Clients' Connections

Measuring the Role of Private Information in Decentralised Markets*

Péter Kondor
London School of Economics, CEPR

Gábor Pintér
Bank of England

15th June 2021

Abstract

We propose a new measure of private information in decentralised markets – connections – which exploits the time-variation in the number of dealers with whom a client trades in a time period. Using trade-level data for the UK government bond market, we show that clients perform better when having more connections as their trades predict future price movements. Time-variation in market-wide connections also helps explain yield dynamics. Given our novel measure, we present two applications suggesting that (i) dealers pass on information, acquired from their informed clients, to their affiliates, and (ii) informed clients better predict the orderflow intermediated by their dealers.

JEL Classification*:* G12, G14, G24

Keywords: Private Information, Client-Dealer Connection, OTC, Government Bonds, Financial Networks

*First version: May 2018. Previous version was titled "Private Information and Client Connection in Government Bond Markets". We thank Christoph Aymanns (discussant), Andy Blake, Andrea Buraschi, Robert Czech, Marco Di Maggio, Wouter Den Haan, Amir Kermani (discussant), Dong Lou, Albert Menkveld, Daniel Paravisini, Paolo Pasquariello, Loriana Pelizzon (discussant), Dmitrii Pugachev, Angelo Ranaldo, Adam Reed (discussant), Lucio Sarno, Norman Schürhoff (discussant), Eric Stafford, Marti Subrahmanyam, Davide Tomio (discussant), Laura Veldkamp, Gertjan Vlieghe and participants at the LSE seminar, 3rd SAFE Market Microstructure Conference, the Conference on Non-Bank financial institutions (BoE), 2019 EFA Conference, 2019 SED Conference, the 2020 AFA Conference and the 7th Sovereign Bond Market Conference for helpful comments. Special thanks to Lena Boneva for helping us understand the ZEN data, to Alex Parsons for sharing the macroeconomic news data, and to Marcin Kacperczyk for fruitful conversations. The views expressed are those of the authors, and not necessarily those of the Bank of England or its committees. Kondor acknowledges financial support from the European Research Council (Starting Grant \#336585). Email: p.kondor@lse.ac.uk and gabor.pinter@bankofengland.co.uk.

1 Introduction

A main role of financial markets is to aggregate private information held by economic agents. Trading activity and subsequent adjustments in asset prices release this information to the wider public, thereby making markets more efficient and increasing the welfare of society. The main challenge facing any scientific study of this mechanism is that neither private information nor the identities of its owners are readily observable.

Our paper proposes a proxy for private information. We combine a detailed dataset of the UK government bond market, covering the identities and transactions of trading parties, with insights from the microstructure literature. The idea is that, just like in centralised markets where informed traders may split their trades over time to slow down information revelation and avoid market-impact (Kyle, 1985), informed traders in decentralised markets may submit orders to different dealers at the same time, thereby splitting their trades in the cross-section. This implies that one should observe a trader obtaining private information to trade with more dealers than usual. Accordingly, our proposed proxy for private information in decentralised markets is the time-variation in the number of dealers that clients trade with, which we will refer to as clients' *connections*.[1]

Our empirical analysis yields two sets of results. First, we confirm that connections serve as a proxy for private information by showing that (i) clients make more profitable trades when having more connections and (ii) time-variation of total client connections in the market helps explain daily innovations in yields. Second, we present two of the many possible applications of our proxy: (i) we find suggestive evidence that dealers learn from their informed clients and pass this information to their affiliates, and (ii) we also show that while the private information proxied by connections includes fundamental information around key macro-events, it also contains information on future orderflows. In particular, more connected clients better predict the orderflow intermediated by the dealers they trade with.

We start with the idea that trading with more dealers may be advantageous because it helps the client hide her private information. This, however, requires the client to reach out for quotes from dealers she does not regularly trade with, which is costly. Therefore, the client will do so only when the benefit of hiding information is sufficiently large, that

[1]Our study focuses on the UK government bond market, because (i) being one of the most liquid decentralised markets, it provides a particularly hard test to measure private information, (ii) our dataset provides a detailed, almost universal coverage of all transactions on this market, and (iii) the government bond market plays a crucial role in the economy as the yield curve serves as a benchmark in many financial transactions, it affects government financing costs and plays an important role for the implementation of monetary policy.

is, when her information is sufficiently precise. In subsequent periods the client should overperform.[2] We expect that when a client is connected to more dealers, her trades are more profitable even after controlling for the volume and the number of her transactions in the given period. This effect should not be driven by favourable transaction prices, but by forecasting future price movements. That is, the price of bonds that connected clients buy (sell) should increase (decrease) in subsequent days. We also expect our baseline result to be driven by more sophisticated clients (e.g. hedge funds and asset managers) who are more likely to trade for speculative reasons compared to less sophisticated clients (e.g. pension funds, foreign central banks etc.). We find empirical evidence for each of these predictions. Including client fixed effects, we identify these results primarily from the within-variation of a given client's activity.

We also consider aggregate implications for yield dynamics. We construct a market-wide measure of private information – the total number of client-dealer connections in the system in a time period. We then measure the response of yields to changes in aggregate connections, and find a significant effect even after controlling for trading volume and the total number of clients in the market.

Given our proxy for private information, we offer two of the many possible applications. As a first application, we provide suggestive evidence that dealers pass on information, acquired from their informed clients, to their affiliates. To show this, we use a novel source of variation in our data: for each dealer, we are able to distinguish between trading accounts that perform a market-making function from trading accounts that correspond to other, client-like arms of the given dealer bank, i.e. the given dealer's *affiliates*. We then test whether dealers' affiliates perform better when the given dealer trades with a larger proportion of high-connection clients. We find that this is indeed the case, suggesting that these affiliates obtain the information that their dealers learn from informed clients.

As a second application, we study the nature of private information in government bond markets. Our main focus is to assess whether the private information captured by the time-variation in clients' connectedness is on fundamentals, or on future order-flow. We find some evidence for both. To illustrate the viability that clients trade with more dealers when obtaining private information on key macroeconomic events, we first analyse trading around the Brexit referendum. Given the large uncertainty before the vote, market participants were motivated to either reduce their exposure radically, or to

[2] To demonstrate that our narrative can work in a standard rational framework, we extend the Glosten and Milgrom (1985) model in the Online Appendix 2.

generate private information and bet on the outcome. In line with our hypothesis, we show that a change in their number of dealer connections helps identify the client group with private information. In particular, the group of clients who were connected with more dealers on the day before the referendum persistently increased the duration of their positions for days before the referendum and, subsequently, outperformed other clients when the yield curve dropped immediately as the outcome of the poll became public. We also show evidence that our main findings are more pronounced around macroeconomic announcements. This reinforces that fundamental information plays some role in our mechanism.

At the same time we find strong evidence that more connected clients can better predict the maturity structure of other clients' orderflow, especially the part of the orderflow received by their own dealers in subsequent days. For instance, when a more connected client's orders are concentrated on the short-end of the yield curve in a given day, her dealer is more likely to receive a disproportionate share of orders for short bonds in the following five days. We also show that trading in line with the maturity structure of clients' future orders can be profitable because of the resulting pressure on prices.

Related Literature While our paper is the first to propose clients' connections as a measure of private information in decentralized markets, our study is related to several streams of the literature.

There is a vast literature on measuring private information in financial markets. A large group of these papers focus on security-based measures (e.g. Easley, Kiefer, O'Hara, and Paperman, 1996; Chakravarty, Gulen, and Mayhew, 2004; Duarte and Young, 2009; Roll, Schwartz, and Subrahmanyam, 2010; Johnson and So, 2018). These papers identify securities for which a large share of transactions are likely to be motivated by private information in a given period, typically using the aggregate volume characteristics of those securities, and study the implied return patterns. Instead, our measure allows to study informed transactions of any given client. As our applications show, this feature changes the range of relevant questions we can address with our approach.

A more related group of papers identify informed transactions focusing on the activity of a specific group of clients such as large shareholder activists or corporate insiders (Cohen, Malloy, and Pomorski, 2012; Collin-Dufresne and Fos, 2015) often during specific episodes (Boulatov, Hendershott, and Livdan, 2013; Hendershott, Livdan, and Schurhoff, 2015). By design, these studies are mostly focusing on the cross-sectional heterogeneity in information, building on ex-ante assumptions of which clients should be more informed and in which periods private information should be concentrated.[3] Instead, we use time-series heterogeneity to identify client specific periods of informed trading. That is, our measure can systematically identify periods of informed trading for any given client, even if these periods are uncorrelated across clients.

Our first application studies potential information leakages across clients, their dealers, and the affiliates of these dealers.[4] While there are many empirical works studying the trading process in decentralised markets (e.g. Gabrieli and Georg, 2014; Hollifield, Neklyudov, and Spatt, 2017; Brancaccio, Li, and Schurhoff, 2017) most of these do not focus on the role of private information.[5] Instead, the most related work to this application is Maggio, Franzoni, Kermani, and Sommavilla (2019). Just as we do in this application, they use the network of transactions across market participants to study the flow of private information among them. Apart from the context – they focus on brokers and their clients in stock markets – their proxy of informed trades and the suggested mechanism are also different from our approach. They identify a client's informed transactions as those which are executed by a more connected broker. The argument is that central brokers gather information by executing informed trades, which is then leaked to their best clients through these transactions. Instead, we identify informed transactions as those which are executed when the client is more connected. Our argument is that the client chooses to be more connected when her information is more precise in order to hide it.

Our second application is related to the literature on price discovery in government bond markets (Fleming and Remolona, 1999; Balduzzi, Elton, and Green, 2001; Green, 2004; Brandt and Kavajecz, 2004; Pasquariello and Vega, 2007; Hortacsu and Kastl, 2012; Valseth, 2013). This literature emphasises the informational role of clients' and/or dealers' orderflow. We add to this literature by highlighting the empirical link between variation in connections and orderflow predictability. We are able to do so due to the important feature of our dataset: for each trade we can observe the identity of both parties. This allows us to map out the dynamics of connections of government market

[3]Another approach is to study the effect of transactions which ex-post turns out to be private information driven. For instance (Meulbroek, 1992; Kacperczyk and Pagnotta, 2019) investigates the effect of transactions that subsequently became subject to SEC investigations of insider trading activities.

[4]There is a related, growing theoretical literature on the role of private information in decentralized markets such as Duffie, Malamud, and Manso (2009), Golosov, Lorenzoni, and Tsyvinski (2014), Babus and Kondor (2018), Brancaccio, Li, and Schurhoff (2017) amongst others.

[5]A notable exception is Hagstromer and Menkveld (2019) which uses short-term comovement across quotes of different dealers to map information percolation, and Collin-Dufresne, Hoffmann, and Vogel (2020) which finds that dealers' markups partially prices in future permanent price impact.

participants and explore their links with the price discovery process.

The remainder of the paper is as follows: Section 2 introduces the environment, concepts and hypotheses. Section 3 describes the data sources and provides summary statistics; Section 4 presents the empirical results on using connections as proxy for private information; Section 5 contrasts our findings with alternative explanations; Section 6 presents the two applications of our measure; Section 7 presents robustness checks; Section 8 concludes.

2 Context and Main Hypotheses

We start this section with a basic description of the micro-structure of the UK gilt market. Then, we discuss our main hypotheses.

2.1 Primary Dealers in the UK Gilt Market

The key actors in the UK gilt market are the primary dealers, also known as gilt-edged market markers (GEMMs). In our sample period between 2011 and 2017, their number fluctuates around 20. From now on, we refer to this group as dealers. The UK Debt Management Office (DMO) tenders new issues of government securities to dealers. Clients (e.g. as asset managers, commercial banks, foreign central banks etc.) buy and sell government securities mostly through bilateral transactions to this group.[6] Primary dealers are committed to make, on demand, continuous and effective two-way prices to their clients by regulation. They must also maintain a minimum market share (DMO, 2011).[7]

When a client trades in the UK gilt market, she can observe quotes of all dealers on electronic trading platforms. However, these observed quotes are merely indicative and only small trades can be executed at these prices. If the client wishes to trade a larger quantity, she directly contacts the dealers typically via the phone. Unlike other, centralised exchanges (e.g. the UK gilt futures market) that are increasingly automated, the gilt cash market, which our study focuses on, continues to retain its traditional OTC characteristics where reputation and trading relationships matter largely for dealers (to continue to attract orderflow and thereby trading profitably) as well as for clients (to receive favourable price quotes).

[6] In our sample, only about 1% of client trades are directly between clients.

[7] See Benos and Zikes (2018) for further details about the institutional arrangements of the UK gilt market.

In our sample, we observe that clients tend to trade with a relatively small and persistent subset of all the dealers. Based on interviews with traders, we understand that clients perceive that requesting quotes from dealers they do not regularly trade with as costly. The source of this cost might be diverse. Especially for larger trades, it might take time to approach multiple dealers for quotes. Also, as dealers' quotes reveal information about their inventory, if it is not reciprocated with trades, the dealer might decide to give less tight quotes to that particular client next time. Perhaps most importantly, clients build complex relationships with the investment banks acting as dealers which goes beyond this particular market. Deviating from their regular dealers might harm this relationship.

2.2 Our Mechanism and Main Implications

Our main conjecture is that the time-variation in clients' connections can be a proxy for the time-varation in the precision of their private information. The underlying mechanism is that, just like in centralised markets where informed traders may split their trades over time to slow down information revelation and reduce market-impact (Kyle, 1985), informed traders in decentralised markets may submit orders to various dealers, thereby spreading out their trades in the cross-section.[8]

However, spreading out trades requires the client to reach out for quotes from those dealers, which might be costly. The client will do so only when the advantage of hiding her private information is large, that is, when its information is sufficiently precise. This implies that a trader obtaining more precise private information should trade with more dealers than usual. This mechanism provides a number of testable implications.

First, consider the time-variation of the performance of a given client. Under our intuition, we should observe that when clients are connected to more dealers, they overperform. However, overperformance could come from multiple sources. For instance, even if connections were not related to information, clients requesting more quotes would expose their dealers to more competition, possibly resulting in more favourable transac-

[8]Since Kyle (1985), the microstructure literature has extensively studied how private information can be concealed by splitting informed orders in smaller amounts over time to avoid market impact (e.g. Garleanu and Pedersen, 2013; Mascio, Lines, and Naik, 2017; Back, Collin-Dufresne, Fos, Li, and Ljungqvist, 2018) Spread out across dealers might be a particularly prominent way to hide private information in markets where positions with multiple legs is the norm (Duarte, Longstaff, and Yu, 2006). In these markets, a trader's individual trades reveal her strategic position less than her full portfolio would. This is the case in treasury markets where speculation on the changing shape of the yield curve often involves the simultaneous buying and selling of multiple bonds of different maturity. This is in line with the the summary statistics (Table 1) on the number of bonds clients traded in our sample.

tion prices. Instead, if connection proxies private information, we should expect that the price of government bonds, purchased by the client in these periods, should increase in subsequent days. That is, a more connected client's overperformance should come from the correlation of the direction of their transactions and future price movements.

Also, our interpretation would be further supported if these effects were stronger among those clients who are usually considered to trade on information, such as hedge funds and asset managers compared to insurance companies, pension funds, commercial banks and government organisations. We will refer to the former (latter) group as more (less) sophisticated traders.

We summarize these predictions in the following hypotheses.[9]

Hypothesis 1 *Periods with more connections for a given client should be associated with higher trading profit.*

Hypothesis 2 *More connections for a client who buys (sells) on a given date should be associated with a larger subsequent price increase (decrease).*

Hypothesis 3 *These effects should be stronger for more sophisticated traders.*

Note that our mechanism does not imply causality between connections and performance in any direction. Instead, both higher performance and higher connectedness are caused by more private information.

Second, consider implications for aggregate connections and price formation. In the absence of news, innovations of bond prices should be driven by private information. Also, under our conjecture, average connection in a given time period is a measure of the amount of private information present in the market. Therefore, we should expect a comovement between this measure and innovations in yields. This gives our last hypothesis.

Hypothesis 4 *Periods with higher aggregate connections should be associated with larger absolute innovations in yields.*

[9]To show that our mechanism is consistent with standard arguments of trading under asymmetric information, we rationalize each of these predictions in a simple extension of the Glosten and Milgrom (1985) model in Online Appendix A.1. The mechanism is as follows. When adverse selection is the main determinant of trading cost, the cost of trading is determined by the information content of that unit. If informed traders spread out their orders across multiple dealers, they can reduce the informed share of orderflow per dealer. This reduces adverse selection for each dealer, which reduces trading cost per unit. This gain, which clients can extract through more favourable bid-ask spreads, balances out the cost of increasing connections.

At this point we do not take a stance on the nature of private information connection is a proxy for. It might be fundamental information on future macroeconomic news or information on the price impact of future orderflow or a mixture of both. All the above hypotheses hold in any of these cases. We will investigate this question further in Section 6.2.

Note also that clients might have other reasons to split their trades across dealers than hiding their private information. Most importantly, large trades might have smaller price impact if split into smaller packages through multiple dealers. In Section 5, we investigate the potential of this alternative mechanism to explain our results.

Next, we introduce our data and present evidence that supports Hypotheses 1-4.

3 Measurement and Summary Statistics

In this section we describe the data and construct the two main variable of interest: clients' connections and performance.

3.1 Data Source

To analyse how the dynamics of client-dealer connections are related to clients' trading performance and information, one needs a detailed transaction-level dataset which contains information on the identity of both sides of a trade. The proprietary ZEN database maintained by the UK Financial Conduct Authority (FCA), fittingly provides this information together with information on the transaction date and time; the execution price and quantity; the International Securities Identification Number (ISIN); the account number, the buyer-seller flag. The ZEN database contains trade reports for all secondary-market transactions, where at least one of the counterparties is an FCA-regulated entity. We focus exclusively on conventional gilts. Given that all dealers in our sample are FCA-regulated, we have at least one report for each dealer-client transaction, thereby giving us virtually full coverage of the client trade universe. Our sample covers the period between October 2011 and June 2017. We match our transaction-level data with information on bond duration and end-of-day closing prices obtained from Datastream.

A key aspect of our empirical analysis is to exploit the time-variation in client-dealer connections which requires the matching of each transaction with a client and a dealer identifier. The names of clients and dealers are recorded as unstructured strings of text in the ZEN database. Moreover, a typical client or dealer tends to have multiple accounts with different variants of the firm name across accounts and also within the same account. We use a textual algorithm that searches through the unstructured strings of names and accounts, and assigns a unique identifier to each transaction. When constructing identifiers, we aim at the highest possible level of consolidation by treating parent companies, affiliates and different arms as one client or dealer.[10] After discarding duplicate trades, we end up with 480 identified clients and about 1.2 million trades transacted by them and their dealers. The trading activity between these clients and dealers covers around 80% of all client activity (in terms of trading volume) in the UK gilt market.

3.2 Client-Dealer Connections

Our baseline measure of connections is the number of dealers a given client is connected to in a given time period. A client is connected to a dealer if she trades with the dealer at least once.[11] Since client connectivity is a key variable in our analysis, we provide some descriptive statistics to describe it.

Table 1 presents summary statistics based on our baseline regression sample that is aggregated to the client-day level. We find that the average client on a given day is connected to three dealers and carries out about 10 transactions with them. There is substantial sample variation: the average difference in connections between the 90th and 10th percentile is 6. To illustrate the extent to which the variation in client connectivity is a cross-sectional phenomenon, we compute the averages of our measures at the client-level, and plot the resulting distribution in a histogram (left panel of Figure 1). We find that the distribution of the connection measure is positively skewed, with the mass of clients having low values and a few clients exhibiting large values.

Clients that are on average more connected can differ from less connected clients along other time-invariant characteristics such as size, business model etc. To control for this, we purge out client fixed effects from our connectivity measures and plot the resulting distribution in a histogram (right panel of Figure 1). We find substantial within-client variation: the average difference in connections between the 90th and 10th percentile is

[10] In Section 6, we take a closer look at the heterogeneity of dealer accounts. For each dealer we distinguish between the market making accounts of a given dealer (the ones which are mainly used for transactions with clients) from the affiliate accounts (the ones which are mainly used for transactions with the market making accounts of the same dealer). We identify signs of information flow from informed clients to dealers' affiliates via their market making activity.

[11] In the previous version of our paper (Kondor and Pinter, 2019), we also use eigenvector centrality (Bonacich and Lloyd, 2001; Maggio, Kermani, and Song, 2017) as an alternative measure of connectivity, yielding very similar results.

Table 1: Summary Statistics – Client-Day Level

(a) All Clients

	(1) Mean	(2) Median	(3) p10	(4) p90	(5) sd	(6) N
First Order Connection	3.28	3.00	1.00	6.00	2.33	100,696
Transaction Number	10.14	6.00	3.00	20.00	12.92	100,696
Volume (£millions)	73.31	12.36	0.11	214.66	158.69	100,696
Number of Bonds Traded	6.45	4.00	2.00	13.00	6.27	100,696

(b) High Connection vs Low Connection Days

	High Connection Days		Low Connection Days	
	Mean	Median	Mean	Median
2-day Performance	0.54	0.27	-0.22	0.09
4-day Performance	0.69	0.31	-0.27	0.18
First Order Connection	5.04	4.00	2.38	2.00
Transaction Number	14.70	10.00	7.80	5.00
Volume (£millions)	114.29	27.80	52.30	8.04
Number of Bonds Traded	8.55	6.00	5.37	4.00

Notes: This table reports summary statistics for our baseline sample, covering 2011m10-2017m6, that is collapsed at the client-day level. Panel A reports summary statistics for all clients. Panel B reports summary statistics on volume-weighted performance measures at the 2-day and 4-day horizons, measured in basis points. Panel B differentiates between high connection and low connection trading days, by placing each client observation into two groups based on the within-variation of connections, i.e. depending on whether the client's first-order connection on a given day is below or above the client's own median connection measure based on the whole sample.

3.45, which is high compared to the corresponding value using cross-client variation (2.42). Similarly, the standard deviation of first-order connections is around 1.12 in the cross-section and as high as 1.47 when using only the within-client variation. This substantial *within-variation* in connections is a key feature of the data, which our empirical analysis will primarily rely on.

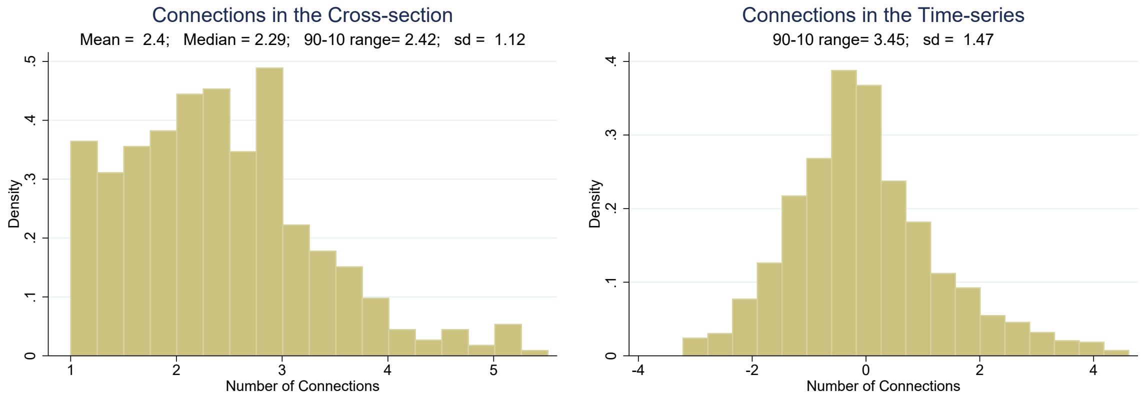
3.3 Trading Performance

3.3.1 Baseline Measure

To measure trading performance, we follow Maggio, Franzoni, Kermani, and Sommavilla (2019) and compute the T-day-horizon return on each trade of client i on day t, measured as the percentage difference between the transaction price and the closing price T days after the transaction date.[12] Formally, for each trade j, we construct the measure

[12]The T-day horizon starts at the start of each day and ends after T days. We use overlapping time windows. For example, to compute one-day performance measures ($T = 1$), we compare all trades on

Figure 1: Cross-Client and Within-Client Variation in Connections



Notes: these figures summarize the cross-client and within-client variation in our first-order connections, defined as the number of dealers with whom a client trades on a given trading day. The left panel plots the distribution of average client connections, after collapsing the data at the client-level. To construct the right panel, we first collapse the data at the client-day level, then run a panel regression to purge out client and day fixed effects ($ClientConnections_{i,t} = \alpha_i + \mu_t + \varepsilon_{i,t}$), and plot the distribution of the residuals ($\varepsilon_{i,t}$).

$Performance_j^T$ as follows:

$$Performance_j^T = \left[\ln\left(P^T\right) - \ln\left(P_j^\star\right)\right] \times \mathbf{1}_{B,S}, \tag{3.1}$$

where $P_j^\star$ is the transaction price, P^T is the T-day ahead closing price of the corresponding gilt, and $\mathbf{1}_{B,S}$ is an indicator function equal to 1 when the transaction is a buy trade, and -1 when it is a sell trade. All transactions-specific cumulative (T-day ahead) returns are then averaged for each client i and day t using the pound value of the trades as weights. (In our baseline sample we include client-day observations that are based on more than two transactions, in order to achieve a less noisy daily performance measurement.) This client-day level measure, denoted by $Performance_{i,t}^T$, will be our left-hand-size variable in our baseline regressions detailed below. As robustness, we also present the results using unweighted daily average returns.

Table 1 summary statistics of the 2-day and 4-day performance measures. Panel B shows that the average client performs better on days with more dealer connections compared to days when the same client has fewer connections. Note that trading volume and numbers of transactions are also higher in high connection days, which motivates the use of these variables as controls in our regressions.

day 1 to the closing price on day 2, and compare all trades on day 2 to the close price on day 3, and so on.

3.3.2 Decomposing Trading Performance

In this part, we propose a decomposition method which extends our baseline performance measurement. The T-day performance of a client on a trade can be high because the given client faces lower price impact compared to other clients trading at the same time. We refer to this as the transaction component of performance. Alternatively, trading performance can be high because the given client can better anticipate future prices changes. We refer to this as the anticipation component of performance. Building on 3.1, we compute the decomposition for each transaction j as follows:

$$\ln\left(P^T\right) - \ln\left(P_j^\star\right) = \underbrace{\left[\ln\left(P^T\right) - \ln\left(\overline{P}\right)\right]}_{Anticipation} + \underbrace{\left[\ln\left(\overline{P}\right) - \ln\left(P_j^\star\right)\right]}_{Transaction}, \tag{3.2}$$

where $\overline{P}$ is the only new term which denotes the average transaction price (based on all available dealer-client trades in the corresponding gilt) measured around the time of transaction j.[13] To estimate $\overline{P}$, we split each trading day into three time-windows, and compute the average transaction price $\overline{P}$ based on all relevant transactions in each window. The intra-day time windows are <11am, 11am-15pm and >15pm, which are set to have an approximately even number of transactions across the windows. Given the trade-level decomposition, we then collapse our dataset at the client-day level using both volume-weighted and unweighted daily average returns.

Note that most of the recent empirical work on financial networks (Afonso, Kovner, and Schoar, 2014; Hendershott, Li, Livdan, and Norman, 2017; Hollifield, Neklyudov, and Spatt, 2017; Maggio, Kermani, and Song, 2017) mainly focused on the transaction component. Distinguishing between the transaction component and the anticipation component allows us to test whether more connections increase performance because clients can achieve more favourable deals (at lower mark-ups) or because clients have private information about future price changes.

[13] When computing the average transaction price $\overline{P}$ for the calculation of the transaction component for client i, we exclude the trades of client i from the construction of $\overline{P}$. In effect, we compute a different average transaction price for each client, $\overline{P}_{i\notin}$, in order to avoid any mechanical effect (say, of the given client's size or trading activity) on the computation of the transaction component.

4 Client Connections as Proxy for Private Information

This section presents our main empirical results, supporting the key message of our paper: time-variation in client-dealer connections can be used to proxy time-variation in private information. First, we present supporting evidence for Hypotheses 1–3. Using panel-data regressions, we show that clients' trading performance systematically increases when the given client trades with more dealers, and that this effect is stronger among more sophisticated clients. Second, we study innovations in yields and Hypothesis 4. We provide evidence that variation in total client-dealer connections in the market comove with the day-to-day innovations in the level and slope of the yield curve. Finally, we argue that our quantitative estimates are economically significant.

4.1 Client Profitability

In this part, we connect the time-variation in clients' connections with the time-variation in their performance along the lines of Hypotheses 1–3.

4.1.1 Baseline Results

To estimate whether a client's trading performance increases when the client increases its connections with the primary dealer sector, we run the following daily panel regression:

$$Performance^T_{i,t} = \beta \times ClientConnections_{i,t} + X_{i,t} + \alpha_{i,year} + \mu_t + \varepsilon_{i,t}, \tag{4.1}$$

where $Performance^T_{i,t}$ is the trading performance (3.1) of client i on day t at horizon T; $ClientConnections_{i,t}$ is the number of dealers the given client is connected to on day t; $\alpha_{i,year}$ and μ_t are client-year and day fixed effects, respectively[14]; $X_{i,t}$ includes controls such as the number of transactions and trading volume. These controls are important for checking that our connections variable is not simply picking up the effect of increased trade size (Easley and O'Hara, 1987; Merrick, Naik, and Yadav, 2005; Maggio, Franzoni, Kermani, and Sommavilla, 2019). Throughout the analysis, in computing standard errors

[14]The client-year fixed effect, $\alpha_{i,year}$, aims control for any low-frequency changes in client's trading activity that would not be captured by a client fixed effect, α_i. For example, successfully growing (and surviving) traders may experience over time a gradual expansion of their trading network as well as an increase in trading performance.

we take a conservative approach, and employ two-way clustering at the client and time level. This allows for arbitrary correlation of the residuals across time and across clients.

The main coefficient of interest in 4.1 is β which captures the relation between client connections and trading performance. Panel a of Table 2 reports our baseline results for value-weighted trading performance. Each column corresponds to a different trading horizon going from $T = 1$ to $T = 5$. We find a positive relationship between client connections and trading performance, whose magnitude and statistical significance increases in the horizon. At the 5-day horizon, an additional connection is worth about 0.5bp.

Panel b of Table 2 decomposes the baseline into the effects going through the transaction component and the anticipation component (3.2). By doing so, we learn whether more connected clients may perform better because they get better deals compared to other clients trading around the same time (transaction component) or because they can better anticipate future price changes over the coming trading days (anticipation component). Our mechanism does not have strong predictions on the earlier, but requires the latter effect to be present.

Table 2: Client Connections and Trading Performance: Baseline Results

	(1) 1-day	(2) 2-day	(3) 3-day	(4) 4-day	(5) 5-day
Client	0.190**	0.268**	0.406***	0.535***	0.569***
Connections	(2.12)	(2.46)	(2.91)	(3.04)	(2.94)
Volume	0.206*	0.281*	0.306*	0.269	0.282
	(1.65)	(1.93)	(1.71)	(1.38)	(1.28)
Tran.	-0.801***	-1.222***	-1.620***	-1.552***	-1.965***
	(-2.95)	(-3.19)	(-3.86)	(-3.36)	(-3.98)
N	100414	100414	100414	100414	100414
R^2	0.057	0.056	0.057	0.057	0.058
Day FE	Yes	Yes	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes	Yes	Yes

(a) Trading Performance over 1-5 Days

	(1) Baseline	(2) Transaction	(3) Anticipation
Client	0.535***	0.098**	0.430**
Connections	(3.04)	(2.43)	(2.39)
Volume	0.269	-0.075	0.338*
	(1.38)	(-1.21)	(1.75)
Tran.	-1.552***	-0.240*	-1.313***
	(-3.36)	(-1.84)	(-2.82)
N	100414	100348	100348
R^2	0.057	0.095	0.055
Day FE	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes

(b) Decomposing 4-day Performance: Transaction vs Anticipation Effect

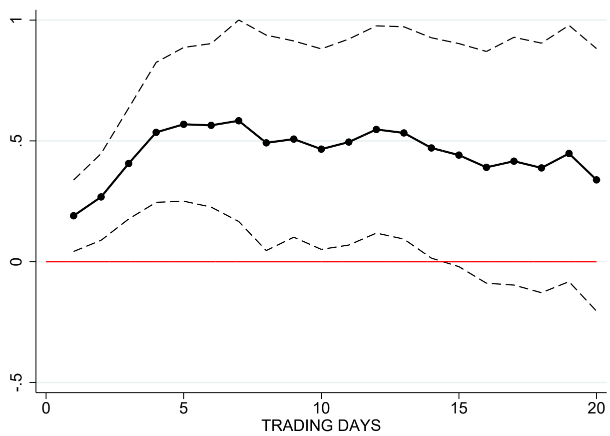
Notes: panel A regresses the value-weighted trading performance at different time horizons on client connections (4.1). The transaction-level data is collapsed at the client-day level. The performance measures are in basis points. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Tran."). Panel B decomposes the 4-day performance effect into a transaction component and an anticipation component (3.2). The results are based on the average transaction price $\overline{P}$ that uses the trades (for the given gilt) in a 3-hour window within the transaction time (excluding the given client's trades from the computation of $\overline{P}$ – see footnote 13). T-statistics in parentheses are based on robust standard errors, using two-way clustering at the day and the client level. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

Our results show that a client, when more connected, tends to perform significantly better in each component. When more connected, she tends to trade at a more favourable price and to the direction of future price movements. Quantitatively, we find that the anticipation component has a much larger role in the overall higher performance of clients when they are more connected. In particular, almost 80% of our baseline effect

(0.53bp) is explained by the anticipation component (0.43bp) as opposed to the transaction component (0.1bp).

Moreover, we assess the persistence of the effect of connections and gradually increase the trading horizon up to 20 days ($T = 20$) while re-estimating our baseline regression 4.1. In Figure 2, we present the 20 estimated βs together with the 90% confidence bands. We find that the effect peaks between the 4-day and 7-day horizon coefficients. While our point estimates remain positive for weeks, the effect gradually loses significance.

Figure 2: Connections and Performance over 1-20 Day Horizons

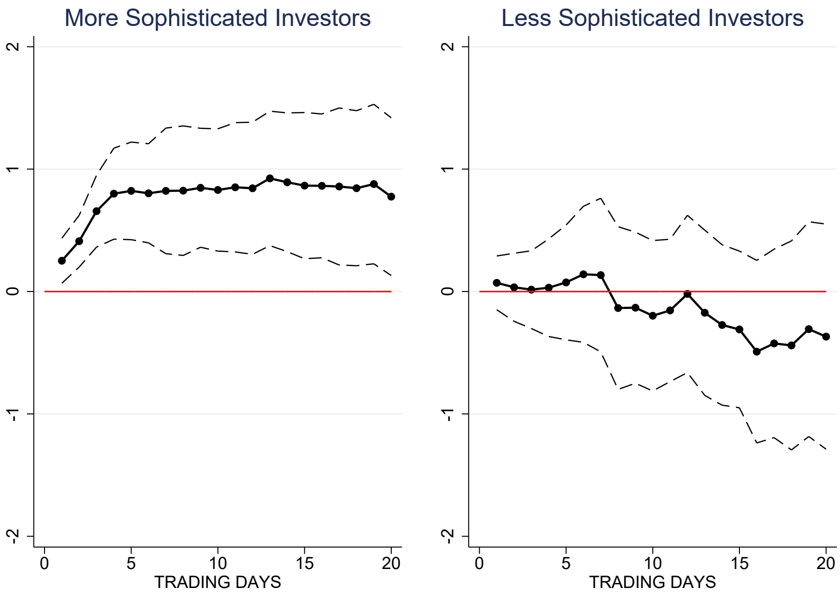


Notes: this figure plots the estimated β coefficients from our baseline regression 4.1 up to 20-day horizon ($T = 20$), using the value weighted performance variable as the regress and, measured in basis points. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Transactions"). The shaded area denotes the 90% confidence band, It is based on robust standard errors, using two-way clustering at the client and day level.

4.1.2 More Sophisticated vs Less Sophisticated Investors

To reinforce the information-based interpretation of our baseline estimates, we next combine the within-variation of client connections with cross-sectional heterogeneity of client types. It is reasonable to conjecture that not all type of clients engage in speculative trading on treasury markets. Pension funds, foreign central banks, other government organisations, insurance companies and commercial banks are less focused on future price movements of treasuries compared to hedge funds and asset managers. In our formal language, if the former group neither seek nor receive private information, then variation in their connections introduces noise in our findings. To assess this possibility, based on the name of the client account in our data, we assign clients to a more sophisticated and a less sophisticated group. We end up having 250 and 230 clients in each group, respectively.[15]

Figure 3: Connections and Performance over 1-20 Day Horizons: More vs. Less Sophisticated Clients



Notes: this figure plots the estimated β coefficients from our baseline regression 4.1 up to 20-day horizon ($T = 20$), for more sophisticated (left panel) and less sophisticated (right panel) clients separately, using the value weighted performance variable as the regress and, measured in basis points. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Transactions"). The shaded area denotes the 90% confidence band, It is based on robust standard errors, using two-way clustering at the client and day level.

Given our classification, we re-estimate our baseline (Figure 2) for both types of clients separately, and plot the estimates on Figure 3 up to the 20-day horizon. We find substantial heterogeneity across the client types with the majority of our baseline effects being driven by the more sophisticated clients. Note also, that our effect is more persistent in the sample of more sophisticated clients.[16]

[15] The guiding principle in our classification was (i) to focus on the main business profile of a given client and (ii) to aim at the highest possible level of consolidation, when determining whether a given client can be regarded as a hedge fund / asset manager (more sophisticated) or other type (less sophisticated). In most cases, this was straightforward. For example, we have 46 government entities (mainly foreign central banks), that can be immediately placed in the group of less sophisticated clients, and around 40 hedge funds that belong to the more sophisticated group. In contrast, there could be some ambiguity in the classification of some asset managers. For example, certain less sophisticated clients (e.g. insurance companies, commercial banks and pension funds) have asset manager branches. In line with our strategy, we regarded these asset manager accounts as part of the parent company with a less sophisticated type. The full list of the names of more sophisticated and less sophisticated clients, used in our paper, can be found in the online replication material.

[16] Table 11 of the Online Appendix shows the performance decomposition into transaction and anticip-

The fact that most of our baseline effect is driven by the increased ability of more sophisticated investors to predict future price movements is important for the information-based interpretation of our reduced-form evidence. For example, this result makes alternative interpretations related to demand pressures of large uninformed traders less likely. Our group of less sophisticated traders features 46 government entities (mainly foreign central banks) who trade infrequently, but when they do, they often trade large quantities. Trading large quantities – big enough to move prices – may require increased number of dealer connections. Therefore, Figure 3 is indicative that when connections increase to reduce the individual order size, it does not translate to over-performance in subsequent days. We return to the discussion of this and other alternative explanations in Section 5.

4.1.3 The Economic Significance

We argue that our baseline results are also economically significant. For example, using the estimate (0.5bp) in Column 4-5 of Table 2, if a client trades with one additional dealer on a day, then the expected increase in her short-term returns is more than twice her average performance (we are using the fact that the median 4-day returns are 0.22bp in our sample). Table 3 further illustrates the economic significance of the performance-connection relationship, by comparing days when clients have few connections to days when they have more connections. Single-sorting by the within-variation in connections, we find that the difference in mean (median) performances is about 1bp (0.25bp), consistent with our baseline regression results (Table 2).

Table 3: Illustrating the Economic Significance of Connections

	Average Daily Volume (in £000s)		Average Daily 4-day Performance		Decomposition of Gross Performance	
	Mean	Median	Mean	Median	Mean	Median
Low Connection Days	50,000	6,900	-0.381	0.136	-33%	9%
High Connection Days	109,000	24,900	0.704	0.393	133%	91%
					100%	100%

Note: This table illustrates the economic significance of the performance-connectedness relationship. The sample (at the client-day level) into two groups using single-sorting, based on the within-variation of connections, i.e. the first (second) group contains the observations for those days when the given client had fewer (more) daily connections compared to its sample average. The 4-day performance measures are in basis points.

Moreover, clients trade more when they are more connected: the median trading

ation components (as panel (b) of Table 2 above) for both client groups. More sophisticated investors' overall benefit from an additional connection concentrates in the anticipation component, with both components being statistically significant. There is some statistical evidence that less sophisticated investors also enjoy more favourable transaction prices when having an additional connection.

volume is about £6.9million (£24.9million) on days when the client has fewer (more) dealer connections than its sample average. The performance difference coupled with the difference in trading volume in high and low connectivity days implies that the majority of positive trading performance is concentrated in high connectivity days.

4.2 Aggregate Connections

Having presented evidence on the positive relationship of a client's connections and her individual performance, we now turn to the information content in aggregate connections. First, we consider the aggregate implication of Hypothesis 1. If many highly connected clients buy (sell) a given bond a given day, that should signal that the price of the given bond will increase (decrease) during the following days or weeks. Second, we show that yield innovations tend to cluster in days when aggregate connection is high, validating Hypothesis 4.

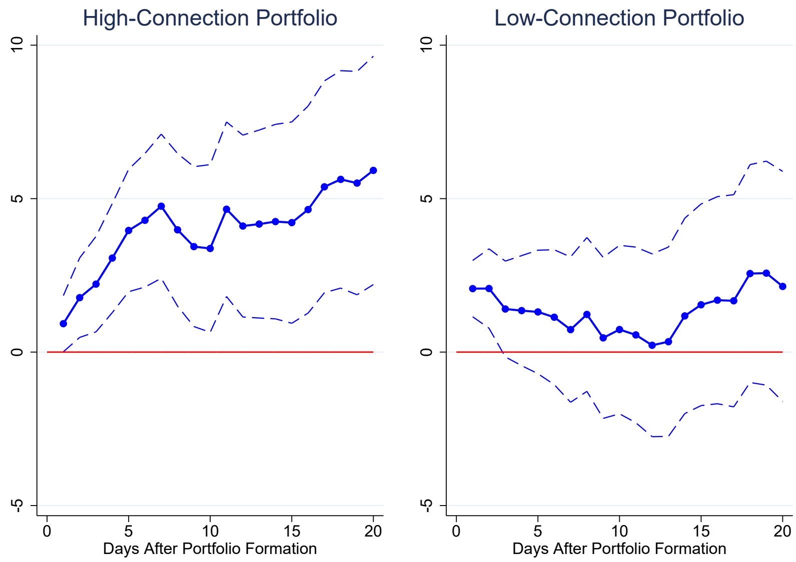
4.2.1 Aggregate Connections and Portfolio Returns

An alternative way to test Hypothesis 1 is by portfolio analysis. In this part, we argue that the aggregate portfolio choice of connected clients is a signal for future bond returns. To show this, we build on the portfolio analysis of Maggio, Franzoni, Kermani, and Sommavilla (2019) and Czech, Huang, Lou, and Wang (2021). For each client, we sort trading days into two groups – high-connection and low-connection days – based on whether the number of connections of the client in that day is above or below the average for the given client. To avoid look-ahead bias, we use extended windows to compute average connections, i.e. we re-compute average connections for each client using a sample up to the given trading day. For the high-connection client-day pairs, we proceed in three steps. First, for each gilt we calculate the daily aggregate orderflow generated by high-connection clients. Second, we rank all gilts on each trading day based on the computed order flow, i.e. gilts that high-connection clients predominantly bought (sold) will feature at the top (bottom) of this daily ranking. Third, we build a long-short portfolio that goes long on the top tertile and goes short on the bottom tertile of gilts by this ranking. We then compute the cumulative daily returns of this portfolio up to 20 days, as shown in the left panel of Figure 4.

For the low-connection client-day pairs, we follow the three steps above, and the cumulative daily returns of this portfolio is shown by the right panel of Figure 4.[17]

[17]It is important to highlight that our sorting of gilts using clients' orderflows is based on the within-

Figure 4: Long-short Portfolio Returns – Sorted by Daily Orderflows of High/Low Connection clients



Notes: The figure shows the cumulative return of a long-short portfolio that is based on the daily orderflow of more sophisticated clients with fewer (left panel) or more (right panel) dealer connections than their sample average. The sorting of clients is based entirely on the within-variation of connections. On each trading days, all government bonds are sorted into two groups based on the orderflow of connected and unconnected clients. To avoid look-ahead bias, we use extended windows to compute average connections, i.e. we re-compute average connections for each client using a sample up to the given trading day, thereby excluding future days from the calculation of the sorting variable. The portfolios are rebalanced every day and are held for 1, 2,... 20 trading days. The dashed lines denote 90% confidence intervals based on robust standard errors.

The results show that the high-minus-low portfolio of high-connection clients yields positive returns, reaching about 5bp after 20 days. In contrast, the high-minus-low portfolio of low connections generates returns that are statistically insignificant in the long horizon. That is, we find that there is a strong information content in the aggregate portfolio choice of connected clients.

4.2.2 Aggregate Connections and the Yield Curve

Specifically, we test whether time-variation in aggregate client-dealer connections in the market can explain variation in yields.

We start by constructing an aggregate measure of connections defined as the total number of unique client-dealer connections on a given trading day. We then examine the relationship between changes in aggregate connections and the absolute changes of the

variation of connections as well, as opposed to sorting clients purely based on characteristics that vary in the cross-section (e.g. Czech, Huang, Lou, and Wang (2021)).

5-year yield. All our regressions control for trading volume, given previous evidence on the relation between price changes and volume (Karpoff, 1987; Bessembinder and Seguin, 1993).

Table 4: Explaining Daily Changes in 10-year Yields with Aggregate Connections

	More Sophisticated Investors			Less Sophisticated Investors		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log (Connections_t)$	0.028***	0.034***		0.021***	0.001	
	(6.22)	(3.57)		(5.88)	(0.07)	
$\Delta \log (NumOfClients_t)$		-0.016*			0.024***	
		(-1.70)			(2.87)	
$\Delta \left(\frac{Connections_t}{NumOfClients_t} \right)$			0.011***			-0.003
			(2.93)			(-0.79)
N	1450	1450	1450	1450	1450	1450
R^2	0.032	0.035	0.029	0.027	0.034	0.019

Notes: this table regresses the absolute value of daily changes in the 10-year yield on daily changes in the logarithm of the total number of aggregate connections, the total number of clients and changes in connections per clients. We run separate regressions for more sophisticated clients (columns 1-3) and for less sophisticated clients (columns 4-6). The transaction-level data is collapsed at the day level yielding 1450 trading days spanning the period 4 Oct 2011 to 30 June 2017. Data on yields are from the Bank of England Financial Database. T-statistics, based on robust standard errors, are in parentheses. All regressions controls for trading volume (Karpoff, 1987; Bessembinder and Seguin, 1993). The coefficients for the deterministic variables (constant, linear and quadratic time trends) and trading volume are not shown. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

Results are summarised in Table 4, distinguishing between more and less sophisticated clients. We find a statistically strong relationship (2.8bp and 2.1bp) between daily changes in aggregate connections and absolute deviations in yields levels both types of clients. However, we detect differences in the sources of explanatory power, as we include the total number of unique clients (Tauchen and Pitts, 1983) as a control in the regression. For more sophisticated clients, connections continue to be significant (column 2), whereas for less sophisticated clients, connections are no longer significant (column 5). This means that in the case of more sophisticated clients it is, effectively, the changes in the total number of dealer connections per client (and not the changing number of clients) that drive day-to-day yield changes; whereas, in the case of less sophisticated clients, it is their increased presence which affects yields. We check this explicitly by including the change in the total number of connections as a regressor (columns 3 and 6). Overall, these results suggest that the aggregate connections of more sophisticated clients are an important conduit by which new (private and public) information gets built into prices, consistent with Hypothesis 4.

So far, our empirical results provide support for Hypotheses 1–4. In the next Section we consider alternative explanations behind our results.

5 Alternative Explanations

The main premise of this paper is that when a client trades with more dealers it is a sign that she has private information about the future price movement. This private information is reflected in an expected price increase (decrease) of a given asset when she buys (sells) from an unusually large number of dealers. In this section, we consider and refute potential alternative explanations for our findings which do not require connected clients to have private information.

5.1 Connections as a Proxy of Large Demand Shocks

Hiding information is not the only possible motivation to split trades across dealers. When trading cost are convex in the size of the order, clients who has to trade large quantities for non-informational reasons might be also motivated to increase their connections and trade with many dealers to save on transaction cost (see O'Hara, 2015; Choi, Larsen, and Seppi, 2019). In this part, we investigate whether such uniformed trades might be explaining our results.

The first sign that high connections is unlikely to proxy for periods when uninformed clients are splitting orders across dealers to reduce transaction costs is our results in panel (b) of Table 2. While such periods might correspond to larger transaction component of performance, they are unlikely to correspond to a larger anticipation component of performance. For instance, a client who is forced to buy a large quantity might be able to do so at a smaller price increase when splitting the order across multiple dealers compared to periods when she trades the same quantity with the one dealer. This would lead to a larger transaction component. However, it is not clear why the price is more likely to increase further in subsequent days.[18]

For a more direct test, note that the main determinant of whether a client reaches out to more dealers is different under our main hypothesis and the alternative. In our case, the client increases connections whenever her information is more precise. Under the alternative hypothesis, the client does so whenever she has to trade a lot. That is, in the latter case the client's trading volume should be able to fully take over the role of her connections. As long as information precision and the client's trading volume is not perfectly correlated[19], under our hypothesis, connections should play an independent

[18]One might wonder whether correlated demand shocks across clients and across time might explain such pattern. In Online Appendix A.2, we provide a formal argument to show that it is unlikely.

[19]We should expect some positive correlation between information precision and order size as informed

role.

Note that all our regressions already include clients' daily trading volume which aims to control for the linear effect of demand shocks. However, it is still possible that volume interacts with connections and performance non-linearly. We test for this possibility as follows.

We are interested in whether the connection-performance relation is stronger during trading days when a given client trades particularly large amounts compared to when she trades relatively little. If connections simply proxy large demand shocks, we would expect the connection-performance relation to be stronger during large volume days. To reinforce that our results are not simply picking up the effect of trading volume (driving both connections and performance), as a first step, in panel (a) of Table 5, we double-sort our sample using the within-variation both in connections and in trading volume. The performance difference on high and low connectivity days is approximately the same irrespective of whether the client's trading volume is high or low, and thereby the majority of positive trading performance continues to concentrate in high connectivity days.

clients are likely to trade more aggressively when their information is more precise. However, it is a reasonable assumption that most clients trade government bonds both for speculative motives and for uninformed motives (e.g. to fill the order of their own clients, to hedge their interest rate exposure on other markets, to obtain collateral for their funding activity etc.). Our test has power as long as the latter motive does not lead to systematically smaller trades than the former motive.

Table 5: The Role of Trading Volume

		Average Daily Volume (in £000s)		Average Daily 4-day Performance		Decomposition of Gross Performance	
Volume	Connections	Mean	Median	Mean	Median	Mean	Median
Low	Low	17,300	2,400	-0.414	0.113	-28%	7%
Low	High	44,800	9,700	0.733	0.403	128%	93%
High	Low	83,100	20,600	-0.347	0.147	-33%	12%
High	High	173,000	57,400	0.676	0.380	133%	88%

(a) Economic Significance of Connections

	1-day	2-day	3-day	4-day	5-day
Client Connections *	0.353***	0.488***	0.668***	0.760***	0.784***
Low Volume Days	(3.46)	(3.85)	(4.05)	(3.62)	(3.37)
Client Connections *	0.136	0.195*	0.319**	0.461***	0.498***
High Volume Days	(1.53)	(1.80)	(2.34)	(2.66)	(2.64)
Volume	0.361***	0.491***	0.555***	0.482**	0.487**
	(2.64)	(3.18)	(2.89)	(2.21)	(2.01)
Tran.	-0.744***	-1.144***	-1.528***	-1.473***	-1.889***
	(-2.74)	(-2.98)	(-3.64)	(-3.18)	(-3.83)
p-value, equality of connection coefficients	0.00004	0.00005	0.00007	0.0109	0.0162
N	100414	100414	100414	100414	100414
R^2	0.057	0.056	0.057	0.057	0.058
Day FE	Yes	Yes	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes	Yes	Yes

(b) Connection and Performance on Low and High Volume Days

Notes: Panel A illustrates the economic significance of the performance-connectedness relationship. The sample (at the client-day level) is split into four groups using double-sorting, based on the within-variation of both connections and trading volume. The numbers in the last two columns decompose gross performance (defined as the product of volume and performance) into the contribution of each group. The 4-day performance measures are in basis points. Panel B regresses the value-weighted trading performance at different time horizons on client connections interacted with a dummy taking value 1 (0) if the client's daily trading volume is above (below) her sample average. The transaction-level data is collapsed at the client-month level. The performance measures are in %-points. T-statistics in parentheses are based on robust standard errors, using two-way clustering at the day and the client level. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

We also formally test the differential effect of volume as follows. Panel b of Table 5 shows the results for a variant of our baseline regression (4.1) where we interacted our measure of connections with a dummy indicating whether a client traded on a given day more or less than her sample average. The results shows that, if anything, connection effects are stronger during low volume days. As shown by Figure 8 in the Appendix, the results are similar for longer horizons irrespective of whether we use the full sample or the subsample consisting only of more sophisticated clients.

5.2 Informed Dealers and Uninformed Clients

An other possibility is that a client trades with many dealers when (some of the) dealers are informed, even if the client is uninformed. For instance, one such mechanism is that a dealer, expecting a price increase, would give unfavourable ask quotes to increase its inventory. This might motivate clients to requests quotes from other dealers too, ultimately increasing their number of connections. To test for these, we perturb our baseline regression 4.1 by including additional, dealer-specific controls. First, we compute that average number of trading relationships (with clients) and the average trading volume of all dealers that a given client trades with on a trading day. These measures are motivated by the recent work Maggio, Franzoni, Kermani, and Sommavilla (2019) which focuses on the heterogeneity in the centrality of stock-market brokers to study information diffusion. Then, we decompose the 4-day trading performance into transaction and anticipation components and regress these components on clients' connections and dealer-level variables similarly to panel (b) of Table 2.

The results are shown in Table 6. We find that the added dealer characteristics mainly affect the transaction component. Including both dealer connections and dealer volume (column 3) somewhat lowers the connection effect on the transaction component (0.082). This result is primarily driven by the inclusion of trading volume, suggesting that some of the connections effect in our baseline is picking up that clients who increase their connections to receive tighter bid-ask spreads tend to trade with larger dealers who provide more favourable prices.

However, turning to the anticipation component (columns 4-6), we find no evidence that dealer characteristics affect our baseline connection results. That is, the direction of trade of connected clients predicts the subsequent price movement, while the characteristics of their dealers do not. This enforces our interpretation that client connections proxy for clients' private information.

While these results suggest that information at the dealer level cannot explain our baseline results, it by no means proves that there is no information flow between clients and dealers. In fact, in the next Section we argue that such two-way information flow is present in our data. In particular, we present evidence that dealers learn from their informed clients and let that information leak to their affiliates.[20]

[20] As a related exercise, in Online Appendix C, we exploit the incidents when two primary dealers ceased their market making functions during our sample period leading to a plausibly exogenous shock to the connections of those clients which previously had been in contact with the exiting dealers. We find no evidence that this exogenous shock significantly reduced the performance of these clients. This reinforces our interpretation that clients proxy for existing private information, as opposed to being the

Table 6: Decomposing 4-day Performance: Controlling for the Average Characteristics of Clients' Dealers

	(1)	(2)	(3)	(4)	(5)	(6)
	Transaction Component			Anticipation Component		
Client	0.098**	0.111***	0.082**	0.467**	0.456**	0.468**
Connections	(2.43)	(2.69)	(2.16)	(2.34)	(2.27)	(2.34)
Mean Connections of		0.034***	-0.005		-0.031	-0.014
Client's Dealers		(2.95)	(-0.41)		(-0.58)	(-0.25)
Dealers' Mean			0.384***			-0.162
Volume			(4.43)			(-0.49)
N	100348	100348	100348	100348	100348	100348
R^2	0.095	0.095	0.096	0.056	0.056	0.056
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: this table decomposes the 4-day performance effect into a transaction component and an anticipation component (3.2) and regresses the components on client connections and average characteristics of the dealers that clients trade with. Mean Connections of Client's Dealers is the average number of trading relationships of *all dealers* that a given client trades with on a trading day. Similarly, Dealers' Mean Volume is the log of average trading volume of all the dealers that a client trades with on a trading day. The transaction-level data is collapsed at the client-day level. The performance measures are in basis points. All regressions include as additional controls (not shown) the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Tran."). The computation of the transaction component is based on the average transaction price $\overline{P}$ that uses the trades (for the given gilt) in a 3-hour window within the transaction time. T-statistics in parentheses are based on robust standard errors, using two-way clustering at the day and the client level. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

6 Applications

Having established that client connections serve well as a proxy for private information in dealer markets, one can use this proxy to empirically investigate a number of long-standing issues in the finance literature.

In this section, we turn our attention to two questions in particular. First, in Subsection 6.1, we explore the leakage of information from dealers to their preferred clients. In particular, we present suggestive evidence that dealers learn from their informed clients and pass this information to their affiliates.[21]

Second, in Subsection 6.2, we are interested in the the nature of private information

medium to transfer the information of dealers to the clients. In Online Appendix C, we also discuss the caveats of this exercise.

[21] In a related set of papers, Maggio, Franzoni, Kermani, and Sommavilla (2019); Barbon, Maggio, Franzoni, and Landier (2017) shows that brokers in stock markets pass on orderflow information to the clients that they have had a strong trading relationships. It is important to keep in mind that our set up is differs from this line of work even beyond the difference of the traded asset. The identification in these papers relies on the heterogeneity of stock market brokers a given client trades with. In the UK government bond market there are fewer dealers and they are more homogeneous, leading to more limited cross-sectional variation in their centrality.

in treasury markets. This is an intriguing topic as this market is often viewed as a market with little role for private information. Our main focus is to assess whether the private information captured by the time-variation in clients' connectedness is on fundamentals, or on future orderflow. We find evidence for both. On the one hand, we present suggestive evidence that connections identify speculative trading activity before the Brexit referendum. We also show evidence that our main findings are more pronounced around macroeconomic announcements. This reinforces that fundamental information plays some role in our mechanism, at least around key macroeconomic events.

On the other hand, we find systematic evidence that more connected clients can better predict the maturity structure of other clients' orders, especially the part of the orderflow that is received by their own dealers in subsequent days. For instance, when a more connected client's orders are concentrated on the short-end of the yield curve in a given day, her dealer is more likely to receive a disproportionate share of orders for short bonds in the following five days. We also show that trading in line with the maturity structure of clients' future orders is profitable because of the resulting pressure on prices.

6.1 Information Leakages

In this part, with the help of our particularly detailed data set, we investigate the information leakage from dealers to their trading affiliates. While our baseline analysis took the approach to consolidate the accounts of dealers (and of clients) at the highest possible level, in this subsection we dive in to explore the heterogeneity in the trading accounts of dealers to study information leakages.

For each dealer, we are able to distinguish between trading accounts that perform a market-making function (trading primarily with clients, executing large number of transactions, participating in primary auctions) from trading accounts that correspond to other, client-like trading affiliates of the given dealer bank (trading primarily with other dealers, executing lower number of transactions, e.g. asset-manager arms). We refer to this latter group of accounts as the given dealer's affiliates.[22]

[22]Specifically, we identify 321 trading accounts, corresponding to 21 dealers, totalling about 662 thousand transactions. These accounts can have orders of magnitude fewer transactions (than the market-making accounts), and these accounts trade predominantly with other dealers instead of other clients. See Table 12 of the Appendix for summary statistics. The names corresponding to these account include the terms 'Asset Management', 'Wealth Management', 'Private Bank', 'Investment Management', 'Managed Funds' among others as well as geographical names such as 'Zurich', 'Europe' etc. One caveat of our classification, therefore, is that we group together multiple trading accounts of a given dealer that may perform diverse functions, leading to possible measurement error. The list of unique BIC identifiers along with the names of these trading accounts can be found in the online replication material.

We then test whether dealers' affiliates perform better when the given dealer trades with a larger proportion of high-connection clients. The idea is that even if spreading out trades across multiple dealers slows down the diffusion of information from clients to their dealers, the dealers with a larger proportion of informed clients should still be able to learn more from their clients' orderflow. Under this conjecture, higher trading profit of affiliates indicates that this informational advantage is passed on towards them.

To use our connectivity measure to proxy the informativeness of client orderflow that a given dealer faces, we construct the following measure for each dealer i on day t:

$$InfShare_{i,t} = \frac{Vol^H_{i,t}}{Vol^L_{i,t} + Vol^H_{i,t}}, \tag{6.1}$$

where $Vol^H_{i,t}$ and $Vol^L_{i,t}$ are the trading volume of dealer i with clients whose connectivity on day t is high and low, respectively. Again, we rely purely on the time-series variation in connectivity when sorting client-day pairs. Specifically, we place each trading day of each client into three tertiles based on connections. If the given client is in the top tertile on a given trading day, then her trading volume with dealer i contributes to $Vol^H_{i,t}$. In all other trading days the same client's trading volume with dealer i contributes to $Vol^L_{i,t}$.

We use measure 3.1 to compute the daily trading performance of dealers' affiliates, $AffilPerformance_{i,t}$. To test whether affiliates perform better when their dealers trade with more connected clients, we estimate the following daily panel regression for each dealer i and day t:

$$AffilPerformance^T_{i,t} = \beta \times InfShare_{i,t} + X_{i,t} + \alpha_{i,year} + \mu_t + \varepsilon_{i,t}, \tag{6.2}$$

where $\alpha_{i,year}$ and μ_t are client-year and day fixed effects; and $X_{i,t}$ includes control variables such as the number of transactions and trading volume of dealers' affiliates as well as the dealers. The main coefficient of interest in 6.2 is β which captures the relation between dealers' enhanced interaction with high-connection clients and the performance of dealers' affiliates.

Table 7: Dealers' Informed Clientele and the Performance of Dealers' Affiliates

	0-day (1)	1-day (2)	2-day (3)	1-day (4)	2-day (5)
InfShare	0.278	1.690**	1.889*	1.657**	1.812*
	(0.59)	(2.26)	(2.03)	(2.29)	(1.95)
DealerVolume	-0.005	-0.377**	-0.328	-0.384*	-0.338
	(-0.04)	(-2.10)	(-1.19)	(-2.08)	(-1.23)
DealerConnections	-0.003	0.016	-0.028	0.016	-0.027
	(-0.11)	(0.31)	(-0.32)	(0.31)	(-0.31)
AffilConnections				0.024	0.015
				(0.14)	(0.07)
InfShare of				-0.548	-1.567
OtherDealers				(-0.44)	(-0.73)
N	20898	20898	20898	20880	20880
R^2	0.081	0.082	0.079	0.082	0.079
Day FE	Yes	Yes	Yes	Yes	Yes
Affil.#Year FE	Yes	Yes	Yes	Yes	Yes

Notes: this table shows the results for regression 6.2, which regresses the value-weighted trading performance of dealers' affiliates at different time horizons on our informativeness measure (6.1). The transaction-level data is collapsed at the affiliate client-day level. The performance measures are in bp-points. For columns (1)-(3), we include as controls the natural logarithm of the pound trade volume of affiliate clients (not shown) and the affiliate's dealers ("DealerVolume"), and the natural logarithm of the number of daily transactions of affiliates (not shown) and total number of client connections of affiliates' dealers ("DealerConnections"). Controls not shown have statistically insignificant coefficients. For columns (4)-(5), we add as controls the connections of the affiliate clients ("AffilConnections"), the informativeness of client orderflow of other dealers that the affiliate client is connected to ("InfShare of OtherDealers"). T-statistics in parentheses are based on robust standard errors, using two-way clustering at the day and the affiliate client level. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

Columns (1)-(3) in Table 7 show that when a dealer's clientele goes from low connectivity to high-connectivity, then the trading performance of the dealer's affiliates improves by around 1.7-1.8 basis points over the 1-2 day horizon. Interestingly, there is no effect at the 0-day horizon suggesting that we are not simply picking up that dealers' affiliates get better information about bid-ask spreads; the information they learn is informative about imminent changes in the yield curve over the coming days. It is worth noting that these results are not just detecting that larger/better dealers who tend to attract more informed (and connected) clients might have higher performing affiliates: the inclusion of affiliate client fixed effects absorbs this type of time-invariant heterogeneity across dealers.

Note that the regressions also control for the trading volume and the connections of the dealer (corresponding to the given affiliate whose performance we aim to explain) with little effect on the baseline. This suggests that the estimated β is not simply picking up that dealers' trading volume, while being correlated to our measure $InfShare_{i,t}$, conveys information about yields to dealers' affiliates in the spirit of Kyle (1985). Instead, it is the composition of dealers' clientele that determines how dealers' affiliates perform (against other dealers). This suggests that affiliates obtain the information that their dealers learn from informed clients.

There are potential caveats to consider. For instance, (i) the information, which increases the profitability of dealers' affiliates, might originate from other dealers that these clients are connected to, or (ii) it might originate from the affiliates themselves. To reinforce that neither is the case, we perturb our research design by adding two control variables to regression 6.2. First, we build on measure 6.1 to compute the average informativeness of all other dealers that the given affiliate is connected to (excluding the given affiliate's own dealer from this average measure). The constructed variable ("InfShare of OtherDealers") is aimed to control for the first identification concern. Second, we include as a control the number of dealer connections of affiliates ("Subsid Connections") to address the second concern. Columns (4)-(5) of Table 7 show that these additional controls are statistically insignificant and their inclusion in the regression makes little difference to the coefficient on $InfShare_{i,t}$.[23] We interpret these results that the information is in fact flowing from dealers to their affiliates and not the other way around.[24]

Other interpretation might include that affiliates are learning from highly connected dealers without the intermediation of dealers or their information is coming from the same source. Because of these caveats, we think of the results in this section as suggestive and subject to future research.

6.2 The Nature of Private Information: Future Fundamentals or Future Orderflow

In this part, we investigate the nature of private information clients' connections may proxy. Our main focus is to assess whether the private information captured by the time-variation in clients' connectedness is on fundamentals, or on future orderflow. We find evidence for both.

[23]Figure 10 in Appendix shows the results over longer horizons from the model corresponding to columns (4)-(5) in Table 7.

[24]Moreover, we also relate this analysis to that of Maggio, Franzoni, Kermani, and Sommavilla (2019) which focuses on the cross-sectional heterogeneity in the eigenvalue-centrality of stock-market brokers to study information diffusion. The main premise of their paper is that more central brokers gather and disseminate more information than less central brokers do. This begs the question of whether the eigenvalue-centrality of a dealer in our application proxies the time-variation in the share of connected clients in the dealer's total client base, measured by $InfShare_{i,t}$. We find no statistically significant effect for dealer centrality and including it in the regression makes little difference to the coefficient on $InfShare_{i,t}$, as shown our working paper (Kondor and Pinter, 2019).

In the first part, using the case study of the Brexit referendum, and re-running our main tests with distinguishing days with and without macroeconomic announcements, we argue that fundamental information plays some role in our mechanism, at least around key macroeconomic events. In the second part, we present evidence that more connected clients can better predict the maturity structure of other clients' orderflow, especially the part of the orderflow received by their own dealers in subsequent days. This suggests that clients' connections in government bond markets might proxy both for private information on price effects of anticipated orderflows and that of information connected to macroeconomic events.

6.2.1 Connections as a Proxy for Information on Future Fundamentals

Betting on Brexit: An Event Study In this section, we take a close look at the connectedness-performance relationship during the days around the Brexit referendum on leaving the European Union. The referendum took place on Thursday 23 June 2016, and the results that 51.9% of the participants voted to leave became public on Friday morning (24 June 2016). Based on polls, the chances of a leave or a remain vote were close to 50-50 leaning slightly towards remain immediately before the vote. Either way market prices were expected to jump. In particular, the common perception was that a leave result would likely trigger a rate cut soon, leading to an immediate downward shift in the yield curve on 24 June. Indeed, this is what happened with the 1-year, 5-year and 25-year yields dropping by 14bp, 30bp and 24bp, respectively, on 24 June. This was followed by the Bank of England cutting the policy rate by 25bp in August.

Given the large uncertainty before the vote, market participants were motivated to either reduce their exposure radically, or to generate private information and bet on the outcome. Reportedly, major hedge funds ordered private opinion polls to generate an informational edge.[25] Our main hypothesis implies that we should be able to separate these two groups from each other based on the change of their connectivity before the

[25] Reportedly, major hedge funds ordered private opinion polls to generate an informational edge for this bet and earned handsomely on those bets:

> "Behind the scenes, a small group of people had a secret – and billions of dollars were at stake. Hedge funds aiming to win big from trades that day had hired YouGov and at least five other polling companies [...]. Their services, on the day and in the days leading up to the vote, varied, but pollsters sold hedge funds critical, advance information, including data that would have been illegal for them to give the public. Some hedge funds gained confidence, through private exit polls, that most Britons had voted to leave the EU, or that the vote was far closer than the public believed – knowledge pollsters provided while voting was still underway and hours ahead of official tallies." ("The Brexit Short: How Hedge Funds Used Private Polls to Make Millions", Bloomberg Businessweek, 25th June, 2018)

vote. We should see that clients with private information increase the number of dealers they trade with to hide this information. Furthermore, they should be the group who, in average, increases the duration of its portfolio to speculate on the leave outcome and when the yield curve eventually drops, they should overperform the others.

To verify this hypothesis, we group all those private clients who traded on the referendum day 23 June into two groups based on how many dealers they traded with compared to their average daily dealer connections during the whole sample (2011 Oct – 2017 Jun). We end up with a total of 131 clients who traded on the day of the referendum. The 5-day trading performance of connected clients were about 0.5bp on average compared to -0.15bp of unconnected clients. While connected clients had a larger average trading volume than unconnected (£39m compared to £18m), the major difference between the client groups is in the direction and magnitude of the duration purchased: connected clients on average bought £121m of duration whereas unconnected clients sold an average £0.2m on 23 June.

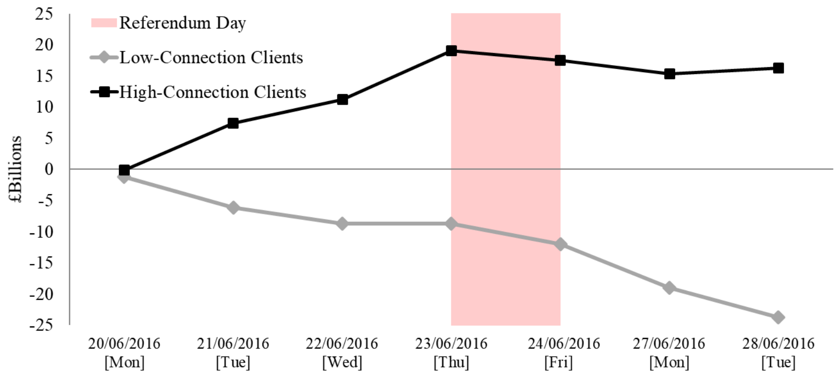
To illustrate these facts visually, panels (a) and (b) of Figure 5 show the total duration and cumulative performance of the two groups in the days around the vote. While this episode is intuitive, note that the differences in performance of the high- and low-connection groups might come from other, unobserved heterogeneity in these two groups. Indeed, it is quite likely that the traders who decide to bet on the outcome of the Brexit vote are very different from those institutions that decide to cut back their exposure in this volatile period. Also, this particular episode might be special. Hence we turn to systematic evidence next where we can include client- and time- fixed effects as well as additional controls to decompose the different forces at play.

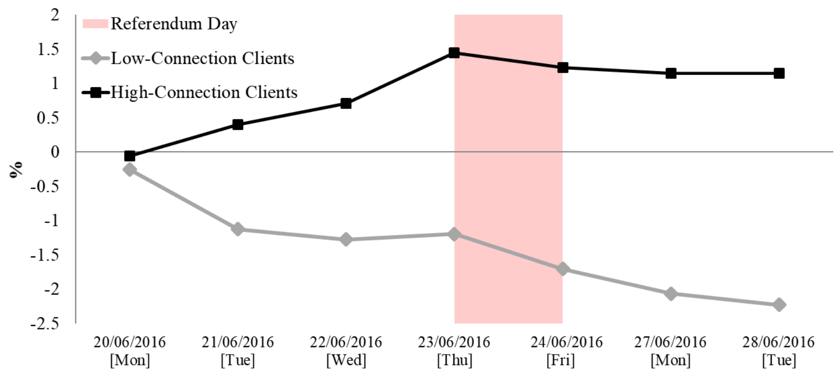
Connections During Informationally Intensive Periods To further test whether connections proxy fundamental private information, we divide our sample into periods with more or less fundamental information. A natural proxy for periods when fundamental private information is especially relevant are trading days that coincide with the release of macroeconomic news.

Macroeconomic news hit markets almost constantly, some of which may contain very little surprise component or little relevance to affecting prices. It is therefore important to identify trading days where macroeconomic news truly surprised markets and moved prices non-trivially. To do so, we build on the high-frequency methodology of Swanson and Williams (2014a,b) to identify the surprise components of macroeconomic announce-

Figure 5: Connections and Performance around the Brexit Referendum

(a) Cumulative Daily Net Duration of High and Low Connection Clients





(b) Cumulative Returns of High and Low Connection Clients

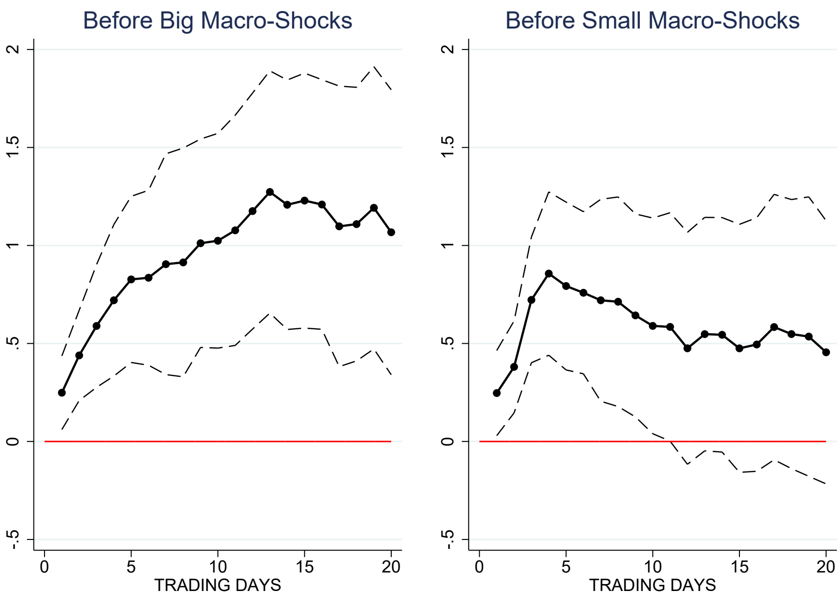
Notes: In Panel a, the black squared line depicts the evolution of the cumulative duration-weighted net position (in £billions) of those 66 clients that have more dealer-connections on the day of the referendum than their sample average. The grey diamond line shows the result for those 65 clients that have low within-connectedness on the day of the referendum. In Panel b, the black squared (gray diamon) line depicts the cumulative average returns (in %) of high-connection (low-connection) clients. The average returns for both groups are weighted by the individual clients' daily trading volume. The returns are computed using the closing price on 29 June 2016 as the reference price.

ments and their effects on bond prices.[26] We sort trading days into two groups depending on whether the magnitude of the macroeconomic surprise on the given day was smaller or bigger than the sample average.

We then re-estimate a variant of our baseline model 4.1, where we interact connections with an indicator variable for small and big macroeconomic news days. Importantly, we lead the indicator variable by one period, thereby asking whether connections matter more for performance *ahead of* big macroeconomic shocks. Thereby we test whether the

[26] Our dataset is obtained from colleagues from the Bank of England (building on Eguren-Martin and McLaren (2015)). The method uses historical tick data to compute the change in the 3-year interest rate in a tight window (five minutes before and five minutes after) around the release of both nominal and real news from both the UK and the US.

Figure 6: Connections and Performance over 1-20 Day Horizons: Before Macroeconomic Surprises



Notes: this figure plots the estimated β coefficients from variant of regression 4.1 up to 20-day horizon ($T = 20$), where we interact connections with an indicator variable for days before small and large macroeconomic surprises. The construction of surprises follows the methodology of Swanson and Williams (2014a,b). We restrict the sample to more sophisticated investors, and include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Transactions"). The shaded area denotes the 90% confidence band associated with the estimated β coefficients, It is based on robust standard errors, using two-way clustering at the day and the client level.

information proxied by connections has predictive power of future macroeconomic news.[27] Figure 6 shows that the point estimates for the effect of connections of more sophisticated clients on their trading performance tend to be higher and are more persistent before big macroeconomic surprises compared to days before small surprises.

Next, we study the possibility that clients' connections may also proxy private information about future orderflow.

6.2.2 Connections as a Proxy of Information on the Price Impact of Future Orderflow

Our starting point is the empirical literature (Fleming and Remolona 1999; Evans and Lyons 2002; Brandt and Kavajecz 2004; Menkveld, Sarkar, and van der Wel 2012) on

[27] In the previous version of our paper (Kondor and Pinter, 2019) we also document that the relationship between connections and trading performance is stronger *during the contemporaneous* release of monetary news, consistent with some market participants having an advantage at processing the newly arrived information (Kandel and Pearson, 1995; Pasquariello and Vega, 2007).

the role of orderflow in driving prices in various dealer markets. As orderflow has a price impact, private information about future orderflow can be used profitably. In this part, we study the possibility that connections are the proxy for such private information.

First, we propose a measure of co-movement of the composition of client's orders with the future aggregate orderflow of a given group of clients. The idea is that whenever this measure is positive, the client, intentionally or by chance, is effectively front-running that group of clients. We test whether this measure identifies profitable trades. We indeed find that whenever the duration composition of a client's trade is similar to that of all the other clients in subsequent days, her performance is higher. Second, we connect our baseline results to orderflow information: we show that whenever a client is more connected, the composition of her trades tend to be more similar to the group of clients in subsequent days who are served by the same dealer. We also show that a client who is a regular counterparty of the given dealer can predict the composition of the orderflow better. This raises the possibility that dealers have a role in disseminating orderflow information towards their own, regular clients.

Measuring Co-movement between Client Trades and Future Orderflow Our proposed measure aims to capture whether a client trades in the same direction as other clients in the subsequent trading days. First, we partition all transactions in K equal-sized segments based on the modified duration of all traded gilts. We then compute the net trading position of client i, on day t, in duration segment k, $W_{i,t,k}$.[28] We then calculate the cumulative net trading position of group g between days $t+1$ and $t+T$ in duration segment k, denoted by $W^g_{t+T,k}$. The identity of group g will play an important role in section 6.2.2 where we identify the group whose orderflow connected clients can forecast. For now, we set g for the group of all the clients in the market. Our daily covariance measure, $\Psi^{T,g}_{i,t}$, is then computed as follows:

$$\Psi^{T,g}_{i,t} = \frac{1}{K}\sum_{k=1}^{K}\left(W_{i,t,k} - \frac{1}{K}\sum_{k=1}^{K} W_{i,t,k}\right)\left(W^g_{t+T,k} - \frac{1}{K}\sum_{k=1}^{K} W^g_{t+T,k}\right). \tag{6.3}$$

When $\Psi^{T,g}_{i,t}$ is high, the given client tends to concentrate her orders in the same segment as group g in the subsequent T days.[29]

Given this measure, we first check whether it is profitable to guess right the segments

[28] Clients' net positions correspond to their orderflow in this market, as client-dealer trades are initiated by clients.

[29] Our working paper (Kondor and Pinter, 2019) provides a visual illustration of this measure.

of the yield curve where future demand pressure will be concentrated. For each client i we partition the trading days into two sets, indicated by the dummy variable $D_{i,t}^{T,g}$, based on whether $\Psi_{i,t}^{T,g}$ for the given day is larger or smaller than the client-specific median in the full sample. Then we estimate the following regression:

$$Performance_{i,t}^{T} = \gamma \times D_{i,t}^{T,g} + \alpha_i + \mu_t + \varepsilon_{i,t}, \tag{6.4}$$

where $Performance_{i,t}^{T}$ is our baseline performance measure (3.1). Table 8 summarises the results. Panel A shows the results when the covariance measure uses the cumulative orderflow of the market ($g = Total$) at the 1-day horizon (columns 1-2) and 5-day horizon (columns 3-4). For the latter, we compute the turnover-weighted performance measures at the 1- and 3-day horizons. For the former, we compute performance at the 5- and 7-day horizons. We find that the trading performance of a client can be 2-3bp higher on high covariance days, i.e. predicting the orderflow of the market is profitable. Panel B shows the results when the covariance measure uses the cumulative orderflow of the subset of the market that is intermediated by the dealers that the given client is connected to ($g = Own$). We find that if a client can predict this subset of the aggregate orderflow, it is still profitable with performance being 1-2bp higher on high covariance days.

Table 8: Trading Performance on Days with High Covariance with Future Orderflow

	1-day Covariance		5-day Covariance	
	1-day Perf.	3-day Perf.	5-day Perf.	7-day Perf.
	(1)	(2)	(3)	(4)
$Q_{i,t}^{Total} = 1$	2.186***	3.211***	2.439***	2.571***
	(5.66)	(5.26)	(3.07)	(2.73)
Volume	0.223*	0.346*	0.347	0.333
	(1.80)	(1.96)	(1.57)	(1.30)
Tran.	-0.526**	-1.033***	-1.133***	-1.312***
	(-2.09)	(-3.01)	(-3.06)	(-2.69)
N	100311	100311	100040	100039
R^2	0.058	0.058	0.058	0.056
Day FE	Yes	Yes	Yes	Yes
Client-Day FE	Yes	Yes	Yes	Yes

(a) Covariance with the *Total Market* Orderflow

	1-day Covariance		5-day Covariance	
	1-day Perf.	3-day Perf.	5-day Perf.	7-day Perf.
	(1)	(2)	(3)	(4)
$Q_{i,t}^{Own} = 1$	0.607**	1.096**	1.532***	2.023***
	(2.37)	(2.58)	(2.86)	(3.47)
Volume	0.211*	0.344*	0.349	0.338
	(1.70)	(1.94)	(1.60)	(1.34)
Tran.	-0.516**	-1.039***	-1.155***	-1.334***
	(-2.04)	(-3.03)	(-3.11)	(-2.76)
N	100407	100407	100407	100406
R^2	0.057	0.057	0.058	0.056
Day FE	Yes	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes	Yes

(b) Covariance with the Market Orderflow Intermediated by *Own Dealers*

Notes: panel A regresses the value-weighted trading performance at different time horizons on a dummy $Q_{i,t}^{Total}$ that takes value 1 if on day t the orderflow of client i has a covariance (see measure 4.1) with the future orderflow of the market that is higher than the median (based on all trading days of the given client). Panel B regresses performance at different time horizons on a dummy $Q_{i,t}^{Own}$ that takes value 1 if on day t the orderflow of client i has a covariance (see measure 4.1) with the future orderflow of its own dealers that is higher than the median (based on all trading days of the given client). Own dealers ($g = Own$) are the ones that the client traded with on the day of the trade (for which the trading performance is calculated) as well as during the past 10 trading days. The transaction-level data is collapsed at the client-day level. The performance measures are in %-points. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Tran."). T-statistics in parentheses are based on robust standard errors, using two-way clustering at the day and the client level. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

Table 9: Client Connectivity and Covariance with the Orderflow

	Regular Connections		Non-Regular Connections	
	1-day Covariance	5-day Covariance	1-day Covariance	5-day Covariance
	(1)	(2)	(3)	(4)
Client	0.0045**	0.0037**	-0.0007	-0.0023
Connections	(2.55)	(2.57)	(-0.42)	(-1.43)
Volume	0.0011	0.0027	0.0046**	0.0044**
	(0.60)	(1.57)	(2.11)	(2.19)
Tran.	0.0061	0.0013	0.0011	0.0033
	(1.36)	(0.26)	(0.22)	(0.75)
N	100407	100407	100407	100407
R^2	0.051	0.054	0.038	0.040
Day FE	Yes	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes	Yes

Notes: this table regresses different versions of the covariance measure 6.3 on our connectivity measure and controls (6.5). The transaction-level data is collapsed at the client-day level. The performance measures are in basis points. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of day transactions ("Tran."). T-statistics in parentheses are based on robust standard errors, using two-way clustering at the day and the client level. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

Connected Clients Predict the Orderflow Let us return to our baseline result that the time-variation in a client's number of connections is a proxy for her level of private information. In this section, we provide evidence that this private information is, at least partially, on the duration composition of the future orderflow of certain group of other clients, as measured by our covariance measure 6.3. In this case, we expect that the covariance measure of a given client on a given day tends to be higher when this client is more connected. Hence, we estimate the following panel regression:

$$D_{i,t}^{T,g} = \phi \times ClientConnections_{i,t} + X_{i,t} + \alpha_i + \mu_t + \varepsilon_{i,t}, \tag{6.5}$$

where the terms on the right-hand-side are identical to our baseline specification 4.1.

Table 9 shows the results after decomposing the aggregate orderflow into the part that is intermediated through the dealers which a given client is connected to ($g = Regular$) and into the part that goes through all the other dealers that the given client does not have a regular relationships with ($g = Non - Regular$). We regard a client-dealer connection regular if the client traded with the given dealer on the given trading day as well as over the previous two weeks. The complement of this set includes relationships that are non-regular.[30]

We find (Columns 1-2) that it is the covariance with *Regular* dealer orderflow that correlates with the client's connectivity, and the effects for $Non - Regular$ dealer orderflow are economically and statistically insignificant (Columns 3-4). Specifically, we find having one additional dealer connection increases the probability that the client is in a high covariance day by about 0.4%. Our interpretation is that dealers, intentionally or unintentionally, disseminate information about future orders towards (some of) their clients. We have little evidence on the exact mechanism. In principle, dealers' private information on their clients expected orders in the subsequent days might be revealed accidentally by the dealers' quotes. Or it might be that there is an intentional information flow from dealers to their best clients helping dealers to keep these clients as suggested by (Maggio, Kermani, and Song, 2017).

Note also that our quantitative findings suggest that this channel explains only a fraction of our baseline effect in Table 2. In our baseline, one additional dealer increases the client's 5-day performance with approximately 0.5 basis-points. This is to compare with the $0.004 \times 1.5 = 0.006$ basis-points for the 0.4% larger probability of a high covariance day with the clients' own dealer 5-day orderflow, and the 1.5 basis-point higher performance on these high covariance days. Whether this is due to our imperfect measurement of orderflow information or due to the importance of fundamental private information should be subject to future research.

7 Robustness

Our baseline performance measure used daily average %-returns weighted by the size of each transaction. To show that our baseline is not driven by this weighting scheme, Table 10 in Appendix re-estimates our baseline model using unweighted performance measures – the results are very similar to our baseline (Table 2).

As a candidate for informationally intensive periods, we used the surprise component in the release of macroeconomic news. As an alternative classification, we also experimented by sorting trading days into two groups based on realised price volatility, measured by the daily dispersion of transaction prices (Jankowitsch, Nashikkar, and Subrahman-

[30]Note that, by the additivity of covariance, our measure is additive in the following sense:

$$YC_{i,t}^{T,Total} = YC_{i,t}^{T,Regular} + YC_{i,t}^{T,Non-Regular}. \tag{6.6}$$

This property helps the interpretation of our results.

yam, 2011). Our motivation for this grouping is that days with high price dispersion proxy for periods of higher market frictions, which may generate more profitable trading opportunities for informed traders. We would therefore expect the relationship between connections and performance to be more pronounced during these periods. Figure 9 in Appendix shows the results for the more sophisticated investors up to the 20-day horizon. We find that the effect of connections of sophisticated clients on their trading performance is stronger, more significant and more persistent on trading days with higher price dispersion compared to days with low price dispersion.

Moreover, all the performance regressions so far were based on data at the client-day level, as opposed to the client-month level. This is to measure the formation and the dynamics of client-dealer connections accurately. This however may lead to problems of oversampling those clients that trade very frequently, i.e. every day. The previous version of our paper Kondor and Pinter (2019) conducted the analysis at the client-month level, and we obtained similar results.

8 Summary and Conclusion

Our paper provided evidence from the UK government bond market that clients better predict future price movements when they have more dealer connections compared to periods when they have fewer connections. This effect is stronger around macroeconomic announcements. We also showed that innovations in the slope and level of the yield curve are associated with days of higher aggregate connections in the market. Based on these findings, we argued that time-variation in client connections serves as an empirical proxy for time-variation in private information. We also presented two applications using this proxy. We found evidence suggesting that dealers leak the information deduced from their client base to their affiliates. We also established that part of the private information identified by connections is related to the maturity structure of the orderflow the given client's dealer is receiving in subsequent days.

These results have several implications. First, our results highlight the relevance of financial network formation to the price discovery process in government bond markets. While the literature has extensively studied the role of private information and aggregate orderflow in determining yield curve dynamics, we find that a better understanding of the network structure can sharpen our understanding of the price discovery process in these markets. Second, while a number of recent papers have studied the core-periphery structure of OTC markets (primarily focusing on the cross-sectional characteristics of dealer-client relationships), our results emphasize the dynamic and endogenous nature of networks. Third, slow trade execution is often regarded as optimal because it minimizes price impact, thereby helping to hide private information (Kyle, 1985). We find that trade execution with multiple primary dealers could serve a similar purpose, suggesting that splitting trades over time and across dealers may be substitutable.

We expect that our insights go beyond treasury markets (as confirmed by a follow-up paper by Czech and Pinter (2020) on the UK corporate bond market), even if it might not be present in all segments of financial markets.

Conceptually, clients are motivated to hide their information by trading with more counterparties in any market where multiple-leg strategies are the norm. This is so, because in this case the trader's individual trades cannot reveal her strategic position, only her portfolio would. Arguably, this is the case in treasury markets where positions on the change of the yield curve consists of long-short positions over multiple durations. This can also be the case in any other markets where traders aim to profit from the relative mispricing of multiple assets. For instance, in the stock market agents might identify stocks with different prices which load on similar risk factors, or in derivative markets agents might profit from complex replicating strategies. However, our insight is less potent when the main element of a strategy is a position in a single asset. For instance, when stock pickers identify a single mispriced stock, trading small amounts with many counterparties might even lead to a faster revelation of their information compared to trading a large amount with a single counterparty.

A clear caveat of our approach is that to calculate clients' connections, a detailed, transaction level data-set, including the identities of market participants, is required. As the trend seems to be that such data-sets are becoming increasingly accessible to the academic community, we expect that our approach opens up new avenues to better understand the role of private information in financial markets.

References

Afonso, G., A. Kovner, and A. Schoar (2014): "Trading partners in the interbank lending market," Staff Reports 620, Federal Reserve Bank of New York. 3.3.2

Angrist, J. D., and J.-S. Prischke (2009): *Mostly Harmless Econometrics*. Princeton University Press, first edn. C.2

Babus, A., and P. Kondor (2018): "Trading and Information Diffusion in Over-the-Counter Markets," *Econometrica*, 86(5), 1727–1769. 4

Back, K., P. Collin-Dufresne, V. Fos, T. Li, and A. Ljungqvist (2018): "Activism, Strategic Trading, and Liquidity," *Econometrica*, 86(4), 1431–1463. 8

Balduzzi, P., E. J. Elton, and T. C. Green (2001): "Economic News and Bond Prices: Evidence from the U.S. Treasury Market," *The Journal of Financial and Quantitative Analysis*, 36(4), 523–543. 1

Barbon, A., M. D. Maggio, F. Franzoni, and A. Landier (2017): "Brokers and Order Flow Leakage: Evidence from Fire Sales," NBER Working Papers 24089, National Bureau of Economic Research, Inc. 21

Benos, E., and F. Zikes (2018): "Funding constraints and liquidity in two-tiered OTC markets," *Journal of Financial Markets*. 7

Bessembinder, H., and P. J. Seguin (1993): "Price Volatility, Trading Volume, and Market Depth: Evidence from Futures Markets," *The Journal of Financial and Quantitative Analysis*, 28(1), 21–39. 4.2.2, **??**

Bonacich, P., and P. Lloyd (2001): "Eigenvector-like measures of centrality for asymmetric relations," *Social Networks*, 23(3), 191 – 201. 11

Boulatov, A., T. Hendershott, and D. Livdan (2013): "Informed Trading and Portfolio Returns," *Review of Economic Studies*, 80(1), 35–72. 1

Brancaccio, G., D. Li, and N. Schurhoff (2017): "Learning by Trading: The Case of the U.S. Market for Municipal Bonds," mimeo. 1, 4

Brandt, M. W., and K. A. Kavajecz (2004): "Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve," *Journal of Finance*, 59(6), 2623–2654. 1, 6.2.2

Chakravarty, S., H. Gulen, and S. Mayhew (2004): "Informed Trading in Stock and Option Markets," *Journal of Finance*, 59(3), 1235–1258. 1

Choi, J. H., K. Larsen, and D. J. Seppi (2019): "Information and trading targets in a dynamic market equilibrium," *Journal of Financial Economics*, 132(3), 22–49. 5.1

COHEN, L., C. MALLOY, AND L. POMORSKI (2012): "Decoding Inside Information," *The Journal of Finance*, 67(3), 1009–1043. 1

COLLIN-DUFRESNE, P., AND V. FOS (2015): "Do Prices Reveal the Presence of Informed Trading?," *Journal of Finance*, 70(4), 1555–1582. 1

COLLIN-DUFRESNE, P., P. HOFFMANN, AND S. VOGEL (2020): "Informed Traders and Dealers in the FX Forward Market," Discussion paper. 5

CZECH, R., S. HUANG, D. LOU, AND T. WANG (2021): "Informed Trading in Government Bond Markets," *Journal of Financial Economics*, Forthcoming. 4.2.1, 17

CZECH, R., AND G. PINTER (2020): "Informed trading and the dynamics of client-dealer connections in corporate bond markets," Bank of England working papers 895, Bank of England. 8

DMO (2011): "A guide to the roles of the DMO and Primary Dealers in the UK government bond market," Discussion paper, Debt Management Office. 2.1

DUARTE, J., F. A. LONGSTAFF, AND F. YU (2006): "Risk and Return in Fixed-Income Arbitrage: Nickels in Front of a Steamroller?," *The Review of Financial Studies*, 20(3), 769. 8

DUARTE, J., AND L. YOUNG (2009): "Why is PIN priced?," *Journal of Financial Economics*, 91(2), 119–138. 1

DUFFIE, D., S. MALAMUD, AND G. MANSO (2009): "Information Percolation With Equilibrium Search Dynamics," *Econometrica*, 77(5), 1513–1574. 4

EASLEY, D., N. M. KIEFER, M. O'HARA, AND J. B. PAPERMAN (1996): "Liquidity, Information, and Infrequently Traded Stocks," *The Journal of Finance*, 51(4), 1405–1436. 1

EASLEY, D., AND M. O'HARA (1987): "Price, trade size, and information in securities markets," *Journal of Financial Economics*, 19(1), 69 – 90. 4.1.1

EGUREN-MARTIN, F., AND N. MCLAREN (2015): "How much do UK market interest rates respond to macroeconomic data news?," *Bank of England Quarterly Bulletin*, 55(3), 259–272. 26

EVANS, M., AND R. LYONS (2002): "Order Flow and Exchange Rate Dynamics," *Journal of Political Economy*, 110(1), 170–180. 6.2.2

FLEMING, M. J., AND E. M. REMOLONA (1999): "Price Formation and Liquidity in the U.S. Treasury Market: The Response to Public Information," *Journal of Finance*, 54(5), 1901–1915. 1, 6.2.2

GABRIELI, S., AND C.-P. GEORG (2014): "A network view on interbank market freezes," Discussion paper. 1

GARLEANU, N., AND L. H. PEDERSEN (2013): "Dynamic Trading with Predictable Returns and Transaction Costs," *Journal of Finance*, 68(6), 2309–2340. 8

Glosten, L. R., and P. R. Milgrom (1985): "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14(1), 71–100. 2, 9, A.1

Golosov, M., G. Lorenzoni, and A. Tsyvinski (2014): "Decentralized Trading with Private Information," *Econometrica*, 82(3), 1055–1091. 4

Green, T. C. (2004): "Economic News and the Impact of Trading on Bond Prices," *Journal of Finance*, 59(3), 1201–1234. 1

Hagstromer, B., and A. Menkveld (2019): "Information Revelation in Decentralized Markets," *The Journal of Finance*, (Forthcoming). 5

Hendershott, T., D. Li, D. Livdan, and S. Norman (2017): "Relationship Trading in OTC Markets," CEPR Discussion Papers 12472, C.E.P.R. Discussion Papers. 3.3.2

Hendershott, T., D. Livdan, and N. Schurhoff (2015): "Are institutions informed about news?," *Journal of Financial Economics*, 117(2), 249–287. 1

Hollifield, B., A. Neklyudov, and C. Spatt (2017): "Bid-Ask Spreads, Trading Networks, and the Pricing of Securitizations," *Review of Financial Studies*, 30(9), 3048–3085. 1, 3.3.2

Hortacsu, A., and J. Kastl (2012): "Valuing Dealers' Informational Advantage: A Study of Canadian Treasury Auctions," *Econometrica*, 80(6), 2511–2542. 1

Jankowitsch, R., A. Nashikkar, and M. G. Subrahmanyam (2011): "Price dispersion in OTC markets: A new measure of liquidity," *Journal of Banking and Finance*, 35(2), 343–357. 7, 9

Johnson, T. L., and E. C. So (2018): "A Simple Multimarket Measure of Information Asymmetry," *Management Science*, 64(3), 1055–1080. 1

Kacperczyk, M., and E. Pagnotta (2019): "Chasing Private Information," *Review of Financial Studies*, Forthcoming. 3

Kandel, E., and N. D. Pearson (1995): "Differential Interpretation of Public Signals and Trade in Speculative Markets," *Journal of Political Economy*, 103(4), 831–872. 27

Karpoff, J. (1987): "The Relation between Price Changes and Trading Volume: A Survey," *Journal of Financial and Quantitative Analysis*, 22(01), 109–126. 4.2.2, **??**

Kondor, P., and G. Pinter (2019): "Clients' Connections: Measuring the Role of Private Information in Decentralised Markets," Cepr discussion paper. 11, 24, 27, 29, 7

Kyle, A. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53(6), 1315–35. 1, 2.2, 8, 6.1, 8

Maggio, M. D., F. Franzoni, A. Kermani, and C. Sommavilla (2019): "The relevance of broker networks for information diffusion in the stock market," *Journal of Financial Economics*. 1, 3.3.1, 4.1.1, 4.2.1, 5.2, 21, 24

MAGGIO, M. D., A. KERMANI, AND Z. SONG (2017): "The value of trading relations in turbulent times," *Journal of Financial Economics*, 124(2), 266 – 284. 11, 3.3.2, 6.2.2

MASCIO, R. D., A. LINES, AND N. Y. NAIK (2017): "Alpha Decay," mimeo. 8

MENKVELD, A., A. SARKAR, AND M. VAN DER WEL (2012): "Customer Order Flow, Intermediaries, and Discovery of the Equilibrium Risk-Free Rate," *Journal of Financial and Quantitative Analysis*, 47(04), 821–849. 6.2.2

MERRICK, J., N. Y. NAIK, AND P. K. YADAV (2005): "Strategic trading behavior and price distortion in a manipulated market: anatomy of a squeeze," *Journal of Financial Economics*, 77(1), 171–218. 4.1.1

MEULBROEK, L. K. (1992): "An Empirical Analysis of Illegal Insider Trading," *Journal of Finance*, 47(5), 1661–99. 3

O'HARA, M. (2015): "High frequency market microstructure," *Journal of Financial Economics*, 116(2), 257–270. 5.1

PASQUARIELLO, P., AND C. VEGA (2007): "Informed and Strategic Order Flow in the Bond Markets," *Review of Financial Studies*, 20(6), 1975–2019. 1, 27

ROLL, R., E. SCHWARTZ, AND A. SUBRAHMANYAM (2010): "O/S: The relative trading activity in options and stock," *Journal of Financial Economics*, 96(1), 1–17. 1

SWANSON, E. T., AND J. C. WILLIAMS (2014a): "Measuring the Effect of the Zero Lower Bound on Medium- and Longer-Term Interest Rates," *American Economic Review*, 104(10), 3154–85. 6.2.1, 6

——— (2014b): "Measuring the effect of the zero lower bound on yields and exchange rates in the U.K. and Germany," *Journal of International Economics*, 92, S2 – S21, 36th Annual NBER International Seminar on Macroeconomics. 6.2.1, 6

TAUCHEN, G., AND M. PITTS (1983): "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, 51(2), 485–505. 4.2.2

VALSETH, S. (2013): "Price discovery in government bond markets," *Journal of Financial Markets*, 16(1), 127–151. 1

Internet Appendix

A Theory

A.1 Connections and Private information

Consider many periods of trading indexed by t and informed clients indexed by i. The end-of-period liquidation value of the asset is $V_t = \mu_t + \rho V_{t-1} + \varepsilon_t$ where μ_t is a known drift term, while ε_t is either 0 or 1 with equal probability, and independent across periods. The innovation, ε_t, becomes public information at the end of the period and all positions are liquidated at the closing price V_t.

Note that the process of V_t can incorporate different interpretations.

1. We can think of V_t as the fundamental value of the asset, perhaps specifying the process as a simple random walk ($\rho = 1, \mu_t = 0$). Then, ε_t is a fundamental innovation, and a signal on ε_t is private fundamental information..

2. Alternatively, we might assume that the fundamental value of the asset is known, but the end-of-period liquidation value, V_t is affected by demand or supply pressure. Perhaps some large liquidity traders tend to submit large orders at the end of the period and dealers need a premium for holding those risky positions betwen periods. For instance, if we rewrite the process as

$$V_t - V_{t-1} = (\rho - 1)\left(V_{t-1} - \left(\frac{-\mu_t}{\rho - 1}\right)\right) + \varepsilon_t$$

choosing $\rho - 1 > 0$, and $\mu_t < 0$, gives a V_t which tends to revert to $\frac{-\mu_t}{1-\rho}$, the fundamental value. Then, ε_t is innovation in demand pressure and a signal on ε_t is private information on future order flow.

In any case, we assume that during the period informed clients with a signal on ε_t, noise traders, and market makers trade determining the mid-period transaction prices P_{it}. The objective of clients is to maximize their trading profit $x_{it}(V_t - P_{it})$ each period by choosing to buy or sell one unit at the prevailing ask or bid prices respectively, i.e., choosing $x_{it} = \{1, -1\}$. The trading protocol is a modified version of Glosten and Milgrom (1985). Clients and noise traders seek bid and ask quotes from one or more risk neutral, market maker (MM) in each period. Just as in Glosten and Milgrom (1985), we assume that MMs are competitive, hence, their quotes are determined by a zero expected profit condition. Sampling quotes from more market makers might be costly.

To convey the intuition we focus on the simplest possible case. We consider four potential MMs, $m = R^i, N^i$ serving two clients $i = 1, 2$. Client i is assigned to MMs R^i, N^i (for regular and potential new comer). We assume that client i's signal, $s_t = B, S$ (for buy and sell) is informative:

$$\Pr\left(\varepsilon_t = 1 | s_t = B\right) = \frac{1}{2} + \Delta_{ti}$$

where $\Delta_{ti} > 0$ might vary across clients and time. Δ_{ti} is observable to clients and has the two point support of $\{\Delta_L, \Delta_H\}$ with $\Delta_H > \Delta_L$. Before a client observes her signal, he commits to a quote request function $\rho\left(\Delta_{ti}\right) : \{\Delta_L, \Delta_H\} \to \{R^i, (R^i, N^i)\}$ which describes the states when dealer i requests quotes from one or both dealers. The cost of the earlier is normalized to 0, while requesting two sets of quotes costs c. We think of c as a non-observable, non-pecuniary cost. It is a reduced form treatment to capture a search cost, or the reputation cost coming from the unmodelled future punishment from the dealer who provided a quote but did not received the trade. Importantly, client i is present in the market at t with only probability $(1 - \alpha)$. Even if the client requests two sets of bid and ask prices, eventually, she can trade only with one of the dealers. After observing the bids and asks she decides whether to buy or sell at one of those prices. Whenever client i is not requesting quotes from a given MM assigned to her, regardless it is by choice or because she is not present at that period, a noise trader requests quotes instead and buys or sells with equal probability. Therefore, MMs receive exactly one request for quotes in any given period, but might or might not trade. We assume that MMs observe Δ_{it} of their assigned client, but they do not observe whether she is present at the given period. That is, they cannot observe whether a quote request comes from client i or a noise trader. After trading, positions are liquidated at the realised true value V_t.

The following Proposition characterizes the equilibrium in this stage game. The client requests two quotes if and only if her information is sufficiently precise. In that case, she gets identical quotes and trades with each of the MMs with equal probability. The intuition relies on a simple idea. For fixed parameters, when the client i asks a quote only from R^i, R^i trades with an informed dealer with probability $(1 - \alpha)$, while N^i trades with only noise traders. Therefore, the bid-ask spread provided by R^i is relatively wide, while the bid ask spread provided by N^i is zero. When i asks a quote from both MMs and trades with only one of them randomly, R^i faces with an informed dealer with a probability $(1-\alpha)\,\pi$ only, where π is the mixing probability. Therefore, she will give better quotes to the client. In equilibrium, π has to adjust in a way that N^i wants to give identical quotes to R^i. Therefore, mixing between two dealers helps the client to hide his information better implying more favourable transaction prices. At the same time asking for two sets of quotes is costly. Hence, the client chooses to do so if and only if Δ_{ti} is sufficiently high.

Proposition 1 *Let*

$$\bar{\Delta} = \frac{(\alpha+1)}{\alpha\,(1-\alpha)}c$$

be within the support of Δ_{ti}.

1. *If* $\Delta_{ti} < \bar{\Delta}$, *the informed trader* i *trades only with* R^i *and the equilibrium bid ask quotes are*

$$\begin{aligned}
A_t^{R^i}\left(\Delta_{ti} < \bar{\Delta}\right) &= \mu_t + \rho V_{t-1} + \frac{1}{2} + \Delta_{ti}\,(1-\alpha) \\
B_t^{R^i}\left(\Delta_{ti} < \bar{\Delta}\right) &= \mu_t + \rho V_{t-1} + \frac{1}{2} - \Delta_{ti}\,(1-\alpha) \\
A_t^{N^i}\left(\Delta_{ti} < \bar{\Delta}\right) &= B^N\left(\Delta_{ti} < \bar{\Delta}\right) = \mu_t + \rho V_{t-1} + \frac{1}{2}.
\end{aligned}$$

2. *If* $\Delta_{ti} > \bar{\Delta}$, *the informed trader* i *seeks quotes from both MM and trades with each with equal probability. The equilibrium bid ask quotes are*

$$\begin{aligned}
A_t^{R^i}\left(\Delta_{ti} > \bar{\Delta}\right) &= A_t^{N^i}\left(\Delta_{ti} > \bar{\Delta}\right) = \mu_t + \rho V_{t-1} + \frac{1}{2} + \Delta_{ti}\frac{(1-\alpha)}{1+\alpha} \\
B_t^{iR}\left(\Delta_{ti} > \bar{\Delta}\right) &= B_t^{N^i}\left(\Delta_{ti} > \bar{\Delta}\right) = \mu_t + \rho V_{t-1} + \frac{1}{2} - \Delta_{ti}\frac{(1-\alpha)}{1+\alpha}.
\end{aligned}$$

Proof. The quotes are derived by Bayes Rule. For example, the ask price provided by R^i to i when the MM understands that i trades with him with probability $\pi = 1$ i

$$E\left(V_{t-1} + \varepsilon_t \middle|\ \text{observing a buy in } t, \pi = 1\right) \quad = \quad V_{t-1} \quad + \quad \frac{1}{2} \quad + \quad \Delta_{ti}\,(1-\alpha)\,.$$

When the trader observes quotes from both MMs, in equilibrium the two MMs has to submit the same quotes given the mixed strategy of acceptance from the trader. For this, the informed trader has to mix with probability half. In this case the ask price is given by

$$E\left(V_{t-1}+\varepsilon_t|\text{ observing a buy in } t, \pi=\frac{1}{2}\right) \quad = \quad V_{t-1} \quad + \quad \frac{1}{2} \quad + \quad \Delta_{ti}\frac{(1-\alpha)}{1+\alpha}.$$

For a fixed Δ_{ti}, the expected benefit of transacting at the more favourable prices implied by mixing, $\pi = \frac{1}{2}$, is

$$\Sigma_{V_t=0,1}\Pr\left(\varepsilon_t|s=H,\Delta_{it}\right)\left(\left(\varepsilon_t-\left(\frac{1}{2}+\Delta_{it}\frac{(1-\alpha)}{1+\alpha}\right)\right)-\left(\varepsilon_t-\left(\frac{1}{2}-\Delta_{it}\left(1-\alpha\right)\right)\right)\right)=$$
$$=\Delta_{it}\alpha\frac{1-\alpha}{\alpha+1},$$

which is increasing in Δ_{ti}. The indifference condition given cost c determines $\bar{\Delta}$. ■

To picture the implied time-series and cross-sectional evaluation of prices and trades, we assume that for each client i, this subgame is repeated in many time-periods. These games are independent from each other because all random variables are redrawn in each period and because the MMs are disjunct across the two clients. Suppose that $\Delta_H > \bar{\Delta} > \Delta_L$. The correlation structure across time and clients in Δ_{ti} can be arbitrary.

To see the implications, it is useful to compare implied profits of clients in the two states. (Because of symmetry, it is sufficient to calculate the implied expected profit conditional on a signal $s = B$.)

$$\Pi\left(\Delta_H\right)=\Sigma_{V_t=0,1}\Pr\left(\varepsilon_t|s=B,\Delta_H\right)\left(\varepsilon_t-\left(\frac{1}{2}+\Delta_H\frac{(1-\alpha)}{1+\alpha}\right)\right)=2\alpha\frac{\Delta_H}{\alpha+1}$$

$$\Pi\left(\Delta_L\right)=\Sigma_{V_t=0,1}\Pr\left(\varepsilon_t|s=B,\Delta_L\right)\left(\varepsilon_t-\left(\frac{1}{2}-\Delta_L\left(1-\alpha\right)\right)\right)=\alpha\Delta_L.$$

Clearly,

$$\Pi\left(\Delta_H\right)-\Pi\left(\Delta_L\right)=\alpha\left(\frac{1-\alpha}{\alpha+1}\Delta_H+\left(\Delta_H-\Delta_L\right)\right)>0$$

for any parameter values. Note also, that $\Pi\left(\Delta_H\right)-\Pi\left(\Delta_L\right)$ is increasing in Δ_H and in $\left(\Delta_H-\Delta_L\right)$.

Consider an interval with multiple, say, D, periods. Within this interval, in each period when $\Delta_{it} = \Delta_L$, client i trades with only R^i with probability 1. In each period when $\Delta_{it} = \Delta_H$, with probability $\frac{1}{2}$ she trades with N^i. Hence, if ξ_{iD} is a counting process for the periods with $\Delta_{it} = \Delta_H$ within D, then the expected number of connections of

i within interval D is $1 + \frac{\xi_{iD}}{2}$, an increasing function of ξ_{iD}. That is, the number of connections within an interval is a proxy for the number of periods with an interval where the information precision of the client is Δ_H. These observations give the Hypotheses 1 and 3.

For later use, note that aggregating connections over clients in a given interval D is a proxy for the total private information present in the market.

Next, observe that the difference in the trading profit across $\Delta_{it} = \Delta_H$ and $\Delta_{it} = \Delta_L$ comes from two sources: the change in probability that the client trades the right direction, and the change in the transaction price. On one hand, the probability that client i is trading at the right direction, buying before the price moves up and selling before the price moves down, is $\frac{1}{2} + \Delta_{ti}$, an increasing function of Δ_{ti}. This gives Hypothesis2, which we refer to in the text as the anticipation component.

On the other hand, the transaction price can be more or less favorable when $\Delta_{it} = \Delta_H$. In particular, if and only if

$$\frac{\Delta_H}{\Delta_L} < 1 + \alpha,$$

a client buys (sells) at a lower (higher) price when she is more informed. This relationship drives the sign of the relationship between connections and the transaction component defined in the text. The reason for the ambiguous result is that the client's ability to hide its higher quality signal better by trading with more dealers is limited. When

$$1 + \alpha < \frac{\Delta_H}{\Delta_L}, \tag{A.1}$$

then a client with higher information precision mixing between the two dealers gets a less favourable price then the client with the lower information precision who trades with one dealer only. Of course, the high precision client's price is still more favourable than if she were to trade with her regular dealer only. Otherwise, there would be profitable deviation in equilibrium.

Finally, we turn to price discovery. For this, we calculate the expected average transaction price when the innovation is $\varepsilon_t = 1$ in each possible scenarios with respect to the type of traders arriving. First, when both traders have high precision signals, both request quotes from both of their assigned MMs, but they trade only at one of those quotes. The average transaction price is

$$E\left(\frac{P_{1t}+P_{2t}}{2}|\varepsilon_t=1,\Delta_H,\Delta_H\right)=$$
$$=V_{t-1}+\left(\frac{1}{2}+\Delta_H\right)\left(\frac{1}{2}+\Delta_H\right)\left(\left(\frac{1}{2}+\Delta_H\frac{(1-\alpha)}{1+\alpha}\right)\right)+$$
$$\left(\frac{1}{2}-\Delta_H\right)\left(\frac{1}{2}-\Delta_H\right)\left(\left(\frac{1}{2}-\Delta_H\frac{(1-\alpha)}{1+\alpha}\right)\right)+2\left(\frac{1}{2}+\Delta_H\right)\left(\frac{1}{2}-\Delta_H\right)\left(\frac{1}{2}\right)$$
$$=V_{t-1}+\frac{1}{2}+\frac{2\Delta_H^2\left(1-\alpha\right)}{\alpha+1}.$$

If both clients have low precision signals, each requests quotes only from R^i and N^i trades with liquidity traders at price $V_{t-1}+\frac{1}{2}$. That is, the average transaction price is:

$$E\left(\frac{P_{1t}\left(R^1\right)+P_{2t}\left(R^2\right)+P_{1t}\left(N^1\right)+P_{2t}\left(N^2\right)}{4}|\varepsilon_t=1,\Delta_L,\Delta_L\right)=$$
$$=V_{t-1}+\left(\frac{1}{2}+\Delta_L\right)\left(\frac{1}{2}+\Delta_L\right)\left(\frac{1}{2}\left(\frac{1}{2}+\Delta_L\left(1-\alpha\right)\right)+\frac{1}{2}\frac{1}{2}\right)$$
$$+\left(\frac{1}{2}-\Delta_L\right)\left(\frac{1}{2}-\Delta_L\right)\left(\frac{1}{2}\left(\frac{1}{2}-\Delta_L\left(1-\alpha\right)\right)+\frac{1}{2}\frac{1}{2}\right)$$
$$+2\left(\frac{1}{2}+\Delta_L\right)\left(\frac{1}{2}-\Delta_L\right)\left(\frac{1}{2}\right)$$
$$=V_{t-1}+\frac{1}{2}+\left(1-\alpha\right)\Delta_L^2.$$

Following similar calculations, when the first client has low precision, while the second one has high precision, the average price is:

$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{1t}}{3}|\varepsilon_t=1,\Delta_L,\Delta_H\right)=V_{t-1}+\frac{1}{2}+\frac{2}{3}\left(1-\alpha\right)\left(\frac{\Delta_H^2}{\alpha+1}+\Delta_L^2\right).$$

Also, we get the expected average price for the case when at least one of the clients is not present (which we denote by $\oslash$), hence replaced by two noise traders:

$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{2t}\left(R^2\right)+P_{2t}\left(N^2\right)}{4}|\varepsilon_t=1,\oslash,\oslash\right)=V_{t-1}+\frac{1}{2}$$
$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{2t}\left(R^2\right)+P_{2t}\left(N^2\right)}{4}|\varepsilon_t=1,\oslash,\Delta_L\right)=V_{t-1}+\frac{1}{2}+\frac{1}{2}\Delta_L^2\left(1-\alpha\right)$$
$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{2t}}{3}|\varepsilon_t=1,\oslash,\Delta_H\right)=V_{t-1}+\frac{1}{2}+\frac{1}{2}\Delta_H^2\frac{1-\alpha}{\alpha+1}.$$

It is easy to check that

$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{2t}\left(R^2\right)+P_{2t}\left(N^2\right)}{4}|\varepsilon_t=1,\oslash,\Delta_L\right)<$$

$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{2t}}{3}|\varepsilon_t=1,\oslash,\Delta_H\right),E\left(\frac{P_{1t}\left(R^1\right)+P_{2t}\left(R^2\right)+P_{1t}\left(N^1\right)+P_{2t}\left(N^2\right)}{4}|\varepsilon_t=1,\Delta_L,\Delta_L\right)$$

$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{2t}}{3}|\varepsilon_t=1,\oslash,\Delta_H\right),E\left(\frac{P_{1t}\left(R^1\right)+P_{2t}\left(R^2\right)+P_{1t}\left(N^1\right)+P_{2t}\left(N^2\right)}{4}|\varepsilon_t=1,\Delta_L,\Delta_L\right)<$$
$$E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{1t}}{3}|\varepsilon_t=1,\Delta_L,\Delta_H\right)<E\left(\frac{P_{1t}+P_{2t}}{2}|\varepsilon_t=1,\Delta_H,\Delta_H\right).$$

Recall that aggregate connections $\Sigma_i\xi_{iD}$ is increasing in the fraction of high precision clients in the market. Therefore, with the caveat that the comparison of $E\left(\frac{P_{1t}\left(R^1\right)+P_{1t}\left(N^1\right)+P_{2t}}{3}|\varepsilon_t=1,\oslash,\right.$ and $E\left(\frac{P_{1t}\left(R^1\right)+P_{2t}\left(R^2\right)+P_{1t}\left(N^1\right)+P_{2t}\left(N^2\right)}{4}|\varepsilon_t=1,\Delta_L,\Delta_L\right)$ depends on the parameters, we form Hypothesis 4.

A.2 Connections and Persistent Demand Shocks

Consider 3 days of trading indexed by $t=1,2,3$. There are two dealers in each period are present in the market, standing ready to trade and indexed by $i=1,2$. Also, on each of dates $t=1,2$ a risk-neutral liquidity trader arrives with a fixed negative demand of 1 or 2 units of the asset, $d_t=\{1,2\}$. In particular, the first trader has to sell one or two units with equal probability, while the second trader has to sell the same number of units as the first with probability $\pi>\frac{1}{2}$. We summarize the resulting distribution as follows:

(d_1,d_2)	-1	-2
1	$\frac{1}{2}\pi$	$\frac{1}{2}(1-\pi)$
2	$\frac{1}{2}(1-\pi)$	$\frac{1}{2}\pi$

Liquidity traders cannot choose the amount or direction of their total demand. However, they can choose whether to sell all their required units from one of the dealers, or share their trades across the two dealers. Asking for quotes from two dealers has an extra cost. The idea is that each trader has a regular dealer, while they do not have a relationship with the other one. Therefore, obtaining a quote from only one is free, but from both requires a non-pecuniary cost of effort of c. At date $t=3$, the fundamental value of the asset V is realized. V is 0 or 1 with equal probability. As a result, at that point any dealer holding any position off-loads that asset at the true value V.

The trading protocol is as follows. A trader arriving in period t with demand d_t can contact her own dealer for free and signal that she would like to trade $a_{t1} = d_1$. The dealer responds with a firm price quote p_{it}. The client can accept, and trade accordingly, or reject. If she rejects she pays the cost c and gives a take-it-or-leave it offers to both dealers in a form of a quantity price pair (a_{ti}, p_{ti}) where $a_{t1} + a_{t2} = d_t$.

Each dealer is risk-averse with a standard, concave utility function over final wealth $u\left(W_{3i}\right)$

$$W_{3i}(a_{1i}, a_{2i}, p_{1i}, p_{2i}) \equiv e_M + \left(e_i + a_{1i} + a_{2i}\right) V - a_{1i}p_{1i} - a_{2i}p_{2i},$$

where e_M is a monetary endowment, while e_i is existing inventory of the asset.

Dealers' equilibrium holdings $a^*_{1i}(d_1), a^*_{2i}(d_1, d_2), \in [0, 2]$ depend on the state d_1, d_2 and whether clients share their demand across the two dealers. As $u(\cdot)$ is concave, the dealer asks larger premium for holding larger positions, implying that the equilibrium prices are decreasing in equilibrium quantities a^*_{ti}.

We are looking for parameters, e_i, e_j, π, c that in equilibrium the trader shares her demand if and only if $d_t = 2$. That is, we guess and verify that in any period with $d_t = 1$, dealer 1 fully absorbs the demand shock of the arriving client. She will do so at the lowest price which still deters the entry of dealer 2. In contrast, in any period $d_t = 2$, the two dealers share the demand shock at an allocation and price which leads to zero surplus to both of them by the logic of Bertrand competition. Now we derive the equilibrium conditions.

Whenever $d_t = 1$ only the first dealer trades,

$$a^*_{11}(1) = a^*_{21}(d_1, 1) = 1, a^*_{12}(1) = a^*_{22}(d_1, 1) = 0.$$

It is easy to see that the corresponding prices are determined by the minimal prices at which the client is deterred from searching for a quote from the second dealer. That is, $p^*_{11}(1) = p^{\min}_1$, $p^*_{21}(d_1, 1) = p^{\min}_{2i}(d_1)$ where

$$p^{\min}_1 \equiv \{p_{12} : E_{V,d_2}\left[u'\left(e_M + \left(e_2 + a^*_{22}(1, d_2)\right) V - a^*_{22}(1, d_2)p^*_{22}(1, d_2)\right)\left(V - p_{12}\right)\right] = 0\} + c,$$

and

$$p^{\min}_2(d_1) \equiv \{p_{22} : E_V\left(u'\left(\left(e_M + e_2 + a^*_{12}(d_1)\right) V - a^*_{12}(d_1)p^*_{12}(d_1)\right)\left(V - p_{22}\right)\right) = 0\} + c.$$

Note that if dealer 1 were to quote higher prices than $p^{\min}_t$, dealer 2 would be willing to buy a small quantity at a price which would be profitalbe to seek out for the client even taking into account the cost c.

A further equilibrium condition is that the resulting prices $p_{11}^{*}(1), p_{21}^{*}(d_1,1)$ are sufficiently low that dealer 1 is willing to participate. That is,

$$E_{V,d_2}[W_{3i}(a_{11}^{*}(1),a_{21}^{*}(1,d2),p_{11}^{*}(1),p_{2i}^{*}(1,d_2)) \geq E_{V,d_2}[W_{3i}(0,a_{21}^{*}(1,d_2),p_{11}^{*}(1),p_{2i}^{*}(1,d_2)) \tag{A.2}$$

$$E_{V}[W_{3i}(a_{11}^{*}(d_1),a_{21}^{*}(d_1,1),p_{11}^{*}(d_1),p_{2i}^{*}(d_1,1)) \geq E_{V}[W_{3i}(a_{11}^{*}(d_1),0,p_{11}^{*}(d_1),p_{2i}^{*}(d_1,1)) \tag{A.3}$$

If a client decides to share her demand shock across the two dealers, the two dealers are effectively in a Bertrand competition. Starting with the second period, demand $a_{21}^{*}\left(d_1,2\right)$=2-$a_{22}^{*}\left(d_1,2\right)$ and price $p_{21}^{*}\left(d_1,2\right) = p_{21}^{*}\left(d_1,2\right)$= p_2 has to solve

$$\begin{aligned}
&\max_{a_{21}\in[0,2],p_2} p_2 \\
&E_V\left(\left(u\left(e_M+\left(e_1+a_{11}^{*}\left(d_1\right)+a_{21}\right)V-a_{11}^{*}\left(d_1\right)p_{11}^{*}\left(d_1\right)-a_{2i}p_2\right)\right)\right)\geq \\
&\quad E\left(\left(u\left(e_M+\left(e_i+a_{11}^{*}\left(d_1\right)\right)V-a_{11}^{*}\left(d_1\right)p_{11}^{*}\left(d_1\right)\right)\right)\right) \\
&E_V\left(\left(u\left(e_M+\left(e_2+\left(2-a_{11}^{*}\left(d_1\right)\right)+\left(2-a_{21}\right)\right)V-\left(2-a_{11}^{*}\left(d_1\right)\right)p_{11}^{*}\left(d_1\right)-\left(2-a_{21}\right)p_2\right)\right)\right)\geq \\
&\quad E\left(\left(u\left(e_M+\left(e_2+\left(2-a_{11}^{*}\left(d_1\right)\right)+\right)V-\left(2-a_{11}^{*}\left(d_1\right)\right)p_{11}^{*}\left(d_1\right)\right)\right)\right).
\end{aligned} \tag{A.4}$$

Similarly, in the first period, taking the solution of problem (A.4) as given, the price $p_{11}^{*}\left(d_1\right) = p_{12}^{*}\left(d_1\right) = p_1$ and the quantities $a_{11}^{*}\left(d_1\right) = 2 - a_{12}^{*}\left(d_1\right)$, have to solve the following problem:

$$\begin{aligned}
&\max_{a_{11}\in[0,2],p_1} p_1 \\
&E_{V,d_2}\left(\left(u\left(e_M+\left(e_1+a_{11}+a_{21}^{*}\left(d_1,d_2\right)\right)V-a_{11}p_1-a_{21}^{*}\left(d_1,d_2\right)p_{21}^{*}\left(d_1,d_2\right)\right)\right)\right)\geq \\
&\quad E_{V,d_2}\left(\left(u\left(e_M+\left(e_i+a_{21}^{*}\left(d_1,d_2\right)\right)V-a_{21}^{*}\left(d_1,2\right)p_{22}^{*}\left(d_1,d_2\right)\right)\right)\right) \\
&E_{V,d_2}\left(\left(u\left(e_M+\left(e_2+\left(2-a_{11}\right)+\left(2-a_{21}^{*}\left(d_1,d_2\right)\right)\right)V-\left(2-a_{11}\right)p_1-\left(2-a_{21}^{*}\left(d_1,2\right)\right)p_{22}^{*}\left(d_1,d_2\right)\right)\right)\right)\geq \\
&\quad E_{V,d_2}\left(\left(u\left(e_M+\left(e_2+\left(2-a_{21}^{*}\left(d_1,d_2\right)\right)\right)V-\left(2-a_{21}^{*}\left(d_1,d_2\right)\right)p_{22}^{*}\left(d_1,d_2\right)\right)\right)\right).
\end{aligned} \tag{A.5}$$

The inequalities in each of these problems are the participation constraints of the two dealers. In an interior solution both participation constraints will bind.

Discussion If π is sufficiently large, the realization of $d_1$is the same as d_2 most of the time. When both demand shocks are low, clients trade only with their own dealer, that

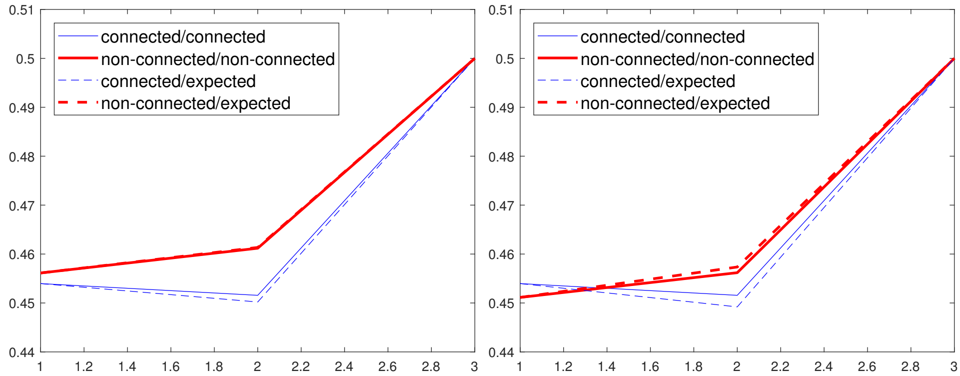


Figure 7: The price and the expected price in each period, for small (left panel) and large (right panel) cost of contacting a second dealer. In each panel, we show the price path when both demand shock is small ($d_t = 1$, thick red solid) hence both clients have a single connection, and when both demand shocks are large ($d_t = 2$, thin blue solid) hence both clients have two connections. To calculate the expected profit for a connected and unconnected client 1, we also show the expected prices in period 1 in each of the cases (dashed). Parameters are $u(\cdot) = \ln(\cdot)$,$e_M = 10, e_1 = 0, e_2 = 1, \pi = 0.8, c = 0.015$ and $c = 0.02$ in the left and right panels, respectively.

is, their connection is low. When both demand shocks are high, connections are high. To see whether higher connections predict higher short-term returns, we have to compare the price pattern in these two scenarios. The two panels of Figure show the result for two sets of parameter values.

The left panel shows equilibrium prices when c is small, while the right panel shows the same prices when c is larger. The thick, red curve corresponds to the case when clients are not connected, because each experience a small demand shock. The thin, blue curves correspond the case when both clients are connected as both experienced a large demand shock. (The dashed curves show the period 2 expected price conditional on the first demand shock.)

As clients are selling, each client makes a short-term gain (loss) if the price decreases (increases) in the period after their entry. Clearly, client 1 makes a gain when connected, and a loss when unconnected, under both of the parameter combinations. This is consistent with the idea that connection is associated with higher profit. However, client 2 makes a loss in the short -term regardless of the parameter values or her connections. In particular, in each of the cases represented in the figure, her loss is larger when connected.

Our empirical design corresponds comparing the short-term average return of client 1 and client 2 when both our connected to the case when both our unconnected. This comparison depends on the parameters. For instance, under the parameters corresponding to the right panel, the average short-term return is higher when both clients are unconnected (-0.044 vs -0.446) while on the right panel the opposite is the case (-0.05 vs -0.044).

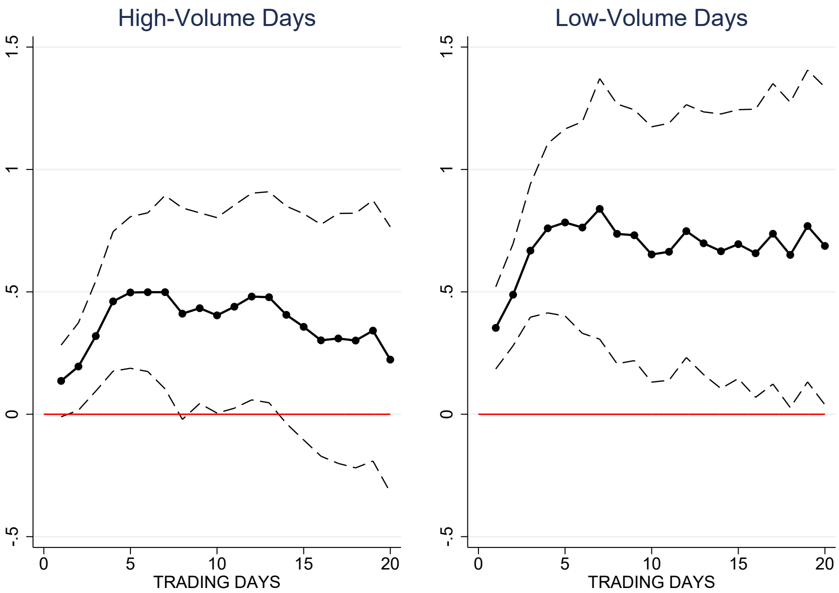
Note , that the average profit is negative in each case, consistently with the idea that dealers are willing to absorb these shocks for a risk-premium, and clients trade for liquidity reasons.

We also emphasize that the total amount of trades of a given client across all the dealers is large exactly when their number of connections are large. That is, in this economy connections are caused by larger demand shocks. As we argue in the main text, this is a testable implication which we can reject in the data.

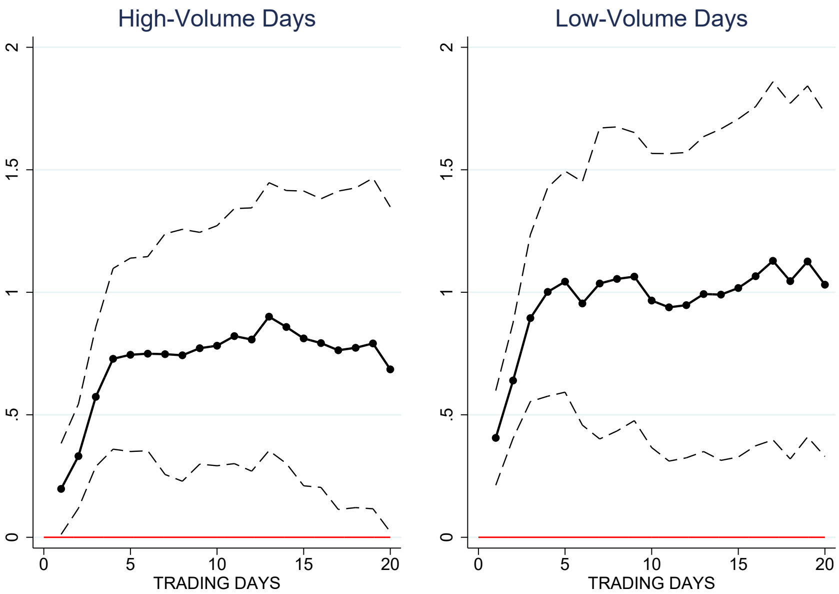
B Additional Tables and Figures

Figure 8: Connections and Performance over 1-20 Day Horizons: High vs Low Volume Days

(a) All Clients

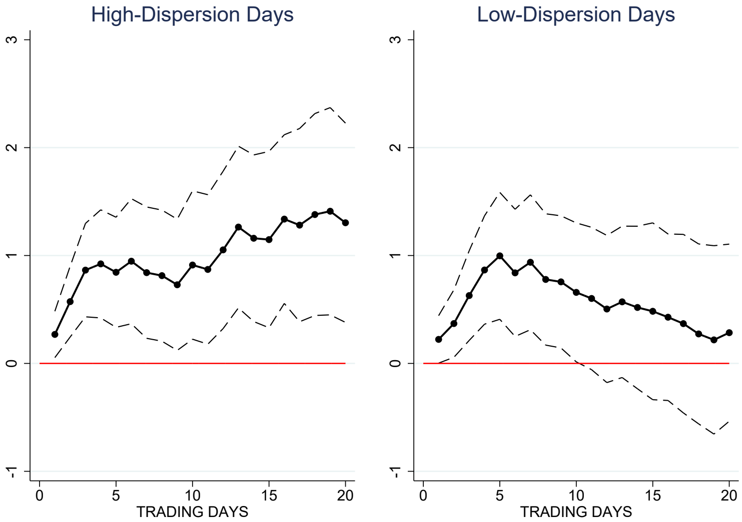


(b) Sophisticated Clients



Notes: this figure plots the estimated β coefficients from variant of regression 4.1 up to 20-day horizon ($T = 20$), where we interact connections with an indicator variable for days above (left panel) and below (right panel) the client's average daily trading volume. Panel 8a presents the results for all clients, and panel 8b restricts the sample to sophisticated investors. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Transactions"). The shaded area denotes the 90% confidence band associated with the estimated β coefficients, It is based on robust standard errors, using two-way clustering at the day and the client level.

Figure 9: Connections and Performance over 1-20 Day Horizons: During Periods of High and Low Price Dispersion



Notes: this figure plots the estimated β coefficients from variant of regression 4.1 up to 20-day horizon ($T = 20$), where we interact connections with an indicator variable for high and low dispersion days. High (low) dispersion days are those trading days where the daily price dispersion (Jankowitsch, Nashikkar, and Subrahmanyam, 2011) is in the highest (lowest) tertile of our sample of trading days. We restrict the sample to sophisticated investors, and include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Transactions"). The shaded area denotes the 90% confidence band associated with the estimated β coefficients, It is based on robust standard errors, using two-way clustering at the day and the client level.

Table 10: Client Connections and Trading Performance: Unweighted Performance Results

	(1)	(2)	(3)	(4)	(5)
	1-day	2-day	3-day	4-day	5-day
Client	0.187**	0.250**	0.322**	0.480***	0.501***
Connections	(2.05)	(2.11)	(2.14)	(2.70)	(2.84)
Volume	-0.004	0.040	0.073	0.099	0.010
	(-0.04)	(0.29)	(0.43)	(0.54)	(0.05)
Tran.	-0.467*	-0.683*	-0.952**	-1.237***	-1.386***
	(-1.77)	(-1.93)	(-2.32)	(-2.67)	(-2.68)
N	100414	100414	100414	100414	100414
R^2	0.064	0.062	0.061	0.062	0.063
Day FE	Yes	Yes	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes	Yes	Yes

(a) Trading Performance over 1-5 Days

	(1)	(2)	(3)
	Baseline	Transaction	Anticipation
Client	0.480***	0.089**	0.387**
Connections	(2.70)	(2.44)	(2.20)
Volume	0.099	-0.088**	0.180
	(0.54)	(-2.21)	(0.99)
Tran.	-1.237***	-0.158	-1.079**
	(-2.67)	(-1.34)	(-2.30)
N	100414	100348	100348
R^2	0.062	0.149	0.059
Day FE	Yes	Yes	Yes
Client*Year FE	Yes	Yes	Yes

(b) Decomposing 4-day Performance: Transaction vs Anticipation Effect

Notes: panel A regresses the unweighted trading performance at different time horizons on client connections (4.1). The transaction-level data is collapsed at the client-day level. The performance measures are in basis points. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Tran."). Panel B decomposes the 4-day performance effect into a transaction component and an anticipation component (3.2). The results are based on the average transaction price $\overline{P}$ that uses the trades (for the given gilt) in a 3-hour window within the transaction time. T-statistics in parentheses are based on robust standard errors, using two-way clustering at the day and the client level. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

Table 11: Decomposing 4-day Performance into Transaction and Anticipation Effect: More vs Less Sophisticated Investors

	(1)	(2)	(3)	(4)	(5)	(6)
	More Sophisticated Investors			Less Sophisticated Investors		
	Baseline	Transaction	Anticipation	Baseline	Transaction	Anticipation
Client	0.799***	0.096*	0.696***	0.031	0.099	-0.076
Connections	(3.54)	(1.83)	(3.03)	(0.13)	(1.60)	(-0.29)
Volume	0.154	-0.096	0.244	0.487*	-0.034	0.512*
	(0.59)	(-1.15)	(0.96)	(1.83)	(-0.40)	(1.82)
Tran.	-2.037***	-0.188	-1.855***	-0.662	-0.319*	-0.328
	(-3.18)	(-1.09)	(-2.96)	(-1.00)	(-1.75)	(-0.46)
N	60694	60660	60660	39720	39688	39688
R^2	0.066	0.097	0.064	0.082	0.126	0.080
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Client#Year FE	Yes	Yes	Yes	Yes	Yes	Yes

(a) Weighted Performance

	(1)	(2)	(3)	(4)	(5)	(6)
	More Sophisticated Investors			Less Sophisticated Investors		
	Baseline	Transaction	Anticipation	Baseline	Transaction	Anticipation
Client	0.745***	0.079	0.663***	-0.050	0.096*	-0.154
Connections	(3.08)	(1.65)	(2.77)	(-0.22)	(1.69)	(-0.71)
Volume	-0.103	-0.126**	0.014	0.391	-0.040	0.422
	(-0.42)	(-2.38)	(0.06)	(1.58)	(-0.71)	(1.65)
Tran.	-1.539**	-0.047	-1.497**	-0.537	-0.299*	-0.218
	(-2.40)	(-0.31)	(-2.35)	(-0.81)	(-1.67)	(-0.32)
N	60694	60660	60660	39720	39688	39688
R^2	0.069	0.148	0.066	0.092	0.184	0.089
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Client#Year FE	Yes	Yes	Yes	Yes	Yes	Yes

(b) Unweighted Performance

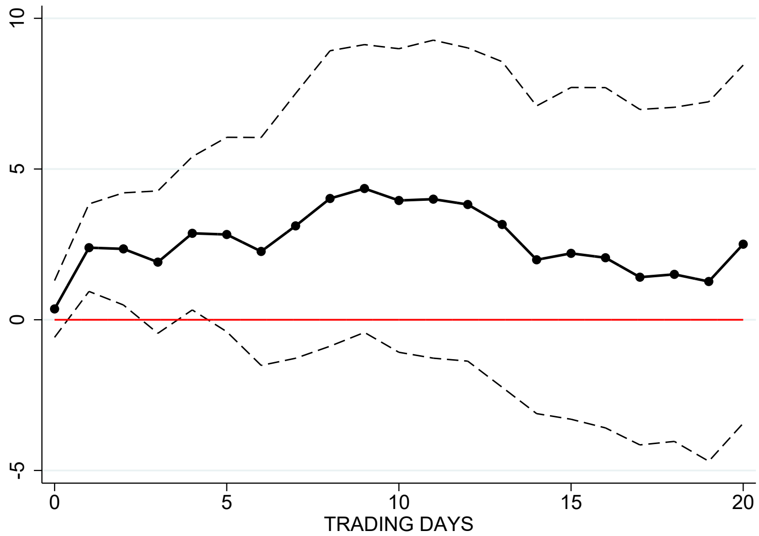
Notes: this table regresses the value-weighted (12a) and unweighted (12b) trading performance at the 4-day horizon, and its transaction and anticipation components (3.2), on client connections (4.1). The transaction-level data is collapsed at the client-day level. The performance measures are in basis points. We include as a control the natural logarithm of the pound trade volume of each client ("Volume") and the natural logarithm of the number of daily transactions ("Tran."). T-statistics in parentheses are based on robust standard errors, using two-way clustering at the client and day level. Columns 1-3 and columns 4-6 show the results for more sophisticated and less sophisticated clients, respectively. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

Table 12: Summary Statistics of Dealers' Affiliates – Client-Day Level

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Median	p10	p90	sd	N
InfShare	0.41	0.39	0.03	0.84	0.29	20,901
First Order Connection	3.01	2.00	1.00	6.00	2.16	20,901
Transaction Number	18.30	9.00	1.00	49.00	26.30	20,901
Volume (£millions)	108.90	17.71	0.33	290.14	261.56	20,901
Number of Bonds Traded	7.41	6.00	1.00	15.00	6.09	20,901

Notes: This table reports summary statistics for the sample of of dealers' affiliates (used in Section 6.1), covering 2011m10-2017m6, that is collapsed at the client-day level.

Figure 10: InfShare and Performance of Dealers' Affiliates over 1-20 Day Horizons



Notes: this figure plots the estimated β coefficients from variant of regression 6.2 up to 20-day horizon ($T = 20$). We include as a control the natural logarithm of the pound trade volume of each affiliate and each affiliate's dealer, the natural logarithm of the number of daily transactions of affiliate, the number of connections of each affiliate and each affiliate's dealer as well as the average $InfShare$ of dealers that a given affiliate trades with. The shaded area denotes the 90% confidence band associated with the estimated β coefficients, It is based on robust standard errors, using two-way clustering at the affiliate and the client level.

C GEMM Exits: 'Shocks' to Clients' Connections

C.1 Background

In our sample, there are two notable incidents of primary dealers ceasing their market making functions. These exits of GEMMs can be used as plausibly exogenous shocks to the connections of those client who traded with these dealers prior to their exits.

On 22 October 2015, it was announced that Credit Suisse would resign as GEMM, which would take effect at the close of business of the following day (23 October). On 29 January 2016, it was announced that Societe General would stop acting as GEMM, which would take effect on the following week (5 Feb). The Debt Management Office cited high capital costs of maintaining stock of British government debt to trade as well as tougher regulation since the Great Recession as reasons for these exits.

C.2 Empirical Analysis

We use these plausibly exogenous shocks to client connections to estimate the causal effect of connections on trading performance using a differences-in-differences (DD) strategy (Angrist and Prischke, 2009). First, we check whether connections of clients who traded with dealers prior to their exits significantly changed after the dealers' exits. We focus on a 20-day period around the exit, i.e. the treated group γ_s includes clients who traded at least once with the dealer during the 20 days prior to the given dealer's exit. Specifically, we estimate the following regression for each client i, group s and time period t:

$$Connections_{i,s,t} = \gamma_s + \lambda_t + \delta D_{s,t} + controls_{i,s,t} + \varepsilon_{i,s,t}, \tag{C.1}$$

where γ_s takes value 1 for the treated group and 0 for the control group, λ_t takes value 1 for the post-treatment period and 0 for the pre-treatment period, $D_{s,t}$ takes value 1 for the treated group during the post-treatment period and 0 otherwise, and $controls_{i,s,t}$ includes the log of trading volume and number of transactions of client i. The coefficient of interest is δ, which captures whether the exit of the given GEMM caused a significant drop in the number of dealer connections of the treated group of clients.

Table 13 shows that clients lose on average one dealer-connection after the given dealer ceases to be a market maker in the gilt market.

Next, we test whether changes in connections, induced by exiting GEMMs, would have an effect on the trading performance of clients. To do that, we use an instrumental

Table 13: The Effect of GEMM Exit on Client's Connections

δ	-1.135**
	(-2.23)
γ_s	1.996***
	(5.25)
λ_t	0.122
	(0.71)
Intensity	1.563***
	(20.04)
Volume	0.297***
	(8.71)
N	811
R^2	0.678

Notes: This table shows the estimation results for regression C.1. The sample includes sophisticated clients only. T-statistics in parentheses are based on robust standard errors. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01).

variable strategy, and take the fitted values from equation C.1 to use it in the following performance regression:

$$Performance^T_{i,s,t} = \gamma_s + \lambda_t + \beta \times Connec\tilde{t}ions_{i,s,t} + controls_{i,s,t} + \varepsilon_{i,s,t}, \tag{C.2}$$

where β is the coefficient of interest, which reveals whether changes in connections, induced by GEMM exits, would have an effect on trading performance among sophisticated clients. Table 14 shows that there is a statistically insignificant relationship between connections and trading performance.

Discussion and Caveats Our baseline result in the full sample shows that there is a positive and significant relationship between trading performance and clients' connection. One possible explanation for this is that clients learn from their new connections, and this newly acquired knowledge is reflected in increased trading performance. If this mechanism is at at play, we would have expected the estimates for β in regression C.2 to be positive and significant.

While this non-result is suggestive that this mechanism is not what is driving the main result of our paper, we acknowledge that this non-result comes from a rather weak test: there are only two data points in our sample for GEMM exits, so the plausibly

Table 14: GEMM Exit, Connections and Trading Performance

	(1)	(2)	(3)	(4)	(5)	(6)
	0-day	1-day	2-day	3-day	4-day	5-day
Connections	-1.444	-5.685	-4.647	-4.804	-9.889	-12.812
	(-0.67)	(-1.28)	(-0.81)	(-0.85)	(-1.31)	(-1.39)
Intensity	3.127	9.956	7.424	9.299	16.762	20.067
	(0.93)	(1.43)	(0.82)	(1.05)	(1.41)	(1.38)
Volume	0.626	1.575	1.698	1.043	2.440	4.188
	(0.94)	(1.14)	(0.95)	(0.59)	(1.04)	(1.45)
N	811	811	811	811	811	811
F-stat	7.431	7.431	7.431	7.431	7.431	7.431

Notes: This table shows the instrumental-variable estimation results for regression C.2. The sample includes sophisticated clients only. T-statistics in parentheses are based on robust standard errors. Asterisks denote significance levels (* p<0.1, ** p<0.05, *** p<0.01). The F-statistics denotes the Cragg-Donald Wald F statistic.

exogenous variation in clients' connections is rather limited. Also, the two dealers who exited are rather peripheral in the dealer network, so the learning mechanism may be much weaker in these cases compared to more central and larger dealers (whose exit we do not observe).