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

TRADE SIZE AND TRANSACTION COSTS IN OVER-THE-COUNTER MARKETS

ALBERT FOCK

Work project carried out under the supervision of

Professor André C. Silva

11/01/2024

I investigate the relationship between trade size and transaction costs in OTC markets. This involves simulating the search-theoretic model developed by Lagos and Rocheteau (2009), which implies that trade size and transaction costs increase together. For empirical analysis, I utilize the TRACE enhanced dataset on corporate bond trades from 2006 to 2022 and compute the bid-ask spread as a measure of transaction costs. My empirical analysis finds a negative influence of trade size on transaction costs. I discuss the mismatch between the theoretical model and empirical evidence considering the results of other empirical studies.

Keywords: trade size, transaction cost, bid-ask spread, over-the-counter market, trading frictions

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).

I. Introduction

This work investigates trade size and transaction costs in over-the-counter (OTC) markets. I examine how they develop over time and how both relate to each other in corporate bond markets. According to theory, transaction costs increase together with the trade size (Lagos and Rocheteau, 2009). My empirical examination does not confirm this relationship. On the contrary, I find a significant negative impact of trade size on transaction costs and discuss potential explanations for this mismatch.

The empirical analysis is based on the TRACE enhanced dataset of secondary U.S. corporate bond trades for the period from 2006 to 2022. I compute the bid-ask spread for each trade as a measure of transaction costs and regress it on the log of trade size as well as several dummy variables for trade characteristics. I obtain statistically significant negative coefficients for the log of trade size in the range of -22.6 to -3.39 for the five most traded bonds in the sample period.

To extract the bid-ask spreads I apply an estimation method that is based on reference prices.[1] I obtain average bid-ask spreads of 86 basis points (bps) for the full sample of all trades. I present how spreads evolve over the sample period and show that spreads behave differently during the two crises in the sample period: the subprime crisis as well as the COVID-19 crisis. I summarize recent literature studying this phenomenon. Additionally, I provide extensive summary statistics for spreads, trade size, and trade volume for the full dataset.

Before diving into the empirical work, I study the search-theoretic model developed by Lagos and Rocheteau (2009) to understand theoretical implications regarding investor and dealer

[1] The estimation method is based on Choi, Huh, Shin (2023). Reference prices are the volume-weighted average price of all interdealer trades for each bond and trade day.

behavior in the presence of trading frictions. I exhibit the model's implication that trade size and transaction costs increase together.

To illustrate this implication, I summarize the model's specification and present the critical equilibrium equations. I develop why the model concludes trade size and transaction costs increase together and present key equations determining their interaction. Additionally, I simulate the behavior of these variables in a two-investor economy and visualize the behavior of trade size and transaction costs for varying parameters.

My work builds on Lagos and Rocheteau's search-theoretic model that exhibits the behavior of investors' asset choice, trade volume and transaction costs in the presence of trading frictions. It incorporates classic OTC market characteristics: trades occur decentralized through bilateral meetings between dealers and investors with the presence of search frictions as well as bargaining processes between investors and dealers. The model is unique as it removes any restrictions on investors' asset holdings, unlike earlier models that restricted holdings to just 0 or 1 units. The resulting model shows that investor accommodate trading frictions by adjusting their asset choices. Specifically, investors asset choice will be more dispersed when trading frictions are low.

Duffie, Gârleanu, and Pedersen (2005) are pioneers in the area of search-theory. They develop a model for OTC markets that captures the impact of search and bargaining frictions on asset prices and intermediation. They conclude that transaction costs are lower if investors have better search options and can easily find other investors or dealers to trade with. They suggest that smaller investors have limited search options. Accordingly, they have lower bargaining power in the negotiations with dealers, and therefore, smaller investors and smaller trades are associated with higher transaction costs.

Numerous empirical studies have examined bond markets as a representation of OTC markets. While many studies explore the extensive TRACE dataset on U.S. corporate bond trades, Pinter, Wang and Zou (2022) examine the relationship between trade size and transaction costs in a dataset for UK gilts. They identify a “size penalty” where costs increase with trade size when accounting for investors identities. Accordingly, their finding is contrary to my empirical result. I relate my empirical findings to theirs in my analysis.

I contribute to these studies by constructing an extensive dataset of U.S. corporate bond trades for the period 2006 to 2022 and computing the key variable bid-ask spread for each trade. I relate the spreads and the trade size to each other and find a clear negative relationship. I imbed this empirical finding in existing search-theoretic models. Moreover, I compare my result to the findings of other empirical studies utilizing different datasets.

II. Search Theory

In the following I introduce the model developed by Lagos and Rocheteau (2009) to illustrate the theoretic implication that transaction costs increase with larger trade size.

Time is continuous, starts at $t = 0$, and goes on forever. There are two types of agents: investors and dealers. Investors have a utility function u_i that is twice continuously differentiable, strictly increasing, strictly concave, and dependent on their asset holdings a.[2] Each investor experiences preference shocks, modeled as a Poisson process with arrival rate δ, that changes their utility function. The probability that the investor draws preference type i is π_i. These preference

[2] More specifically, investors utility is dependent on a consumption good, one unit of the asset produces one unit of the consumption good. Accordingly, good consumption coincides with the investor’s asset holdings a. (Lagos and Rocheteau, 2009)

shocks reflect the concept that investors value the asset differently over time and therefore want to rebalance their asset positions. Dealers do not hold assets. (Lagos and Rocheteau, 2009)

Investors can adjust their asset positions through dealers who continuously trade the asset in an interdealer market on behalf of investors. Meetings between investors and dealers occur randomly, following a Poisson process with arrival rate α. When a dealer and an investor meet, they negotiate the quantity of assets that the dealer will acquire for the investor and the intermediation fee charged by the dealer. Dealers have bargaining power η. From the investor's perspective, the analyzed environment is equivalent to one where an investor meets a dealer according to a Poisson process with arrival rate $\kappa = \alpha\ (1 - \eta)$ and has all the bargaining power. Accordingly, κ is the inverse of trading frictions. (Lagos and Rocheteau, 2009)

1. Investors and Dealers Utility

The value function $V_i(a, t)$ denotes the maximum expected discounted utility attainable by an investor with preference type *i* holding *a* assets at time *t*:

$$(1) \qquad V_i(a,t) = E_i\left[\int_t^{T_\alpha} e^{-r(s-t)} u_{k(s)}(a)ds + e^{-r(T_\alpha - t)}\left\{V_{k(T_\alpha)}\left[a_{k(T_\alpha)}(T_\alpha), T_\alpha\right] - p(T_\alpha)\left[a_{k(T_\alpha)}(T_\alpha) - a\right] - \varphi_{k(T_\alpha)}(a, T_\alpha)\right\}\right]$$

where T_α denotes the next time the investor contacts a dealer and $k(s)$ denotes the investors preference type at time s.[3] The first term on the right side of the equation represents the discounted utility of the investor that he attains from his current asset holdings a until the point when he contacts a dealer, T_α. The second term represents the expected discounted utility attained by the new level of asset holdings $a_{k(T_\alpha)}(T_\alpha)$ from time T_α onwards. At T_α the dealer purchases $a_{k(T_\alpha)}(T_\alpha) - a$ in the market at price $p(T_\alpha)$ on behalf of the investor; the investor

[3] E_i is over the random variables T_α and $k(s)$, and is indexed by *i* to indicate that it is conditional on $k(t) = i$.

readjusts his asset holdings from a to $a_{k(T_\alpha)}(T_\alpha)$ and pays the dealer an intermediation fee $\varphi_{k(T_\alpha)}(a, T_\alpha)$. (Lagos and Rocheteau, 2009)

Let W(t) denote the maximum expected discounted utility attainable by a dealer. It satisfies

(2) $$W(t) = E\left\{e^{-r(T_\alpha - t)}\left[\int_S \varphi_i(a, T_\alpha) dH_{T_\alpha} + W(T_\alpha)\right]\right\}$$

where the first term in the squared bracket represents the expected discounted utility for the dealer from trading the next time he meets an investor T_α that is a random draw from H_{T_α}, the distribution of investors across preference types and asset holdings at time T_α. $W(T_\alpha)$ is the dealers expected utility after that point of time. (Lagos and Rocheteau, 2009)

2. Equilibrium

In the described environment, the unique equilibrium is characterized by allocations and prices converging to the steady-state values $\{a_i, q, p, \varphi_i(a)\}$[4] that satisfy the following system of equations:

(3) $$p = \frac{q}{r},$$

(4) $$\bar{u}'_i(a_i) \leq q \text{ "="} if\ a_i > 0,$$

(5) $$\sum_{i=1}^{I} \pi_i\, a_i = A,$$

(6) $$\varphi_i(a) = \frac{\eta[\bar{u}_i(a_i) - \bar{u}_i(a) - q(a_i - a)]}{r + \kappa}$$

where $\bar{u}_i(a_i)$ is the expected discounted utility for investor *i* which is a combination of the investor's current utility profile $u_i(a_i)$ and the average utility profile

(7) $$\bar{u}_i(a_i) = \frac{(r+\kappa)u_i(a_i) + \delta \sum_j \pi_j u_j(a_i)}{r + \kappa + \delta}$$

[4] The model in Lagos and Rocheteau additionally considers the measure of Investors *H*. As the variable is not of interest for my investigation, it has been omitted.

and q represents the present value of the expected capital loss to the investor from holding a_i from t until the next time he readjusts his assets T_α. (Lagos and Rocheteau, 2009)

Accordingly, in the equilibrium investors choose their asset holdings such that their resulting marginal utility equals the expected capital loss from holding the asset. Additionally, prices are determined by the expected capital loss from holding the asset and the interest rate. The parameter A denotes the asset supply in the interdealer market. The transaction costs φ are equal to the utility gain for the investor resulting from rebalancing his asset position from his initial asset holdings a to a_i, divided between the dealer and the investor according to the dealer's bargaining power η. (Lagos and Rocheteau, 2009)

III. Simulation of Theory

I simulate this equilibrium in an economy with two investors and a set of utility functions in the form of $u_i = \varepsilon_i \ln(a_i)$, where ε_i is a vector of values between 0.55 and 1.45 with equal probability π_i, the average utility profile is $u = \ln(a_i)$. Investor 1 has the current utility profile $\varepsilon_i = 0.75$ and investor 2 has $\varepsilon_i = 1.25$. Preference shocks δ occur with an arrival rate of 0.1, and investors have initial asset holdings $a = 50$.[5]

The simulation illustrates that transaction costs increase together with trade size. I focus my simulation on this implication of the model as I investigate the relationship empirically. In Appendix A.2 simulations of the investor's choice of asset holdings in dependence of trading frictions are shown as this is another key implication of the model: a decrease in trading frictions encourages investors to take on more extreme asset positions, prioritizing their current valuation

[5] Furthermore, I choose the following parameters: Asset supply in the interdealer market $A = 10$ and interest rate $r = 0.1$.

of asset holdings over their expected valuation. Furthermore, in Appendix A.3 I show how the trade volume reacts to varying trading frictions.

To understand how transaction costs react to varying trade sizes consider the equilibrium level of transaction costs

$$\varphi_i(a) = \frac{\eta[\bar{u}_i(a_i) - \bar{u}_i(a) - q(a_i - a)]}{r + \kappa} \tag{8}$$

and its derivative with respect to a

$$\frac{\partial \varphi_i(a)}{\partial a} = -\frac{\eta}{r + \kappa}[\bar{u}_i{}'(a) - q]. \tag{9}$$

Additionally, consider the transaction costs per asset traded

$$\frac{\varphi_i(a)}{a_i - a} = \frac{\eta}{r + \kappa}\left[\frac{\bar{u}_i(a_i) - \bar{u}_i(a)}{a_i - a} - q\right] \tag{10}$$

and its derivative with respect to a

$$\frac{\partial}{\partial a}\left[\frac{\varphi_i(a)}{a_i - a}\right] = \frac{\eta}{r + \kappa}\left[\frac{\bar{u}_i(a_i) - \bar{u}_i(a) - \bar{u}_i{}'(a)(a_i - a)}{(a_i - a)^2}\right]. \tag{11}$$

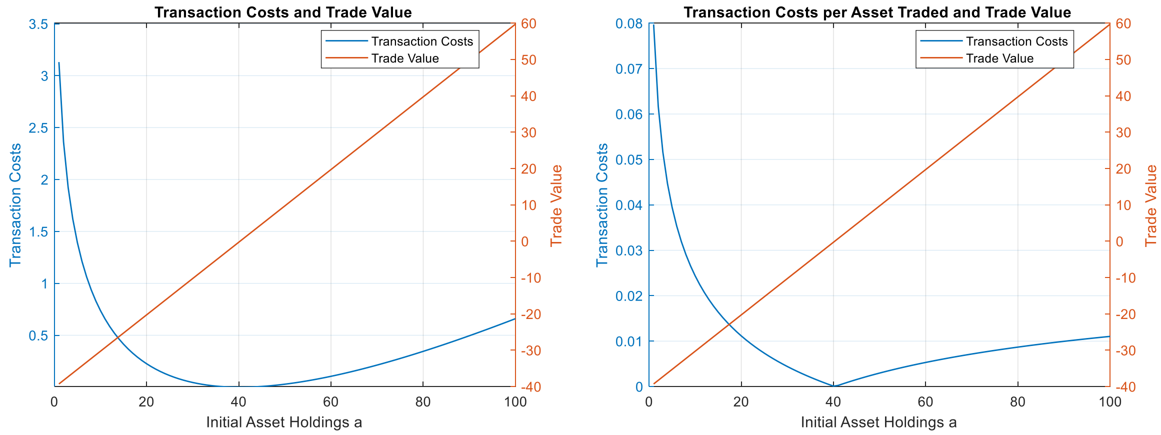
Both derivatives have the same sign as $a - a_i$[6]. In other terms, for an investor holding asset position a and wanting to trade $a_i - a$, the reaction of transaction costs – both total and per unit of asset traded – to a change in the initial asset holding a has the same sign as $a - a_i$.

I illustrate this relationship in Figure 4. It follows that transaction costs increase with larger trade sizes $|a_i - a|$. Note that the red line in Figure 4 is the trade value $a - a_i$ whose absolute value $|a_i - a|$ is the trade size. When trade size – the difference between the chosen asset a_i and the initial asset holdings a – rises, transaction costs increase.

[6] For an investor who holds asset position $a \geq 0$ and whishes to trade $|a_i - a| > 0$. I provide the proof in Appendix A.1.

As shown, the transaction costs are the utility gain of the investor from rebalancing his asset holdings that are divided between the investor and the dealer according to their bargaining power. With increasing trade size, the investors utility gain from trading increases and accordingly the dealers share is increasing. The derivatives above show this effect is true for both the total transaction costs as well as the transaction costs per asset traded. It is crucial to note that this analytical examination maintains constant characteristics of both investor and dealers.

Figure 1: Transaction Costs, Transaction Costs per Asset Traded, and Trade Value on varying Asset Holdings



IV. Trade Size and Transaction Costs in Empirics

My empirical analysis shows that transaction costs decrease for increasing trade size. I compute bid-ask spreads for U.S. corporate bond trades as a measure of transaction costs for the period from 2006 to 2022.[7] I then regress the resulting spreads on the log of trade size as well as a vector of dummy-variables for various trade characteristics for the most traded bonds in this period and obtain significant negative coefficients for trade size in the range of -3.39 to -22.6. The results of this regression are presented in Table 1. Accordingly, my empirical analysis does

[7] I provide a detailed description of the dataset and its cleaning in Section IV.3. Furthermore, I describe how I compute the bid-ask spreads in detail in Section IV.4.

not confirm the implication of Lagos and Rocheteau that transaction costs increase with the size of the trade.

I investigate their relationship by regressing $Spread_i$, the bid-ask spread for each trade *i*, according to the following model for each of the five most frequently traded bonds within my sample period[8]:

$$(12) \qquad Spread_i = \alpha + \beta_1\, log(Trade\ Size_i) + \beta_2\, Dealer - Buy_i + \beta_3\, Customer\ Trade_i + \beta_4\, Agency\ Trade_i + \varepsilon_i$$

The dummy variables Dealer-Buy, Customer Trade, and Agency Trade capture trade characteristics and act as control variables.

Table 1: Regressions of Bid-Ask Spreads on Trade Size and Trade Characteristics

Cusip	Observations	log(Trade Size)	Dealer-Buy	Customer Trade	Agency Trade	R-squared
61748AAE6	191,983	-8.23***	-53.97***	148.29***	-88.28***	0.279
		(0.208)	(0.995)	(0.649)	(0.763)	
369604BQ5	187,520	-12.76***	-118.46***	223.33***	-116.79***	0.4560
		(0.184)	(0.907)	(0.730)	(0.818)	
92978AAA0	148,586	-22.6***	-158.68***	283.92***	-183.11***	0.4775
		(0.243)	(1.056)	(0.946)	(1.109)	
369604BC6	144,147	-12.79***	-95.4***	182.04***	-117.48***	0.3511
		(0.190)	(1.059)	(0.929)	(0.750)	
037833AK6	130,766	-3.39***	-13.96***	43.68***	-24.93***	0.1032
		(0.086)	(0.489)	(0.450)	(0.401)	

Notes : *, **,and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

I use the log of the trade size due to the wide dispersion of the variable, ranging up to $330 million. The coefficients for the log of trade size range from -3.39 to -22.6. For the most traded

[8] I identify the most traded bonds by the most frequent TRACE Bond ID as the cusip is not indicated for 32m of the raw trades.

bond with a coefficient β_1 of -3.39 for example, this means that an 1% increase in the trade size is associated with a 3.39% decrease in the spread.

All Dummy variables are significant, and the sign of their coefficients are consistent across the evaluated bonds. Bid-ask spreads are lower for buy-trades from the perspective of the dealer and for agency trades. Customer trades, however, have significantly higher spreads.

The coefficients in the regression analysis are all highly significant with t-values varying up to 300 and with values above 50 in average. These very large t-values are partly due to the enormous number of observations. The obtained R-squared varying from 0.10 to up to 0.47 show that the regression is explaining a considerable part of the very large variation in the data for some of the bonds. I provide robustness checks that validate my results in Appendix A.4.

1. Empirics and Theory

Previous studies exploring data on OTC market found similar results. For instance, Feldhütter (2012) using TRACE enhanced data from 2004 to 2009 found that investors making large trades buy at lower prices or sell at higher prices. Green, Hollifield, and Schurhoff (2007) study data on municipal bond markets and conclude that dealers earn lower average profits on larger trades. They explain this with varying bargaining power for investors. Larger trades get better prices because dealers' bargaining power decreases with trader size, and larger traders often trade larger amounts (Green, Hollifield, and Schurhoff, 2007).

This dynamic is the limitation of my empirical analysis as I do not consider differences in investor characteristics between individual trades, however, the theoretic model assumes investor characteristics to be constant. These unaccounted-for characteristics may contribute to the observed relationship between trade size and transaction costs. (Choi, Huh, Shin, 2023; Feldhütter, 2012; Pinter, Wang, and Zou, 2022)

To control for this effect, it would be ideal to control for investors characteristics. However, the TRACE enhanced dataset lacks information on investors. Pinter, Wang, and Zou (2022) used the ZEN database on secondary UK government bond trades for their empirical analysis of OTC markets.[9] Different to other datasets, the ZEN database provides not only detailed information on all individual transactions, but also the identities of both counterparties in each transaction (Czech et al., 2021; Pinter, Wang, and Zou, 2022).

Pinter, Wang, and Zou find that larger trades have lower trading costs in government bond markets than smaller trades, consistent with my findings for U.S. corporate bond trades. However, when controlling for investors identities, they identify a "size penalty" that lets transaction costs increase with the size of trade.[10] Accordingly, the implication by Lagos and Rocheteau is empirically substantiated: transaction costs increase with the trade size, reflecting the increasing utility gain from trading that is split between the investor and the dealer.

Additionally, Pinter, Wang, and Zou and other studies such as Rapp and Waibel's (2023) explain this effect with higher costs for the dealers to execute trades with large trade size. They might have to split orders incurring additional search costs. Accordingly, these costs are passed on to the investors, thus increasing the transaction costs with larger trade sizes.

2. Data Summary

In the following I investigate how spreads, trade size, and trade volume have evolved throughout the sample period and discuss why spreads behave differently in the two crisis periods in the sample. I show summary statistics for transaction costs, trade size, and trade volume for the

[9] Even though gilts are listed in the London Stock Exchange, most of the trades take place over the counter (Czech et al., 2021).

[10] Additionally, they find that this effect is larger for more sophisticated investors (hedge funds and asset managers) and is smaller for less sophisticated investors (tpension funds, foreign central banks, insurance companies).

sample period from 2006 to 2022 in Table 2 and 3. The average trade size approximates $600,000. Notably, the average trade size for customer trades is approximately $800,000. The full clean dataset covers a total trade volume of $115 trillion and daily trade volume averages to $23 billion. The average spread in the full sample is 86bps and – consistent throughout the sample period – spreads for customer trades are significantly higher with 151bps.

Table 2: Summary Statistics of the TRACE enhanced dataset

	total number of trades in million		trade size in $		daily trade volume in $bn		number of trades per day	
	all trades	customer trades	all trades	customer trades	all trades	customer trades	all trades	customer trades
Full Sample	198.3	111.4	593,768	794,682	23.6	17.8	42,845	22,405
Pre-Crisis: Jan 2006 - Jun 2007	6.9	4.4	841,118	994,935	15.4	11.7	18,364	11,740
Crisis: Jul 2007 - Apr 2009	11.0	6.7	615,674	799,502	14.6	11.5	23,771	14,350
Post-Crisis: May 2009 - May 2012	31.7	18.6	768,925	981,134	29.9	22.6	38,933	23,006
Pre-Covid: Jun 2012 - Feb 2020	99.2	57.1	557,111	748,552	22.6	17.6	40,551	23,468
Covid: Mar 2020 - Dec 2022	49.5	24.6	515,716	723,069	28.7	20.0	55,609	27,696

Table 3: Summary Statistics of Bid-Ask Spreads

	total number of trades in million		mean spread in bps		percentage of negative spreads	
	all trades	customer trades	all trades	customer trades	all trades	customer trades
Full Sample	118.1	63.8	86.20	150.63	26.80	16.15
Pre-Crisis: Jan 2006 - Jun 2007	2.9	1.6	77.57	146.33	30.04	17.10
Crisis: Jul 2007 - Apr 2009	5.9	3.4	143.05	225.69	23.36	13.05
Post-Crisis: May 2009 - May 2012	18.2	10.2	109.64	184.82	25.00	14.00
Pre-Covid: Jun 2012 - Feb 2020	58.3	32.5	86.00	146.87	26.05	15.97
Covid: Mar 2020 - Dec 2022	32.7	16.1	64.00	121.36	29.48	18.40

The data shows distinct periods within the sample period. Notably, the financial crisis period (July 2007 to April 2009) is characterized by low trade volume and small trade sizes, however, spreads peak during that period. Conversely, the post-crisis period (May 2009 to May 2012) exhibits high trade volume and large trade sizes. The COVID period, starting from March 2020, initiates with a peak in daily trade volume and shows low trade sizes. Spreads gradually decrease since the financial crisis and are lowest in 2022.

I compute the average spread of 86bps for the full sample with available reference prices for 118 million trades. In the sample restricted to customer trades with a volume of more than $1 million the average spread is significantly lower with 27bps[11]. Spreads for customer trades consistently exceed those for the entire sample. The percentage of negative spreads inversely correlates with spread heights, reaching its lowest point during the financial crisis.

Figure 2: Average Spread in bsp (30-days rolling average)

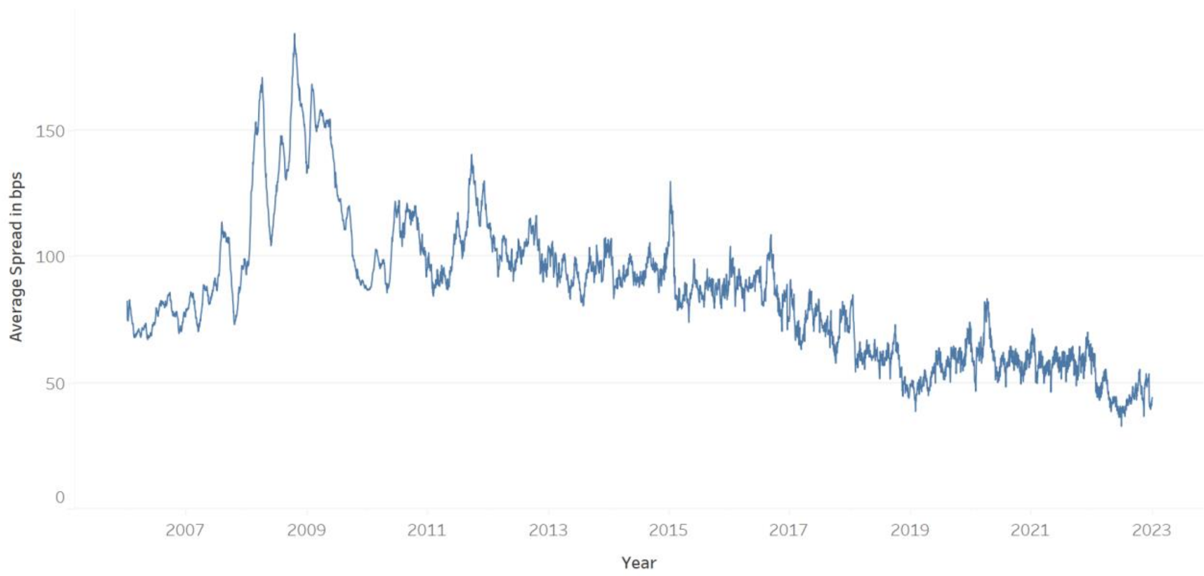


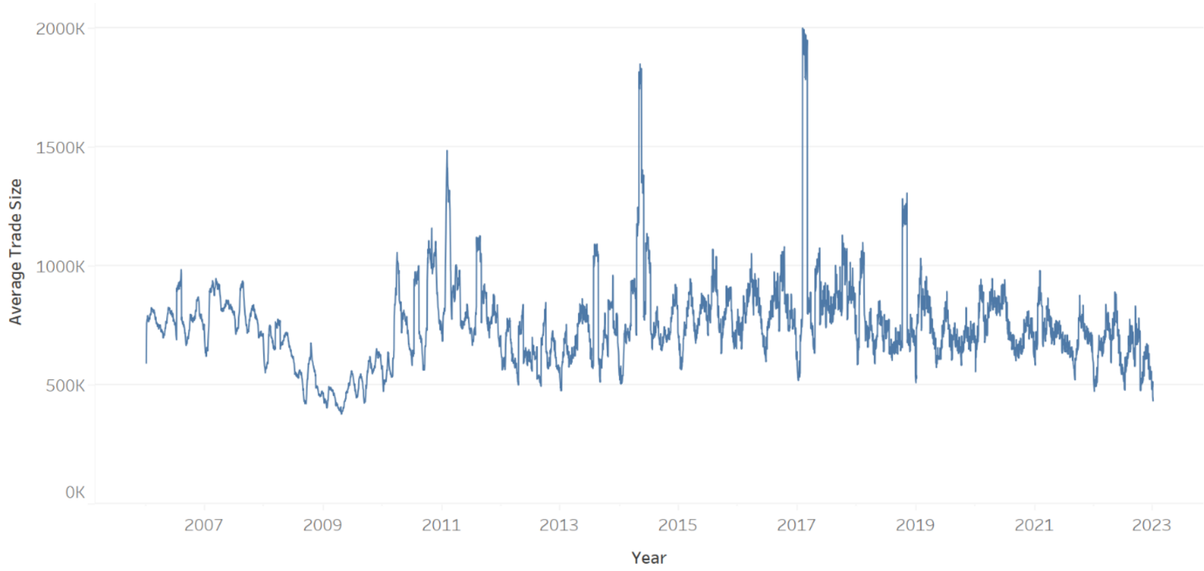
Figure 2 shows that spreads peak during the 2009 financial crisis and exhibit a consistent decrease in subsequent years. Specifically, average corporate bond spreads amount to 143bps during the crisis period from July 2007 to April 2009. This effect is associated with declining liquidity on corporate bond markets during the crisis (Dick-Nielsen, Feldhütter, and Lando, 2012).

[11] This is in line with the result in Choi, Huh, and Shin. They obtain average spreads of 33bps for investment grade bond trades and 41 bps for high yield bond trades, considering only customer trades with par value above $1 million for a sample period from January 2006 to December 2016. I obtain an average spread of 35bps across investment grade bond and high yield bond trades for the same period and restrictions.

However, during the peak phase of the COVID-19 crisis spreads do not increase significantly and for the whole COVID-19 period from March 2020 to December 2022 they decline to 63bps in average. Empirical literature indicates the COVID-19 pandemic as an exogenous shock also caused a significant decline in liquidity. However, this initial shock receded quickly and almost fully after the announcement of the Fed's intention to intervene on bond markets and purchase bonds. (Kargar et al., 2021; O'Hara, Zhou, 2021)

From January 2006 to June 2007, before the subprime crisis, average trade sizes are high at $841,118 in average. During the subprime crisis trade sizes are low but increase again during the post-crisis period from May 2009 to May 2012 as observed in Figure 3. The years 2013 to 2018 show some outliers with unusually large trade sizes. Starting in 2011, trade sizes decrease gradually and are lowest in the COVID-19 period from March 202 until the end of the sample period.

Figure 3: Average Trade Size – 30 days moving average



Note: Outliers disrupt the moving average of the time series. For a version with a shorter moving average, please refer to Appendix A.6.

Figure 4: Daily Trade Volume – 30 days moving average

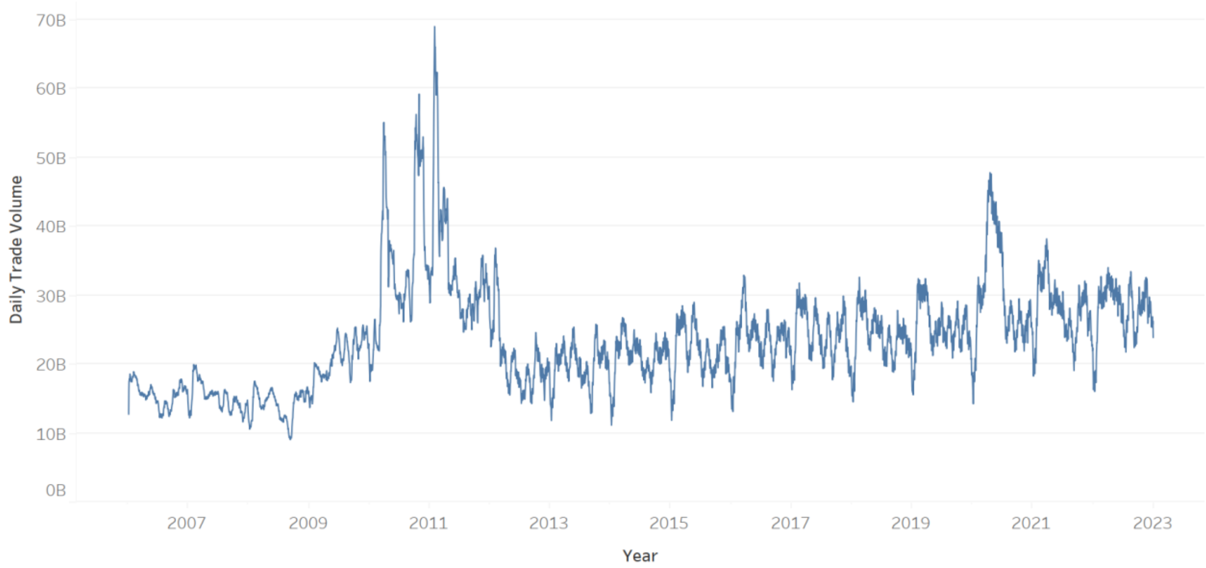


Figure 4 illustrates that trade volumes reach their highest levels in the post financial crisis period from March 2010 to June 2011. Since 2013 there is a gradual increase with a peak during the onset of the COVID-19 crisis. Seasonal trends are apparent, with trade volumes reducing during the Christmas season and peaking again in February and March each year. Notably, the financial crisis period (July 2007 to April 2009) is characterized by low trade volume. Conversely, trade volume increases a little during the onset of the COVID-19 crisis.

3. The Dataset

My data source is the TRACE enhanced – Bond Trades (BTDS) dataset from the Financial Industry Regulatory Authority (FINRA) of U.S. corporate bond trades.[12] The dataset provides comprehensive trade information, including trade date, time, volume, price, and trading capacity (principal or agent). Trades are categorized as interdealer or customer trades. Additional

[12] Starting from July 1, 2002 dealers are required to report over-the-counter corporate bond trades. The dataset is a result of regulatory initiatives to increase price transparency in secondary corporate bond markets. (Bao, Pan, Wang, 2011)

details include whether a trade is a buy or sell, information about the counterpart, and technicalities such as a unique message sequence number for each trade and day.

My sample period is from January 2006 to December 2022.[13] My initial sample consists of 299,379,700 trades. After deleting 42,746 trades with uncomplete data my raw dataset consists of 299 million US corporate bond transactions and provides 42 variables for each transaction.

I filter the dataset according to the method of Dick-Nielsen (2014). The filter is designed to delete transactions which are already known and marked as being errors. The filter runs in 3 steps. The first step deletes corrections and cancellations, the second step deletes removals. Trades identified in the first two steps are actual errors. In the third step interdealer transaction pairs are identified and the doubles removed. Finally, agency transactions are removed and special transactions that are potential outliers identified. I provide a detailed description of the individual steps in Appendix A.5.

I delete 12 million trades as corrections and cancellations in the first step and 1.5 million trades as reversals in the second step of the filter. In step 3 of the filter, 72 million interdealer-doubles are deleted, and 15 million trades identified as agency-double trades. In total, I remove 33.8% of the trades in the raw dataset. Additionally, I classify 17 million trades as trades with special circumstances.[14] The number of deleted trades is in line with the fraction of filtered trades indicated by Dick-Nielsen in the amount of “almost 35% of the raw transactions” in 2007.

[13] I start my sample in January 2006 as the FINRA (formerly the National Association of Securities Dealers) implemented phase III of the bond transaction reporting during the year 2005, then requiring reporting on approximately 99% of all public transactions. (Bao, Pan, Wang, 2011; Feldhütter, 2012)

[14] As the exclusion of these special trades does not affect the outcome of the regression analysis, they are not excluded from the analysis.

4. Bid-Ask Spread Estimation

I use the obtained clean dataset to compute bid-ask spreads according to the method by Choi, Huh, Shin (2023) as follows; First I compute a reference price for each bond and trade day as the volume-weighted average price of all interdealer trades larger than $100,000[15] on that day. In a second step I compute the spreads for individual trades according to the following formula:

$$Spread = 2Q * \frac{traded\ price - reference\ price}{reference\ price} \qquad (13)$$

where Q is +1 for a customer buy and -1 for a customer sell. All spreads are winsorized on a 1% level on both tails.

The precision of the calculation of *Spread* depends on an accurate estimation of the reference prices. This approach proxies the reference prices employing interdealer prices as they are more likely to stay within the bid-ask spreads than trades involving customers. Nevertheless, it is to be noted that they may also deviate from the fundamental values and interdealer can be missing for some days and bonds. Choi, Huh, Shin (2023) clear these concerns by performing several robustness tests and showing that their results involving *Spread* are not sensitive to their choice of reference prices. (Choi, Huh, Shin, 2023)

V. Conclusion

In this work I explore the relationship between trade size and transaction costs in OTC markets, both theoretically and empirically. Contrary to the search-theoretic implication by Lagos and Rocheteau, which predicts an increase in transaction costs with larger trade sizes, my empirical analysis finds a significant negative influence of trade size on transaction costs. While other empirical studies find a similar relationship, Pinter, Wang, and Zou that utilize a different

[15] As interdealer trades tend to be smaller there a low cutoff date of $100,000 is chosen. (Choi, Huh, Shin, 2023)

dataset and control for investor characteristics identify a “size penalty” that lets transaction costs increase with larger trade size. Accordingly, the tested implication by Lagos and Rocheteau is empirically substantiated.

I present extensive summary statistics for the TRACE enhanced dataset and specifically for the utilized variables transaction costs and trade size as well as trade volume. I discuss observed patterns throughout the sample period and most interestingly the different behaviour of transaction costs during the two crises in the sample period: the subprime crisis and the COVID-19 crisis.

References

Helen Allen, John Hawkins, and Setsuya Sato. 2001. “Electronic Trading and its Implications for Financial Systems”. BIS papers, 7: 30-52.

Jack Bao, Jun Pan, and Jiang Wang. 2011. “The Illiquidity of Corporate Bonds”. The Journal of Finance, 66(3): 911-946.

Jaewon Choi, Yesol Huh, and Sean Seunghun Shin. 2023. “Customer Liquidity Provision: Implications for Corporate Bond Transaction Costs”. Management Science, 70(1): 187-206.

Robert Czech, Shiyang Huang, Dong Lou, and Tianyu Wang. 2021. “Informed Trading in Government Bond Markets”. Journal of Financial Economics 142(3): 1253-1274.

Jens Dick-Nielsen, Peter Feldhütter, David Lando. 2012. “Corporate bond liquidity before and after the onset of the subprime crisis”. Journal of Financial Economics, 103: 471-492.

Jens Dick-Nielsen. 2014. “How to clean Enhanced TRACE data”. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2337908

Darrel Duffie, Nicolae Gârleanu, and Lasse Heje Pedersen. 2005. “Over-the-Counter Markets”. Econometrica, 73(6): 1815-1847.

Peter Feldhütter. 2012. „The Same Bond at Different Prices: Identifying Search Frictions and Selling Pressures“. The Review of Financial Studies, 25(4): 1155-1206.

Richard C. Green, Burton Hollifield, and Norman Schurhoff, 2007. “Financial Intermediation and the Costs of Trading in an Opaque Market”. The Review of Financial Studies, 20(2): 275-314.

Julien Hugonnier, Benjamin Lester, and Pierre-Olivier Weill. 2019. „Frictional Intermediation in Over-the-Counter Markets”. Review of Economic Studies, 87: 1432-1469.

Mahyar Kargar, Benjamin Lester, David Lindsay, Shuo Liu, Pierre-Olivier Weill, Diego Zúñiga. 2021. „Corporate Bond Liquidity during the COVID-19 Crisis". The Review of Financial Studies, 34(11): 5352-5401.

Ricardo Lagos and Guillaume Rocheteau. 2009. "Liquidity in Asset Markets with Search Frictions". Econometrica, 77(2): 403-426.

Maureen O'Hara, Xing (Alex) Zhou. 2021. "Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis". Journal of Financial Economics, 142: 46-68.

Gabor Pinter, Chaojun Wang, and Junyuan Zou. 2022. "Size discount and size penalty: Trading costs in bond markets". Bank of England Staff Working Paper No. 970

Andreas C. Rapp and Martin Waibel. 2023. "Managing Regulatory Pressure: Bank Regulation and its Impact on Corporate Bond Intermediation". Swedish House of Finance Research Paper No. 23-12.

Hans Stoll. 2006. "Electronic Trading in Stock Markets". Journal of Economic Perspectives, 20(1): 153-174.

Appendix

A.1 Proof of the Behavior of Trade Value and Marginal Transaction Costs

Lemma 4 in Lagos and Rocheteau (2019) states:

For an investor who holds asset position $a \geq 0$ *and wishes to trade* $|a_i - a| > 0$, *both* $\partial \varphi_i(a)/\partial a$ *and* $\partial/\partial a[\varphi_i(a)/|a_i - a|]$ *have the same sign as* $a - a_i$.

I provide the proof according to Lagos and Rocheteau in the following:

Consider the derivative of the transaction costs in respect to a

$$(14) \qquad \frac{\partial \varphi_i(a)}{\partial a} = -\frac{\eta}{r+\kappa}[\bar{u}_i{}'(a) - q]$$

Suppose that the nonnegativity constraint on a_i is slack. Then, since $\bar{u}_i$ is strictly concave and $\bar{u}_i'(a_i) - q = 0$, we know that $\bar{u}_i'(a) - q < 0$ if and only if $a - a_i > 0$, and $\partial \varphi_i(a)/\partial a$ has the same sign as $a - a_i$. If $a_i = 0$, then $a > a_i$ and $\bar{u}_i'(a) - q < \bar{u}_i'(a_i) - q \leq 0$, so $\partial \varphi_i(a)/\partial a > 0$, which is the same sign as $a - a_i = a > 0$. This establishes the first part. To show the second we consider the derivative of the transaction costs per asset traded with respect to a

$$(15) \qquad \frac{\partial}{\partial a}\left[\frac{\varphi_i(a)}{a_i - a}\right] = \frac{\eta}{r+\kappa}\left[\frac{\bar{u}_i(a_i) - \bar{u}_i(a) - \bar{u}_i{}'(a)(a_i - a)}{(a_i - a)^2}\right]$$

which is strictly negative, since $\bar{u}_i$ is strictly concave.

A.2 Simulation of the Distribution of Asset Holdings

In the steady state, as defined by equations (3) and (4), the choice of asset holdings is determined by

$$(16) \qquad \bar{u}_i{}'(a_i) = rp$$

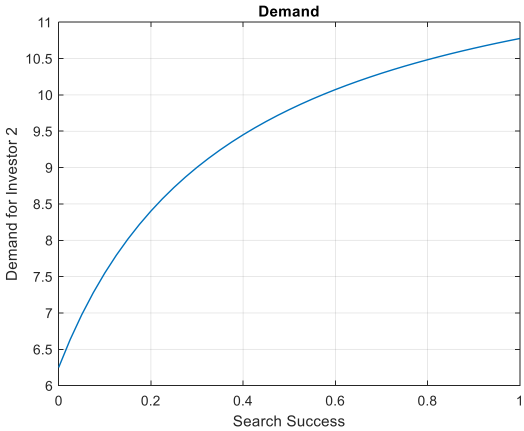
Let $a_i = g_i(\kappa; p)$ denote the choice of asset holdings characterized by the above equation. Then

(17) $$\frac{\partial g_i(\kappa;p)}{\partial \kappa} = \frac{\delta\left[\bar{u}_i{}'(a_i) - \sum_{j=1}^{I} \pi_j \bar{u}_j(a_i)\right]}{-\bar{u}_i{}''(a_i)(r+\kappa+\delta)^2}$$

has the sign of $u_i{}'(a_i) - \sum_{j=1}^{I} \pi_j u_j{}'(a_i)$ (Lagos and Rocheteau, 2009). Consequently, an investor whose current marginal utility from holding the asset exceeds his expected marginal valuation over the holding period increases his demand when the trading frictions κ increase.

In Figure 5, I illustrate this relationship for investor 2, characterized by a current utility profile of $\varepsilon_i = 1.25$. κ denotes the trading frictions and is determined as $\kappa = \alpha\,(1-\eta)$, where η represents the dealer bargaining power and α is the Poisson arrival rate at which investors meet dealers. I refer to α as search success. The plotted graph depicts the demand of investor 2 for varying values of search success α, with dealer bargaining power η held constant, meaning the investors demand is depicted for varying values of trading frictions κ.

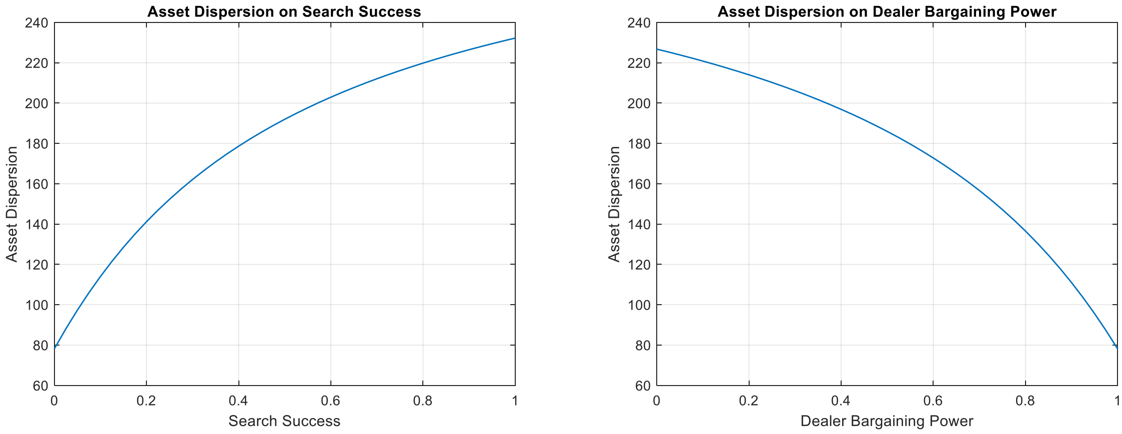
Figure 5: Demand of Investor 2 depending on Search Success



In the case of investor 2 $u_i'(a_i) > \sum_{j=1}^{I} \pi_j u_j{}'(a_i)$. The investor anticipates that his valuation of the asset holdings is likely to revert toward $\sum_{j=1}^{I} \pi_j u_j{}'(a_i)$ in the future. As rebalancing may be delayed due to trading frictions, his choice of asset holdings is lower than in a world with no trading frictions. An increasing search success rate α means the investor is more likely to find a dealer quickly and therefore will give more weight to his current marginal utility, leading to an increase in demand. (Lagos and Rocheteau, 2009)

This means that the asset holdings become more dispersed as trading frictions decrease. Figure 2 illustrates how asset dispersion[16] behaves with varying values for search success and dealer bargaining power: Asset dispersion increases with search success (denoting decreasing trading frictions), and an increase in the dealer bargaining power (denoting increasing trading frictions) leads to diminishing asset dispersion.

Figure 6: Asset Dispersion and Trading Frictions



Note: When observing the effect of search success, I keep bargaining power constant at its midpoint and vice versa.

A.3 Simulation of Trade Volume

The observed behavior of asset dispersion implies that with decreasing trading frictions investors need to trade larger quantities when rebalancing their asset holdings meaning the trade size is expected to increase with decreasing trading frictions. However, an increase in search success α has additional effects on the total trade volume: Trade volume increases as more investors gain access to the market and can trade. Finally, the measure of investors that are mismatched to their desired asset position and seek to trade decreases. The two latter effects combined lead to an increase in trade volume. Moreover, in case of logarithmic preferences, the first effect lets

[16] I compute the asset dispersion as the variance of the equilibrium asset holdings for each investor.

trade volume unambiguously increase with a reduction in trade volume (Lagos and Rocheteau, 2009).

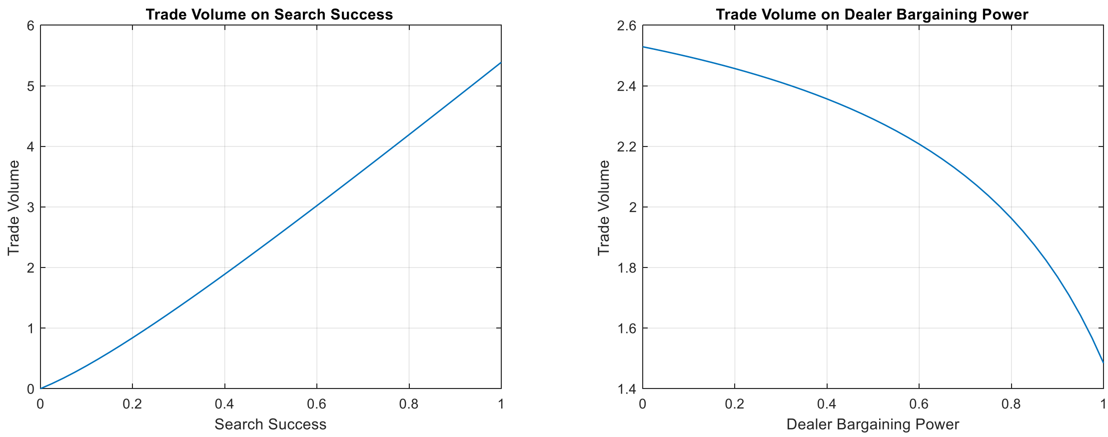
I show this effect of varying search delay as well as dealer bargaining power on Trade Volume in Figure 7. I define a Trade Volume as

$$V = \frac{\alpha}{2} \sum_i^I n_i |a_i - a| \tag{18}$$

where n_i denotes the density of investors with current preference type i and a_i is the choice of asset holdings of an investor with current preference type i.

As implied, decreasing trading frictions (increased search success and decreasing dealer bargaining power) have a positive effect on the trade volume.

Figure 7: Trade Volume and Trading Frictions



A.4 Robustness tests

I utilized robust standard errors since both the White-test and the Breusch-Pagan-test indicated heteroscedastic errors. Additionally, I run this regression on varying subsamples for different periods and trade sizes. Results of the regression for a restricted sample of trade sizes below $1 million is shown in Table 4. I consistently obtain negative coefficients for trade size. Notably, the regression does not obtain significant results for trade sizes above $1 million.

Table 4: Regressions of Bid-Ask Spreads on Trade Size and Trade Characteristics for Trade Size below $1m

Cusip	Observations	log(Trade Size)	Dealer-Buy	Customer Trade	Agency Trade	R-squared
61748AAE6	187,098	-6.45***	-54.11***	150.17***	-87.91***	0.282
		(0.251)	(1.014)	(0.657)	(0.768)	
369604BQ5	176,840	-7.03***	-126.30***	234.67***	-114.20***	0.484
		(0.220)	(0.936)	(0.740)	(0.832)	
92978AAA0	137,696	-17.84***	-166.80***	299.26***	-179.43***	0.490
		(0.322)	(1.082)	(0.968)	(1.136)	
369604BC6	138,353	-9.69***	-98.93***	187.73***	-115.20***	0.360
		(0.243)	(1.096)	(0.957)	(0.764)	
037833AK6	121,253	-2.26***	-14.18***	46.30***	-24.85***	0.107
		(0.122)	(0.518)	(0.476)	(0.410)	

Notes : *, **,and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Furthermore, I regress spread on a varying selection of independent variables as for example on the absolute value of trade size. These robustness test confirm a negative influence of trade size on the spreads.

Finally, I estimate *Spread* on all bond trades for bond that are traded more than 20,000 times with fixed effects for each bond *j*:

$$Spread_i = \alpha_j + \beta_1 \, log(Trade\ Size_i) + \beta_2 Dealer - Buy_i + \beta_3 Customer\ Trade_i + \beta_4 Agency\ Trade_i + \varepsilon_i \quad (19)$$

The result of this regression, presented in Table 5, shows a significant negative influence of trade size on the bid-ask spreads. Furthermore, all coefficients are consistent with my baseline regressions in their sign and size.

Table 5: Regression of Bid-Ask Spreads on Trade Size and Trade Characteristics for Bonds with at least 20,000 Trades with Fixed Effects for each bond

Observations	log(Trade Size)	Dealer-Buy	Customer Trade	Agency Trade	R-squared
29,802,525	-15.56***	-62.19***	162.01***	-81.64***	0.303
	(0.013)	(0.076)	(0.064)	(0.076)	

A.5 Dick-Nielsen Filter

The Dick-Nielsen Filter runs in 3 steps. I explain each of the steps in the following.

The first step removes corrections and cancellations reported on the same date. These can be uniquely identified by the link between the message sequence number and the original message sequence number as this variable is unique on an intra-reporting day level. (Dick-Nielsen, 2014)

The second step removes reversals and the matching original trade report. Reversals are cancelations reported on a later date than the date on which the original transaction took place. Each reversal should be linked to exactly one original transaction. If for some reason more than one transaction can be linked (by matching) to the same reversal, only one of the matching reports is deleted. The reversal report should exactly match the original report and its As-Of indicator contains an 'R'. (Dick-Nielsen, 2014)

The third step identifies interdealer transaction pairs (interdealer transactions that are reported by both sides), removes one of the reports and classifies the retained report as an interdealer transaction. Furthermore, this step removes agency transactions where the principal transaction has the same price as the agency transaction (a sort of double counting). Finally, I identify special transactions that are potential outliers such as commissioned transactions, trades with odd number of days to settlement etc.[17](Dick-Nielsen, 2014)

[17] Specifically, these are WI trades, trades that don't take place on the secondary market, trades of equity linked notes, trades with more than 6 days to settlement, not cash sales, commissioned trades and automatic give ups.

A.6 Average Trade Size – 5 Days Moving Average