A Structural Model of Liquidity in Over-the-Counter Markets*

Jamie Coen[†] Patrick Coen[‡]

November 2021

Link to latest version

Abstract

We study how firm heterogeneity determines liquidity in over-the-counter markets. Using a rich dataset on trading in the secondary market for sterling corporate bonds, we build and estimate a flexible model of search and trading in which firms have heterogeneous search costs. We show that the 8% most active traders supply as much liquidity as the remaining 92%. Liquidity is thus vulnerable to shocks to these firms: if the 4% most active traders stop trading, liquidity falls by over 60%. Bank capital regulation reduces the willingness of these active traders to hold assets and thus reduces liquidity. However, trader search, holdings and intermediation respond endogenously to reduce the welfare costs of regulation by 30%. These costs are greater in a stress, when these margins of adjustment are constrained. The introduction of trading platforms, which homogenise the ability of traders to trade frequently, improves aggregate welfare but harms the most active traders who currently profit from supplying liquidity.

Keywords: Liquidity, over-the-counter markets, financial intermediation.

*The views in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. We are particularly grateful to our supervisor, Alessandro Gavazza, and to Anil Kashyap and Dimitri Vayanos for many helpful discussions. We are also grateful for comments by Giulia Brancaccio, Eddie Gerba, Péter Kondor, Will Matcham, Ben Moll, Martin Oehmke, Martin Pesendorfer, Mark Schankerman, Pasquale Schiraldi, Nick Vause and Shengxing Zhang, as well as seminar participants at the Bank of England and London School of Economics. The authors acknowledge the financial support of the Economic and Social Research Council, the Paul Woolley Centre at LSE, the French National Research Agency and the TSE-Partnership foundation.

[†]London School of Economics, w.j.coen@lse.ac.uk.

[‡]Toulouse School of Economics, patrick.coen@tse-fr.eu.

1 Introduction

Key financial assets trade in over-the-counter (OTC) markets, where firms seeking to trade must search for trading counterparties. Market liquidity is the extent to which these firms can trade quickly and at low cost, and is a key part of the efficient functioning of markets. It is an equilibrium outcome of firms' trading behaviour, with firms supplying liquidity by enabling their counterparties to trade. Trading activity varies significantly across firms: most trade very rarely, but some trade frequently.

We study the implications of this firm heterogeneity for market liquidity. What is the distribution of liquidity supply across frequent and infrequent traders? This leads us to study how this heterogeneity interacts with financial stress, policy and technology. How resilient is liquidity to shocks to these frequent traders? What is the impact of bank capital regulation, where banks are amongst the most frequent traders in these markets? What is the impact of introducing trading platforms, which homogenise the ability of traders to trade frequently?

We build our response to these questions around a rich dataset on trading in the secondary market for sterling corporate bonds. We begin by establishing some key empirical facts regarding heterogeneity in how frequently traders trade, who supplies liquidity, and how traders vary their trading behaviour to manage their balance sheets. Based on these facts, we develop a flexible model of search and trading in an OTC market. In the model, traders with *ex ante* heterogeneity in search costs and *ex post* heterogeneity in asset holdings and valuations search for and trade an asset with each other. Traders' search intensity, asset holdings, and trading quantities and prices are endogenously determined. Intermediation is endogenous, in that any trader can supply liquidity. We estimate the model using the identifying information in our data on corporate bonds, and show it fits these data well.

Our model enables us to quantitatively study market liquidity, who supplies it, and how it changes in counterfactual scenarios, whilst estimating the model ensures our results are grounded in the data. This enables us to make the following contributions. We show that 8% of traders supply as much liquidity as the remaining 92%. This asymmetry in the supply of liquidity shapes the impacts of stress, policy, and technology. We show that liquidity is highly vulnerable to shocks to these key liquidity suppliers: if the 4% most active traders stop trading, liquidity falls by over 60%. We show that bank capital regulation negatively affects liquidity by reducing the incentives of these key suppliers to hold assets, but search, holdings, and intermediation adjust to reduce the welfare costs of regulation by 30%. These adjustments are constrained during stress, which causes the negative effects of capital regulation to increase. We show that whilst trading platforms increase aggregate welfare, they harm the most active traders who see their returns from supplying liquidity decrease. We offer this as an explanation for why platforms have not supplanted bilateral trading as the dominant trading mechanism in OTC markets.

Additionally, we are the first to set out a search model of trading in an OTC market with endogenous search and unconstrained holdings. We show that traders adjust their search intensity to manage their balance sheets, searching harder when they are far away from their target holdings. We show this can explain key trading patterns in OTC markets.

The starting point for our analysis is rich data on the transactions, holdings, and identities of traders of sterling corporate bonds. The data are novel, relative to datasets commonly used in the literature, in two key ways: (1) They include the identities of *all traders*,[1] which means we can study how the characteristics of all traders drive their trading decisions; and (2) They include data on the *holdings* of firms as well as their transactions, enabling us to study the role of firms' inventories in driving trading patterns.

We present a number of empirical facts which inform our work. We show that the distribution of trading frequencies across traders is heavily skewed. On average, the 6% most frequent traders in a bond are responsible for 50% of the trades. We then show that the most frequent traders act as intermediaries, channelling assets between traders and sellers and earning a spread by doing so. Intermediation is not limited to firms traditionally seen as dealers in these markets: over a quarter of trades do not involve a dealer, suggesting that customers may be supplying liquidity to each other and to dealers.

Finally we show that firms actively vary their trading frequency to manage their balance sheets. In particular, traders offset trades by buying and then selling (or selling and then buying) the same bond within a short interval. They do this far more frequently than would be the case if trade orders arrived exogenously through time. We then compare dealers' tendency to offset their purchases of bonds to variation in capital regulation across bonds and through time, and show that offsetting is more frequent when capital regulation is tighter. This is consistent with capital regulation leading dealers to manage their balance sheets more tightly.

Our data and empirical findings guide our modelling decisions in the following ways. The heterogeneity in trading frequency across traders leads us to allow for *heterogeneity*

[1] Comparable datasets in the US, for example, only include the identities of firms that traditionally act as dealers in these markets.

in traders' search costs. The fact that a trader's trading frequency appears to respond to their state and the characteristics of the asset being traded leads us to consider *endogenous search intensity*. Given that trading occurs both between dealers and between customers, and the fact that our data contain the trading behaviour of all types of agents, we treat *intermediation as endogenous*.

We develop and estimate a model of trading in a decentralised financial market, building on Üslü (2019) by endogenising search intensity. In the model, a continuum of forward-looking traders trade an asset with each other in a market characterised by search frictions. Traders' holdings of the asset are unrestricted. Traders face random shocks to the utility they derive from holding the asset, which creates heterogeneity in liquidity needs and thus gains from trade. Traders choose how frequently to meet a counterparty subject to a convex cost of searching. These search costs vary across traders. A trader who meets a counterparty draws this counterparty randomly from the trading population. Intermediation in this model is endogenous, in the sense that rather than designating traders as intermediaries, we study how *ex ante* heterogeneity in search cost and *ex post* heterogeneity in liquidity needs across traders determine intermediation.

We estimate the model by matching theoretical moments to those we observe in the data. The novel aspects of our data, notably the presence of identifiers for all market participants, allows us to exploit the identifying information contained in a broad range of moments across and within traders. The search cost distribution is identified by the trading frequency distribution, whilst utility parameters and shock frequency are identified by the distribution of prices, trading quantities, and asset holdings.

We study how traders search in equilibrium. Traders increase their search intensity when their trading needs are high. This creates a convex pattern of trading frequency in traders' asset holdings, with traders searching harder when they are far from a target asset holding. In equilibrium there is wide heterogeneity in trading costs, with some traders finding it much more costly than others to meet counterparties. Traders with lower search costs are more willing to trade to extreme asset positions, as they are able to return to target holdings faster. This means that traders with low search costs naturally emerge as intermediaries.

We then run counterfactual analyses about how firm composition, financial stress, regulation, and technology interact with firm heterogeneity to determine liquidity. In the first counterfactual, we quantify traders' contributions to market liquidity. To do this, we 'withdraw' sets of traders from markets, which entails them selling their asset and stopping searching for counterparties. By studying how market liquidity changes when traders are withdrawn, we can quantify how much liquidity they supply.

The 8% most frequent traders supply as much liquidity as all other traders combined. As a result, market functioning is highly vulnerable to shocks that cause these traders to withdraw from markets. If only the 4% most frequent traders withdraw, liquidity falls by over 60% and price volatility more than doubles. This result is driven by the skewed distribution of estimated search costs, which in turn is driven by the skewed distribution of trading we observe in the data.

Second, we examine the potential effects of capital regulation on banks. Many of the most active traders in OTC markets are banks. Thus whilst there is broad agreement that tighter capital regulation enhances the stability of the banking system, there are concerns that, by reducing banks' willingness to hold assets, it also reduces market liquidity. To study this potential unintended consequence of capital regulation, we simulate an increase in the cost of holding assets for the most frequent traders only, whom we term dealers.

Liquidity declines, as dealers are less willing to take large positions in a bond. Dealers shrink their asset holdings and adjust their search intensity to more tightly manage their inventories. Unregulated traders pick up the slack by increasing their asset holdings, causing their welfare to increase. These endogenous responses mean that the welfare cost of regulation is 30% lower than it would be if traders did not adjust their behaviour. However, these impacts increase in periods when non-dealers are subject to stress, as they are less able to offset dealers' reduced willingness to hold the asset.

These findings rationalise a number of trends in markets in recent years, reported both in this paper and in the literature. Firstly, dealers have reduced their corporate bond inventories (Dick-Nielsen and Rossi, 2019). Secondly, dealers increasingly tightly manage their holdings, organising trades that offset incoming trade orders either instantly or soon after the trade (Schultz, 2017). Third, liquidity during stress events appears to have deteriorated relative to before the financial crisis (Bao, O'Hara and Zhou, 2018; Dick-Nielsen and Rossi, 2019). Fourth, traders who traditionally do not operate as intermediaries can make money by supplying liquidity (Choi and Huh, 2018; BlackRock, 2015; Li, 2021). Each of these can be rationalised as the product of regulation increasing dealers' costs of holding inventory, and the endogenous responses of traders.

In our third counterfactual analysis we study the impacts of technologies that decrease search costs, such as electronic many-to-many trading platforms. The bulk of corporate bond trades are organised bilaterally, between traders who communicate and bargain by phone. The same is true of the markets for derivatives and structured products. In recent years trading platforms have emerged as alternative trading mechanisms in these markets, where trades take place electronically and offered trades are posted to all participating traders rather than communicated bilaterally. These types of trading mechanisms offer clear efficiency benefits, though their progress has been relatively slow (The Economist, 2020).

We study the effects of these platforms by undertaking two counterfactuals: (a) reducing and homogenising search costs across traders; and (b) studying the outcome in a frictionless, Walrasian case. In each simulation, we find that platforms improve aggregate welfare, but reduce the welfare of traders with the lowest search costs. We argue that this helps to explain the relatively slow adoption of these technologies: the most frequent traders—whose participation is required for a platform to be viable—would lose out under platform-based trading.

Below we discuss our contribution to the literature. In Section 2 we describe our data and describe the institutional setting. In Section 3 we set out some key patterns in trading, intermediation and regulation in the sterling corporate bond market. In Section 4 we describe our model, and in Section 5 we describe how we estimate the model. In Section 6 we show the results of our structural estimation and discuss some key implications. In Section 7 we undertake counterfactual analyses, before concluding in Section 8.

1.1 Literature

This paper's contribution is to quantify the heterogeneity in the supply of liquidity across traders, and to study the implications of this for the resilience of liquidity to stress, the impact of bank capital regulation, and the impact of changes in trading technologies. By doing this we contribute to three strands of literature: (a) empirical work documenting the structure of OTC markets; (b) empirical work on the determinants of market liquidity, including the roles of trader stress, regulation and technology; and (c) work modeling search frictions in OTC markets.

An empirical literature uses transaction-level data to document features of OTC markets. Most notably, a series of papers document a robust pattern across OTC markets: there exists a core of traders who trade frequently and a periphery who trade rarely (Di Maggio, Kermani and Song, 2017; Li and Schürhoff, 2019; Hollifield, Neklyudov and Spatt, 2017). Our contribution is that, with our estimated model, we can quantify what this implies for the heterogeneity in firms' *supply of market liquidity*. Liquidity is not something that can be measured directly in the data, and a firm's role in determining liquidity is effectively a counterfactual question about what liquidity would be if that firm did not exist. Quantifying this, as our paper does, is key to understanding how liquidity changes when conditions in markets change.

Our ability to quantify traders' roles in supplying market liquidity enables us to contribute the study of OTC markets under stress. Eisfeldt, Herskovic, Rajan and Siriwardane (2020) study the impact of intermediary exit in the market for credit default swaps (CDS). They develop a network model of the CDS market, and show that the removal of a single intermediary can increase credit spreads by 20%. Our contribution is to quantify the reliance of *liquidity* on a set of key firms.

Our paper contributes to a literature studying banking regulation and market liquidity.[2] A set of empirical papers including Adrian, Fleming, Shachar and Vogt (2017), Bessembinder, Jacobsen, Maxwell and Venkataraman (2018), Bao, O'Hara and Zhou (2018), Schultz (2017), Dick-Nielsen and Rossi (2019) and Choi and Huh (2018) seeks to understand the effect of post-crisis regulatory changes on market liquidity by comparing measures of liquidity before and the after the financial crisis. A related set of papers studies violations of no-arbitrage conditions—including covered-interest parity (Du, Tepper and Verdelhan, 2018) and the relationship between bond yields and credit default swap rates (Duffie, 2010*b*)—and relates these to post-crisis banking regulation. Finally, two theoretical papers study the impact of capital regulation on liquidity. Cimon and Garriott (2019) and Saar, Sun, Yang and Zhu (2020) show that capital regulation incentivises dealers to intermediate in a way that minimises the inventory they hold. In particular, both show that capital regulation increases the extent to which dealers operate as 'matchmakers' between buyers and sellers, and decreases their incentives to take assets onto their balance sheets.

To the best of our knowledge, we are the first to study banking regulation and market liquidity in a structural context. This enables us to quantitatively identify the mechanisms by which regulation affects liquidity, and thus explain a number of recent trends in markets. It also enables us to study the impact of capital regulation in counterfactual scenarios, most notably in times of stress.

We also contribute to a literature studying how different trading mechanisms impact financial market outcomes (Allen and Wittwer, 2021; Hendershott and Madhavan, 2015;

[2]See Vayanos and Wang (2013) for a survey.

Plante, 2018). We show that distributional effects mean platform-based trading—even if it improves efficiency—may not be implemented.

A large theoretical literature studies search frictions in financial markets. Examples include Afonso and Lagos (2015), Brancaccio, Li and Schürhoff (2020), Duffie, Gârleanu and Pedersen (2005, 2007), Farboodi (2021), Farboodi, Jarosch and Shimer (2021),Gavazza (2016), Gârleanu (2009), Gromb and Vayanos (2018), Hugonnier, Lester and Weill (2020), Lagos and Rocheteau (2007), Lagos and Rocheteau (2009), Liu (2020), Neklyudov (2019), Sambalaibat (2018) and Vayanos and Weill (2008).[3] Our model builds most closely on Üslü (2019), who sets out a theoretical model of endogenous intermediation between traders with different fixed trading speeds. We show empirically that traders appear to condition their trading frequency on their state and the state of the market. We thus endogenise firms' trading speeds, which gives us a more flexible model with which to run counterfactuals and enables us to explain key trading patterns.[4] We show how to identify and estimate this class of model, and in so doing are able to provide quantitative results. We demonstrate how firms adjust their search intensity to manage their trading portfolios, and how this changes in counterfactual scenarios.[5]

[3] An alternative literature takes a network-based approach to studying OTC markets, relaxing the assumption that traders search randomly and studying the network of relationships between traders. Examples include Babus and Kondor (2018), Chang and Zhang (2021), Coen and Coen (2019) and Craig and Ma (2021).

[4] Liu (2020) also studies endogenous search, but in a setting with binary asset holdings, where traders are pre-assigned as dealers—who do not receive liquidity shocks—or customers. The fact that we leave holdings unconstrained means we can study how traders condition their search intensity on their asset holdings and how this impacts the distribution of their asset holdings, the fact that ours is a model of exogenous intermediation means we can study how search costs determine who intermediates, and the fact that we allow all traders to be hit by shocks enables us to study how traders use search to respond to shocks. Farboodi, Jarosch and Shimer (2021) study homogeneous traders' ex ante investment in a search technology, whereas we allow a trader's choice of search intensity to depend on both their type and their state.

[5] In independent, contemporaneous work Brancaccio and Kang (2021) study endogenous search with unconstrained holdings in the context of the US municipal bond market. They estimate a model of exogenous intermediation to study how search frictions interact with the features of bonds that municipalities issue.

2 Data and Institutional Setting

2.1 Data

We use a combination of four datasets. Our primary dataset is a database on corporate bond transactions maintained by the Financial Conduct Authority (FCA).[6] This contains trade-level data on secondary-market trades in bonds where at least one of the firms is an FCA-regulated entity. In practice this means that almost all financial firms with a legal entity in the UK (including subsidiaries of foreign banks) appear in the dataset, including both banks and non-banks. Our sample covers trading in sterling corporate bonds from January 2012 to December 2017. For each trade, we see who is buying and selling the bond, the price, the quantity traded, the instrument traded, and the time of the trade.

Relative to other datasets typically used in the literature, the advantage of this dataset is that it includes the identity of *all traders*. Traders in OTC markets are typically referred to as belonging to one of two groups: *dealers*—who traditionally supply liquidity in these markets—and *customers* consisting of all other traders. In our data, we take dealers to be firms that are permitted to trade in the primary market with national banks as well as inter-dealer brokers, who exist to facilitate trades between dealers. We refer to all other traders as customers. Dealers tend to be banks. Existing studies based on US data typically observe the identity of dealers, but not customers.[7] In our data we observe both dealers and customers.

This feature of the data is crucial for our paper. It means we can characterise the trading activity of all traders, and use this to study which traders take on different roles in the market. In particular, it allows us to study intermediation as an endogenous outcome, rather than assuming that dealers are the intermediaries. Our estimated impacts of stress, bank capital regulation and trading platforms on liquidity will depend critically on parameters in our model that are identified by the distribution of trading frequencies across traders. Without knowing the identities of all traders, we could not characterise this distribution.

[6] Other studies using these data include Czech and Roberts-Sklar (2019), Czech, Huang, Lou and Wang (2021) and Mallaburn, Roberts-Sklar and Silvestri (2019).

[7] Existing studies typically rely on TRACE data on US corporate bond transactions (for example Choi and Huh, 2018; Trebbi and Xiao, 2017; Kargar, Lester, Lindsay, Liu, Weill and Zuniga, 2021) or US municipal bond transactions (for example Brancaccio, Li and Schürhoff, 2020; Hugonnier, Lester and Weill, 2020). TRACE data comes in two forms: an academic version which includes anonymised dealer identifiers, and a regulatory version which reveals the dealer identities. In both cases, any customer is simply marked as 'C', meaning customers can neither be identified nor tracked through the data. The municipal bond dataset includes dealer identities, but again provides no information on the customer.

We add to this data on the bond *holdings* of seven banks at end-2016 and end-2017, using Bank of England data, and 300 mutual funds quarterly from 2012 to 2017, using data from Morningstar. This enables us to separate who holds a bond from who trades it, and to relate trading decisions to traders' inventories of assets. Finally, we match these three datasets with information on bond characteristics and primary issuance from Thomson Reuters' Eikon database.

2.2 The Secondary Market for Sterling Corporate Bonds

The sterling corporate bond market is a key source of financing for both UK and non-UK firms. It has increased in importance since the financial crisis, with virtually all net financing raised by UK private non-financial firms from 2009 to 2016 coming in the form of bonds rather than bank lending (Bank of England, 2016). It is largely an over-the-counter market, with traders determining the terms of trade bilaterally, typically by phone (Anderson, Webber, Noss, Beale and Crowley-Reidy, 2015; Czech and Roberts-Sklar, 2019). Traders in the market consist of dealers, asset managers, insurance companies, hedge funds, and non-dealer banks. Dealers are a counterparty in around three-quarters of trades in our dataset. 80% of trades are carried out on a principal basis—where a firm who holds a bond sells the bond directly to a counterparty—rather than on an agency basis—where a firm acts as a middleman between two other trading counterparties, and at no point takes the bond onto its own balance sheet. Of these agency trades, the bulk are a trading counterparty buying for a non-trading client—for example a wealth manager buying bonds on behalf of their clients. 92% of trades between trading firms are carried out on a principal basis.

Table 1 gives key summary stats from our dataset. The average trading price is 108% of a bond's face value with significant variation around this figure. The median trade size is £989,000 with a tail of larger 'block trades' that skews the distribution right. Across the market there are around 1,000 firms trading each month. There is significant heterogeneity in trading activity both across instruments and across traders.

The market is illiquid. Trading tends to be rare, with the median bond trading once a month. There is large variation in price across bonds, and dispersion in price for the same bond: the R-squared from a regression of trading price on instrument fixed effects is 72%, with the remaining 28% reflecting within-bond price dispersion. Where a trader buys and sells a bond in the same week—a measure of the spread the trader earns—the median difference between the purchase price and the sale price is 0.13% of the bond's face value.

Table 1: Summary Statistics

	Mean	Std. Dev.	Median	25^{th} pctile	75^{th} pctile
Aggregate					
Price (% par)	108	13	106	101	114
Trade size (£000)	475	989	100	14	405
Monthly volume (£bn)	22	4	22	19	25
Monthly traders	975	68	980	929	1,021
Instrument-level					
Issuance (£mn)	135	403	10	3	200
Trades per month	10	26	1	0	7
Number of traders	35	66	4	2	31
Trader-level					
Monthly volume (£000)	10,456	102,915	30	2	458
Instruments traded	64	210	7	2	27
Trades per instrument traded	5	34	2	1	3

Note: This table gives some key descriptive statistics from our data. Aggregate statistics are computed across all instruments and all traders. Instrument-level statistics show how issuance and trading vary across instruments. Trader-level statistics show how trading activity varies across traders. Trades per instrument shows the distribution of the ratio between a trader's total trades and the number of instruments in which they trade.

In the rest of this paper, our analysis is performed at the bond level. As such, we only include bonds where we have a meaningful number of trades with which to characterise trading. In what follows, we restrict the dataset to bonds that have been traded at least 10 times over our sample period of 6 years. This removes under 1% of the trades in our dataset. Further details on how we prepare the data are given in Section A1.

2.3 Banking Regulation

Banks are subject to two types of capital regulation: a leverage requirement and a risk-weighted capital requirement. The leverage requirement states that a bank's equity must exceed a given fraction of its total assets. The risk-weighted capital requirement requires that a bank's equity exceeds a given fraction of its total risk-weighted assets.[8] Risk weights vary across bonds according to their creditworthiness, and capital requirements have more than doubled since the 2008 global financial crisis.[9]

[8]Where the measures of equity differ across the two requirements (BCBS, 2017).

[9]See Figure A1 for details.

3 Empirical Facts

We exploit the richness of our data to uncover a number of empirical facts that motivate our questions and guide our modelling. A key theme in these facts will be heterogeneity in the roles of different traders, both in terms of how frequently they trade and the roles they play. This heterogeneity will shape our modelling assumptions, and in our results it will play a major role in determining the effects on markets of shocks to traders, capital regulation, and the introduction of trading platforms.

Fact 1: 6% of traders trade as frequently as the remaining 94% combined.

Figure 1 shows that in each bond a small subset of traders are responsible for the majority of trading. To show this, for each bond we compute the average trading frequency of each trader, compute the distribution of this across traders, and plot the average of this across bonds. The distribution is heavily positively skewed: a small set of firms trades very frequently, and a long tail of firms trades infrequently.

This fact leads us to treat traders' search costs as heterogeneous in our model, and in estimation to model these search costs as a Gamma distribution, which is well suited to capturing positively skewed distributions. In our results the shape of this distribution will play a major role in determining the effects of shocks to traders, capital regulation, and the introduction of trading platforms.

Fact 2: Fast traders intermediate.

Figure 2 shows that the traders that trade the most intermediate in these markets. Figure 2a shows how the ratio of a firm's net trading volume to its gross trading volume in a bond depends on its trading frequency. A firm's net-to-gross trading ratio is decreasing in the frequency with which it trades a bond. This implies that firms who trade a bond frequently trade in a balanced fashion—buying and selling the bond—whilst those who trade less frequently tend to trade in one direction only. Figure 2b shows how the price at which a trader traders a bond depends on how often they trade it. The purchase price is decreasing in a trader's trading frequency, whilst the sale price is increasing. As a consequence, infrequent traders typically pay a spread, whilst frequent traders earn a spread. This is consistent with evidence of a centrality premium documented by Di Maggio, Kermani and Song (2017) in the US corporate bond market, where more central dealers earn higher spreads. Intuitively, Figure 2b shows a similar relationship between prices and trading frequency holds across all traders rather than just dealers.

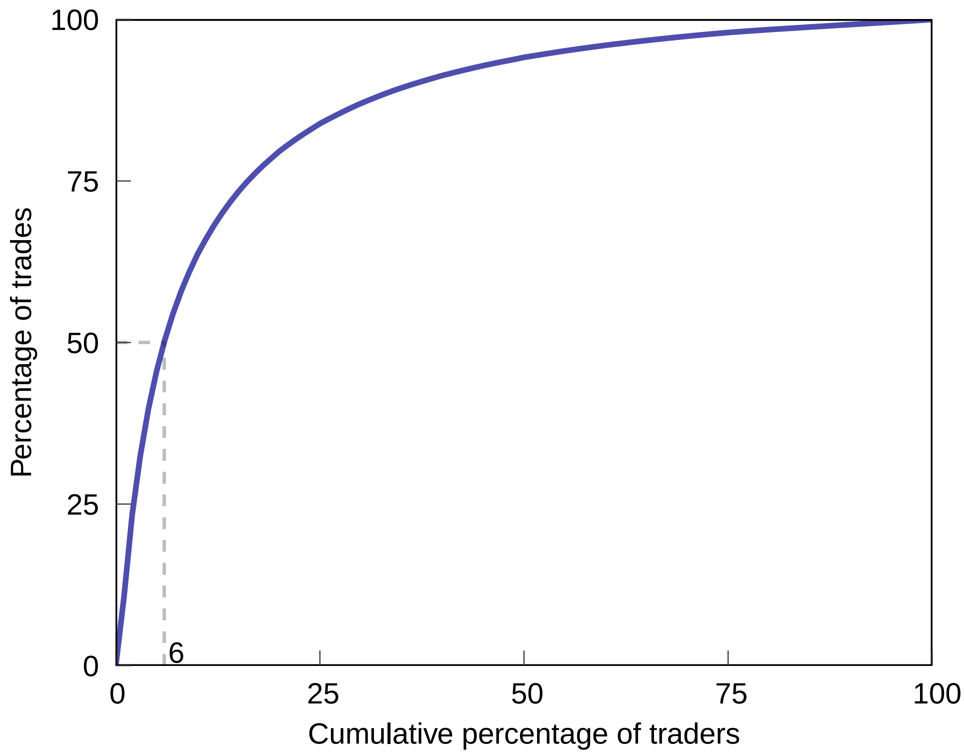


Figure 1: Trade frequency across traders

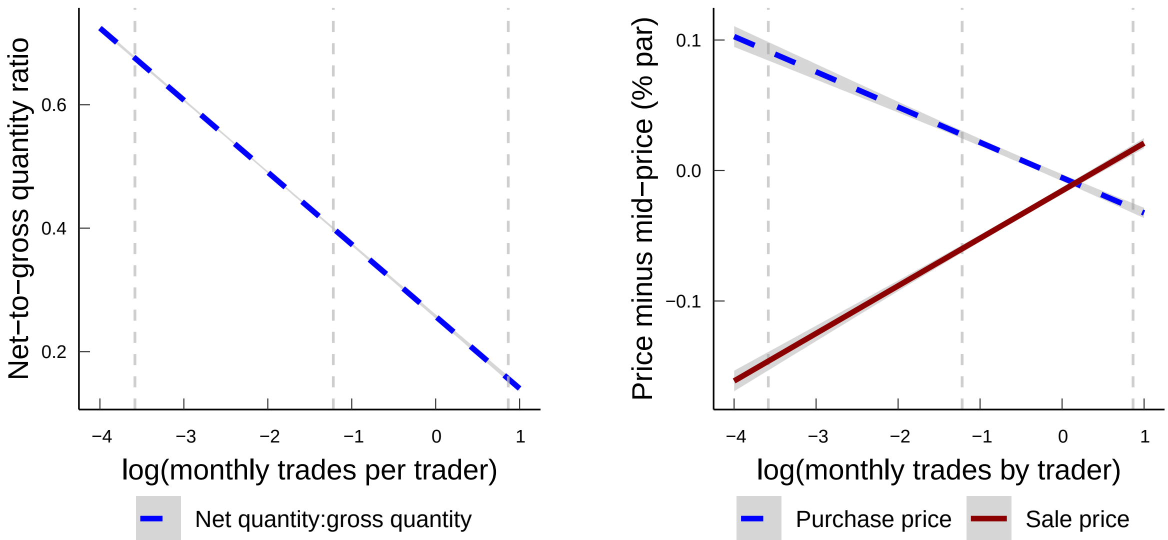
Note: This chart shows the distribution of trades across traders. For each instrument, we calculate the percentage of total trades carried out by the x% most frequent traders in that instrument. We then take the cumulative sum of this across all traders in the instrument. The chart shows the average of this across instruments, weighted by the number of firms trading in each instrument. The dashed lines show that on average the top 6% of traders are responsible for half of the trades in a bond.

These patterns are consistent with fast traders intermediating: trading to facilitate other traders' trading needs and earning a spread as a result. In our model we will show that fast traders emerge as intermediaries.

Fact 3: Dealers and customers both supply and demand liquidity.

This fact shows that liquidity supply is not restricted to traders traditionally thought of as dealers, whilst liquidity demand is not restricted to traders thought of as customers. Table 2 summarises the percentage of trades between dealers and customers. 23% of trades do not involve a dealer as counterparty, whilst 14% of trades are between two dealers. On the basis that all trades are the result of liquidity demand and liquidity supply, it follows that in some trades customers supply liquidity and in others dealers demand liquidity.

This is consistent with the institutional features of the firms that trade in bond markets.



(a) Net-to-gross trading ratio

(b) Trading prices

Figure 2: Intermediation and trading speed

Note: These figures show that the most frequent traders tend to intermediate. For the left figure, we take the absolute value of a firm's net trading in an instrument and divide it by their gross trading in the instrument. We then regress this figure—which lies between 0 and 1—on the log of the firm's average trades per month in that instrument. For the right figure, we take each trade by a trader, and subtract the mean trading price for that instrument in that month. We then regress this variable on the log of the average trades per month in that instrument by the firm buying the asset (blue line) and selling the asset (red line), and plot the fitted values in the figure. Firms on the left-hand side of the figure on average pay a spread, whilst those on the right-hand side earn a spread. The dashed vertical lines show the 25^{th}, 50^{th} and 75^{th} percentile of the variable on the x-axis, and the grey shaded areas are 95% confidence intervals.

The entities typically identified as dealers are generally parts of universal banking groups.[10] These firms do seek to supply liquidity and make money by doing so, but have a number of other motives for trading that they share with other types of financial firm. For example, these banks are big players in derivative markets. As their derivative positions change in value banks are subject to margin calls, which mean they must post cash collateral to their counterparties (Heller and Vause, 2012). Such shocks give banks reason to demand, rather than supply, liquidity in bond markets. Likewise, firms typically treated as customers can and do supply liquidity. As described in Choi and Huh (2018) and BlackRock (2015), firms such as asset managers increasingly seek to make money by supplying liquidity in these markets. This is consistent with evidence in other markets of non-dealers supplying liquidity (Biais, Declerck and Moinas, 2016; Franzoni, Plazzi and Çötelioğlu, 2019).

[10]See Duffie (2010*a*) for a list of major dealer banks.

Table 2: Trading by type (%)

Buyer\ Seller	Customer	Dealer
Customer	23	30
Dealer	33	14

Note: This table summarises the trading frequencies between dealers and customers.

Much of the literature treats dealers and customers as distinct entities, with the former supplying liquidity and the latter demanding liquidity.[11] This is formalised in models by (a) assuming customers receive shocks to their asset valuations that give them a motive to trade, whilst dealers do not (Grossman and Miller, 1988; Duffie, Gârleanu and Pedersen, 2005); and (b) assuming customers do not trade with each other. Intermediation is thus exogenous in these models: dealers supply liquidity and customers demand it. Our data paint a more nuanced picture of the sterling corporate bond market. As a result, we will treat intermediation as endogenous: we will not pre-specify who intermediates, but will instead study which traders intermediate in equilibrium. In particular, all traders will face shocks that lead them to demand liquidity and will be able to supply liquidity. And rather than separating traders out into dealers and customers, we will instead capture the differences between traders with heterogeneous model parameters.

Fact 4: Traders vary their trading frequency to manage their balance sheets. Table 3 shows that traders vary their trading frequency to manage their balance sheets. In particular, we show that traders trade the same bond on the same day much more frequently than would be implied by a model where trades arrive exogenously. To do this, we first compute the average empirical probability that, conditional on trading a given bond on a given day, a trader (a) trades that bond more than once that day ('paired trade'); (b) both buys and sells that bond that day ('offsetting trade'); and (c) trades the bond more than once that day where the trades are either all purchases or all sales ('amplifying trades'). We then show how often this would occur if trades arrived exogenously according to a Poisson process, where the arrival rate is equal to the trader's average trading frequency in that bond, and a trade is equally likely to be a buy or a sell. The probability of paired trades is much higher in the data than according to the model of exogenous trading. This is mostly

[11] Theoretical papers such as Üslü (2019) and Farboodi, Jarosch and Shimer (2021) study endogenous intermediation in theory, but papers that take search models to data typically treat intermediation as exogenous, partly due to a lack of data on the trades of non-dealers (Hugonnier, Lester and Weill, 2020; Brancaccio, Li and Schürhoff, 2020). Our novel data on customer identities allows us to consider endogenous intermediation empirically.

Table 3: Endogenous trading frequency

	Probability: Theory (%)	Probability: Data (%)
Paired trades	6.9	44
Offsetting trades	3.5	31
Amplifying trades	3.4	13

Note: This tables shows that traders vary their trading frequency to manage their balance sheets. Paired trades are when a trader trades the same bond more than once on the same day. Offsetting trades are when a trader both buys and sells the same bond on the same day. Amplifying trades are when a trader trades more than once in the same bond on the same day, and all these trades are in the same direction. The theoretical probabilities are computed assuming that trades arrive at Poisson rate λ, where λ is the average trading rate in the data, and buys and sells are equally likely. All probabilities are computed conditional on the trader trading that bond at least once on that day, and are computed separately for each bond before averaging across bonds.

driven by traders offsetting their trades.

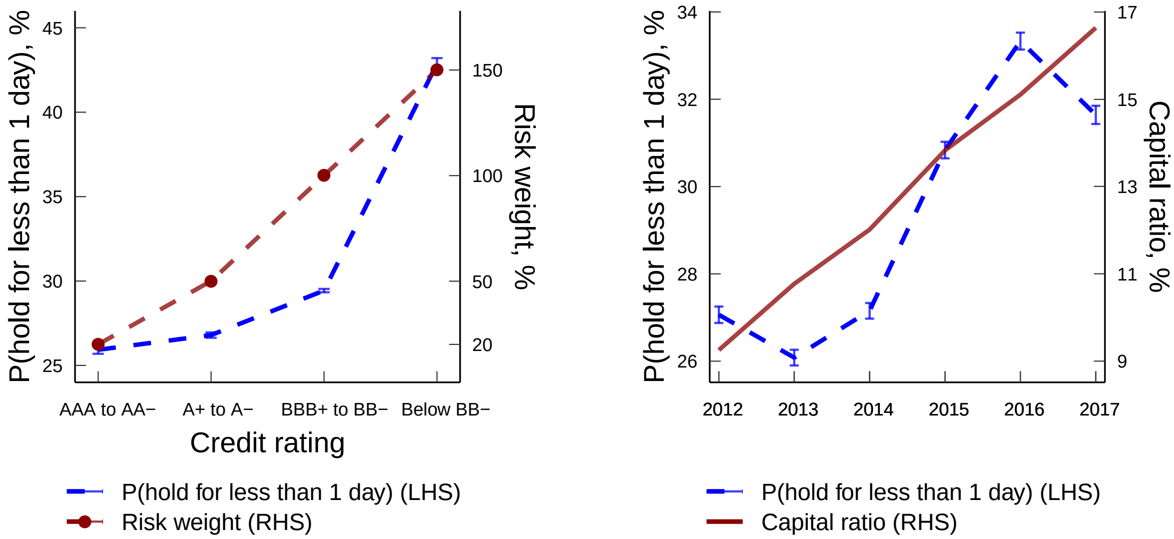
Most models of search assume traders contact each other exogenously. In particular, a trader's meeting rate does not depend on the gains to trade.[12] In models with unconstrained asset holdings such as Lagos and Rocheteau (2009) and Üslü (2019) every meeting results in a trade, and so trading rates are constant through time. Table 3 suggests traders vary their trading rates to manage their inventory. To rationalise this, in our model we will treat search intensity as an endogenous variable that traders choose and can condition on both their type and their state.[13]

Fact 5: Dealer trading frequency varies with capital regulation.

Figure 3 shows that dealers manage their bond inventories more tightly when capital requirements are high. To show this, for each purchase of a bond by a dealer, we compute the probability it offsets this purchase by selling the same bond on the same day. We then compare how this probability varies through time and across bonds, after controlling for the average frequency with which a bond trades. Figure 3a shows how the average offsetting probability varies across bonds with different credit ratings (blue line), and how these credit ratings map into different risk weights in bank capital regulation (red line). Figure 3 shows how the average offsetting probability varies through time (blue line) as well as UK banks' average capital ratios through time (red line). Traders were significantly more likely to offset

[12] A notable exception is Liu (2020), who endogenises dealer search intensity in a model with binary asset holdings. Farboodi, Jarosch and Shimer (2021) allow traders to make an ex ante investment to choose their meeting rate, but this meeting rate is then fixed through time.

[13] We note that this is not the only approach that could rationalise Table 3. Instead, one could introduce a fixed cost of executing a trade, which would mean that not all trader meetings result in a trade.



(a) Holding period by rating

(b) Holding period through time

Figure 3: Patterns in asset holding periods

Note: In this figure we show that the period of time for which a trader holds a bond is negatively correlated with capital regulation, both in the cross-section and inter-temporally. For each purchase by a dealer, we compute the length of time before they next sell that bond, or the bond matures. We create a dummy variable equal to 1 if this is less than 1 day, average this across all purchases and all dealers, and show how this varies by year (blue line in right panel) and by the risk weight the bond receives in banking regulation (blue line in left panel), after controlling for the bond's average trading frequency. The vertical lines show a 95% confidence interval around the point estimate. The red line in the left panel shows the average risk weight by bond credit rating according to the standardised approach to risk weighting, as described in BCBS (2006). The red line in the right panel shows UK banks' average capital ratios through time, taken from Bank of England (2021).

bond purchases at the end of the sample—when capital regulation was higher—than at the start of the sample, as well as in bonds with the highest capital requirement.

Capital regulation, to the extent that it increases traders' costs of holding inventory, increases the incentive for traders to minimise their asset holdings. Figure 3 is consistent with capital regulation increasing dealers' costs of holding assets, and dealers adjusting their trading behaviour to minimise these costs. This motivates our counterfactual in Section 7, where we model the impact of capital regulation on banks as an increase in dealers' costs of holding an asset. The fact that dealers appear to respond to these higher costs is a fact that we will seek to rationalise with our model.

4 A Model of Trading in an Over-the-Counter Market

In this section we set out a model of search and trading in a decentralised asset market. Informed by the empirical facts above, our model has the following features: (1) Trading frequencies are endogenously determined and heterogeneous; (2) Intermediation is endogenous, in that any trader in any trade can supply or demand liquidity; and (3) Traders' asset holdings are unlimited and endogenous. Üslü (2019) presents a model of endogenous intermediation with unlimited holdings and exogenous meeting rates between traders. We take this model and endogenise trader search intensity.

In our model, a set of traders trade a single asset. Traders face random shocks to their marginal utility of holding the asset that lead them to want to trade. Search frictions mean it takes time for a trader to find a counterparty with whom to trade. Traders can choose the rate at which they meet counterparties, but increasing this rate is costly. The costs of searching vary across traders, and at each point in time traders choose how hard to search to balance the costs of searching against the prospective gains from trade. Once two traders meet, they bargain over the trading price and quantity. An equilibrium is a steady state of this system, where search, the terms of trade, and asset holdings are the results of traders' optimising decisions.

4.1 The Setting

Agents and assets Time is continuous and the horizon infinite. There is a continuum of infinitely-lived traders with measure 1, who discount the future at rate $r > 0$. These traders derive utility from holding a numéraire good with marginal utility equal to 1, and can hold an unconstrained amount of a durable asset with fixed supply a. A trader whose asset holding is h and net consumption of the numéraire good is c receives utility flow $u(\beta, h) + c$, where

$$u(\beta, h) = \beta h - \kappa \frac{h^2}{2} \tag{1}$$

is the utility flow from holding the asset. A trader's lifetime utility is the present discounted value of expected utility flows, net of payments for transactions.

A trader draws a new valuation parameter β from the distribution $G(\beta)$ at Poisson arrival rate η. These random shocks to β create a motive to trade. These shocks can be interpreted as shocks to a firm's funding position. For example, a trader who receives a shock that lowers their valuation could be an asset manager who faces investor outflows, and thus needs to sell the asset to raise funds. An increase in valuation would then represent an asset manager facing investor inflows.

This utility function over holdings can be interpreted as a reduced form of a more fundamental setting where traders exhibit constant absolute risk aversion, and utility is derived from consumption rather than asset holdings (Duffie, Gârleanu and Pedersen, 2007; Üslü, 2019). In this setting, risk-averse traders seeking to maximise their utility choose how much wealth to allocate between a risky asset and a risk-free asset, and how much to consume. They have a stochastic income stream which is correlated with the returns to the risky asset, and this correlation changes stochastically, changing the desirability of holding the risky asset. These shocks underly the shocks to β in the reduced-form utility in (1). The parameter κ in equation (1) is a linear function of the trader's coefficient of absolute risk aversion. In the rest of this paper we will refer to κ as risk aversion. We work with the reduced-form utility, rather than the more fundamental setting, as this captures the main dynamics at play, and when we turn to estimation the parameters we can identify are those of the reduced-form utility.

Endogenous search, matching and trade Traders search for and meet counterparties with whom to trade. Traders choose the rate at which they meet counterparties, but meeting counterparties frequently is costly. Let γ denote a trader's search intensity. A trader searching with intensity γ incurs cost $s(z, \gamma)$, where parameter $z \sim F(z)$ capture a trader's cost of searching, which varies across traders and is fixed through time. The search cost function $s(z, \gamma)$ is a twice continuously differentiable, increasing and convex function of γ. Optimal search γ will be a function of the search parameter z, as well as the trader's valuation β and their current holding h.

The benefit to searching is that it results in meetings with other traders, and potential gains from trade. Meetings between traders searching with intensity γ and γ' are governed by a matching function $m(\gamma, \gamma')$, which is symmetric and linearly increasing in both arguments. This captures the fact that traders can both contact and be contacted by other traders, and that a trader can increase its meeting rate by increasing its search intensity γ.

Together, the search cost function and matching function capture the trade-off a trader faces in choosing how hard to search. A trader benefits by increasing the frequency at which it can trade, as this enables it to better optimise its portfolio and respond to shocks. But doing so, for example by hiring more people to the firm's trading desk or paying them to work overtime, is costly. Firms will choose search intensity to balance these costs and benefits. The benefits to trading are state-dependent—for example a trader may be eager to trade after receiving a shock to their asset valuation, whilst after having traded they may have little desire to trade again. As a result, traders will condition their search intensity on their state—their asset holdings h and their valuation β—as well as their search parameter z.

Once two traders have been matched they engage in bilateral Nash bargaining over the quantity traded q and the unit price p. All traders have bargaining strength of one half. When a trader of type (z, β, h) meets a trader of type (z', β', h'), the signed trading quantity is $q((z, \beta, h),(z', \beta', h'))$ and the unit price is $p((z, \beta, h),(z', \beta', h'))$.

4.2 Solving the Model

In this section we define and characterise a stationary equilibrium. Doing so requires that we analyse the three-way feedback between traders' search decisions, the distribution of the asset across traders, and the terms at which they trade. A trader's search decision depends on the distribution of the asset across traders and the terms of trade as these determine the gains to searching. Distributions depend on traders' search decisions and the terms of trade as these determine the flows of the asset between traders. Finally, the terms of trade depend on distributions and search as they determine future trading opportunities. To proceed we write down expressions for traders' value functions, their trading decisions, their search decisions, and the distributions of assets across traders. An equilibrium is then a mutually consistent, stationary relationship between these endogenous variables.

4.2.1 Trading and Value Function

We begin by setting out the equations that govern the quantity and price at which traders trade, before writing down their value function. Let $V(z, \beta, h)$ be the value function of a trader with search parameter z, valuation β and holding h. We assume $V(z, \beta, h)$ is differentiable, increasing and concave. In equilibrium it will inherit these properties from the flow utility function $u(\beta, h)$. When a trader of type $\Delta \equiv (z, \beta, h)$ meets a trader of type $\Delta' \equiv (z', \beta', h')$, the total surplus from trading quantity q is given by the sum of the two traders' changes in value after the trade:

$$V(z, \beta, h-q) - V(z, \beta, h) + V(z', \beta', h'+q) - V(z', \beta', h'). \tag{2}$$

Nash bargaining implies that the quantity they trade $q(\Delta, \Delta')$ and the unit price $p(\Delta, \Delta')$ jointly solve:

$$\begin{aligned} \underset{p,q}{\text{maximise}} \quad & (V(z,\beta,h-q) - V(z,\beta,h) + pq)(V(z',\beta',h'+q) - V(z',\beta',h') - pq) \\ \text{subject to} \quad & V(z,\beta,h-q) - V(z,\beta,h) + pq \geq 0, \\ & V(z',\beta',h'+q) - V(z',\beta',h') - pq \geq 0. \end{aligned} \tag{3}$$

The terms of trade maximise the product of the traders' surpluses, subject to each trader weakly preferring the agreed trade to not trading at all.[14]

The quantity that solves this problem must by definition maximise total trading surplus. The asset is thus sold by the trader who values the asset less to the trader who values it more, and the quantity sold is that which equalises the post-trade marginal value of the two traders:

$$V_3(z,\beta,h-q(\Delta,\Delta')) = V_3(z',\beta',h'+q(\Delta,\Delta')),$$

where $V_3(z,\beta,h)$ is the derivative of the value function with respect to holdings h. Intuitively, this means trading quantity is larger when the pre-trade marginal values of traders are more spread out, and when the curvature of the value function is lower (as then the trading quantity that equalises traders' marginal values must be greater). When we estimate the model in Section 5, these facts will mean the parameters governing the slope and curvature of the value function (value β and risk aversion κ) are identified by trading quantities.

The trading surplus when two traders meet is then simply the surplus (equation (2)) evaluated at the optimal trading quantity:

$$S\big((z,\beta,h),(z',\beta',h')\big) = V(z,\beta,h-q(\Delta,\Delta')) - V(z,\beta,h) + V(z',\beta',h'+q(\Delta,\Delta')) - V(z',\beta',h').$$

Evaluating the first-order-condition of equation (3) with respect to price and plugging in the optimal quantity $q(\Delta, \Delta')$ gives the following expression for the optimal unit price:

$$p(\Delta,\Delta') = \frac{1}{2}\Bigg(\underbrace{\frac{V(z',\beta',h'+q(\Delta,\Delta')) - V(z',\beta',h')}{q(\Delta,\Delta')}}_{\approx \text{ slope of } \Delta' \text{ value in } h'} + \underbrace{\frac{V(z,\beta,h) - V(z,\beta,h-q(\Delta,\Delta'))}{q(\Delta,\Delta')}}_{\approx \text{ slope of } \Delta \text{ value in } h} \Bigg).$$

[14] Given the traders' bargaining strengths are equal, they do not appear in the optimisation.

The price thus depends on the slopes of the traders' value functions. As the optimal trading quantity approaches zero, the price approaches the average of the two traders' marginal values. The variance of trading prices will therefore be greater when traders' marginal values are more spread out, and when the curvature of the value function is greater (as this increases the variation in the slope of the value function). These facts will mean that trading prices, as well as trading quantities, will help to identify the value shocks β and risk aversion κ in Section 5.

We can now write down an expression for traders' value functions, taking as given their trading decisions, their search decisions and the distribution of the asset across traders. The Hamilton-Jacobi-Bellman equation that governs traders' optimal behaviour can be written as:

$$\underbrace{r\,V(z,\beta,h)}_{\text{Value}} = \underbrace{u(\beta,h) - s(z,\gamma(z,\beta,h))}_{\substack{\text{Flow utility \&}\\ \text{search costs}}} + \underbrace{\eta \int (V(z,\beta',h) - V(z,\beta,h))G(d\beta') +}_{\text{Switch type}}$$
$$\frac{1}{2}\iiint \underbrace{m(\gamma(z,\beta,h),\gamma(z',\beta',h'))}_{\substack{\text{Meeting}\\ \text{probability}}}\underbrace{S((z,\beta,h),(z',\beta',h'))}_{\text{Surplus}}\,\Phi(dz',d\beta',dh'), \quad (4)$$

where $\Phi(z,\beta,h)$ is the cdf of traders of type (z,β,h), $\gamma(z,\beta,h)$ is the optimal search intensity for trader type (z,β,h), and $s(z,\beta,h) \equiv s(z,\gamma(z,\beta,h))$ are the corresponding search costs.

The value function has the usual asset pricing interpretation. A trader of type (z,β,h) gets flow utility $u(\beta,h)$ from holding the asset and incurs search costs $s(z,\gamma(z,\beta,h))$. At intensity η the trader switches type, drawing a new value β' from $G(\beta')$. At intensity $m(\gamma(z,\beta,h),\gamma(z',\beta',h'))\Phi(dz',d\beta',dh')$ the trader meets a counterparty with type (z',β',h'), in which case they trade and extract half the total surplus.

Traders' values thus depend on the search and trading decisions of both themselves and their potential future counterparties, and the distributions of the asset across traders. A trader's search determines the search cost it incurs as well as the expected gains to trade it enjoys. The distribution of the asset across traders, as well as these traders' search decisions, then determine how frequently the trader meets a counterparty, and which type of counterparty they meet. The Nash bargaining protocol then determines the gains to trade when they do meet a counterparty.

4.2.2 Trader Search

We now characterise traders' optimal search decisions, given their trading decisions and the distribution of the asset. To do so, we take the first-order condition of the value function in equation (4) and apply the envelope theorem, yielding the equation governing the optimal search of type $\Delta \equiv (z, \beta, h)$:

$$\underbrace{s_2(z, \gamma(\Delta))}_{\substack{\text{Marginal cost}\\ \text{of search}}} = \frac{1}{2} \int \underbrace{\frac{\partial m(\gamma(\Delta), \gamma(\Delta'))}{\partial \gamma(\Delta)}}_{\substack{\text{Increase}\\ \text{in meetings}}} \underbrace{S(\Delta, \Delta')}_{\substack{\text{Surplus from}\\ \text{meeting}}} \Phi(d\Delta'). \tag{5}$$

The left-hand side is the marginal cost of searching, which is increasing in the trader's search intensity. The right-hand side is the marginal benefit to searching. Increasing search intensity increases the rate at which the trader meets a counterparty and enjoys the expected gains from trade. The expected gains from trade depend on the trading surplus when the trader meets a given counterparty, and the likelihood of meeting each counterparty type. Traders will thus search harder if the expected gains from trade are higher, or if their search costs are lower. When we estimate the model in Section 5, this negative relationship between search costs and meeting rates will mean that the distribution of search costs $F(z)$ is identified by the distribution of average trading rates across traders.

4.2.3 Type Distributions

We now close the model by providing expressions for the equilibrium distributions of assets across trader types. For the system to be in steady state, the net flows from trading and value shocks must be equal to zero across all trader types (z, β, h). The measure of traders with search costs z, valuation $\beta \leq \beta^*$ and holding $h \leq h^*$ must satisfy:

$$\begin{aligned}
&\iint_{\underline{\beta}}^{\beta^*} \int_{h^*}^{\infty} m(\gamma(z, \beta, h), \gamma(\Delta'))\phi(z, \beta, h)\phi(\Delta')\mathbb{1}(q(z, \beta, h, \Delta') \geq h - h^*) dh d\beta d\Delta' - \\
&\quad \iint_{\underline{\beta}}^{\beta^*} \int_{-\infty}^{h^*} m(\gamma(z, \beta, h), \gamma(\Delta'))\phi(z, \beta, h)\phi(\Delta')\mathbb{1}(q(z, \beta, h, \Delta') < h - h^*) dh d\beta d\Delta' \\
&\qquad = \eta(1 - G(\beta^*)) \int_{\underline{\beta}}^{\beta^*} \int_{-\infty}^{h^*} \phi(z, h, \beta) dh d\beta - \eta G(\beta^*) \int_{\beta^*}^{\bar{\beta}} \int_{-\infty}^{h^*} \phi(z, h, \beta) dh d\beta
\end{aligned} \tag{6}$$

for all (z, β^*, h^*). The left-hand side represents net inflows due to trade. This consists of the flow of traders with type $(z, \beta \leq \beta^*, h > h^*)$ who meet a counterparty and sell enough of the asset such that their holdings fall below h^*, minus the flow of traders with type $(z, \beta \leq \beta^*, h \leq h^*)$ who meet a counterparty and buy enough that their holdings exceed h^*. The right-hand side consists of net outflows due to value shocks. The outflows consist of the stock of traders with type $(z, \beta \leq \beta^*, h \leq h^*)$, who with intensity η receive a new value shock, and with probability $1 - G(\beta^*)$ draw a new value above β^*. The inflows are traders with type $(z, \beta > \beta^*, h \leq h^*)$ who with intensity $\eta G(\beta^*)$ receive a shock that takes their value below β^*.

Finally, the distribution of types in equilibrium must be consistent with the ex ante distribution of types:

$$\iint \Phi(z, \beta, h) d\beta dh = F(z) \quad \forall z. \tag{7}$$

4.2.4 Equilibrium

We now define an equilibrium based on the system of equations derived above. An equilibrium is a set of value functions, search intensities, terms of trade and distributions of agents and assets that are mutually consistent in steady state. Let the space of trader types (z, β, h) be $\mathcal{T} = \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}$. A steady state equilibrium is:

- value function $V : \mathcal{T} \to \mathbb{R}$;
- pricing function $p : \mathcal{T}^2 \to \mathbb{R}^+$;
- trading quantity function $q : \mathcal{T}^2 \to \mathbb{R}$;
- density function $\phi : \mathcal{T} \to \mathbb{R}^+$; and
- search intensity functions $\gamma : \mathcal{T} \to \mathbb{R}^+$

that solve the following equilibrium conditions:

- value functions (equation (4));
- distribution functions (equations (6) and (7));
- optimal search (equation (5));
- optimal prices and quantity (equation (3)); and
- market clearing:

$$\iiint h \Phi(dz, d\beta, dh) = a. \tag{8}$$

An equilibrium is thus a set of equations that solves the three-way feedback between traders' search decisions, the distribution of the asset across traders, and the terms at which they trade. We solve for this equilibrium numerically in the sections that follow.

4.3 Properties of the Equilibrium

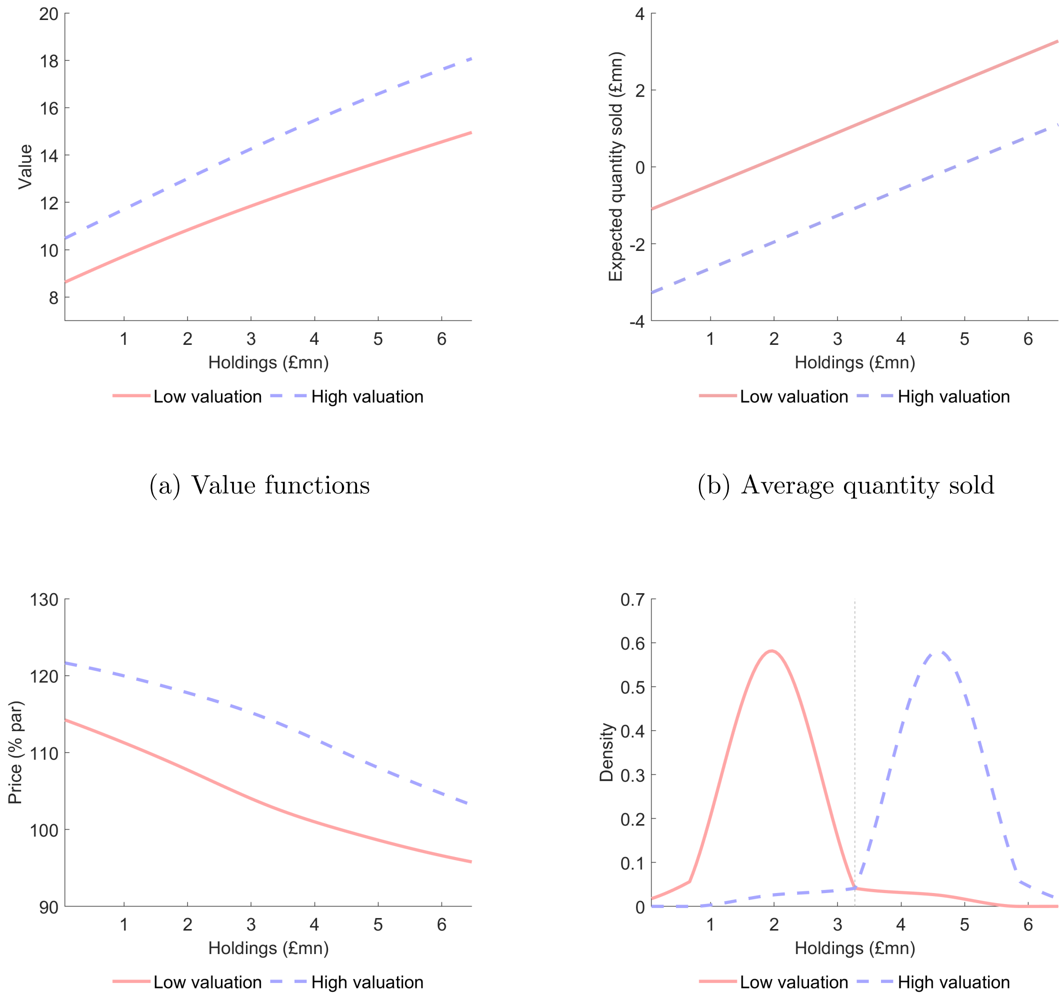
Figure 4 displays some properties of the equilibrium, showing the value functions, prices, trading quantity and distributions of traders with different valuations β as a function of holdings, for the parameter vector we estimate in Section 5. The value function is increasing and concave in holdings, inheriting these properties from the quadratic flow utility (Figure 4a).[15] A trader with a low valuation β values the asset less, and so at any level of holding is more likely to be a net seller than a trader with a high valuation (Figure 4b), and trades the asset at a lower price (Figure 4c). As a result of these differences in trading patterns, the low-value trader holds less of the asset (Figure 4d). With quadratic utility and a symmetric distribution of β, the equilibrium distributions are symmetric around the per-capita supply of the asset a, with the distributions of holdings for traders with low valuation the mirror image of those with high valuation.

4.3.1 Search and Trading

Figure 5a shows how traders vary their search and trading behaviour according to their valuation and asset holdings. For a given valuation β a trader's search intensity is a convex function of its asset holdings. Search intensity takes its minimum at a quantity which we call the trader's *target holding*, defined as the level of asset holding $h^*(z, \beta)$ at which the returns to search are minimised.

Trader search thus takes an intuitive form: traders search least when the gains from trade are low, and search harder as the gains from trade increase. Traders' values are concave in their holdings, which means that in equilibrium the gains to trade are higher when a trader has very low or high holdings. Search intensity is thus U-shaped in holdings. This mechanism in our model can rationalise the empirical fact in Table 3, which shows that traders offset their trades more frequently than if their search was exogenous. When search is chosen optimally, traders increase search intensity whenever a trade takes them away from their

[15]This concavity is necessary for the model to have a solution, as it ensures traders will always trade and hold a finite amount of the asset.



(a) Value functions

(b) Average quantity sold

(c) Average trading price

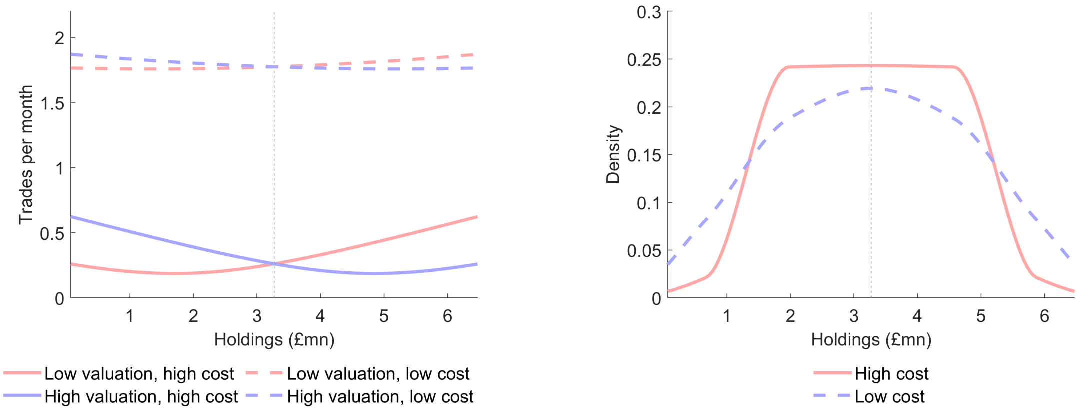
(d) Distributions

Figure 4: Terms of trade and distributions by trader type

Note: These figures summarise key features of the model's equilibrium, for the parameter values estimated in Section 5. Each panel plots one of the model's outcome variables against a trader's holdings for different levels of valuation β, for a given search parameter z. Panel (a) shows the value function $V(z, \beta, h)$. Panel (b) shows the average quantity sold per trade. Panel (c) shows the average trading price. Panel (d) shows the distribution of holdings. The grey vertical line in panel (d) shows the per-capita supply of the asset a.

target holdings, or a shock changes this target.

Traders with low search costs search harder, and trade more frequently as a result (Figure 5a). Their search intensity is less sensitive to their valuation β. Put differently, the difference in search intensity between low- and high-cost searchers is greatest when traders are at their target holdings, and have no reason to demand liquidity. Thus low-cost traders do not



(a) Trading frequency by shock & holding

(b) Holding distribution by type

Figure 5: Search and distributions by trader type

Note: These figures summarise some key quantities in our estimated model. Panel (a) shows trading frequency by asset holding h and valuation β, for a trader with a given search cost. Panel (b) shows the distribution of agents across different holdings for low-search-cost traders and high-search-cost traders. The grey vertical lines show per-capita supply of the asset.

leverage their cost advantage to respond faster to shocks, but instead to supply liquidity.

Figure 5b shows how a trader's search cost influences its asset holdings. Traders who are able to search at lower cost allow their holdings distribution to spread out more.[16] They are willing to trade to more extreme holdings because their superior search technology means they will be able to offset this trade relatively quickly.

4.3.2 Liquidity

Liquidity in this context is the ease with which traders can change their holdings when they wish to. Traders who wish to change their asset holdings demand liquidity, and those who facilitate this supply liquidity. Traders demanding liquidity face the following types of friction: (a) the time they must wait to trade, (b) the cost they pay to search, (c) the extent to which the quantity they trade is rationed relative to what they would trade in a market without frictions, and (d) the difference in the price at which they trade relative to that in a market without frictions. Aggregate welfare is determined by how the assets are distributed across traders with different values and the search costs traders incur, with trading profits

[16] Üslü (2019) finds a similar result in a setting with exogenous trading frequency.

netting out across traders.

In what follows, we will track a summary measure of liquidity which we call *market depth*. We define this to be the amount a trader could hypothetically sell per unit time without moving the price by more than a given amount D. This measure is in the spirit of definitions of market depth used in the literature on market microstructure (Foucault, Pagano and Röell, 2013). To construct this measure, we first note that the maximum price a trader of type $\Delta \equiv (z, \beta, h)$ would be willing to pay to buy amount q of a bond is:

$$\frac{V(z,\beta,h+q) - V(z,\beta,h)}{q}.$$

As q goes to 0, this becomes the marginal value of holdings in h, denoted $V_3(z, \beta, h)$.

We can then define the maximum amount $q^\pi(\Delta, D)$ that could be sold to this trader at a price no more than D below $V_3(z, \beta, h)$:

$$\frac{V(z,\beta,h+q^\pi(\Delta,D)) - V(z,\beta,h)}{q^\pi(\Delta,D)} = V_3(z,\beta,h) - D,$$

where $V_3(z, \beta, h)$ is the derivative of the trader's value with respect to their holdings.

Intuitively, $V_3(z, \beta, h)$ is the 'current' price at which the trader would be willing to trade a trivial amount ϵ before the trade, and the term on the left-hand side is the price at which they are willing to trade amount $q^\pi(\Delta, D)$. The difference between the two represents a cost for the trader selling the asset, as it captures the decline in the price they receive for the bond that is caused by their own sale. Our measure $\Pi(D)$ is then the trade-weighted average of this across traders multiplied by 2Γ, which is the average frequency with which a trader gets to trade:

$$\Pi(D) = 2\Gamma \int \frac{\gamma(\Delta)\phi(\Delta)}{\Gamma} q^\pi(\Delta, D) d\Delta. \tag{9}$$

This measure captures both the extensive margin of liquidity—how often traders get the opportunity to trade—and the intensive margin—how much they are able to trade upon meeting a counterparty without incurring large costs. A larger value of $\Pi(D)$ indicates a higher level of market depth, and hence a more liquid market.

5 Estimation

We first set out the parametric assumptions we make to take the model to the data. We then describe our estimation procedure, before summarising the moments we use. Finally we describe the key variation that identifies each of our model's parameters.

5.1 Parametric Assumptions

We use the following linear matching function:

$$m(\gamma, \gamma') = 2\gamma \frac{\gamma'}{\Gamma}, \tag{10}$$

where

$$\Gamma = \iiint \gamma(z, \beta, h)\Phi(dz, d\beta, dh).$$

This technology captures the fact that a trader can both contact and be contacted by another trader. Conditional on contact the likelihood of meeting a given type of counterparty is proportional to their search intensity.

Another common linear matching function is $m(\gamma, \gamma') = \gamma + \gamma'$ (Shimer and Smith, 2001), which assumes that counterparties are drawn uniformly from the distribution of traders. This matching function could not fit our data. In particular (1) Trading frequency is heavily right-skewed as shown in Figure 1; and (2) Many traders have a monthly trading frequency per instrument that is close to 0. As a result, this matching function would imply negative search intensities for the left-tail of the trading frequency distribution.

We take the search cost function to be:

$$s(z, \gamma) = (\gamma - z)^2, \tag{11}$$

where $z \sim F(z)$. A trader can costlessly meet traders at rate z, but must pay a constant marginal cost to meet traders more frequently. Intuitively, z can be thought of as a flow of contacts a trader makes as part of its everyday business. For example, if the trader is an investment bank it meets its clients regularly as part of its broader investment banking activity. However, to increase its meeting rate above this minimum level is costly, as it must contact other traders or hire more staff to make the contacts. For simplicity, we take the marginal cost of these contacts to be linear. In what follows, we will refer to z as search efficiency, and refer to traders with high z as having high search efficiency or low search costs.

We assume the search parameter z follows a Gamma distribution, with shape parameter k_z and scale parameter θ_z. We choose this distribution as it can match the skewness of the trading rate distribution (Figure 1), has strictly positive support, and is relatively parsimonious. We assume the shock distribution $G(\beta)$ is uniform, with mean μ_β and variance σ_β^2. We choose a uniform distribution as it is simple and parsimonious.

5.2 Estimation Procedure

We estimate the model using the data described in section 2.1, assuming these data are generated by the model in steady state. The unit of time is 1 month and the monthly discount rate is 0.5%. We fix the asset supply a to be equal to the mean total amount outstanding in a bond, normalised by the number of traders trading in that bond.

We estimate the parameter vector $\psi = \{\mu_\beta, \sigma_\beta, \kappa, \eta, k_z, \theta_z\}$, where $\{\mu_\beta, \sigma_\beta\}$ are the parameters of the β distribution, κ is risk aversion, η is shock frequency and $\{k_z, \theta_z\}$ are parameters of the distribution of the search parameter z. We use a minimum-distance estimator that matches theoretical moments implied by the model to their empirical counterparts. In practice this takes the form of a nested loop: for any given ψ we numerically solve the model. We then calculate a vector of theoretical moments $m(\psi)$ and compare this to its empirical counterpart m_0. We choose the parameter vector that minimises the sum of squared percentage differences between the theoretical and empirical moments:

$$\hat{\psi} = \arg\min_{\psi} (m(\psi) - m_0)'\Omega^{-1}(m(\psi) - m_0)$$

where $\Omega = m_0 m_0'$.

To find the model's equilibrium we solve the model's equations at a finite set of points in the type space $\mathcal{T}$. The distributions of valuations β and search cost parameters z are discrete versions of the continuous distributions in section 5.1. We use interpolation splines to ensure continuity of functions in holdings h. We then use a nonlinear solver to numerically solve the equilibrium equations. For full details of the estimation procedure, see Appendix A2.

The model is defined at the *instrument level*. When computing the empirical moment vector we compute moments separately for each bond, before averaging across bonds to give m_0.

5.3 Moments

Our data contain the key information we need to identify the model: trading frequencies, trading prices and quantities, and asset holdings. The data contain the identity of all traders, and not just dealers as is commonly the case, which means we can characterise trader heterogeneity across all traders.

We match two sets of moments: *across-trader* moments and *within-trader* moments. Within-trader moments measure the distribution of quantities for a given trader in a given bond—for example how much its inventory varies through time. Across-trader moments measure the distribution of average quantities across traders for a given bond—for example comparing how the average trading frequency of one trader differs from the average trading frequency of other traders.[17] We choose moments covering the means, variances and correlations of traders' trading frequencies, trading prices, trading quantities and holdings to identify our parameters. We give analytical expressions for each of these moments below.

5.3.1 Theoretical Moments

Expectations

1. Average trading frequency:

$$\mathbb{E}(n) = 2 \int \gamma(\Delta)\phi(\Delta)\pi(\Delta)d\Delta,$$

   where $\pi(\Delta)$ is the fraction of Δ's meetings that result in a trade:

$$\pi(\Delta) = \int \mathbb{1}(q(\Delta, \Delta') \neq 0)\frac{\gamma(\Delta')\phi(\Delta')}{\Gamma}d\Delta'.$$

2. Average trade size:

[17] Note that the means of a quantity are the same within and across traders.

$$\mathbb{E}(|q|) = \int \frac{\gamma(\Delta)\phi(\Delta)}{\Gamma} \int |q(\Delta, \Delta')| \frac{\gamma(\Delta')\phi(\Delta')}{\Gamma} d\Delta' d\Delta.$$

3. Average price:

$$\mathbb{E}(p) = \int \frac{\gamma(\Delta)\phi(\Delta)}{\Gamma} \int p(\Delta, \Delta') \frac{\mathbb{1}(q(\Delta, \Delta') \neq 0)}{\pi(\Delta)} \frac{\gamma(\Delta')\phi(\Delta')}{\Gamma} d\Delta' d\Delta.$$

Across-trader moments

4. Standard deviation of trading frequency across traders:

$$SD^A(n) = \sqrt{\int (n(z) - \mathbb{E}(n))^2 f(z) dz},$$

where $n(z)$ is type z's average trading frequency:

$$n(z) = 2 \iint \gamma(z, \beta, h) \frac{\phi(z, \beta, h)}{f(z)} \pi(z, \beta, h) d\beta dh.$$

Within-trader moments

5. Standard deviation of holdings, within traders:

$$SD^W(h) = \int \sqrt{\mathbb{V}(h|z)} f(z) dz,$$

where $\mathbb{V}(h|z)$ is the variance in holdings for a trader with search parameter z:[18]

$$\mathbb{V}(h|z) = \iint (h - a)^2 \frac{\phi(z, \beta, h)}{f(z)} d\beta dh.$$

6. Standard deviation of trading prices, within traders:

$$SD^W(p) = \int \sqrt{\mathbb{V}(p|z)} f(z) dz,$$

where $\mathbb{V}(p|z)$ is the variance of trading prices for a trader with search parameter z:

[18] Note that the symmetry of the equilibrium shown in Figure 4d means that each trader on average holds the asset's per-capita supply, a.

$$\mathbb{V}(p|z) = \iint \frac{\gamma(z,\beta,h)\phi(z,\beta,h)}{\Gamma(z)f(z)} \times$$
$$\int \left(p(z,\beta,h,\Delta') - \mathbb{E}(p|z)\right)^2 \frac{\mathbb{1}(q(z,\beta,h,\Delta') \neq 0)}{\pi(z,\beta,h)} \frac{\gamma(\Delta')\phi(\Delta')}{\Gamma} d\Delta' d\beta dh;$$

$\mathbb{E}(p|z)$ is type z's mean trading price:

$$\mathbb{E}(p|z) = \iint \frac{\gamma(z,\beta,h)\phi(z,\beta,h)}{\Gamma(z)f(z)} \times$$
$$\int p(z,\beta,h,\Delta') \frac{\mathbb{1}(q(z,\beta,h,\Delta') \neq 0)}{\pi(z,\beta,h)} \frac{\gamma(\Delta')\phi(\Delta')}{\Gamma} d\Delta' d\beta dh;$$

and $\Gamma(z)$ is type z's average search intensity:

$$\Gamma(z) = \iint \gamma(z,\beta,h)\phi(z,\beta,h) d\beta dh.$$

7. Correlation between quantity sold and holdings, within-traders:

$$corr^W(h,q) = \int \frac{cov(h,q|z)}{\sqrt{\mathbb{V}_{TW}(h|z)\mathbb{V}(q|z)}} f(z)dz,$$

where expressions for the conditional (trade-weighted) covariance of holdings and quantity sold $cov(h,q|z)$, trade-weighted holdings variance $\mathbb{V}_{TW}(h|z)$ and variance of quantity sold $\mathbb{V}(q|z)$ are given in Appendix A3.

8. Correlation between absolute inventory $inv \equiv |h-a|$ and trading frequency, within traders:

$$corr^W(inv,n) = \int \frac{cov(inv,n|z)}{\sqrt{\mathbb{V}(inv|z)\mathbb{V}(n|z)}} f(z)dz,$$

where expressions for the conditional covariance of absolute inventory and trading frequency $cov(inv,n|z)$, variance of absolute inventory $\mathbb{V}(inv|z)$ and variance of trading frequency $\mathbb{V}(n|z)$ are given in Appendix A3.

5.3.2 Empirical Moments

We compute the empirical counterpart of each of these moments to construct the empirical moment vector m_0. As the model is defined at the instrument level, we compute each of the moments separately for each bond, and the empirical moment vector m_0 contains the averages of these moments across bonds. The size of our sample is then the number of bonds in our data. To compute bootstrap standard errors, we resample from the set of bonds and repeat the estimation for each bootstrap sample.[19]

The data are likely to include the effects of economic forces outside of our model. Firstly, there may be heterogeneity in preferences for a given bond across traders. For example, a trader who was an underwriter on a bond's initial issuance may hold more of it throughout its life. To control for this, we compute our within-trader moments based only on the residual variation after stripping out trader-level variation. This is equivalent to including trader-level fixed effects. In the example of a trader holding more of a bond it underwrote than other traders, our approach strips out this variation when computing the variance of trader holdings. Secondly, news about the fundamentals of a bond will affect its price. To control for this, when computing the within-trader variance in prices we further control for the current credit rating of the instrument, and compute the variation in prices based on the residual variance after controlling for this. This strips out any news about a bond that is captured in its credit rating.

5.4 Identification

The moments defined above provide a mapping from the model's parameters to the data. In this section we show that there is sufficient variation to invert this mapping and infer the parameters from the data. The model's nonlinearity means that all but one parameters affect all moments, but for each of the model's parameters there are key moments that pin it down. Table 4 summarises this key variation for each parameter.

The distribution of the search parameter z is identified by the distribution of average trading frequencies across traders, as the optimal search equation (5) ensures that a trader's average trading frequency is a monotonic function of its search parameter z. As a result the parameters of the distribution of z are pinned down by the moments of the distribution of traders' average trading frequencies.

[19] Standard errors will be added to the results in a later version of this paper.

Table 4: Identification

Parameter	Moment
Search efficiency $z \sim \Gamma(k_z, \theta_z)$	
$\mathbb{E}(z) = k_z\theta_z$	$\overset{[+]}{\mathbb{E}(n)}$
$\mathbb{V}(z) = k_z\theta_z^2$	$\overset{[+]}{SD^A(n)}$
Utility $u(h) = \beta h - 0.5\kappa h^2$; $\beta \sim U(\mu_\beta, \sigma_\beta)$	
μ_β	$\overset{[+]}{\mathbb{E}(p)}$
σ_β	$\overset{[+]}{SD^W(p)}$, $\overset{[+]}{\mathbb{E}(\lvert q\rvert)}$, $\overset{[+]}{SD^W(h)}$
κ	$\overset{[+]}{SD^W(p)}$, $\overset{[-]}{\mathbb{E}(\lvert q\rvert)}$, $\overset{[-]}{SD^W(h)}$
Shock frequency η	
η	$\overset{[+]}{corr^W(\lvert h-a\rvert, n)}$, $\overset{[+]}{corr^W(h, q)}$

Note: This table summarises the key variation that identifies our parameters. n is the trading frequency, p is the trading price, q is the amount sold by a trader and h is a trader's holding. $SD^W(x)$ is the standard deviation of x for a given trader, whilst $SD^A(x)$ takes the average value of x for each trader and computes the standard deviation of this across traders. $corr^W(x, y)$ is the correlation of x and y for a given trader. Shocks β are distributed according to a uniform distribution with mean μ_β and variance σ_β^2. Parameter κ governs traders' risk aversion. Parameter η is the frequency (per month) at which traders draw new values β. Parameter z in search cost function $z = (\gamma - z)^2$ is distributed according to a Gamma distribution with shape parameter k_z and scale parameter θ_z. A [+] above a moment indicates the moment is increasing in the relevant parameter, whilst a [−] indicates it is decreasing.

The mean of the value shock distribution μ_β is identified by the average trading price. This is because it affects the mean price, but no other moments in the model. As set out in Section 4, when two traders meet, the quantity q they trade is such that their post-trade marginal values for the asset are equalised: $V_3(z, \beta, h - q) = V_3(z', \beta', h' + q)$. Moving the location of the valuation distribution $G(\beta)$ has no effect on this, as it affects all traders equally. As a result, the location of the β distribution does not affect trading surplus, and so does not affect search intensity or the asset distributions. The average trading price, which depends on the average slope of traders' value functions, does depend on the location of β, and as a result pins it down exactly.[20]

The variance of the value shock σ_β^2 and the level of risk aversion κ are identified by the distributions of holdings, trade sizes and price. A decrease in σ_β^2 causes the value functions

[20] In practical terms this makes estimation a simpler task, as we can estimate the mean of the shock distribution separately from the other parameters. In a first stage we search over $(\sigma_\beta, \kappa, \eta, k_z, \theta_z)$ to match all moments except the mean price. In a second stage we choose μ_β to perfectly fit the mean price.

across different levels of β to move closer together, whilst an increase in risk aversion increases the curvature of the value function in holdings for a given β. Both these shifts cause traders to shrink the variance of their asset holdings. Intuitively, both a smaller shock variance and increased risk aversion reduce a trader's incentive to trade to extreme asset holdings. For the same reason, they also both decrease the average trade size. As explained in Section 4 trade size is smaller when traders' marginal valuations are less spread out (as is the case when σ_β^2 is small) and when the curvature of the value function is greater (as is the case when κ increases).

An increase in risk aversion κ and a decrease in the variance of shocks σ_β^2 have opposite predictions, however, for the variance of prices. As set out in Section 4, the price at which two traders trade is governed by the slopes of their value functions. A reduction in σ_β^2 reduces the variance of these slopes across different valuations β and hence reduces the variance of prices. An increase in κ, however, increases the curvature of utility for a given β, which increases the variation in the slope of the value function across holdings. This increases the variance of prices, and enables us to separate the variance of shocks from traders' risk aversion.

The shock frequency, η, is identified by the correlation between traders' holdings of the asset and their trading price and quantity. An increase in the frequency of shocks moves the value functions of traders with different levels of β closer together. As a trader's value becomes more transient, they place less weight on their current value in their trading decisions and more on their expected future value, which is simply the mean of the distribution of β. In the context of Figure 4b and Figure 5a, this causes the average trading quantities and trading frequencies of traders with different β to move closer together. This strengthens the positive correlation between their holdings and the amount they sell, and their inventories—defined as the absolute difference between their holdings and the per-capita supply of the asset—and the amount they trade. These moments thus pin down the frequency of shocks.

6 Results

Table 5 shows the estimated parameters. Below we discuss the interpretation of each of the parameters in turn.

Figure 6 shows the estimated distribution of the search parameter z, and how the estimated trading rates fit those observed in the data. The model fits the empirical distribution

Table 5: Parameter Estimates

Parameter	Estimate
Search efficiency $z \sim \Gamma(k_z, \theta_z)$	
k_z	0.545
θ_z	0.374
Shock frequency η	
η	0.040
Utility $u(h) = \beta h - 0.5\kappa h^2$; $\beta \sim U(\mu_\beta, \sigma_\beta)$	
μ_β	0.031
σ_β	0.015
κ	0.008

Note: Parameter z in search cost function $z = (\gamma - z)^2$ is distributed according to a Gamma distribution with shape parameter k_z and scale parameter θ_z. Parameter η is the frequency (per month) at which traders draw new values β. Shocks β are distributed according to a uniform distribution with mean μ_β and variance σ_β^2. Parameter κ governs traders' risk aversion.

of trading quite well, replicating the heterogeneous and skewed distribution of trading rates we observe in the data (Figure 1). The model rationalises this with a heterogeneous distribution of search efficiency (the inverse of search costs). Intuitively, this means some traders have a meeting technology that enables them to meet counterparties frequently at low cost, whereas others do not. Institutionally, these active traders could be large investment banks that have large client networks who they contact regularly. This large client network could come from some more efficient technology—these traders are simply more effective at being middlemen—or could be due to the fact that they often have other business lines which help them create contacts. The less active traders do not enjoy the same technologies or client contacts, and are thus less able to trade frequently.

Shocks are infrequent—a trader faces a shock on average 0.5 times a year. These infrequent shocks are consistent with the infrequent trading we observe in corporate bond markets. Traders trade more frequently—0.4 times a month (Figure 6)—than they are shocked. In a frictionless market these two frequencies—the trading frequency and the shock frequency—would be the same. This difference can be explained by two factors: (1) When it meets a counterparty, a trader is unable to trade the full amount it would like, and will thus trade again; and (2) Traders do not just trade to demand liquidity when shocked, but also to supply liquidity to other traders.

The parameters of the utility function $u(\beta, h) = \beta h - 0.5\kappa h^2$ are reduced-form parameters,

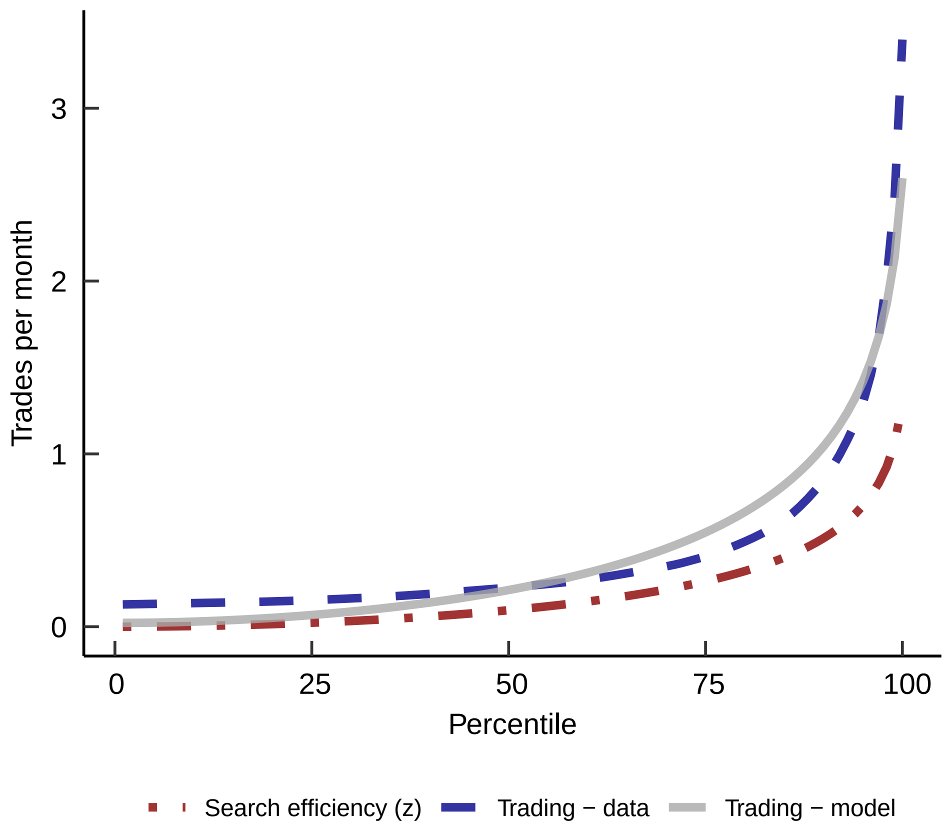


Figure 6: Distribution of trading frequencies and search parameter z

Note: This chart shows (a) the estimated distribution of the parameter z in traders' search cost function $s(z, \gamma) = (\gamma - z)^2$, where z follows a Gamma distribution (red dashed line), (b) the distribution of traders' average trading rates in the estimated model (grey line); and (c) the empirical distribution of traders' average trading rates (blue dashed line). To compute the empirical distribution, we first compute the distribution of traders' average trading rates per bond, and then take the weighted average of this across bonds, weighted by the number of traders in each bond.

and thus their magnitudes are difficult to interpret in terms of fundamental preferences.[21] Here we limit ourselves to a discussion of the relative magnitudes of the utility parameters β and κ, and note that below we will show the model fits the data very well. Both the variance of β and the level of κ govern variation in the slope of the utility function: κ creates variation in marginal utility across holdings for a given β, whilst a shock to β changes the marginal utility for a given holding. The variation caused by β is quantitatively larger than that caused by κ. Concretely, a one standard deviation shock to β causes a change in marginal

[21] As explained in Section 4 $u(\beta, h)$ is a reduced-form version of a more fundamental model with CARA preferences over consumption. However, given we cannot identify the parameters of this more fundamental setup, it is hard to map our parameter estimates into this setting in a quantitative sense.

Table 6: Model fit

Moment	Data	Model
Mean price, %	1.09	1.09
Std. dev. price within traders, %	0.04	0.04
Mean holdings, £mn	3.27	3.27
Std. dev. holdings within traders, £mn	1.23	1.31
Mean trade size, £mn	0.67	0.55
Mean trading frequency, per month	0.44	0.43
Std. dev. trading frequency across traders, per month	0.55	0.55
Correlation inventory & trading frequency	0.08	0.09
Correlation holdings & quantity sold	0.33	0.30

Note: This table shows empirical moments and simulated moments calculated using the parameter estimates in Table 5.

utility that is 50% greater than that caused by a one standard deviation change in holdings, which operates via κ. The concavity of a trader's utility function, which is governed by κ, is thus small relative to the distance between the utility functions of traders with different valuations β. This limited concavity of traders' flow utility is what lies behind the limited concavity of the value function displayed in Figure 4a. As a result, transferring the asset between traders with the same valuation β has a limited effect on their welfare.

Table 6 summarises the model fit. The model fits price, holdings, trading quantity and trading frequency moments well. Two moments merit further discussion. In the data, variation in price is more limited than variation in asset holdings: the standard deviation of the prices a trader pays for a bond is around 4% of the mean, whilst the standard deviation of their holdings of the bond is 40% of the mean. Fitting these two moments is what drives the risk aversion parameter κ to be quite low in our results, as low risk aversion reduces the variability of prices but increases the variability of holdings.

7 Counterfactuals

In this section we run counterfactual simulations to study the determinants of liquidity. We begin by establishing which traders supply liquidity. We show that a small set of traders with low search costs supplies the bulk of liquidity in these markets, and that as a result market outcomes are highly sensitive to the actions of these traders. We then study two counterfactuals that affect these traders' incentives to supply liquidity. The first – bank capital regulation – changes the cost to these liquidity suppliers of holding the asset. The second – the introduction of trading platforms – changes the cost advantage these liquidity suppliers enjoy over the rest of the market.

7.1 Resilience of Market Liquidity

In this section we study who supplies liquidity in these markets, and what this means for the resilience of liquidity to shocks. To do this, we 'withdraw' sets of traders from markets, which entails them selling the asset and stopping searching for counterparties.[22] We then solve for the new equilibrium in the market with the remaining traders, and show how liquidity has changed. This offers a convenient way to quantify each trader's contribution to market liquidity, and also to study the impact of shocks to traders that lead them to sell their assets and stop trading.

Figure 7 shows the relative contributions to market liquidity of the lowest- and highest-search cost traders in our model. The blue dashed line shows the fall in market depth when we withdraw the low-cost traders. Liquidity decreases sharply as we withdraw the traders with the lowest costs, with the pace of the decline in liquidity decreasing as we remove more and more traders. The red line shows the fall in market depth when we withdraw the high-cost traders. Liquidity barely changes as we withdraw the least efficient traders, but progressively deteriorates as we withdraw more traders.

The top 8% most frequent traders supply as much liquidity as all other traders combined (Figure 7). This asymmetry in the provision of liquidity is driven by the data. Our model maps traders' average trading frequencies into search efficiency parameters z. The stark heterogeneity observed in traders' average trading frequencies (Figure 1) thus drives the highly skewed distribution of z we obtain in Figure 7. When we counterfactually remove low- and high-cost traders from the market, this skewed distribution in z results in very different effects on market liquidity.

This result has direct implications for the resilience of market liquidity to shocks to traders. As a result of their disproportionate contribution to market liquidity, the market is vulnerable to shocks to the lowest-cost traders that lead them to withdraw from the market. We demonstrate in Table 7 by showing how welfare and liquidity would change if the 4% of traders with the lowest search costs withdrew from the market. The impacts of this small

[22]This is equivalent to assuming they still meet counterparties, but refuse to trade with them.

Figure 7: Trader contributions to market liquidity

Note: In this figure we show that low-search-cost traders are disproportionately important for market liquidity. We summarise each trader's contribution to market liquidity by withdrawing sets of traders and computing the effect on liquidity. The x-axis shows the percentage of traders we withdraw. The y-axis shows the resulting fall in market depth relative to its level in the estimated equilibrium, where depth $\Pi(D)$ is defined in equation (9), and D is set to 1%. The red line shows the fall in liquidity when we sequentially remove traders starting from those with the highest search costs. The blue dashed line shows the fall in liquidity when we start from those with the lowest search costs. The grey dashed lines show the points where removing the 8% of traders with the lowest costs has the same effect on liquidity as removing the 92% with the highest costs.

change in the trader population are extreme. The remaining traders, with their inferior search technologies, are able to meet each other less frequently. Market depth falls by 62%, whilst price volatility increases by 140%. The remaining traders are forced to absorb the exiting traders' sales, which, due to the concavity of their value functions, lowers the price. In response to the poorer liquidity conditions traders shrink the variance of their holdings. In total the welfare of traders falls by 5%.

This stark dependence of liquidity on a few key traders can help explain patterns in trading and liquidity during the Covid-19 crisis in March 2020. In the early parts of March, dealers were unwilling to expand their inventories (Kargar, Lester, Lindsay, Liu, Weill and

Table 7: Effects of 4% of traders withdrawing

	% change in counterfactual
Market depth	-62
Holdings variance	-9
Mean trading frequency	-17
Mean price	-23
Price variance	140
Aggregate utility	-5

Note: This table shows the impact of the 4% of traders with the lowest search costs selling their asset holdings and stopping trading. Holdings and price variance are within-trader variances: we take the variance for a given trader and average across traders. Market depth, which we define in equation (9) is a measure of how much traders are able to sell without reducing the trading price by more than 1%.

Zuniga, 2021). During this period, market conditions deteriorated rapidly, with prices falling and the cost of transacting increasing. From 18$^{\text{th}}$ March, the Federal Reserve took large scale policy action to boost liquidity in markets, at which point dealers began to allow their inventories to increase and markets stabilised (Kargar, Lester, Lindsay, Liu, Weill and Zuniga, 2021). This is consistent with the findings with Table 7, which show that bond markets are highly reliant on the appetite for trading of a small proportion of traders.

7.2 Regulating Dealers

Tighter capital regulation increases the fraction of a bank's balance sheet that must be funded by equity, rather than debt. If a bank applies this evenly across all its business, this increases the fraction of any position—short or long—that must be funded by equity.[23] There is relatively broad agreement that capital regulation enhances financial stability by increasing bank resilience, but there are concerns that it also harms liquidity in financial markets (Duffie, 2018). In this section we study this potential cost of capital regulation.

There are three mechanisms by which tighter capital regulation can reduce a bank's incentive to hold assets. The first is via violations of the Modigliani-Miller theorem (Modigliani and Miller, 1958). A bank's cost of equity is typically higher than its cost of debt. In the absence of any frictions, increasing a bank's equity funding will not increase its overall cost of capital, as its costs of equity and of debt will fall. In a world with frictions, its cost of

[23] Whilst capital regulation applies at the bank level, banks' internal capital markets are typically organised such that increases in capital requirements take effect across the bank's business lines (Bajaj, Binmore, Dasgupta and Vo, 2018).

capital will increase.[24] The second is via a violation of the debt overhang problem (Myers, 1977) discussed by Andersen, Duffie and Song (2019) and Duffie (2018), by which capital regulation reduces a bank's willingness to take a new, relatively safe asset onto its balance sheet. Finally, in the short run banks are unable to issue new equity. As a result, a binding capital requirement in the short run sets an upper limit on banks' asset holdings.

Given dealer-affiliated banks have typically acted as intermediaries in fixed income markets and are amongst the most frequent traders in the sterling corporate bond market, there are concerns that bank capital regulation could harm market functioning. In the context of our model, bank capital regulation affects the incentives of the low-search cost traders that supply liquidity to hold the asset. To study this, we simulate an increase in the cost of holding a long or short position in an asset for the traders with the lowest search costs. Their flow utility function becomes:

$$u(\beta, h) = \beta h - \kappa \frac{h^2}{2} - \tau |h|, \tag{12}$$

where $\tau \geq 0$ represents the increase in costs due to regulation.

Table 8 summarises how market quantities change when we set $\tau = 1\%$ for the 15% of traders with the lowest search costs. These figures are such that if trader behaviour did not adjust at all, dealers' utility would fall by 40%. Given that it is difficult to map a level of τ into a level of capital regulation, the figures that follow are best interpreted in relative terms, rather than as the absolute impact of a given level of capital regulation.

Tighter regulation of dealers leads them to significantly decrease their holdings of the asset, as they find holding inventory more costly. Market clearing implies unregulated traders increase their holdings. Tighter regulation, by reducing the marginal utility of holding the asset for dealers, reduces the average trading price.

Dealers adjust their trading frequency in two ways, as shown in Figure 8. Firstly, their target holding—where search is at its minimum—shifts to the left, reflecting their reduced desire to hold inventory. Secondly, the slope of the trading frequency curve becomes steeper

[24] Potential violations of the Modigliani-Miller assumptions include the tax advantage of debt (Kashyap, Stein and Hanson, 2010), asymmetric information leading to a 'pecking order' theory of financing (Myers and Majluf, 1984), agency problems of bank management (Diamond and Rajan, 2001) and the possibility that banks' short-term debt enjoys a 'money-like' convenience yield (Stein, 2012; Kashyap, Stein and Hanson, 2010). Note that in the case of the tax shield on debt, a full welfare analysis would need to take into account the increase in tax revenues resulting from higher capital requirements (Admati, DeMarzo, Hellwig and Pfleiderer, 2013).

Table 8: Capital counterfactual

	% change in counterfactual
Dealers	
Expected holdings	-33
Spread received	17
Expected utility	-37
Other traders	
Expected holdings	6
Spread paid	17
Expected utility	1
Aggregate	
Price	-19
Market depth	-10
Aggregate welfare	-5

Note: This table summarises the changes in key variables in the capital counteractual, relative to our baseline equilibrium. Capital regulation is modelled as setting $\tau = 1\%$ for the 15% of traders with the lowest search costs, where their flow utility from holding the asset is given by $u(\beta, h) = \beta h - 0.5\kappa h^2 - \tau|h|$. Market depth, which we define in equation (9) is a measure of how much traders are able to sell without reducing the trading price by more than 1%.

when they are away from their target, meaning regulation leads them to search harder when away from target inventory. This is the mechanism driving the empirical finding in Figure 3 that dealers' tendency to offset trades is related to capital regulation, and stems from the fact that regulation has increased the costs of being away from target holding.

Tighter regulation increases the spreads earned by dealers (Table 8). This reflects the balance of two forces. On the one hand, regulation means dealers are less inclined to trade away from target holding, as the costs of taking large positions are greater. All else equal, this would increase the spread required to persuade them to do so. On the other hand, conditional on being away from target holding, dealers are more anxious to trade back to target. This would decrease the spread they pay, as in these cases they are demanding, rather than supplying, liquidity. The former effect dominates the latter, meaning dealers' realised spreads increase.

Tighter capital regulation reduces market depth, our summary measure of liquidity, by 10%. This is driven by dealers' decreased willingness to take large positions in the asset now inventory costs have increased. Aggregate welfare in this market falls by 5%. If traders were unable to adjust their trading or holdings decision, the welfare cost would be 7%. In equilibrium, the costs of capital regulation are mitigated by the effects of endogenous

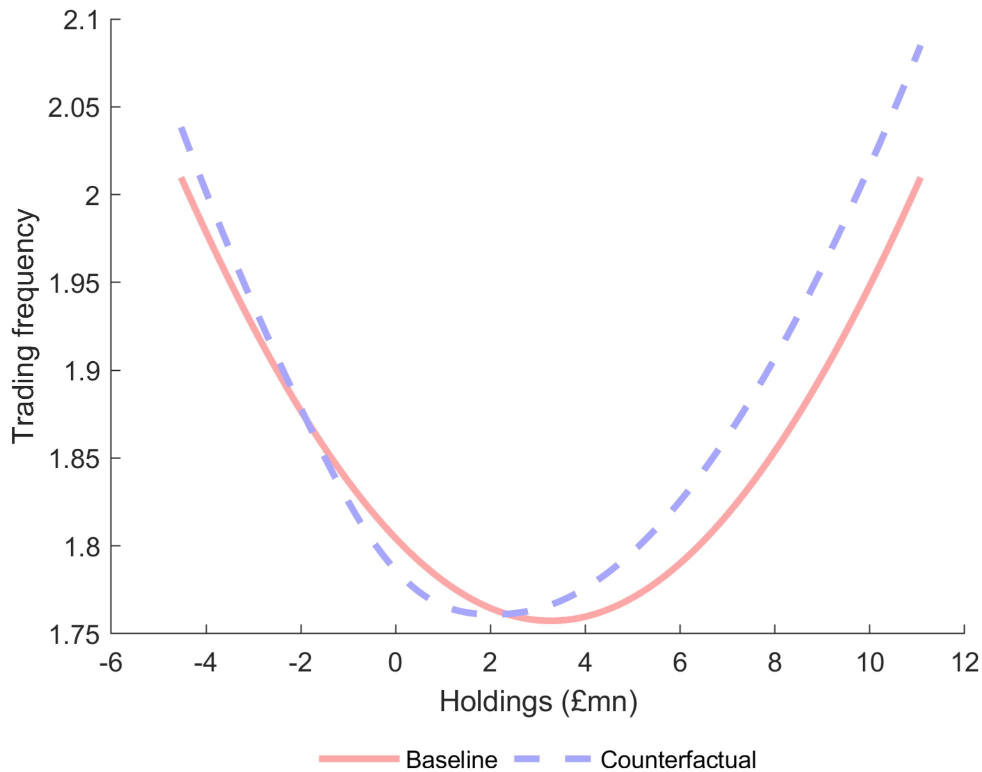


Figure 8: Dealer capital regulation and search

Note: This figure shows how capital regulation affects dealer search. We show that capital regulation shifts dealers' target holdings downwards and causes them to search harder when they are away from their target holdings. Capital regulation is modelled as setting $\tau = 1\%$ for the 15% of traders with the lowest search costs, where their flow utility from holding the asset is given by $u(\beta, h) = \beta h - 0.5\kappa h^2 - \tau|h|$.

search—with dealers optimally adjusting their search behaviour to minimise the harm caused by tighter regulation—and the endogenous reallocation of asset holdings from dealers to non-dealers.

The magnitudes of these effects are driven by the distributions of the search efficiency parameter z and the level of risk aversion κ. The fact that search efficiencies are highly skewed, and that capital regulation is applied to dealers, mean that capital regulation affects traders who play a key role in supplying liquidity. If the distribution of search efficiency were more homogeneous, the effect on liquidity would be more limited. Risk aversion κ governs the curvature of the value function. The greater is κ, the less willing non-dealers are to increase their holdings of the asset in response to capital regulation. The fact that κ is relatively low means that the transfer of holdings from dealers to non-dealers is relatively large, which limits the welfare costs of capital regulation.

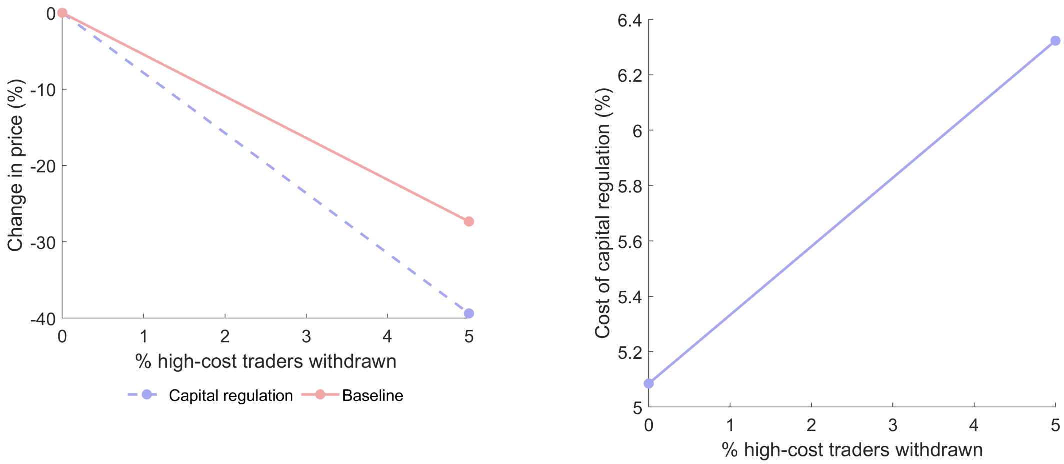
7.2.1 Liquidity, Regulation and Trading in a Sell-off

The negative effects of illiquidity can be particularly large during stress, amplifying the effects of financial shocks (Brunnermeier and Pedersen, 2009; Gromb and Vayanos, 2002). There is some evidence that liquidity during stress events—for example when a bond is downgraded or is removed from a market index—deteriorated after the 2008 financial crisis, as dealers were subject to more stringent regulation (Bao, O'Hara and Zhou, 2018; Dick-Nielsen and Rossi, 2019). In each of these cases, the stress event is a shock that causes traders who are not dealers—such as traders who track market indices—to sell the bond. An important advantage of a structural model is that we are able to counterfactually simulate how the effects of capital regulation vary during these types of sell-offs.

To do this, we subject dealers to capital regulation at the same time as assuming high-cost traders set their asset holdings to zero and withdraw from the market. Figure 9 shows the results of this. Figure 9a shows the change in prices as a function of high-cost trader withdrawals both in the baseline (red line) and when dealers are subject to capital regulation (blue line). With capital regulation, the fall in price is greater. Figure 9b shows the fall in aggregate welfare due to capital regulation as a function of high-cost trader withdrawals. The cost of capital regulation is greater in a stress. Intuitively, the costs of capital regulation are limited by the endogenous responses of customers, who increase their asset holdings by trading with dealers who wish to offload the bond. As these traders stop trading, these margins of adjustment are constrained, and the costs of capital regulation increase.

7.3 Trading Technologies

Trading in corporate bond markets is predominantly undertaken via traditional methods: dealers intermediate the market, and trades are organised bilaterally between traders, who often communicate on the phone. In recent years there has been a gradual increase in the use of electronic trading platforms (Anderson, Webber, Noss, Beale and Crowley-Reidy, 2015), which replace bilateral negotiation with a multilateral system where a trade quote is posted to all platform members, any of whom can bid to be counterparty to the trade. These types of platforms offer potential efficiency savings, particularly for those who do not have large client networks with which to trade. Such trading innovations require the participation of sufficient traders in order to succeed—there needs to be sufficient supply of assets and trading activity in order to make setting up such a platform worthwhile. The successful



(a) Prices in a stress

(b) Welfare in a stress

Figure 9: Impact of capital regulation in a stress

Note: In this figure we show that the negative effect of capital regulation on welfare and prices is worse in a stress. Stress—along the x-axis—is the percentage of traders that withdraw from the market, selling their asset and stopping trading. Traders are withdrawn in descending order of their search costs, starting from the highest-cost traders. The left-hand chart plots the change in price as a percentage of par against the level of stress, when dealers are subject to capital regulation (blue dashed line) and when they are not (red line). The right-hand chart shows the level of aggregate welfare when dealers are subject to capital regulation vs when they are not, as a function of the level of stress. Capital regulation is modeled as setting $\tau = 1\%$ for the 15% of traders with the lowest search costs, where their flow utility from holding the asset is given by $u(\beta, h) = \beta h - 0.5\kappa h^2 - \tau|h|$.

implementation of these innovations thus depends on the participation of the most efficient traders, as these are the largest and most frequent traders.

The take-up of these new technologies has been relatively slow (The Economist, 2020). In this section we offer an explanation for this: the most efficient traders in corporate bond markets lose out under these new technologies. To show this, we run two counterfactual simulations. First, we *reduce* and *homogenise* search costs in traders. Specifically, we take the highest value of search efficiency z in our (discretised) distribution of search costs, and counterfactually assign all traders this value of z. This captures the fact that platform-based trading has the potential to increase search efficiency, particularly for the most inefficient traders. Secondly, we study the Walrasian equilibrium, where trading is frictionless and prices are no longer bilaterally negotiated. This captures the increase in search efficiency on platforms and also the change in trading mechanism they introduce.

Table 9 summarises the effect of these changes in search technologies on trading and welfare. The penultimate column shows the effects of reducing and homogenising traders'

Table 9: Counterfactual search technologies

	Baseline		Homogenous	Walrasian
	Low cost	High cost	Aggregate	Aggregate
Trades per month	1.76	0.20	1.75	0.03
Spread, bps	223	-223	0	0
Utility	16.3	13.3	14.5	14.5
Aggregate				
Price variance, bps	378		90	0
Trade size, £mn	0.5		0.5	1.9
Gross volume, £mn	0.07		0.23	0.08
Utility	13.8		14.5	14.5

Note: This table shows the effect of platform-based trading on trading, spreads and welfare The first two columns show results in our baseline equilibrium. The next column shows results when all traders' search costs are set to those of the lowest-cost trader. The final column shows results in the Walrasian equilibrium. Gross volume is instantaneous trading volume per trader. The spread is the difference between a trader's average selling price and buying price.

search costs. Traders who initially have high search costs take advantage of the improvement in their meeting technologies by increasing their trading frequency. These traders benefit from the change in technology. Low-cost traders, however, lose out from it. The reason for this can be seen in the effect on spreads. Initially, low-cost traders on average earn trading revenues by charging a spread—they sell the asset at a high price and buy it at a low price. They are able to do this because of their technological advantage over higher-cost traders. When platforms erode this advantage, their gains from trade decrease. As a result, whilst aggregate welfare increases in the counterfactual, the welfare of the lowest-cost traders decreases. As shown in the final column of Table 9, the lowest-cost traders lose out even in the case where trading is frictionless.

This result is driven by the shape of the distribution of traders' search efficiencies z, which in turn is driven by the distribution of trading rates we observe in the data (Figure 1). The skewness of the estimated distribution of z means there are traders where the costs of frictionless trading exceed the benefits. All traders derive some benefit from frictionless trading, as they can respond instantly to shocks. The most efficient traders also incur a cost, as they no longer earn spreads. The heterogeneity of traders' search efficiencies means the spreads earned by the most efficient traders—and hence the losses they incur from frictionless trading—are large. Additionally, the fact that these efficient traders can meet counterparties much more frequently than the rate at which they are shocked means the benefits of frictionless markets—being able to respond to shocks instantly—are relatively small. As a result, these traders stand to lose out from platforms. This result is driven by the data in the sense that if we had estimated a distribution of z that was lower and more homogeneous, no traders would lose out from frictionless trading.

The most efficient traders thus have an incentive to resist shifts to trading platforms. This can explain why trading platforms have for a long time struggled to gain a foothold in corporate bond markets (Bessembinder, Spatt and Venkataraman, 2020).[25]

8 Conclusion

Liquidity is a key aspect of financial market functioning, but is hard to pin down. It is an equilibrium outcome that cannot be directly observed in the data. It varies with financial market conditions, and is impacted by regulatory and technological change. Liquidity, and how resilient it is in times of stress, is a key area of concern for academics (Duffie, 2018) and policymakers (Powell, 2015).

In this paper we present a quantitative model of liquidity in an OTC market. This enables us to document the reliance of market liquidity on a small set of traders, and to study the implications of this for the resilience of liquidity to trader stress, the impact of bank capital regulation, and the effects of new trading technologies.

[25]For example, in 2001 the online trading platform BondBook closed its operations on the basis that 'the behavioural change among market participants required for the platform to take off' was unlikely to materialise sufficiently quickly (see Finextra, "BondBook shuts down its trading operations', available at `https://www.finextra.com/newsarticle/3605/bondbook-shuts-down-trading-operations`). Similarly, in 2013 the major asset manager BlackRock shelved plans to develop its own in-house trading platform (see Grind, L. and T. Demos "BlackRock Shelves Platform For Bonds", available at `https://www.wsj.com/articles/SB10001424127887323551004578441053526969438`).

References

Admati, A. R., DeMarzo, P. M., Hellwig, M. F. and Pfleiderer, P. C. (2013), 'Fallacies, irrelevant facts, and myths in the discussion of capital regulation: Why bank equity is not socially expensive', *Max Planck Institute for Research on Collective Goods* 23.

Adrian, T., Fleming, M., Shachar, O. and Vogt, E. (2017), 'Market liquidity after the financial crisis', *Annual Review of Financial Economics* 9, 43–83.

Afonso, G. and Lagos, R. (2015), 'Trade dynamics in the market for federal funds', *Econometrica* 83(1), 263–313.

Allen, J. and Wittwer, M. (2021), Centralizing over-the-counter markets, Working Paper, Stanford University.

Andersen, L., Duffie, D. and Song, Y. (2019), 'Funding value adjustments', *The Journal of Finance* 74(1), 145–192.

Anderson, N., Webber, L., Noss, J., Beale, D. and Crowley-Reidy, L. (2015), 'The resilience of financial market liquidity', Bank of England Financial Stability Paper.

Babus, A. and Kondor, P. (2018), 'Trading and information diffusion in over-the-counter markets', *Econometrica* 86(5), 1727–1769.

Bajaj, R., Binmore, A., Dasgupta, R. and Vo, Q.-A. (2018), 'Banks' internal capital markets: how do banks allocate capital internally?', *Bank of England Quarterly Bulletin* p. Q2.

Bank of England (2016), 'Financial stability report: November 2016', Technical report, Bank of England.

Bank of England (2021), 'Financial Policy Committee's core indicators', https://www.bankofengland.co.uk/financial-stability. Accessed: 05/08/2021.

Bao, J., O'Hara, M. and Zhou, X. A. (2018), 'The volcker rule and corporate bond market making in times of stress', *Journal of Financial Economics* 130(1), 95–113.

BCBS (2006), 'International convergence of capital measurement and capital standards', Technical report, Bank for International Settlements.

BCBS (2017), 'Basel III: finalising post-crisis reforms', Technical report, Bank for International Settlements.

Bessembinder, H., Jacobsen, S., Maxwell, W. and Venkataraman, K. (2018), 'Capital commitment and illiquidity in corporate bonds', *The Journal of Finance* 73(4), 1615–1661.

Bessembinder, H., Spatt, C. and Venkataraman, K. (2020), 'A survey of the microstructure of fixed-income markets', *Journal of Financial and Quantitative Analysis* 55(1), 1–45.

Biais, B., Declerck, F. and Moinas, S. (2016), 'Who supplies liquidity, how and when?', Bank for International Settlements Working Paper 563.

BlackRock (2015), 'Addressing market liquidity', Technical report, BlackRock.

Brancaccio, G. and Kang, K. (2021), 'Intermediaries and product selection in the municpal bond market', Unpublished Working Paper.

Brancaccio, G., Li, D. and Schürhoff, N. (2020), 'Learning by trading: The case of the US market for municipal bonds', Working Paper, New York University Stern School of Business.

Brunnermeier, M. K. and Pedersen, L. H. (2009), 'Market liquidity and funding liquidity', *The review of financial studies* 22(6), 2201–2238.

Cai, Y. and Judd, K. L. (2013), 'Shape-preserving dynamic programming', *Mathematical Methods of Operations Research* 77(3), 407–421.

Chang, B. and Zhang, S. (2021), Endogenous market making and network formation, Working Paper, London School of Economics.

Choi, J. and Huh, Y. (2018), 'Customer liquidity provision: Implications for corporate bond transaction costs', Working paper, University of Illinois Urbana-Champaign.

Cimon, D. and Garriott, C. (2019), 'Banking regulation and market making', *Journal of Banking & Finance* 109, 105653.

Coen, P. and Coen, J. (2019), 'A structural model of interbank network formation and contagion', Working paper, London School of Economics.

Craig, B. R. and Ma, Y. (2021), 'Intermediation in the interbank lending market', Working Paper, Columbia Business School.

Czech, R., Huang, S., Lou, D. and Wang, T. (2021), 'Informed trading in government bond markets', *Journal of Financial Economics* .

Czech, R. and Roberts-Sklar, M. (2019), 'Investor behaviour and reaching for yield: Evidence from the sterling corporate bond market', *Financial Markets, Institutions & Instruments* 28(5), 347–379.

Di Maggio, M., Kermani, A. and Song, Z. (2017), 'The value of trading relations in turbulent times', *Journal of Financial Economics* 124(2), 266–284.

Diamond, D. W. and Rajan, R. G. (2001), 'Liquidity risk, liquidity creation, and financial fragility: A theory of banking', *Journal of Political Economy* 109(2), 287–327.

Dick-Nielsen, J. and Rossi, M. (2019), 'The cost of immediacy for corporate bonds', *The Review of Financial Studies* 32(1), 1–41.

Du, W., Tepper, A. and Verdelhan, A. (2018), 'Deviations from covered interest rate parity', *The Journal of Finance* 73(3), 915–957.

Duffie, D. (2010*a*), 'The failure mechanics of dealer banks', *Journal of Economic Perspectives* 24(1), 51–72.

Duffie, D. (2010*b*), 'Presidential address: Asset price dynamics with slow-moving capital', *The Journal of finance* 65(4), 1237–1267.

Duffie, D. (2018), 'Financial regulatory reform after the crisis: An assessment', *Management Science* 64(10), 4835–4857.

Duffie, D., Gârleanu, N. and Pedersen, L. H. (2005), 'Over-the-counter markets', *Econometrica* 73(6), 1815–1847.

Duffie, D., Gârleanu, N. and Pedersen, L. H. (2007), 'Valuation in over-the-counter markets', *The Review of Financial Studies* 20(6), 1865–1900.

Eisfeldt, A. L., Herskovic, B., Rajan, S. and Siriwardane, E. (2020), OTC intermediaries, Working paper, UCLA.

Farboodi, M. (2021), 'Intermediation and voluntary exposure to counterparty risk', Working Paper, Massachusetts Institute of Technology.

Farboodi, M., Jarosch, G. and Shimer, R. (2021), 'The emergence of market structure', Working Paper, Massachusetts Institute of Technology.

Foucault, T., Pagano, M. and Röell, A. (2013), *Market liquidity: theory, evidence, and policy*, Oxford University Press.

Franzoni, F. A., Plazzi, A. and Çötelioğlu, E. (2019), 'What constrains liquidity provision? Evidence from hedge fund trades', CEPR discussion paper 13645.

Gârleanu, N. (2009), 'Portfolio choice and pricing in illiquid markets', *Journal of Economic Theory* 144(2), 532–564.

Gavazza, A. (2016), 'An empirical equilibrium model of a decentralized asset market', *Econometrica* 84(5), 1755–1798.

Goodman, T. (2001), 'Shape preserving interpolation by curves', Working paper, Dundee University.

Gromb, D. and Vayanos, D. (2002), 'Equilibrium and welfare in markets with financially constrained arbitrageurs', *Journal of Financial Economics* 66(2-3), 361–407.

Gromb, D. and Vayanos, D. (2018), 'The dynamics of financially constrained arbitrage', *The Journal of Finance* 73(4), 1713–1750.

Grossman, S. J. and Miller, M. H. (1988), 'Liquidity and market structure', *The Journal of Finance* 43(3), 617–633.

Heller, D. and Vause, N. (2012), Collateral requirements for mandatory central clearing of over-the-counter derivatives, Working paper, Bank for International Settlements.

Hendershott, T. and Madhavan, A. (2015), 'Click or call? auction versus search in the over-the-counter market', *The Journal of Finance* 70(1), 419–447.

Hollifield, B., Neklyudov, A. and Spatt, C. (2017), 'Bid-ask spreads, trading networks, and the pricing of securitizations', *The Review of Financial Studies* 30(9), 3048–3085.

Hugonnier, J., Lester, B. and Weill, P.-O. (2020), 'Frictional intermediation in over-the-counter markets', *The Review of Economic Studies* 87(3), 1432–1469.

Kargar, M., Lester, B., Lindsay, D., Liu, S., Weill, P.-O. and Zuniga, D. (2021), 'Corporate bond liquidity during the Covid-19 crisis', *The Review of Financial Studies* 34(11), 5352–5401.

Kashyap, A. K., Stein, J. C. and Hanson, S. (2010), 'An analysis of the impact of 'substantially heightened' capital requirements on large financial institutions', Working paper, University of Chicago Booth School of Business.

Lagos, R. and Rocheteau, G. (2007), 'Search in asset markets: Market structure, liquidity, and welfare', *American Economic Review* 97(2), 198–202.

Lagos, R. and Rocheteau, G. (2009), 'Liquidity in asset markets with search frictions', *Econometrica* 77(2), 403–426.

Li, D. and Schürhoff, N. (2019), 'Dealer networks', *The Journal of Finance* 74(1), 91–144.

Li, J. (2021), 'The importance of investor heterogeneity: An examination of the corporate bond market', Working Paper, The University of Chicago.

Liu, S. (2020), 'Dealers' search intensity in US corporate bond markets', Working paper, Tsinghua University.

Mallaburn, D., Roberts-Sklar, M. and Silvestri, L. (2019), 'Resilience of trading networks: evidence from the sterling corporate bond market', Bank of England Working Paper 813.

Modigliani, F. and Miller, M. H. (1958), 'The cost of capital, corporation finance and the theory of investment', *The American Economic Review* 1, 3.

Myers, S. C. (1977), 'Determinants of corporate borrowing', *Journal of financial economics* 5(2), 147–175.

Myers, S. C. and Majluf, N. S. (1984), 'Corporate financing and investment decisions when firms have information that investors do not have', *Journal of financial economics* 13(2), 187–221.

Neklyudov, A. (2019), 'Bid-ask spreads and the over-the-counter interdealer markets: Core and peripheral dealers', *Review of Economic Dynamics* 33, 57–84.

Plante, S. (2018), 'Should corporate bond trading be centralized? Theory and evidence', Working Paper, Wisconsin School of Business.

Powell, J. H. (2015), 'Structure and liquidity in treasury markets', Speech at the Brookings Institution, Washington, DC, August 3, 2015, Technical report, Board of Governors of the Federal Reserve System (US).

Saar, G., Sun, J., Yang, R. and Zhu, H. (2020), 'From market making to matchmaking: Does bank regulation harm market liquidity?', Working Paper, Cornell University.

Sambalaibat, B. (2018), 'A theory of liquidity spillover between bond and CDS markets', Working paper, Princeton University.

Schultz, P. (2017), 'Inventory management by corporate bond dealers', Working paper, University of Notre Dame.

Shimer, R. and Smith, L. (2001), 'Matching, search, and heterogeneity', *The BE Journal of Macroeconomics* 1(1).

Stein, J. C. (2012), 'Monetary policy as financial stability regulation', *The Quarterly Journal of Economics* 127(1), 57–95.

The Economist (2020), 'Electronic trading platforms are challenging bond broker-dealers', 2020-03-07.

Trebbi, F. and Xiao, K. (2017), 'Regulation and market liquidity', *Management Science* 65(5).

Üslü, S. (2019), 'Pricing and liquidity in decentralized asset markets', *Econometrica* 87(6), 2079–2140.

Vayanos, D. and Wang, J. (2013), 'Market liquidity - theory and empirical evidence', *in* 'Handbook of the Economics of Finance', Vol. 2, Elsevier, pp. 1289–1361.

Vayanos, D. and Weill, P.-O. (2008), 'A search-based theory of the on-the-run phenomenon', *The Journal of Finance* 63(3), 1361–1398.

Appendices

A1 Further Details on Data and Empirics

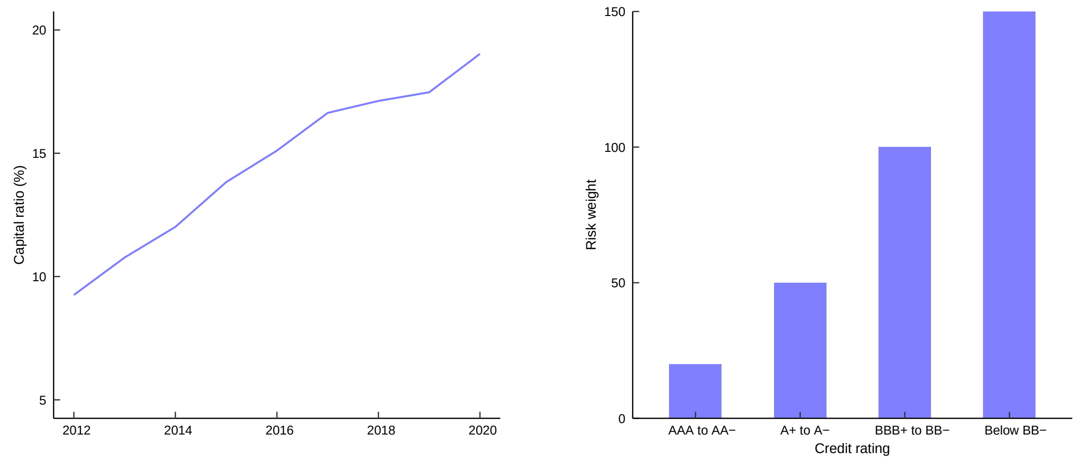
Traders are consolidated to the group level, with the exception of the asset management arms of large banking groups, who we separate from the banking business. We do this as these two business lines report separately in the data, and they tend to operate with separate balance sheets and trading strategies. We winsorize both trading prices and quantities at the 1^{st} and 99^{th} percentiles. We exclude instruments that have been traded fewer than 10 times in total in our dataset, to focus on instruments where we have a meaningful number of observations. Finally, we restrict the sample to trades that take place at least 10 days after a bond is issued, to remove any trading in the primary market.

The data present three practical challenges, common when dealing with transactions data: (a) duplicate reporting where, for example, both the buyer and the seller report the transaction; (b) how to treat agency trades, where a firm trades a bond on behalf of another firm; and (c) how to deal with trades with missing counterparty names and IDs.

To remove duplicate trade reports, we identify all trades where the instrument, trading quantity and price match across two trade reports, and the firms reporting the trades differ. We then remove one of these duplicate reports.

With agency trades, we aim to distinguish between two types. The first is agency trading where the client firm on whose behalf the bond is being traded is a non-trading firm, for example a client of a wealth manager or a mutual fund within an asset manager's group. For the purposes of the model, we consolidate the trading firm and its clients into a single group represented by its trading entity - the wealth management firm or the asset management firm in the examples above. The second is agency trading between two trading firms. For the purposes of the model, these are trades between two distinct trading firms, and are thus not consolidated in any way. In the empirical results in Section 3 we (a) do not include agency trades when computing spreads, as agency trades do not earn a spread; and (b) treat each leg of an agency trade as a distinct trade. For example, if A buys a bond from B for C, we count this as two trades: one between A and B, and one between A and C.

In some cases a counterparty in a trade is identified only by an internal code, and not by name. This will most commonly be the case when the counterparty is not a trading firm but, for example, a client of an wealth manager. When computing firm-level summary



(a) UK banks' capital ratios

(b) Risk weights by credit rating

Figure A1: Bank capital ratios and risk weights.

Note: These figures summarise the risk-weighted capital regulations faced by banks. The left panel shows risk-weighted capital ratios for the UK banking system, taken from Bank of England (2021). Risk weights are for exposures to corporates according to the standardised approach to risk weighting, as described in BCBS (2006).

statistics such as the number of trading firms or the distribution of trading frequencies, we do not include unnamed firms as trading firms. However, we do take into account trades with unnamed firms when computing these statistics for named firms. For example, when computing a firm's holding period for a bond, we include their trades with unnamed counterparties to accurately reflect the changes in their asset portfolios.

A1.1 Further Details on Institutional Setting

Figure A1 shows intertemporal and cross-sectional variation in capital regulation. Figure A1a shows how UK banks' average risk-weighted capital ratios have increased through time. Figure A1b shows how capital requirements vary according to a bond's credit rating.

A2 Computational Details

In this section we describe in detail how we solve the model.

At each step of the estimation procedure we solve for the unknown value, pricing, trading quantity, density and search intensity functions. The value shocks are assumed to take discrete values $\beta_1, ..., \beta_{n_\beta}$ and the search costs discrete values $z_1, ..., z_{n_z}$. Each of the functions are continuous in holdings h, and take the form of interpolation splines between the values $h_1, ..., h_{n_h}$. We thus need to solve for each of the functions at $n_T = n_z n_\beta n_h$ points.

We use a nonlinear solver to search over the search intensities γ and densities ϕ at each point on our grid. Conditional on these quantities, and the spline functions we fit through the holdings dimension, we can solve for the value functions V, the price function p and the quantity function q directly without resorting to numerical methods. This significantly reduces the dimensionality of the problem. Below we explain how we compute the value, pricing and quantity functions, before setting out equations we solve numerically to solve for the search and density functions.

Given our matching function (equation (10)), we can solve for the value function solely in terms of the search intensities. To see this, first plug the matching function into the expression for optimal search (equation (5)) and multiply by $\gamma(z, \beta, h)$, yielding:

$$\begin{aligned}\gamma(z, \beta, h) s_2(z, \gamma(z, \beta, h)) = \\ \frac{1}{2} \iiint m(\gamma(z, \beta, h), \gamma(z', \beta', h')) S((z, \beta, h), (z', \beta', h')) \Phi(dz', d\beta', dh'). \quad (13)\end{aligned}$$

Using this equation we can substitute out the final term of the value function (equation 4):

$$\begin{aligned}rV(z, \beta, h) = u(\beta, h) - s(z, \gamma(z, \beta, h)) + \\ \eta \int (V(z, \beta', h) - V(z, \beta, h)) G(d\beta') + \gamma(z, \beta, h) s_2(z, \gamma(z, \beta, h)).\end{aligned}$$

All terms involving the value function V enter this equation linearly, the distribution function ϕ does not enter, and the functional form $s(z, \gamma)$ is known. As a result, conditional on knowing the function γ we have a closed form expression for the value function V at each of the points on our grid. Having computed the value function, we then interpolate along the holdings dimension using a cubic spline to give value functions $V(z, \beta, h)$ that are continuous in h. We can then solve exactly for trading quantity and price at each point on our grid using the Nash bargaining solution, which depends only on the value functions of the two traders that meet.

The variables we pass to the nonlinear solver thus consist of a set of search intensities γ and densities ϕ at each point on the grid. We then fit a cubic interpolation spline through the holdings grid to get a function for $\gamma(z, \beta, h)$ that is continuous in h. For the density $\phi(z, \beta, h)$

we fit a cubic Hermite spline through the holdings values, constraining the function to be a valid density function.[26]

The remaining equations that need solving numerically are the market clearing equation (equation (8)), the search intensity equation (equation (5)) and the distribution equation (equation (6)). The market clearing equation is straightforward to compute given the density function. We compute the terms of the search intensity equation based on a version of the discretized version of the density function ϕ at the points of the grid, with the surplus following directly from the value function and trading quantity.

The distribution function (equation 6) involves a double integral over the holdings of a trader and their potential counterparties involving the quantity traded, their measures and their search intensities, which is potentially difficult to evaluate. However, a property of the Nash bargaining solution simplifies this significantly. In particular, the post-trade holdings of trader (z, β, h) after meeting another trader (z', β', h') depend only on the sum of the two traders' pre-trade holdings, $h^T \equiv h + h'$. We can thus define $h^1(z, \beta, z', \beta', h^T)$ as the post-trade holdings of trader (z, β, h) after meeting trader $(z', \beta', h^T - h)$. Fitting a cubic spline through the holdings dimension of $h^1()$, for any given h^* we can solve for the level of h^T such that $h^1(z, \beta, z', \beta', h^T) = h^*$. The flows into and out of holdings $h < h^*$ from trades between types (z, β) and (z', β') are then as shown in Figure A2, and to get the trading flows we simply integrate meeting rates over the relevant areas. Given we have splines for the relevant expressions, this is an analytical integral, and is straightforward to compute. This enables us to compute the terms of the distribution function (equation (6)).

This process enables us to solve the model at each step of the estimation procedure. We then compute the theoretical moments described in Section 5.1 based on the model solution, and choose the parameters to minimise the distance between these moments and their theoretical counterparts.

A3 Moments

In this section we provide further details on some of the moments used in estimation.

- Correlation between quantity sold and holdings, within-traders:

[26] Hermite splines are particularly convenient for modelling densities as they are shape-preserving, meaning the interpolated curve and the points through which we are interpolating have the same local minima. This makes it easy to constrain the functions to be positive whilst still ensuring they are smooth. See Cai and Judd (2013) and Goodman (2001) for further details.

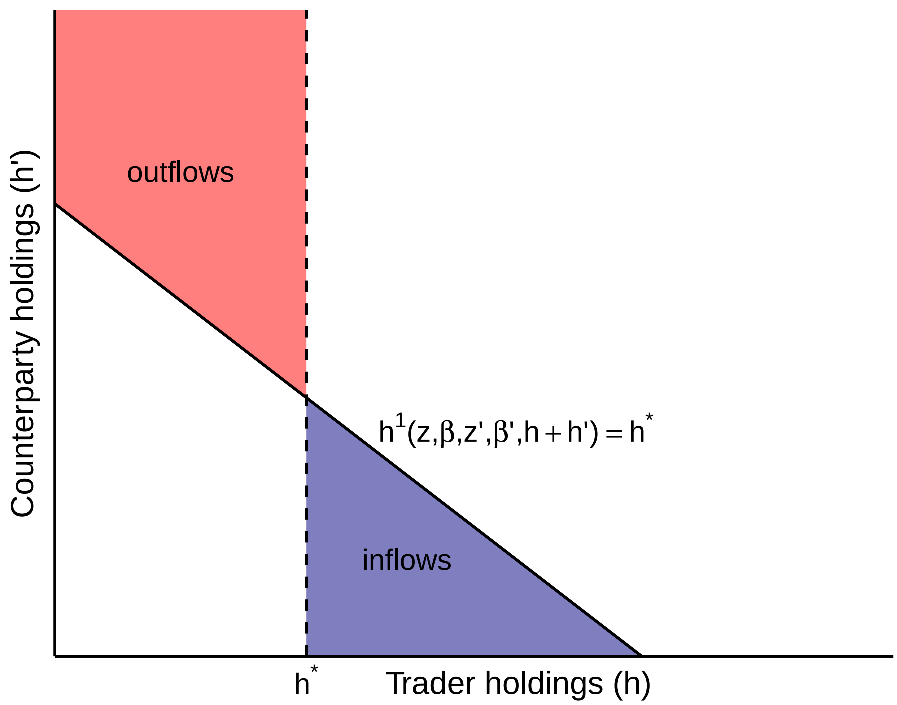


Figure A2: Trading inflows to and outflows from $h \leq h^*$

Note: This figure shows the inflows to and outflows from holdings $h < h^*$. For given (z, β, z', β'), if a trader and counterparty meet with pre-trade holdings in the blue shaded area, the trader's post-trade holdings will be beneath h^*, representing an inflow. If their holdings are in the red shaded area, the trader's post-trade holdings will be above h^*, representing an outflow. The solid line, with gradient -1, denotes meetings that will result in the trader having post-trade holdings of exactly h^*.

$$corr^W(h, q) = \int \frac{cov(h, q|z)}{\sqrt{\mathbb{V}(h|z)\mathbb{V}(q|z)}} f(z) dz,$$

where:

$$cov(h, q|z) = \iint \frac{\gamma(z, \beta, h)\phi(z, \beta, h)}{\Gamma(z) f(z)} \times \int \big(q(z, \beta, h, \Delta') - \mathbb{E}(q|z)\big)\big(h - a\big) \frac{\gamma(\Delta')\phi(\Delta')}{\Gamma} d\Delta' d\beta dh,$$

$$\mathbb{V}_{TW}(h|z) = \iint \frac{\gamma(z, \beta, h)\phi(z, \beta, h)}{\Gamma(z) f(z)} \int \big(h - a\big)^2 \frac{\gamma(\Delta')\phi(\Delta')}{\Gamma} d\Delta' d\beta dh,$$

$$\mathbb{V}(q|z) = \iint \frac{\gamma(z,\beta,h)\phi(z,\beta,h)}{\Gamma(z)f(z)} \int \big(q(z,\beta,h,\Delta') - \mathbb{E}(q|z)\big)^2 \frac{\gamma(\Delta')\phi(\Delta')}{\Gamma} d\Delta' d\beta dh.$$

- Correlation between absolute inventory $inv \equiv |h-a|$ and trading frequency, within traders:

$$corr^W(inv,s) = \int \frac{cov(inv,n|z)}{\sqrt{\mathbb{V}(inv|z)\mathbb{V}(n|z)}} f(z)dz,$$

  where:

$$\mathbb{V}(inv|z) = \iint inv^2 \frac{\phi(z,\beta,h)}{f(z)} d\beta dh,$$

$$cov(inv,n|z) = \iint \big(2\gamma(z,\beta,h)\pi(z,\beta,h) - n(z)\big) inv \frac{\phi(z,\beta,h)}{f(z)} d\beta dh,$$

$$\mathbb{V}(n|z) = \iint \big(2\gamma(z,\beta,h)\pi(z,\beta,h) - n(z)\big)^2 \frac{\phi(z,\beta,h)}{f(z)} d\beta dh.$$