Comparing Search and Intermediation Frictions Across Markets*

Gábor Pintér
Bank of England

Semih Üslü
Johns Hopkins Carey

June 17, 2022

Abstract

In intermediated markets, trading takes time and intermediaries extract rents. We estimate a structural search-and-bargaining model to quantify these trading delays, intermediaries' ability to extract rents, and the resulting welfare losses in government and corporate bond markets. Using transaction-level data from the UK, we identify a set of clients who are active in both markets. We exploit the cross-market variation in the distributions of these clients' trading frequency, prices, and trade sizes to estimate our structural model. We find that trading delays and dealers' market power both play a crucial role in explaining the differences in liquidity across the two markets. Dealers' market power is more severe in the government bond market, while trading delays are more severe in the corporate bond market. We find that the welfare loss from frictions in the government and corporate bond markets are 7.8% and 12.2%, respectively, and our decomposition implies that this loss is almost exclusively caused by trading delays in the corporate bond market, while trading delays and dealers' market power split the welfare loss equally in the government bond market. Using data from the COVID-19 crisis period, we also find that these welfare losses might more than triple during turbulent times, revealing the fragility of the OTC market structure.

JEL Classification*:* D40, G10, G11, L10

Keywords: Search Frictions, Market Power, Government Bonds, Corporate Bonds, Over-the-Counter Markets

*We would like to thank, for helpful comments and suggestions, Ana Babus, Briana Chang, Darrell Duffie, Mark Egan, Thanasis Geromichalos, Foti Grigoris, Joel Hasbrouck, Yunzhi Hu, Liang Ma, John Power, Zhaogang Song, Davide Tomio, Nicholas Trachter, Pierre-Olivier Weill, Milena Wittwer, Liyan Yang, and the seminar participants at Johns Hopkins University, Bank of England, Federal Reserve Bank of Richmond, Universidad Diego Portales, the 2021 Macro-Finance Workshop at the Bank of England, the 2021 Fixed Income and Financial Institutions (FIFI) Conference at USC-Columbia, the 2022 CFM London Macro Workshop, the 12th Financial Markets and Corporate Governance Conference, the 3rd DC Juniors Finance Conference, and the 11th Annual Stern/Salomon Center Microstructure Conference. The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees.

"... liquidity—the ability to trade when and where you want to trade, in significant size and without much cost."

— Larry Harris, TRADING & ELECTRONIC MARKETS: WHAT INVESTMENT PROFESSIONALS NEED TO KNOW (2015, P. 1)

1 Introduction

How large are trading frictions in over-the-counter (OTC) financial markets? How much do these frictions and the associated welfare losses vary across markets and across normal and turbulent times? What is the role of search frictions facing clients and of intermediaries' market power in explaining these cross-market differences? To give quantitative answers to these questions, we estimate a dynamic structural model of OTC market liquidity for the corporate as well as government bond markets—two of the most important markets in the financial system.

We find that search frictions in the UK government bond market are quantitatively modest with the estimated welfare loss amounting to less than 4% compared to the frictionless benchmark. In contrast, search frictions in the UK corporate bond market are more severe and estimated to generate a welfare loss of around 12%. Perhaps surprisingly, we find that dealers' ability to extract rents is significantly larger in the government bond market leading to a welfare loss of almost 4%, while the welfare loss from dealers' rent extraction is negligible in the corporate bond market. In sum, the welfare impact of OTC trading frictions is quite sizable in both markets with a total welfare loss of 7.79% in the government bond market and of 12.17% in the corporate bond market. A re-estimation of the model parameters to fit data from the COVID-19 crisis implies that these percentage welfare losses became 3.2 times as large in both markets during the crisis, pointing to the fragility of the OTC market structure when faced with a large negative shock.

To arrive at these estimates, our paper makes a contribution to both the theoretical and empirical literatures on OTC markets. Our empirical analysis uses a non-anonymous transaction-level dataset which covers close to the universe of all secondary market trades in the UK government and corporate bond markets. Importantly, we are able to identify a common set of clients who are actively trading in both markets and who drive the majority of the trading volume in the client-dealer segment of both markets. This unique feature of the data allows us to exploit *cross-market* differences in all three main dimensions of market liquidity: trading frequency, trade size, and transaction price. Comparing trading frictions across markets is a hard task because client composition is endogenous to the given market in question. Our approach of keeping the set of clients fixed in the two markets goes a long way in addressing these selection issues, allowing for a comparison of the trading frictions that the *same clients* face in two different markets.

The theoretical contribution of the paper is to develop a model of OTC trading with two-sided search and bilateral bargaining that can be structurally estimated using transaction-level data with client identities. The model is based on the stationary version of Lagos and Rocheteau (2009), which we extend to allow for multi-dimensional heterogeneity in client characteristics. These extensions are motivated by the substantial client-level heterogeneity we document in both the government bond and the corporate bond market. We introduce client heterogeneity both in the frequency of preference shocks and of the arrival of trading opportunities that, in turn, translates into heterogeneity in clients' endogenous trade frequencies as in the data.

In our dynamic model, clients have heterogeneous and time-varying marginal utility for holding a perfectly divisible homogenous asset, and so, they ideally want to hold different amounts of the asset from one another and over time. In the estimated version of our model, clients' marginal utility types are binary, high or low, and change over time following client-specific continuous-time Markov chains. Thus, a given client has an ideal high-type asset position and an ideal low-type asset position in mind. If a client's current taste matches her ideal asset position, the client is happy and does not need to trade. However, a switch of the client's taste from high to low, or vice versa, makes the client want to update her asset position accordingly. Importantly, these two ideal asset positions and the resulting trade size the client wants to trade are affected by how frequently she expects to switch to the opposite taste type and how frequently she can match a dealer to trade. These are the novel client heterogeneity dimensions of our model relative to Lagos and Rocheteau (2009).

Clients in our model can trade only with dealers and only in a bilateral fashion subject to search and bargaining frictions, while dealers can trade amongst themselves multilaterally and without frictions. Bilateral bargaining in our model leads to client-specific bid and ask prices. When switching from the low-type asset position to the high-type position, the client pays a negotiated ask price per share of the asset bought from the dealer. Vice versa, when switching from the high-type asset position to the low-type position, the client receives a negotiated bid price per share of the asset sold to the dealer. In the end, our model generates client-specific trade frequencies, trade sizes, and prices as in the data. This close resemblance of our model's endogenous outcomes with the transaction-level data from the real-world fixed-income markets allows us to identify the deep parameters of our model.

Despite rich heterogeneity in the model, we derive a number of equilibrium objects that are readily observable from the data as (integral transforms of) closed-form expressions. These include trading volume, price dispersion, the distribution of clients' average trade frequencies, and client-specific average trade size. We then use these data moments to estimate some of the deep parameters of our structural model. The estimation results reveal that trading delays are vastly different across the two markets. Average trading delays in the government bond market can be measured in minutes, whereas in the corporate bond market they can be measured in hours with a median of around three quarters of a trading day. These results highlight the large differences in the severity of search frictions that the two bond markets exhibit. Moreover, the estimation implies that the share of transaction surpluses that dealers capture is about three times as large in the government bond market as in the corporate bond market, which means that dealers have significantly higher market power in the government bond market compared to the corporate bond market.[1] These findings pose a challenge to the common view that high dealer market power and severe search frictions must co-exist in an intermediated market. We find that, despite the reasonably low level of search frictions, government bond dealers can exert a significant market power over their clients.

Our analysis highlights the importance of utilizing information regarding all three dimensions of market liquidity, trade frequency, trade size, and price, to have a full understanding of the drivers of market liquidity, which are the deep parameters of our model. For example, it would be impossible to distinguish between how often clients need to trade (clients' preference shock frequency) and how often they can trade (clients' matching efficiency with dealers) by looking at their trade frequencies only. However, we show that endogenous trade frequency increases with both preference shock frequency and matching efficiency, while endogenous trade sizes decline with preference shock frequency but increase with matching efficiency. Thus, by using both trade frequency and trade size information, it is possible to simultaneously identify preference shock frequency and matching efficiency.

Similarly, it would be impossible to distinguish between search frictions (inverse of matching efficiency) and dealers' market power by looking at price dispersion only because both search frictions and dealers' market power are positively related with the endogenous price dispersion. However, as argued, we are able to identify the level of search frictions using information on trade frequency and trade size. Thus, the additional unique information on price dispersion allows us to identify dealers' ability to extract rents that we name *dealers' market power*. Our empirical analysis shows that average trade frequency in the government bond market is more than six times that of the corporate bond market. This contributes to a large difference between the estimated matching efficiency levels in the two markets. Despite this large difference between the two markets in terms of search frictions, price dispersion of government bonds is only around half of the corporate bond price dispersion. Thus, our estimation "rationalizes" this with a significantly larger dealer market power in the government bond market.

To compare the two markets in terms of the welfare consequences of their friction levels, we use our parameter estimates and calculate the market participants' welfare in equilibrium, in the first-best allocation (the solution to an unconstrained planner's problem), and in the second-best allocation (the solution to a constrained planner's problem). We confirm that the first-best allocation coincides with the frictionless benchmark allocation and that the second-best coincides

[1] Share of transaction surpluses captured by sellers in bargaining with buyers is widely used in monetary economics literature to model sellers' market power, as it provides a convenient way to model market power in the presence of continuum of agents. See, for example, Lagos and Wright (2005) and Choi and Rocheteau (2021). In modeling OTC financial markets, it is not only convenient, but accurate: "These search-and-bargaining features are empirically relevant in many markets, such as those for mortgage-backed securities, corporate bonds, emerging market debt, bank loans, derivatives, and certain equity markets" (Duffie, Gârleanu, and Pedersen, 2005, p. 1815).

with the allocation which would obtain if dealers did not have any market power. This analysis allows us to decompose the total welfare loss into a component caused by dealers' market power and another component caused purely by search frictions. We find that the total welfare losses in the government and corporate bond markets are 7.79% and 12.17%, respectively, and our decomposition implies that this loss is almost exclusively caused by search frictions in the corporate bond market, while search frictions and dealers' market power split the welfare loss equally in the government bond market. To gauge the reliability of our welfare loss estimations, we calculate the minimum and the maximum welfare loss implied by the 95% confidence intervals of our baseline parameter estimates. The resulting bounds are 6.63% and 9.51% for the welfare loss in the government bond market and 11.09% and 13.64% in the corporate bond.

In the last part of the paper, we re-estimate the model parameters to match trading activity during the COVID-19 crisis. Our parameter estimates imply that the total welfare losses in the government and corporate bond markets are 25.44% and 39.63%, respectively, and our decomposition implies that this loss is almost exclusively caused by search frictions in the corporate bond market as in normal times, while dealers' market powers' share in the welfare loss in the government bond market increased from around 50% to 60%. Overall, these estimates imply that the welfare losses from OTC market frictions are especially severe during turbulent times. One counterfactual exercise we conduct is about what the resulting welfare losses would be if the OTC market structure was not particularly fragile during turbulent times. To this end, we calculate the welfare losses by keeping matching efficiency and dealers' market power exactly at the level of normal times, but accommodating the clients' preference parameters that reflect the COVID-19 shock. We find that the total welfare loss in the government and corporate bond markets would be 10.72% and 15.25%, respectively. This counterfactual analysis implies that the vast majority of the additional welfare loss during turbulent times is because of the fragility of the OTC market structure when faced with a large negative shock.

1.1 Related Literature

There is a vast literature on empirical analysis of OTC financial markets. See, for example, Garbade and Silber (1976), Edwards, Harris, and Piwowar (2007), Jankowitsch, Nashikkar, and Subrahmanyam (2011), Di Maggio, Kermani, and Song (2017), O'Hara, Wang, and Zhou (2018), Li and Schürhoff (2019), Dick-Nielsen, Poulsen, and Rehman (2021), and Kondor and Pintér (2022). While these papers offer reduced-form empirical analyses to determine stylized facts specific to OTC markets and to test some economic hypotheses that may explain those stylized facts, we offer a structural empirical analysis of the determinants of OTC market liquidity through the lens of a search-based model. Hotchkiss and Jostova (2017), Cestau, Hollifield, Li, and Schürhoff (2019), and Bessembinder, Spatt, and Venkataraman (2020) provide recent surveys of the empirical literature.

A theoretical literature following Duffie, Gârleanu, and Pedersen (2005) and Lagos and Rocheteau (2009, henceforth LR) that modeled OTC financial markets with a search-based framework focused mostly on offering qualitative insights. A non-exhaustive list of such papers includes Duffie, Gârleanu, and Pedersen (2007), Vayanos and Wang (2007), Vayanos and Weill (2008), Chang and Zhang (2016), Chiu and Koeppl (2016), Bethune, Sultanum, and Trachter (2018), Farboodi, Jarosch, and Shimer (2018), Üslü (2019), Üslü and Velioğlu (2019), and Hugonnier, Lester, and Weill (2020), among others. See Weill (2020) for a recent survey. In this paper, we use insights from this literature to quantitatively compare the severity of search frictions and the market power of intermediaries across different OTC markets. Inspired by the empirical studies documenting a high level of heterogeneity among clients, we develop an intermediated OTC market model with ex-ante heterogeneity by extending LR and structurally estimate it.[2]

The list of papers that adopt an intermediated OTC market framework with a competitive inter-dealer market similar to LR includes Lagos and Rocheteau (2006), Lester, Rocheteau, and Weill (2015), Sultanum (2018), Chang and Zhang (2021), Colliard, Foucault, and Hoffmann (2021), Chiu, Davoodalhosseini, and Jiang (2022), and Kargar, Lester, and Weill (2022), among others. These papers focus mainly on offering theoretical insights and a few of them incorporate calibrated numerical examples. In contrast, we structurally estimate our model to compare the distributions of clients' exposure levels to frictions and the dealers' market power in the UK government bond and corporate bond markets. Accordingly, we also quantify the welfare loss from search and intermediation frictions in these markets.

Structural estimation of dynamic decentralized trading models is an understudied area. Examples include Feldhütter (2012), Gavazza (2016), Brancaccio, Li, and Schürhoff (2017), Buchak, Matvos, Piskorski, and Seru (2020), Hendershott, Li, Livdan, and Schürhoff (2020), Liu (2020), and Coen and Coen (2021). With the exception of Coen and Coen (2021), these papers model the trade of an indivisible unit of an asset, while we model trading a divisible asset with trade sizes optimally chosen. Accordingly, while these papers can be regarded as estimating a version of the Duffie, Gârleanu, and Pedersen (2005) model (an indivisible asset model), we estimate a version of the LR model (a divisible asset model). There are several reasons behind this modeling choice we make. First, trade size heterogeneity is a prevalent empirical fact in markets for financial assets as we show in the UK government bond and corporate bond markets and as Üslü (2019) shows in the US corporate bond market. Second, as LR argues, optimally choosing their trade size provides clients with a flexibility to respond to trading frictions, which, in turn, affects the endogenous outcomes like liquidity and welfare. Third, and relatedly, this allows us to use information on trade sizes to identify the deep parameters of our model. Trade sizes are especially important in identifying simultaneously the clients' preference shock frequency and matching efficiency with

[2]Empirical work that studied heterogeneity among clients in OTC markets includes O'Hara, Wang, and Zhou (2018) and Hendershott, Li, Livdan, and Schürhoff (2020), among others.

dealers from transaction-level data.[3] In their independent, contemporaneous work, Coen and Coen (2021) also allow for endogenous trade sizes in a decentralized market with all-to-all trading *à la* Üslü (2019). They analyze the quantitative impact of post-crisis regulations on the functioning of the inter-dealer market and the degree of substitutability of liquidity provision by dealers and by clients in the UK corporate bond market. We abstract away from the inter-dealer market frictions, and focus exclusively on the frictions in the client-dealer segment of the UK government and corporate bond markets, instead. As in Üslü (2019), Coen and Coen (2021) assume that all market participants split transaction surpluses in half. This precludes the possibility of identifying dealers' market power within the context of their model, while it is one of our main motivations in this paper.[4]

There is also a literature on structural estimation of static decentralized trading models. See, for example, Eisfeldt, Herskovic, Rajan, and Siriwardane (2018), Allen, Clark, and Houdec (2019), Allen and Wittwer (2021), Hendershott, Livdan, and Schürhoff (2021), and Beltran (2022). Because of their static nature, these models do not feature explicit trading delays, and typically rely on reduced-form search costs. Our dynamic model, instead, features trading delays, and search is costly because of delayed trade, not because there are physical search costs. This is arguably a realistic approach to search in financial markets.[5] Another strand of literature estimates auction models to analyze primary markets. We instead study the secondary market trading in OTC financial markets. For examples of the estimation of primary market auction models, see Kang and Puller (2008), Hortaçsu and McAdams (2010), and Kastl (2011), among others. Clark, Houdec, and Kastl (2021) provide a survey of this literature.

Finally, there is a tradition of estimating structural search models in labor economics, industrial organization, and financial intermediation literatures. A non-exhaustive list of papers includes Eckstein and Wolpin (1990) and Gautier and Teulings (2015) from labor economics; Hong and Shum (2006), De los Santos, Hortaçsu, and Wildenbeest (2012), and Galenianos and Gavazza (2017) from industrial organization; and Hortaçsu and Syverson (2004), Woodward and Hall (2012), and Egan (2019) from financial intermediation. See Eckstein and van den Berg (2007) and Gavazza and Lizzeri (2021) for more comprehensive surveys. Apart from studying different markets and having a special focus on comparing markets, another important feature of our analysis is

[3] For example, Gavazza (2016) can also identify simultaneously the preference shock frequency and the matching efficiency in the market for business aircraft, in which the traded asset is naturally indivisible and trade sizes are fixed at one. The fraction of aircraft for sale is an observed variable in his dataset, which helps him identify the preference shock frequency and the matching efficiency in the absence of trade size variation. In transaction-level data from financial markets, there is no counterpart of the fraction of aircraft for sale, but there is trade size variation, which motivates our modeling choice.

[4] In addition, the *semi-centralized* market structure in our model inherited from LR allows us to obtain the theoretical moments in closed form or as integral transforms of some closed-form expressions. Hence, our generalized method of moments (GMM) estimation makes use of theoretical moments of analytical form, while Coen and Coen (2021) rely on numerical solutions of the equilibrium objects and so their GMM estimation takes the form of a nested loop where a nonlinear solver is used to search over endogenous objects as well as model parameters.

[5] See Weill (2020) for a discussion.

that we place a particular emphasis on endogenous trade sizes and financial asset divisibility, while these papers model negotiating over an indivisible unit of labor or of a financial or consumer product.

The rest of the paper is organized as follows. Section 2 introduces the dataset we use and provides descriptive statistics for the data moments we employ in the structural estimation stage. Section 3 introduces the structural model environment, characterizes its equilibrium, and derives formulae for the theoretical moments. Section 4 describes the estimation procedure, reports the parameter estimates and the bootstrap standard errors, and presents our counterfactual analyses regarding welfare. Section 5 re-estimates the model parameters to fit data from the COVID-19 crisis. Section 6 concludes.

2 Data and Measurement

2.1 Data Source and Sample Selection

To compare trading frictions across government and corporate bond markets, we use a regulatory, trade-level dataset, which covers close to the universe of secondary market trades in the UK markets. A main advantage of the so-called ZEN database is that, unlike other datasets typically employed in the literature (e.g. TRACE), it contains the *identities of both counterparties* for each transaction in addition to information on the time stamp, the transaction amount and price, the International Securities Identification Number, the account number, and buyer-seller flags.[6] The granularity of the dataset also allows us to identify clients who trade in both government and corporate bonds in a given time period. This feature of the data enables us not only to explore heterogeneity in search frictions across clients, but *across markets* as well.[7]

Our baseline sample covers the period between Aug 2011 and Dec 2017. We filter out duplicates and erroneous entries, and exclude all inter-dealer trades as well as client-to-client trades. We identify 574 clients who are active in both markets, and whose trades cover the majority of total client trading volume. Moreover, our definition of dealers is the set of gilt-edged market makers (GEMMs) who perform market-making functions in both bond markets. Their number fluctuates around 20 in our sample.[8] We end up with approximately 1.15 million trades in 57 nominal

[6] See Czech, Huang, Lou, and Wang (2021), Czech and Pintér (2020), and Kondor and Pintér (2022) for further details and recent applications of this dataset that mainly focused on identifying informed trading in corporate and government bond markets.

[7] This is a major advantage of our dataset compared to those used in the recent OTC literature using structural search models. For example, Hendershott, Li, Livdan, and Schürhoff (2020) are only able to observe a subset (insurance companies) of clients in one market (corporate bonds) only.

[8] Certain large clients (particularly in the corporate bond market) have emerged to perform market making functions. We exclude them from our set of dealers, and focus on GEMMs in order to have a common set of dealers across the two markets. Note that the ZEN database is maintained by the UK's Financial Conduct Authority (FCA), and the database contains all secondary market transactions, where at least one of the counterparties is an FCA-regulated entity. Given that all GEMMs as well as many of the active clients are FCA-regulated, our

government bonds and about 1.12 million trades in about 2700 corporate bonds. Further details on our data construction procedure are provided in Appendix B. Another common name in the UK for nominal government bonds is conventional *gilts*. Thus, we use the gilt market and the UK government bond market interchangeably throughout the paper.

Risk characteristics are an important dimension that can make the representative corporate bond distinct from government bonds. To mitigate this issue, we aim to control for heterogeneity in risk profiles across the two markets by excluding high-yield corporate bonds.[9] In addition, from the set of investment-grade bonds we keep the 57 bonds that have the highest number of transactions, thereby obtaining a sample of about 187,000 corporate bond transactions that has the same number of distinct assets as our sample of government bonds.[10] Our sample selection aims to minimize cross-market heterogeneity in payoff risk and adverse selection risk, thereby facilitating a better identification of the cross-market differences in search and intermediation frictions.

Moreover, identifying a common set of clients as well as dealers who operate in corporate and government bonds also mitigates some of the selection problems that may impede any cross-market analysis. For example, certain clients such as foreign central banks may specialize in trading government bonds and these clients may have very different characteristics as well, i.e., the composition of clients may be endogenous to the given market, which would make the comparison of frictions across markets more difficult.

2.2 Data Description

Table 1 presents summary statistics of the main variables for both markets that will be used in the empirical analysis. The variables are computed separately for each market on each trading day.

To measure price dispersion, we first compute the absolute deviation of each transaction price in our dataset from the hourly average transaction price. We scale this deviation by the hourly average price so that we can compute daily averages of the transaction-specific absolute deviations across all available bonds in the given market. We compute daily average dispersion (across all available bonds) by weighting each observation by the size of the corresponding trade (Jankowitsch, Nashikkar, and Subrahmanyam, 2011). As shown in Table 1, average price dispersion is

dataset covers virtually the entire secondary market trading activity in UK corporate and government bonds. For further details on the identities of GEMMs, see https://www.dmo.gov.uk/responsibilities/gilt-market/market-participants/.

[9] We match our corporate bond dataset with information on corporate bond ratings from Thomson Reuters Eikon, covering the three major rating agencies Moody's, Standard & Poor's (S&P), and Fitch. Ratings of Moody's are used as the default option because of the firm's large market coverage. S&P ratings are used if ratings from Moody's are not available for the given bond. Fitch ratings are used as a third option.

[10] Corporate bonds that are traded less frequently are more likely to be subject to adverse selection risk (Ronen and Zhou, 2013; Benmelech and Bergman, 2018).

about 6.4 bps in the government bond market and about 10.7 bps in the corporate bond market.[11] Since Garbade and Silber (1976), price dispersion has often been used as a proxy for the severity of trading frictions in decentralized financial markets. The newly documented fact that price dispersion is about 40% larger in corporate bonds than government bonds is suggestive of corporate bond markets being more frictional than government bond markets. However, without additional information on clients' trading intensities and other quantities, price dispersion alone is not informative on the nature of the trading frictions (search vs. intermediation frictions) and on the exact difference in welfare losses due to trading frictions across the two markets.

To measure average intensity, we compute the mean of the total number of transactions of the each of the 574 clients on each trading day. In addition, we compute intensity dispersion as the mean of the absolute deviation of clients' total number of transactions from average intensity. Both measures are scaled by the number of assets (57) in each sample. The second and third rows of Table 1 shows that average intensity and intensity dispersion are 0.023 and 0.036, respectively, in the government bond market, and they are 0.0038 and 0.0068 in the corporate bond market. That is, average intensity and intensity dispersion are about 5-6 times larger in the government bond market compared to corporate bonds. While the recent empirical literature (O'Hara, Wang, and Zhou, 2018) studied intensity in corporate bond markets, the cross-market comparison in intensities is novel. The large difference in intensity measures across the two markets is again indicative of the corporate bond market being more frictional than the government bond market. However, without a structural model and additional empirical moments, the challenge remains to identify whether clients in corporate bond markets are *unable* or *unwilling* to trade more than in government bond markets.

We measure bond turnover as total daily trading volume scaled by our proxy for asset supply, the amount of issued bond outstanding. The natural logarithm of turnover in our sample is about 2.1 log points larger in government bonds (-4.8) than in corporate bonds (-6.9).

Finally, recent search models of OTC markets (LR) predict that the severity of trading frictions faced by a client is an important determinant of the trade size demanded by the given client. To measure average trade size, we first compute the daily mean of the nominal size of each clients' trades, and then compute the mean across the clients. We scale this measure by asset supply as well. We find that the natural logarithm of scaled average trade size is similar (-11.38 vs. -11.40) across the two markets. The similarity of these two values is driven by the fact asset supply is

[11] To link these estimates to the existing measures of price dispersion in the literature, note that Jankowitsch, Nashikkar, and Subrahmanyam (2011) used a sample from the TRACE database to find a mean price dispersion of about 50 bps in the US corporate bond market. Their estimate is measured in standard deviations and used end-of-day price quotes as the benchmark price. As a cross-check, we tried to compare our results and sample to theirs, by (i) including all available corporate bonds in our calculation, (ii) converting our absolute deviation based measure into standard deviation, (iii) and using end-of-day price quotes as benchmark price. We find price dispersion to be about 40 bps in government bonds and 72 bps in corporate bonds. Our choice of using higher-frequency (hourly instead of end-of-day) benchmarks aims to mitigate the over-estimation of dispersion that can be caused by the arrival of intra-day news in the market.

Table 1: Summary Statistics

	Government Bonds		Corporate Bonds		
Variable	Mean	sd	Mean	sd	N
Price Dispersion	0.0006368	0.0002587	0.0010714	0.0009618	1614
Average Intensity	0.0231884	0.0051587	0.0038278	0.0009539	1614
Intensity Dispersion	0.0361403	0.0076654	0.00681	0.0015803	1614
Log Turnover	-4.803826	0.4256643	-6.90944	0.6709676	1614
Average Log Trade Size	-11.38151	0.3800972	-11.40813	0.6642129	1614

Notes: This table summarizes the empirical moments that we use in the structural estimation. Price dispersion is the scaled mean absolute deviation of transaction prices from the average transaction price in the given hour. Average intensity is the mean of the clients' number of transactions. Intensity dispersion is the mean absolute deviation of clients' number of transactions from average intensity. Turnover is computed as the daily trading volume scaled by asset supply. Average trade size is the mean (across clients) of clients' mean trade size scaled by asset supply. The sample includes 1614 trading days over the period Aug 2011 - Dec 2017.

considerably larger in government bonds than in corporate bonds.[12] We scale both volume and average trade size in both markets to be able to estimate our model parameters using scale-free empirical moments.

3 A Model of Intermediated OTC Markets

To capture the rich client heterogeneity observed in fixed-income markets in practice, we extend the stationary version of LR along two dimensions. Namely, clients in our model are ex-ante heterogeneous in the frequency at which they switch valuation types as well as the frequency at which they receive trade opportunities. In addition, we allow these cross-sectional frequency distributions to have a continuous unbounded support. While this introduces some technical hurdles for the equilibrium characterization, it allows us to make use of parsimonious and realistically skewed distributions in the structural estimation stage.

3.1 The Economic Environment

Time is continuous and runs forever. The economy is populated by a continuum of clients and a continuum of dealers whose measures are both normalized to one.

There is a long-lived asset in exogenous supply $A > 0$. There is also a perishable good, called *numéraire*, that all agents consume and produce. Negative net consumption is allowed in the sense that if an agent produces more numéraire good to support her purchase of the asset than she consumes, her net consumption becomes negative. The flow utility of a client is

$$c + \varepsilon u(a),$$

[12] Asset supply is about £750 billion and £60 billion in our estimation subsample for the UK government and corporate bonds, respectively. Unscaled average trade size would be more than 10 times larger in government bonds, consistent with Belsham, Rattan, and Maher (2017).

where c is her net consumption of the numéraire, a is her asset position, and ε is her current taste type. The felicity function $u : [-M, M] \to \mathbb{R}$ is twice continuously differentiable, strictly increasing, strictly concave, and defined for an arbitrarily large $M > 0$. Clients also discount the future at rate $r > 0$.

Clients are heterogeneous in their taste types whose variation over time is governed by a continuum of pair-wise independent Poisson processes. Upon an arrival, the shocked client's new taste type, ε', is drawn from the cdf $F : [\varepsilon_l, \varepsilon_h] \to \mathbb{R}$. These Poisson taste shocks generate time-varying exogenous heterogeneity across clients, and so, generate the fundamental motive to trade. In addition to the current taste heterogeneity, clients are heterogeneous with respect to two permanent characteristics, $\chi_1 > 0$ and $\chi_2 > 0$, where χ_1 refers to the rate at which a client receives taste shocks and χ_2 the rate at which she receives trade opportunities with dealers. We denote with $\chi = (\chi_1, \chi_2)$ a client's two-dimensional permanent characteristics. The cross-sectional distribution of clients' characteristics is represented by the joint cdf $G : \mathbb{R}^2_+ \to \mathbb{R}$. We follow Vayanos and Wang (2007) in interpreting a large-χ_1 client as a liquidity trader and a small-χ_1 as a buy-and-hold trader. Similarly, we follow O'Hara, Wang, and Zhou (2018) in interpreting a large-χ_2 client as an active trader and a small-χ_2 as a passive trader.

Dealers' utility flow is assumed to be c; i.e., they do not derive utility from holding the asset. Dealers can trade the asset instantaneously in a competitive inter-dealer market at the market-clearing price P. Accordingly, we assume without loss of generality that dealers do not hold any position in the asset as they do not derive utility from holding the asset. Clients can trade only with dealers, infrequently, and in a bilateral fashion, i.e., with one dealer at a time. Clients receive trading opportunities at the arrival times of a continuum of pair-wise independent Poisson processes with client-specific arrival rates of χ_2.

Upon the arrival of a trade opportunity shock, the shocked client is matched with a dealer picked randomly and uniformly from the pool of dealers. Each meeting between a dealer and a client is followed by a Pareto-optimal bargaining game during which the bargaining parties determine a trade quantity that maximizes the joint-surplus of the trade and a trade price that splits the maximized joint-surplus between the dealer and the client. In the end, the dealer captures a fraction $\eta \in [0, 1]$ of the surplus, while the client captures the remaining share $1 - \eta$.[13]

3.2 Equilibrium Definition

Let $V(\varepsilon, a, \chi)$ refer to the continuation utility of a client with the current taste type of ε, current asset position a, and the permanent characteristic vector χ. If this client meets a dealer at this

[13] This Pareto-optimal bargaining solution can be understood as the generalized Nash solution or the proportional Kalai solution. Aruoba, Rocheteau, and Waller (2007) show that the Nash solution and the Kalai solution may be very different when agents trade divisible assets with a payment constraint. We do not run into this problem because our agents are assumed to be able to produce the numéraire good, and so, our agents do not have any payment constraint. Therefore, both the Kalai and the Nash solutions generate the same Pareto-optimal result we describe.

moment, the number of shares of the asset the client buys, $q(\varepsilon, a, \chi)$, solves

$$\max_{q \in \mathbb{R}} V(\varepsilon, a+q, \chi) - V(\varepsilon, a, \chi) - Pq, \tag{3.1}$$

where $V(\varepsilon, a+q, \chi) - V(\varepsilon, a, \chi)$ represents the client's contribution to surplus creation and is equal to the change in her continuation utility after she has bought q units of the asset (or after she has sold $-q$ units if q is negative). The last term, $-Pq$, represents the dealer's contribution to the surplus creation and is equal to the cost of obtaining q units of the asset from the inter-dealer market (or the benefit of selling $-q$ units in the inter-dealer market if q is negative). Note that the trade price between the client and the dealer does not enter the joint-surplus formula because it is simply a transfer of the numéraire from one party to the other when both have a linear utility in that.

Trade price that the client pays to the dealer per unit of the asset traded is denoted by $p(\varepsilon, a, \chi)$ and is equal to

$$p(\varepsilon, a, \chi) = P + \eta \frac{V(\varepsilon, a+q(\varepsilon, a, \chi), \chi) - V(\varepsilon, a, \chi) - Pq(\varepsilon, a, \chi)}{q(\varepsilon, a, \chi)}. \tag{3.2}$$

Since the joint-surplus in the numerator of the fraction in (3.2) is non-negative by the optimal choice of $q(\varepsilon, a, \chi)$, this means that the dealer charges a markup over the inter-dealer price P when selling the asset to the client ($q(\varepsilon, a, \chi) > 0$). Vice versa, the dealer obtains a markdown when buying the asset from the client ($q(\varepsilon, a, \chi) < 0$).

The continuation utility $V(\varepsilon, a, \chi)$ satisfies the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned} rV(\varepsilon, a, \chi) = \varepsilon u(a) + \chi_1 \int_{\varepsilon_l}^{\varepsilon_h} \left[V(\varepsilon', a, \chi) - V(\varepsilon, a, \chi)\right] dF(\varepsilon') \\ + \chi_2 \left[V(\varepsilon, a+q(\varepsilon, a, \chi), \chi) - V(\varepsilon, a, \chi) - q(\varepsilon, a, \chi)\, p(\varepsilon, a, \chi)\right]. \end{aligned} \tag{3.3}$$

A client's "flow" continuation utility, $rV(\varepsilon, a, \chi)$, equals the sum of three terms. The first term is the client's current utility flow from holding a units of the asset when of taste type ε. The second term is the flow value of switching to another taste type ε' at a Poisson rate of χ_1. The last term is the flow value of changing the asset position from a to $a + q(\varepsilon, a, \chi)$ by paying $q(\varepsilon, a, \chi)\, p(\varepsilon, a, \chi)$ units of the numéraire, which is also an infrequent possibility arriving at a Poisson rate of χ_2.

Let $\Phi_\chi(\varepsilon, a)$ denote the stationary joint cdf of clients' taste types and asset positions conditional their permanent characteristics. The stationarity of this cdf is guaranteed by the following inflow-outflow equation:

$$\chi_1 F(\varepsilon) \int_{\varepsilon}^{\varepsilon_h} \Phi_\chi(d\varepsilon', a) + \chi_2 \int_{\varepsilon_l}^{\varepsilon} \int_{-M}^{M} \mathbb{I}_{\{a \leq a' + q(\varepsilon', a', \chi)\}} \Phi_\chi(d\varepsilon', da')$$
$$= \chi_1 (1 - F(\varepsilon)) \Phi_\chi(\varepsilon, a) + \chi_2 \Phi_\chi(\varepsilon, a). \quad (3.4)$$

The first term of the LHS and of the RHS stand in for the inflow and the outflow due to taste shocks, respectively. Similarly, the second terms stand in for the inflow and the outflow due to trade.

To understand the first term of the LHS, note that a client with characteristics $\chi = (\chi_1, \chi_2)$, an asset position smaller than a, and a taste type larger than ε switches to a taste type smaller than or equal to ε with probability $F(\varepsilon)$ following a taste shock that occurs at rate χ_1. Thus, the multiplication of χ_1, the probability $F(\varepsilon)$, and the measure of such clients gives us the inflow to $\Phi_\chi(\varepsilon, a)$ due to taste shocks. Similarly, the first term of the RHS, the outflow due to tastes shocks, is equal to the multiplication of the Poisson intensity of taste shock, χ_1, the probability, $1 - F(\varepsilon)$, that the new taste type is larger than ε, and the measure of clients from whom the outflow is originating.

The second term of the LHS, the (gross) inflow to $\Phi_\chi(\varepsilon, a)$ due to trade, is the multiplication of the Poisson intensity of trade opportunities for this class of clients, χ_2, and the measure of a subset of clients with characteristics χ. The indicator function inside the integral makes sure that a candidate inflow client wants to hold an asset position less than a so she indeed creates an inflow. The second term of the RHS, similarly, stands in for the (gross) outflow from $\Phi_\chi(\varepsilon, a)$, which is equal to the multiplication of the Poisson rate of trade opportunities and the measure of clients with characteristics χ who have an asset position smaller than a and a taste type smaller than ε.

In addition to the stationarity condition (3.4), there are three additional feasibility or accounting identities that $\Phi_\chi(\cdot, \cdot)$ must satisfy:

$$\int_0^{\chi_2} \int_0^{\chi_1} \int_{-M}^{M} \int_{\varepsilon_l}^{\varepsilon_h} \Phi_{\chi'}(d\varepsilon, da)\, G(d\chi_1', d\chi_2') = G(\chi_1, \chi_2) \quad (3.5)$$

for all $\chi \in \operatorname{supp}(dG)$,

$$\int_{-M}^{M} \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi(d\varepsilon, da) = 1, \quad (3.6)$$

for all $\chi \in \operatorname{supp}(dG)$, and

$$\int_0^\infty \int_0^\infty \int_{-M}^{M} \int_{\varepsilon_l}^{\varepsilon_h} a \Phi_\chi (d\varepsilon, da)\, G(d\chi_1, d\chi_2) = A. \tag{3.7}$$

Equation (3.5) implies that the equilibrium conditional distribution of clients' states is consistent with the exogenous distribution of client characteristics, (3.6) follows from the fact that $\Phi_\chi(\cdot,\cdot)$ is a conditional cdf, and (3.7) guarantees that all units of the asset are held by clients, i.e., the inter-dealer market clears and dealers do not hold inventory.

Taking stock, we define a stationary equilibrium as follows.

Definition 1 *A stationary equilibrium is (i) a function $V : [\varepsilon_l, \varepsilon_h] \times [-M, M] \times \mathbb{R}^2_+ \to \mathbb{R}$ for clients' continuation utilities, (ii) a function $q : [\varepsilon_l, \varepsilon_h] \times [-M, M] \times \mathbb{R}^2_+ \to \mathbb{R}$ for clients' trade sizes, (iii) a function $p : [\varepsilon_l, \varepsilon_h] \times [-M, M] \times \mathbb{R}^2_+ \to \mathbb{R}$ for clients' transaction prices, (iv) a joint cdf $\Phi_\chi : [\varepsilon_l, \varepsilon_h] \times [-M, M] \to \mathbb{R}$ for clients' taste types and asset positions conditional on their characteristics $\chi \in \operatorname{supp}(dG)$, and (v) an interdealer market price $P \in \mathbb{R}$ such that*

- *Given (ii) and (iii), (i) solves the HJB equation (3.3).*
- *Given (i) and (v), (ii) maximizes the joint surplus (3.1).*
- *Given (i), (ii), and (v), (iii) splits the maximized joint surplus between the client and the dealer according to (3.2).*
- *Given (ii), (iv) satisfies the stationarity and feasibility conditions (3.4)-(3.7).*
- *(v) is implied by (ii) and (iv), i.e., the interdealer market clears by Walras' law.*

3.3 Equilibrium Characterization

Substituting (3.2) into (3.3),

$$\begin{aligned} rV(\varepsilon, a, \chi) = \varepsilon u(a) + \chi_1 \int_{\varepsilon_l}^{\varepsilon_h} \left[V(\varepsilon', a, \chi) - V(\varepsilon, a, \chi)\right] dF(\varepsilon') \\ + \chi_2 (1-\eta) \left[V(\varepsilon, a + q(\varepsilon, a, \chi), \chi) - V(\varepsilon, a, \chi) - q(\varepsilon, a, \chi) P\right]. \end{aligned}$$

Using (3.1) and with a change of variable,

$$\begin{aligned} rV(\varepsilon, a, \chi) = \varepsilon u(a) + \chi_1 \int_{\varepsilon_l}^{\varepsilon_h} \left[V(\varepsilon', a, \chi) - V(\varepsilon, a, \chi)\right] dF(\varepsilon') \\ + \chi_2 (1-\eta) \max_{a'} \left\{V(\varepsilon, a', \chi) - V(\varepsilon, a, \chi) - (a' - a) P\right\}. \end{aligned}$$

Rearrangement implies,

$$[r+\chi_1+\chi_2(1-\eta)]\,V(\varepsilon,a,\chi)=\varepsilon u(a)+\chi_1\int_{\varepsilon_l}^{\varepsilon_h}V(\varepsilon',a,\chi)\,dF(\varepsilon')$$
$$+\chi_2(1-\eta)\max_{a'}\left\{V(\varepsilon,a',\chi)-(a'-a)P\right\} \quad (3.8)$$

or

$$V(\varepsilon,a,\chi)=\frac{r\frac{\varepsilon u(a)}{r}+\chi_1\int_{\varepsilon_l}^{\varepsilon_h}V(\varepsilon',a,\chi)\,dF(\varepsilon')+\chi_2(1-\eta)\max_{a'}\left\{V(\varepsilon,a',\chi)-(a'-a)P\right\}}{r+\chi_1+\chi_2(1-\eta)}.$$

The auxiliary HJB equation (3.8) shows that a client's current continuation value is equal to the weighted average of three values: the value of holding a units of the asset forever while keeping the current taste type ε, the value of keeping a units forever with a randomly drawn taste type from the cdf F, and the value of holding the optimal amount by trading at the interdealer price P while keeping the taste type ε. The first one affects the continuation utility because the client is impatient ($r>0$), and so, its weight is $r/(r+\chi_1+\chi_2(1-\eta))$. The second one affects the continuation value because the client receives taste shocks at the Poisson rate χ_1, and so, its weight is $\chi_1/(r+\chi_1+\chi_2(1-\eta))$. The last one affects the continuation value because the client gets to trade at the Possion rate χ_2. The weight of this term is, however, weighted down by $1-\eta$ because the intermediating dealer will capture a share, η, of the trade surplus. Thus, the weight of the last term turns out to reflect a bargaining-adjusted Poisson rate $\chi_2(1-\eta)$: $\chi_2(1-\eta)/(r+\chi_1+\chi_2(1-\eta))$.

The first key step in the characterization of the equilibrium is to solve the auxiliary HJB equation (3.8) given P. One challenge in establishing the existence and uniqueness of the solution to (3.8) is that it does not define a contraction mapping because we allow χ_1 and χ_2 to be arbitrarily large, $\chi\in[0,\infty)^2$. This eliminates the possibility of a uniform contraction modulus. To overcome this challenge, we exploit local contraction techniques developed by Martins-da Rocha and Vailakis (2010) to show the existence and uniqueness of the solution to (3.8). Then, we follow a method of undetermined coefficients to obtain this unique solution in closed form.

Proposition 1 *Given P, the unique solution to the auxiliary HJB equation (3.8) has the following functional form:*

$$V(\varepsilon,a,\chi)=\frac{\chi_2(1-\eta)}{r+\chi_2(1-\eta)}Pa+\frac{1}{r+\chi_2(1-\eta)}\frac{(r+\chi_2(1-\eta))\,\varepsilon+\chi_1\bar{\varepsilon}}{r+\chi_1+\chi_2(1-\eta)}u(a)+\kappa_0(\varepsilon,\chi), \quad (3.9)$$

where

$$\bar{\varepsilon} \equiv \int_{\varepsilon_l}^{\varepsilon_h} \varepsilon dF(\varepsilon).$$

and for some $\kappa_0 : [\varepsilon_l, \varepsilon_h] \times \mathbb{R}_+^2 \to \mathbb{R}$.

Now, one can easily determine the terms of trade between a client and a dealer using (3.1) and (3.2).

The FOC of the joint surplus maximization (3.1) is

$$V_2(\varepsilon, a', \chi) = P,$$

where $V_2(\cdot,\cdot,\cdot)$ refers to the derivative with respect to the second argument. Using (3.9) and after rearranging, one obtains that the optimal asset position of a client with the taste type of ε and characteristics χ is

$$a^\star(\varepsilon, \chi) = (u')^{-1}\left[\frac{r + \chi_1 + \chi_2(1-\eta)}{(r + \chi_2(1-\eta))\varepsilon + \chi_1\bar{\varepsilon}} rP\right], \tag{3.10}$$

i.e., a client with these characteristics will end up holding $a^\star(\varepsilon, \chi)$ units of the asset after meeting a dealer regardless of her initial asset position a. Thus, the bilateral trade quantity between this client and the dealer is

$$q(\varepsilon, a, \chi) = (u')^{-1}\left[\frac{r + \chi_1 + \chi_2(1-\eta)}{(r + \chi_2(1-\eta))\varepsilon + \chi_1\bar{\varepsilon}} rP\right] - a. \tag{3.11}$$

Given P, (3.9) and (3.11), (3.2) gives the negotiated price, $p(\varepsilon, a, \chi)$, between the given client and dealer.

Equipped with (3.10) and (3.11), we derive the stationary equilibrium distribution in closed form in Proposition 2.

Proposition 2 *Let* $a^\star(\cdot,\cdot)$ *be given by* (3.10)*, and let the function* $\tilde{\varepsilon}_\chi : [-M, M] \to supp\,(dF)$ *be defined such that* $\tilde{\varepsilon}_\chi(a) = \sup\{\varepsilon \in supp\,(dF) : a \geq a^\star(\varepsilon, \chi)\}$*. Then, for any* $\chi \in supp\,(dG)$*,*

$$\Phi_\chi(\varepsilon, a) = \frac{\chi_1 F(\varepsilon) F(\tilde{\varepsilon}_\chi(a)) + \chi_2 F(\min\{\varepsilon, \tilde{\varepsilon}_\chi(a)\})}{\chi_1 + \chi_2}. \tag{3.12}$$

As part of the proof of Proposition 2, we show in the Appendix that the fraction (or density) of type-χ clients whose current asset position is $a^\star(\varepsilon, \chi)$ is equal to the fraction (or density) of type-χ clients who have a current taste type of ε, which is $dF(\varepsilon)$. This allows us to re-write (3.7) as

$$\int_0^\infty \int_0^\infty \int_{\varepsilon_l}^{\varepsilon_h} a^\star(\varepsilon, \chi)\, dF(\varepsilon)\, G(d\chi_1, d\chi_2) = A.$$

Using (3.10), this equation allows us to determine the interdealer market price P, which is the only remaining equilibrium object that is yet to be determined.

Proposition 3 *Assume $u'(\infty) = 0$ and $u'(0) = \infty$. Then, the unique equilibrium interdealer market price P solves*

$$\int_0^\infty\int_0^\infty\int_{\varepsilon_l}^{\varepsilon_h} (u')^{-1}\left[\frac{r+\chi_1+\chi_2(1-\eta)}{(r+\chi_2(1-\eta))\varepsilon+\chi_1\bar{\varepsilon}} rP\right] dF(\varepsilon)\, G(d\chi_1, d\chi_2) = A. \tag{3.13}$$

Taking stock, the equilibrium defined in Definition 1 is characterized by Propositions 1–3 in closed form up to the interdealer market price P. With some additional parametric assumptions, P can be obtained in closed form as well. For example, as in the original LR model, P is available in closed form when $u(a) = \log(a)$. In the next subsection, we assume an iso-elastic utility function, which nests the log utility as a special case, and a binary taste-type structure to obtain explicit formulas for endogenous liquidity measures such as bid-ask spread, trade volume, and price dispersion.

3.4 A Special Case

We let $u(a) = \frac{a^{1-\gamma}}{1-\gamma}$. We assume that clients' taste types take on binary values $\varepsilon \in \{0, 2\sigma\}$ for $\sigma > 0$. This setup provides an important simplification as there will be a one-to-one mapping between client characteristics χ and unsigned trade sizes due to binary taste types. That is, fixing characteristics χ, a high-type client currently stuck with the low-type optimal position will buy $q(\chi)$ units of the asset upon meeting a dealer; vice versa, a low-type client currently stuck with the high-type optimal position will sell $q(\chi)$ units, where

$$q(\chi) \equiv a^\star(2\sigma, \chi) - a^\star(0, \chi).$$

For further simplicity, we impose some symmetry conditions: the distribution F has two equal mass points at $\varepsilon = \mathbb{E}[\varepsilon] \pm \sigma$, which means that one natural interpretation for parameter σ is *preference volatility*. This also implies that upon receiving a taste shock a client preserves her current taste with probability $1/2$ or switches to the opposite taste type with probability $1/2$. In other words, a low-type (resp. high-type) client with characteristics $\chi = (\chi_1, \chi_2)$ switches to high type (resp. low type) at the Poisson rate of $\chi_1/2$.

Let us start by calculating the optimal holding of each client type in this special case. Using (3.10) and our assumptions on the utility function and the taste types,

$$a^\star(\mathbb{E}[\varepsilon] \pm \sigma, \chi) = \left[\left(1 \pm \frac{r+\chi_2(1-\eta)}{r+\chi_1+\chi_2(1-\eta)}\right)\frac{\sigma}{rP}\right]^{1/\gamma}. \tag{3.14}$$

Note that this optimal holding takes as given the interdealer market price P. Now we calculate the equilibrium value of P so that we can write all equilibrium objects in terms of exogenous parameters and distributions. With a slight abuse of notation "$dG(\chi') = G(d\chi'_1, d\chi'_2)$", Equation (3.13) can be re-written as

$$\frac{1}{2}\int\limits_{\mathbb{R}^2_+}\left[(1+\Sigma(\chi'))\frac{\sigma}{rP}\right]^{1/\gamma} dG(\chi') + \frac{1}{2}\int\limits_{\mathbb{R}^2_+}\left[(1-\Sigma(\chi'))\frac{\sigma}{rP}\right]^{1/\gamma} dG(\chi') = A,$$

where

$$\Sigma(\chi) \equiv \frac{r+\chi_2(1-\eta)}{r+\chi_1+\chi_2(1-\eta)}$$

is the endogenous demand rescaling coefficient that represents clients' individually optimal response to frictions. After rearranging,

$$P = \frac{\sigma}{r}\left\{\frac{1}{2A}\int\limits_{\mathbb{R}^2_+}\left[(1+\Sigma(\chi'))^{1/\gamma} + (1-\Sigma(\chi'))^{1/\gamma}\right] dG(\chi')\right\}^{\gamma}. \tag{3.15}$$

Substituting (3.15) into (3.14), the equilibrium optimal holding of each client type is

$$a^\star(\mathbb{E}[\varepsilon] \pm \sigma, \chi) = A\frac{(1 \pm \Sigma(\chi))^{1/\gamma}}{\frac{1}{2}\int\limits_{\mathbb{R}^2_+}\left[(1+\Sigma(\chi'))^{1/\gamma} + (1-\Sigma(\chi'))^{1/\gamma}\right] dG(\chi')}, \tag{3.16}$$

and, in turn, the equilibrium trade quantity of client χ is

$$q(\chi) = A\frac{(1+\Sigma(\chi))^{1/\gamma} - (1-\Sigma(\chi))^{1/\gamma}}{\frac{1}{2}\int\limits_{\mathbb{R}^2_+}\left[(1+\Sigma(\chi'))^{1/\gamma} + (1-\Sigma(\chi'))^{1/\gamma}\right] dG(\chi')}. \tag{3.17}$$

In the context of search and matching models, exposure to frictions is conveniently represented by a Poisson intensity, χ_2. It stands in for many unmodeled features like the level of connectivity to the interdealer market, search ability, or ability to quickly close a deal with dealers perhaps thanks to repeat relations. The equilibrium trade quantity (3.17) reveals the standard relationship between trade aggressiveness and exposure to frictions. Controlling for χ_1, a more active trader (i.e., with larger χ_2) trades in larger quantities because she is less afraid of being stuck with a suboptimal position following a taste shock. This is represented by a larger $\Sigma(\chi)$ as χ_2 increases. On the other hand, controlling for χ_2, a larger χ_1 leads to lower $\Sigma(\chi)$ and lower trade quantities, because a large-χ_1 trader switches her taste type very frequently and trades in a way to hedge herself against the risk of being stuck with a suboptimal position following a taste shock. The fact that χ_1 and χ_2 have opposite effect on trade sizes makes trade size information very useful to simultaneously identify χ_1 and χ_2 from the data. Other endogenous objects such as trading intensity are typically increasing in both χ_1 and χ_2. Therefore, adding trade-size-related information to the list of data moments brings some unique information and helps us identify the deep parameters of our model.

Next, we determine the equilibrium transaction prices in our special case. Let us start by defining $\alpha(\chi) \equiv p(2\sigma, a^\star(0,\chi), \chi)$ and $\beta(\chi) \equiv p(0, a^\star(2\sigma,\chi), \chi)$ as the ask and the bid price between a client of type χ and a dealer. Note that these two are the only realized transaction prices for this client in our special case: The client pays $\alpha(\chi)$ per unit to the dealer when she buys $q(\chi)$ units from him, and vice versa, she receives $\beta(\chi)$ per unit from the dealer when she sells $q(\chi)$ units to him. We calculate these prices using (3.2):

$$\alpha(\chi) = \frac{(1-\eta)\, q(\chi)\, P + \eta \left\{V(2\sigma, a^\star(2\sigma,\chi), \chi) - V(2\sigma, a^\star(0,\chi), \chi)\right\}}{q(\chi)}$$

and

$$\beta(\chi) = \frac{(1-\eta)\, q(\chi)\, P + \eta \left\{V(0, a^\star(2\sigma,\chi), \chi) - V(0, a^\star(0,\chi), \chi)\right\}}{q(\chi)}.$$

These formulas for bid and ask prices are already intuitive in their current form. When the client has all the bargaining power ($\eta = 0$), both the bid and the ask price are equal to P, which means that the bid-ask spread is zero and the dealer's transaction surplus is zero. When the dealer has all the bargaining power ($\eta = 1$), bid and ask prices are equal to the client's "reservation prices," which leaves the client with zero transaction surplus and maximizes the bid-ask spread that goes to the dealer.

As an intermediate step, note that Equation (3.9) implies

$$\begin{aligned} V(\varepsilon, a^\star(2\sigma,\chi), \chi) - V(\varepsilon, a^\star(0,\chi), \chi) = {} & \frac{\chi_2(1-\eta)}{r + \chi_2(1-\eta)} q(\chi)\, P \\ & + \frac{1}{r+\chi_2(1-\eta)} \frac{(r+\chi_2(1-\eta))\,\varepsilon + \chi_1 \sigma}{r + \chi_1 + \chi_2(1-\eta)} \Delta_u(\chi), \end{aligned}$$

where

$$\begin{aligned} \Delta_u(\chi) & \equiv \frac{\left[a^\star(2\sigma,\chi)\right]^{1-\gamma}}{1-\gamma} - \frac{\left[a^\star(0,\chi)\right]^{1-\gamma}}{1-\gamma} \\ & = \frac{1}{1-\gamma} \left(\frac{A}{\frac{1}{2} \int_{\mathbb{R}_+^2} \left[(1+\Sigma(\chi'))^{1/\gamma} + (1-\Sigma(\chi'))^{1/\gamma}\right] dG(\chi')} \right)^{1-\gamma} \left[(1+\Sigma(\chi))^{\frac{1}{\gamma}-1} - (1-\Sigma(\chi))^{\frac{1}{\gamma}-1}\right]. \end{aligned}$$

Thus,

$$\alpha(\chi) = (1-\eta)\, P + \frac{\eta}{r + \chi_2(1-\eta)} \left(\chi_2 (1-\eta)\, P + (1+\Sigma(\chi))\, \sigma \frac{\Delta_u(\chi)}{q(\chi)} \right)$$

and

$$\beta\left(\chi\right)=\left(1-\eta\right)P+\frac{\eta}{r+\chi_{2}\left(1-\eta\right)}\left(\chi_{2}\left(1-\eta\right)P+\left(1-\Sigma\left(\chi\right)\right)\sigma\frac{\Delta_{u}\left(\chi\right)}{q\left(\chi\right)}\right).$$

3.4.1 Digression: Bid-Ask Spreads

Define

$$s\left(\chi\right)\equiv\alpha\left(\chi\right)-\beta\left(\chi\right).$$

From the formulas above,

$$s\left(\chi\right)=\eta\frac{2\sigma}{r+\chi_{1}+\chi_{2}\left(1-\eta\right)}\frac{\Delta_{u}\left(\chi\right)}{q\left(\chi\right)}.$$

Assuming $\gamma=1$ (log utility) and using the equilibrium trade quantity (3.17),

$$s\left(\chi\right)=\frac{\eta\sigma}{A}\frac{\Delta_{u}\left(\chi\right)}{r+\chi_{2}\left(1-\eta\right)}. \tag{3.18}$$

To analyze the qualitative properties of $s\left(\chi\right)$, we use the lemma below.

Lemma 1 *Assume $\gamma=1$. Then, $\Delta_{u}\left(\chi\right)/\left[r+\chi_{2}\left(1-\eta\right)\right]$ is strictly decreasing in both χ_1 and χ_2.*

Lemma 1 has an unambiguous prediction for the relationship between how frequently a client trades and the bid-ask spread she faces. A client trades frequently either because she is a liquidity trader (high χ_1) or an active trader (high χ_2). Either way, the second factor of the bid-ask spread (3.18) is lower for a client who trades more frequently compared to that of a client who trades less frequently, as stated by Lemma 1. First, $\Delta_u\left(\chi\right)$ is decreasing in χ_1, and so is the second factor of $s\left(\chi\right)$. Second, it is decreasing in χ_2, because while both its numerator and its denominator increase with χ_2, the numerator is concave but the denominator is linear. Thus, the effect coming from the denominator is stronger and makes the second factor of $s\left(\chi\right)$ a decreasing function of χ_2 as well. Taking stock, a client who trades frequently faces a smaller bid-ask spread. This result is consistent with the findings of O'Hara, Wang, and Zhou (2018) on the US corporate bond market and our findings in Appendix C on the UK government bond and corporate bond markets.

3.4.2 Going Back to Characterization: Equilibrium Distribution

So far, we have derived the equilibrium inter-dealer market price, value functions, bilateral dealer-client trade quantities and prices. What remains is to derive the equilibrium distribution, and then, we can derive price- and quantity-related moment formulae, allowing us to connect the model to the data.

The result (3.10) that there is a unique optimal holding for any client (ε,χ) implies the following support for the equilibrium asset holding distribution for clients with characteristics χ:

$$\mathcal{A}\left(\chi\right)=\left\{a^{\star}\left(\varepsilon,\chi\right):\varepsilon\in\operatorname{supp}\left(dF\right)\right\}.$$

Indeed, any client (ε, χ) either holds her optimal position $a^\star(\varepsilon, \chi)$ or is still stuck with a previously optimal position $a^\star(\varepsilon', \chi)$ for some $\varepsilon' \neq \varepsilon$ because she is yet to trade with a dealer after her latest taste shock. For our special case with the binary taste types, this set contains two elements:

$$\mathcal{A}(\chi) = \{a^\star(0, \chi), a^\star(2\sigma, \chi)\}.$$

As a result, the equilibrium distribution admits a simple representation as a conditional probability mass function (pmf). Let $\phi_\chi(\varepsilon, a)$ be the equilibrium joint pmf of taste types and asset positions conditional on client characteristics. Equation (3.12) implies for our special case that

$$\phi_\chi(\varepsilon, a) = \begin{cases} 0 & \text{if } a \notin \mathcal{A}(\chi) \\ \frac{1}{4}\frac{\chi_1+2\chi_2}{\chi_1+\chi_2} & \text{if } a = a^\star(\varepsilon, \chi) \\ \frac{1}{4}\frac{\chi_1}{\chi_1+\chi_2} & \text{if } a = a^\star(\varepsilon', \chi) \text{ for } \varepsilon' \neq \varepsilon. \end{cases} \tag{3.19}$$

Equation (3.19) effectively shows the mass of clients who are happy with their current holding and the mass of those who are unhappy with their current holding. Considering that only the latter type will trade in equilibrium, the rate at which the χ-clients trade with dealers is

$$\chi_2 \frac{1}{4}\frac{\chi_1}{\chi_1+\chi_2} dG(\chi) + \chi_2 \frac{1}{4}\frac{\chi_1}{\chi_1+\chi_2} dG(\chi) = \chi_2 \frac{1}{2}\frac{\chi_1}{\chi_1+\chi_2} dG(\chi),$$

where χ_2 is the per-client meeting rate, $dG(\chi)$ is the mass of those clients, and $\frac{1}{2}\frac{\chi_1}{\chi_1+\chi_2}$ is the fraction of χ-clients who are unhappy with their current holding: Half of them are of low taste type but holding a high position and the other half are of high taste type but holding a low position. Thus, although a client with characteristics $\chi = (\chi_1, \chi_2)$ meets a dealer at the exogenous Poisson intensity of χ_2, her endogenous trading intensity is

$$\theta(\chi) \equiv \frac{\chi_2}{2}\frac{\chi_1}{\chi_1+\chi_2}. \tag{3.20}$$

Price dispersion In our model and in the data, different dealer-client pairs trade at different prices. The magnitude of this deviation from the law of one price is quantified by price dispersion. We calculate the equilibrium price dispersion as the trade size-weighted mean absolute deviation of dealer-client transaction prices from the inter-dealer market price normalized by the inter-dealer market price:

$$\sigma_p = \frac{\int_{\mathbb{R}_+^2}\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h} \theta(\chi)\, q(\chi)\, |p(\varepsilon, a, \chi) - P|\, \Phi_\chi(d\varepsilon, da)\, dG(\chi)}{P\int_{\mathbb{R}_+^2}\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h} \theta(\chi)\, q(\chi)\, \Phi_\chi(d\varepsilon, da)\, dG(\chi)}.$$

And, using the fact that half of the transactions happen at client-specific bid prices and the other half at client-specific ask prices in our special case,

$$\sigma_p = \frac{\int\limits_{\mathbb{R}_+^2} \theta(\chi)\, q(\chi)\, \frac{\alpha(\chi)-\beta(\chi)}{2} dG(\chi)}{P \int\limits_{\mathbb{R}_+^2} \theta(\chi)\, q(\chi)\, dG(\chi)}.$$

Hence, in our special case, the equilibrium price dispersion is equal to the half of (a weighted average of) the realized bid-ask spreads.

3.4.3 Welfare

One of our main motivations to write this model of OTC market is to quantify the welfare loss caused by the frictions characteristic of OTC markets. Accordingly, we calculate various measures of social welfare in our model environment. Naturally, the first one is social welfare evaluated at the equilibrium allocation:

$$\mathbb{W}^{Eq} \equiv \frac{1}{r} \int\limits_{\mathbb{R}_+^2} \int\limits_{-M}^{M} \int\limits_{\varepsilon_l}^{\varepsilon_h} \varepsilon u(a)\, \Phi_\chi(d\varepsilon, da)\, dG(\chi) - \frac{1}{r} \int\limits_{\varepsilon_l}^{\varepsilon_h} \varepsilon u(A)\, dF(\varepsilon). \tag{3.21}$$

Let us highlight the key properties of our model environment that make $\mathbb{W}^{Eq}$ a sensible measure of social welfare. The first term represents the present value of all utility benefits stemming from clients' asset holdings. Any transfer of numéraire between clients and dealers net out to zero thanks to quasi-linear and transferable utility, and so, the welfare is generated by clients' utility flows only. In our model, clients' utility flows change over time due to exogenous time variation in their ε and endogenous time variation in their a. However, the distribution of utility flows across clients stays the same in the stationary equilibrium. Hence, the present value calculation reduces to dividing the aggregate utility flow by the discount rate r.[14] Because we want $\mathbb{W}^{Eq}$ to capture the welfare created by OTC trading opportunities, we subtract a baseline level of welfare from the first term of $\mathbb{W}^{Eq}$. Our choice of baseline welfare is the level of welfare in an "autarky" allocation in which every client's holding is always equal to the per-capita supply, A, of the asset and none of the clients trades as they switch from one taste type to another. That is, our baseline welfare measures the level of welfare obtained when clients forego all the gains from trade.

The second welfare measure we calculate is the unconstrained efficient or *first-best* welfare, i.e., the level of welfare when a benevolent social planner decides the allocation of assets across agents

[14] To be more precise, we make use of the equality $\frac{1}{r} = \int\limits_0^\infty e^{-rt} dt$ for $r > 0$.

without any constraint apart from the usual resource constraints:

$$\mathbb{W}^{FB} \equiv \max_{\{a(\varepsilon)\}_{\varepsilon\in\mathrm{supp}(dF)}} \frac{1}{r}\int_{\varepsilon_l}^{\varepsilon_h} \varepsilon u\left[a\left(\varepsilon\right)\right] dF\left(\varepsilon\right) - \frac{1}{r}\int_{\varepsilon_l}^{\varepsilon_h} \varepsilon u\left(A\right) dF\left(\varepsilon\right),$$

subject to

$$\int_{\varepsilon_l}^{\varepsilon_h} a\left(\varepsilon\right) dF\left(\varepsilon\right) = A$$

and

$$-M \leq a\left(\varepsilon\right) \leq M$$

for all $\varepsilon \in \mathrm{supp}\left(dF\right)$. Again, in writing down this welfare measure, we use the fact that how numéraire is allocated across agents is irrelevant to social welfare due to transferable utility and the fact that the distribution of taste types across clients is stationary.

The last welfare measure we calculate is the constrained efficient or *second-best* welfare, i.e., the level of welfare when a benevolent social planner can modify only the terms of trade when agents get to trade, but otherwise is subject to the constraints stemming from the OTC market structure as well as the usual resource constraints:

$$\mathbb{W}^{SB} \equiv \int_0^\infty e^{-rt}\left\{\int_{\mathbb{R}_+^2}\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h} \varepsilon u\left(a\right) \Phi_\chi^*\left(d\varepsilon, da \,|\, t\right) dG\left(\chi\right)\right\} dt - \frac{1}{r}\int_{\varepsilon_l}^{\varepsilon_h} \varepsilon u\left(A\right) dF\left(\varepsilon\right). \tag{3.22}$$

The planner maximizes $\mathbb{W}^{SB}$ with respect to controls, $q\left(\varepsilon, a, \chi \,|\, t\right)$, subject to the laws of motion for the state variables, $\Phi_\chi^*\left(d\varepsilon, da \,|\, t\right)$, and to the feasibility condition of asset reallocation,

$$\int_{\mathbb{R}_+^2}\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h} \chi_2 q\left(\varepsilon, a, \chi \,|\, t\right) \Phi_\chi^*\left(d\varepsilon, da \,|\, t\right) dG\left(\chi\right) = 0. \tag{3.23}$$

Note that using prices as control variable is redundant for the usual reason that any transfer of the numéraire good from one agent to another does not affect $\mathbb{W}^{SB}$ because of quasi-linear preferences.

The next proposition presents these three welfare notions as functions of model primitives.

Proposition 4 *In our special case with iso-elastic utility and binary taste types, the values of social welfare evaluated at the first-best, the second-best, and the equilibrium allocations are*

$$\mathbb{W}^{FB} = \frac{\sigma}{r}\frac{A^{1-\gamma}}{1-\gamma}\left(2^{1-\gamma} - 1\right),$$

$$\mathbb{W}^{SB}=\frac{\sigma}{r}\frac{A^{1-\gamma}}{1-\gamma}\left[\frac{\int_{\mathbb{R}_+^2}\left[\left(\frac{2r+\chi_1+2\chi_2}{r+\chi_1+\chi_2}\right)^{\frac{1}{\gamma}-1}\frac{\chi_2+\chi_1/2}{\chi_2+\chi_1}+\left(\frac{\chi_1}{r+\chi_1+\chi_2}\right)^{\frac{1}{\gamma}-1}\frac{\chi_1/2}{\chi_2+\chi_1}\right]dG\left(\chi\right)}{\left\{\frac{1}{2}\int_{\mathbb{R}_+^2}\left[\left(\frac{2r+\chi_1+2\chi_2}{r+\chi_1+\chi_2}\right)^{1/\gamma}+\left(\frac{\chi_1}{r+\chi_1+\chi_2}\right)^{1/\gamma}\right]dG\left(\chi\right)\right\}^{1-\gamma}}-1\right],$$

and

$$\mathbb{W}^{Eq}=\frac{\sigma}{r}\frac{A^{1-\gamma}}{1-\gamma}\left[\frac{\int_{\mathbb{R}_+^2}\left[\left(\frac{2r+\chi_1+2\chi_2(1-\eta)}{r+\chi_1+\chi_2(1-\eta)}\right)^{\frac{1}{\gamma}-1}\frac{\chi_2+\chi_1/2}{\chi_2+\chi_1}+\left(\frac{\chi_1}{r+\chi_1+\chi_2(1-\eta)}\right)^{\frac{1}{\gamma}-1}\frac{\chi_1/2}{\chi_2+\chi_1}\right]dG\left(\chi\right)}{\left\{\frac{1}{2}\int_{\mathbb{R}_+^2}\left[\left(\frac{2r+\chi_1+2\chi_2(1-\eta)}{r+\chi_1+\chi_2(1-\eta)}\right)^{1/\gamma}+\left(\frac{\chi_1}{r+\chi_1+\chi_2(1-\eta)}\right)^{1/\gamma}\right]dG\left(\chi\right)\right\}^{1-\gamma}}-1\right],$$

respectively.

One lesson from Proposition 4 is that the efficiency implications of LR apply to our generalized setup as well. That is, the second-best welfare would obtain in equilibrium if dealers' share of transaction surplus were zero. As in the model of LR, clients in our model reduce their trade quantities by inefficiently large amount because they cannot internalize all the gains from trade due to $\eta > 0$. This means that while the source of inefficiency in this model is search frictions ($\chi_2 < \infty$), the source of *constrained* inefficiency is dealers' market power ($\eta > 0$).

In our quantitative analysis below, Proposition 4 plays a key role. After we estimate the model parameters for the gilt market and the UK corporate bond market, we use Proposition 4 to calculate $\mathbb{W}^{FB}$, $\mathbb{W}^{SB}$, and $\mathbb{W}^{Eq}$ for each market. This allows us to understand and compare the extent to which OTC market frictions affect the participants' well-being in two of Europe's largest fixed-income markets. In particular, we calculate $\left(\mathbb{W}^{FB}-\mathbb{W}^{Eq}\right)/\mathbb{W}^{FB}$ to quantify the welfare loss from the real-world frictions relative to what could obtain in a perfect world. Then, we decompose this relative welfare loss to a component due to dealers' market power, $\left(\mathbb{W}^{SB}-\mathbb{W}^{Eq}\right)/\mathbb{W}^{FB}$, and a component due purely to search frictions, $\left(\mathbb{W}^{FB}-\mathbb{W}^{SB}\right)/\mathbb{W}^{FB}$.

4 Estimating the Model

4.1 Bringing the Model to the Data

In what follows, we estimate the special case of our model with iso-elastic utility and binary taste types (presented in Section 3.4) for the gilt market and the UK corporate bond market. We have to determine five parameters, r, σ, γ, A, and η, and one distribution, $G\left(\chi_1,\chi_2\right)$, for each market.

4.1.1 Summary of Model Equations for Estimation

The equations for the observable empirical objects to be used in the estimation are:

$$\overline{\theta} = \int_{\mathbb{R}_+^2} \theta(\chi)\, dG(\chi), \tag{4.1}$$

$$\sigma_\theta = \int_{\mathbb{R}_+^2} \left|\theta(\chi) - \overline{\theta}\right| dG(\chi), \tag{4.2}$$

$$\log\frac{\bar{q}}{A} = \log\int_{\mathbb{R}_+^2} \frac{q(\chi)}{A} dG(\chi) = \log\int_{\mathbb{R}_+^2}\left(\frac{(1+\Sigma(\chi))^{1/\gamma} - (1-\Sigma(\chi))^{1/\gamma}}{\frac{1}{2}\int_{\mathbb{R}_+^2}\left[(1+\Sigma(\chi'))^{1/\gamma} + (1-\Sigma(\chi'))^{1/\gamma}\right] dG(\chi')}\right) dG(\chi) \tag{4.3}$$

$$\begin{aligned}\log\mathcal{T} &= \log\int_{\mathbb{R}_+^2}\left(\frac{\theta(\chi)\, q(\chi)}{A}\right) dG(\chi) \\ &= \log\int_{\mathbb{R}_+^2}\left(\theta(\chi)\frac{(1+\Sigma(\chi))^{1/\gamma} - (1-\Sigma(\chi))^{1/\gamma}}{\frac{1}{2}\int_{\mathbb{R}_+^2}\left[(1+\Sigma(\chi'))^{1/\gamma} + (1-\Sigma(\chi'))^{1/\gamma}\right] dG(\chi')}\right) dG(\chi),\end{aligned} \tag{4.4}$$

$$\sigma_p = \frac{\int_{\mathbb{R}_+^2} \theta(\chi)\, q(\chi)\, \frac{\alpha(\chi)-\beta(\chi)}{2} dG(\chi)}{P\int_{\mathbb{R}_+^2} \theta(\chi)\, q(\chi)\, dG(\chi)}, \tag{4.5}$$

Going over the intensity-based, quantity-based, or price-based moments listed above, the entire distribution G is necessary to calculate any endogenous measure. Therefore, we try to determine it entirely. In doing so, we impose further parametric assumptions. First, we parameterize clients' characteristics, $\chi = (\chi_1, \chi_2)$, by assuming a one-to-one relationship between χ_1 and χ_2:

$$\chi_2 = \lambda\log(1+\chi_1), \tag{4.6}$$

where the parameter $\lambda > 0$ will be estimated. Intuitively, this assumption imposes that clients, who have a higher exposure to taste shocks, are likely to receive more trading opportunities as well. This positive relationship is micro-founded in the literature, for example, by Vayanos and Wang (2007) with endogenous market segmentation and by Hendershott, Li, Livdan, and Schürhoff (2020) with endogenous client-dealer relationships.[15] Here, we take this positive association as

[15] Vayanos and Wang (2007) let agents with different taste shock intensity choose the segment of the market in which to trade. They show that an asymmetric equilibrium exists in which high-taste-shock-intensity agents trade in a liquid segment and low-taste-shock-intensity agents trade in the illiquid segment. Hendershott, Li, Livdan, and Schürhoff (2020) allow clients with different taste shock intensity to choose the number of dealers in their relationship set. The optimal number of dealers is increasing in taste shock intensity, which allows high-taste-shock-intensity clients to trade more frequently than low-taste-shock-intensity clients.

given and let the data tell us about its strength (λ). Equation (4.6) effectively reduces the two-dimensional client heterogeneity by one dimension, which simplifies the estimation of the model.

With this simplification, it suffices to pin down the marginal distribution of χ_1, $G(\chi_1)$. Our second parametric assumption is related to this. We assume an exponential distribution, $G(\chi_1) \sim Exp(\delta)$. Then, to estimate four deep parameters, δ, λ, η, and γ, we use information regarding five different data moments: (i) the mean (4.1) and (ii) the mean absolute deviation (4.2) of clients' trade intensities, (iii) the natural logarithm of average trade size normalized by the per-capita asset supply (4.3), (iv) the natural logarithm of bond turnover, $\log \mathcal{T}$, calculated as above in (4.4), and (v) price dispersion scaled by the inter-dealer price (4.5).

The remaining parameters are the preference parameters and the per-capita asset supply: r, σ, and A. We set r equal to 5% per annum as is common.[16] Since we use an iso-elastic utility function, the only role A plays is to scale up and down the equilibrium quantities and prices. Thus, we normalize A to be 1.[17] What remains to be determined is the preference volatility parameter, σ. It is easy to see that the only role of σ is to scale up and down all the price levels, $\alpha(\chi)$, $\beta(\chi)$, and P. Hence, it does not affect our *normalized* price dispersion measure (4.4), and so, it does not affect any of the endogenous moments we match in our estimation procedures. Similarly, the welfare measures stated in Proposition 4 are only scaled up or down by σ, and so, the relative welfare measures we are interested in are not affected by σ either. Therefore, we leave the parameter σ out of our estimation.

4.1.2 Estimation

A unique feature of the dataset we use is that it contains the identities of both counterparties for each trade. This allows us to design an empirical framework that compares the market structures, and not the different pool of participants or assets the markets contain. That is to say, the subsample we use for estimation contains trades of the same clients both in the gilt and the corporate bond markets. In addition, it contains trades of a subset of corporate bonds that are similar to gilts in terms of payoff and adverse selection risks. Therefore, allowing for different utility curvatures (γ) for the gilt and the corporate bond markets would give our model an unfair advantage in matching the data moments at best, and would lead to identification problems at worst. For this reason, we assume a common utility curvature parameter for the two markets.[18]

We use generalized methods of moments (GMM) to estimate the model parameters for the gilt market and the corporate bond market. We have eight moment conditions to estimate the parameter vector $\psi = (\gamma, \delta_g, \lambda_g, \eta_g, \delta_c, \lambda_c, \eta_c)$, which consists of seven parameters. Given that the

[16]See Duffie, Gârleanu, and Pedersen (2007), Feldhütter (2012), Lester, Rocheteau, and Weill (2015), and Hugonnier, Lester, and Weill (2020) for example.

[17]See Lester, Rocheteau, and Weill (2015).

[18]We thank our discussant Liang Ma for this suggestion.

structural model is over-identified, we use GMM for the estimation:

$$\min_{\psi \in \Psi} \left[\left(\hat{m}\left(\psi\right) - m_S \right) \oslash m_S \right]' \hat{W} \left[\left(\hat{m}\left(\psi\right) - m_S \right) \oslash m_S \right], \tag{4.7}$$

where $\oslash$ is Hadamard division, $\hat{m}(\psi)$ is the vector of theoretical moments computed from the model evaluated at the parameter vector ψ, m_S is the vector of corresponding sample moments, and $\hat{W}$ is a weighting matrix. Note that we follow Gavazza (2016) in using moments in percentage deviation from their empirical targets to ensure that they have the same scale. In practice, we use simple differences for the logged moments (i.e., we do not calculate percentage deviations for the logged moments) because percentage deviations and log deviations are already of the same scale.

Six of our eight theoretical moments are average trade intensity, log bond turnover, and price dispersion given by (4.1), (4.4), and (4.5), respectively, calculated separately for the gilt and the corporate bond markets each. Our seventh moment is the ratio of intensity dispersion in the gilt market to intensity dispersion in the corporate bond market that we name *relative intensity dispersion* and calculate using (4.2). The Monte Carlo evidence from Section 4.5 implies that the sample intensity deviation calculated from any dataset is a biased estimator of the theoretical intensity dispersion implied by the parameters of the underlying data generating process. Therefore, we do not use the intensity dispersion levels from the two markets as separate moments. Nevertheless, the Monte Carlo evidence also reveals that relative intensity dispersion calculated using two datasets is an unbiased estimator of the theoretical relative intensity dispersion implied by the parameters of the respective data generating processes of the two markets. This allows us to use relative intensity dispersion as a moment for estimation. Finally, our eighth moment is the natural logarithm of average normalized trade size in the gilt market minus the natural logarithm of average normalized trade size in the corporate bond market, which we name *relative log trade size* and calculate using (4.3). Our parsimonious distributional assumptions for $G(\chi_1)$ in the two markets and the use of a common utility curvature parameter preclude the possibility of identifying both the level of log average trade size and the level of log bond turnover in the two markets at the same. Hence, we use the levels of log turnover as separate moments from the two markets, and only use relative log trade size as a way of further disciplining the parameter estimates for our cross-market analysis.

The sample moments, m_S, are computed at daily frequency, yielding 1614 observations for each moment condition. Given the relatively small sample, we follow Altonji and Segal (1996) by using equally weighted moments, i.e., we use an identity matrix for $\hat{W}$. By doing so, we avoid having to estimate the optimal weighting matrix that can have poor finite sample properties.[19]

[19] See Altonji and Segal (1996) as well as Honore, Jorgensen, and de Paula (2020) for a recent discussion. In Appendix D, we report the parameter estimates from a two-step GMM estimation, which are virtually identical to the results of our baseline one-step GMM estimation.

4.2 Empirical Results

The estimation is conducted jointly for the government bond market and the corporate bond market. The parameter estimates are shown in column 1 and column 2 of Table 2 along with bootstrap standard errors in parentheses for the government bond market and the corporate bond market, respectively. There are three notable differences in the estimates across the two markets. First, the average taste shock intensity in the government bond market ($1/\delta_g$) is about 5.5 times larger than the average taste shock intensity in the corporate bond market ($1/\delta_c$). That is, clients in the government bond market need to trade 5.5 times as often as they need to trade in the corporate bond market. Moreover, the elasticity of trading opportunities with respect to taste shock intensity (λ) as well as dealers' market power (η) are substantially higher in government bonds than in corporate bonds. This is primarily driven by the estimation algorithm trying to match the empirical moments for the clients' intensity distribution, log turnover, and price dispersion that are markedly different across the two bond markets.

Table 2: Parameter Estimates

	Government Bonds	Corporate Bonds
	(1)	(2)
γ – Curvature of the utility function	17.95	17.95
	(0.343)	(0.343)
δ – Exponential distn. parameter for $G(\chi_1)$	22.58	124.9
	(0.154)	(0.907)
λ – Matching efficiency	4123.83	232.65
	(570.264)	(26.095)
η – Dealers' market power	0.95	0.28
	(0.0038)	(0.0076)

Notes: This table reports the estimates of the parameters. The sample contains 1614 trading days covering the period 2011m8-2017m12. The parameter estimates are obtained by minimizing the objective function (4.7). Bootstrap standard errors, shown in parentheses, are based on 200 simulated datasets.

To illustrate the fit of the estimated model, Table 3 shows the moments computed from the data as well as those implied by the model evaluated at the estimated parameters. Overall, the model evaluated at the estimated parameters matches the data reasonably well in both markets. The mean absolute difference between the eight empirical moments and the corresponding theoretical moments is about 2.58%.

While Table 3 illustrates the fit of the targeted moments, Figure 1 plots the entire distribution of trading intensity in the cross section of clients in both markets, which are essentially un-targeted functions apart from their implied first two moments. Figure 1 shows that the distributions of clients' trading intensity in both markets are highly skewed both in the data and in our theory evaluated at the estimated parameters. This implies that our exogenous structural assumptions (i.e., exponential distribution for clients' taste shock intensity and a logarithmic relation between clients' taste shock intensity and meeting intensity with dealers) do a good job in generating

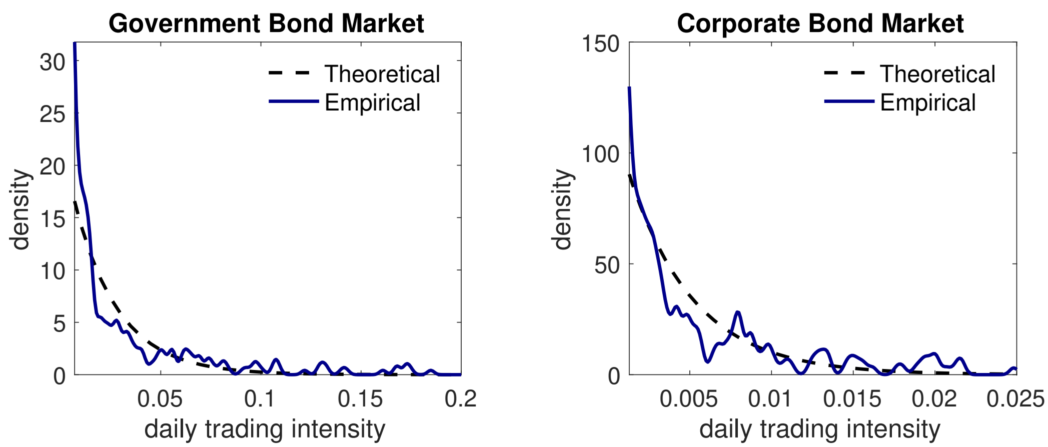
Table 3: Model Fit

Moments	Government Bonds Empirical (1)	Government Bonds Theoretical (2)	Corporate Bonds Empirical (3)	Corporate Bonds Theoretical (4)
Price Dispersion	0.00064	0.00064	0.0011	0.0011
Average Intensity	0.0232	0.0221	0.0038	0.0040
Log Turnover	-4.8038	-4.9235	-6.9094	-6.6619
Relative Intensity Dispersion	5.3069	5.5498	1	1
Relative Log Trade Size	0.0266	0.0266	0	0

Notes: This table reports the values of the empirical moments and of the theoretical moments calculated at the estimated parameters. Columns (1)-(2) and Columns (3)-(4) show the results for the government bond and corporate bond markets, respectively.

realistically skewed distributions of endogenous trade intensity in both markets.

Figure 1: Probability Density Functions of Clients' Daily Trading Intensity



Notes: This figure shows the theoretical and empirical probability density functions (pdf) of client-specific average daily trading intensity for the gilt market in the left panel and for the UK corporate bond market in the right. Theoretical pdfs are the pdf of the endogenous variable (3.20), implied by the estimated parameter values (as reported in Table 2). Empirical pdfs are one-dimensional kernel density estimates of the cross-sectional empirical variable which is the time-series average of each client's daily number of trades per bond.

4.3 Sources of Identification

Search-based models typically follow a general equilibrium approach. While individual agents take the equilibrium distribution as given in calculating the option value of continuing search, their individual actions generate the equilibrium distribution in question. As a result, a model structure emerges whereby (almost) all exogenous parameters affect all endogenous outcomes, as pointed out by Eckstein and van den Berg (2007) and Gavazza (2016). Although this is the case in our model as well, it is instructive to study how model parameters affect certain key moments to get a better understanding of the sources of identification. To that end, in Figure 2 and Figure 3, we inspect how the eight moment conditions change as we perturb a given parameter around its estimated value.

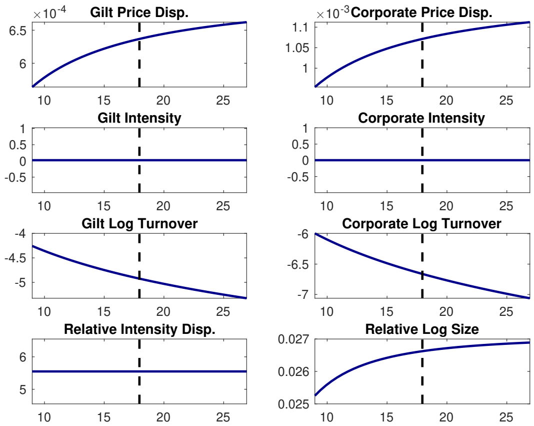
To begin with, Figure 2 shows how the theoretical moments change as we perturb each of the four gilt market parameters around their estimated values (denoted by the vertical line), while keeping all other parameters fixed at their estimated values. Panels 2a, 2b, 2c and 2d show the results for the curvature parameter (γ) for the utility function, the exponential distribution parameter (δ_g) for $G(\chi_1)$, matching efficiency (λ_g), and dealers' market power parameter (η_g), respectively.

Panel 2a shows that increasing the curvature parameter increases price dispersion and decreases turnover in both markets; it also increases average trade size in the gilt market relative to the corporate bond market, and leaves intensity-related variables unaffected. While all size- and price-related moments contribute to the identification of the curvature of the utility function, we place a particular emphasis on turnover because it is an unambiguously decreasing function of γ. As the curvature, γ, goes up, trade sizes shrink, as can be seen from (3.17), but trade intensity stays unaffected, leading to lower turnover. In turn, the estimation identifies 17.95 as the curvature level that matches the empirical log turnover in the gilt market.

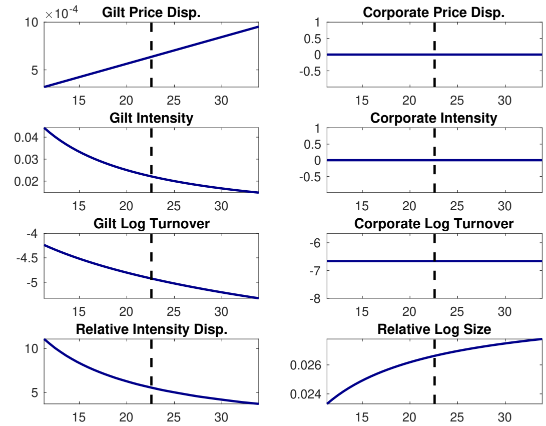
We view δ_g and λ_g as jointly identified by three data moments: the gilt average trade intensity, intensity dispersion in the gilt market relative to the corporate bond market, and average trade size in the gilt market relative to the corporate bond market. Panel 2c shows that an obvious effect of increasing λ_g is to raise the gilt average trade intensity and the intensity dispersion in the gilt market relative to the corporate bond market because of the increase in the meeting rates $\chi_2 = \lambda_g \log(1+\chi_1)$. Thus, the intensity-related moments can be considered the main identifier of λ_g, given the exponential distribution parameter, δ_g, for $G(\chi_1)$. However, the intensity-related moments have an unambiguous relationship with δ_g, too: increasing δ_g contributes to a reduction in the gilt average trade intensity and in the intensity dispersion in the gilt market relative to the corporate bond market, as can be seen in Panel 2b. This is because increasing δ_g shifts the whole distribution of the taste shock intensity inwards, thereby lowering its first and second moments as well as lowering the meeting rates $\chi_2 = \lambda_g \log(1+\chi_1)$, keeping λ_g and all other parameters constant. Therefore, if we only relied on intensity-related moments, there would be many combinations of δ_g and λ_g that generates the same level of intensity moments, leading to a joint identification issue. Fortunately, the relative trade size moment can help resolve this issue. Although δ_g and λ_g have opposite effect on the intensity moments, their effect on trade size is on the same direction: average trade size in the gilt market relative to the corporate bond market is an increasing function of both δ_g and λ_g. Both a reduced taste shock frequency (higher δ_g) and an increased frequency of meetings with dealers (higher λ_g) increase clients' aggressiveness, leading to larger gilt trade sizes. Hence, δ_g and λ_g can be jointly identified once intensity- and trade-size-related moments are both used.

Figure 2: Illustrating the Sources of Identification for the Gilt Market Parameters

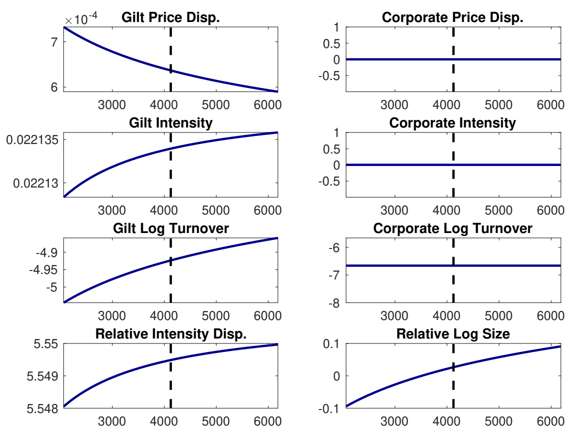
(a) Curvature of the Common Utility (γ)



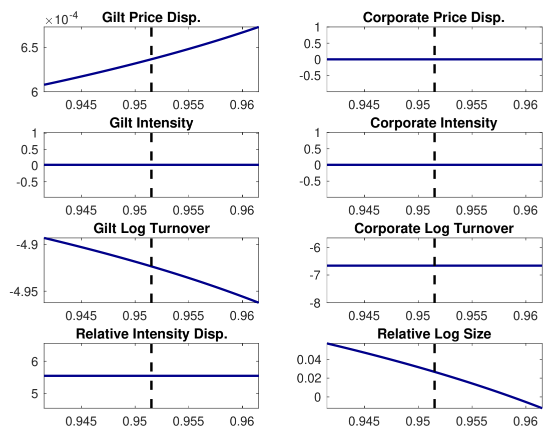
(b) Exponential Distn. Parameter for $G(\chi_1)$ (δ_g)



(c) Matching Efficiency (λ_g)



(d) Dealers' Market Power (η_g)



Notes: This figure shows how the theoretical moments change as we perturb one of the seven parameters around its estimated value (denoted by the vertical line), while keeping all the other parameters fixed at their estimated values. Panels 2a, 2b, 2c, and 2d show the results for the curvature parameter (γ), the exponential parameter for distribution $G(\chi_1)$, for λ_g, and for η_g, respectively.

Panel 2d shows that when the share, η_g, of transaction surplus captured by dealers rises, price dispersion rises, ceteris paribus. Accordingly, we view the price dispersion moment as the main identifier of η_g. Average trade size in the gilt market relative to the corporate bond market falls with η_g because clients pay higher rent to gilt dealers, which makes them cut down their demand (standard effect of avoiding price impact in the LR environment). Intensity-related moments are not impacted, as dealer bargaining power does not affect how often clients trade. This is a common result of models in which endogenous transaction costs are proportional to transaction surpluses such as LR and Üslü (2019).

Finally, Figure 3 shows how the theoretical moments change as we perturb each corporate bond market parameter around their estimated values. The economics behind the results of Figure 3 mirrors that of Figure 2.

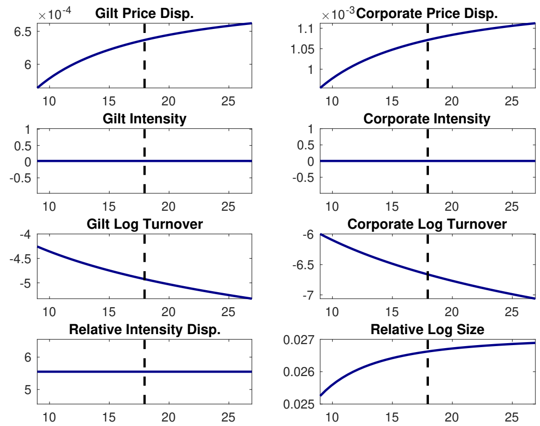
The discussion of the sources of identification above highlights the importance of using all three dimensions of market liquidity, trade frequency, trade size, and price, to have a full understanding of the drivers of market liquidity, which are the deep parameters of our model. For example, it would be impossible to distinguish between how often clients need to trade and how often they can trade by looking at their trade frequencies alone. However, our estimation can simultaneously identify δ_g and λ_g. The particular combination of $\delta_g = 22.58$ and $\lambda_g = 4123.83$ generates the gilt average trade intensity level of 0.0221. Panels 2b and 2c imply that one could increase δ_g and λ_g simultaneously and generate the same level of average intensity. In doing so, however, one would increase the gilt trade sizes, and so, would overshoot the empirical relative log trade size moment. Thus, given γ and η_g, only a unique combination of δ_g and λ_g can match the empirical trade intensity and the relative log trade size at the same time.

Similarly, it would be impossible to distinguish between search frictions and dealers' market power by looking at clients' trade frequencies alone or price dispersion alone. However, our estimation can simultaneously identify λ_g and η_g. The particular combination of $\lambda_g = 4123.83$ and $\eta_g = 0.95$ generates the gilt price dispersion level of 0.00064. Panels 2c and 2d imply that one could increase λ_g and η_g simultaneously and generate the same level of price dispersion. In doing so, however, one would increase the average trade intensity, and so, would overshoot the empirical average trade intensity. Thus, given γ and δ_g, only a unique combination of λ_g and η_g can match the empirical trade intensity and the empirical price dispersion at the same time.

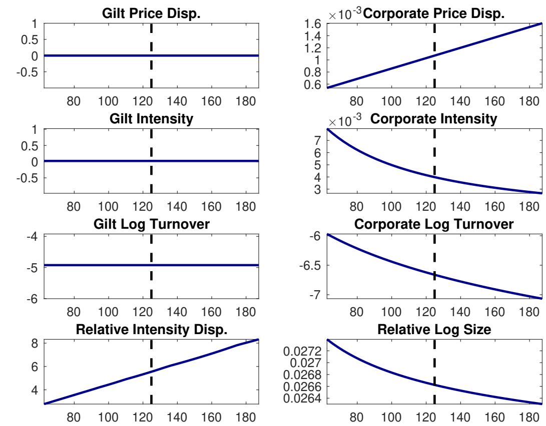
This realization helps one understand the differences of search frictions and dealers' market power across the government bond and the corporate bond market. Average trade intensity in the gilt market is more than six times that of the corporate bond market. This contributes to a large difference between the estimated matching efficiency parameters in the two markets, $\lambda_g >> \lambda_c$. Despite this large difference between the two markets in terms of search frictions, the gilt price dispersion is only around half of the corporate bond price dispersion. Thus, our estimation "rationalizes" this with a significantly larger dealer market power in the gilt market, $\eta_g >> \eta_c$. One may be concerned that our finding of significantly larger dealer market power in the gilt market may be specific to the particular way we calculate price dispersion from transaction data. In Appendix E, we show that more conventional ways of calculating price dispersion would actually exacerbate this difference. We view our less extreme results more reliable because our way of calculating price dispersion does a better job at eliminating some intra-day price volatility caused by intra-day arrival of news.

Figure 3: Illustrating the Sources of Identification for the Corporate Bond Market Parameters

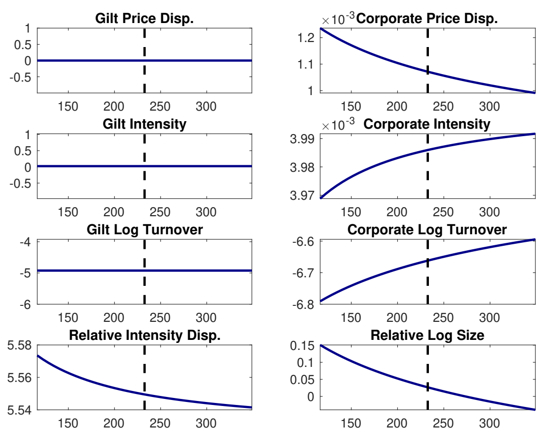
(a) Curvature of the Common Utility (γ)



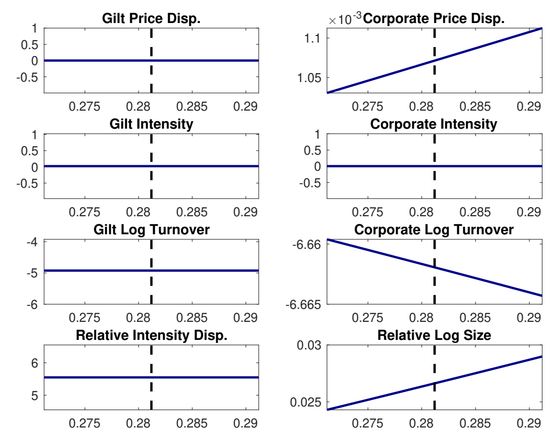
(b) Exponential Distn. Parameter for $\tilde{G}(\chi_1)$ (δ_c)



(c) Matching Efficiency (λ_c)



(d) Dealers' Market Power (η_c)



Notes: This figure shows how the theoretical moments change as we perturb one of the seven parameters around its estimated value (denoted by the vertical line), while keeping all the other parameters fixed at their estimated values. Panels 3a, 3b, 3c and 3d show the results for the curvature parameter (γ), the exponential parameter $G(\chi_1)$, for λ_c and for η_c.

4.4 Quantitative Assessment of Frictions and Welfare

A main advantage of using our structural model to study OTC markets is that we can estimate two key objects that characterize the severity of trading frictions: average trading delays and dealers' bargaining power. Table 4 shows the results for the two markets. Trading delays are expressed in terms of days, and dealers' bargaining power is expressed as the fraction of the surplus that dealers get from each trade. Mathematically, the expected trading delay of a client with taste shock intensity χ_1 is $1/(\lambda \log(1+\chi_1))$ and the share of transaction surplus a dealer captures is η.

We find that trading delays could be measured in minutes with a median value of less than ten minutes, and are thereby negligible in government bond markets —consistent with the calibration results of Vayanos and Weill (2008) for the US government bond market. In contrast, trading delays in corporate bonds are substantially larger with a median value of around three quarters of a day.

Table 4: Welfare Results I: Estimated Trading Delays and Dealers' Bargaining Power

	Government Bonds	Corporate Bonds
	(1)	(2)
Average Trading Delays		
Median	0.0080215	0.77668
p25	0.0040705	0.38941
p75	0.019158	1.8683
Dealers' Bargaining Power	95.15%	28.12%

Notes: This table reports summary statistics for trading delays (upper panel) and dealers' bargaining power (lower panel), implied by the theoretical model evaluated at the estimated parameter values (as reported in Table 2). Trading delays are expressed as a fraction of a trading day.

The lower panel of Table 4 shows the estimated shares of transaction surplus captured by dealers in both markets, which we refer to as dealers' bargaining power in short. The respective bargaining power estimates of 95.15% and 28.12% in the gilt and the corporate bond markets confirm the common view that dealers enjoy a high market power in bilateral OTC markets.[20] Interestingly, our parameter estimates show that the dealer market power is significantly larger in the gilt market where dealers provide liquidity more actively compared to the corporate bond market.[21]

Table 5 estimates, separately for each bond market, the relative welfare loss caused by the presence of OTC market frictions. For the calculation, we use the formulas presented in Proposition 4. The top panel of Table 5 reports a "95% confidence interval" for the welfare loss relative to the first-best benchmark in each market. These intervals are calculated by finding the minimum and the maximum welfare loss that a set of parameter combinations from the 95% confidence intervals of our parameter estimates can generate. The medium panel reports the relative welfare loss levels in both markets implied by the estimated parameter values exactly. These estimates imply that the welfare loss caused by OTC frictions is quite sizable in both markets, while it is significantly (almost 50%) larger in the corporate bond market than in the gilt market.

[20]See Duffie (2011), Di Maggio, Kermani, and Song (2017), Li and Schürhoff (2019), Schultz and Song (2019), and Hendershott, Li, Livdan, and Schürhoff (2020) for a discussion.

[21]Feldhütter (2012) finds that dealers capture 97% of trade surpluses in the US corporate bond market. Similarly, Hendershott, Li, Livdan, and Schürhoff (2020) find that the US corporate bond dealers capture 98% of the trade surplus when selling to an insurance company and capture 94% when buying from an insurance company. Our estimates imply that the US corporate bond market and the gilt market are very similar in terms of dealers' market power, but dealers in the UK corporate bond market have significantly lower market power. One reason behind this difference is arguably the relatively more active client-to-client segment of the UK corporate bond market, which acts as a competitive fringe and limits dealers' market power. See Appendix F for more information. Another explanation is the endogeneity of market power, η. Choi and Rocheteau (2021) show that as search frictions mitigate, market participants have incentive increase their rent-seeking activities in order to have a higher market power.

We also decompose the welfare loss into two parts with different economic meanings: the part attributed to technological constraints that cannot be relaxed unless the market structure itself is changed (i.e., the welfare loss caused by the search frictions, characteristic of OTC markets) and the remaining part that is due to imperfect competition between dealers (i.e., the welfare loss caused by inefficient price impact avoidance of clients that a benevolent social planner could eliminate via Pigouvian taxation). The bottom panel of Table 5 shows that, in the corporate bond market, almost the entire loss can be attributed to search frictions with dealers' market power making a small contribution. In the gilt market, the decomposition of the welfare loss is more even across the two components with dealers' market power being the slightly more contributor. This is not surprising given the high dealer bargaining power estimate of 95.15% in the gilt market.

Table 5: Welfare Results II: Estimated Welfare Losses

	Government Bonds	Corporate Bonds
	(1)	(2)
Min. - Max. Welfare Loss	6.637% - 9.511%	11.092% - 13.637%
Welfare Loss	7.7899%	12.174%
Due to Search Frictions	3.8871%	12.027%
Extensive Margin	1.4664%	4.9834%
Intensive Margin	2.4207%	7.0436%
Due to Dealers' Market Power	3.9028%	0.1467%

Notes: This table reports the welfare losses in the government bond and corporate bond markets implied by the estimated parameter values and standard errors (as reported in Table 2). The top panel reports the minimum and the maximum possible relative welfare loss in each market implied by the 95% confidence intervals of the estimated parameter values. The medium panel reports the relative welfare loss levels in each market implied by the estimated parameter values exactly. The bottom panel reports various decompositions.

Table 5 highlights two important lessons. First, even if the estimated median trading delays sound reasonably small (a few minutes in the gilt market), the welfare loss caused by them can be sizable. This again points to the importance of utilizing all dimensions of market liquidity (frequency, trade size, and price) to correctly capture the market participants' preferences and multi-dimensional heterogeneity among them. Indeed, the curvature of the utility functions determines how costly it is to be stuck with an undesirable asset position due to frictions. Further, heterogeneity determines how variable welfare losses are across market participants, which affects the quantification of aggregate effects in the presence of skewed distributions. The second lesson that Table 5 highlights is that severe search frictions and high dealer market power do not have to be coupled together. Indeed, our estimations imply that there are significantly less search friction in the gilt market, while gilt dealers exert a high bargaining power against their clients at the same time.

The last exercise we perform using our estimation is to further decompose the welfare loss caused purely by search frictions into two components: a component that stems from the clients' endogenous demand cutting behavior in response to search frictions (i.e., intensive margin effect) and a component that stems from the fact that some clients will be stuck with wrong asset positions for some time even if they choose their second-best trade quantities (i.e., extensive margin effect). Our estimation implies that in both markets, the extensive and intensive margins amount to 40% and 60%, respectively. This points to the importance of taking into account the clients' flexibility to adjust their demand in response to frictions in financial markets. Existing structural models of OTC markets such as Gavazza (2016) and Hendershott, Li, Livdan, and Schürhoff (2020) did not consider this intensive margin effect by fixing trade sizes at one with their indivisible asset assumption.

4.5 Monte Carlo Evidence

To confirm that the seven parameters in our model are recoverable using the eight moment conditions, we have conducted a series of Monte Carlo simulations. Table 6 reports the details of an example. In this particular simulation, we simulate 100 trade-level datasets, whereby each dataset consists of 800 days, with 800 clients trading in each dataset. We use the parameter values described by column 1 of Table 6.

Table 6: Monte Carlo Results

	True Values (1)	Mean of Simulated (2)	Std of Simulated (3)
γ – Curvature of the utility function	10	10.0005	0.11559
First Market			
δ_I – Exponential distn. parameter for $G(\chi_1)$	22	22.118	0.1122
λ_I – Matching efficiency	3000	3000.5956	1.9891
η_I – Dealers' market power	0.95	0.94922	0.0026106
Second Market			
δ_{II} – Exponential distn. parameter for $G(\chi_1)$	125	123.9817	1.2947
λ_{II} – Matching efficiency	250	256.2822	14.3315
η_{II} – Dealers' market power	0.3	0.30576	0.0077124

Notes: This table shows the results from a Monte Carlo simulation, where the data generation used the parameters shown in column 1. The mean (column 2) and standard deviation (column 3) of estimated parameters are based on the 100 Monte Carlo simulated samples.

For each market, on each day and in each dataset, the simulation proceeds in five steps: (i) the cross-sectional distribution of taste shock intensities of clients is simulated using exponential distributions and parameters δ_I and δ_{II} in the first and the second market, respectively; (ii) the Poisson arrival rates implied by (i) are used to simulate client-specific taste shock processes in the two markets; (iii) the Poisson arrival rates implied by (i) and $\chi_2 = \lambda \log(1 + \chi_1)$ are used to simulate client-specific random trading times in the two markets; (iv) assuming random initial asset holding across clients consistent with the stationary equilibrium distribution, the information obtained in steps (i)-(iii) is used to simulate the time-series of transactions (trade size and trading costs) of each client, on each day in each market and in each dataset; (v) the simulated datasets are used to compute the eight simulated moments that correspond to the empirical moments used in our baseline estimation in Section 4.1.2. Columns 2-3 of Table 6 summarize the results from the Monte Carlo exercise. We find that the eight moment conditions are sufficient to recover the seven model parameters with high precision and without bias.

Table 7: Monte Carlo Results: Absolute Intensity Deviation

Average Absolute Intensity Deviation	Theoretical Moments (1)	Mean of Simulated (2)	Std of Simulated (3)
First Market	0.016716	0.044444	0.00035533
Second Market	0.0029313	0.013251	0.053014
Ratio of First to Second	5.7026	5.5329	0.55362

Notes: This table shows the results from a Monte Carlo simulation, where the data generation used the parameter values shown in column 1 of Table 6. Column 1 shows the implied theoretical moments for average absolute intensity deviation for the first market (first row), the second market (second row), and for the ratio of these values (row 3). Column 2 (resp. 3) shows the values that are obtained by taking the means (resp. standard deviations) across the 100 datasets.

In addition, we employ the Monte Carlo exercise to justify why we use the ratio of absolute intensity deviations across the two markets as an empirical moment in the estimation, instead of using the market-specific absolute intensity deviations as separate moments. As can be seen from the table, average simulated absolute intensity deviations are a biased estimate of the absolute intensity deviations implied by the true data generating processes. In contrast, we find that the estimation comes very close to recovering the ratio of absolute intensity deviations across the two markets (last row of Table 7). For this reason, we use the ratio of intensity deviations (instead of the levels) as an empirical moment when estimating the model on real data.

5 Frictions during Turbulent Times

As an application of our framework, we present an analysis of the COVID-19 episode in the UK through the lens of our structural model. More specifically, we first calculate how the empirical moments used in the baseline estimation changed during February-April 2020. We then re-estimate the model parameters, and quantify how trading delays, dealers' market power, and the resulting welfare losses from these frictions have changed during that turbulent period.

Background and literature The spread of the COVID-19 pandemic in early 2020 presented a major shock to the global financial system, including fixed-income markets. The crisis was characterized by large and persistent selling pressures across many asset classes. Recent studies documented that these selling pressures were driven by bond mutual funds that suffered large outflows (Falato, Goldstein, and Hortacsu, 2021; Ma, Xiao, and Zeng, 2022). Other papers emphasized the inability of the dealer sector to absorb inventory onto their balance sheets (Kargar, Lester, Lindsay, Liu, Weill, and Zuniga, 2021). As a consequence, liquidity dried up both in government bond markets (Duffie 2020; He, Nagel, and Song 2022) and in corporate bond markets featuring large increases in trading costs (O'Hara and Zhou, 2021). Only the quick and large-scale interventions by central banks across the world helped restore liquidity and avoid a prolonged worsening of financing conditions (Haddad, Moreira, and Muir 2021).

In what follows, we utilize our structural framework to understand how clients' preferences and trading delays they face as well as dealers' market power changed in the UK government and corporate bond markets. This provides us with a unique opportunity to quantitatively assess the resilience of different market structures to the large negative shocks to the financial system such as the COVID-19 shock.

Data and stylized facts To conduct this analysis, we employ the MiFID II bond transaction data, which covers the period from January 2018 to May 2020. While ZEN is generally regarded as the predecessor of the MiFID II database, there are differences in the reporting requirements, which makes it difficult to consistently merge it with our baseline sample. To obtain empirical moments corresponding to the COVID-19 crisis, that can be used to re-estimate our model for this period, we proceed as follows. We compute how much each of the eight moments changed from the 2018-2019 period to the February-April 2020 period. We then use these percentage or log changes to adjust the moment values from our baseline sample (reported in Table 1 and Table 3).

Table 8: Empirical Moments during COVID-19

Variable	Government Bonds Change (1)	Government Bonds Implied Moment (2)	Corporate Bonds Change (3)	Corporate Bonds Implied Moment (4)
Price Dispersion	+53.6%	0.000978	+146.3%	0.002639
Average Intensity	+17.9%	0.027338	-8.0%	0.003520
Log Turnover	+0.223	-4.580826	+0.058	-6.851440
Relative Intensity Dispersion	+26.0%	6.686867	0%	1
Relative Log Trade Size	+0.052	0.078623	0	0

Notes: This table summarizes the empirical moments that are used for the structural estimation for the COVID-19 period (Feb-April 2020). Column (1) and (3) summarize how much the moments changed from the period 2018-2019 to the period February-April 2020. Column (2) and (4) report the implied moments for the COVID-19 period. Price dispersion is the scaled mean absolute deviation of transaction price from the average transaction price in the given hour. Average intensity is the mean of the clients' number of transactions. Turnover is computed as the daily trading volume scaled by asset supply. Intensity dispersion is the mean absolute deviation of clients' number of transactions from average intensity. Average trade size is the mean (across clients) of clients' mean transaction size scaled by asset supply.

Table 8 reports the corresponding changes in our estimation moments during the COVID-19 crisis. Price dispersion experienced the largest change with 53.6% and 146.3% increases in the government and corporate bond markets, respectively. Both average client intensity in the gilt market (17.9%) and intensity dispersion in the gilt market relative to the corporate bond market (26%) increased, but average client intensity decreased in corporate bonds (-8%). Turnover increased in both markets, but the increase was more pronounced in government bonds. Finally, average trade size has the same pattern as turnover, and so, relative size increased.

The large increase in price dispersion is indicative of substantial worsening of trading frictions during the crisis, with some notable cross-market differences emerging: (i) the increase in price dispersion was almost three times as large in corporate bonds; (ii) the corporate bond trade intensity fell while an increase in intensity in government bonds was observed; and (iii) the corporate bond volume and trade size only moderately changed compared to the government bond market. The combination of these facts and the intuition behind the identification properties of our structural model (see Section 4.3) suggest that the worsening of trading frictions during the COVID-19 crisis was likely more severe in the corporate bond market, compared to the government bond market which likely absorbed selling pressures to a larger extent. To quantify these effects, we now turn to structural estimation.

Estimation results Using the moments reported in Table 8, we re-estimate our structural model. The parameter estimates for the turbulent COVID-19 period is reported in Table 9. In addition, Table 10 shows the fit of the model for this period.

Table 9: Parameter Estimates for the COVID-19 Period

	Government Bonds		Corporate Bonds	
	Normal	Turbulent	Normal	Turbulent
	(1)	(2)	(3)	(4)
γ – Curvature of the utility function	17.95	9.10	17.95	9.10
δ – Exponential distn. parameter for $G(\chi_1)$	22.58	19.27	124.9	128.05
λ – Matching efficiency	4123.83	735.06	232.65	17.19
η – Dealers' market power	0.95	0.98	0.28	0.37

Notes: This table reports the estimates of the parameters. The parameter estimates are obtained by minimizing the objective function (4.7). Results in columns (1) and (3) are based on empirical moments from the period 2011m8-2017m12, as reported in Table 1 and Table 3. Results in columns (2) and (4) are based on empirical moments from the COVID-19 period as reported in Table 8.

Table 10: Model Fit (COVID-19)

	Government Bonds		Corporate Bonds	
Moments	Empirical	Theoretical	Empirical	Theoretical
	(1)	(2)	(3)	(4)
Price Dispersion	0.000978	0.000978	0.002639	0.002639
Average Intensity	0.027338	0.025911	0.003520	0.003688
Log Turnover	-4.580826	-4.656439	-6.851440	-6.682323
Relative Intensity Dispersion	6.686867	7.025747	1	1
Relative Log Trade Size	0.078623	0.078644	0	0

Notes: This table reports the values of the empirical moments and of the theoretical moments calculated at the estimated parameters. Columns (1)-(2) and Columns (3)-(4) show the results for the government bond and corporate bond markets, respectively.

Table 9 reveals that there are some notable changes in the parameter estimates across the two time periods. First, the elasticity of clients' utility function, $1/\gamma$, increased from 0.056 to 0.11. This doubling of the elasticity naturally relates to clients' heightened desire to trade due to the selling pressures during the COVID-19 crisis. Table 8 shows that log turnover increased in both markets. Our model estimation accounts for this by increasing the estimated elasticity of the utility function during COVID-19. Because of this particular change in preferences, being able to trade is inherently more important for clients during the crisis. Thus, dealers' immediacy provision becomes even more crucial from a welfare standpoint.

Second, our estimation implies that the average taste shock frequency of the average client, $1/\delta$, increased in the gilt market, while it declined in the corporate bond market during the COVID-19 crisis. This is consistent with the evidence that during turbulent times, investors rely more on liquid assets and less on illiquid assets to manage their overall portfolio, leading to an apparent "reverse flight to liquidity." See, for example, Choi, Han, Shin, and Yoon (2020) and Ma, Xiao, and Zeng (2022).

A comparison of matching efficiencies in Table 9 reveals that dealers significantly refrained from immediacy provision in both markets. Matching efficiency declined by 83% and 93% in the gilt and the corporate bond market, respectively. This finding lends support to the view that the OTC market structure is not sufficiently resilient to large negative shocks such as the COVID-19 period around March 2020.

Finally, dealers exert more market power during the COVID-19 crisis according to the comparison of the market power estimates from Table 9. However, the magnitudes of the change are noticeably different in the two markets. The share of the transaction surplus captured by dealers in the gilt market increased by less than 3 percentage points but increased by almost 10 percentage points in the corporate bond market. A strict comparison of these two numbers may not be very revealing because the gilt market dealers already had a market power of 95% during the normal times, without much room for further increase. However, combined with the difference in changes in matching efficiency, the gilt market seems to be relatively more stable than the corporate bond market.

This then raises the question: could one conclude that the gilt market is more stable from a welfare perspective as well? Our results in Tables 11 and 12 shed some light on this question.

Table 12 shows that the percentage welfare loss tripled in both markets during the COVID-19 crisis compared to normal times. This means that although by looking at the changes in the friction parameters λ and η, one might conclude that the gilt market is more resilient to negative shocks, the explicit welfare accounting shows that the relative increase in welfare losses are the same in the two markets. This is mainly because clients rely more on the gilt market during turbulent times due to their reverse flight to liquidity behavior; as a result, a relatively mild worsening of frictions in the gilt market leads to a relative welfare loss that is comparable to that in the corporate bond market. Hence, from a relative welfare perspective, both markets seem

Table 11: Welfare Results I: Estimated Trading Delays and Dealers' Bargaining Power (COVID-19)

	Government Bonds		Corporate Bonds	
	Normal	Turbulent	Normal	Turbulent
	(1)	(2)	(3)	(4)
Average Trading Delays				
Median	0.0080215	0.038495	0.77668	10.7769
p25	0.0040705	0.019582	0.38941	5.403
p75	0.019158	0.091801	1.8683	25.9252
Dealers' Bargaining Power	95.15%	97.99%	28.12%	36.71%

Notes: This table reports summary statistics for trading delays (upper panel) and dealers' bargaining power (lower panel), implied by the theoretical model evaluated at the estimated parameter values. Trading delays are expressed as a fraction of a trading day. Results in columns (1)-(4) are based on parameter values from the respective columns of Table 9.

Table 12: Welfare Results II: Estimated Welfare Losses (COVID-19)

	Government Bonds		Corporate Bonds	
	Normal	Turbulent	Normal	Turbulent
	(1)	(2)	(3)	(4)
Welfare Loss	7.7899%	25.44%	12.174%	39.625%
Due to Search Frictions	3.8871%	10.378%	12.174%	38.837%
Due to Dealers' Market Power	3.9028%	15.061%	0.1467%	0.7878%

Notes: This table reports the welfare losses in the government bond and corporate bond markets implied by the estimated parameter values. The top panel reports the relative welfare loss levels in each market implied by the estimated parameter values exactly. The bottom panel reports various decompositions. Results in columns (1)-(4) are based on parameter values from the respective columns of Table 9.

equally fragile during turbulent times.

Counterfactual analysis Now that we have estimated the deep parameters of our model for the COVID-19 period, we turn to a counterfactual analysis to investigate how resilient the OTC market structure is when faced with a large negative shock. Table 9 shows that clients' preferences and market frictions both changed during the COVID-19 crisis. Two natural questions are, then, how much of the additional welfare losses in turbulent times can be explained by the change in clients' preferences and how much of it is caused by the worsening of the OTC markets' functioning. To shed light on these questions, we calculate the welfare losses in the government and corporate bond markets in a counterfactual scenario.

Table 13 shows the welfare losses that would have realized in the government and corporate bond markets during the COVID-19 period if market frictions had not exacerbated. More specifically, we use clients' preference parameters (i.e., the elasticity of clients' utility function and the exponential distribution parameter for taste shock intensity) from the COVID-19 period that reflect the change in clients' inherent trading needs during turbulent times, but we keep market friction parameters (i.e., matching efficiency and dealers' market power) the same as normal times. Comparing Table 12 and Table 13 implies that the welfare loss would increase from 7.79%

Table 13: Estimated Welfare Losses (Counterfactual)

	Government Bonds	Corporate Bonds
	(1)	(2)
Welfare Loss	10.717 %	15.254%
Due to Search Frictions	6.1495%	15.092%
Due to Dealers' Market Power	4.5671%	0.1617%

Notes: This table reports the welfare losses in the government bond and corporate bond markets implied by the parameter values chosen for our counterfactual analysis. In both markets, parameters γ and δ are from columns (2) and (4) of Table 9 and parameters λ and η are from columns (1) and (3) of Table 9.

to 10.72% in the government bond market and from 12.17% to 15.25% in the corporate bond market, if the OTC market structure was fully resilient to large negative shocks like the COVID-19 shock of 2020. Considering that welfare losses actually rose to 25.44% and 39.63%, the majority of the additional welfare losses during turbulent times occurs because OTC market intermediaries are not as able or willing to supply liquidity as they are during normal times.

Note that our analysis is mainly concerned with quantifying search frictions and dealers' market power but is essentially agnostic about the sources of these frictions. Therefore, while we can quantitatively demonstrate that the quality of intermediation for clients' trades worsens during turbulent times, understanding the role of OTC market dealers' incentives, of frictions they face in inter-dealer trade, and of regulations imposed on them goes beyond the scope of this paper and is subject of an ongoing research agenda (e.g. Coen and Coen, 2021; Chiu, Davoodalhosseini, and Jiang, 2022; Kargar, Lester, and Weill, 2022).

6 Conclusion

A theoretical literature following Duffie, Gârleanu, and Pedersen (2005) and Lagos and Rocheteau (2009) generated a wealth of qualitative insights into understanding the role of search frictions and dealers' market power in OTC financial markets. We have developed a dynamic structural model that uses transaction-level data with client identities and quantifies the search frictions and dealers' market power in question. Utilizing data on the UK government and corporate bond markets, we find that clients' search times in the government bond market can be measured in minutes, whereas in the corporate bond market they can be measured in hours with a median of around three quarters of a trading day. In addition, our estimation implies that the share of transaction surpluses that dealers capture is about 95% in the government bond market, while it is about 28% in the corporate bond market. That is, despite the reasonably low level of search frictions, government bond dealers can exert a significant market power over their clients. These findings pose a challenge to the common view that high dealer market power and severe search frictions must co-exist in an intermediated market. Furthermore, we find that the welfare loss from frictions in the government and corporate bond markets are 7.8% and 12.2%, respectively, and our decomposition implies that this loss is almost exclusively caused by search frictions in the corporate bond market, while search frictions and dealers' market power split the welfare loss equally in the government bond market. One lesson from these findings is that even if search times are reasonably small (a few minutes in government bonds), the welfare loss from search frictions can be sizable (almost 4% in government bonds). Finally, using data from the COVID-19 crisis period, we find that the welfare losses might more than triple during turbulent times, underlining the fragility of the OTC market structure.

A future avenue for research is to incorporate a frictional inter-dealer market into our framework to study the role of dealers' costs and benefits in intermediation provision incentives and the liquidity and welfare implications of regulations imposed on them. Another related avenue is to leverage the dealer identities provided in the ZEN dataset to see how heterogeneous dealer characteristics translate into their endogenous intermediation provision behavior or to understand the contribution of dealer heterogeneity vis-a-vis client heterogeneity to market-wide liquidity measures such as price dispersion.

References

Jason Allen and Milena Wittwer. Centralizing over-the-counter markets? Working paper, 2021.

Jason Allen, Robert Clark, and Jean-François Houdec. Search frictions and market power in negotiated price markets. *Journal of Political Economy*, 127(4):1550–1598, 2019.

Joseph Altonji and Lewis M Segal. Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366, 1996.

S Borağan Aruoba, Guillaume Rocheteau, and Christopher Waller. Bargaining and the value of money. *Journal of Monetary Economics*, 54(8):2636–2655, 2007.

Thomas Belsham, Alex Rattan, and Rebecca Maher. Corporate Bond Purchase Scheme: design, operation and impact. *Bank of England Quarterly Bulletin*, 57(3):170–181, 2017.

Paula Beltran. A lending network under stress: A structural analysisof the money market funds industry. Working paper, 2022.

Efraim Benmelech and Nittai Bergman. Debt, Information, and Illiquidity. NBER Working Papers 25054, National Bureau of Economic Research, Inc, September 2018.

Evangelos Benos and Filip Zikes. Funding constraints and liquidity in two-tiered otc markets. *Journal of Financial Markets*, 2018.

Hendrik Bessembinder, Chester Spatt, and Kumar Venkataraman. A survey of the microstructure of fixed-income markets. *Journal of Financial and Quantitative Analysis*, 55(1):1–45, 2020.

Zachary Bethune, Bruno Sultanum, and Nicholas Trachter. An information-based theory of financial intermediation. Working paper, 2018.

BoE. The gilt-edged market: the Bank of England's relationship with the gilt-edged market makers and inter-dealer brokers. *Bank of England Quarterly Bulletin*, pages 198–203, June 1997.

Giulia Brancaccio, Dan Li, and Norman Schürhoff. Learning by trading: The case of the u.s. market for municipal bond. Working paper, 2017.

Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru. Why is intermediating houses so difficult? evidence from ibuyers. Working paper, 2020.

Dario Cestau, Burton Hollifield, Dan Li, and Norman Schürhoff. Municipal bond markets. *Annual Review of Financial Economics*, 11:65–84, 2019.

Briana Chang and Shengxing Zhang. Endogenous market making and network formation. 2016.

Briana Chang and Shengxing Zhang. Risk concentration and interconnectedness in otc markets. Working paper, 2021.

Jonathan Chiu and Thorsten V. Koeppl. Trading dynamics with adverse selection and search: Market freeze, intervention and recovery. *Review of Economic Studies*, 83(3):969–1000, 2016.

Jonathan Chiu, Mohammad Davoodalhosseini, and Janet Hua Jiang. Dealer inventory and market liquidity. Mimeo, 2022.

Jaewon Choi, Jungsuk Han, Sean Seunghun Shin, and Ji Hee Yoon. The more illiquid, more expensive: A search-based explanation of the illiquidity premium. Working paper, 2020.

Michael Choi and Guillaume Rocheteau. Foundations of market power in monetary economies. Working paper, 2021.

Robert Clark, Jean-François Houdec, and Jakub Kastl. The industrial organization of financial markets. *Handbook of Industrial Organization*, 5(1):427–520, 2021.

Jamie Coen and Patrick Coen. A structural model of liquidity in over-the-counter markets. Working paper, 2021.

Jean-Edouard Colliard, Thierry Foucault, and Peter Hoffmann. Inventory management, dealers' connections, and prices in otc markets. *Journal of Finance*, 76(5):2199–2247, 2021.

Robert Czech and Gábor Pintér. Informed trading and the dynamics of client-dealer connections in corporate bond markets. Working paper, 2020.

Robert Czech, Shiyang Huang, Dong Lou, and Tianyu Wang. Informed trading in government bond markets. *Journal of Financial Economics*, 142(3):1253–1274, 2021.

Babur De los Santos, Ali Hortaçsu, and Matthijs R Wildenbeest. Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102 (6):2955–2980, 2012.

Marco Di Maggio, Amir Kermani, and Zhaogang Song. The value of trading relations in turbulent times. *Journal of Financial Economics*, 124(2):266–284, 2017.

Jens Dick-Nielsen, Thomas K. Poulsen, and Obaidur Rehman. Dealer networks and the cost of immediacy. Working paper, 2021.

DMO. A guide to the roles of the DMO and Primary Dealers in the UK government bond market. Technical report, Debt Management Office, August 2011.

DMO. GEMM Guidebook: A guide to the roles of the DMO and Primary Dealers (GEMMs) in the UK government bond market. Technical report, United Kingdom Debt Management Office, 2021.

Darrell Duffie. *Dark Markets*. Princeton University Press, Princeton, NJ, 2011.

Darrell Duffie. Still the world's safe haven? Redesigning the U.S. Treasury market after the COVID-19 crisis. Working paper, Brookings, 2020.

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Over-the-counter markets. *Econometrica*, 73(6):1815–1847, 2005.

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Valuation in over-the-counter markets. *Review of Financial Studies*, 20(6):1865, 2007.

Zvi Eckstein and Gerard J. van den Berg. Empirical labor search: A survey. *Journal of Econometrics*, 136(2):531–564, 2007.

Zvi Eckstein and Kenneth I. Wolpin. Estimating a market equilibrium search model from panel data on individuals. *Econometrica*, 58(4):783–808, 1990.

Amy K. Edwards, Lawrence E. Harris, and Michael S. Piwowar. Corporate bond market transaction costs and transparency. *Journal of Finance*, 62(3):1421–1451, 2007.

Mark Egan. Brokers versus retail investors: Conflicting interests and dominated products. *Journal of Finance*, 74(3):1217–1260, 2019.

Andrea L. Eisfeldt, Bernard Herskovic, Srimam Rajan, and Emil N. Siriwardane. Otc intermediaries. Working paper, 2018.

Antonio Falato, Itay Goldstein, and Ali Hortacsu. Financial fragility in the COVID-19 crisis: The case of investment funds in corporate bond markets. *Journal of Monetary Economics*, 123(C): 35–52, 2021.

Maryam Farboodi, Gregor Jarosch, and Robert Shimer. The emergence of market structure. Working paper, 2018.

Peter Feldhütter. The same bond at different prices: Identifying search frictions and selling pressures. *Review of Financial Studies*, 25(4):1155–1206, 2012.

Manolis Galenianos and Alessandro Gavazza. A structural model of the retail market for illicit drugs. *American Economic Review*, 107(3):858–896, 2017.

Kenneth D Garbade and William L Silber. Price dispersion in the government securities market. *Journal of Political Economy*, 84(4):721–740, 1976.

Pieter A. Gautier and Coen N. Teulings. Sorting and the output loss due to search frictions. *Journal of the European Economic Association*, 13(6):1136–1166, 2015.

Alessandro Gavazza. An empirical equilibrium model of a decentralized asset market. *Econometrica*, 84:1755–1798, 2016.

Alessandro Gavazza and Alessandro Lizzeri. Frictions in product markets. *Handbook of Industrial Organization*, 4:433–484, 2021.

George-Levi Gayle and Andrew Shephard. Optimal taxation, marriage, home production, and family labor supply. *Econometrica*, 87(1):291–326, 2019.

Valentin Haddad, Alan Moreira, and Tyler Muir. When Selling Becomes Viral: Disruptions in Debt Markets in the COVID-19 Crisis and the Fed's Response. *The Review of Financial Studies*, 34(11):5309–5351, 01 2021.

Larry Harris. *Trading and Electronic Markets: What Investment Professionals Need to Know*. The CFA Institute Research Foundation, Charlottesville, VA, 2015.

Zhiguo He, Stefan Nagel, and Zhaogang Song. Treasury inconvenience yields during the covid-19 crisis. *Journal of Financial Economics*, 143(1):57–79, 2022.

Terrence Hendershott, Dan Li, Dmitry Livdan, and Norman Schürhoff. Relationship trading in otc markets. *Journal of Finance*, 75(2):683–734, 2020.

Terrence Hendershott, Dmitry Livdan, and Norman Schürhoff. All-to-all liquidity in corporate bonds. Working paper, 2021.

Han Hong and Matthew Shum. Using price distributions to estimate search costs. *RAND Journal of Economics*, 37(2):257–275, 2006.

Bo Honore, Thomas Jorgensen, and Aureo de Paula. The informativeness of estimation moments. *Journal of Applied Econometrics*, 35(7):797–813, 2020.

Ali Hortaçsu and David McAdams. Mechanism choice and strategic bidding in divisible good auctions: An empirical analysis of the turkish treasury auction market. *Journal of Political Economy*, 118(5):833–865, 2010.

Ali Hortaçsu and Chad Syverson. Product differentiation, search costs, and competition in the mutual fund industry: A case study of s&p 500 index funds. *Quarterly Journal of Economics*, 119(2):403–456, 2004.

Edith Hotchkiss and Gergana Jostova. Determinants of corporate bond trading: A comprehensive analysis. *Quarterly Journal of Finance*, 7(02):1750003, 2017.

Julien Hugonnier, Benjamin Lester, and Pierre-Olivier Weill. Frictional intermediation in over-the-counter markets. *Review of Economic Studies*, 87:1432–1469, 2020.

Plamen Ivanov, Alexei Orlov, and Michael Schihl. Bond liquidity and dealer inventories: Insights from US and European regulatory data. Occasional Paper 52, Financial Conduct Authority, 2021.

Rainer Jankowitsch, Amrut Nashikkar, and Marti G. Subrahmanyam. Price dispersion in otc markets: A new measure of liquidity. *Journal of Banking and Finance*, 35:343–357, 2011.

Boo-Sung Kang and Steven L Puller. The effect of auction format on efficiency and revenue in divisible goods auctions: A test using korean treasury auctions. *Journal of Industrial Economics*, 56(2):290–332, 2008.

Mahyar Kargar, Benjamin Lester, David Lindsay, Shuo Liu, Pierre-Olivier Weill, and Diego Zuniga. Corporate Bond Liquidity during the COVID-19 Crisis. *The Review of Financial Studies*, 34(11):5352–5401, 05 2021. ISSN 0893-9454.

Mahyar Kargar, Benjamin Lester, and Pierre-Olivier Weill. Inventory, market making, and liquidity: Theory and application to the corporate bond market. Mimeo, 2022.

Jakub Kastl. Discrete bids and empirical inference in divisible good auctions. *Review of Economic Studies*, 78(3):974–1014, 2011.

Péter Kondor and Gábor Pintér. Clients' connections: Measuring the role of private information in decentralised markets. *Journal of Finance*, 77(1):505–544, 2022.

Ricardo Lagos and Guillaume Rocheteau. Search in asset markets. FRB of Cleveland Working Paper No. 06-07, 2006.

Ricardo Lagos and Guillaume Rocheteau. Liquidity in asset markets with search frictions. *Econometrica*, 77(2):403–426, 2009.

Ricardo Lagos and Randall Wright. A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113(3):463–484, 2005.

Benjamin Lester, Guillaume Rocheteau, and Pierre-Olivier Weill. Competing for order flow in otc markets. *Journal of Money, Credit, and Banking*, 47:77–126, 2015.

Dan Li and Norman Schürhoff. Dealer networks. *Journal of Finance*, 74(1):91–144, 2019.

Shuo Liu. Dealer's search intensity in u.s. corporate bond markets. Working paper, 2020.

Yiming Ma, Kairong Xiao, and Yao Zeng. Mutual Fund Liquidity Transformation and Reverse Flight to Liquidity. *The Review of Financial Studies*, 02 2022. ISSN 0893-9454.

V. Filipe Martins-da Rocha and Yiannis Vailakis. Existence and uniqueness of a fixed point for local contractions. *Econometrica*, 78(3):1127–1141, 2010.

Maureen O'Hara and Xing (Alex) Zhou. Anatomy of a liquidity crisis: Corporate bonds in the covid-19 crisis. *Journal of Financial Economics*, 142(1):46–68, 2021.

Maureen O'Hara, Yihui Wang, and Xing (Alex) Zhou. The execution quality of corporate bonds. *Journal of Financial Economics*, 130(2):308–326, 2018.

Tavy Ronen and Xing Zhou. Trade and information in the corporate bond market. *Journal of Financial Markets*, 16(1):61–103, 2013.

Paul Schultz and Zhaogang Song. Transparency and dealer networks: Evidence from the initiation of post-trade reporting in the mortgage backed security market. *Journal of Financial Economics*, 133(1):113–133, 2019.

Nancy L. Stokey and Robert E. Jr. Lucas. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, MA, 1989.

Bruno Sultanum. Financial fragility and over-the-counter markets. *Journal of Economic Theory*, 177:616–658, 2018.

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

Semih Üslü and Güner Velioğlu. Liquidity in the cross section of otc assets. Working paper, 2019.

Evert van Imhoff. *Optimal Economic Growth and Non-stable Population*. Springer-Verlag, Berlin, Germany, 1982.

Dimitri Vayanos and Tan Wang. Search and endogenous concentration of liquidity in asset markets. *Journal of Economic Theory*, 136(1):66–104, 2007.

Dimitri Vayanos and Pierre-Olivier Weill. A search-based theory of the on-the-run phenomenon. *Journal of Finance*, 63:1361–1398, 2008.

Pierre-Olivier Weill. The search theory of over-the-counter markets. *Annual Review of Economics*, 12:747–773, 2020.

Susan E. Woodward and Robert E. Hall. Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review*, 102(7):3249–3276, 2012.

Supplement to "Comparing Search and Intermediation Frictions Across Markets"

This online appendix contains proofs and additional empirical results omitted from the printed manuscript.

Gábor Pintér[1] **Semih Üslü**[2]

A Proofs

A.1 Proof of Proposition 1

First, we establish that the functional equation (3.8) admits a unique real solution, taking as given the inter-dealer market price P. Our argument is adapted from the existence and uniqueness proofs of the earlier models with unrestricted asset holdings, especially Lagos and Rocheteau (2006) and Üslü (2019), and uses the fixed point tools for dynamic programming from Martins-da Rocha and Vailakis (2010).

Rewrite (3.8) as

$$V(\varepsilon, a, \chi) = \frac{\varepsilon u(a) + \chi_1 \int_{\varepsilon_l}^{\varepsilon_h} V(\varepsilon', a, \chi)\, dF(\varepsilon') + \chi_2 (1-\eta) \max_{a' \in [-M,M]} \{V(\varepsilon, a', \chi) - (a' - a) P\}}{r + \chi_1 + \chi_2 (1-\eta)}. \tag{A.1}$$

From (A.1), one can define the mapping O such that

$$(OV)(\varepsilon, a, \chi) = \frac{1}{r + \chi_1 + \chi_2(1-\eta)} \Bigg(\varepsilon u(a) + \chi_1 \int_{\varepsilon_l}^{\varepsilon_h} V(\varepsilon', a, \chi)\, dF(\varepsilon') + \chi_2 (1-\eta) \max_{a' \in [-M,M]} \{V(\varepsilon, a', \chi) - (a' - a) P\} \Bigg). \tag{A.2}$$

Then, showing (3.8) has a unique solution is equivalent to showing O has a unique fixed point. Let $\mathcal{T} = [\varepsilon_l, \varepsilon_h] \times [-M, M] \times \mathbb{R}^2_+$ and let $C(\mathcal{T})$ be the set of bounded continuous functions on $\mathcal{T}$. Suppose $V \in C(\mathcal{T})$, then the *theorem of the maximum* implies that the maximization on the

[1]Bank of England, e-mail: gabor.pinter@bankofengland.co.uk

[2]Carey Business School, Johns Hopkins University, e-mail: semihuslu@jhu.edu

RHS of (A.2) has a continuous solution (Theorem 3.6 of Stokey and Lucas, 1989, p. 62). Then, because $u(\cdot)$ is a continuous function defined on a compact set $[-M, M]$, $O : C(\mathcal{T}) \to C(\mathcal{T})$. We next prove Lemma 2 and Lemma 3 as intermediate steps to invoke the fixed point arguments of Martins-da Rocha and Vailakis (2010).

Lemma 2 *Let $\mathcal{D} = (d_j)_{j\in\mathbb{R}_+^2}$ be a family of semidistances such that*

$$d_j(f, g) = \sup_{x\in[\varepsilon_l,\varepsilon_h]\times[-M,M]\times[0,j]^2} |f(x) - g(x)|.$$

Let σ be the weak topology on $C(\mathcal{T})$ defined by the family $\mathcal{D}$. Then, $(C(\mathcal{T}), \mathcal{D})$ is sequentially σ-complete.

Proof. We need to show that every σ-Cauchy sequence in $C(\mathcal{T})$ converges to an element of $C(\mathcal{T})$ for the σ-topology (Martins-da Rocha and Vailakis, 2010, p. 1128). We proceed with three steps. First, we determine a candidate limiting function f for any arbitrary σ-Cauchy sequence $\{f_n\}$ in $C(\mathcal{T})$. Then, we establish that $\{f_n\}$ indeed converges to f for the σ-topology. Finally, we show $f \in C(\mathcal{T})$.

1. Fix a σ-Cauchy sequence $\{f_n\}$ in $C(\mathcal{T})$. Then, the σ-Cauchy criterion implies that $d_j(f_n, f_m) \to 0$ as $n, m \to \infty$ for all $j \in \mathbb{R}_+$. Given $x \in \mathcal{T}$, the sequence of real numbers $\{f_n(x)\}$ satisfies the Cauchy criterion; and by the completeness of real numbers, it converges in the uniform distance to a limit point—call it $f(x)$. Note that while it is possible that one or both elements of $x(3)$ may be ∞, we still have $f_n(x) < \infty$ because $\{f_n\}$ belongs to a space of bounded functions, and so, $f(x) < \infty$. Thus, the limiting values define a function $f : \mathcal{T} \to \mathbb{R}$. We then take our "candidate" limiting function for $\{f_n\}$ to be f.

2. Since $\{f_n(x)\}$ converges in the uniform distance to $f(x)$ for all $x \in \mathcal{T}$,

$$d_j(f_n, f) = \sup_{x\in[\varepsilon_l,\varepsilon_h]\times[-M,M]\times[0,j]^2} |f_n(x) - f(x)|$$

   approaches zero for all $j \in \mathbb{R}_+$ as n diverges, which means $\{f_n\}$ converges to f for the σ-topology.

3. We need to show that f is bounded and continuous. For continuity, it suffices to show that for every $j \in \mathbb{R}_+$, every $\epsilon > 0$ and every $x \in \mathcal{T}$, there exists $\zeta > 0$ such that

$$|f(x) - f(y)| < \epsilon \text{ if } d_j(x, y) < \zeta.$$

Fix j, x, and ϵ. Choose k such that $d_j(f, f_k) < \epsilon/3$, which is possible because $\{f_n\}$ converges to f for the σ-topology. Then choose ζ such that

$$d_j(x, y) < \zeta \text{ implies } |f_k(x) - f_k(y)| < \epsilon/3,$$

which is possible as f_k is continuous. Then,

$$\begin{aligned}|f(x) - f(y)| \leq |f(x) - f_k(x)| + |f_k(x) - f_k(y)| + |f_k(y) - f(y)| \\ \leq 2d_j(f, f_k) + |f_k(x) - f_k(y)| < \epsilon,\end{aligned}$$

which establishes the continuity of f. Lastly, the boundedness of f follows because $f_n \in C(\mathcal{T})$, and so, there exists $N \in \mathbb{R}$ such that $\sup_{x\in\mathcal{T}} |f_n(x)| \leq N$ and $\leq$ is a continuous relation.

■

Lemma 3 *The mapping $O : C(\mathcal{T}) \to C(\mathcal{T})$ is a 0-local contraction with respect to $\mathcal{D}$.*

Proof. We need to show that for every $j \in \mathbb{R}_+$, there exists $\beta \in [0, 1)$ such that $d_j\left(OV^A, OV^B\right) \leq \beta d_j\left(V^A, V^B\right)$ for all $V^A, V^B \in C(\mathcal{T})$ (Martins-da Rocha and Vailakis, 2010, pp. 1128-1130). We start by fixing $V^A, V^B \in C(\mathcal{T})$. Then,

$$\begin{aligned}\left(OV^A - OV^B\right)(\varepsilon, a, \chi) &= \frac{\chi_1}{r + \chi_1 + \chi_2(1-\eta)} \int_{\varepsilon_l}^{\varepsilon_h} \left[V^A(\varepsilon', a, \chi) - V^B(\varepsilon', a, \chi)\right] dF(\varepsilon') \\ &+ \frac{\chi_2(1-\eta)}{r + \chi_1 + \chi_2(1-\eta)} \left[\max_{a'\in[-M,M]} \left\{V^A(\varepsilon, a', \chi) - a'P\right\} - \max_{a''\in[-M,M]} \left\{V^B(\varepsilon, a'', \chi) - a''P\right\}\right] \\ &\leq \frac{\chi_1}{r + \chi_1 + \chi_2(1-\eta)} \int_{\varepsilon_l}^{\varepsilon_h} \left[V^A(\varepsilon', a, \chi) - V^B(\varepsilon', a, \chi)\right] dF(\varepsilon') \\ &\quad + \frac{\chi_2(1-\eta)}{r + \chi_1 + \chi_2(1-\eta)} \left[V^A(\varepsilon, a^*, \chi) - V^B(\varepsilon, a^*, \chi)\right] \\ &\leq \frac{\chi_1}{r + \chi_1 + \chi_2(1-\eta)} \int_{\varepsilon_l}^{\varepsilon_h} \left|V^A(\varepsilon', a, \chi) - V^B(\varepsilon', a, \chi)\right| dF(\varepsilon') \\ &\quad + \frac{\chi_2(1-\eta)}{r + \chi_1 + \chi_2(1-\eta)} \left|V^A(\varepsilon, a^*, \chi) - V^B(\varepsilon, a^*, \chi)\right|,\end{aligned}$$

where $a^* = \arg\max_{a' \in [-M,M]} \left\{ V^A(\varepsilon, a', \chi) - a'P \right\}$. Following the same steps, one can also show

$$\left(OV^B - OV^A\right)(\varepsilon, a, \chi) \leq \frac{\chi_1}{r + \chi_1 + \chi_2(1-\eta)} \int_{\varepsilon_l}^{\varepsilon_h} \left| V^A(\varepsilon', a, \chi) - V^B(\varepsilon', a, \chi) \right| dF(\varepsilon') + \frac{\chi_2(1-\eta)}{r + \chi_1 + \chi_2(1-\eta)} \left| V^A(\varepsilon, a^*, \chi) - V^B(\varepsilon, a^*, \chi) \right|.$$

Thus,

$$\left| \left(OV^A - OV^B\right)(\varepsilon, a, \chi) \right| \leq \frac{\chi_1}{r + \chi_1 + \chi_2(1-\eta)} \int_{\varepsilon_l}^{\varepsilon_h} \left| V^A(\varepsilon', a, \chi) - V^B(\varepsilon', a, \chi) \right| dF(\varepsilon') + \frac{\chi_2(1-\eta)}{r + \chi_1 + \chi_2(1-\eta)} \left| V^A(\varepsilon, a^*, \chi) - V^B(\varepsilon, a^*, \chi) \right|.$$

In turn,

$$\sup_{(\varepsilon,a) \in [\varepsilon_l, \varepsilon_h] \times [-M,M]} \left| \left(OV^A - OV^B\right)(\varepsilon, a, \chi) \right| \leq \frac{\chi_1 + \chi_2(1-\eta)}{r + \chi_1 + \chi_2(1-\eta)} \sup_{(\varepsilon,a) \in [\varepsilon_l, \varepsilon_h] \times [-M,M]} \left| \left(V^A - V^B\right)(\varepsilon, a, \chi) \right|,$$

and finally,

$$\sup_{x \in [\varepsilon_l, \varepsilon_h] \times [-M,M] \times [0,j]^2} \left| \left(OV^A - OV^B\right)(x) \right| \leq \frac{j + j(1-\eta)}{r + j + j(1-\eta)} \sup_{x \in [\varepsilon_l, \varepsilon_h] \times [-M,M] \times [0,j]^2} \left| \left(V^A - V^B\right)(x) \right|,$$

which more compactly can be written as

$$d_j\left(OV^A, OV^B\right) \leq \beta d_j\left(V^A, V^B\right),$$

where

$$\beta \equiv \frac{j + j(1-\eta)}{r + j + j(1-\eta)},$$

which concludes the proof. ■

Now we are ready to apply Corollary 2.1 of Martins-da Rocha and Vailakis (2010) to argue that O has a unique fixed point in $C(\mathcal{T})$. Lemma 2 shows that $C(\mathcal{T})$ is sequentially σ-complete. Lemma 3 shows that $O : C(\mathcal{T}) \to C(\mathcal{T})$ is a 0-local contraction. Hence, Corollary 2.1 of Martins-da Rocha and Vailakis (2010) implies that O has a unique fixed point in $C(\mathcal{T})$.

We next follow a guess-and-verify approach to determine the *unique* value function V. We conjecture that

$$V(\varepsilon, a, \chi) = \kappa_0(\varepsilon, \chi) + \kappa_1(\varepsilon, \chi) u(a) + \kappa_2(\chi) a,$$

where $\kappa_0(\varepsilon, \chi)$, $\kappa_1(\varepsilon, \chi)$, and $\kappa_2(\chi)$ are the coefficients to be determined. Because $D(\varepsilon, \chi)$ does not interact with the asset position, it does not affect the terms of trade as will be clear shortly. Thus, we primarily focus on determining $\kappa_1(\varepsilon, \chi)$ and $\kappa_2(\chi)$.

One can apply the envelope theorem to the auxiliary HJB equation (3.8) to obtain:

$$[r + \chi_1 + \chi_2(1-\eta)] V_2(\varepsilon, a, \chi) = \varepsilon u'(a) + \chi_1 \int_{\varepsilon_l}^{\varepsilon_h} V_2(\varepsilon', a, \chi) dF(\varepsilon') + \chi_2(1-\eta) P,$$

where $V_2(\cdot, \cdot, \cdot)$ refers to the derivative with respect to the second argument. By using our conjectured value function and matching coefficients, we obtain

$$\kappa_2(\chi) = \frac{\chi_2(1-\eta)}{r + \chi_2(1-\eta)} P$$

and

$$\kappa_1(\varepsilon, \chi) = \frac{1}{r + \chi_2(1-\eta)} \frac{(r + \chi_2(1-\eta))\varepsilon + \chi_1 \bar{\varepsilon}}{r + \chi_1 + \chi_2(1-\eta)},$$

which complete the proof.

A.2 Proof of Proposition 2

With a change of variable $a = a^\star(\tilde{\varepsilon}, \chi)$, we re-write the stationarity condition (3.4) as

$$\chi_1 F(\varepsilon) \int_{\varepsilon}^{\varepsilon_h} \Phi_\chi(d\varepsilon', \tilde{\varepsilon}) + \chi_2 \int_{\varepsilon_l}^{\min\{\varepsilon, \tilde{\varepsilon}\}} \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi\left(d\varepsilon', d\tilde{\tilde{\varepsilon}}\right) = \chi_1(1 - F(\varepsilon)) \Phi_\chi(\varepsilon, \tilde{\varepsilon}) + \chi_2 \Phi_\chi(\varepsilon, \tilde{\varepsilon}),$$

where $(\varepsilon, \tilde{\varepsilon})$ refers to the individual state of the client who is currently of type ε but holding the target position associated with type $\tilde{\varepsilon}$. We take the differential of both sides with respect to ε:

$$\begin{aligned}\chi_1 dF(\varepsilon) \int_{\varepsilon}^{\varepsilon_h} \Phi_\chi(d\varepsilon', \tilde{\varepsilon}) - \chi_1 F(\varepsilon) \Phi_\chi(d\varepsilon, \tilde{\varepsilon}) + \mathbb{I}_{\{\varepsilon \le \tilde{\varepsilon}\}} \chi_2 \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi\left(d\varepsilon, d\tilde{\tilde{\varepsilon}}\right) \\ = -\chi_1 dF(\varepsilon) \Phi_\chi(\varepsilon, \tilde{\varepsilon}) + \chi_1(1 - F(\varepsilon)) \Phi_\chi(d\varepsilon, \tilde{\varepsilon}) + \chi_2 \Phi_\chi(d\varepsilon, \tilde{\varepsilon}).\end{aligned}$$

After rearranging,

$$\chi_1 dF(\varepsilon) \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi(d\varepsilon', \tilde{\varepsilon}) + \mathbb{I}_{\{\varepsilon \leq \tilde{\varepsilon}\}} \chi_2 \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi\left(d\varepsilon, d\tilde{\tilde{\varepsilon}}\right) = (\chi_1 + \chi_2) \Phi_\chi(d\varepsilon, \tilde{\varepsilon}). \tag{A.3}$$

Aggregating over $\varepsilon \in [\varepsilon_l, \varepsilon_h]$,

$$\chi_1 \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi(d\varepsilon', \tilde{\varepsilon}) + \chi_2 \int_{\varepsilon_l}^{\tilde{\varepsilon}} \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi\left(d\varepsilon, d\tilde{\tilde{\varepsilon}}\right) = (\chi_1 + \chi_2) \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi(d\varepsilon, \tilde{\varepsilon}).$$

After cancellations, this equation holds if and only if

$$\int_{\varepsilon_l}^{\tilde{\varepsilon}} \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi\left(d\varepsilon, d\tilde{\tilde{\varepsilon}}\right) = \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi(d\varepsilon, \tilde{\varepsilon}).$$

The LHS is equal to $F(\tilde{\varepsilon})$ because χ and ε are independently distributed in the cross section of clients. Then,

$$F(\tilde{\varepsilon}) = \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi(d\varepsilon, \tilde{\varepsilon}). \tag{A.4}$$

In words, this equality means that in the stationary equilibrium, the fraction of type-χ clients who have a current taste type of $\tilde{\varepsilon}$ or lower is equal to the fraction of type-χ clients whose current asset holding is associated with $\tilde{\varepsilon}$ or lower. This "symmetry" between the marginal distribution of taste types and the marginal distribution of asset holdings obtains in the original Lagos and Rocheteau (2006) environment as well (p. 13). More specifically, the Lagos and Rocheteau (2006) result is a special case of ours because (i) we have an arbitrary distribution of taste types that may be discrete, continuous, or mixed, while Lagos and Rocheteau (2006) allowed only for discrete distribution and (ii) we have heterogeneity in χ, and so, our equilibrium object is a *conditional* distribution, while in the Lagos and Rocheteau (2006) environment all clients have the same χ.

We next substitute (A.4) into (A.3),

$$\chi_1 dF(\varepsilon) F(\tilde{\varepsilon}) + \mathbb{I}_{\{\varepsilon \leq \tilde{\varepsilon}\}} \chi_2 dF(\varepsilon) = (\chi_1 + \chi_2) \Phi_\chi(d\varepsilon, \tilde{\varepsilon}).$$

After rearranging and integrationg over $\varepsilon \in [\varepsilon_l, \varepsilon]$,

$$\chi_1 F(\varepsilon) F(\tilde{\varepsilon}) + \chi_2 F(\min\{\varepsilon, \tilde{\varepsilon}\}) = (\chi_1 + \chi_2) \Phi_\chi(\varepsilon, \tilde{\varepsilon}).$$

After reverting our initial change of variable, the formula in the proposition obtains.

A.3 Proof of Proposition 3

Define

$$A^{d}(P) \equiv \iint\limits_{\mathbb{R}_{+}^{2}} \int_{\varepsilon_l}^{\varepsilon_h} (u')^{-1}\left[\frac{r+\chi_1+\chi_2(1-\eta)}{(r+\chi_2(1-\eta))\varepsilon+\chi_1\bar{\varepsilon}} rP\right] dF(\varepsilon)\, dG(\chi).$$

An application of the inverse function theorem implies that A^d is strictly decreasing because u is strictly concave. Hence, the market-clearing interdealer price P such that $A^d(P) = A$ is unique whenever it exists. From the standard Inada limit conditions on u stated in the proposition, $A^d(0) = \infty$ and $A^d(\infty) = 0$ and hence the existence of $P \in (0, \infty)$ is guaranteed by the intermediate value theorem.

A.4 Proof of Lemma 1

When $\gamma = 1$, $\Delta_u(\chi) = \log[\chi_1 + 2x(\chi_2)] - \log(\chi_1)$, where $x(\chi_2) \equiv r + \chi_2(1-\eta)$. It is easy to verify that $\Delta_u(\chi)$ is strictly decreasing in χ_1, and so is $\Delta_u(\chi)/[r+\chi_2(1-\eta)]$. To complete the proof, we need to show that $\{\log[\chi_1 + 2x(\chi_2)] - \log(\chi_1)\}/x(\chi_2)$ is strictly decreasing in χ_2. It suffices to show that

$$\frac{2x(\chi_2)}{\chi_1 + 2x(\chi_2)} < \log[\chi_1 + 2x(\chi_2)] - \log(\chi_1).$$

By taking the exponential of both sides and rearranging, it suffices to show that

$$e^{\frac{2x(\chi_2)}{\chi_1+2x(\chi_2)}} - \frac{2x(\chi_2)}{\chi_1} < 1.$$

The LHS and the RHS are equal to each other when $\chi = 0$. Thus, it suffices to show that the LHS is strictly decreasing. That is, we need to show that the first derivative of the LHS is negative. Equivalently,

$$\frac{1}{[\chi_1 + 2x(\chi_2)]^2} e^{\frac{2x(\chi_2)}{\chi_1+2x(\chi_2)}} - \frac{1}{\chi_1^2} < 0.$$

Again, the LHS and the RHS are equal to each other when $\chi = 0$. Thus, the proof is done if the LHS is strictly decreasing, which is obviously true because its first derivative is negative:

$$-e^{\frac{2x(\chi_2)}{\chi_1+2x(\chi_2)}} \frac{\chi_1 + 4x(\chi_2)}{[\chi_1 + 2x(\chi_2)]^4} < 0.$$

A.5 Proof of Proposition 4

The formula for $\mathbb{W}^{Eq}$ follows easily by substituting (3.19), (3.16) and $u(a) = \frac{a^{1-\gamma}}{1-\gamma}$ into (3.21).

To calculate $\mathbb{W}^{FB}$, we solve the problem of an unconstrained planner. Because we are in our special case, the planner's problem is:

$$\mathbb{W}^{FB}\left(\varepsilon_l\right)=\max_{a(0),a(2\sigma)}\frac{\varepsilon_l}{r}\frac{\left[a\left(0\right)\right]^{1-\gamma}}{1-\gamma}\frac{1}{2}+\frac{2\sigma}{r}\frac{\left[a\left(2\sigma\right)\right]^{1-\gamma}}{1-\gamma}\frac{1}{2}-\frac{\sigma}{r}\frac{A^{1-\gamma}}{1-\gamma}, \tag{A.5}$$

subject to

$$\frac{1}{2}\left(a\left(0\right)+a\left(2\sigma\right)\right)=A,$$

$$-M\leq a\left(0\right)\leq M,$$

$$-M\leq a\left(2\sigma\right)\leq M.$$

In our special case $\varepsilon_l=0$ for simplicity, but setting up the planner's problem directly for this special case would be unnatural and create "discontinuities" with respect to frictional welfare measures, because the planner would not care for the asset position of the low-type clients. But this is true only for $\varepsilon_l=0$, and the planner would have interior solution for low-type clients for any other $\varepsilon_l>0$. Thus, we first solve the planner's problem for a generic $\varepsilon_l>0$. Then, we calculate the limit $\mathbb{W}^{FB}\equiv\lim_{\varepsilon_l\to 0}\mathbb{W}^{FB}\left(\varepsilon_l\right)$.

Because M is an arbitrarily large number that can be chosen to allow for interior solutions, we are fine as long as $a\left(0\right),a\left(2\sigma\right)<\infty$. Because the second term on the RHS of (A.5) does not depend on the control variables, the Lagrangian is

$$\mathcal{L}^{FB}=\frac{\varepsilon_l}{2r}\frac{\left[a\left(0\right)\right]^{1-\gamma}}{1-\gamma}+\frac{\sigma}{r}\frac{\left[a\left(2\sigma\right)\right]^{1-\gamma}}{1-\gamma}+\mu\left(A-\frac{1}{2}\left(a\left(0\right)+a\left(2\sigma\right)\right)\right).$$

The FOCs for interior solution are

$$\frac{\varepsilon_l}{2r}\left[a\left(0\right)\right]^{-\gamma}-\frac{\mu}{2}=0$$

and

$$\frac{\sigma}{r}\left[a\left(2\sigma\right)\right]^{-\gamma}-\frac{\mu}{2}=0.$$

Then,

$$a\left(0\right)=\left(\frac{\varepsilon_l}{\mu r}\right)^{1/\gamma}$$

and

$$a\left(2\sigma\right)=\left(\frac{2\sigma}{\mu r}\right)^{1/\gamma}.$$

Substituting these into the resource constraint, the Lagrange multiplier is

$$\mu = \frac{1}{r}\left(\frac{\varepsilon_l^{1/\gamma}+(2\sigma)^{1/\gamma}}{2A}\right)^{\gamma},$$

which in turn implies

$$a\left(0\right) = 2A\frac{\varepsilon_l^{1/\gamma}}{\varepsilon_l^{1/\gamma}+(2\sigma)^{1/\gamma}}$$

and

$$a\left(2\sigma\right) = 2A\frac{(2\sigma)^{1/\gamma}}{\varepsilon_l^{1/\gamma}+(2\sigma)^{1/\gamma}}.$$

Substituting into (A.5),

$$\mathbb{W}^{FB}\left(\varepsilon_l\right) = \frac{\varepsilon_l}{2r}\frac{(2A)^{1-\gamma}}{1-\gamma}\left[\frac{\varepsilon_l^{1/\gamma}}{\varepsilon_l^{1/\gamma}+(2\sigma)^{1/\gamma}}\right]^{1-\gamma}+\frac{(2A)^{1-\gamma}}{1-\gamma}\frac{\sigma}{r}\left[\frac{(2\sigma)^{1/\gamma}}{\varepsilon_l^{1/\gamma}+(2\sigma)^{1/\gamma}}\right]^{1-\gamma}-\frac{\sigma}{r}\frac{A^{1-\gamma}}{1-\gamma}.$$

By taking the limit as $\varepsilon_l \to 0$, one obtains $\mathbb{W}^{FB}$ stated in the proposition.

What remains to show is the formula for $\mathbb{W}^{SB}$ in the proposition. Since the endogenous conditional asset holding distribution can be continuous as well as the exogenous distribution of ε and χ, we potentially have a continuum of control variables as in Üslü (2019) and Farboodi, Jarosch, and Shimer (2018); we follow these papers in appealing to van Imhoff (1982) and interpret the integrals in the objective function as summation over discrete intervals with lengths $d\varepsilon$ and da approaching zero.

Keeping in mind van Imhoff (1982)'s interpretation and because the second term of (3.22) does not depend on the control variables, the planner's current-value Hamiltonian can be written as

$$\begin{aligned}\mathcal{L}^{SB}\left(q|\Phi\right) = &\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\varepsilon u\left(a\right)\Phi_\chi\left(d\varepsilon,da\right)dG\left(\chi\right)\\
&+\chi_1\int_{\varepsilon_l}^{\varepsilon_h}\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\left(\vartheta\left(\varepsilon',a,\chi\right)-\vartheta\left(\varepsilon,a,\chi\right)\right)\Phi_\chi\left(d\varepsilon,da\right)dG\left(\chi\right)dF\left(\varepsilon'\right)\\
&+\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\chi_2\left\{\vartheta\left(\varepsilon,a+q\left(\varepsilon,a,\chi\right),\chi\right)-\vartheta\left(\varepsilon,a,\chi\right)\right\}\Phi_\chi\left(d\varepsilon,da\right)dG\left(\chi\right)\\
&-\zeta\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\chi_2 q\left(\varepsilon,a,\chi\right)\Phi_\chi\left(d\varepsilon,da\right)dG\left(\chi\right),\end{aligned}$$

$\vartheta(\varepsilon, a, \chi)$ denotes the current-value co-state variable associated with $\Phi_\chi(d\varepsilon, da)$; and ζ is the Lagrange multiplier associated with the condition (3.23).

The FOC for optimization is

$$\vartheta_2(\varepsilon, a + q(\varepsilon, a, \chi), \chi) = \zeta, \tag{A.6}$$

if $\chi > 0$, $\Phi_\chi(d\varepsilon, da) > 0$, and $dG(\chi) > 0$.

In any optimum q^e, the co-state variables must satisfy the ODEs,

$$\nabla_{n(\varepsilon,a,\chi)}\mathcal{L}^{SB}(q^e|\Phi) = r\vartheta(\varepsilon, a, \chi) - \dot{\vartheta}(\varepsilon, a, \chi), \tag{A.7}$$

where $n(\varepsilon, a, \chi)$ is the degenerate measure which puts all the probability on the type (ε, a, χ) and ∇_n denotes the Gâteaux differential in the direction of measure n:

$$\nabla_n\mathcal{L}^{SB}(q^e|\Phi) = \lim_{\epsilon\to 0}\frac{\mathcal{L}^{SB}(q^e|\Phi + \epsilon n) - \mathcal{L}^{SB}(q^e|\Phi)}{\epsilon}.$$

For small ϵ, we obtain up to second-order terms:

$$\begin{aligned}
\mathcal{L}^{SB}(q^e|\Phi + \epsilon n) - \mathcal{L}^{SB}(q^e|\Phi) = \epsilon\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\varepsilon u(a)\, n(d\varepsilon, da, d\chi) \\
+ \epsilon\chi_1\int_{\varepsilon_l}^{\varepsilon_h}\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}(\vartheta(\varepsilon', a, \chi) - \vartheta(\varepsilon, a, \chi))\, n(d\varepsilon, da, d\chi)\, dF(\varepsilon') \\
+ \epsilon\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\chi_2\left\{\vartheta(\varepsilon, a + q^e(\varepsilon, a, \chi), \chi) - \vartheta(\varepsilon, a, \chi)\right\} n(d\varepsilon, da, d\chi) \\
- \epsilon\zeta\int_0^\infty\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\chi_2 q^e(\varepsilon, a, \chi)\, n(d\varepsilon, da, d\chi).
\end{aligned}$$

Thus,

$$\begin{aligned}
\nabla_{n(\varepsilon,a,\chi)}\mathcal{L}^{SB}(q^e|\Phi) = \varepsilon u(a) + \chi_1\int_{\varepsilon_l}^{\varepsilon_h}(\vartheta(\varepsilon', a, \chi) - \vartheta(\varepsilon, a, \chi))\, dF(\varepsilon') \\
+ \chi_2\left\{\vartheta(\varepsilon, a + q^e(\varepsilon, a, \chi), \chi) - \vartheta(\varepsilon, a, \chi) - \zeta q^e(\varepsilon, a, \chi)\right\}.
\end{aligned}$$

Using (3.23), (A.7), and the FOC (A.6), the following ODE for the co-state variables obtains in any optimum:

$$r\vartheta(\varepsilon,a,\chi)-\dot{\vartheta}(\varepsilon,a,\chi)=\varepsilon u(a)+\chi_1\int_{\varepsilon_l}^{\varepsilon_h}\left(\vartheta(\varepsilon',a,\chi)-\vartheta(\varepsilon,a,\chi)\right)dF(\varepsilon')$$
$$+\chi_2\max_{q\in[-M,M]}\left\{\vartheta(\varepsilon,a+q,\chi)-\vartheta(\varepsilon,a,\chi)-\zeta q\right\}$$

s.t.

$$\int_0^{\infty}\int_{-M}^{M}\int_{\varepsilon_l}^{\varepsilon_h}\chi_2 q(\varepsilon,a,\chi)\,\Phi_\chi(d\varepsilon,da)\,dG(\chi)=0.$$

Checking that the planner's optimality conditions do not coincide with the equilibrium conditions is easy. More specifically, the comparison with (3.8) reveals that the planner's optimality conditions and the equilibrium conditions would be identical if $\eta=0$ and $P=\zeta$ in the equilibrium condition, which means that the efficiency implications of LR apply to our generalized setup as well. It also means $\mathbb{W}^{SB}$ stated in the proposition obtains by substituting $\eta=0$ into $\mathbb{W}^{Eq}$.

B Data Construction

The primary data source for our baseline estimation is the ZEN dataset which was the UK's transaction reporting system administered by the Financial Conduct Authority during our sample period 2011m8-2017m12. There is no public version of ZEN, which is why this dataset has been used relatively rarely in the academic literature (recent exceptions include Benos and Zikes (2018), Czech, Huang, Lou, and Wang (2021), and Kondor and Pintér (2022)). The structure of the ZEN dataset is similar to the TRACE dataset often used to study the US corporate bond market, with the important exception that almost all trade reports in the ZEN include the identities of the counterparties. (See Ivanov, Orlov, and Schihl (2021) for a recent comparison between the ZEN dataset and the TRACE dataset, using a common set of corporate bonds traded in both the UK and US.)

All secondary market trades are reported in the ZEN dataset, where at least one of the counterparties is an FCA-regulated entity. We drop duplicate trade reports as well as trade reports with missing client identifiers. We also exclude trades with inter-dealer brokers (IDBs) as well as trades of less than £1,000 in par value, and remove trades with erroneous price entries. We end up with 1,277,555 and 1,231,507 observations for the sterling corporate and government bond markets, respectively, covering 57 nominal government bonds and 2796 corporate bonds. From our sample of clients, we are able to identify 574 clients who trade in both markets, leaving us with 1,225,347 and 1,222,272 observations in the two markets. About half of these clients can be classified as hedge funds and asset managers, and the remaining half comprises of pension funds, insurance companies, commercial banks, international organizations and others. From the sample of corporate bonds, we identify the 57 most frequently traded bonds that have an investment grade rating according to Moody's.[3] This leaves us with 186,832 observations for corporate bonds, covering 1614 trading days.

We use these two samples of government and corporate bonds (each including trades for 57 assets) to compute empirical moments related to average trade size, average client intensity, average absolute intensity deviation and market turnover. Days in which a client does not trade enters the intensity computations as zeros. To compute price dispersion, we keep trading hours with at least two transactions occurring in that time window according to the time stamp of the trade report. This leaves us with 1,037,595 and 77,858 transactions in the government and corporate bond markets, respectively. To obtain daily measures of price dispersion, we first compute (normalized) absolute deviations from the hourly average transaction price, and then average within the day these deviations using the size of the trade as weights.

C On the Empirical Relation between Clients' Execution Costs and Trading Intensity

In Section 3.4.1 of the main text, we argued that one of the implications of our model is that a client who trades more frequently faces a smaller bid-ask spread. This feature is consistent with the empirical evidence of O'Hara, Wang, and Zhou (2018) that used the TRACE dataset on a subset of US corporate bond transactions.

This appendix provides additional empirical evidence on the relation between clients' execution costs and their trading intensity using our datasets for the UK government and corporate bond markets. We adopt the framework of O'Hara, Wang, and Zhou (2018) which studied this relationship on a subset of the US corporate bond market which includes trades by insurance companies. They find that an insurance company who trades in a similar size, on the same side (i.e buying or selling), the same bond on the same day with the same dealer will face a more favorable execution cost if the given company is a more active trader than if it is a less active trader. We revisit this result in the context of the UK bond markets with an emphasis on exploring whether there are

[3] Information on corporate bond ratings is from Thomson Reuters Eikon, which covers the three major rating agencies Moody's, Standard & Poor's (S&P), and Fitch. Ratings of Moody's are used as the default option because of the firm's large market coverage. S&P ratings are used if ratings from Moody's are not available for the given bond. Fitch ratings are used as a third option.

any significant differences in the cost-intensity relationship across government and corporate bond markets.

To determine whether clients are more or less active, we use the last six months of clients' trading volume to place clients into two groups.[4] More active clients are in the top quintile of the distribution, while the rest of the clients are classified as less active.[5] The sample we use in this exercise is the same as the one used in the estimation of our structural model.

Given that execution costs vary with the direction of trades, we first divide our sample of trades into buys and sales. To control for the effect of trade size, we then divide buys and sales into three trade-size categories based on the nominal value of trades: small (£1–£100,000), medium (£100,000–£1,000,000) and large (above £1,000,000).[6] For each trade-size category, in each bond, on each day and for buy and sell trades separately, we then compute the average prices at which more active and less active clients traded.

If execution costs were similar across more and less active traders, then this measure of price difference should not be significantly different from zero. The baseline estimates of O'Hara, Wang, and Zhou (2018) (reported in Table 2 of their paper) show a price difference of –0.17% for buys and 0.36% for sales, based on their sample of insurance companies on the US corporate bond markets. They argue that search and intermediation frictions are likely explanations behind these price differences.

It is an interesting question on its own right whether these effects are present in other markets (such as the government bond market) and whether they are specific only to insurance companies. Our findings, based on a sample more representative of the trading universe, are summarised in Table 14. The price difference in our full sample of government bonds is -0.0183% for buys and 0.0193% for sales; whereas the estimates using the full baseline sample of corporate bonds is -0.0627% for buys and 0.0601% for sales. Comparing our estimates for the corporate bond market to those of O'Hara, Wang, and Zhou (2018) reveals that our estimates are significantly smaller. This is likely driven by the fact that in our baseline sample of corporate bonds we exclude bonds that are infrequently traded in order to have a set of assets that are comparable to government bonds in terms of payoff and adverse selection risk.

Comparing our estimates across the government and corporate bond markets reveals that the

[4] Using past trading volume as a sorting variable is motivated by Section 4.3.2. of O'Hara, Wang, and Zhou (2018), which uses trading volume as a robustness check. Their baseline sorting variable is clients' bond portfolio, which is not available in our sample.

[5] This cutoff for classification delivers comparable groups (in terms of number of transactions) of more active and less active clients in both markets.

[6] Note that O'Hara, Wang, and Zhou (2018) uses four size categories with the third and fourth categories being "Round-lot" ($1,000,000–$5,000,000), and "Block" (above $5,000,000). Given the smaller trade sizes in the UK, we would not have sufficient number of observations to estimate relative execution costs in the fourth size category.

Table 14: Differences in Execution Costs for More Active and Less Active Clients

	Government Bonds			Corporate Bonds		
	PriceDiff	p-value	Nobs	PriceDiff	p-value	Nobs
Buys						
Full sample by trade-size category	-0.0183%	0.000	51584	-0.0627%	0.000	9681
£1–£100,000	-0.0251%	0.000	20510	-0.0701%	0.001	5073
£100,000–£1,000,000	-0.0142%	0.000	23574	-0.0210%	0.052	1513
above £1,000,000	0.0013%	0.565	25478	-0.0121%	0.349	311
Sells						
Full sample by trade-size category	0.0193%	0.000	51533	0.0601%	0.000	12494
£1–£100,000	0.0398%	0.000	20895	0.0729%	0.000	7471
£100,000–£1,000,000	0.0107%	0.000	25400	0.0271%	0.004	1468
above £1,000,000	0.0006%	0.803	25323	0.0280%	0.174	285

Notes: The table presents the differences in execution costs for more active and less active traders. A client is classified as more active (less active) if its last six month of trading volume is inside (outside) the top quintile of the client distribution. The sample ,covering the period of 2011m8–2017m12, is the same as the one used in the structural estimation of our baseline model. To obtain the price difference estimates, we first separate our full sample of bond trades into clients' buys and sells. Within each subsample, we then classify each trade into three size categories based on its pound transaction value: small (£1 to £100,000), medium (£100,000 to £1,000,000) and large (above £1,000,000. Then for each bond, trading day, and size category, we calculate the average price (as a percentage of the par) for more active traders and for less active traders, and compute the difference. The p-values are based on standard errors that are clustered at the bond level.

price difference is about three times larger for corporate bonds in the full sample. This is indicative of more severe trading frictions in the corporate bond market compared to the government bond market. Examining how the price difference varies across the trade size distribution reveals that the majority of the effect in the full sample concentrates in the lower segment of the trade size distribution. We find that the price difference is statistically insignificant for the largest size category, consistent with the evidence by O'Hara, Wang, and Zhou (2018) for the US.

D Two-step GMM

Our baseline results are obtained using a one-step GMM estimator. This is suitable in our framework, given that all our moments are approximately in percentage deviations, so assigning equal weights implies invariance to changes in units. The Monte Carlo evidence in Section 4.5 also confirms that a one-step GMM is appropriate in recovering the structural parameters of our model (for a reasonable set of parameters). As a robustness check, this section reports how much the baseline parameter estimates would change if we used a two-step GMM estimator instead.

We follow recent practice (Gayle and Shephard, 2019; Honore, Jorgensen, and de Paula, 2020) and choose the weighting matrix to be a diagonal matrix, whose elements are the inverse of the diagonal variance-covariance matrix of the empirical moments. This is to avoid the well-known problems with using the optimal weighting matrix, which can have poor small-sample properties

Table 15: Parameter Estimates: Two-step GMM

	Government Bonds (1)	Corporate Bonds (2)
γ – Curvature of the utility function	18.02	18.02
δ – Exponential distn. parameter for $G(\chi_1)$	22.05	123.84
λ – Matching efficiency	3836.16	186.07
η – Dealers' market power	0.95	0.27

Notes: this table reports the point estimates of the parameters using a two-step GMM estimation. As weighting matrix, we use inverse of the diagonal of the variance-covariance matrix of the empirical moments implied by our baseline point estimates using a one-step GMM (Gayle and Shephard, 2019; Honore, Jorgensen, and de Paula, 2020).

(Altonji and Segal, 1996). Table 15 reports the point estimates from the two-step GMM estimation. The results are similar to our baseline estimates implied by the one-step GMM (Table 2).

E Sensitivity Analysis

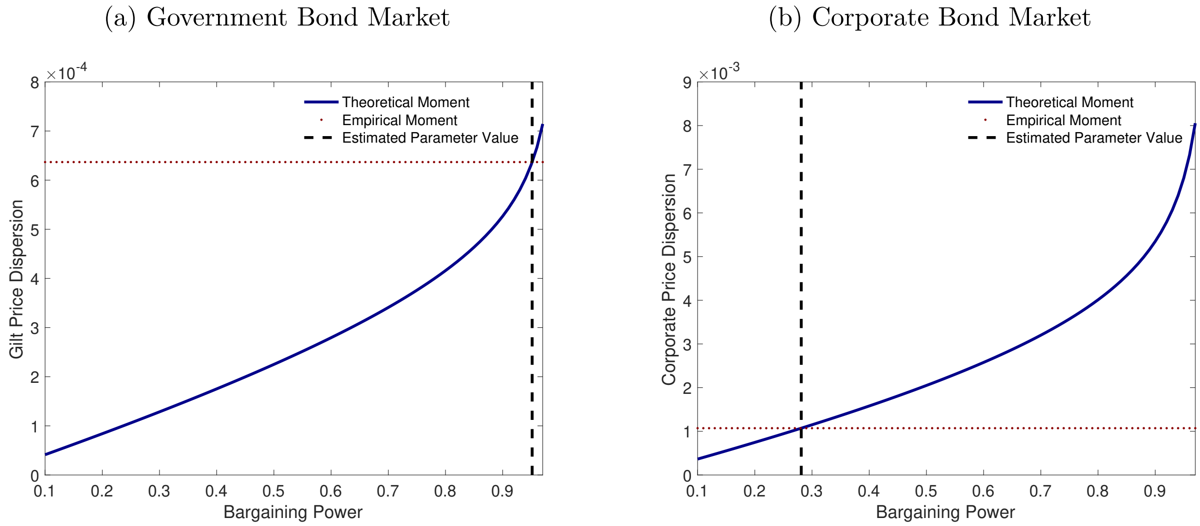
E.1 Price Dispersion and Dealers' Bargaining Power

An interesting result implied by our structural estimation is the large estimated bargaining power that dealers in the government bond market ($\eta_g = 0.95$) have compared to the corporate bond market ($\eta_c = 0.28$). As argued in Section 4.3, empirical measures of price dispersion is a key source of variation in the data that helps identify dealers' bargaining power in the model. While price dispersion in the corporate bond market is almost twice as high as in the government bond market (11 bps vs 6.4 bps), this difference is dwarfed by the difference in aggregate turnover (about 2.1 log points) across the two markets. The model's quantitative interpretation of these cross-market differences leads to the higher estimated bargaining power of dealers in government bonds.

To further explore the relationship between bargaining power and price dispersion, we ask: how much would price dispersion need to change in the model such that dealers' bargaining power in the government bond market would fall below that in the corporate bond market. To answer this question, Figure 4 shows the counterfactual levels of price dispersion in both markets as we change dealers' bargaining power in the given market, while keeping all other parameters fixed.

Inspecting the solid blue line in panel (a) of Figure 4 reveals that price dispersion in government bonds would have to fall from the observed 6.4 bps (horizontal dotted line) to about 1 bp, so that the model-implied bargaining power of dealers is reduced from the estimated value of 0.95 (vertical dash line) to 0.28 - the level estimated for corporate bonds. Using a similar reasoning, panel (b) of Figure 4 shows that price dispersion in the corporate bond market would have to rise from 11 bps to more than 60 bps so that the model-implied bargaining power in corporate bonds would

Figure 4: Price Dispersion and Dealers' Bargaining Power



Notes: The blue solid lines in the left and right panels show the counterfactual measures of price dispersion in the government and corporate bond markets, respectively, as we change dealers' bargaining power from the estimated values (marked by the vertical dashed black lines) while keeping all other parameters fixed. The horizontal dotted lines show the empirical levels of price dispersion computed from the micro-data covering the period 2011m8-2017m12.

be comparable to that in the government bond market.

Overall, the comparative statics of Figure 4 suggests that in our structural framework and given the constellation of empirical moments, it is unlikely that the result regarding dealers' relative bargaining power across the two markets could easily be reversed.

E.2 The Effects of Price Dispersion Measurement

In our baseline estimation, we measure price dispersion as the absolute deviation of transaction prices from a benchmark price. As benchmark, we used the hourly average transaction price of the given bond. The literature often uses benchmark prices that vary at the daily frequency (e.g. Jankowitsch, Nashikkar, and Subrahmanyam (2011) and Üslü and Velioğlu (2019)). Our choice of using higher-frequency (hourly instead of end-of-day) benchmarks aims to mitigate the over-estimation of dispersion due to the gradual arrival of news in the market, which might dominate price dispersion measures using a benchmark price with lower-frequency variation.

In this section, we check how our baseline estimation changes if we used the daily average transaction price as a benchmark price when measuring price dispersion in the corporate and government bonds. Measured price dispersion in the government bond market would increase from 0.00064 to 0.00127, while it would increase from 0.0011 to 0.0016 in the corporate bond market. That is, measured price dispersion in government bond markets would almost double, whereas the corresponding moment in the corporate bond market would increase by almost 50%

compared to the values used in the baseline estimation. The question remains about the extent to which these dramatic changes in the empirical moments would change the parameter estimates and the associated welfare results. To answer this, we re-estimate the model after modifying the price dispersion moments in both markets.

Table 16: Parameter Estimates

	Government Bonds	Corporate Bonds
	(1)	(2)
γ – Curvature of the utility function	9.42	9.42
δ – Exponential distn. parameter for $G(\chi_1)$	22.59	116.56
λ – Matching efficiency	2483.26	13.20
η – Dealers' market power	0.99	0.23

Notes: this table reports the point estimates of the parameters using a one-step GMM estimation. Compared to our baseline set of empirical moments, price dispersion is set to be 0.00127 and 0.0016 in the government and the corporate bond market, respectively.

The new parameter estimates are presented in Table 16. The empirical fit of the re-estimation is virtually identical to that of our baseline. The re-estimation is able to match the increased price dispersion moments by lowering the estimated matching efficiency parameters in both markets (λ_g and λ_c). The estimate for parameter λ_g falls from 4124 to 2483 and λ_c falls from 233 to 13.2. That is, the model's re-estimation interprets the changed dispersion moments as clients's value function becoming less stable in the time-varying arguments (asset position and taste) which increases transaction surpluses, leading to a rise in price dispersion. Moreover, the partial effect of reducing the matching efficiency parameters is to reduce turnover in both markets as well. Given that the quantity moments are unchanged during this re-estimation, a reduction in the estimated curvature parameter γ (from 17.9 to 9.4) is used to fit the data.

We next compute the total welfare loss under the new constellation of parameters. The results in Table 17 show a sizeable increase in welfare loss compared to our baseline specification. Total welfare loss in the government and corporate bond markets increases to 27% and 43%, respectively, from the values 7.8% and 12% implied by our baseline. Now the source of welfare loss in the government bond market tilts significantly towards dealer market power, while search frictions and dealer market power split the welfare loss in half in our baseline estimation. This reveals that although dealers' market power naturally increases by small percentage points (from 95% to 99%), its welfare impact is significant.

Overall, these results highlight the importance of price dispersion measurement when estimating quantitative models of OTC markets and associated welfare losses. The ideal empirical counterpart of our theoretical price dispersion measure would compare clients who would trade the same bond at the same time. Our choice of using hourly average prices around which to compute

Table 17: Estimated Welfare Losses

	Government Bonds (1)	Corporate Bonds (2)
Total Welfare Loss	27.423%	42.954%
Due to Search Frictions	6.963%	42.646%
Due to Dealers' Market Power	20.46%	0.30844%

Notes: the table reports summary statistics for trading delays (upper panel) and dealers' bargaining power (lower panel), implied by the theoretical model evaluated at the estimated parameter values (as reported in Table 16). Trading delays are expressed as a fraction of a trading day.

price dispersion aims to come close to this ideal. Reducing the time window further would be possible in the government bond market, where assets are traded frequently enough to carry out such a calculation. However, we face data limitations in the corporate bond market, where assets are not traded frequently enough to estimate price dispersion based on time windows less than an hour.

F Further Details on the Dealer Sector

This section provides some institutional background on the dealer sector in the UK government and corporate bond markets. We also present novel stylized facts on the role of the dealer sector in providing liquidity for clients in both markets. The institutional background as well as the stylized facts are aimed at motivating some of the modeling choices we make in our structural estimation; they also motivate why we focus on a common set of dealers in our empirical design for cross-market comparison.

Specifically, the dealer sector in our model is taken to be exogenous and we do not model the endogenous nature of intermediation as Üslü (2019) and others. This is consistent with the empirical observation that market making in the UK government bond market is performed by designated primary dealers, also known as gilt-edged market makers (GEMMs). Their number hovers around 20 during our sample. GEMMs have a number of obligations and enjoy certain privileges. A quick summary of these based on BoE (1997); DMO (2011, 2021) is as follows. There are three main obligations of GEMMs: (i) they are obliged to make on demand continuous two-way prices, thereby providing continuous liquidity for clients; this also includes maintaining some minimum level of secondary market share; (ii) they are expected to plan an active role in primary auctions and to purchase a minimum level of new issuance; (iii) they are obliged to provide the DMO with data and information as well as market intelligence. In return, the GEMMs enjoy certain privileges such as (i) the eligibility to submit competitive bids directly to the DMO and to purchase a specified fraction of gilts on offer at auctions; (ii) preferred counterparty status (e.g. DMO would deal directly only with GEMMs); (iii) access to the to the services of Inter-Dealer Brokers (IDBs).

Table 18: Intermediation by Primary Dealers

Government Bonds				
	Volume		Number of Transactions	
	billion £s	%	n	%
Client Trades (2011m8-2017m12)	11,431.5	100%	1,439,093	100%
Primary Dealer as Counterparty	11,250.8	98.4%	1,277,555	88.8%
Other Counterparty	180.6	1.6%	161,538	11.2%
Corporate Bonds				
Client Trades (2011m8-2017m12)	1944.8	100%	1,961,510	100%
Primary Dealer as Counterparty	1644.9	84.5%	1,231,507	62.8%
Other Counterparty	299.9	15.5%	730,003	37.2%

Notes: The table presents summary statistics based on the trade-level ZEN dataset for the government and corporate bonds markets covering the sample 2011m8-2017m12. In the sample we include all trades where at least one of the counterparties is a client. These client trades are then split into two subsamples: one where the other counterparty is a primary dealer, also known as a gilt-edged market marker (GEMMs), and another subsample where the other counterparty is not a primary dealer. All identified clients are included in the construction of the table, i.e. not only those that trade in both markets, as in our baseline. All assets are included in the computations, i.e., 57 assets in the government bond market and 2796 assets in the corporate bond market. Further information on GEMMs can be found on the website of the Debt Management Office https://www.dmo.gov.uk/responsibilities/gilt-market/market-participants/.

While GEMMs do not have the same status in the sterling corporate bond market as they have in the gilt market, they continue to play an important role in providing liquidity for corporate bond traders. To provide some stylized facts on this, Table 18 reports the fraction of all available client trades in our sample that are intermediated by GEMMs and by non-GEMMs. In terms of trading volume, we find that about 98% of the trades are intermediated by GEMMs in the government bond market, while this share amounts to a still sizable 85% in the corporate bond market. In terms of number of transactions, GEMMs intermediate about 89% and 63% of trades in the government and corporate bond markets, respectively.[7] The lower share of GEMMs in corporate bond intermediation is suggestive of GEMMs facing relatively fiercer market-making competition from non-GEMMs (e.g. other large clients) in this market. This is consistent with the result, implied by the structural estimation, that dealers have lower market power in the corporate bond market compared to the government bond market.

Overall, the dominant role of GEMMs in the provision of liquidity in both markets and their stable presence in market making are a primary reason why we propose a theoretical framework with exogenous intermediation structure and why we fix the identities of dealers (in addition to

[7]Note that the lower share of dealer intermediation when measured in number of transactions is indicative of non-GEMM counterparties playing an increased role of intermediating smaller sized trades. This is consistent with the extending role of trading platforms in bond markets (Bessembinder, Spatt, and Venkataraman, 2020).

fixing the identities of clients) when comparing the two markets.