

Comparing Search and Intermediation Frictions Across Markets*

Gábor Pintér[†] Semih Üslü[‡] Jean-Charles Wijnandts[§]

Abstract

We develop a two-asset search-and-bargaining model of OTC trading to estimate frictions and welfare losses in the UK government and corporate bond markets. Using transaction-level data and a matched client sample, we find that both trading delays and intermediation frictions are more pronounced in corporate bonds. Welfare losses due to these frictions are 2.4% in government bonds and 5.0% in corporate bonds—driven primarily by trading delays. Using data from the COVID-19 crisis, we find that these losses might more than double 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, the editor Dimitris Papanikolaou, one anonymous associate editor, one anonymous referee, Hengjie Ai, Rui Albuquerque, Ana Babus, Briana Chang, Alex Chinco, Darrell Duffie, Mark Egan, Peter Feldhütter, Thanasis Geromichalos, Foti Grigoris, Joel Hasbrouck, Burton Hollifield, Edith Hotchkiss, Yunzhi Hu, Karam Kang, Tommy Leung, Weiling Liu, Liang Ma, Joshua Mollner, John Power, Batchimeg Sambalaibat, Andriy Shkilko, Fabricius Somogyi, Zhaogang Song, Davide Tomio, Fabrice Tourre, Nicholas Trachter, Pierre-Olivier Weill, Milena Wittwer, Randall Wright, and Liyan Yang, and audience at various seminars and conferences. The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England, the Bank for International Settlements, or their committees.

[†]Monetary and Economic Department, the Bank for International Settlements, Centralbahnplatz 2, 4051 Basel, Switzerland, e-mail: gabor.pinter@bis.org

[‡]Carey Business School, Johns Hopkins University, 100 International Drive, Baltimore, MD 21211, USA, e-mail: semihuslu@jhu.edu

[§]Financial Stability Strategy and Risk, the Bank of England, Threadneedle Street, London, EC2R 8AH, UK, e-mail: jean-charles.wijnandts@bankofengland.co.uk

“... 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 provide quantitative answers to these questions, we estimate a novel 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 2.38% relative to a frictionless benchmark. In contrast, search frictions in the UK corporate bond market are more pronounced, generating a welfare loss of 4.95%. Notably, the majority of welfare losses in both markets stem from search delays, with intermediation frictions contributing negligibly to the welfare loss in the government bond market and accounting for only 0.10% in the corporate bond market. In total, the welfare impact of OTC trading frictions is 2.38% in the government bond market and 5.05% in the corporate bond market under normal market conditions. When we re-estimate the model using data from the COVID-19 crisis, the welfare losses rise sharply to 3.63% in the government bond market and 11.35% in the corporate bond market.¹ This increase—particularly in the role of intermediation frictions in the corporate bond market—highlights the fragility of the OTC market structure in times of stress.

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 trade cost. 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

¹For more details, see Appendix B.

the trading frictions that the *same clients* face in two different markets.

The theoretical contribution of the paper is to develop a structural model of OTC trading with two-sided search and bilateral bargaining that can be estimated using transaction-level data with client identities. Building on the stationary version of [Lagos and Rocheteau \(2009\)](#), we extend the framework along three key dimensions to better capture the realities of fixed-income markets. First, clients in our model hold positions in two OTC-traded assets—government and corporate bonds—allowing us to analyze cross-market interactions and conduct quantitative comparisons. Second, we introduce a “core” of intermediaries who bargain with dealers in the inter-dealer market, generating endogenous inter-dealer price dispersion and allowing us to quantify the passthrough of inter-dealer market frictions to clients’ trade costs. Finally, we incorporate multi-dimensional heterogeneity in client characteristics, which is 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 utilities for holding the two assets, leading them to prefer different portfolios from one another and to adjust their holdings over time. 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 portfolio and an ideal low-type portfolio in mind. If a client’s current taste matches her ideal portfolio, 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 portfolio accordingly. Importantly, these two ideal portfolios and the resulting trade sizes 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 either asset. These features represent the novel client heterogeneity dimensions of our model relative to [Lagos and Rocheteau \(2009\)](#), and they are tightly linked to our model’s two-asset structure.

Clients in our model trade exclusively with dealers in a bilateral fashion subject to search frictions. Dealers, in turn, can offload undesirable inventory positions—whether long or short—to core broker-dealers without search frictions but still via bilateral bargaining. Core broker-dealers, however, trade amongst themselves multilaterally and without frictions. The presence of bilateral bargaining both in the inter-dealer segment and the client-dealer segment leads to inter-dealer price dispersion as well as client-specific bid and ask prices. When switching from the low-type portfolio to the high-type portfolio, the client pays a negotiated ask price per share of the asset bought from the dealer. Vice versa, when switching from the high-type portfolio to the low-type portfolio, 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, the equilibrium can be characterized in closed form in a special case of the model with iso-elastic utility. This allows us to derive a number of equilibrium objects that are readily observable from the data as (distributions or integral transforms of) closed-form expressions. These include, from each market, trading volume, client-dealer price dispersion, inter-dealer price dispersion, average client trade cost, the distribution of client-specific average trade frequencies, and the distribution of 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 substantial differences in trading delays across the two markets. The median client spends less than five minutes searching to trade a government bond, compared to approximately 45 minutes for a corporate bond. At the 75th percentile, clients require about half a day to complete a government bond transaction, and nearly 1.4 days for a corporate bond. These findings underscore the stark contrast in search frictions between the two markets. Furthermore, the estimated share of transaction surplus lost by clients is more than nine times higher in the corporate bond market than in the government bond market. We then decompose these client losses into two components: dealers' market power and the passthrough of inter-dealer market frictions. In the government bond market, nearly all client losses stem from the passthrough of inter-dealer frictions. In contrast, in the corporate bond market, roughly one-third of the lost surplus is attributable to dealer market power, and the remaining two-thirds to inter-dealer frictions.

To fully exploit the granularity of our dataset and the rich client heterogeneity in our model, we non-parametrically estimate the joint distribution of three client-specific characteristics: taste shock frequency, meeting rate with government bond dealers, and meeting rate with corporate bond dealers. We begin by formally proving that, given the market-wide parameters, observing clients' trading frequencies in government bonds, trading frequencies in corporate bonds, and trade sizes in either market is sufficient to uniquely identify this joint distribution. We then embed this identification result into a fixed-point algorithm, which becomes a core component of our estimation strategy and enables a non-parametric recovery of the joint distribution of clients' three-dimensional characteristics. The intuition behind this result is as follows. Let χ_0 denote a client's taste shock frequency, χ_1 her meeting rate with government bond dealers, and χ_2 her meeting rate with corporate bond dealers. A client's government bond trade frequency increases with χ_0 and χ_1 , but decreases with χ_2 ; her corporate bond trade frequency increases with χ_0 and χ_2 , but decreases with χ_1 ; and her trade size in either market increases with both χ_1 and χ_2 , but decreases with χ_0 . As a result, each of the three observable client-level moments provides distinct information about the underlying vector (χ_0, χ_1, χ_2) . This structure effectively yields a system of three equations in three unknowns for each client, allowing us to identify each individual's taste shock frequency and matching efficiencies in the two markets.

The identification of client-specific characteristics, as described above, forms the inner loop of our estimation procedure. In the outer loop, we recover the market-wide parameters using a

minimum-distance estimation strategy that leverages both market-level moments and client-level moments not utilized in the inner loop. With both the client-level heterogeneity and market-wide parameters in hand, we next proceed to conduct our welfare analyses.

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 with the allocation which would obtain if intermediaries did not capture any transaction surplus. This analysis allows us to decompose the total welfare loss into a component caused by intermediation frictions and another component caused purely by search frictions. We find that the total welfare losses in the government and corporate bond markets are 2.38% and 5.05%, respectively, and our decomposition implies that these losses are almost exclusively caused by search frictions in both markets. 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 2.14% and 3.00% for the welfare loss in the government bond market and 4.36% and 6.31% 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 3.63% and 11.35%, respectively, and our decomposition implies that this loss is almost exclusively caused by search frictions in the government bond market as in normal times, while intermediation frictions' share in the welfare loss in the corporate bond market increased from around 2% to 30%. 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 were not particularly fragile during turbulent times. To this end, we calculate the welfare losses by keeping matching efficiency and intermediation frictions 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 2.54% and 5.24%, 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), Kondor and Pintér (2022), and Kargar, Lester, Plante, and Weill (2023). 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. Notably, Kargar, Lester, Plante, and Weill (2023) is the only empirical study to directly measure trading delays. Their identification strategy exploits fragmentation in the US corporate bond market with centralized auctions and decentralized bilateral trade. They use the time of a failed centralized auction as the time of an investor's first effort to execute a particular decentralized transaction. In the absence of data on centralized auctions, we follow a structural approach to measure trading delays in the UK's government and corporate bond markets.

A theoretical literature following Duffie, Gârleanu, and Pedersen (2005) and Lagos and Rocheteau (2009, henceforth LR) that modeled OTC financial markets with a dynamic search-based framework focused mostly on offering qualitative insights. See Weill (2020) for a recent survey. 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 featuring an inter-dealer market without search frictions similar to LR includes Lagos and Rocheteau (2006), Lester, Rocheteau, and Weill (2015), Sultanum (2018), Kargar, Passadore, and Silva (2020), Chang and Zhang (2021), Colliard, Foucault, and Hoffmann (2021), Chiu, Davoodalhosseini, and Jiang (2022), Cohen, Kargar, Lester, and Weill (2023), and Li (2024), among others. These studies primarily offer theoretical insights, with a few incorporating calibrated numerical exercises. 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. In terms of theoretical contribution, to the best of our knowledge, our paper offers the only LR extension with multiple interdependent assets, which is essential to shed light on the cross-asset liquidity interactions.³

Structural estimation of dynamic decentralized trading models is an understudied but fast-growing area. Examples include Feldhütter (2012), Albuquerque and Schroth (2015), Gavazza (2016), Brancaccio, Li, and Schürhoff (2017), Buchak, Matvos, Piskorski, and Seru (2020), Hendershott, Li, Livdan, and Schürhoff (2020), Liu (2020), Coen and Coen (2021), Lu, Puzzello, and

²Empirical 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.

³Li (2024) also studies an LR extension with multiple assets. Her model assumes independent asset payoffs, and so, it does not result in any cross-asset interaction liquidity-wise.

Zhu (2023), and Brancaccio and Kang (2024). Our key contribution relative to these papers is that we develop and estimate a multi-asset model, while these papers' model feature trading of a single asset. A second contribution is that, with the exception of Coen and Coen (2021), Lu, Puzzello, and Zhu (2023), and Brancaccio and Kang (2024), this literature models the trade of an indivisible unit of an asset, while we model trading of divisible assets with trade sizes optimally chosen. The main benefit of this modeling choice in the context of structural estimation is that asset divisibility allows us to use information on trade sizes to identify the deep parameters of our model.⁴ Trade sizes are especially important in identifying simultaneously the clients' preference shock frequency and matching efficiency with dealers from transaction-level data.⁵ In their independent, contemporaneous work, Coen and Coen (2021) and Brancaccio and Kang (2024) also allow for endogenous trade sizes in a decentralized market with all-to-all trading *à la* Üslü (2019). As in Üslü (2019), these papers 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. In addition, we develop a multi-asset model to be able to capture liquidity spillovers between the government and the corporate market, while these papers all offer single-asset models.

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.⁶ Another strand of literature estimates auction models to analyze primary markets. In contrast, our focus is on 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

⁴In addition, 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.

⁵For 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.

⁶See Weill (2020) for a discussion.

(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. Another important feature of our analysis—in addition to our special focus on studying and comparing different markets—is the emphasis we place on endogenous trade sizes and asset divisibility. These other contributions model instead 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 concludes. Appendix A presents the formal results on the identification of clients’ characteristics. Appendix B re-estimates the model parameters to fit data from the COVID-19 crisis.

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.⁷ 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.⁸

Our baseline sample covers the period between Aug 2011 and Dec 2017. We apply standard filters to exclude duplicate trades, transactions below £1,000 par value, missing identifiers, and erroneous entries. We identify 526 clients who are active in both markets, and whose trades account for 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

⁷See 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.

⁸This 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.

markets. Their number fluctuates around 20 in our sample.⁹ We end up with approximately 2.96 million trades in 57 nominal government bonds.¹⁰ Around 63% of these transactions take place in the inter-dealer segment and 37% in the client-dealer segment. Further details on our data construction procedure are provided in Appendix F.

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.¹¹ 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 229,180 corporate bond transactions that has the same number of distinct assets as our sample of government bonds.¹² The breakdown of the transactions between the inter-dealer and client-dealer segments is respectively 35% and 65% in this case. 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.

We measure price dispersion in both the client-dealer and inter-dealer segments of the gov-

⁹Certain 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 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/>.

¹⁰Another 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.

¹¹We 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.

¹²Corporate bonds that are traded less frequently are more likely to be subject to adverse selection risk (Ronen and Zhou, 2013; Benmelech and Bergman, 2018).

ernment and corporate bond markets using the same approach. We first compute the absolute deviation of each transaction price from the hourly average transaction price of a given market segment. 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 segment. 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 the first row of Table 1, average client-dealer price dispersion is about 6.4 bps in the government bond market and about 12.3 bps in the corporate bond market.¹³ For inter-dealer transactions, average price dispersion is around 6.01 bps in the government bond market and 8.3 bps in the corporate bond market.¹⁴ 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 client-dealer price dispersion is almost twice as large in corporate bonds than government bonds is suggestive of corporate bond markets being more frictional than government bond markets. In addition, the smaller gap in price dispersion observed for inter-dealer transactions (around 38%) is indicative of different market power dynamics in each market segment.

A complementary approach to measuring the trading frictions faced by clients is with the trading cost measure used in O’Hara and Zhou (2021) and Pintér, Wang, and Zou (2024). Specifically, for each trade v we compute the following measure:

$$Cost_v = \left[\log(P_v) - \log(\bar{P}) \right] \times \mathbf{1}_{B,S},$$

where P_v is the transaction price, $\mathbf{1}_{B,S}$ is an indicator function equal to 1 when the client buys, and equal to -1 when the client sells, and $\bar{P}$ is a benchmark price.¹⁵ The third row of Table 1 reports the cross-sectional average of client-specific mean trade costs in each market. The average client trade cost is respectively 0.77 bps in the government bond market and 7.08 bps in the corporate bond market—consistent with the evidence from the price dispersion measures that the corporate

¹³To 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.

¹⁴Eisfeldt, Herskovic, and Liu (2024) report an average inter-dealer price dispersion of 40.5 bps for the US corporate bond market. However, their measure is computed at a monthly frequency on a sample covering the period 2004–2018—with elevated levels of price dispersion during the GFC—and includes a much larger cross-section of bonds (10,537) than our set of 57 highly-liquid investment-grade bonds.

¹⁵We follow the methodology outlined in Equation (1) of Pintér, Wang, and Zou (2024), using the average price of all transactions at the bond-day level as the benchmark price against which transaction costs are computed.

bond market is more frictional than the government bond market. However, without additional information on clients' trading intensities and other quantities, price dispersions and client trade costs alone are 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 for each of the 526 clients on each trading day. In addition, we compute intensity dispersion as the mean of the absolute deviation of each client's total number of transactions from average intensity. Both measures are scaled by the number of assets (57) in each sample. The fourth and fifth rows of Table 1 show that average intensity and intensity dispersion are 0.034 and 0.053, respectively, in the government bond market, and they are 0.0041 and 0.0073 in the corporate bond market. That is, average intensity and intensity dispersion are about 7-8 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.

Table 1: Summary Statistics

Variable	Government Bonds		Corporate Bonds		N
	Mean	sd	Mean	sd	
Price Dispersion	0.0006445	0.0002072	0.0012262	0.0006705	1,440
Inter-dealer Price Dispersion	0.0006051	0.0001689	0.0008283	0.0009383	1,440
Clients' Trade Costs (bps)	0.7699433	4.7900271	7.0814378	10.8234570	1,440
Average Intensity	0.0343979	0.0075582	0.0040584	0.0012159	1,440
Intensity Dispersion	0.0531009	0.0111978	0.0072671	0.0020633	1,440
Trade Volume (£)	281,460	137,987	2,744	2,645	1,440
Average Trade Size (£)	6,467,011	4,301,335	277,808	244,438	1,440

Notes: This table summarizes the empirical moments that inform our structural estimation. Price dispersion is the scaled mean absolute deviation of transaction prices from the average transaction price in the given hour, calculated by using client-dealer trade prices. Inter-dealer price dispersion is calculated in the same way as price dispersion, using inter-dealer trade prices. Clients' trade costs are calculated using the methodology of Pintér, Wang, and Zou (2024) and averaged across clients. To calculate trade costs, we use the benchmark price computed as the average price of all transactions at the bond-day level. 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. Trade volume is total daily trading volume in terms of par value per bond per client. Average trade size is the mean (across clients) of clients' mean trade size. The sample includes 1440 trading days over the period Aug 2011 - Dec 2017.

We measure trade volume as the total daily trading volume in terms of par value, scaled by the number of clients and the number of assets in each market. The average daily trade volume in our sample is £281,460 per bond per client in the government bond market and £2,744 in the corporate bond market.

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 find that the average trade size in the government bond market is around 23 times larger than in the corporate bond market.¹⁶

3 A Model of Intermediated OTC Markets

To capture cross-market interactions, inter-dealer price dispersion, and the rich client heterogeneity observed in fixed-income markets in practice, we extend the stationary version of LR along three dimensions. First, clients hold positions in two OTC-traded assets as opposed to one in LR. This is essential because we want to quantitatively compare the government bond and the corporate bond markets. Second, although we keep the LR assumption that there are no search frictions in the inter-dealer market, we introduce a “core” who bargain with dealers when providing intermediation services in the inter-dealer market.¹⁷ This departure from LR offers two primary benefits: (i) we obtain equilibrium inter-dealer price dispersion as in the data and (ii) we use information from the inter-dealer price dispersion to figure out the costs dealers face for the intermediation services they consume in the inter-dealer market. Finally, 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.

3.1 The Economic Environment

Time is continuous and runs forever. The economy is populated by a continuum of clients, a continuum of dealers, and a continuum of core broker-dealers. The measures of all three types of agents are normalized to one each.

There are two long-lived assets in exogenous per-client-capita supplies $A_1 > 0$ and $A_2 > 0$. Although the model treats these two as generic assets, in our quantitative analyses we associate asset 1 with government bonds and asset 2 corporate bonds. 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 an asset than

¹⁶This is consistent with Belsham, Rattan, and Maher (2017) who document that typical trade sizes in the UK government bond market are typically 10 to 25 times larger than in the sterling-denominated investment-grade corporate bond market. We also calculate the mean (across clients) of q_{gilt}^i/q_{corp}^i , where q_k^i is the time-series average of client i 's trade sizes in market $k \in \{gilt, corp\}$, and find that it equals 14.8085.

¹⁷This *three-tiered* market structure is reminiscent of that in Colliard, Foucault, and Hoffmann (2021). The exogenously specified core of this market structure can be thought of as a reduced-form way to capture the endogenous core-periphery structure and inter-dealer price dispersion of Hugonnier, Lester, and Weill (2020), Farboodi, Jarosch, and Shimer (2018), and Üslü (2019), among others.

she consumes, her net consumption becomes negative. The flow utility of a client is

$$c + \varepsilon u(a_1 + \rho a_2),$$

where c is her net consumption of the numéraire, ε her current taste type, and a_1 and a_2 her positions in asset 1 and asset 2, respectively. The felicity function $u : [-M, M] \rightarrow \mathbb{R}$ is twice continuously differentiable, strictly increasing, strictly concave, and defined for an arbitrarily large $M > 0$. The parameter $\rho > 0$ is the constant marginal rate of substitution (MRS) between the two assets.¹⁸ 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 uniform cdf $U(\{\varepsilon_l, \varepsilon_h\})$. That is, 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. 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 three permanent characteristics, $\chi_0 > 0$, $\chi_1 > 0$, and $\chi_2 > 0$, where χ_0 refers to the rate at which a client receives taste shocks and χ_k the rate at which she receives trade opportunities with dealers to trade asset $k \in \{1, 2\}$. We denote with $\chi = (\chi_0, \chi_1, \chi_2)$ a client's three-dimensional permanent characteristics. The cross-sectional distribution of clients' characteristics is represented by the joint cdf $G : [\chi_l, \chi_h]^3 \rightarrow \mathbb{R}$. We follow [Vayanos and Wang \(2007\)](#) in interpreting a large- χ_0 client as a liquidity trader and a small- χ_0 as a buy-and-hold trader. Similarly, for $k \in \{1, 2\}$, we follow [O'Hara, Wang, and Zhou \(2018\)](#) in interpreting a large- χ_k client as an active trader of asset k and a small- χ_k as a passive trader.

Dealers' and core broker-dealers' utility flow is assumed to be c ; i.e., they do not derive any utility from holding the assets. Core broker-dealers can trade asset k amongst themselves instantaneously in a frictionless market at the market-clearing price P_k . Accordingly, we assume without loss of generality that core broker-dealers do not hold any position in the assets as they do not derive any utility from holding the assets. We assume that dealers can trade with core broker-dealers without delay but only bilaterally. That is, if a dealer has an unwanted position in an asset, he is immediately matched with a core broker-dealer to offload this unwanted position. However, the terms of trade will generically be different from P_k because we assume some positive bargaining power for the core. Nevertheless, this instantaneous access to the core allows dealers to offload any unwanted position to the core instantly, and so, dealers do not hold any position in the assets either. Finally, the end users of the assets, i.e. clients, can trade only with dealers, infrequently, and in a bilateral fashion, i.e., with one dealer at a time. Clients receive trading

¹⁸[Üslü and Velioğlu \(2019\)](#) micro-found the constant marginal rate of substitution assumption in a CARA-Brownian model with multiple assets. We, instead, hard-wire this assumption to our model with a general “warm-glow” concave utility.

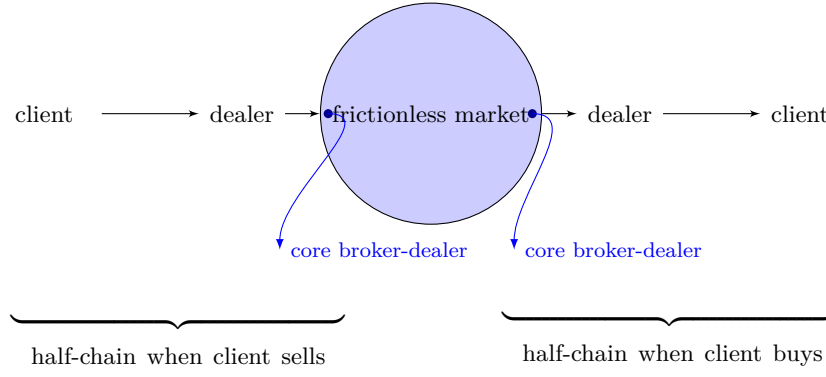


Figure 1: The left half of the figure shows the flow of assets after a client with a selling need meets a dealer. Vice versa, the right half of the figure shows the flow of assets after a client with a buying need meets a dealer.

opportunities at the arrival times of a continuum of pair-wise independent Poisson processes with client-specific arrival rates of χ_k for $k \in \{1, 2\}$. Figure 1 demonstrates typical intermediation (half-)chains started by a client-seller (the left side) and a client-buyer (the right side) with our assumed three-tiered market structure.

Upon the arrival of a trade opportunity shock for asset $k \in \{1, 2\}$, the shocked client is matched with a dealer picked randomly and uniformly from the pool of dealers. Simultaneously, the dealer is matched with a core broker-dealer picked randomly and uniformly from the core. The terms of trade is, then, set by a multilateral Pareto-optimal bargaining game during which the bargaining parties determine a trade quantity that maximizes the joint-surplus of the trade, a trade price between the client and the dealer, and a trade price between the dealer and the core. These two trade prices are determined such that the maximized joint-surplus is split among the client, the dealer, and the core. In the end, the core captures a fraction $\eta_k^c \in [0, 1]$ of the surplus and the dealer $\eta_k^d \in [0, 1]$, while the client captures the remaining share $1 - \eta_k$, where $\eta_k \equiv \eta_k^c + \eta_k^d$.¹⁹ Figure 2 demonstrates our assumed multilateral bargaining solution with a pie chart.

3.2 Equilibrium Definition

Let $V(\varepsilon, a_1, a_2, \chi)$ refer to the continuation utility of a client with the current taste type of ε , current positions a_1 and a_2 in asset 1 and 2, respectively, and the permanent characteristic vector χ . If this client meets a dealer to trade asset $k \in \{1, 2\}$ at this moment, the dealer simultaneously meets a core broker-dealer. Then, $q_k(\varepsilon, a_1, a_2, \chi)$, which stands in for both the number of shares of asset k the client buys from the dealer and the number of shares of the asset k the dealer buys

¹⁹This Pareto-optimal bargaining solution can be understood as the generalized Nash solution for *multilateral* bargaining. For example, it can be derived from the cooperative bargaining problem of [Lensberg \(1988\)](#), [Krishna and Serrano \(1996\)](#), and [Suh and Wen \(2006\)](#) with asymmetric bargaining powers. See Appendix C for more details.

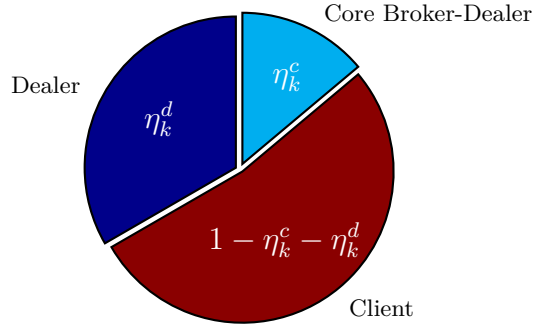


Figure 2: Exogenously specified surplus shares in the multilateral bargaining game in the market for asset k

from the core, solves

$$\max_{q \in \mathbb{R}} V(\varepsilon, a_k + q, a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - P_k q, \quad (3.1)$$

where $V(\varepsilon, a_k + q, a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi)$ 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 asset k (or after she has sold $-q$ units if q is negative). The last term, $-P_k q$, represents the core's contribution to the surplus creation and is equal to the cost of obtaining q units of asset k from the frictionless market (or the benefit of selling $-q$ units in the frictionless market if q is negative). Note that neither the client-dealer nor dealer-core trade price enters the joint-surplus formula because they are simply a transfer of the numéraire from one party to another when each have a linear utility in that.

Trade price that the client pays to the dealer per unit of the asset traded is denoted by $p_k(\varepsilon, a_1, a_2, \chi)$ and is equal to

$$p_k(\varepsilon, a_1, a_2, \chi) = P_k + \eta_k^c \frac{V(\varepsilon, a_k + q_k(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - P_k q_k(\varepsilon, a_1, a_2, \chi)}{q_k(\varepsilon, a_1, a_2, \chi)} \quad (3.2)$$

and, similarly, trade price that the dealer pays to the core per unit of the asset traded is denoted by $p_k^d(\varepsilon, a_1, a_2, \chi)$ and is equal to

$$p_k^d(\varepsilon, a_1, a_2, \chi) = P_k + \eta_k^c \frac{V(\varepsilon, a_k + q_k(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - P_k q_k(\varepsilon, a_1, a_2, \chi)}{q_k(\varepsilon, a_1, a_2, \chi)}. \quad (3.3)$$

Since the joint-surplus in the numerator of the fraction in (3.3) is non-negative by the optimal choice of $q_k(\varepsilon, a_1, a_2, \chi)$, this means that the core charges a markup over the frictionless price P_k when selling the asset to the dealer ($q_k(\varepsilon, a_1, a_2, \chi) > 0$). Vice versa, the core obtains a markdown when buying the asset from the dealer ($q_k(\varepsilon, a_1, a_2, \chi) < 0$). Similarly, $p_k(\varepsilon, a_1, a_2, \chi) -$

$p_k^d(\varepsilon, a_1, a_2, \chi)$ has the same sign as $q_k(\varepsilon, a_1, a_2, \chi)$ because $\eta_k \geq \eta_k^c$. This implies that the dealer charges a markup when selling the asset to the client. Vice versa, the dealer obtains a markdown when buying the asset from the client.

The continuation utility $V(\varepsilon, a_1, a_2, \chi)$ satisfies the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\begin{aligned} rV(\varepsilon, a_1, a_2, \chi) &= \varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} [V(\varepsilon', a_1, a_2, \chi) - V(\varepsilon, a_1, a_2, \chi)] \\ &+ \sum_{k \in \{1, 2\}} \chi_k [V(\varepsilon, a_k + q_k(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - q_k(\varepsilon, a_1, a_2, \chi) p_k(\varepsilon, a_1, a_2, \chi)]. \end{aligned} \quad (3.4)$$

A client's "flow" continuation utility, $rV(\varepsilon, a_1, a_2, \chi)$, equals the sum of three terms. The first term is the client's current utility flow from holding a_1 and a_2 units of asset 1 and 2, respectively, 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_0 . The last term is the flow value of changing the position in asset k from a_k to $a_k + q_k(\varepsilon, a_1, a_2, \chi)$ by paying $q_k(\varepsilon, a_1, a_2, \chi) p_k(\varepsilon, a_1, a_2, \chi)$ units of the numéraire, which is also an infrequent possibility arriving at a Poisson rate of χ_1 and χ_2 for asset 1 and asset 2, respectively.

Let $\Phi_\chi(\varepsilon, a_1, a_2)$ denote the stationary joint cdf of clients' taste types and asset positions conditional on their permanent characteristics. The stationarity of this cdf is guaranteed by the following inflow-outflow equation:

$$\begin{aligned} \chi_0 U(\varepsilon) \int_{\varepsilon}^{\varepsilon_h} \Phi_\chi(d\varepsilon', a_1, a_2) &+ \chi_1 \int_{-M-M}^M \int_{-M-M}^M \int_{\varepsilon_l}^{\varepsilon} \mathbb{I}_{\{a_1 \leq a'_1 + q_1(\varepsilon', a'_1, a'_2, \chi)\}} \Phi_\chi(d\varepsilon', da'_1, da'_2) \\ &+ \chi_2 \int_{-M-M}^M \int_{-M-M}^M \int_{\varepsilon_l}^{\varepsilon} \mathbb{I}_{\{a_2 \leq a'_2 + q_2(\varepsilon', a'_1, a'_2, \chi)\}} \Phi_\chi(d\varepsilon', da'_1, da'_2) \\ &= \chi_0 (1 - U(\varepsilon)) \Phi_\chi(\varepsilon, a_1, a_2) + (\chi_1 + \chi_2) \Phi_\chi(\varepsilon, a_1, a_2). \end{aligned} \quad (3.5)$$

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 remaining 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 $\chi = (\chi_0, \chi_1, \chi_2)$, asset positions smaller than a_1 and a_2 , and a taste type larger than ε switches to a taste type smaller than or equal to ε with probability $U(\varepsilon)$ following a taste shock that occurs at rate χ_0 . Thus, the multiplication of χ_0 , the probability $U(\varepsilon)$, and the measure of such clients gives us the inflow to $\Phi_\chi(\varepsilon, a_1, a_2)$ 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_0 , the

probability, $1 - U(\varepsilon)$, 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 $\Phi_\chi(\varepsilon, a_1, a_2)$ due to trading asset 1, is the multiplication of the Poisson intensity of opportunities to trade asset 1 for this class of clients, χ_1 , 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 a position in asset 1 less than a_1 so she indeed creates an inflow. The third term has a similar interpretation but regarding asset 2. Finally, the second term of the RHS stands in for the (gross) outflow from $\Phi_\chi(\varepsilon, a_1, a_2)$, which is equal to the multiplication of the Poisson rate of total trade opportunities and the measure of clients with characteristics χ who have a positions smaller than a_1 and a_2 in asset 1 and 2, respectively, and a taste type smaller than ε .

In addition to the stationarity condition (3.5), there are two additional feasibility or accounting identities that $\Phi_\chi(\cdot, \cdot, \cdot)$ must satisfy:

$$\int_{\chi_l}^{\chi_2} \int_{\chi_l}^{\chi_1} \int_{\chi_l}^{\chi_0} \int_{-M}^M \int_{-M}^M \int_{\varepsilon_l}^{\varepsilon_h} \Phi_{\chi'}(d\varepsilon, da_1, da_2) G(d\chi'_0, d\chi'_1, d\chi'_2) = G(\chi_0, \chi_1, \chi_2) \quad (3.6)$$

and

$$\int_{-M}^M \int_{-M}^M \int_{\varepsilon_l}^{\varepsilon_h} \Phi_\chi(d\varepsilon, da_1, da_2) = 1 \quad (3.7)$$

for all $\chi \in \text{supp}(dG)$. Equation (3.6) implies that the equilibrium conditional distribution of clients' states is consistent with the exogenous distribution of client characteristics and (3.7) follows from the fact that $\Phi_\chi(\cdot, \cdot, \cdot)$ is a conditional cdf. Finally, the frictionless markets in which the core broker-dealers trade amongst themselves must clear:

$$\int_{\chi_l}^{\chi_h} \int_{\chi_l}^{\chi_h} \int_{\chi_l}^{\chi_h} \int_{-M}^M \int_{-M}^M \int_{\varepsilon_l}^{\varepsilon_h} \chi_k q_k(\varepsilon, a_1, a_2, \chi | P_k) \Phi_\chi(d\varepsilon, da_k, da_{-k}) G(d\chi_0, d\chi_1, d\chi_2) = 0 \quad (3.8)$$

for all $k \in \{1, 2\}$. Equation (3.8) guarantees that all shares of the assets brought to the frictionless markets by the core-broker dealers net out to zero in each market. Because this is the market-clearing condition, we have made the dependence of q_k on P_k explicit in our notation. Mathematically, this dependence stems from the fact that the optimization problem (3.1) takes P_k as given.

Taking stock, we define a stationary equilibrium as follows.

Definition 1. Let $\mathcal{T} = \{\varepsilon_l, \varepsilon_h\} \times [-M, M]^2 \times [\chi_l, \chi_h]^3$. A stationary equilibrium is (i) a function $V : \mathcal{T} \rightarrow \mathbb{R}$ for clients' continuation utilities, (ii) a set of functions $q_k : \mathcal{T} \rightarrow \mathbb{R}$ for clients' trade sizes for $k \in \{1, 2\}$, (iii) a set of functions $p_k : \mathcal{T} \rightarrow \mathbb{R}$ for transaction prices between clients and

dealers for $k \in \{1, 2\}$, (iv) a set of functions $p_k^d : \mathcal{T} \rightarrow \mathbb{R}$ for transaction prices between dealers and the core for $k \in \{1, 2\}$, (v) a joint cdf $\Phi_\chi : \{\varepsilon_l, \varepsilon_h\} \times [-M, M]^2 \rightarrow \mathbb{R}$ for clients' taste types and asset positions conditional on their characteristics $\chi \in \text{supp}(dG)$, and (vi) a pair of frictionless market prices $P_k \in \mathbb{R}$ for $k \in \{1, 2\}$ such that

- Given (ii) and (iii), (i) solves the HJB equation (3.4).
- Given (i) and (vi), (ii) maximizes the joint surplus (3.1).
- Given (i), (ii), and (vi), (iii) satisfies (3.2).
- Given (i), (ii), and (vi), (iv) satisfies (3.3).
- Given (ii), (v) satisfies the stationarity and feasibility conditions (3.5)-(3.7).
- Given (ii) and (v), (vi) satisfies (3.8).

3.3 Equilibrium Characterization

Substituting (3.2) into (3.4),

$$\begin{aligned} rV(\varepsilon, a_1, a_2, \chi) &= \varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} [V(\varepsilon', a_1, a_2, \chi) - V(\varepsilon, a_1, a_2, \chi)] \\ &+ \sum_{k \in \{1, 2\}} \chi_k (1 - \eta_k) [V(\varepsilon, a_k + q_k(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - q_k(\varepsilon, a_1, a_2, \chi) P_k]. \end{aligned}$$

Using (3.1) and with a change of variable,

$$\begin{aligned} rV(\varepsilon, a_1, a_2, \chi) &= \varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} [V(\varepsilon', a_1, a_2, \chi) - V(\varepsilon, a_1, a_2, \chi)] \\ &+ \sum_{k \in \{1, 2\}} \chi_k (1 - \eta_k) \max_{a'_k} \{V(\varepsilon, a'_k, a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - (a'_k - a_k) P_k\}. \end{aligned}$$

Rearrangement implies

$$\begin{aligned} [r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)] V(\varepsilon, a_1, a_2, \chi) &= \varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V(\varepsilon', a_1, a_2, \chi) \\ &+ \sum_{k \in \{1, 2\}} \chi_k (1 - \eta_k) \max_{a'_k} \{V(\varepsilon, a'_k, a_{-k}, \chi) - (a'_k - a_k) P_k\} \quad (3.9) \end{aligned}$$

or

$$V(\varepsilon, a_1, a_2, \chi) = \frac{r \frac{\varepsilon u(a_1 + \rho a_2)}{r} + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V(\varepsilon', a_1, a_2, \chi)}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} + \frac{\sum_{k \in \{1, 2\}} \chi_k (1 - \eta_k) \max_{a'_k} \{V(\varepsilon, a'_k, a_{-k}, \chi) - (a'_k - a_k) P_k\}}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}.$$

The auxiliary HJB equation (3.9) shows that a client's current continuation value is equal to the weighted average of four values: the value of holding (a_1, a_2) units of the assets forever while keeping the current taste type ε , the value of keeping (a_1, a_2) units forever with a randomly drawn taste type from the cdf $U(\{\varepsilon_l, \varepsilon_h\})$, and the two values associated with holding the optimal amount of asset $k \in \{1, 2\}$ by trading it at the frictionless price P_k while keeping the taste type ε . The first of the four values affects the continuation utility because the client is impatient ($r > 0$), and so, its weight is $r / (r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))$. The second one affects the continuation value because the client receives taste shocks at the Poisson rate χ_0 , and so, its weight is $\chi_0 / (r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))$. The last two affect the continuation value because the client gets to trade asset k at the Poisson rate χ_k . The weight of this term is, however, weighted down by $1 - \eta_k$ because the intermediating dealer and core broker-dealer together will capture a share, $\eta_k = \eta_k^d + \eta_k^c$, of the trade surplus. Thus, the weights of the last two terms turn out to reflect a *bargaining-adjusted* Poisson rate $\chi_k(1 - \eta_k)$: $\chi_k(1 - \eta_k) / (r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))$.

The first key step in characterizing the equilibrium is to solve the auxiliary HJB equation (3.9) given some restriction on P_1 and P_2 . This step uses standard fixed-point tools for dynamic programming because (3.9) defines a contraction mapping, and so, we prove that there exists a unique solution to (3.9) given P_1 and P_2 . Then, with some additional restriction on P_1 and P_2 , we obtain this unique solution in closed form by following a method of undetermined coefficients.

Proposition 1. *Let $\theta = a_1 + \rho a_2$ denote the composite asset position of a client who holds a_1 and a_2 units of asset 1 and 2, respectively. Let $\bar{\varepsilon} = (\varepsilon_l + \varepsilon_h) / 2$. Given P_1 and P_2 such that $\rho P_1 = P_2$,²⁰ the unique solution to the auxiliary HJB equation (3.9) has the following functional form:*

$$V(\varepsilon, \theta, \chi) = \frac{\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} P_1 \theta + \frac{1}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \frac{(r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)) \varepsilon + \chi_0 \bar{\varepsilon}}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} u(\theta) + C(\varepsilon, \chi) \quad (3.10)$$

for some $C : \{\varepsilon_l, \varepsilon_h\} \times [\chi_l, \chi_h]^3 \rightarrow \mathbb{R}$.

²⁰This assumption is verified ex post. Namely, this restriction is implied by the market-clearing conditions in the frictionless markets. Intuitively, because of our constant MRS assumption between the two assets, any price ratio deviating from the MRS, ρ , would lead to an excess demand for one asset and an excess supply of the other, and so, it is necessarily the case that $\rho P_1 = P_2$ in any general equilibrium.

Proposition 1 reveals the tractability benefit of the constant MRS between the assets. As in Üslü and Velioğlu (2019), we are able to define the sufficient statistic θ for a client's multi-asset portfolio, even though we use a general concave utility function as opposed to CARA utility in Üslü and Velioğlu (2019). Given the value function (3.10) in Proposition 1, one can now easily determine the terms of trade between a client and a dealer using (3.1) and (3.2) and those between the given dealer and the core using (3.1) and (3.3).

Applying the chain rule, the FOC of the joint surplus maximization (3.1) is

$$V_{\theta}(\varepsilon, \theta', \chi) \theta_k = P_k,$$

where θ' is the client's post-trade composite asset position, $V_{\theta}(\cdot, \cdot, \cdot)$ refers to the derivative with respect to the second argument, and θ_k refers to the derivative of the composite asset position with respect to asset k for $k \in \{1, 2\}$. Thus, $\theta_1 = 1$ and $\theta_2 = \rho$. Using (3.10) and after rearranging, one obtains that the optimal composite asset position of a client with the taste type of ε and characteristics χ is

$$\theta^*(\varepsilon, \chi) = (u')^{-1} \left[\frac{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{(r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))\varepsilon + \chi_0\bar{\varepsilon}} r P_1 \right], \quad (3.11)$$

i.e., a client with these characteristics will end up holding positions in asset 1 and 2 that generates $\theta^*(\varepsilon, \chi)$ after meeting a dealer regardless of her initial asset positions a_1 and a_2 and no matter which asset is being traded with this particular dealer. Thus, the bilateral trade quantity between this client and the dealer (and, accordingly, between the dealer and the core) is

$$q_k(\varepsilon, \theta, \chi) = \frac{1}{\theta_k} \left\{ (u')^{-1} \left[\frac{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{(r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))\varepsilon + \chi_0\bar{\varepsilon}} r P_1 \right] - \theta \right\}. \quad (3.12)$$

Given a client's composite asset position θ , Equation (3.12) implies that the only asset-specific determinant of trade size is θ_k . That is, the larger the θ_k , the smaller the trade size of asset k . This also means that if $\rho > 1$ (resp. $\rho < 1$), asset 1 has larger (resp. smaller) trade sizes than asset 2.²¹ Thus, in our quantitative analyses in the next section, we will use the aggregate government bond trade size relative to the aggregate corporate bond trade size to calibrate the parameter ρ based on the data.

Going back to equilibrium characterization, given P_1 , P_2 , (3.10), and (3.12), (3.2) gives the negotiated price, $p_k(\varepsilon, \theta, \chi)$, between a given client and a dealer and (3.3) gives the negotiated price, $p_k^d(\varepsilon, \theta, \chi)$, between the dealer and the core, where we again write the client's composite asset position θ as an argument of the pricing functions because a_1 and a_2 affect these prices only

²¹This is reminiscent of the Üslü and Velioğlu (2019) result that the riskier assets are traded in smaller sizes. That is, in their CARA-Brownian model, assets' payoff processes generate the endogenous MRS between the assets, and so, Üslü and Velioğlu (2019) tie the assets' systematic risks to their trade size. In our model with an asset-in-the-utility function approach, the exogenous MRS parameter ρ determines the relative trade size.

as part of θ .

Taking stock, the clients' value functions, trade sizes, and transaction prices are all affected by their a_1 and a_2 only through their composite asset position θ . Therefore, instead of characterizing the joint distribution $\Phi_\chi(\varepsilon, a_1, a_2)$ of clients' taste types, positions in asset 1, and positions in asset 2, it suffices to characterize the joint distribution of clients' taste types and composite asset positions, which we denote with $F_\chi(\varepsilon, \theta)$. Equipped with (3.11) and (3.12), we derive the stationary equilibrium distribution $F_\chi(\varepsilon, \theta)$ in closed form in Proposition 2.

Proposition 2. *Let $\theta^*(\cdot, \cdot)$ be given by (3.11), and let $f_\chi(\cdot, \cdot)$ be the equilibrium joint probability mass function (pmf) of taste types and composite asset positions conditional on client characteristics. Then, for any $\chi \in \text{supp}(dG)$,*

$$f_\chi(\varepsilon, \theta) = \begin{cases} 0 & \text{if } \theta \notin \Theta(\chi) \\ \frac{1}{4} \frac{\chi_0 + 2\chi_1 + 2\chi_2}{\chi_1 + \chi_2} & \text{if } \theta = \theta^*(\varepsilon, \chi) \\ \frac{1}{4} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2} & \text{if } \theta = \theta^*(\varepsilon', \chi) \text{ for } \varepsilon' \neq \varepsilon, \end{cases} \quad (3.13)$$

where

$$\Theta(\chi) = \{\theta^*(\varepsilon_l, \chi), \theta^*(\varepsilon_h, \chi)\}.$$

The intuition behind Proposition 2 follows from the result (3.11) that there is a unique optimal composite holding for any client (ε, χ) , which implies the support $\Theta(\chi)$ for the equilibrium composite holding distribution for clients with characteristics χ . Namely, any client (ε, χ) either holds her optimal composite position $\theta^*(\varepsilon, \chi)$ or is still stuck with a previously optimal position $\theta^*(\varepsilon', \chi)$ for $\varepsilon' \neq \varepsilon$ because she is yet to trade with a dealer after her latest taste shock. As a result, the equilibrium distribution admits a simple representation, which is the conditional pmf (3.13).

Finally, the only remaining equilibrium objects yet to be determined are the frictionless market prices P_1 and P_2 . To determine these, we re-write the market-clearing condition (3.8) as

$$\int \sum_{\theta \in \Theta(\chi)} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_k q_k(\varepsilon, \theta, \chi | P_k) f_\chi(\varepsilon, \theta) dG(\chi) = 0$$

for $k \in \{1, 2\}$, with a slight abuse of notation “ $dG(\chi) = G(d\chi_0, d\chi_1, d\chi_2)$ ”. Using (3.12) and (3.13), this equation allows us to determine P_1 and P_2 in Proposition 3.

Proposition 3. *Assume $u'(\infty) = 0$ and $u'(0) = \infty$. Then, the unique equilibrium prices P_1 and P_2 solve*

$$\int \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} (u')^{-1} \left[\frac{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{(r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))\varepsilon + \chi_0\bar{\varepsilon}} r P_1 \right] dG(\chi) = A_1 + \rho A_2 \quad (3.14)$$

and

$$P_2 = \rho P_1.$$

In precis, the equilibrium defined in Definition 1 is characterized by Propositions 1–3 in closed form up to the frictionless prices P_1 and P_2 . With some additional parametric assumptions, P_1 and P_2 can be obtained in closed form as well. For example, similar to the original LR model, P_1 and P_2 are available in closed form when $u(\theta) = \log(\theta)$. In the next subsection, we assume an iso-elastic utility function, which nests the log utility as a special case, and show that together with our assumed binary taste-type structure it allows us to obtain explicit formulas for endogenous liquidity measures such as trade volume, price dispersion, and clients' trade costs.

3.4 A Special Case

We let $u(\theta) = \frac{\theta^{1-\gamma}}{1-\gamma}$. In addition, we assume $\varepsilon_l = 0$ and $\varepsilon_h = 2\sigma$ for some $\sigma > 0$. That is, the distribution U has two equal mass points at $\varepsilon = \mathbb{E}[\varepsilon] \pm \sigma$, which means that one natural interpretation for parameter σ is *preference volatility*.

Let us start by calculating the optimal composite holding of each client type in this special case. Using (3.11) and our assumptions on the utility function and the taste types,

$$\theta^*(\mathbb{E}[\varepsilon] \pm \sigma, \chi) = \left[\left(1 \pm \frac{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \right) \frac{\sigma}{rP_1} \right]^{1/\gamma}. \quad (3.15)$$

Note that this optimal composite holding takes as given the frictionless price P_1 . Now we calculate the equilibrium value of P_1 so that we can write all equilibrium objects in terms of exogenous parameters and distributions. Equation (3.14) can be re-written as

$$\frac{1}{2} \int \left[(1 + \Sigma(\chi')) \frac{\sigma}{rP_1} \right]^{1/\gamma} dG(\chi') + \frac{1}{2} \int \left[(1 - \Sigma(\chi')) \frac{\sigma}{rP_1} \right]^{1/\gamma} dG(\chi') = A_1 + \rho A_2,$$

where

$$\Sigma(\chi) \equiv \frac{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}$$

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

$$P_1 = \frac{\sigma}{r} \left\{ \frac{1}{2(A_1 + \rho A_2)} \int \left[(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma} \right] dG(\chi') \right\}^\gamma. \quad (3.16)$$

Substituting (3.16) into (3.15), the equilibrium optimal composite holding of each client type is

$$\theta^*(\mathbb{E}[\varepsilon] \pm \sigma, \chi) = (A_1 + \rho A_2) \frac{(1 \pm \Sigma(\chi))^{1/\gamma}}{\frac{1}{2} \int \left[(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma} \right] dG(\chi')}, \quad (3.17)$$

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

$$q_k(\chi) = \frac{A_1 + \rho A_2}{\theta_k} \frac{(1 + \Sigma(\chi))^{1/\gamma} - (1 - \Sigma(\chi))^{1/\gamma}}{\frac{1}{2} \int [(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma}] dG(\chi')} \quad (3.18)$$

for $k \in \{1, 2\}$. Therefore, this setup provides an important simplification as there is 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 composite position will buy $q_k(\chi)$ units of asset k upon meeting a dealer; vice versa, a low-type client currently stuck with the high-type optimal composite position will sell $q_k(\chi)$ units, where $q_k(\chi) \equiv [\theta^*(2\sigma, \chi) - \theta^*(0, \chi)] / \theta_k$ is given by the explicit formula (3.18).

The equilibrium trade quantity (3.18) reveals the relationship between trade aggressiveness and exposure to frictions standard in search models with divisible assets. Controlling for χ_0 , a more active trader (i.e., with larger χ_1 or 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_1 or χ_2 increase. On the other hand, controlling for χ_1 and χ_2 , a larger χ_0 leads to lower $\Sigma(\chi)$ and lower trade quantities, because a large- χ_0 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 an effect on trade sizes opposing the effect of χ_0 makes trade size information very useful to simultaneously identify χ_0 , χ_1 , and χ_2 from the data. Other endogenous objects such as trading intensity are typically increasing in χ_0 . Therefore, adding trade-size-related information for asset 1 or asset 2 (government and corporate bonds in practice) 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 $\alpha_k(\chi) \equiv p_k(2\sigma, \theta^*(0, \chi), \chi)$ and $\beta_k(\chi) \equiv p_k(0, \theta^*(2\sigma, \chi), \chi)$ as the ask and the bid price a client of type χ faces when trading with a dealer and $\mathcal{A}_k(\chi) \equiv p_k^d(2\sigma, \theta^*(0, \chi), \chi)$ and $\mathcal{B}_k(\chi) \equiv p_k^d(0, \theta^*(2\sigma, \chi), \chi)$ as the ask and the bid price this dealer faces when trading with the core to offload the unwanted positions stemming from trading with the client- χ . We calculate these prices using (3.2) and (3.3):

$$\alpha_k(\chi) = \frac{(1 - \eta_k) q_k(\chi) P_k + \eta_k \{V(2\sigma, \theta^*(2\sigma, \chi), \chi) - V(2\sigma, \theta^*(0, \chi), \chi)\}}{q_k(\chi)}, \quad (3.19)$$

$$\beta_k(\chi) = \frac{(1 - \eta_k) q_k(\chi) P_k + \eta_k \{V(0, \theta^*(2\sigma, \chi), \chi) - V(0, \theta^*(0, \chi), \chi)\}}{q_k(\chi)}, \quad (3.20)$$

$$\mathcal{A}_k(\chi) = \frac{(1 - \eta_k^c) q_k(\chi) P_k + \eta_k^c \{V(2\sigma, \theta^*(2\sigma, \chi), \chi) - V(2\sigma, \theta^*(0, \chi), \chi)\}}{q_k(\chi)}, \quad (3.21)$$

and

$$\mathcal{B}_k(\chi) = \frac{(1 - \eta_k^c) q_k(\chi) P_k + \eta_k^c \{V(0, \theta^*(2\sigma, \chi), \chi) - V(0, \theta^*(0, \chi), \chi)\}}{q_k(\chi)} \quad (3.22)$$

for $k \in \{1, 2\}$. These formulas for bid and ask prices in the client-dealer market and the inter-dealer market are already intuitive in their current forms.²² When the client has all the bargaining power ($\eta_k \equiv \eta_k^d + \eta_k^c = 0$), all bid and ask prices are equal to P_k , which means that there is no price dispersion as dealers and core broker-dealers cannot capture any surplus. On the other extreme, when the core has all the bargaining power ($\eta_k^c = 1$), bid and ask prices are equal to the client's respective reservation prices, which leaves clients and dealers with zero transaction surplus and maximizes the bid-ask spread that goes to the core. Finally, if the dealer has all the bargaining power ($\eta_k^d = 1$), then, the bid and ask prices in the client-dealer market are equal to the client's reservation prices, which leaves the client with zero transaction surplus, but there is no bid-ask spread in the inter-dealer market as the dealer can buy and sell at P_k , which leaves the core with zero transaction surplus as well.

3.4.1 Concluding Characterization: The Three Dimensions of Market Liquidity

So far, we have derived *in closed form* all the theoretical equilibrium objects including the frictionless prices, clients' value functions, bilateral trade quantities and prices both in the client-dealer market and the inter-dealer market, as well as the equilibrium distribution in Proposition 2. Next, we derive a number of equilibrium objects that are readily observable from the data as (integral transforms of) closed-form expressions, which relate to the three main dimensions of market liquidity: trading frequencies, trade sizes, and transaction prices.

Clients' trade frequency Equation (3.13) effectively shows the mass of clients who are happy with their current composite holding and the mass of those who are unhappy with their current composite holding. Considering that only the latter type will trade in equilibrium, the rate at which the χ -clients trade asset k with dealers is

$$\chi_k \frac{1}{4} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2} dG(\chi) + \chi_k \frac{1}{4} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2} dG(\chi) = \chi_k \frac{1}{2} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2} dG(\chi),$$

where χ_k is the per-client meeting rate, $dG(\chi)$ is the mass of those clients, and $\frac{1}{2} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2}$ is the fraction of χ -clients who are unhappy with their current composite holding: Half of them are of low taste type but holding a high composite position and the other half are of high taste type but holding a low composite position. Thus, although a client with characteristics $\chi = (\chi_0, \chi_1, \chi_2)$ meets a dealer at the exogenous Poisson intensity of χ_k to trade asset k , her endogenous trading

²²In Appendix D, we use the closed-form formulae of the continuation utilities and derive more explicit expressions for these bilateral bid and ask prices, which are later used in our estimation algorithms.

intensity is

$$\lambda_k(\chi) \equiv \frac{\chi_k}{2} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2} \quad (3.23)$$

for $k \in \{1, 2\}$.

Trade sizes in the client-dealer market Equation (3.18) implies that the client- χ 's trade sizes of asset 1 and 2 are

$$q_1(\chi) = (A_1 + \rho A_2) \frac{(1 + \Sigma(\chi))^{1/\gamma} - (1 - \Sigma(\chi))^{1/\gamma}}{\frac{1}{2} \int [(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma}] dG(\chi')} \quad (3.24)$$

and

$$q_2(\chi) = \frac{A_1 + \rho A_2}{\rho} \frac{(1 + \Sigma(\chi))^{1/\gamma} - (1 - \Sigma(\chi))^{1/\gamma}}{\frac{1}{2} \int [(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma}] dG(\chi')},$$

respectively. That is, $q_1(\chi) = \rho q_2(\chi)$ holds for all $\chi \in \text{supp}(dG)$. This follows from our constant MRS assumption between the two assets. This result implies that the correlation coefficient between clients' trade sizes in the two markets must be +1. Empirically, we find that this correlation coefficient between clients' trade sizes in the UK's government and corporate bond markets is +0.8186 when we use the 57 most liquid corporate bonds and +0.8128 when we use all corporate bonds in our dataset. Therefore, although the empirical correlation is not as strong as the model-implied one, its level is broadly comparable if treated as a non-targeted moment for our estimation exercise below.

We utilize the relation $q_1(\chi) = \rho q_2(\chi)$ to calibrate the parameter ρ as the ratio of the aggregate trade size in the government bond market to the aggregate trade size in the corporate bond market. The previous two equations imply

$$2q_1(\chi) = 2\rho q_2(\chi) = q_1(\chi) + \rho q_2(\chi) = 2(A_1 + \rho A_2) \frac{(1 + \Sigma(\chi))^{1/\gamma} - (1 - \Sigma(\chi))^{1/\gamma}}{\frac{1}{2} \int [(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma}] dG(\chi')} \equiv \bar{q}(\chi), \quad (3.25)$$

which reveals that trade sizes provide us with essential information to recover each client's χ_0 , χ_1 , and χ_2 from the transactions data. Namely, given r , A_1 , A_2 , ρ , γ , η_1 , and η_2 and the empirical counterparts of $\lambda_1(\chi)$, $\lambda_2(\chi)$, and $\bar{q}(\chi)$, the system implied by (3.23) and (3.25) provides us with three equations to pin down χ_0 , χ_1 , and χ_2 .²³ Note that r and ρ are calibrated parameters. Below we design a minimum-distance estimation procedure to estimate the inverse elasticity, γ , of the utility function, the asset-specific bargaining power parameters, η_1 and η_2 , and the number of tradable shares of each asset, A_1 , A_2 , by using additional aggregate and client-specific moments.

²³Lemma 3 in Appendix A shows that this system of equations has a unique positive real solution for χ_0 , χ_1 , and χ_2 .

Hence, given these parameters, our model allows us to recover the client-specific parameters entirely transparently by using data on clients' trade frequencies and composite trade size.

Clients' trade costs Because of dealers' and core broker-dealers' ability to extract rents, clients trade at prices different from the frictionless price at which the core can trade amongst themselves, which means clients face endogenous trade costs in our model. To quantify these costs, we calculate the theoretical counterpart of the trade cost measure of [Pintér, Wang, and Zou \(2024\)](#) averaged across all clients:

$$\phi_k = \int \sum_{\theta \in \Theta(\chi)} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} (\log p_k(\varepsilon, \theta, \chi) - \log P_k) \left(\mathbb{I}_{\{\varepsilon = \varepsilon_h\}} - \mathbb{I}_{\{\varepsilon = \varepsilon_l\}} \right) f_\chi(\varepsilon, \theta) dG(\chi).$$

And, using the fact that half of the transactions happen at client-specific bid prices and the other half at client-specific ask prices thanks to our symmetric binary taste-type assumption,

$$\phi_k = \int \frac{\log \alpha_k(\chi) - \log \beta_k(\chi)}{2} dG(\chi). \quad (3.26)$$

Hence, the aggregate trade cost equals half the spread between the logged ask and logged bid prices, averaged across clients.

Price dispersion in the client-dealer market 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 frictionless core price normalized by the frictionless core price:

$$\sigma_{p_k} = \frac{\int \sum_{\theta \in \Theta(\chi)} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_k q_k(\chi) |p_k(\varepsilon, \theta, \chi) - P_k| f_\chi(\varepsilon, \theta) dG(\chi)}{P_k \int \sum_{\theta \in \Theta(\chi)} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_k q_k(\chi) f_\chi(\varepsilon, \theta) dG(\chi)}.$$

And, using the fact that half of the transactions happen at client-specific bid prices and the other half at client-specific ask prices thanks to our symmetric binary taste-type assumption,

$$\sigma_{p_k} = \frac{\int \lambda_k(\chi) q_k(\chi) \frac{\alpha_k(\chi) - \beta_k(\chi)}{2} dG(\chi)}{P_k \int \lambda_k(\chi) q_k(\chi) dG(\chi)}.$$

Hence, in the client-dealer market, the equilibrium price dispersion is equal to the half of (a weighted average of) the realized bid-ask spreads.

Price dispersion in the inter-dealer market Following the same steps as in the previous paragraph, one can show that the equilibrium price dispersion in the inter-dealer market is

$$\sigma_{p_k}^d = \frac{\int \lambda_k(\chi) q_k(\chi) \frac{\mathcal{A}_k(\chi) - \mathcal{B}_k(\chi)}{2} dG(\chi)}{P_k \int \lambda_k(\chi) q_k(\chi) dG(\chi)}.$$

Hence, in the inter-dealer market, the equilibrium price dispersion is equal to the half of (a weighted average of) the realized bid-ask spreads that dealers face when trading with core broker-dealers.

By using (3.19), (3.20), (3.21), and (3.22), one finds that

$$\frac{\sigma_{p_k}^d}{\sigma_{p_k}} = \frac{\eta_k^c}{\eta_k^c + \eta_k^d}. \quad (3.27)$$

That is, the ratio of the inter-dealer price dispersion to the client-dealer price dispersion observed in the data is informative about how the parameter $\eta_k = \eta_k^d + \eta_k^c$ is distributed across dealers and the core. More precisely, let $DR_k = \sigma_{p_k}^d / \sigma_{p_k}$ denote this price dispersion ratio observed in the data. Then, $\eta_k^d = (1 - DR_k) \eta_k$ and $\eta_k^c = DR_k \eta_k$. In our estimation exercises below, we will utilize these two identities to separately determine the market power that dealers have against clients and the market power that dealers face when trading in the inter-dealer market.

3.4.2 Welfare

One of our main motivations to write this model of OTC markets is to quantify the welfare loss caused by the frictions characteristic of these 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:

$$\mathbb{W}^{Eq} \equiv \frac{1}{r} \int \sum_{\theta \in \Theta(\chi)} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \varepsilon u(\theta) f_\chi(\varepsilon, \theta) dG(\chi) - \frac{1}{r} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} \varepsilon u(A_1 + \rho A_2). \quad (3.28)$$

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 transfers of numéraire between clients and dealers and those between dealers and the core 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 $\theta = a_1 + \rho a_2$. 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 .²⁴ Because we want $\mathbb{W}^{Eq}$ to capture the welfare created by OTC trading opportunities,

²⁴To be more precise, we make use of the equality $\frac{1}{r} = \int_0^\infty e^{-rt} dt$ for $r > 0$.

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 of asset 1 and 2 are always equal to the assets’ respective per-capita supplies, A_1 and A_2 , 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 without any constraint apart from the usual resource constraints:

$$\mathbb{W}^{FB} \equiv \max_{a_1(\varepsilon_l), a_1(\varepsilon_h), a_2(\varepsilon_l), a_2(\varepsilon_h)} \frac{1}{r} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} \varepsilon u[a_1(\varepsilon) + \rho a_2(\varepsilon)] - \frac{1}{r} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} \varepsilon u(A_1 + \rho A_2),$$

subject to

$$\sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} a_1(\varepsilon) = A_1,$$

$$\sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} a_2(\varepsilon) = A_2,$$

and

$$-M \leq a_k(\varepsilon) \leq M$$

for all $\varepsilon \in \{\varepsilon_l, \varepsilon_h\}$ and $k \in \{1, 2\}$. 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 third welfare measure we calculate is a 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:

$$\mathbb{W}^{SB} \equiv \int_0^\infty e^{-rt} \left\{ \int_{-M}^M \int_{-M}^M \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \varepsilon u(a_1 + \rho a_2) \Phi_\chi^*(\varepsilon, da_1, da_2 | t) dG(\chi) \right\} dt - \frac{1}{r} \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} \varepsilon u(A_1 + \rho A_2). \quad (3.29)$$

The planner maximizes $\mathbb{W}^{SB}$ with respect to controls, $q_k(\varepsilon, a_1, a_2, \chi | t)$, subject to the laws of motion for the state variables, $\Phi_\chi^*(\varepsilon, da_1, da_2 | t)$, and to the feasibility conditions of asset reallocation,

$$\int_{-M}^M \int_{-M}^M \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_1 q_1(\varepsilon, a_1, a_2, \chi | t) \Phi_\chi^*(\varepsilon, da_1, da_2 | t) dG(\chi) = 0 \quad (3.30)$$

and

$$\int \int \int \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_2 q_2(\varepsilon, a_1, a_2, \chi | t) \Phi_\chi^*(\varepsilon, da_1, da_2 | t) dG(\chi) = 0. \quad (3.31)$$

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, the values of social welfare evaluated at the first-best, the second-best, and the equilibrium allocations are*

$$\mathbb{W}^{FB} = \frac{\sigma (A_1 + \rho A_2)^{1-\gamma}}{r} \frac{1}{1-\gamma} (2^{1-\gamma} - 1),$$

$$\begin{aligned} \mathbb{W}^{SB} = & \frac{\sigma (A_1 + \rho A_2)^{1-\gamma}}{r} \frac{1}{1-\gamma} \\ & \times \left[\frac{\int \left[\left(\frac{2r+\chi_0+2\chi_1+2\chi_2}{r+\chi_0+\chi_1+\chi_2} \right)^{\frac{1}{\gamma}-1} \frac{\chi_1+\chi_2+\chi_0/2}{\chi_1+\chi_2+\chi_0} + \left(\frac{\chi_0}{r+\chi_0+\chi_1+\chi_2} \right)^{\frac{1}{\gamma}-1} \frac{\chi_0/2}{\chi_1+\chi_2+\chi_0} \right] dG(\chi)}{\left\{ \frac{1}{2} \int \left[\left(\frac{2r+\chi_0+2\chi_1+2\chi_2}{r+\chi_0+\chi_1+\chi_2} \right)^{1/\gamma} + \left(\frac{\chi_0}{r+\chi_0+\chi_1+\chi_2} \right)^{1/\gamma} \right] dG(\chi) \right\}^{1-\gamma}} - 1 \right], \end{aligned}$$

and

$$\begin{aligned} \mathbb{W}^{Eq} = & \frac{\sigma (A_1 + \rho A_2)^{1-\gamma}}{r} \frac{1}{1-\gamma} \\ & \times \left[\frac{\int \left[\left(\frac{2r+\chi_0+2\chi_1(1-\eta_1)+2\chi_2(1-\eta_2)}{r+\chi_0+\chi_1(1-\eta_1)+\chi_2(1-\eta_2)} \right)^{\frac{1}{\gamma}-1} \frac{\chi_1+\chi_2+\chi_0/2}{\chi_1+\chi_2+\chi_0} + \left(\frac{\chi_0}{r+\chi_0+\chi_1(1-\eta_1)+\chi_2(1-\eta_2)} \right)^{\frac{1}{\gamma}-1} \frac{\chi_0/2}{\chi_1+\chi_2+\chi_0} \right] dG(\chi)}{\left\{ \frac{1}{2} \int \left[\left(\frac{2r+\chi_0+2\chi_1(1-\eta_1)+2\chi_2(1-\eta_2)}{r+\chi_0+\chi_1(1-\eta_1)+\chi_2(1-\eta_2)} \right)^{1/\gamma} + \left(\frac{\chi_0}{r+\chi_0+\chi_1(1-\eta_1)+\chi_2(1-\eta_2)} \right)^{1/\gamma} \right] dG(\chi) \right\}^{1-\gamma}} - 1 \right], \end{aligned}$$

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 clients' share of transaction surplus were 100%. 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 $\eta_k > 0$. This means that while the source of inefficiency in this model is search frictions ($\chi_1 < \infty$ and $\chi_2 < \infty$), the source of *constrained* inefficiency is intermediation frictions ($\eta_1 > 0$ or $\eta_2 > 0$).

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 $\mathbb{W}^{Eq}$. 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 $(\mathbb{W}^{FB} - \mathbb{W}^{Eq}) / \mathbb{W}^{FB}$ to quantify the welfare loss from the real-world frictions relative to what would obtain in a perfect world. Then, we decompose this relative welfare loss to a component due to intermediation frictions, $(\mathbb{W}^{SB} - \mathbb{W}^{Eq}) / \mathbb{W}^{FB}$, and a component due purely to search frictions, $(\mathbb{W}^{FB} - \mathbb{W}^{SB}) / \mathbb{W}^{FB}$.

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 presented in Section 3.4 for the gilt market (asset 1) and the UK corporate bond market (asset 2). We have to determine ten parameters, r , σ , γ , ρ , A_1 , A_2 , η_1^d , η_2^d , η_1^c , and η_2^c and one distribution, $G(\chi_0, \chi_1, \chi_2)$.

4.1.1 Summary of Model Equations for Calibration and Estimation

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

$$\lambda_k(\chi) = \frac{\chi_k}{2} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2}, \quad (4.1)$$

$$\bar{q}(\chi) = 2(A_1 + \rho A_2) \frac{(1 + \Sigma(\chi))^{1/\gamma} - (1 - \Sigma(\chi))^{1/\gamma}}{\frac{1}{2} \int [(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma}] dG(\chi')}, \quad (4.2)$$

$$\begin{aligned} \mathcal{V}_k &= \int \lambda_k(\chi) q_k(\chi) dG(\chi) \\ &= \int \left(\lambda_k(\chi) \frac{A_1 + \rho A_2}{\theta_k} \frac{(1 + \Sigma(\chi))^{1/\gamma} - (1 - \Sigma(\chi))^{1/\gamma}}{\frac{1}{2} \int [(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma}] dG(\chi')} \right) dG(\chi), \end{aligned} \quad (4.3)$$

$$\phi_k = \int \frac{\log \alpha_k(\chi) - \log \beta_k(\chi)}{2} dG(\chi), \quad (4.4)$$

$$\sigma_{p_k} = \frac{\int \lambda_k(\chi) q_k(\chi) \frac{\alpha_k(\chi) - \beta_k(\chi)}{2} dG(\chi)}{P_k \int \lambda_k(\chi) q_k(\chi) dG(\chi)}, \quad (4.5)$$

$$\sigma_{p_k}^d = \frac{\int \lambda_k(\chi) q_k(\chi) \frac{\mathcal{A}_k(\chi) - \mathcal{B}_k(\chi)}{2} dG(\chi)}{P_k \int \lambda_k(\chi) q_k(\chi) dG(\chi)}. \quad (4.6)$$

Going over the endogenous measures listed above, the entire distribution G is necessary to calculate any quantity-based or price-based moment. Therefore, we try to determine it entirely. In particular, we estimate G “non-parametrically” by calculating the three-dimensional characteristics (χ_0, χ_1, χ_2) of each of the 526 clients in our sample. Lemma 3 of Appendix A shows that, given r , A_1 , A_2 , ρ , γ , η_1 , and η_2 , the equation system implied by (4.1) and (4.2) is sufficient to uniquely

pin down χ_0 , χ_1 , and χ_2 for each client, under mild regularity conditions.²⁵ As the empirical counterpart of $\bar{q}(\chi)$, we rely solely on corporate bond trade sizes, $\bar{q}(\chi) = 2\rho q_2(\chi)$, together with a calibrated value of ρ . Then, we let the empirical government bond trade sizes, $q_1(\chi)$, inform the estimation of market-wide parameters.

More precisely, to estimate five market-wide deep parameters, γ , η_1 , η_2 , A_1 , and A_2 , we use information from 529 different data moments: (i) trade volume, $\mathcal{V}_1$, calculated as above in (4.3) for the government bond market, (ii) aggregate trade cost, ϕ_1 , in the government bond market, (iii) aggregate trade cost, ϕ_2 , in the corporate bond market, and (iv) time-series average of government bond trade sizes, $q_1(\chi)$, for each of the 526 clients in our sample.

The remaining parameters are the preference parameters and the MRS between the two assets: r , σ , and ρ , as well as the decomposed bargaining powers for dealers and the core, η_1^d and η_1^c in the gilt market and η_2^d and η_2^c in the corporate bond market. We determine these parameters by calibration. We set r equal to 2.5% per annum, which approximately equals the yields on AAA-rated corporate bonds in our sample. This choice reflects the view that the government bond yield is a poor estimate for the true risk-free rate because of government bonds' convenience yield.²⁶ The MRS parameter, $\rho = 14.8085$, is calibrated to match the ratio of the aggregate trade size in the government bond market to the aggregate trade size in the corporate bond market. More precisely, we set ρ equal to the mean (across clients) of q_1^i/q_2^i , where q_k^i is the time-series average of client i 's trade sizes in market k . Given the estimated η_1 and η_2 , we calibrate η_1^d , η_1^c , η_2^d , and η_2^c to match the ratio of the inter-dealer price dispersion to the client-dealer price dispersion, that we name DR_k , both in the government bond and the corporate bond markets by using the model-implied identities $\eta_k^d = (1 - DR_k)\eta_k$ and $\eta_k^c = DR_k\eta_k$. It is straightforward to infer $DR_1 = 0.9389$ and $DR_2 = 0.6755$ from Table 1. 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, $\alpha_k(\chi)$, $\beta_k(\chi)$, $\mathcal{A}_k(\chi)$, $\mathcal{B}_k(\chi)$, and P_k . Hence, it does not affect our trade cost and price dispersion measures (4.4), (4.5), and (4.6), 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

We set up a minimum-distance estimation to estimate the (non-calibrated) parameters of our two-asset equilibrium model. Remember that we associate asset 1 with the gilt market and asset

²⁵More precisely, Lemma 3 states two sufficient conditions for uniqueness. First, the intertemporal elasticity of substitution, $1/\gamma$, must be less than 1. This assumption is well-supported by empirical studies in macroeconomics and asset pricing. See Vissing-Jørgensen (2002) and Hansen, Heaton, Lee, and Roussanov (2007), among others. Second, the highest values of $\bar{q}(\chi)$ should not be too large relative to the per-capita supplies of the assets. Our estimation algorithm is designed to ensure that these theoretical bounds are satisfied at the estimated parameters.

²⁶We thank Thanasis Geromichalos for this suggestion.

2 the corporate bond market. We have 529 moment conditions to estimate the parameter vector $\psi = (\gamma, \eta_1, \eta_2, \bar{A})$, which consists of four elements, where $\bar{A} \equiv A_1 + \rho A_2$.²⁷ Mathematically, ψ solves

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

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.

Our theoretical moments consist of three market-wide moments and 526 client-level moments. The three market-wide moments are trade volume in the government bond market and trade costs in the government bond and the corporate bond markets, computed according to (4.3) and (4.4). The 526 client-level moments are each client's government bond trade size computed according to (3.24). The sample moments, m_S , are computed at daily frequency, yielding 1440 observations for each moment condition. Given the relatively small sample, we follow [Altonji and Segal \(1996\)](#) in avoiding to estimate the optimal weighting matrix, which can exhibit poor finite sample properties.²⁸ More specifically, we “hand select” a diagonal matrix, where $\hat{W}_{ii} = 1$ for aggregate moments and $\hat{W}_{jj} = 1/N^2 \ll 1$ for client-level moments. The logic behind this choice is that client-level moments are already used for non-parametric exact identification of client characteristics, χ . The goal of the minimum-distance estimation stage is to identify the market-wide parameters, ψ . Hence, we place greater emphasis on market-wide moments by explicitly down-weighting client-level moments. This approach is in the spirit of macroeconomic studies that incorporate micro-level heterogeneity. Such studies often explicitly or implicitly give more weight to aggregate moments (e.g., GDP growth, inflation, aggregate investment) because these are often the primary targets of the model, and micro moments (e.g., household consumption distribution) are treated as secondary. See, for example, [Browning, Hansen, and Heckman \(1999\)](#), [Guvenen \(2009\)](#), and [Kaplan and Violante \(2014\)](#).

4.2 Empirical Results

As our structural model is a two-market model, the estimation is conducted jointly for the government and corporate bond markets. The parameter estimates are shown in Table 2 along with bootstrap standard errors in parentheses. Three observations are in order regarding the estimated parameters. First, the implied intertemporal elasticity of substitution (IES), $1/\gamma$, is approximately 0.025, indicating that UK fixed-income clients exhibit an exceptionally low willingness to intertemporally substitute their asset holdings. Second, clients' lost surplus share as measured

²⁷Going over the endogenous measures (4.1)-(4.6), one sees that $A_1 + \rho A_2$ shows up always as a block because of the constant MRS property of our model. Thus, it is possible to estimate $\bar{A}$, but not A_1 and A_2 separately.

²⁸See [Altonji and Segal \(1996\)](#) as well as [Honore, Jorgensen, and de Paula \(2020\)](#) for a recent discussion.

by η is substantially higher in the corporate bond market. More specifically, because they lack direct access to “frictionless” markets and must trade through intermediaries, clients lose 4% and 38% of trade surpluses in the government and corporate bond markets, respectively. Finally, the estimated number of tradable shares of the composite asset, $\bar{A} = A_1 + \rho A_2$, is around 38% of the total supply of composite assets in our dataset (the sum of the supply of government bonds and ρ times the supply of corporate bonds).²⁹ This aligns with the common view of fixed-income markets that once bonds exit their on-the-run stage, most holdings are locked away in the vaults of buy-and-hold investors, leaving little free float available to the clients who trade more actively.

Table 2: Parameter Estimates

γ – Curvature of the utility function	42.296 (6.068)
η_1 – Lost surplus share in gilt ($\eta_1^c + \eta_1^d$)	0.0415 (0.096)
η_2 – Lost surplus share in corporate ($\eta_2^c + \eta_2^d$)	0.383 (0.139)
$\bar{A}$ – Tradable portion of the asset supply (composite)	2.6715×10^7 (2.2476×10^7)

Notes: This table reports the estimates of the parameters. The sample contains 1440 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. The distributions of government bond trade intensities, corporate bond trade intensities, and corporate bond trade sizes are exactly matched. This results from the non-parametric exact identification of the joint distribution of client characteristics by using Lemma 3. In other words, 526×3 client-based data moments are used to identify 526×3 client-specific parameters in the model and Lemma 3 guarantees that the identification relies on a system of equations with unique solution. Then, the remaining 526 client-specific moments and three aggregate moments determine the remaining four aggregate parameters. Two of these aggregate moments, average trade cost in each market, are almost exactly matched. This is because these two moments are uniquely important for the identification of η_1 and η_2 . Finally, γ and $\bar{A}$ are identified by the remaining $526 + 1$ moments, the client-specific government bond trade sizes and the total trade volume of government bonds. Given the substantial over-identification of the system, the alignment of the moment conditions is necessarily imperfect. The misalignment is most pronounced in the client-based moments, owing to their relatively low weighting in the specification of our minimum-distance objective function (4.7).

²⁹The supply of government bonds is 4.9127×10^7 per client per bond. The supply of corporate bonds is 1.3815×10^6 per client per bond. These imply a supply of composite assets of 6.9585×10^7 per client per composite bond.

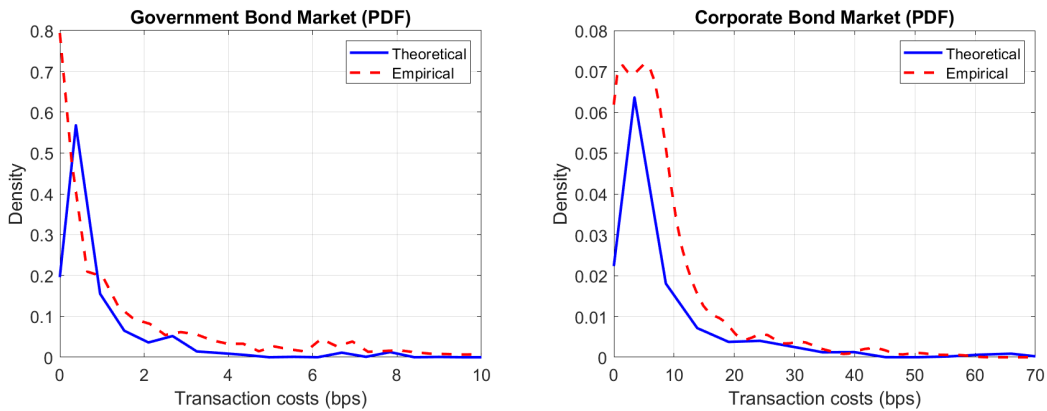
Table 3: Model Fit

Moments	Government Bonds		Corporate Bonds	
	Empirical (1)	Theoretical (2)	Empirical (3)	Theoretical (4)
Trade Intensities				
Median	0.0068	0.0068	0.0007	0.0007
p25	0.0022	0.0022	0.0002	0.0002
p75	0.0221	0.0221	0.0031	0.0031
Trade Sizes				
Median	2,236,224	4,067,963	274,704	274,704
p25	365,051	819,130	55,315	55,315
p75	6,954,749	10,875,793	734,427	734,427
Average Trade Cost	0.7699	0.7699	7.0814	7.0815
Trade Volume	281,460	258,758	—	—

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.

While Table 3 illustrates the fit of the targeted moments, Figure 3 plots the entire distribution of trade costs in the cross section of clients in both markets, which are non-targeted functions apart from their implied first moments. Figure 3 shows that the distributions of clients' trade cost 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 (e.g., market-wide bargaining power parameters and common preference parameters, instead of client-specific ones) do a good job in generating realistically skewed distributions of endogenous trade cost in both markets.

Figure 3: Probability Density Functions of Clients' Trade Costs



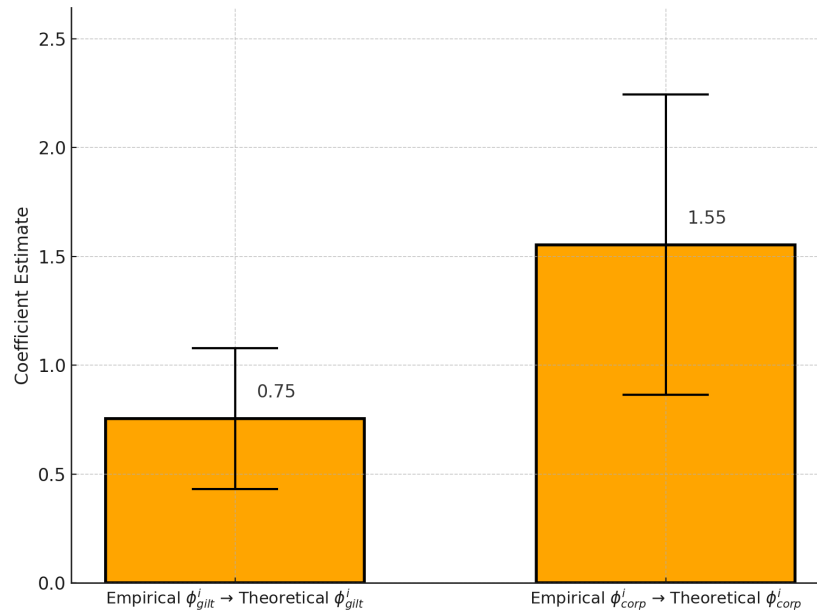
Notes: This figure shows the theoretical and empirical probability density functions (pdf) of client-specific average trade costs for the gilt market in the left panel and for the UK corporate bond market in the right. Theoretical pdfs are the pdf of the endogenous object $(\log \alpha_k(\chi) - \log \beta_k(\chi)) / 2$ as seen inside the integral (3.26), 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 average trade cost.

Although Figure 3 visually demonstrates the close alignment between the theoretical and empirical distributions of clients' trade costs, we also estimate the following cross-sectional regression to convey this match more rigorously:

$$\text{theoretical } \phi_k^i = \beta_k \times (\text{empirical } \phi_k^i) + \epsilon_k^i, \quad (4.8)$$

where ϕ_k^i is client i 's trade cost in market k . We run the regression separately for the government bond ($k = \text{gilt}$) and the corporate bond markets ($k = \text{corp}$). Note that a perfect alignment of the theoretical and empirical trade costs would imply $\beta_{\text{gilt}} = 1$ and $\beta_{\text{corp}} = 1$, which constitute our null hypotheses in this regression analysis. Figure 4 presents the regression results. The coefficient estimates, $\hat{\beta}_{\text{gilt}} = 0.75$ and $\hat{\beta}_{\text{corp}} = 1.55$, are both in the ballpark of one. Moreover, the value of one falls within the 95% confidence intervals for both markets' regressions, indicating that our null hypotheses are not rejected. We interpret this as strong evidence of the model's ability to match the distribution of clients' trade costs—despite these being non-targeted moments. Notably, our estimation procedure only targets the average trade cost in the cross section of clients. Thus, any theoretical client heterogeneity our model estimation captures are based on clients' trade frequencies and trade sizes as well as market-wide moments including the average trade cost in the cross section of clients. However, as demonstrated in Figures 3 and 4, the model realistically reproduces the entire distribution of clients' trade costs as well.

Figure 4: Regression Estimates with 95% Confidence Intervals



Notes: Regression estimates of theoretical ϕ^i on empirical ϕ^i for gilt and corporate bonds as defined in (4.8). For gilt bonds, the coefficient estimate is 0.75 with a robust standard error of 0.17, a 95% confidence interval of [0.43, 1.08], and an R^2 of 0.724. For corporate bonds, the coefficient estimate is 1.55 with a robust standard error of 0.35, a 95% confidence interval of [0.86, 2.24], and an R^2 of 0.517. Error bars denote 95% confidence intervals based on robust standard errors.

4.2.1 Estimated Client Characteristics

In this section, we present the results related to the estimated joint distribution of clients' taste shock intensity, meeting rate with dealers to trade government bonds, and meeting rate with dealers to trade corporate bonds. Table 4 provides the summary statistics for the marginal distributions of these three client characteristics. The median client's daily taste shock rate is 0.0275, which implies that the median client wants to trade a particular government or corporate bond once every 36 days. The average taste shock rate is about 4.2 times as large as the median taste shock rate, implying a right-skewed distribution. When we look at the meeting rates, the median client gets to meet a gilt dealer 6.77 times per day and a corporate bond dealer 0.56 times a day. The averages are substantially higher than their respective medians in each market, which again imply substantial right-skewness. This can also be easily confirmed with the substantial asymmetries of the 25th percentile and the 75th percentile from their respective medians in each market. That is, the median meeting rate is closer to the 25th percentile than the 75th percentile in each market.

Table 4: Distribution of Client Characteristics

Variable	Taste Shock Rate (1)	Gilt Meeting Rate (2)	Corporate Meeting Rate (3)
Average	0.11688	4,350,993	199,241
Standard Deviation	0.23721	1.41×10^7	744,350
Median	0.027535	6.7716	0.56291
p25	0.0090767	0.033986	0.012819
p75	0.10694	1,195,102	39,966

Notes: This table summarizes the marginal distributions for each of the three dimensions of clients' characteristics, recovered by our estimation algorithm. Columns (1), (2), and (3) report summary statistics for the marginal distribution of clients' taste shock intensity (χ_0), clients' meeting rate with dealers to trade a government bond (χ_1), and clients' meeting rate with dealers to trade a corporate bond (χ_2), respectively.

Note that our structural model treats all three dimensions of client characteristics as exogenous parameters, and we estimate these client-specific parameters by effectively “letting the data speak” on the stage set by our structural model. However, one could take a step back and argue that clients must have some degree of control over their meeting rates. In fact, the literature provides micro-foundations for endogenous meeting rates. For example, [Vayanos and Wang \(2007\)](#) model endogenous market segmentation, allowing agents with different taste shock intensities to choose the market segment in which they trade. They show that an asymmetric equilibrium emerges in which high-intensity agents trade in more liquid segments, while low-intensity agents trade in less liquid ones. Similarly, [Hendershott, Li, Livdan, and Schürhoff \(2020\)](#) endogenize client-dealer relationships by allowing clients to choose the number of dealers in their network. In their model, the optimal number of dealer relationships increases with taste shock intensity, enabling high-intensity clients to trade more frequently than those with lower intensity. We next analyze whether our estimated client characteristics are consistent with these theories.

Table 5 presents results for regressions where we used taste shock intensity as the explanatory variable for meeting rates in the government and corporate bond markets. Panel 1 and 3 confirm that clients, who have a higher exposure to taste shocks, are likely to receive more trading opportunities in both markets, compatible with existing OTC market theories of endogenous meeting rates. Panel 2 and 4 investigate if the relation between taste shock intensity and meeting rates is concave or convex. The negative slope coefficients for χ_0^2 imply that the relation is concave in both markets. We then use the slope coefficients reported in Panel 2 and 4 and determine the level of taste shock intensities, $\chi_0^{g*} \approx 1.1$ and $\chi_0^{c*} \approx 1.5$, as maximizing the meeting rate in the government bond and the corporate bond market, respectively. This implies that taste shock intensity has a positive impact on meeting rates for more than 95% of clients in the government bond market and for more than 99% of clients in the corporate bond market. Hence, we conclude that the relation between taste shock intensity and meeting rates is positive and concave, but not really hump-shaped.

Table 5: Explaining Meeting Rates

	Gilt Meeting Rate (χ_1)		Corporate Meeting Rate (χ_2)	
	(1)	(2)	(3)	(4)
χ_0	2.62×10^7 (3.72)	5.43×10^7 (3.85)	1,165,433 (3.23)	1,881,284 (2.59)
χ_0^2		-2.47×10^7 (-1.95)		$-629,419.3$ (-0.90)
Intercept	1,290,179 (2.62)	$-270,156$ (-0.47)	63,025.05 (2.14)	23,304.65 (0.74)
Observations	526	526	526	526
R^2	0.1953	0.2394	0.1379	0.1481

Notes: This table presents OLS regression estimates examining the relationship between meeting rates and the variables χ_0 and χ_0^2 , with separate models for gilt meeting rates (χ_1) and corporate meeting rates (χ_2). Columns (1) and (2) report estimates for gilt meeting rates, while columns (3) and (4) report estimates for corporate meeting rates. The t -statistics are based on robust standard errors and shown in parentheses.

4.3 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 intermediaries' bargaining power. Table 6 shows the results for the two markets. Trading delays are measured in days, while bargaining power is defined as the fraction of trade surplus captured by the agent. Mathematically, the expected trading delay of a client before she trades some government bond is $1/(57\chi_1)$ and the share of transaction surplus she loses to intermediaries is $\eta_1 = \eta_1^d + \eta_1^c$. The same numbers in the corporate bond market are $1/(57\chi_2)$ and $\eta_2 = \eta_2^d + \eta_2^c$, respectively.

Trading delays in the UK government bond market are minimal, typically on the order of

minutes. The median client searches for less than five minutes to trade a representative gilt, rendering such delays negligible—consistent with the calibration results of [Vayanos and Weill \(2008\)](#) for the US government bond market. In contrast, trading delays in the corporate bond market are more pronounced: the median client spends approximately 45 minutes searching to trade a representative corporate bond. Moreover, trading delays vary significantly across clients in both markets. At the 25th percentile, clients face virtually no delay in either market. However, at the 75th percentile, clients require half a day to execute a gilt transaction and nearly 1.4 days to execute a corporate bond transaction.

Table 6: Welfare Results I: Estimated Trading Delays and Dealers’ Bargaining Power

	Government Bonds (1)	Corporate Bonds (2)
Average Trading Delays		
Median	0.0026	0.0314
p25	0.0000	0.0000
p75	0.5162	1.3686
Client’s Lost Surplus Share	4.15%	38.30%
Dealer’s Bargaining Share	0.25%	12.43%
Core Broker-Dealer’s Barg. Share	3.89%	25.87%

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 day.

The lower panel of Table 6 shows the estimated shares of transaction surplus captured by dealers and core broker-dealers in each market, which we refer to as intermediaries’ bargaining power in short. The respective estimates of $\eta_1 = 4.15\%$ and $\eta_2 = 38.3\%$ for intermediaries’ bargaining power in the gilt and the corporate bond market confirm the common view that intermediaries generally have larger market power in corporate bond markets than in government bond markets.³⁰ If the interdealer market were frictionless, the estimates of clients’ lost surplus shares in the two markets would fully reflect the market power dealers exert over their clients. However, the substantial inter-dealer price dispersion documented in Table 1 for both the government and corporate bond markets indicates that interdealer trading is itself subject to frictions. In our theoretical model, we capture this with a non-zero bargaining share that goes to core-broker dealers as dealers offload unwanted positions in the interdealer market. Using equation (3.27), the lower panel of Table 6 decomposes the clients’ lost surplus share into the bargaining shares of dealers and core broker-dealers. The quantitative results suggest that the trade costs faced by clients in the UK’s fixed-income markets arise primarily from frictions in the interdealer market, rather than from dealers exercising market power over clients.

Table 7 estimates the relative welfare loss caused by the presence of OTC market frictions.

³⁰See [Di Maggio, Kermani, and Song \(2017\)](#), [Biais and Green \(2019\)](#), and [Hendershott, Li, Livdan, and Schürhoff \(2020\)](#) for a discussion.

To calculate the welfare loss separately for each bond market, we use the single-asset counterpart of the formulas presented in Proposition 4. More precisely, when calculating the welfare loss in the gilt market, we set all clients' χ_2 s and the tradable shares of corporate bonds A_2 to zero, as well as halving each client's χ_0 . Similarly, when calculating the welfare loss in the corporate bond market, we set all clients' χ_1 s and the tradable shares of gilts A_1 to zero, as well as halving each client's χ_0 .

The top panel of Table 7 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 larger (more than twice) in the corporate bond market than in the gilt market.

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 intermediation frictions including imperfect competition between dealers and interdealer market frictions. The bottom panel of Table 7 shows that, in both markets, almost the entire loss can be attributed to search frictions with dealers' market power making a small contribution.

Table 7: Welfare Results II: Estimated Welfare Losses

	Government Bonds	Corporate Bonds
	(1)	(2)
Min. - Max. Welfare Loss	2.14% - 3.00%	4.36% - 6.31%
Welfare Loss	2.3778%	5.0463%
Due to Search Frictions	2.3774%	4.9464%
Extensive Margin	2.3774%	4.9464%
Intensive Margin	0.0000%	0.0000%
Due to Intermediation Frictions	0.0004%	0.0999%

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.

The final exercise using our estimation is to further decompose the welfare loss arising purely from search frictions into two components. The first is an intensive margin effect, which captures the clients' endogenous demand cutting behavior in response to search frictions. The second is an extensive margin effect, which reflects the fact that some clients remain stuck with suboptimal asset positions for a period of time—even if they choose trade quantities as though they were trading in a Walrasian market. Our estimates suggest that in both markets, the intensive margin

effect is negligibly small. This indicates that the main source of welfare loss is the inevitable delay between a client experiencing a taste shock and her next meeting with a willing dealer.

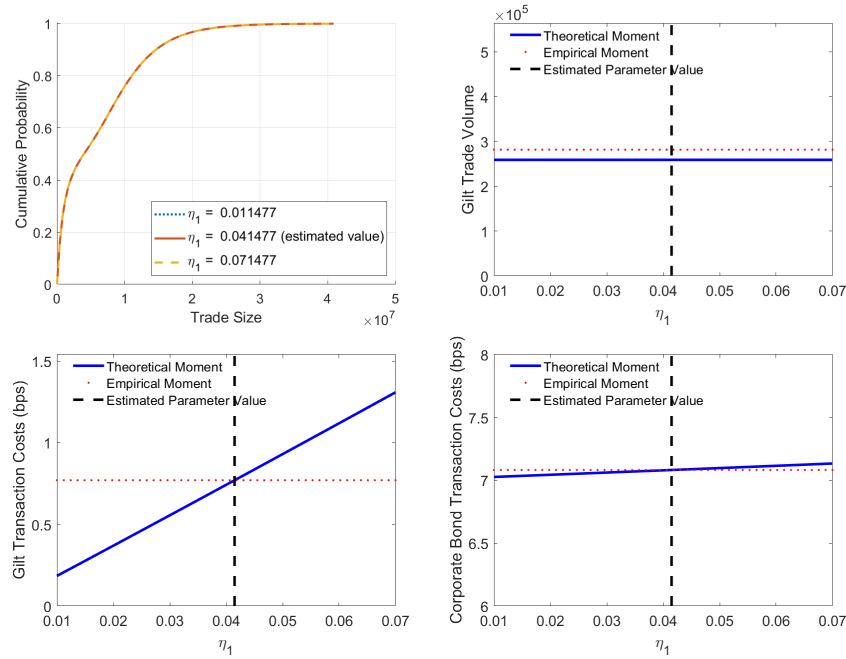
Taking stock, Table 7 underscores two key insights. First, even if the estimated median trading delays sound reasonably small (a few minutes in the gilt market), the welfare loss caused by them can be sizable. In line with the findings in Table 6, a significant share of the welfare loss arises from clients in the 75th percentile and above in terms of trading delays. This highlights the importance of incorporating client heterogeneity into structural models to ensure accurate counterfactual and normative analyses. Second, in the UK's fixed-income markets, the primary source of welfare loss arises from the extensive margin of search frictions, while frictions affecting trade sizes—namely, the intensive margin of search frictions and intermediation frictions—contribute negligibly. This is due to the remarkably low estimate of the IES, $1/\gamma = 0.0236$, which implies that even in the absence of frictions, clients would already prefer to avoid aggressive trading; as a result, frictions do not affect trade sizes in a way that substantially contributes to welfare losses. This low IES estimate is consistent with the findings of Hall (1988) and Campbell and Mankiw (1989), who report that household consumption-saving behavior implies an IES statistically indistinguishable from zero.³¹

4.4 Sources of Identification for the Market-wide Parameters

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 Figures 5–8, we inspect how the four (sets of) moment conditions change as we perturb a given parameter around its estimated value.

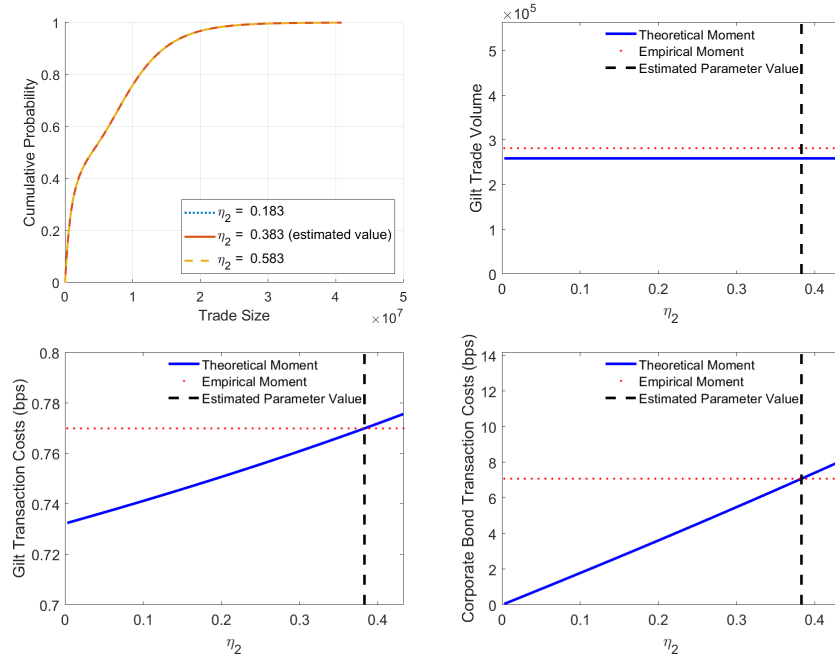
³¹More moderate estimates of the IES appear in the macroeconomics and finance literature. For instance, Yogo (2004) estimates values between 0.03 and 0.06 (for the US), while Vissing-Jørgensen (2002) reports values in the range of 0.3 to 0.4. Accordingly, in Appendix G, we conduct a robustness exercise by recalculating welfare losses for IES values of 0.05 ($\gamma = 20$) and 0.5 ($\gamma = 2$), holding all other parameters constant. Although total welfare losses are higher under these alternative values, intermediation frictions and the intensive margin of search frictions continue to contribute only marginally. Thus, even an IES as high as 0.5 is sufficiently low to preserve the robustness of our qualitative welfare conclusions.

Figure 5: Gilt Dealers' Market Power (η_1)



Note: This figure shows how the theoretical moments change as we perturb the surplus share, η_1 , that clients lose to dealers and core-broker dealers when bargaining to trade a government bond around its estimated value, while keeping all the other parameters fixed at their estimated values. The top-left, top-right, bottom-left and bottom-right panels show the results for the distribution of clients' trade size in the gilt market, total trade volume in the gilt market, average trade cost in the gilt market, and average trade cost in the corporate bond market, respectively.

Figure 6: Corporate Dealers' Market Power (η_2)



Note: This figure shows how the theoretical moments change as we perturb the surplus share, η_2 , that clients lose to dealers and core-broker dealers when bargaining to trade a corporate bond around its estimated value, while keeping all the other parameters fixed at their estimated values. The top-left, top-right, bottom-left and bottom-right panels show the results for the distribution of clients' trade size in the gilt market, total trade volume in the gilt market, average trade cost in the gilt market, and average trade cost in the corporate bond market, respectively.

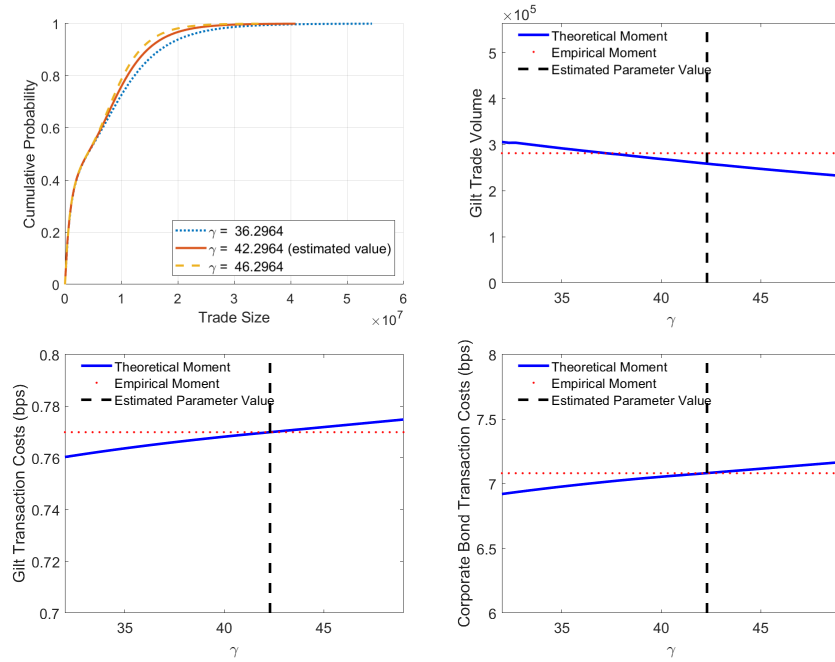
To begin with, Figure 5 illustrates how the theoretical moments respond to perturbations in the gilt dealers' market power parameter, η_1 , around its estimated value, holding all other parameters constant. Similarly, Figure 6 shows the sensitivity of the theoretical moments to variations in the corporate bond dealers' market power parameter, η_2 , again keeping all other parameters fixed at their estimated values. The bottom panels of both figures demonstrate that increasing either η_1 or η_2 leads to higher average trade costs in both markets. In contrast, the top panels indicate that changes in η_1 or η_2 have negligible effects on clients' trade sizes and total trade volume in the gilt market. This pattern suggests that the average trade cost moments in the two markets are the primary identifiers of η_1 and η_2 . Indeed, as also shown in Table 3, the empirical average trade costs closely align with their theoretical counterparts. Figures 5 and 6 confirm that this close match is achieved precisely because the estimation algorithm successfully recovers the true values of η_1 and η_2 .

The remaining two parameters to be determined are the curvature parameter of the utility function, γ , and the tradable portion of the composite asset, $\bar{A}$.³² Figures 7 and 8 show that these parameters are jointly identified by the distribution of clients' trade sizes in the gilt market and the total trade volume in the gilt market. As implied by Equations (3.18) and (4.3), each client's gilt trade size and the aggregate gilt trade volume scale proportionally with $A_1 + \rho A_2$. This relationship is reflected in the top panel of Figure 8, where an increase in the tradable portion of the assets results in a rightward shift of the cdf of trade sizes and a corresponding rise in total gilt trade volume. In contrast, the bottom panel of Figure 8 shows that changes in $\bar{A}$ have a negligible effect on average trade costs in both markets. We therefore conclude that the identification of $\bar{A}$ relies primarily on the distribution of trade sizes and the total volume in the gilt market, rather than on trade cost moments.

Finally, Figure 7 demonstrates that the distribution of trade sizes and the total volume in the gilt market contribute to the identification of the curvature parameter of clients' utility function, γ . An increase in γ makes the utility function more concave, rendering clients more reluctant to deviate from a "balanced" asset holding. Accordingly, the top panel of Figure 7 shows that a higher γ reduces both individual gilt trade sizes and aggregate gilt trade volume. Taken together with Figure 8, we conclude that γ and $\bar{A}$ are jointly identified by the distribution of clients' trade sizes and the total volume in the gilt market. The bottom panel of Figure 7 indicates that changes in γ also influence average trade costs in both the gilt and corporate bond markets. However, since these price-based moments—i.e., average trade costs—are primarily informative for identifying η_1 and η_2 , we regard the size- and volume-based moments as the key drivers for the identification of γ .

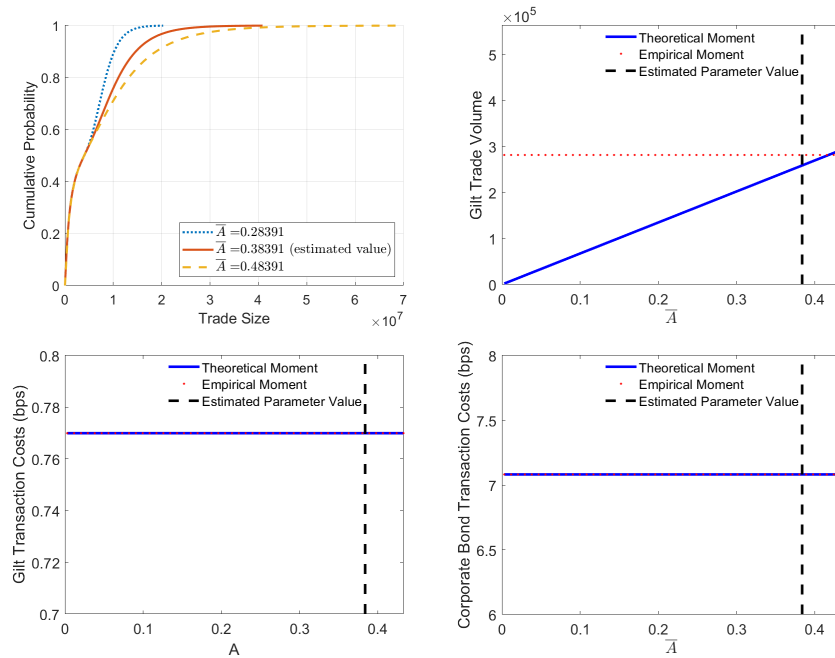
³²In this section and the accompanying tables, we scale the number of tradable shares of the composite asset by the empirical supply of composite assets, which is 6.9585×10^7 per client per composite bond. Accordingly, $\bar{A}$ should be interpreted as the fraction of assets that are in free float.

Figure 7: Curvature of the Common Utility (γ)



Note: This figure shows how the theoretical moments change as we perturb the curvature parameter, γ , of clients' common utility function around its estimated value, while keeping all the other parameters fixed at their estimated values. The top-left, top-right, bottom-left and bottom-right panels show the results for the distribution of clients' trade size in the gilt market, total trade volume in the gilt market, average trade cost in the gilt market, and average trade cost in the corporate bond market, respectively.

Figure 8: Tradable Portion of the Assets ($\bar{A}$)



Note: This figure shows how the theoretical moments change as we perturb the tradable portion, $\bar{A}$, of the assets around its estimated value, while keeping all the other parameters fixed at their estimated values. The top-left, top-right, bottom-left and bottom-right panels show the results for the distribution of clients' trade size in the gilt market, total trade volume in the gilt market, average trade cost in the gilt market, and average trade cost in the corporate bond market, respectively.

4.5 Monte Carlo Evidence

To assess whether the four market-wide parameters in our model are recoverable by minimizing (4.7), we have conducted a series of Monte Carlo simulations. Table 8 reports the details of an example. In this particular simulation, we simulate 100 trade-level datasets, whereby each dataset consists of 20 days, with 526 clients trading in each dataset. We use the parameter values described by column 1 of Table 8. To gauge whether our estimation is robust to small levels of unobserved heterogeneity, we simulated the Monte Carlo datasets from an extended version of our model in which clients are also heterogeneous in their utility function parameter γ .

Table 8: Monte Carlo Results

	True Values	Median of Rec.	Mean of Rec.	Std of Rec.
	(1)	(2)	(3)	(4)
γ – Curvature of the utility function	42	40.0060	40.0168	10.2517
η_1 – Lost surplus share in market 1	0.04	0.0371	0.0371	0.0048
η_2 – Lost surplus share in market 2	0.38	0.3586	0.3592	0.045
$\bar{A}$ – Tradable portion of the assets	2.642×10^7	2.7458×10^7	2.7617×10^7	2.1829×10^7

Notes: This table shows the results from a Monte Carlo simulation, where the data generation used the parameters shown in column 1, with the client-specific utility function parameters drawn from $U([40, 44])$. The median (column 2), mean (column 3), and standard deviation (column 4) of recovered parameters are based on the 100 Monte Carlo simulated samples.

We set the parameters η_1 , η_2 , and $\bar{A}$ to values stated in column 1 of Table 8. We also set each client's characteristics, χ , to our point estimates from the actual data. For each dataset, each client's utility curvature γ is simulated from the uniform distribution on the interval $[40, 44]$. That is, while clients have the same characteristics, χ , in each dataset, their γ differs from dataset to dataset, with its cross-sectional distribution $U([40, 44])$ staying constant. For each market, on each day and in each dataset, the simulation proceeds in four steps: (i) the Poisson arrival rates implied by clients' χ_0 parameters are used to simulate client-specific taste shock processes; (ii) the Poisson arrival rates implied by clients' χ_1 and χ_2 parameters are used to simulate client-specific random times of trade opportunities in the two markets; (iii) assuming random initial asset holding across clients consistent with the stationary equilibrium distribution, the information obtained in steps (i)-(ii) 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; (iv) the simulated datasets are used to compute the 526×4 client-specific and three market-wide moments that correspond to the empirical moments used in our baseline estimation in Section 4.1.2.

Columns 2-4 of Table 8 present the results from the Monte Carlo exercise. We find that minimizing (4.7) is sufficient to recover the four market-wide parameters with reasonable precision.³³ Compared to the bootstrap standard errors reported in Table 2, the Monte Carlo estimates exhibit

³³Note that our estimation strategy exactly recovers the client-specific parameters, χ_0 , χ_1 , and χ_2 , for each client in each dataset.

smaller standard deviations for η_1 , η_2 , and $\bar{A}$, but a larger standard deviation for γ . This discrepancy arises because the simulated data incorporate “noisy” client-specific values of γ , whereas the estimation model used to minimize (4.7) assumes a common γ across all clients. These results also suggest that while unobserved heterogeneity in client preferences may introduce noise into the empirical estimate of γ , the extent of this heterogeneity in the actual data is likely less pronounced than in our Monte Carlo setup, which assumes a uniform distribution over an interval of length 4.

5 Conclusion

A theoretical literature following Duffie, Gârleanu, and Pedersen (2005) and Lagos and Rocheteau (2009) generated a wealth of qualitative insights into the role of search and intermediation frictions in OTC financial markets. We have developed a dynamic structural model that uses transaction-level data with client identities and quantifies the search and intermediation frictions in question. Utilizing data on the UK government and corporate bond markets, we find that the median client’s search time in the corporate bond market is 12 times longer than that in the government bond market. Our estimates imply that clients lose approximately 4% of the transaction surplus in government bonds and 38% in corporate bonds. That is, government bond clients experience relatively small losses, mostly due to the passthrough of inter-dealer frictions, while corporate bond clients are exposed to both more substantial passthrough of inter-dealer frictions and dealers’ market power. Furthermore, we find that the welfare losses from frictions in the government and corporate bond markets are 2.38% and 5.05%, respectively, and our decomposition implies that these welfare losses are almost exclusively caused by search frictions in both markets. Finally, using data from the COVID-19 crisis period, we find that the welfare losses might more than double during turbulent times—rising to 3.63% in the government bond market and 11.35% in the corporate bond market—with intermediation frictions playing a much larger role in the latter. These estimates underline the fragility of the OTC market structure in the face of large aggregate shocks.³⁴

A future avenue for research is to incorporate additional inter-dealer market frictions 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.

³⁴For more details, see Appendix B.

References

- Rui Albuquerque and Enrique Schroth. The value of control and the costs of illiquidity. *Journal of Finance*, 70(4):1405–1455, 2015.
- 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.
- 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 analysis of 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.
- Bruno Biais and Richard C. Green. The microstructure of the bond market in the 20th century. *Review of Economic Dynamics*, 33:250–271, 2019.
- Giulia Brancaccio and Karam Kang. Search frictions and product design in the municipal bond market. *Econometrica*, Forthcoming, 2024.
- Giulia Brancaccio, Dan Li, and Norman Schürhoff. Learning by trading: The case of the u.s. market for municipal bond. Working paper, 2017.
- Martin Browning, Lars Peter Hansen, and James J Heckman. Micro data and general equilibrium models. In J.B. Taylor and M. Woodford, editors, *Handbook of Macroeconomics*, volume 1, chapter 8, pages 543–633. Elsevier Science B.V., 1999.
- Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru. Why is intermediating houses so difficult? evidence from ibuyers. Working paper, 2020.

- John Y Campbell and N Gregory Mankiw. Consumption, income, and interest rates: Reinterpreting the time series evidence. *NBER Macroeconomics Annual*, 4:185–246, 1989.
- 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. Risk concentration and interconnectedness in otc markets. Working paper, 2021.
- Jonathan Chiu, Mohammad Davoodalhosseini, and Janet Hua Jiang. Dealer inventory and market liquidity. Mimeo, 2022.
- 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.
- Assa Cohen, Mahyar Kargar, Benjamin Lester, and Pierre-Olivier Weill. Inventory, market making, and liquidity in otc markets. Working paper, 2023.
- 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.
- 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.

- 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.
- Andrea L Eisfeldt, Bernard Herskovic, and Shuo Liu. Interdealer price dispersion and intermediary capacity. Technical report, National Bureau of Economic Research, 2024.
- 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.
- Fatih Guvenen. An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12:58–79, 2009.

- 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.
- Robert E. Hall. Intertemporal substitution in consumption. *Journal of Political Economy*, 96(2): 339–357, 1988.
- Lars Peter Hansen, John Heaton, Junghoon Lee, and Nikolai Roussanov. Intertemporal substitution and risk aversion. In *Handbook of Econometrics*, volume 6A, pages 3967–4056. Elsevier B.V., North-Holland, Amsterdam, 2007.
- 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.
- Greg Kaplan and Giovanni L Violante. A model of the consumption response to fiscal stimulus payments. *Econometrica*, 82(4):1199–1239, 2014.
- Mahyar Kargar, Juan Passadore, and Dejanir Silva. Liquidity and risk in otc markets: A theory of asset pricing and portfolio flows. Working paper, 2020.
- 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, Sébastien Plante, and Pierre-Olivier Weill. Sequential search for corporate bonds. Working paper, 2023.
- 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.
- Vijay Krishna and Roberto Serrano. Multilateral bargaining. *Review of Economic Studies*, 63(1): 61–80, 1996.
- 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.
- Terje Lensberg. Stability and the nash solution. *Journal of Economic Theory*, 45:330–341, 1988.
- 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.
- Jessica S Li. Price and liquidity in frictional markets with strategic bargaining. Working paper, 2024.
- Shuo Liu. Dealer’s search intensity in u.s. corporate bond markets. Working paper, 2020.
- Dong Lu, Daniela Puzzello, and Aiyong Zhu. Search frictions in over-the-counter foreign exchange markets: Theory and evidence. Working paper, 2023.
- 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.
- Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic Theory*. Oxford University Press, Oxford, UK, 1995.
- 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.
- Gábor Pintér, Chaojun Wang, and Junyuan Zou. Size discount and size penalty: Trading costs in bond markets. *Review of Financial Studies*, 37(7):2156–2190, 2024.
- Tavy Ronen and Xing Zhou. Trade and information in the corporate bond market. *Journal of Financial Markets*, 16(1):61–103, 2013.
- Nancy L. Stokey and Robert E. Jr. Lucas. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, MA, 1989.
- Sang-Chul Suh and Quan Wen. Multi-agent bilateral bargaining and the nash bargaining solution. *Journal of Mathematical Economics*, 42:61–73, 2006.
- 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.
- Annette Vissing-Jørgensen. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853, 2002.
- 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.
- Motohiro Yogo. Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics*, 86(3):797–810, 2004.

A Identification of Clients' Characteristics

Lemma 1. Assume $\gamma > 1$. Let $L_\gamma = \max_{0 \leq y \leq 2^{1/\gamma}} y^{\gamma-1} (2-y^\gamma)^{\frac{1}{\gamma}-1}$. Let $\hat{q} : [\chi_l, \chi_h]^3 \rightarrow [0, Q]$ be a bounded measurable function for $Q < \frac{1-L_\gamma}{1+L_\gamma}$. Then, there exists a unique $\Sigma : [\chi_l, \chi_h]^3 \rightarrow [0, 1]$ that solves

$$\hat{q}(\chi) = \frac{(1 + \Sigma(\chi))^{1/\gamma} - (1 - \Sigma(\chi))^{1/\gamma}}{\int \left[(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma} \right] dG(\chi')}. \quad (\text{A.1})$$

Proof. Define $\delta(\chi) \equiv (1 + \Sigma(\chi))^{1/\gamma}$ and $h(y) \equiv (2 - y^\gamma)^{1/\gamma}$. Then, (A.1) can be re-written as

$$\delta(\chi) = h[\delta(\chi)] + \hat{q}(\chi) \int \{ \delta(\chi') + h[\delta(\chi')] \} dG(\chi'). \quad (\text{A.2})$$

Define the operator T acting on δ as

$$T\delta(\chi) \equiv h[\delta(\chi)] + \hat{q}(\chi) \int \{ \delta(\chi') + h[\delta(\chi')] \} dG(\chi').$$

The functional equation (A.2) can now be written as

$$T\delta = \delta.$$

To prove the existence and uniqueness of δ , one must analyze whether T is a contraction mapping under the stated conditions. Let $\mathcal{B}' = \{f : [\chi_l, \chi_h]^3 \rightarrow [0, 1]\}$. Note that h is Lipschitz continuous with the Lipschitz constant $L_\gamma < 1$ for $\gamma > 1$. Consider an arbitrary pair $\delta^1, \delta^2 \in \mathcal{B}'$ and fix $\chi \in [\chi_l, \chi_h]^3$. Then,

$$\begin{aligned} & |T\delta^1(\chi) - T\delta^2(\chi)| \\ &= \left| h[\delta^1(\chi)] - h[\delta^2(\chi)] + \hat{q}(\chi) \int \{ \delta^1(\chi') - \delta^2(\chi') + h[\delta^1(\chi')] - h[\delta^2(\chi')] \} dG(\chi') \right| \\ &\leq |h[\delta^1(\chi)] - h[\delta^2(\chi)]| + |\hat{q}(\chi)| \int \{ |\delta^1(\chi') - \delta^2(\chi')| + |h[\delta^1(\chi')] - h[\delta^2(\chi')]| \} dG(\chi') \\ &\leq L_\gamma |\delta^1(\chi) - \delta^2(\chi)| + Q \int \{ |\delta^1(\chi') - \delta^2(\chi')| + L_\gamma |\delta^1(\chi') - \delta^2(\chi')| \} dG(\chi') \\ &= L_\gamma |\delta^1(\chi) - \delta^2(\chi)| + Q(1 + L_\gamma) \int |\delta^1(\chi') - \delta^2(\chi')| dG(\chi') \\ &\leq [L_\gamma + Q(1 + L_\gamma)] \|\delta^1 - \delta^2\|, \end{aligned}$$

where $\|\cdot\|$ denotes the *sup norm*. Taking the *sup* over $[\chi_l, \chi_h]^3$ on the left-hand side of the inequality above, we get $\|T\delta^1 - T\delta^2\| \leq [L_\gamma + Q(1 + L_\gamma)] \|\delta^1 - \delta^2\|$, where $L_\gamma + Q(1 + L_\gamma) \in [0, 1]$. Choosing $\delta^2 = 0$, we have $T : \mathcal{B}' \rightarrow \mathcal{B}'$. Since $L_\gamma + Q(1 + L_\gamma) \in [0, 1]$, T is a contraction with modulus $L_\gamma + Q(1 + L_\gamma)$ on $\mathcal{B}'$. Since $(\mathcal{B}', \|\cdot\|)$ is complete (because the set of bounded functions with the

usual sup norm is complete), it follows from the *Banach fixed-point theorem* that T has a unique fixed point $\delta \in \mathcal{B}'$ (Theorem 3.2 of [Stokey and Lucas, 1989](#), p. 50). $\square$

Lemma 2. Let $\chi = (\chi_0, \chi_1, \chi_2)$. Let $\Sigma : [\chi_l, \chi_h]^3 \rightarrow [0, 1]$, $\lambda_1 : [\chi_l, \chi_h]^3 \rightarrow [0, \frac{\chi_1}{2}]$, and $\lambda_2 : [\chi_l, \chi_h]^3 \rightarrow [0, \frac{\chi_2}{2}]$ be bounded measurable functions for $\chi_l > 0$ arbitrarily close to zero and $\chi_h < \infty$ arbitrarily large. Then, the unique solution $(\chi_0, \chi_1, \chi_2) \in [\chi_l, \chi_h]^3$ to the system of following three equations,

$$\Sigma(\chi) = \frac{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}, \quad (\text{A.3})$$

$$\lambda_1(\chi) = \frac{\chi_1}{2} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2}, \quad (\text{A.4})$$

and

$$\lambda_2(\chi) = \frac{\chi_2}{2} \frac{\chi_0}{\chi_0 + \chi_1 + \chi_2}, \quad (\text{A.5})$$

is

$$\begin{aligned} \chi_0 = [2\Sigma(\chi)]^{-1} & \left\{ 2\Sigma(\chi) \lambda_1(\chi) R(\chi) + (1 - \Sigma(\chi)) (r + 2\lambda_1(\chi) D(\chi)) \right. \\ & \left. + \sqrt{8r\lambda_1(\chi) D(\chi) [1 - \Sigma(\chi)]^2 + [2\Sigma(\chi) \lambda_1(\chi) R(\chi) + (1 - \Sigma(\chi)) (-r + 2\lambda_1(\chi) D(\chi))]^2} \right\}, \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} \chi_1 = [2D(\chi)(1 - \Sigma(\chi))]^{-1} & \left\{ 2\Sigma(\chi) \lambda_1(\chi) R(\chi) + (1 - \Sigma(\chi)) (-r + 2\lambda_1(\chi) D(\chi)) \right. \\ & \left. + \sqrt{8r\lambda_1(\chi) D(\chi) [1 - \Sigma(\chi)]^2 + [2\Sigma(\chi) \lambda_1(\chi) R(\chi) + (1 - \Sigma(\chi)) (-r + 2\lambda_1(\chi) D(\chi))]^2} \right\}, \end{aligned} \quad (\text{A.7})$$

and

$$\chi_2 = \frac{\lambda_2(\chi)}{\lambda_1(\chi)} \chi_1, \quad (\text{A.8})$$

where

$$R(\chi) \equiv 1 + \frac{\lambda_2(\chi)}{\lambda_1(\chi)}$$

and

$$D(\chi) \equiv 1 - \eta_1 + \frac{\lambda_2(\chi)}{\lambda_1(\chi)} (1 - \eta_2).$$

Proof. Equations (A.4) and (A.5) imply (A.8). That is, given inputs, determining χ_1 is sufficient to determine χ_2 . Then, we need to determine χ_0 and χ_1 uniquely to complete the proof. Re-write (A.3) and (A.4):

$$\Sigma(\chi) = \frac{r + \chi_1 D(\chi)}{r + \chi_0 + \chi_1 D(\chi)} \quad (\text{A.9})$$

and

$$\lambda_1(\chi) = \frac{\chi_1}{2} \frac{\chi_0}{\chi_0 + \chi_1 R(\chi)}. \quad (\text{A.10})$$

Rearranging (A.9) implies

$$\chi_0 = \frac{1 - \Sigma(\chi)}{\Sigma(\chi)} (r + \chi_1 D(\chi)). \quad (\text{A.11})$$

By plugging into (A.10), one obtains a quadratic equation in χ_1 , which has two real solutions. Its positive solution is given by (A.7). Then, (A.6) follows from (A.11), which completes the proof. $\square$

Lemma 3. *Assume $\gamma > 1$. Let $\chi = (\chi_0, \chi_1, \chi_2)$ and $L_\gamma = \max_{0 \leq y \leq 2^{1/\gamma}} y^{\gamma-1} (2 - y^\gamma)^{\frac{1}{\gamma}-1}$. Let $\bar{q} : [\chi_l, \chi_h]^3 \rightarrow [0, 4(A_1 + \rho A_2)Q]$, $\lambda_1 : [\chi_l, \chi_h]^3 \rightarrow [0, \frac{\chi_1}{2}]$, and $\lambda_2 : [\chi_l, \chi_h]^3 \rightarrow [0, \frac{\chi_2}{2}]$ be bounded measurable functions for $Q < \frac{1-L_\gamma}{1+L_\gamma}$, $\chi_l > 0$ arbitrarily close to zero, and $\chi_h < \infty$ arbitrarily large. Then, the three-equation system implied by (4.1) and (4.2) has a unique solution $(\chi_0, \chi_1, \chi_2) \in [\chi_l, \chi_h]^3$.*

Proof. The result follows immediately from Lemma 1 and Lemma 2 by letting

$$\hat{q}(\chi) = \frac{\bar{q}(\chi)}{4(A_1 + \rho A_2)}.$$

$\square$

B 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. 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' bargaining 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, particularly 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' bargaining 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 the predecessor of the MiFID II database, differences in reporting requirements preclude a consistent merge with our baseline sample. To obtain empirical moments for re-estimating the model during the COVID-19 period, we proceed as follows. We compute how much each of our empirical moments changed from the 2018-2019 period to the February-April 2020 period. We then use these percentage changes to adjust the moment values from our baseline sample (reported in Table 1 and Table 3). For scalar moments, we simply calculate the percentage difference between the pre-COVID-19 period and during the COVID-19 crisis. For vector-based moments (i.e., the vector of clients' trade intensities and the vector of clients' trade sizes), we run the following cross-sectional regression separately for clients' trade intensities and trade sizes in each market:

$$\text{Covid } m_k^i = \beta_k \times (\text{Pre-covid } m_k^i) + \epsilon_k^i,$$

where m_k^i is client i 's trade intensity (resp. trade size) in market k .

Table 9: Empirical Moments during COVID-19

Variable	Government Bonds		Corporate Bonds	
	Change (1)	Implied Moment (2)	Change (3)	Implied Moment (4)
Dispersion Ratio	-34.6%	0.6140406	+16.1%	0.78425550
Average Trade Cost (bps)	+208.7%	2.3771544	+297.6%	28.1582863
Average Intensity	+22.9%	0.0422750	-11.6%	0.0035871
Intensity Dispersion	+22.9%	0.0652610	-11.6%	0.0064232
Trade Volume	+20.8%	339,912	-16.0%	2,304
Average Trade Size	-37.5%	4,041,558	-19.7%	223,008

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. Dispersion ratio is the ratio between interdealer and dealer-to-client price dispersion. Clients' trade costs are calculated using the methodology of Pintér, Wang, and Zou (2024) and averaged across clients. To calculate trade costs, we use the benchmark price computed as the average price of all transactions at the bond-day level. 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. Trade volume is total daily trading volume in terms of par value per bond per client. Average trade size is the mean (across clients) of clients' mean trade size.

Table 9 reports the corresponding changes in our estimation moments during the COVID-19

crisis. Average trade costs experienced the largest change with 208.7% and 297.6% increases in the government and corporate bond markets, respectively. Clients' trade intensities in the gilt market increased (+22.9%), but clients' trade intensities decreased in corporate bonds (-11.6%). Similarly, trade volume in the gilt market increased while it decreased in the corporate bond market. Clients' trade sizes decreased in both markets, but the decline was more pronounced in government bonds. Finally, the ratio between interdealer and dealer-to-client price dispersion decreased in gilts, while it increased in corporates.

The large increase in clients' trade costs is indicative of substantial worsening of trading frictions during the crisis, with some notable cross-market differences emerging: (i) the relative increase in average trade cost was more than 40% larger 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 adjusted compared to the changes in the government bond market. The combination of these facts and the intuition behind the identification properties of our structural model (see Section 4.4) 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 9, we re-estimate our structural model. Because clients' trade sizes changed during the COVID-19 crisis, we update the calibrated MRS parameter, $\rho = 11.5288$, to match the ratio of the aggregate trade size in the government bond market to the aggregate trade size in the corporate bond market during the COVID-19 period. Our baseline estimation strategy of minimizing (4.7) "fails" in this case. That is, it returns $\eta_2 = 1$ because the model is not able to match extremely high average trade cost in the corporate bond market during the COVID-19 crisis. This mechanically leads to an extremely high welfare loss in the corporate bond market because clients lose all the gains from trade due to dealer sector becoming effective monopolists. To overcome this issue, we add a penalty term, $+0.01/(1 - \eta_2)^2$, to the objective function (4.7).³⁵

The parameter estimates for the turbulent COVID-19 period is reported in Table 10. In addition, Table 11 shows the fit of the model for this period.

A comparison of trading delays in Table 12 reveals that dealers significantly refrained from immediacy provision in both markets. Average trading delay of the median client increased by 400% and 316% 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.

Moreover, dealers were able to exert more market discrimination during the COVID-19 crisis

³⁵In the appendix section H, we repeat our COVID-19 estimation with a significantly smaller penalty term, $+0.0001/(1 - \eta_2)^2$, and find that results change only slightly.

Table 10: Parameter Estimates for the COVID-19 Period

	Normal (1)	Turbulent (2)
γ – Curvature of the utility function	42.296	42.5759
η_1^d – Gilt dealers’ bargaining power	0.0025	0.0482
η_1^c – Gilt core broker-dealers’ barg. power	0.0389	0.0767
η_2^d – Corp. dealers’ bargaining power	0.1243	0.2005
η_2^c – Corp. core broker-dealers’ barg. power	0.2587	0.7290
$\bar{A}$ – Tradable portion of assets	2.6715×10^7	2.4134×10^7

Notes: This table reports the estimates of the parameters. The parameter estimates are obtained by minimizing the objective function (4.7) augmented with an additional penalty term, $+0.01/(1 - \eta_2)^2$. 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 9.

Table 11: Model Fit (COVID-19)

Moments	Government Bonds		Corporate Bonds	
	Empirical	Theoretical	Empirical	Theoretical
	(1)	(2)	(3)	(4)
Trade Intensities				
Median	0.0084	0.0084	0.0006	0.0006
p25	0.0028	0.0028	0.0001	0.0001
p75	0.0272	0.0272	0.0027	0.0027
Trade Sizes				
Median	1,397,528	2,511,444	217,841	217,841
p25	228,139	505,708	43,865	43,865
p75	4,346,370	9,655,933	837,548	837,548
Average Trade Cost	2.3771	2.3966	28.1583	17.7264
Trade Volume	339,912	247,365	–	–

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.

according to the comparison of the bargaining power estimates from Table 12. Clients’ lost surplus share approximately triples in both markets. However, the decomposition of the clients’ lost surplus share reveals that the change was mainly caused by dealers’ exerting more market power over clients in the gilt market, while it was mainly caused by core broker-dealers’ exerting more market power over dealers in the corporate bond market. Empirically, this is evident from the decline in the ratio between interdealer and dealer-to-client price dispersion in gilts during the COVID-19 crisis, while the same ratio increased in corporates.

Taking stock, in terms of the *relative* change of trading delays and clients’ lost surplus shares between the pre-COVID-19 period and during COVID-19, the two markets appear to be roughly equally fragile. This then raises the question: could one conclude that the two markets are equally fragile from a welfare perspective as well? Our results in Table 13 shed some light on this question.

Table 12: 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	0.0026	0.0130	0.0314	0.1306
p25	0.0000	0.0000	0.0000	0.0000
p75	0.5162	0.4849	1.3686	1.9279
Client's Lost Surplus Share	4.15%	12.48%	38.30%	92.96%
Dealer's Bargaining Share	0.25%	4.82%	12.43%	20.05%
Core Broker-Dealer's Barg. Share	3.89%	7.67%	25.87%	72.90%

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 10.

Table 13: Welfare Results II: Estimated Welfare Losses (COVID-19)

	Government Bonds		Corporate Bonds	
	Normal	Turbulent	Normal	Turbulent
	(1)	(2)	(3)	(4)
Welfare Loss	2.3778%	3.6338%	5.0463%	11.3484%
Due to Search Frictions	2.3774%	3.6277%	4.9464%	7.9122%
Due to Intermediation Frictions	0.0004%	0.0061%	0.0999%	3.4362%

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 10.

Table 13 shows that the relative welfare loss increased by 53% in the government bond market, while it more than doubled in the corporate bond market. By looking at the decomposition of the welfare loss, it is evident that this asymmetry is mostly due to intermediation frictions becoming an important contributor to the welfare loss in the corporate bond market, while it is still a marginal contributor in the gilt market. Given the changes in clients' trade frequencies and trade sizes, the model interprets the empirical four-fold increase in the corporate bond trade costs as being "too high" so it can be rationalized only by clients' losing close to entirety of transaction surpluses. This, in turn, implies an almost 3.5% welfare loss from intermediation frictions in the corporate bond market during the COVID-19 crisis, while it was only 0.1% during normal times. Hence, from a relative welfare perspective, the corporate bond market overall appears to be more fragile than the government bond market 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 10 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 14: Estimated Welfare Losses (Counterfactual)

	Government Bonds	Corporate Bonds
	(1)	(2)
Welfare Loss	2.5132 %	5.2389%
Due to Search Frictions	2.5128%	5.1356%
Due to Dealers' Market Power	0.0004%	0.1033%

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 $\bar{A}$ and each client's χ_0 are from the COVID-19 estimates, parameters η_1 and η_2 and each clients' χ_1 and χ_2 are from the baseline estimates for normal times.

Table 14 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 intensified. More specifically, we use clients' preference parameters (i.e., the elasticity of clients' utility function and the distribution of clients' taste shock intensity) and the tradable portion of assets $\bar{A}$ from the COVID-19 period to reflect the change in clients' inherent trading needs during turbulent times, but we keep market friction parameters (i.e., clients' lost surplus shares in each market and the distribution clients' meeting rates with dealers in each market) the same as normal times. Comparing Table 13 and Table 14 implies that the welfare loss would increase from 2.38% to 2.51% in the government bond market and from 5.05% to 5.24% in the corporate bond market, if the OTC market structure were fully resilient to large negative shocks like the COVID-19 shock of 2020. Considering that welfare losses actually rose to 3.63% and 11.35%, 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 and intermediation frictions 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; Cohen, Kargar, Lester, and Weill, 2023).

Supplement to “Comparing Search and Intermediation Frictions Across Markets”

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

Gábor Pintér¹ Semih Üslü² Jean-Charles Wijnandts³

C Multilateral Bargaining

In this appendix, we show how our assumed bargaining solution (3.1)-(3.3) could be obtained as the generalized Nash solution for multilateral bargaining. Mathematically, we define the following cooperative bargaining solution:

$$\begin{aligned} & \left\{ q_k(\varepsilon, a_1, a_2, \chi), p_k(\varepsilon, a_1, a_2, \chi), p_k^d(\varepsilon, a_1, a_2, \chi) \right\} \\ & = \arg \max_{q, p, p^d} [V(\varepsilon, a_k + q, a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - pq]^{1-\eta_k^c - \eta_k^d} [pq - p^d q]^{\eta_k^d} [p^d q - P_k q]^{\eta_k^c}, \quad (\text{C.1}) \end{aligned}$$

subject to

$$V(\varepsilon, a_k + q, a_{-k}, \chi) - V(\varepsilon, a_1, a_2, \chi) - pq \geq 0,$$

$$pq - p^d q \geq 0,$$

and

$$p^d q - P_k q \geq 0,$$

where these three constraints are the individual rationality constraints of the client, the dealer, and the core broker-dealer, respectively. The *multilateral* Nash product (C.1) is defined similarly to Theorem 1 of Lensberg (1988), Theorem 1' of Krishna and Serrano (1996), and Proposition 5 of Suh and Wen (2006) with two crucial differences. First, (C.1) allows for asymmetric bargaining powers, while those papers study the symmetric Nash solution. Second, while these papers study bargaining games with a fixed total surplus (i.e. splitting a “fixed pie”), our bargaining problem has a variable total surplus and the determinant, q , of the size of the total surplus is bargained over as well.

¹Bank for International Settlements, e-mail: gabor.pinter@bis.org

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

³Bank of England, e-mail: jean-charles.wijnandts@bankofengland.co.uk

In anticipation of the function $V(\varepsilon, \cdot, a_{-k}, \chi)$ being continuously differentiable and strictly concave for all ε , a_{-k} , and χ , we set up the Lagrangian of the optimization problem (C.1) and find the first-order necessary and sufficient conditions (see Theorem M.K.2, p. 959, and Theorem M.K.3, p. 961, in [Mas-Colell, Whinston, and Green \(1995\)](#)) for optimality by differentiating the Lagrangian. After simplification, the first-order condition with respect to q implies that the trade size, $q_k(\varepsilon, a_1, a_2, \chi)$, solves

$$V_k(\varepsilon, a_k + q, a_{-k}, \chi) = P_k, \quad (\text{C.2})$$

where $V_k(\varepsilon, \cdot, a_{-k}, \chi)$ refers to the first derivative of the function $V(\varepsilon, \cdot, a_{-k}, \chi)$ given ε , a_{-k} , and χ . The continuous differentiability and strict concavity of $V(\varepsilon, \cdot, a_{-k}, \chi)$ guarantees the existence and uniqueness of the trade quantity $q_k(\varepsilon, a_1, a_2, \chi)$. Notice that the quantity that solves (C.2) is also the maximizer of the total trade surplus (3.1), which also verifies the Pareto optimality of our assumed bargaining solution. Then, the first-order conditions with respect to p and p^d imply that the transaction prices, $p_k(\varepsilon, a_1, a_2, \chi)$ and $p_k^d(\varepsilon, a_1, a_2, \chi)$, are determined such that the total trade surplus is split among the parties to reflect their Nash bargaining powers, which yields (3.2) and (3.3).

D Bid and Ask Prices in Our Special Case

As an intermediate step, note that Equation (3.10) implies

$$\begin{aligned} V(\varepsilon, \theta^*(2\sigma, \chi), \chi) - V(\varepsilon, \theta^*(0, \chi), \chi) &= \frac{\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} P_k q_k(\chi) \\ &+ \frac{1}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \frac{(r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))\varepsilon + \chi_0\sigma}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \Delta_u(\chi), \end{aligned}$$

where

$$\begin{aligned} \Delta_u(\chi) &\equiv \frac{[\theta^*(2\sigma, \chi)]^{1-\gamma}}{1-\gamma} - \frac{[\theta^*(0, \chi)]^{1-\gamma}}{1-\gamma} \\ &= \frac{1}{1-\gamma} \left(\frac{A_1 + \rho A_2}{\frac{1}{2} \int [(1 + \Sigma(\chi'))^{1/\gamma} + (1 - \Sigma(\chi'))^{1/\gamma}] dG(\chi')} \right)^{1-\gamma} \left[(1 + \Sigma(\chi))^{\frac{1}{\gamma}-1} - (1 - \Sigma(\chi))^{\frac{1}{\gamma}-1} \right]. \end{aligned}$$

Thus, the ask and the bid price in the client-dealer market are

$$\alpha_k(\chi) = (1 - \eta_k) P_k + \frac{\eta_k}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left((\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)) P_k + (1 + \Sigma(\chi)) \sigma \frac{\Delta_u(\chi)}{q_k(\chi)} \right)$$

and

$$\beta_k(\chi) = (1 - \eta_k) P_k + \frac{\eta_k}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left((\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)) P_k + (1 - \Sigma(\chi)) \sigma \frac{\Delta_u(\chi)}{q_k(\chi)} \right),$$

respectively, and the ask and the bid price in the inter-dealer market are

$$\mathcal{A}_k(\chi) = (1 - \eta_k^c) P_k + \frac{\eta_k^c}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left((\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)) P_k + (1 + \Sigma(\chi)) \sigma \frac{\Delta_u(\chi)}{q_k(\chi)} \right)$$

and

$$\mathcal{B}_k(\chi) = (1 - \eta_k^c) P_k + \frac{\eta_k^c}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left((\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)) P_k + (1 - \Sigma(\chi)) \sigma \frac{\Delta_u(\chi)}{q_k(\chi)} \right),$$

respectively.

E Proofs

E.1 Proof of Proposition 1

First, we establish in Lemma 4 that the functional equation (3.9) admits a unique real solution, taking as given the frictionless market prices P_1 and P_2 . Our argument is adapted from the existence and uniqueness proofs of the earlier models with unrestricted asset holdings, especially Lagos and Rocheteau (2006), and uses the standard fixed point tools for dynamic programming.

Lemma 4. *Given P_1 and P_2 , the auxiliary HJB equation (3.9) has a unique solution.*

Proof. Rewrite (3.9) as

$$V(\varepsilon, a_1, a_2, \chi) = \frac{\varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V(\varepsilon', a_1, a_2, \chi)}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} + \frac{\sum_{k \in \{1, 2\}} \chi_k(1 - \eta_k) \max_{a'_k \in [-M, M]} \{V(\varepsilon, a'_k, a_{-k}, \chi) - (a'_k - a_k) P_k\}}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}. \quad (\text{E.1})$$

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

$$(OV)(\varepsilon, a_1, a_2, \chi) = \frac{1}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left(\varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V(\varepsilon', a_1, a_2, \chi) + \sum_{k \in \{1, 2\}} \chi_k(1 - \eta_k) \max_{a'_k \in [-M, M]} \{V(\varepsilon, a'_k, a_{-k}, \chi) - (a'_k - a_k) P_k\} \right). \quad (\text{E.2})$$

Then, showing (3.9) has a unique solution is equivalent to showing O has a unique fixed point. Let $\mathcal{T} = \{\varepsilon_l, \varepsilon_h\} \times [-M, M]^2 \times [\chi_l, \chi_h]^3$ and let $C(\mathcal{T}) = \{g : \mathcal{T} \rightarrow \mathbb{R} \mid g(\varepsilon, a_1, a_2, \chi) \text{ is bounded and continuous in } a_1 \text{ and } a_2\}$. Suppose $V \in C(\mathcal{T})$, then the *theorem of the maximum* implies that the maximization on the RHS of (E.2) has a solution continuous in a_1 and a_2 . (Theorem 3.6 of Stokey and Lucas, 1989, p. 62). Then, because $u(\cdot)$ is a continuous function defined on a compact set $[-M, M]$, $O : C(\mathcal{T}) \rightarrow C(\mathcal{T})$. Consider the metric space $(C(\mathcal{T}), \|\cdot\|)$, where $\|\cdot\|$ denotes the *sup norm*. We next show that O is a contraction mapping on $(C(\mathcal{T}), \|\cdot\|)$. To this end, we follow the usual procedure of checking the Blackwell's sufficient conditions for a contraction, i.e., we show that O satisfies *monotonicity* and *discounting* properties. To establish monotonicity, we need to show that $V^A \leq V^B$ implies $OV^A \leq OV^B$. Fix $(\varepsilon, a_1, a_2, \chi)$ and let

$a_k^i \in \arg \max_{a'_k \in [-M, M]} \{V^i(\varepsilon, a'_k, a_{-k}, \chi) - (a'_k - a_k) P_k\}$ for $i \in \{A, B\}$. Then, $V^A \leq V^B$ implies

$$\begin{aligned}
(OV^A)(\varepsilon, a_1, a_2, \chi) &= \frac{1}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left(\varepsilon u(a_1 + \rho a_2) \right. \\
&\quad \left. + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V^A(\varepsilon', a_1, a_2, \chi) + \sum_{k \in \{1, 2\}} \chi_k(1 - \eta_k) \left\{ V^A(\varepsilon, a_k^A, a_{-k}, \chi) - (a_k^A - a_k) P_k \right\} \right) \\
&\leq \frac{1}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left(\varepsilon u(a_1 + \rho a_2) \right. \\
&\quad \left. + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V^B(\varepsilon', a_1, a_2, \chi) + \sum_{k \in \{1, 2\}} \chi_k(1 - \eta_k) \left\{ V^B(\varepsilon, a_k^A, a_{-k}, \chi) - (a_k^A - a_k) P_k \right\} \right) \\
&\leq \frac{1}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \left(\varepsilon u(a_1 + \rho a_2) \right. \\
&\quad \left. + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V^B(\varepsilon', a_1, a_2, \chi) + \sum_{k \in \{1, 2\}} \chi_k(1 - \eta_k) \left\{ V^B(\varepsilon, a_k^B, a_{-k}, \chi) - (a_k^B - a_k) P_k \right\} \right) \\
&= (OV^B)(\varepsilon, a_1, a_2, \chi),
\end{aligned}$$

which establishes monotonicity. To verify discounting, consider $c \geq 0$. Then,

$$[O(V + c)](\varepsilon, a_1, a_2, \chi) \leq (OV)(\varepsilon, a_1, a_2, \chi) + \beta c,$$

where

$$\beta = \frac{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \in (0, 1).$$

To prove that O is a contraction mapping, consider two arbitrarily chosen functions $V^A, V^B \in C(\mathcal{T})$. By the definition of sup norm,

$$V^A \leq V^B + \|V^A - V^B\|.$$

Since O has the monotonicity property,

$$OV^A \leq O(V^B + \|V^A - V^B\|).$$

Using the discounting property,

$$OV^A \leq OV^B + \beta \|V^A - V^B\|.$$

Applying the same procedure in reverse establishes

$$OV^B \leq OV^A + \beta \|V^A - V^B\|.$$

Therefore,

$$\|OV^A - OV^B\| \leq \beta \|V^A - V^B\|,$$

which implies that O is a contraction mapping, with modulus β , on the complete metric space $(C(\mathcal{T}), \|\cdot\|)$. Hence, it follows from the *Banach fixed-point theorem* that O has a unique fixed point $V \in C(\mathcal{T})$ (Theorem 3.2 of [Stokey and Lucas, 1989](#), p. 50). $\square$

We next assume $\rho P_1 = P_2$ and follow a guess-and-verify approach to determine the *unique* value function V . We conjecture that

$$V(\varepsilon, a_1, a_2, \chi) = C(\varepsilon, \chi) + D(\varepsilon, \chi) u(a_1 + \rho a_2) + E(\chi) a_1 + H(\chi) a_2,$$

where $C(\varepsilon, \chi)$, $D(\varepsilon, \chi)$, $E(\chi)$, and $F(\chi)$ are the coefficients to be determined. Because $\kappa_0(\varepsilon, \chi)$ 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 $D(\varepsilon, \chi)$, $E(\chi)$, and $F(\chi)$.

One can apply the envelope theorem to the auxiliary HJB equation (3.9) to obtain:

$$\begin{aligned} [r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)] V_k(\varepsilon, a_1, a_2, \chi) &= \varepsilon u'(a_1 + \rho a_2) \theta_k \\ &+ \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} V_k(\varepsilon', a_1, a_2, \chi) + \chi_k(1 - \eta_k) P_k + \chi_{-k}(1 - \eta_{-k}) V_k(\varepsilon, a_k, a'_{-k}, \chi), \end{aligned}$$

where a'_{-k} solves $V_{-k}(\varepsilon, a_k, a'_{-k}, \chi) = P_{-k}$; $V_k(\cdot, \cdot, \cdot, \cdot)$ refers to the derivative with respect to a_k ; and $\theta_1 = 1$ and $\theta_2 = \rho$. By using our conjectured value function and matching coefficients, we obtain

$$\begin{aligned} H(\chi) &= \frac{\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \rho P_1, \\ E(\chi) &= \frac{\chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} P_1, \end{aligned}$$

and

$$D(\varepsilon, \chi) = \frac{1}{r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)} \frac{(r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)) \varepsilon + \chi_0 \bar{\varepsilon}}{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)},$$

which complete the proof.

E.2 Proof of Proposition 2

Let $f_\chi(\varepsilon, \theta) = \lim_{h \rightarrow 0} \int_{-M}^M \int_{\theta - \rho a_2 - h}^{\theta - \rho a_2} \int_{\varepsilon - h}^{\varepsilon} \Phi_\chi(d\varepsilon, da_1, da_2)$ for $\varepsilon \in \{\varepsilon_l, \varepsilon_h\}$. Using the optimal trading rule (3.12), the stationarity condition (3.5) implies

$$\chi_0 \frac{1}{2} \sum_{\tilde{\varepsilon} \in \{\varepsilon_l, \varepsilon_h\}} f_\chi(\tilde{\varepsilon}, \tilde{\varepsilon}) - (\chi_0 + \chi_1 + \chi_2) f_\chi(\varepsilon, \tilde{\varepsilon}) = 0 \quad (\text{E.3})$$

for $\varepsilon \neq \tilde{\varepsilon}$ and

$$\chi_0 \frac{1}{2} \sum_{\tilde{\varepsilon} \in \{\varepsilon_l, \varepsilon_h\}} f_\chi(\tilde{\varepsilon}, \varepsilon) + (\chi_1 + \chi_2) \sum_{\tilde{\varepsilon} \in \{\varepsilon_l, \varepsilon_h\}} f_\chi(\varepsilon, \tilde{\varepsilon}) - (\chi_0 + \chi_1 + \chi_2) f_\chi(\varepsilon, \varepsilon) = 0, \quad (\text{E.4})$$

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 $\tilde{\varepsilon}$. One can use (E.3) and (E.4) to derive $f_\chi(\varepsilon_l, \varepsilon_l)$, $f_\chi(\varepsilon_l, \varepsilon_h)$, $f_\chi(\varepsilon_h, \varepsilon_l)$, and $f_\chi(\varepsilon_h, \varepsilon_h)$. Then, by re-writing, one obtains (3.13).

E.3 Proof of Proposition 3

Lemma 5. *In any general equilibrium, $\rho P_1 = P_2$.*

Proof. Applying the envelope theorem to the auxiliary HJB equation (3.9) implies:

$$\begin{aligned} \left[r + \frac{\chi_0}{2} + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2) \right] V_k(\varepsilon, a_1, a_2, \chi) &= \varepsilon u'(a_1 + \rho a_2) \theta_k \\ &+ \frac{\chi_0}{2} V_k(\varepsilon_-, a_1, a_2, \chi) + \chi_k(1 - \eta_k) P_k + \chi_{-k}(1 - \eta_{-k}) V_k(\varepsilon, a_k, a'_{-k}, \chi), \end{aligned}$$

where a'_{-k} solves $V_{-k}(\varepsilon, a_k, a'_{-k}, \chi) = P_{-k}$; $V_k(\cdot, \cdot, \cdot, \cdot)$ refers to the derivative with respect to a_k ; and $\theta_1 = 1$ and $\theta_2 = \rho$. Also letting a'_k solve $V_k(\varepsilon, a'_k, a_{-k}, \chi) = P_k$,

$$\begin{aligned} \left[r + \frac{\chi_0}{2} + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2) \right] V_k(\varepsilon, a_1, a_2, \chi) &= \varepsilon u'(a_1 + \rho a_2) \theta_k \\ &+ \frac{\chi_0}{2} V_k(\varepsilon_-, a_1, a_2, \chi) + \chi_k(1 - \eta_k) V_k(\varepsilon, a'_k, a_{-k}, \chi) + \chi_{-k}(1 - \eta_{-k}) V_k(\varepsilon, a_k, a'_{-k}, \chi). \end{aligned}$$

Letting $k = 1$ and $k_- = 2$, one obtains

$$\begin{aligned} \left[r + \frac{\chi_0}{2} + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2) \right] V_1(\varepsilon, a_1, a_2, \chi) &= \varepsilon u'(a_1 + \rho a_2) \\ &+ \frac{\chi_0}{2} V_1(\varepsilon_-, a_1, a_2, \chi) + \chi_1(1 - \eta_1) V_1(\varepsilon, a'_1, a_2, \chi) + \chi_2(1 - \eta_2) V_1(\varepsilon, a_1, a'_2, \chi) \quad (\text{E.5}) \end{aligned}$$

and

$$\begin{aligned} \left[r + \frac{\chi_0}{2} + \chi_1 (1 - \eta_1) + \chi_2 (1 - \eta_2) \right] V_2(\varepsilon, a_1, a_2, \chi) &= \varepsilon u'(a_1 + \rho a_2) \rho \\ &+ \frac{\chi_0}{2} V_2(\varepsilon_-, a_1, a_2, \chi) + \chi_1 (1 - \eta_1) V_2(\varepsilon, a'_1, a_2, \chi) + \chi_2 (1 - \eta_2) V_2(\varepsilon, a_1, a'_2, \chi). \end{aligned} \quad (\text{E.6})$$

Multiplying the former with ρ and subtracting it from the latter, one obtains

$$\begin{aligned} &\rho \left[\left(r + \frac{\chi_0}{2} + \chi_1 (1 - \eta_1) + \chi_2 (1 - \eta_2) \right) V_1(\varepsilon, a_1, a_2, \chi) - \frac{\chi_0}{2} V_1(\varepsilon_-, a_1, a_2, \chi) \right. \\ &\quad \left. - \chi_1 (1 - \eta_1) V_1(\varepsilon, a'_1, a_2, \chi) - \chi_2 (1 - \eta_2) V_1(\varepsilon, a_1, a'_2, \chi) \right] \\ &= (r + \chi_0 + \chi_1 + \chi_2) V_2(\varepsilon, a_1, a_2, \chi) - \frac{\chi_0}{2} V_2(\varepsilon_-, a_1, a_2, \chi) \\ &\quad - \chi_1 (1 - \eta_1) V_2(\varepsilon, a'_1, a_2, \chi) - \chi_2 (1 - \eta_2) V_2(\varepsilon, a_1, a'_2, \chi). \end{aligned}$$

The structure of the equation indicates symmetry between V_1 and V_2 with identical weights. For the equality to hold for arbitrary $(\varepsilon, a_1, a_2, \chi)$, the terms must align identically. This symmetry implies that:

$$V_1(\varepsilon, a_1, a_2, \chi) \rho = V_2(\varepsilon, a_1, a_2, \chi). \quad (\text{E.7})$$

Evaluating (E.5) at a_1 such that $V_1(\varepsilon, a_1, a_2, \chi) = P_1$,

$$\begin{aligned} \left[r + \frac{\chi_0}{2} + \chi_1 (1 - \eta_1) + \chi_2 (1 - \eta_2) \right] P_1 &= \varepsilon u'(a_1 + \rho a_2) \\ &+ \frac{\chi_0}{2} V_1(\varepsilon_-, a_1, a_2, \chi) + \chi_1 (1 - \eta_1) P_1 + \chi_2 (1 - \eta_2) \frac{P_2}{\rho}, \end{aligned}$$

which holds for all a_2 . Similarly, evaluating (E.6) at a_2 such that $V_2(\varepsilon, a_1, a_2, \chi) = P_2$,

$$\begin{aligned} \left[r + \frac{\chi_0}{2} + \chi_1 (1 - \eta_1) + \chi_2 (1 - \eta_2) \right] P_2 &= \varepsilon u'(a_1 + \rho a_2) \rho \\ &+ \frac{\chi_0}{2} V_2(\varepsilon_-, a_1, a_2, \chi) + \chi_1 (1 - \eta_1) P_1 \rho + \chi_2 (1 - \eta_2) P_2, \end{aligned}$$

which holds for all a_1 . Then, evaluating the last two equations jointly at a_1 and a_2 such that $V_1(\varepsilon, a_1, a_2, \chi) = P_1$ and $V_2(\varepsilon, a_1, a_2, \chi) = P_2$, (E.7) implies $P_1 \rho = P_2$, which proves the Lemma. $\square$

Lemma 5 justifies the closed-form formula (3.10) and the resulting optimal composite holding

(3.11). Then, the frictionless markets clear if and only if $\theta^d(P_1) = A_1 + \rho A_2$, where

$$\theta^d(P_1) \equiv \int \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} (u')^{-1} \left[\frac{r + \chi_0 + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2)}{(r + \chi_1(1 - \eta_1) + \chi_2(1 - \eta_2))\varepsilon + \chi_0\bar{\varepsilon}} r P_1 \right] dG(\chi).$$

An application of the inverse function theorem implies that θ^d is strictly decreasing because u is strictly concave. Hence, the frictionless price P_1 such that $\theta^d(P_1) = A_1 + \rho A_2$ is unique whenever it exists. From the standard Inada limit conditions on u stated in the proposition, $\theta^d(0) = \infty$ and $\theta^d(\infty) = 0$ and hence the existence of $P_1 \in (0, \infty)$ is guaranteed by the intermediate value theorem. Finally, P_2 is pinned down by Lemma 5: $P_2 = \rho P_1$.

E.4 Proof of Proposition 4

The formula for $\mathbb{W}^{Eq}$ follows easily by substituting (3.13), (3.17) and $u(\theta) = \frac{\theta^{1-\gamma}}{1-\gamma}$ into (3.28).

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:

$$\begin{aligned} & \mathbb{W}^{FB}(\varepsilon_l) \\ = & \max_{a_1(0), a_1(2\sigma), a_2(0), a_2(2\sigma)} \frac{\varepsilon_l [a_1(0) + \rho a_2(0)]^{1-\gamma}}{r} \frac{1}{1-\gamma} \frac{1}{2} + \frac{2\sigma [a_1(2\sigma) + \rho a_2(2\sigma)]^{1-\gamma}}{r} \frac{1}{1-\gamma} \frac{1}{2} - \frac{\sigma (A_1 + \rho A_2)^{1-\gamma}}{r} \frac{1}{1-\gamma}, \end{aligned} \quad (\text{E.8})$$

subject to

$$\begin{aligned} & \frac{1}{2} (a_k(0) + a_k(2\sigma)) = A_k, \\ & -M \leq a_k(0) \leq M, \end{aligned}$$

and

$$-M \leq a_k(2\sigma) \leq M$$

for all $k \in \{1, 2\}$. 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 $\mathbb{W}^{FB} \equiv \lim_{\varepsilon_l \rightarrow 0} \mathbb{W}^{FB}(\varepsilon_l)$.

Because M is an arbitrarily large number that can be chosen to allow for interior solutions, we are fine as long as $a_k(0), a_k(2\sigma) < \infty$. Because the second term on the RHS of (E.8) does not

depend on the control variables, the Lagrangian is

$$\begin{aligned}\mathcal{L}^{FB} = & \frac{\varepsilon_l}{r} \frac{[a_1(0) + \rho a_2(0)]^{1-\gamma}}{1-\gamma} \frac{1}{2} + \frac{2\sigma}{r} \frac{[a_1(2\sigma) + \rho a_2(2\sigma)]^{1-\gamma}}{1-\gamma} \frac{1}{2} \\ & + \mu_1 \left(A_1 - \frac{1}{2} (a_1(0) + a_1(2\sigma)) \right) + \mu_2 \left(A_2 - \frac{1}{2} (a_2(0) + a_2(2\sigma)) \right).\end{aligned}$$

The FOCs for interior solution are

$$\frac{\varepsilon_l}{2r} [a_1(0) + \rho a_2(0)]^{-\gamma} - \frac{\mu_1}{2} = 0,$$

$$\frac{\varepsilon_l}{2r} [a_1(0) + \rho a_2(0)]^{-\gamma} \rho - \frac{\mu_2}{2} = 0,$$

$$\frac{\sigma}{r} [a_1(2\sigma) + \rho a_2(2\sigma)]^{-\gamma} - \frac{\mu_1}{2} = 0,$$

and

$$\frac{\sigma}{r} [a_1(2\sigma) + \rho a_2(2\sigma)]^{-\gamma} \rho - \frac{\mu_2}{2} = 0.$$

Then,

$$a_1(0) + \rho a_2(0) = \left(\frac{\varepsilon_l}{\mu_1 r} \right)^{1/\gamma},$$

$$a_1(2\sigma) + \rho a_2(2\sigma) = \left(\frac{2\sigma}{\mu_1 r} \right)^{1/\gamma},$$

and

$$\mu_2 = \rho \mu_1.$$

Then, using the resource constraints, the Lagrange multiplier μ_1 is

$$\mu_1 = \frac{1}{r} \left(\frac{\varepsilon_l^{1/\gamma} + (2\sigma)^{1/\gamma}}{2(A_1 + \rho A_2)} \right)^\gamma,$$

which in turn implies

$$a(0) = 2(A_1 + \rho A_2) \frac{\varepsilon_l^{1/\gamma}}{\varepsilon_l^{1/\gamma} + (2\sigma)^{1/\gamma}}$$

and

$$a(2\sigma) = 2(A_1 + \rho A_2) \frac{(2\sigma)^{1/\gamma}}{\varepsilon_l^{1/\gamma} + (2\sigma)^{1/\gamma}}.$$

Substituting into (E.8),

$$\mathbb{W}^{FB}(\varepsilon_l) = \frac{\varepsilon_l}{2r} \frac{2^{1-\gamma} (A_1 + \rho A_2)^{1-\gamma}}{1-\gamma} \left[\frac{\varepsilon_l^{1/\gamma}}{\varepsilon_l^{1/\gamma} + (2\sigma)^{1/\gamma}} \right]^{1-\gamma} + \frac{2^{1-\gamma} (A_1 + \rho A_2)^{1-\gamma}}{1-\gamma} \frac{\sigma}{r} \left[\frac{(2\sigma)^{1/\gamma}}{\varepsilon_l^{1/\gamma} + (2\sigma)^{1/\gamma}} \right]^{1-\gamma} - \frac{\sigma (A_1 + \rho A_2)^{1-\gamma}}{r (1-\gamma)}.$$

By taking the limit as $\varepsilon_l \rightarrow 0$, 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 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 (3.29) does not depend on the control variables, the planner's current-value Hamiltonian can be written as

$$\begin{aligned} \mathcal{L}^{SB}(q_1, q_2 | \Phi) &= \int \int \int \sum_{-M-M}^M \varepsilon u(a_1 + \rho a_2) \Phi_\chi(\varepsilon, da_1, da_2) dG(\chi) \\ &\quad + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \int \int \int \sum_{-M-M}^M \frac{1}{2} (\vartheta(\varepsilon', a_1, a_2, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi)) \Phi_\chi(\varepsilon, da_1, da_2) dG(\chi) \\ &\quad + \sum_{k \in \{1, 2\}} \int \int \int \sum_{-M-M}^M \chi_k \{ \vartheta(\varepsilon, a_k + q_k(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi) \} \Phi_\chi(\varepsilon, da_1, da_2) dG(\chi) \\ &\quad - \sum_{k \in \{1, 2\}} \zeta_k \int \int \int \sum_{-M-M}^M \chi_k q_k(\varepsilon, a_1, a_2, \chi) \Phi_\chi(\varepsilon, da_1, da_2) dG(\chi), \end{aligned}$$

$\vartheta(\varepsilon, a_1, a_2, \chi)$ denotes the current-value co-state variable associated with $\Phi_\chi(\varepsilon, da_1, da_2)$; and ζ_1 and ζ_2 are the Lagrange multipliers associated with the conditions (3.30) and (3.31), respectively.

The FOC for optimization is

$$\vartheta_k(\varepsilon, a_k + q_k(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) = \zeta_k, \quad (\text{E.9})$$

if $\chi_k > 0$, $\Phi_\chi(\varepsilon, da_1, da_2) > 0$, and $dG(\chi) > 0$.

In any optimum (q_1^e, q_2^e) , the co-state variables must satisfy the ODEs,

$$\nabla_{n(\varepsilon, a_1, a_2, \chi)} \mathcal{L}^{SB}(q_1^e, q_2^e | \Phi) = r\vartheta(\varepsilon, a_1, a_2, \chi) - \dot{\vartheta}(\varepsilon, a_1, a_2, \chi), \quad (\text{E.10})$$

where $n(\varepsilon, a_1, a_2, \chi)$ is the degenerate measure which puts all the probability on the type $(\varepsilon, a_1, a_2, \chi)$ and ∇_n denotes the Gâteaux differential in the direction of measure n :

$$\nabla_n \mathcal{L}^{SB}(q_1^e, q_2^e | \Phi) = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{L}^{SB}(q_1^e, q_2^e | \Phi + \epsilon n) - \mathcal{L}^{SB}(q_1^e, q_2^e | \Phi)}{\epsilon}.$$

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

$$\begin{aligned} \mathcal{L}^{SB}(q_1^e, q_2^e | \Phi + \epsilon n) - \mathcal{L}^{SB}(q_1^e, q_2^e | \Phi) &= \epsilon \int \int \int \sum_{-M-M}^M \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \varepsilon u(a_1 + \rho a_2) n(\varepsilon, da_1, da_2, d\chi) \\ &\quad + \epsilon \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \int \int \int \sum_{-M-M}^M \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} (\vartheta(\varepsilon', a_1, a_2, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi)) n(\varepsilon, da_1, da_2, d\chi) \\ &\quad + \epsilon \sum_{k \in \{1, 2\}} \int \int \int \sum_{-M-M}^M \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_k \{ \vartheta(\varepsilon, a_k + q_k^e(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi) \} n(\varepsilon, da_1, da_2, d\chi) \\ &\quad - \epsilon \sum_{k \in \{1, 2\}} \zeta_k \int \int \int \sum_{-M-M}^M \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_k q_k(\varepsilon, a_1, a_2, \chi) n(\varepsilon, da_1, da_2, d\chi). \end{aligned}$$

Thus,

$$\begin{aligned} \nabla_{n(\varepsilon, a_1, a_2, \chi)} \mathcal{L}^{SB}(q_1^e, q_2^e | \Phi) &= \varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} (\vartheta(\varepsilon', a_1, a_2, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi)) \\ &\quad + \sum_{k \in \{1, 2\}} \chi_k \{ \vartheta(\varepsilon, a_k + q_k^e(\varepsilon, a_1, a_2, \chi), a_{-k}, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi) - \zeta_k q_k^e(\varepsilon, a_1, a_2, \chi) \}. \end{aligned}$$

Using (3.30), (3.31), (E.10), and the FOC (E.9), the following ODE for the co-state variables obtains in any optimum:

$$\begin{aligned} r\vartheta(\varepsilon, a_1, a_2, \chi) - \dot{\vartheta}(\varepsilon, a_1, a_2, \chi) &= \varepsilon u(a_1 + \rho a_2) + \chi_0 \sum_{\varepsilon' \in \{\varepsilon_l, \varepsilon_h\}} \frac{1}{2} (\vartheta(\varepsilon', a_1, a_2, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi)) \\ &\quad + \sum_{k \in \{1, 2\}} \chi_k \max_{q_k \in [-M, M]} \{ \vartheta(\varepsilon, a_k + q_k, a_{-k}, \chi) - \vartheta(\varepsilon, a_1, a_2, \chi) - \zeta_k q_k \} \end{aligned}$$

s.t.

$$\int_{-M}^M \int_{-M}^M \int \sum_{\varepsilon \in \{\varepsilon_l, \varepsilon_h\}} \chi_k q_k(\varepsilon, a_1, a_2, \chi) \Phi_\chi(\varepsilon, da_1, da_2) dG(\chi) = 0.$$

Checking that the planner's optimality conditions do not coincide with the equilibrium conditions is easy. More specifically, the comparison with (3.9) reveals that the planner's optimality conditions and the equilibrium conditions would be identical if $\eta_k = 0$ and $P_k = \zeta_k$ 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 $\mathbb{W}^{Eq}$.

F 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. Since ZEN is not publicly available, it has been used only sparingly 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, trade reports with missing client identifiers, trades of less than £1,000 in par value, and remove trades with erroneous price entries. Risk characteristics are an important dimension that can make the representative corporate bond distinct from government bonds. To mitigate this issue, we exclude high-yield corporate bonds and keep the 57 bonds that have the highest number of transactions.⁴ We also want to avoid biasing the empirical moments by incorporating days with low trading activity such as trading days around the end-of-year holidays. We adopt a data-driven approach to identify these low trading activity days. For inter-dealer transactions, we keep only days where the number of active dealers is larger than half of the maximum number of daily active dealers observed on a trading day during our sample.⁵ We follow the same approach for client-dealer transactions, but using

⁴Information 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.

⁵For the government bond market, the maximum number of dealers trading in the inter-dealer market is 26,

the number of active clients instead.⁶ The final sample comprises 1,440 trading days satisfying these four criteria and we end up with 2,958,554 and 229,180 transactions in the government and corporate bond markets, respectively. The breakdown between the inter-dealer and client-dealer segments is 63% and 37% for the government bond market, and 35% and 65% for the corporate bond market. We identify 526 clients active in both markets over the sample period, and whose trades cover the majority of total client trading volume.

We use these two samples of government and corporate bonds (each including trades for 57 assets) to compute the empirical moments of the main variables that will be used in the analysis. First, we compute price dispersion in the inter-dealer and client-dealer segments of each market. We keep trading hours with at least two transactions occurring in that time window according to the time stamp of the trade report. For the inter-dealer segment, this leaves 1,749,693 transactions in the government bond market and 37,393 transactions in the corporate bond market. The corresponding numbers for the client-dealer segment are respectively 953,257 and 59,353 transactions. Then, we obtain daily measures of price dispersion by first computing (normalized) absolute deviations from the hourly average transaction price, and then averaging within the day these deviations using the size of the trade as weights.

Second, we compute a series of measures from the client-dealer segment of each market, including average trading costs, average client intensity and intensity dispersion, average daily trading volume, and average trade size. For average trading costs, we follow the approach outlined in [Pintér, Wang, and Zou \(2024\)](#) and use the average price of all transactions at the bond-day level as the benchmark price against which transaction costs are calculated. To compute the intensity-based measures, days in which a client does not trade enter as zeros. Average client intensity is calculated as the mean of the total number of transactions for each of the 526 clients on each trading day. Intensity dispersion is then computed as the mean of the absolute deviation of each client's total number of transactions from average intensity. Both intensity measures are scaled by the number of assets (57) in each sample. Total daily trading volume is computed in terms of par value, scaled by the number of clients and the number of assets in each market. Finally, average trade size is calculated by first computing the daily mean of the nominal size of each client's trades, and then averaging across clients.

while for the corporate bond market it is 21.

⁶The maximum number of clients trading in the client-dealer market on any given day is 190 for the government bond market and 132 for the corporate bond market. Given the higher incidence of low trading activity days in the dealer-client segment of the corporate bond market, we set the threshold to 25% of the maximum number of active clients in this case.

G Robustness to Alternative IES Values

Table 15: Counterfactual Welfare Losses ($\gamma = 20$)

	Government Bonds	Corporate Bonds
	(1)	(2)
Welfare Loss	4.3596%	7.7865%
Due to Search Frictions	4.3590%	7.6573%
Extensive Margin	4.3590%	7.6573%
Intensive Margin	0.0000%	0.0000%
Due to Intermediation Frictions	0.0006%	0.1291%

Notes: This table reports the counterfactual welfare losses in the government bond and corporate bond markets implied by the estimated parameter values as reported in Table 2, apart from the counterfactual utility curvature $\gamma = 20$. The top panel reports the relative welfare loss levels in each market. The bottom panel reports various decompositions.

Table 16: Counterfactual Welfare Losses ($\gamma = 2$)

	Government Bonds	Corporate Bonds
	(1)	(2)
Welfare Loss	32.1538%	38.7879%
Due to Search Frictions	32.1520%	38.4990%
Extensive Margin	31.1741%	36.9335%
Intensive Margin	0.9779%	1.5655%
Due to Intermediation Frictions	0.0018%	0.2889%

Notes: This table reports the counterfactual welfare losses in the government bond and corporate bond markets implied by the estimated parameter values as reported in Table 2, apart from the counterfactual utility curvature $\gamma = 2$. The top panel reports the relative welfare loss levels in each market. The bottom panel reports various decompositions.

H Robustness of the COVID-19 Estimation Results

Table 17: Parameter Estimates for the COVID-19 Period

	Normal	Turbulent
	(1)	(2)
γ – Curvature of the utility function	42.296	42.5752
η_1^d – Gilt dealers' bargaining power	0.0025	0.0475
η_1^c – Gilt core broker-dealers' barg. power	0.0389	0.0756
η_2^d – Corp. dealers' bargaining power	0.1243	0.2125
η_2^c – Corp. core broker-dealers' barg. power	0.2587	0.7723
$\bar{A}$ – Tradable portion of assets	2.6715×10^7	2.4134×10^7

Notes: This table reports the estimates of the parameters. The parameter estimates are obtained by minimizing the objective function (4.7) augmented with an additional penalty term, $+0.0001/(1 - \eta_2)^2$. 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 9.

Table 18: Model Fit (COVID-19)

Moments	Government Bonds		Corporate Bonds	
	Empirical	Theoretical	Empirical	Theoretical
	(1)	(2)	(3)	(4)
Trade Intensities				
Median	0.0084	0.0084	0.0006	0.0006
p25	0.0028	0.0028	0.0001	0.0001
p75	0.0272	0.0272	0.0027	0.0027
Trade Sizes				
Median	1,397,528	2,511,444	217,841	217,841
p25	228,139	505,708	43,865	43,865
p75	4,346,370	9,655,933	837,548	837,548
Average Trade Cost	2.3771	2.3982	28.1583	19.0470
Trade Volume	339,912	247,365	–	–

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 19: 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	0.0026	0.0130	0.0314	0.1228
p25	0.0000	0.0000	0.0000	0.0000
p75	0.5162	0.4672	1.3686	1.8523
Client's Lost Surplus Share	4.15%	12.31%	38.30%	98.48%
Dealer's Bargaining Share	0.25%	4.75%	12.43%	21.25%
Core Broker-Dealer's Barg. Share	3.89%	7.56%	25.87%	77.23%

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 17.

Table 20: Welfare Results II: Estimated Welfare Losses (COVID-19)

	Government Bonds		Corporate Bonds	
	Normal	Turbulent	Normal	Turbulent
	(1)	(2)	(3)	(4)
Welfare Loss	2.3778%	3.6080%	5.0463%	15.1916%
Due to Search Frictions	2.3774%	3.6021%	4.9464%	7.8237%
Due to Intermediation Frictions	0.0004%	0.0059%	0.0999%	7.3669%

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 17.