Liquidity_{l,i,t})] \\
&+ \sum_{m,s} \eta_z D(Microstructure_{m,i,t} * Market_{s,i,t}) + \\
&+ \sum_{s,s} \rho_s D(Market_{s,i,t} * Market_{s,i,t}) + \varepsilon_{i,t}
\end{aligned}
$$

Let $R_{i,t}$ denote the daily bond price return of bond i on day t. I consider several groups of explanatory variables: (1) the lagged dependent variable (the 1 day lag of daily bond price returns) and the 5 day rolling bond price volatility (both treated as endogenous), (2) market microstructure variables, (3) liquidity proxies, (4) bond characteristics, (5) market condition (shock) variables, and interaction terms between these sets of variables. Each group undergoes specific transformations to address unobserved heterogeneity and to void perfect collinearity.

Let $D(X_{i,t}) = X_{i,t} - \bar{X}_i$ denote the entity-level demeaned version of variable X where $\bar{X}_\iota$ is the time-average for bond i . Let $C(X_{i,t}) = \bar{X}_t$ denote the cross-sectional average at time t , where $\overline{X_t} = \frac{1}{N}\sum_{i=1}^{N} X_{i,t}$. When both transformations are applied to the same variable, the resulting transformed regressor is $D(X_{i,t}) - C(X_{i,t})$. Note on instrumentation: $R_{i,t-1}$ and $Vol_{i,t}$ are instrumented by their second lags, i.e. $R_{i,t-3}$ and $Vol_{i,t-2}$.

Appendix 4: Detailed Regression Analysis

Appendix 4.1: Detailed Model Regression Results

After estimating the full model using a range of transformed and differenced explanatory variables, I applied a ranking methodology to assess the significance and magnitude of coefficients. Each regressor and group of regressors was ranked in ascending order based on their relevance. This straightforward approach enables a quick evaluation of the importance of individual variables and variable groups. The results show that while market microstructure variables are consistently statistically significant, their overall economic impact and relative ranking remain more moderate compared to key macroeconomic and bond-level drivers, indicating their role as secondary determinants.

Market shocks

An analysis of market shock determinants reveals that deviations in the daily change of the 10-year Treasury yield are among the strongest predictors of bond price movements. The negative coefficient indicates that when the 10-year yield rises above its average level, bond returns tend to decline relative to their mean. For instance, a one basis point (0.01%) increase in the yield beyond its typical level corresponds to a slight drop in daily bond returns. This outcome aligns with expectations, as bond prices and yields are inversely related. While the individual impact may be modest, these deviations cumulatively exert a significant influence, underscoring the role of short-term benchmark yield fluctuations in shaping daily bond pricing.

While deviations in yield curve curvature are also statistically significant, their effect is more complex due to the cubic transformation applied. The coefficient of -0.0006 captures the non-linear relationship between curvature changes and bond returns. This transformation accounts for asymmetric effects, which can occur during bull steepening or bear steepening scenarios. The negative coefficient suggests that significant deviations in yield curve curvature—whether upward or downward—are associated with lower daily bond returns relative to their means.

This implies that markets react negatively to pronounced shifts in the yield curve shape, likely reflecting increased uncertainty and risk.

To better capture these dynamics, the interaction term between deviations in the daily change of the 10-year Treasury yield and the cubic-transformed yield curve curvature offers valuable insights. The significant coefficient of -0.001 indicates that simultaneous deviations in the 10-year yield and curvature amplify the negative impact on bond returns. In a bear steepening scenario—where long-term rates rise above their average levels, steepening the yield curve—the interaction term intensifies the decline in bond returns, reflecting market concerns about rising long-term interest rates and their adverse effect on bond prices. Conversely, in a bull steepening scenario—where short-term yields fall below their average levels, also steepening the curve—the interaction term may moderate or counteract the typically positive impact of lower short-term rates. Overall, the negative interaction coefficient highlights the complexity introduced by the yield curve's shape, as it influences how deviations in short-term and long-term rates collectively affect bond returns.

The interaction term reveals that deviations in yield level shifts (10-year Treasury) and changes in the yield curve shape are interdependent rather than purely additive. This underscores the importance of considering both the magnitude and direction of yield deviations when assessing their impact on bond returns. While deviations in the 10-year Treasury yield alone account for much of the daily return variation, the interaction term highlights the market's sensitivity to broader yield curve dynamics, particularly in steepening environments. This suggests that while yield curve shape changes do influence bond returns, their effect is more nuanced and generally less dominant than benchmark yield deviations.

Changes in volatility appear to matter only to a limited extent, revealing that the market's underlying price formation processes may place more weight on core yield deviations than on volatility shocks. The effect of market volatility is not significant across estimation methods and might be due to randomness.

Market microstructure

An examination of market microstructure variables reveals that most metrics exhibit statistically significant and economically meaningful relationships with daily bond returns. These effects are measured as deviations relative to bond-specific and time-specific averages, emphasizing the role of short-term shocks and anomalies in market microstructure.

Commissions demonstrate a nuanced role in price formation. The positive and statistically significant coefficient indicates that a 1% increase in the share of commission-incorporated trades—relative to their average levels—leads to a slight increase of 0.0045% in daily bond returns. This suggests that higher transaction costs may be embedded in trade prices, effectively raising the bond's price and, consequently, its calculated day-on-day return. Alternatively, it may reflect a compensatory mechanism where dealers receive higher commissions for providing liquidity during periods of market stress, with bond prices adjusted to offset these increased costs. Despite its statistical significance, the effect size remains modest, indicating that while commissions influence price formation, their impact is secondary to broader macroeconomic yield changes or intrinsic bond characteristics.

The interaction between commissions and daily trading frequency shows a negative and statistically significant coefficient, robust under both clustered and HAC standard errors. This indicates that the combined effect of higher commissions and increased trading frequency results in a more pronounced decline in daily bond returns. Specifically, when both commission-incorporated trades and trading frequency rise simultaneously, bond returns decrease by approximately 0.04% for each unit increase in the interaction term relative to its average levels. This relationship may suggest that frequent trading in such scenarios involves bonds with lower-than-average embedded commissions, leading to less inflated bond prices.

Alternatively, it could reflect a liquidity compensation mechanism: when liquidity is ample, dealers require less compensation for providing liquidity, reducing the need to embed commissions into bond prices.

The interaction between commissions and ATS usage shows a positive and highly significant coefficient, indicating an additive positive effect on daily bond returns when both factors increase above their average levels. Specifically, a one-unit increase in the interaction term corresponds to an approximate 0.000666% rise in bond returns. This suggests a synergistic relationship where the efficiency gains from ATS usage help offset the higher transaction costs associated with increased commissions, leading to a net positive impact on returns.

The interaction between commissions and market volatility exhibits a negative coefficient that is marginally significant under clustered standard errors but highly significant under HAC robust standard errors. This suggests that when market volatility rises above average levels, the presence of commission-incorporated trades amplifies the negative impact on bond returns by approximately 0.15% for each unit increase in the interaction term. This effect may reflect the heightened sensitivity of bond prices to transaction costs during periods of elevated volatility, exerting greater downward pressure on returns. Alternatively, it could indicate that increased volatility coincides with higher-than-average explicit commissions, as dealers demand greater compensation for providing liquidity in more uncertain market conditions.

ATS Usage also shows a positive and significant relationship with daily bond returns. The coefficient indicates that a one-unit increase in the logit-transformed share of trades executed via Alternative Trading Systems—relative to average levels—corresponds to an approximate 0.0083% increase in returns. This likely reflects the enhanced trading efficiency and improved price discovery facilitated by electronic platforms. Greater ATS usage promotes transparency and tighter bid-ask spreads, reducing transaction costs and improving execution quality for both dealers and investors. However, the modest effect size suggests that while ATS usage refines the price formation process, it remains secondary to the broader determinants of bond returns.

The interaction between ATS usage and daily trading frequency yields a negative coefficient, though it is not statistically significant under either estimation method. This suggests that when ATS usage is above average and trading frequency is simultaneously elevated, bond returns tend to decrease slightly. However, the lack of statistical significance indicates that this relationship remains inconclusive. One possible interpretation is that beyond a certain threshold, higher trading activity may erode the marginal benefits of ATS platforms, potentially due to cumulative transaction costs or diminishing informational advantages.

Customer Trade Flow adds a more nuanced and insightful layer to the analysis. The coefficient indicates that a one-unit increase in the logit-transformed share of trades between customers and dealers—relative to average levels—corresponds to an approximate 0.0011% increase in bond returns. Practically, this means that a 1% rise in customer participation (or a 1% decline in inter-dealer trades) is associated with a 0.0011% increase in daily bond returns. This relationship can be interpreted in two ways. Greater customer participation may reflect stronger demand from end investors, driving bond prices—and consequently, returns—higher. Conversely, a decline in inter-dealer trades could signal reduced competitive pressure among dealers, enabling tighter pricing and contributing to higher bond returns.

Additionally, this metric acts as a proxy for the balance between customer-driven and dealer-driven trading activity. A higher rate of customer participation may suggest that bonds are being traded more actively with investors rather than among dealers, potentially signaling broader market confidence or shifts in trading strategies. Conversely, a prevalence of inter-dealer trades could reflect more subdued market conditions or increased inventory management by dealers.

However, the significance of this coefficient varies depending on the estimation method, indicating sensitivity to the assumed error structure. Specifically, while the coefficient is not significant under clustered standard errors it attains significance under HAC robust standard errors. This variability suggests that the relationship between customer trade flow deviations and bond returns is less robust compared to other microstructure variables. The limited and conditional influence of customer trade flow implies that, although increased customer-dealer interactions can contribute to minor price adjustments, they do not play a dominant role in determining daily bond returns.

The interaction between Customer Flow and price spread exhibits a negative coefficient, with significance levels at the borderline. This suggests that increased customer flow combined with wider price spreads may lead to a more substantial decrease in bond returns – relative to average levels- by approximately 0.00905% for each unit increase in the interaction term. This could indicate that when customer participation is high alongside wider spreads, the cost of trading increases, thereby negatively impacting returns. This could also imply that when price spreads increase above average levels which might be due to less liquidity in the market or higher than average intraday price volatility, caused by decreasing inter-dealer activity or increasing customer flow, daily returns tend to suffer.

Customer Selling Pressure (CSP), exhibits the most robust and consistent relationship with bond returns. The negative and highly significant coefficient indicates that a one-unit increase in the logit-transformed percentage of sell trades relative to buy trades—beyond average levels—correlates with a 0.02% decrease in daily bond returns. This relationship aligns with intuitive expectations: heightened selling pressure typically signals increased liquidity demands, compelling dealers to absorb inventory risk by adjusting bond prices downward. The relatively larger magnitude of this effect effectively captures real-time liquidity frictions and supply-demand imbalances more directly than commissions or ATS usage.

The interaction between Customer Selling Pressure and daily trading volume presents a positive coefficient, which is highly significant. This implies that when both selling pressure and trading volume increase above average levels, the negative impact of CSP on bond returns is mitigated by approximately 0.00373% for each unit increase in the interaction term. In other words, higher trading volumes may alleviate some of the downward pressure exerted by increased selling, possibly through enhanced liquidity provision that helps stabilize prices. Alternatively it could imply that large sell order flow of customers might be taken up by dealers at favorable prices maybe to secure customer relationship. This interaction highlights the role of trading volume in buffering the adverse effects of selling pressure, emphasizing the dynamic interplay between liquidity and market pressures.

When evaluating the relative importance of these market microstructure variables collectively, Customer Selling Pressure stands out as the most influential factor, both statistically and economically. It highlights the critical role of dealer responses to liquidity imbalances in determining bond returns. ATS Usage and Commissions maintain consistent significance, reflecting the importance of trading infrastructure and transaction costs in the nuanced process of price formation. In contrast, Customer Trade Flow contributes only marginally, with its influence being weaker and more conditional.

Overall, while these microstructure variables enhance our understanding of bond price formation by capturing intraday trading environments, liquidity frictions, and dealer behavior, their effect sizes remain moderate compared to dominant macro-level drivers such as yield changes. For instance, a one basis point (0.01%) increase in the 10-year Treasury yield deviation leads to a more substantial impact on bond returns, as outlined in the analysis of market shocks. Nevertheless, the statistical significance and clear directional relationships of these microstructure coefficients affirm that market microstructure dynamics are essential for a comprehensive analysis of bond returns.

Liquidity

Based on Ranaldo (2001), liquidity is decomposed into four dimensions: Depth, Tightness, Trading Time, and Resilience. Proxies for these dimensions include trading volume for depth, price spreads for tightness, daily trading frequency and trading inactivity for trading time, and intraday return variance for resilience. Liquidity metrics show significant yet moderate effects on daily bond returns. For daily trading volume, the coefficient is negative and highly significant, indicating that a 1% increase in trading volume above its average level is associated with a 0.01% decrease in daily bond returns. This result suggests that higher market depth may foster more competitive pricing, where dealers compete for order flow, leading to reduced embedded commissions and lower price markups.

Trading inactivity, measured by longer average waiting times between trades, has a positive and highly significant coefficient of 0.0000949, indicating that a 1% increase in inactivity above the mean is associated with a 0.00949% increase in daily bond returns. Extended periods of inactivity may promote greater price stability and more efficient price discovery, allowing bond prices to better reflect their fundamental values. Further, trading inactivity can capture passage of time which might be a more important driver of price movements once closer to maturity. Conversely, daily trading frequency carries a negative coefficient of -0.0004, which is moderately significant under HAC robust standard errors. This suggests that a 1% increase in trading frequency above its average level is linked to a 0.04% decrease in daily bond returns. Higher trading frequency may introduce market noise, where frequent trades disrupt price stability and obscure signals of intrinsic value. Overtrading could encourage speculative behavior, moving prices away from fundamentals and eroding returns over time. Alternatively, higher trading frequency can indicate higher liquidity which allows for more competitive pricing and less embedded transaction costs from dealers which could inflate prices. These results highlight an inverse relationship between trading inactivity and trading frequency, both of which influence price stability and efficiency in distinct but closely aligned ways.

Price spread and intraday return variance do not exhibit statistically significant effects on bond returns with this model, thus tightness and resilience seem to be less relevant in bond pricing.

While liquidity variables do not dominate the ranking of predictors, their significant and moderate effects underscore their role in fine-tuning bond price formation. These findings suggest that liquidity acts as a secondary adjustment mechanism, refining bond returns by influencing the efficiency and stability of price movements rather than serving as primary drivers of returns.

Bond characteristics

Bond-specific characteristics, including time-to-maturity, coupon rate, credit grade, and yield spread often outrank many microstructure variables.

The yield spread exhibits a significant negative coefficient indicating that a 1% increase in the average yield spread across bonds is associated with a 0.1324% decrease in daily bond returns. This coefficient is significant across both estimation methods. This inverse relationship underscores the fundamental principle that narrower yield spreads typically reflect improved credit conditions and lower risk premiums, enhancing the relative value of bonds and boosting their attractiveness to investors. Conversely, wider yield spreads may signal deteriorating credit quality or increased perceived risk, leading to suppressed bond returns.

Time-to-maturity is modeled with a squared transformation, resulting in a significant negative coefficient. This suggests that a 1% increase in the average cubic log-transformed time-to-maturity across bonds is associated with a 0.0262% decrease in daily bond returns. This effect is only significant when using HAC robust standard errors. The non-linear transformation captures the nuanced impact of maturity on bond returns, indicating that as bonds approach their maturity dates, their price sensitivity to market conditions diminishes. The reduced sensitivity is likely attributed to diminishing uncertainty and shorter exposure periods, which help stabilize returns as bonds approach redemption. Additionally, the relationship between time-to-maturity and price sensitivity evolves in a consistent manner; for instance, as a bond nears its maturity date, mechanisms like the pull-to-par effect may become more significant in influencing its price behavior.

The coupon rate carries a positive coefficient meaning that a 1% increase in the average coupon rate across bonds correlates with a 0.66% increase in daily bond returns. This effect is only significant when using HAC robust standard errors. Higher coupon rates enhance the periodic income generated from bonds. As bond prices are fundamentally determined by discounting their expected future cashflows (coupon payment and principal) at prevailing market interest rates, this relationship is as expected (Eom et al. 2004).

Credit grade is associated with a negative coefficient, indicating that a 1% increase in the average credit grade across bonds leads to a 0.0090% decrease in daily bond returns (only for HAC robust standard errors). As higher numerical values represent higher credit quality, this negative relationship might be explained due to higher liquidity for investment grade bonds, dampening returns of these bonds. As High Yield bonds might be more speculative, naturally higher returns might be requested.

The bond type flags—Zero Coupon Bond Flag, Floating Coupon Bond Flag, and Other Coupon Bond Flag—are dummy variables that indicate specific bond types relative to the fixed coupon bonds (the reference category). Since these flags are cross-sectionally averaged, their coefficients represent the percentage change in daily bond returns associated with a 1 percentage point (1%) increase in the proportion of each bond type within the bond universe.

The zero coupon bond flag carries a highly negative coefficient. A 1% increase in the proportion of zero-coupon bonds within the bond universe is associated with a 64.45% decrease in daily bond returns relative to fixed coupon bonds. This substantial negative impact reflects the inherent characteristics of zero-coupon bonds, such as the absence of periodic interest payments, leading to higher price sensitivity to interest rate changes and greater price volatility

(Eom et al. 2004). Consequently, zero-coupon bonds may underperform fixed coupon bonds in terms of daily returns.

Conversely, floating coupon bonds display a positive coefficient, indicating that % increase in the proportion of floating-rate bonds within the bond universe corresponds to a 10.41% increase in daily bond returns relative to fixed coupon bonds. Floating-rate bonds adjust their coupon payments based on prevailing interest rates, which can mitigate interest rate risk and enhance price stability in rising rate environments. This adaptability makes them more resilient compared to fixed coupon bonds, thereby contributing positively to daily returns.

The "Other Coupon" bond flag carries a negative coefficient, implying that a 1% increase in the proportion of bonds with alternative coupon structures is associated with a 26.25% decrease in daily bond returns relative to fixed coupon bonds. This category of bonds includes convertible bonds and callable bonds, which encapsulate additional complexities and risks. Convertible bonds may depend on the performance of the issuing company's stock, while callable bonds introduce the risk of early redemption by the issuer. These factors can reduce the attractiveness and predictability of returns for investors, leading to lower daily bond returns compared to traditional fixed or floating-rate bonds.

Lastly, lagged daily returns and rolling 5-day bond price volatility are examined. Both variables are treated as endogenous and have been instrumentalized using their second lags to address potential endogeneity concerns. The effect of a one-unit increase of lagged daily returns on daily returns is associated with a 0.3624% decrease in the current day's bond return, holding other factors constant. The negative sign implies a potential mean reversion effect, where past positive returns may lead to lower returns in the subsequent period, and vice versa. Given the high p-value there is insufficient evidence to conclude that lagged daily returns have a meaningful impact on current bond returns within the context of this model. Given that the effectiveness of the instrument was proved to be strong (see **Appendix 2: Statistical Tests**), the lagged returns seem to not have strong influence on current returns. The effect of a one-unit increase in the rolling 5-day bond price volatility - above average levels - is associated with a 0.06% increase in daily bond returns, holding other factors constant. The positive sign suggests that higher volatility over the past five days may be associated with higher current returns, potentially reflecting higher risk premiums or compensatory mechanisms for taking on increased volatility. Given the high p-values there is insufficient evidence to assert that rolling 5-day volatility has a significant impact on daily bond returns within the scope of this analysis. Given that the effectiveness of the instrument was proved to be strong (see **Appendix 2: Statistical Tests**), the lagged returns seem to not have strong influence on current returns.

These structural aspects of the bond's valuation framework appear more fundamental, exerting a steadier and more potent influence than the more situational factors of daily trading conditions.

Appendix 4.2. Subset Analysis

Subset Overview

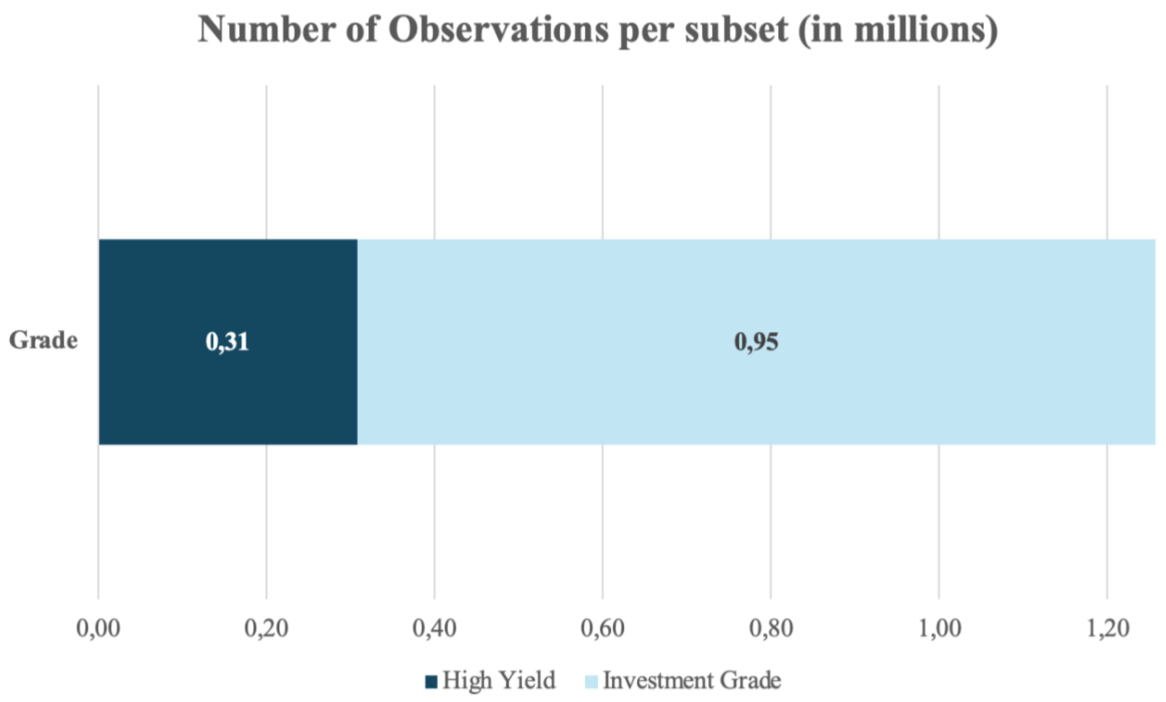

Figure 18: Subset Overview - Division of data

Credit Grade Subset

Subset Analysis	Grade subsets				Grade subsets				Individual Rank High Yield		Group Rank High Yield		Individual Rank Investment Grade		Group Rank Investment Grade	
	High Yield		Investment Grade		High Yield		Investment Grade		Bond-Date Two-way clustered robust standard errors model	HAC robust standard errors mode	Bond-Date Two-way clustered robust standard errors model	HAC robust standard errors mode	Bond-Date Two-way clustered robust standard errors model	HAC robust standard errors mode	Bond-Date Two-way clustered robust standard errors model	HAC robust standard errors mode
R-squared	0,0741	0,0741	0,1036	0,1036	0,0741	0,0741	0,1036	0,1036								
Adj. R-squared	0,0741	0,0741	0,1036	0,1036	0,0741	0,0741	0,1036	0,1036								
F-statistic	-48,861	4055,5	886,41	1,11E+04	-48,861	4055,5	886,41	1,11E+04								
P-value (F-stat)	1,0000	1,0000	0,0000	0,0000	1,0000	1,0000	0,0000	0,0000								
Distribution	chi2(28)	chi2(28)	chi2(28)	chi2(28)	chi2(28)	chi2(28)	chi2(28)	chi2(28)								
Estimator	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS								
No. Observations	302972	302972	933716	933716	302972	302972	933716	933716								
Cov. Estimator	clustered (1572, 322)	kernel (bandwith 15)	clustered (4833, 322)	kernel (bandwith 111)	clustered (1572, 322)	kernel (bandwith 15)	clustered (4833, 322)	kernel (bandwith 111)								
Δ 10Y Treasury Yield	-0,0019		-0,0016		0,0017	0,0000	0,0003	0,0000	2	5	1	1	2	7	1	2
Δ Market Volatility	-0,0007		0,0005		0,6246	0,4650	0,3744	0,0000	10	19			10	10		
Δ Treasury Yield Curvature (TYC)	-0,0008		-0,0005		0,0275	0,0000	0,0421	0,0000	4	8			4	10		
Commission	7,22E-05		7,35E-06		0,0597	0,0195	0,2715	0,1140	10	17	2	5	10	19	3	5
ATS Usage	0,0002		1,35E-05		0,0264	0,0079	0,3770	0,1030	7	15			10	19		
Customer Flow	3,80E-05		3,44E-06		0,1864	0,0850	0,6528	0,2476	10	19			10	19		
Customer Selling Pressure (CSP)	-0,0004		-0,0001		0,0000	0,0000	0,0000	0,0000	5	11			5	13		
Δ Daily Trading Volume	-0,0003		-8,76E-05		0,0832	0,0073	0,0157	0,0000	10	12	5	4	7	16	2	4
Δ Average Trading Inactivity	0,0003		1,81E-05		0,0032	0,0009	0,3215	0,0590	6	12			10	19		
Δ Daily Trading Frequency (DTF)	-0,0007		-7,42E-05		0,1153	0,0347	0,5819	0,1915	10	10			10	19		
Δ Price Spread	-4,69E-05		8,89E-05		0,8269	0,7701	0,1895	0,0001	10	19			10	15		
Δ Intraday Return Variance	-0,0004		-0,0001		8,3808	0,3742	0,0000	0,0025	10	19			5	13		
ATS - DTF Interaction	-0,0001		-5,79E-07		0,1978	0,1472	0,9674	0,9647	10	19	3	3	10	19	3	3
Commission - DTF Interaction	-0,0008		-9,58E-07		0,0673	0,0155	0,2562	0,0721	10	8			10	19		
Commission - ATS Interaction	1,47E-05		6,94E-07		0,0131	0,0025	0,5645	0,2550	9	18			10	19		
CSP - Daily Trading Volume Interaction	8,16E-05		2,03E-05		0,0246	0,0061	0,0012	0,0000	8	16			9	18		
Customer Flow - Price Spread Interaction	-7,44E-05		-8,20E-05		0,6164	0,5445	0,0002	0,0000	10	19			8	17		
Commission - Market Volatiltiy Interaction	-0,0016		-3,88E-05		0,1344	0,0036	0,8809	0,7283	10	6			10	19		
Δ 10Y Treasury Yield - TYC Interaction	-0,0011		-0,0010		0,0271	0,0000	0,0256	0,0000	3	7			3	8		
Δ 10Y Treasury Yield - TTM Interaction	-0,0003		-0,0003		0,0507	0,0000	0,1370	0,0000	10	12			10	12		
Time-to-maturity (TTM)	-0,0019		-0,0120		0,8749	0,6034	0,2228	0,0000	10	19	4	2	10	5	5	1
Coupon Rate	-0,0048		0,0002		0,3208	0,0375	0,9599	0,3253	10	4			10	19		
Credit Grade	dropped***		dropped***		dropped***	dropped***	dropped***	dropped***	10	19			10	19		
Δ Yield Spread	-0,0879		-0,0938		0,0000	0,0000	0,0672	0,0000	1	2			10	4		
Zero Coupon Bond Flag	-0,4307		-0,9534		0,0844	0,0169	0,2419	0,0000	10	1			10	1		
Floating Coupon Bond Flag	0,031		0,0106		0,1173	0,0256	0,0031	0,0000	10	3			1	6		
Other Coupon Bond Flag	-0,1642		-0,3103		0,2933	0,0510	0,1854	0,0000	10	19			10	2		
Lag 1 Daily Price Returns	-0,3552		-0,2239		0,4585	0,3545	0,5428	0,0001	10	19			10	3		
Rolling 5-day Bond Price Volatility	0,001		0,0010		0,2977	0,2875	0,1719	0,0000	10	19			10	8		

*Endogeous: L1_log_daily_return, log_volatility_squared_dev

**Instruments: L1_log_daily_return_lag2, log_volatility_squared_lag2

***dropped due to perfect mutlicolinearity

Table 4: Grade Subset Analysis - Model Results and Ranking

Appendix 5: Potential Criticism

While the regression model employed in this study offers robust insights into the determinants of bond returns, it is essential to acknowledge and address potential criticisms to contextualize the findings effectively.

The parameter selection process, based on VIF, pairwise correlations, and low variance filters, may not fully capture complex interdependencies and can reduce interpretability when combining highly correlated variables. However, this approach was carefully designed to balance multicollinearity mitigation with the inclusion of diverse variables. VIF and correlation checks are standard practices in econometric modeling, ensuring robust results. Combining market volatility shocks using PCA was a deliberate choice to simplify the model while capturing the overall volatility impact from both bond and equity markets. Retaining key market microstructure variables, despite slightly elevated VIF values, was justified by their central role in addressing transaction costs and trading efficiency dynamics. Any resulting multicollinearity is kept within manageable limits, ensuring the model remains focused and aligned with the research objectives.

Endogeneity concerns, such as bias from endogenous regressors like the lagged dependent variable and rolling volatility, may appear insufficiently addressed due to iterative instrument testing. However, these issues were systematically tackled by identifying endogeneity through strong economic reasoning and diagnostic tests, including Autocorrelation Function (ACF) and Partial ACF plots (**Appendix 2.2: ACF/PACF**). Instruments were carefully chosen using their second lags, validated through Hausman tests to ensure theoretical justification and empirical stability. While the process involved iterative testing, it was guided by economic principles to satisfy exclusion restrictions and enhance model validity. This rigorous approach resulted in a stable and meaningful instrument configuration that effectively mitigates endogeneity concerns (**Appendix 2: Statistical Tests**).

Incorporating non-linear transformations, such as the cubic transformation of yield curve curvature, may raise concerns about overfitting and complicating coefficient interpretation, potentially limiting the model's generalizability. However, these transformations were introduced to better capture the complex relationships inherent in bond return dynamics, improving model flexibility and fit. The use of partial regression plots, within-entity and between-entity scatter plots, and strong economic justifications ensured that the transformations reflect meaningful economic relationships rather than merely fitting the data.

Robustness checks were performed using alternative covariance estimators and subset analyses by credit grades to test the model's stability across different specifications. While some variables show conditional significance, this reflects the complexity and heterogeneity of the bond market. The structured robustness checks confirm that the core findings remain stable, while the conditional significance highlights the nuanced roles of certain variables across market segments. This thorough approach enhances the model’s credibility and demonstrates its reliability under varying conditions..

Translating theoretical hypotheses into measurable relationships within the regression framework can risk oversimplifying complex interactions. However, the hypotheses were carefully operationalized by linking theoretical constructs to well-defined, testable variables. Interaction terms and non-linear transformations were included based on theoretical expectations and statistical tests to capture the multi-dimensional nature of bond returns. This structured approach ensures that the empirical analysis directly addresses the research questions, maintaining both the theoretical relevance and empirical validity of the study.

In summary, while the model is subject to potential criticisms related to parameter selection, endogeneity, variable transformations, sensitivity analyses, and model specification, these limitations are addressed through a structured and methodical approach. Each decision in the model-building process was guided by standard econometric practices, rigorous diagnostic testing, and strong economic justifications. By acknowledging and mitigating these potential issues, the study ensures that the findings are both robust and reliable, providing valuable insights into the determinants of bond returns. Future research can build on this foundation by exploring alternative methods and further refining the model to enhance its validity and generalizability.

Appendix 6: Detailed Data Pre-Processing Pipeline and Key Variable Transformations

Data deduplication

I followed the data deduplication process for FINRA TRACE data outlined by Dick-Nielsen (2009) and (2014) using SAS in the WRDS Cloud environment. The cleaning procedure tackles trade cancellation, correction, reversal and eliminates double counting. After deduplicating the data, the data is exported from the cloud and loaded into Python for all following pre-processing steps.

Data Loading and Type Conversion

Once loaded, each variable was converted to its appropriate data type to facilitate accurate computations and analyses. This conversion included transforming numerical representations, dates, and categorical variables to their respective formats, ensuring that subsequent operations could be performed without type-related discrepancies.

Binary Categorical Variable Encoding

Binary categorical variables, which inherently possess two distinct states, were encoded into binary numerical formats (0 and 1). This transformation is essential for enabling these variables to be effectively utilized in regression models and other analytical techniques that require numerical input.

Coupon Type Mapping

Coupon codes within the dataset were systematically mapped to categorical classifications: Fixed, Float, Zero, and Other. This mapping standardizes the representation of coupon types, facilitating consistent interpretation and analysis across the dataset.

Merging Trade-Level Data with Master Dataset

To enrich the trade-level dataset with comprehensive bond characteristics, a merge operation was performed with the historic master dataset. This process entailed several critical sub-steps mimicking an As-Of Merge:

a. Accurate Matching of Securities: Each trade was matched with the corresponding security details that were effective on or before the trade report date (trd_rpt_dt). This ensured that the bond characteristics reflected the state of the security at the time of the trade.
b. Validity of Security Characteristics: Only trades where the trade report date was on or after the effective date of the security details (trd_rpt_dt >= trd_rpt_efctv_dt) were retained. This criterion guarantees that the security characteristics were valid and applicable at the time of the trade.
c. Resolving Multiple Matches: In instances where multiple masterfile records were applicable, the most recent record effective before the trade date was retained for each trade. This approach resolved potential ambiguities arising from overlapping effective dates.
d. Exclusion of Invalid Trades: Trades lacking valid masterfile records—such as those with no records effective before the trade date, missing data, or mismatched identifiers—were excluded from the analysis. This conservative filtering prioritized data quality, ensuring that only reliable and accurate trades were included in the regression analysis.

Calculation of Time to Maturity

The time to maturity for each bond was calculated based on the current date and the bond's maturity date. For perpetual bonds, which do not have a maturity date, the time to maturity was set to a constant value of 1000. This standardized approach accommodates both finite and perpetual bonds within the dataset.

Imputation of Missing Yields and Coupon Rates

Addressing missing values in critical variables such as yields (yld_pt) and coupon rates (cpn_rt) was essential for maintaining dataset integrity:

a. Coupon Rate Interpolation: For each bond on each day, missing coupon rates were interpolated linearly between the previous and next valid coupon rates, ensuring continuity and logical progression. This was only performed where coupon rates were expected and could be assess easily, thus only where missingness was present due to data quality issues.
b. Yield Interpolation: Time-based interpolation was employed to estimate missing yields, maintaining smooth and plausible estimates across temporal gaps in the data. This was only performed for bonds where yields are expected and usually reported and where interpolation was easy, e.g. where the immediate prior trade and the immediate next trade were reporting the yield of the particular bond.
c. Yield to Maturity (YTM) Estimation for Zero Coupon Bonds: For zero coupon bonds, YTM was estimated using the formula: ((F / P)^(1/n) - 1) * 100, where all information is available.
d. Fixed, Floating, and Other Coupon Bonds: For bonds with fixed, floating, or other coupon types, YTM estimation was not performed due to insufficient information, thereby preserving data accuracy by avoiding unreliable imputations.

Handling Missing Values

An analysis of missing values revealed that the missingness in cpn_rt and yld_pt was not completely at random (MCAR). Instead, it appeared to be missing at random (MAR) or not at random (MNAR), potentially related to other variables within the dataset. Given the imperative relevance of these variables observations which could not be aggregated to daily level without incurring NaNs will be dropped prior to the regression as both yields and bond will be entering the regression. Thus, it is important to note that through the inclusion of yields and coupons non-random dropping is incurred. Analyzing the missingness of yields and coupons it is clear that mostly the “Other Coupon Bond” category of bonds will be dropped. This is the case as these bonds are more complex. FINRA TRACE calculates yields themselves, thus where they cannot infer yields, they will assign NANs. As most bonds in the “Other Coupon Bond” category are exotic bonds with callable options or stock options, FINRA does not calculate their yields or no coupon is reported. Therefore, by decision to integrate yields or coupon in the regression we will have some systematic dropping of bonds, which is accepted as yields and coupons are assessed to be valuable regressors. Also, post-exclusion, the dataset retained sufficient size to ensure statistical power. This approach upheld analytical integrity by avoiding the inclusion of incomplete or imputed data that could compromise the reliability of the analysis.

Outlier Detection and Treatment

Outlier detection was performed to identify and address extreme values that could potentially skew the analysis:

a. Per-Variable Outlier Analysis: Each variable was examined for outliers across all bonds over time.
b. Statistical Outlier Detection: Using the ±3 standard deviation rule, no significant outliers were identified.
c. STL Decomposition: Seasonal-Trend decomposition using Loess (STL) (Cleveland and Cleveland 1990) was applied to detect anomalies, and no outliers were found through this method.
d. Justification for Extreme Values: The presence of extreme values was deemed realistic, as they could be explained by inherent bond characteristics rather than indicative of data errors.
e. Specific Outlier Removal: Only instances where zero coupon bonds had a coupon rate greater than zero were identified as outliers and subsequently excluded from the dataset.

Computation of Total Commission

A new variable representing the total commission was computed by summing the buy and sell commissions for each bond on a daily basis. This aggregation provided a comprehensive measure of transactional costs associated with each bond. This variable was further merged with the “Special Price” indicator which encoded embedded commissions in the price. Thus the total commission flag variable finally encodes whether a bond trade included commission or not (either explicit commission or embedded commission)

Aggregation of Trade-Level Data to Daily Panel Data

To facilitate time-series analysis while preserving the panel data structure, trade-level data was aggregated to the daily level with the following considerations:

a. Trade Frequency: The count of trades per bond per day was calculated to gauge trading activity.

b. Summation of Trade Volume and Commission: Daily sums of trade volume and commissions were computed for each bond.

c. Volume-Weighted Averages: Variables such as price, yield, trading market platform, ATS indicator, commission flag, and coupon rate were aggregated using volume-weighted averages to reflect the weighted central tendency.

d. Price Spread Calculation: The daily price spread for each bond was determined by calculating the difference between the minimum and maximum prices observed.

e. Waiting Time Between Trades: The average time in seconds between consecutive trades was computed to assess trading frequency and liquidity.

f. Intraday Metrics: Variance and standard deviation of intraday returns and log intraday returns were calculated to measure volatility and return dispersion within the trading day.

g. Market Footprint: Defined as (last_price−first_price)/entrd_vol_qt_sum, this metric captured the price movement relative to the total entered volume quantity.

h. Proportion of Buy Trades: The weighted percentage of trades with cntra_mp_id = 'C' (indicating buy trades) was computed to quantify buy-side activity of dealers.

i. Weighted Average Counterparty: This measure provided an average counterparty value, weighted by trade volume, to assess counterparty diversity and concentration. It effectively shows customer participation or customer flow

j. Modes were taken for variables where no aggregation was sensible as the variable reflects an intraday time-invariant characteristic, i.e. credit grade, Zero coupon bond flag, ect.

k. Handling Missing Aggregations: Rows resulting in missing values due to aggregation (e.g., single trade days per bond) were excluded to prevent zero inflation and maintain consistency. This exclusion was non-random, ensuring that only observations with incomplete trade data were removed.

Computation of Daily-Level Metrics

Additional metrics were derived at the daily level to enrich the dataset:

a. Price and Volatility Metrics: Daily price returns, rolling 5-day volatility, daily price changes, day counts between consecutive trades were computed to capture various aspects of price dynamics and liquidity.

Yield Spread Calculation

Yield spread, a critical measure of risk premium, was calculated through the following steps:

a. Merging Yield Curve Data: Yield curve information was integrated into the dataset to provide a benchmark for risk-free rates. Yield Curve data was taken from S&P Capital IQ constant maturity yield curve.

b. Interpolation and Extrapolation: Cubic spline interpolation was employed to estimate yields between known data points, while the Nelson-Siegel-Svensson (NSS) model (Annaert et al. 2024) was used to extrapolate yields beyond the 30-year maturity. This approach ensures comprehensive coverage of the yield curve. The cubic spline method was selected due to its ability to accurately interpolate the data, which provides yield information at discrete maturity steps. However, combining cubic spline interpolation for the observed range and the NSS model for extrapolation introduces a minor discontinuity (or "jump") at the 30-year maturity point. This occurs because the NSS model does not necessarily align perfectly with the 30-year bond yield, leading to a slight mismatch between the 30-year yield and the yield for maturities just beyond 30 years. See below a snapshot of a combined cubic spline and NSS fitted curve vs. actual yields. The curve fitting is robust across time and thus suits the purpose of inte- and extrapolating yields for the yield spread analysis.

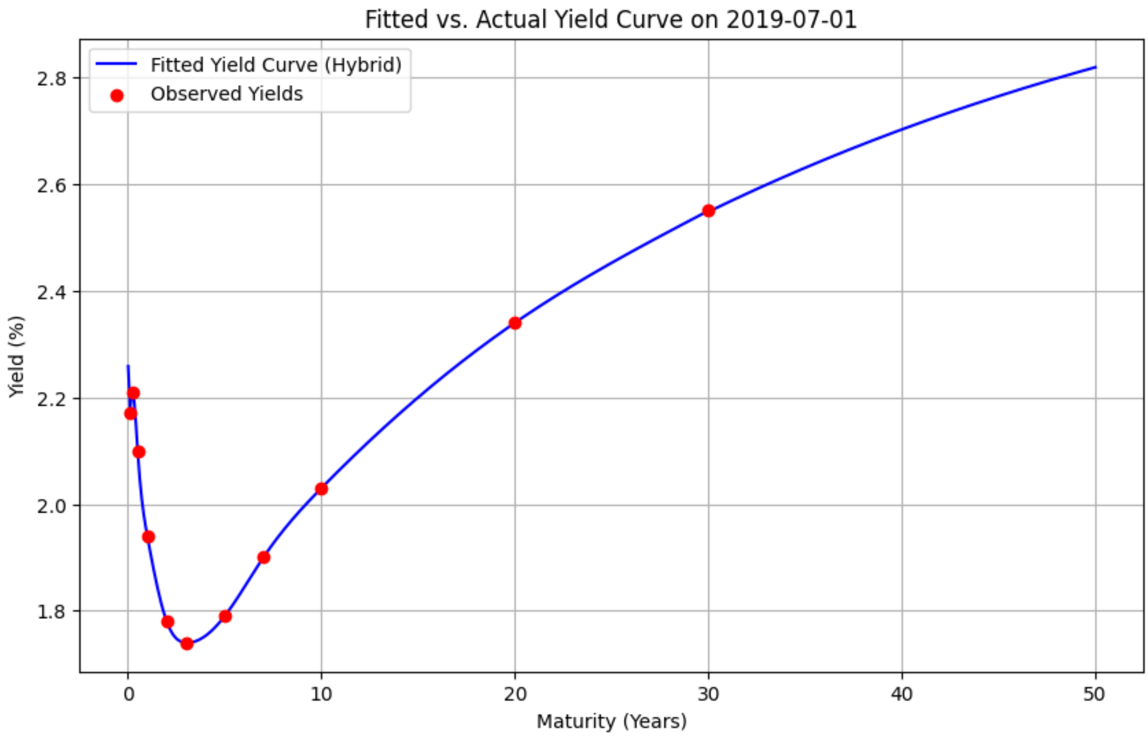

Figure 19: Yield Curve Fitting Snapshot

c. Yield Spread Determination: The yield spread was computed as the difference between the bond yield and the interpolated or extrapolated risk-free yield at the same maturity, thereby quantifying the additional yield demanded by investors for bearing credit risk.

Calculation of Risk-Free Yield Curvature

Risk-free yield curvature, an indicator of the shape of the yield curve, was calculated using the NSS model parameters: The curvature was derived by summing the second and third coefficients of the NSS model on each day, providing a measure of the yield curve's concavity or convexity.

Integration of Market Volatility Indicators

To incorporate market sentiment and volatility, data from the VIX and MOVE indices were merged into the dataset. These indices serve as proxies for overall market volatility and interest rate volatility, respectively, and are instrumental in capturing macroeconomic influences on bond trading.

Variable Transformations

a. To enhance the suitability of the data for regression analysis, several transformations were applied to address distributional characteristics and improve model assumptions. Many variables exhibited right-skewed distributions, which can bias model estimates and violate regression assumptions. To mitigate this, logarithmic, log1p transformations, logit transformations, and Yeo-Johnson transformations were employed to reduce skewness and approximate normality, thereby enhancing model robustness.

b. Transformations were strategically applied to linearize relationships between predictors and the dependent variable, a fundamental assumption in regression analysis. This step ensured that the relationships modeled were both meaningful and statistically valid.

Log-transformed variables included: daily trading frequency, daily trading volume and time to maturity. These continuous variables exhibited wide ranges and significant skewness. Applying the natural logarithm (log(x)) compressed their scales and mitigated the impact of outliers. Log1p-transformed variables (to circumvent issues of log-transform where zeros are present) included: spread size, trading inactivity (average waiting time between trades in seconds), current yield, volatility, intraday return volatility. For variables containing zero values, the standard logarithmic transformation is undefined. The log(1+x) transformation accommodates zero values while still addressing skewness. Logit-transformations were applied to bounded variables (bounded between 0 and 1) since still large skewness was present. Yeo-Johnson transformation was performed on the yield spread variable where also negative values could occur. Variables with low skewness were not transformed, including coupon rate, yield curvature, VIX and MOVE. Variables exhibiting minimal skewness do not benefit significantly from transformation and were thus retained in their original form to maintain simplicity and interpretability.

Lastly, cubic transformations were applied to a selected set of variables—including VIX, MOVE, , bond price volatility, yield curvature, time to maturity, and intraday return variance—to effectively capture their non-linear relationships with daily returns. These variables inherently exhibit complex dynamics that linear models fail to fully encapsulate, potentially leading to underfitted models and biased estimates. By incorporating cubic terms, the model gains the flexibility to represent more intricate patterns and interactions, thereby enhancing its ability to accurately reflect the true underlying relationships. This transformation allows for the accommodation of curvature and inflection points within the data, which are critical for understanding how changes in these variables influence daily returns in a non-proportional manner. The relationship of these variables with daily returns were assess via partial regression plots, within-entity scatter plots and between entity scatter plots. Consequently, these transformations contribute to a more robust and reliable analytical framework, ultimately leading to more insightful and actionable findings in the study.

The comprehensive data pre-processing pipeline outlined above meticulously addressed data loading, type conversion, categorical encoding, merging with master datasets, handling missing values, outlier detection, aggregation, and variable transformation. Each step was carefully designed to enhance data quality, ensure the validity of regression assumptions, and maintain the integrity of the analytical framework. This robust pre-processing foundation is pivotal for conducting reliable and insightful regression analyses within the thesis.

Final Variable Set

The final variable set, thus consists of the following variables:

- Daily Trading Frequency: The number of trades executed for each bond on a given trade date. This metric provides an indication of the trading activity and liquidity of the bond. This measure is in changes, not in levels, thus representing the change in daily trading frequency.
- Daily Trading Volume: the total sum of the uncapped par value volume reported for each bond on each trade date. This represents the aggregate traded volume, reflecting the bond's market activity and liquidity. This measure is a proxy for market depth (Ranaldo 2002). This measure is in changes, not in levels, thus representing the change in daily trading volume.
- Commission: represents the portion of trades for each bond on each day which included either explicit commission or embedded commission in the trade price.
- Customer Selling Pressure: Measures the proportion of dealer purchases for each bond from customers on a given day, weighted by trade volume.
- ATS Usage: Represents the volume-weighted average proportion of trades executed on Alternative Trading Systems (ATS) platforms versus those not on ATS platforms for each bond on each day.
- Customer Flow: represents the volume-weighted average proportion of trades involving costumers and dealers. It presents the customer participation rate or alternatively the inverse of the inter-dealer portion of trades weighted by volume of trade.
- Coupon: The coupon rate of each bond on each day.
- Credit grade: A binary indicator, representing whether a bond is rated as investment grade (1) or high yield (0).
- Coupon types: Floating, Zero, and Other coupon bond flags all indicate whether a bond is part of that category (1) or not (0). By default bonds are encoded as fixed coupon (when all other flags are indicating 0).

- Time to maturity: represents the time to maturity in years for each bonds on each day. For perpetual bonds this value is set to 1000.
- Spread size: represents the spread between the minimum and maximum reported prices for each bonds on each day and serves as a proxy for market tightness, indicating the range of price fluctuations within the day. This measure is in changes, not in levels, thus representing the change spread size.
- Trading inactivity: represents the average waiting time in seconds between consecutive trades for each bond on each day. This metric assesses the trading frequency and liquidity by measuring the intervals between trades. This measure is in changes, not in levels, thus representing the change in trading inactivity.
- Intraday return variance: Measures the variance of log intraday returns of each bond on each day. This measure is a proxy for market resilience. This measure is in changes, not in levels, thus representing the change in intraday return variance.
- Volatility: represents the rolling 5 day log price return volatility of each bond on each day. This metric captures short-term volatility of the bond and is a short-term proxy for risk of that bond.
- Daily return: measures as the log daily return. Is the dependent variable.
- Yield curvature: The yield curve curvature for each bond on each day, calculated as the sum of the second ($\beta 2$) and third ($\beta 3$) coefficients from the Nelson-Siegel-Svensson (NSS) model applied to the US Treasury yield curve. This variable measures the concavity or convexity of the yield curve, reflecting market expectations about future interest rates and economic conditions. This variable is in changes, not in levels, thus representing the change of the yield curvature.
- Yield spread: The yield spread for each bond on each trade date, calculated as the difference between the bond's yield and the interpolated or extrapolated risk-free Treasury yield. This

metric quantifies the additional yield demanded by investors for bearing the credit risk associated with the bond. This variable is in changes, not in levels, thus representing the change of the yield spread.

- 10Y Treasury yield: represents the daily 10Y US Treasury yield. This variable is in changes, not in levels, thus representing the change of the 10Y yield. The 10Y maturity is chosen as the appropriate benchmark yield as the average time to maturity of the bonds in the entire dataset equals approx. 12.7 years, making the 10Y benchmark suitable.
- VIX: The VIX index measures the market volatility expectation and serves as a proxy for overall market sentiment and investor uncertainty. This variable is in changes, not in levels, thus representing the change of the VIX index.
- MOVE: The MOVE index, similar to the VIX index, measures the implied volatility of US Treasury bonds, reflecting expectations of interest rate volatility and serving as an indicator of market stress in the bond market.