Thèse n° 7878

Trading mechanisms in over-the-counter markets

Présentée le 25 août 2020

au Collège du management de la technologie
Chaire du Prof. ordinaire Collin-Dufresne
Programme doctoral en finance

pour l'obtention du grade de Docteur ès Sciences

par

Sebastian VOGEL

Acceptée sur proposition du jury

Prof. J. Hugonnier, président du jury
Prof. P. Collin Dufresne, directeur de thèse
Prof. N. Schürhoff, rapporteur
Prof. A. Neklyudov, rapporteur
Prof. S. Malamud, rapporteur

École
polytechnique
fédérale
de Lausanne

2020

Abstract in German

Im Kapitel "When to Introduce Electronic Trading Platforms in Over-the-Counter Markets?" Bestimme ich ein Gleichgewicht in einem Markt, in dem Händler die Wahl haben, eine elektronische Plattform mit einem RFQ-Protokoll zu verwenden oder einen Dealer direkt anzurufen. Die wichtigste Erkenntnis ist, dass Plattformen die Verhandlungsmacht von Händlern mit Suchkosten erhöhen und dadurch ihre Marktbeteiligung und möglicherweise auch die Wohlfahrt erhöhen. Eine weitere Reibung, die den Handel auf OTC-Märkten beeinflusst, sind Informationsreibungen. Marktteilnehmer können aus Handelsaktivitäten auf Plattformen lernen und mit diesen Informationen an anderen Handelsplätzen handeln. Das Zusammenspiel zwischen einer elektronischen Handelsplattform auf dem Händler-Kunden-Markt und dem Interdealer-Markt wird im Kapitel "Electronic Trading in OTC Markets vs. Centralized Exchange" analysiert. In diesem Kapitel wird ein zweistufiger OTC-Markt mit einem vollständig zentralisierten Markt verglichen und es wird gezeigt, dass informierte Händler möglicherweise die zweistufige Struktur bevorzugen, um von ihren Informationen zu profitieren. Diese Fähigkeit der dezentralen OTC-Märkte, mit asymmetrischen Informationen umzugehen, ist auch ein großer Vorteil der OTC-Märkte im Vergleich zu den Börsenmärkten, wenn Wohlfahrt betrachtet wird. In dem Kapitel "Informed Traders and Dealers in the FX Forward Market" wird analysiert, ob solche Informationsprobleme bestehen, ob Händler in der Lage sind, diese zu verringern, indem sie unterschiedlichen Kunden unterschiedliche Aufschläge in Rechnung stellen, und ob das Verhalten der Marktteilnehmer mit der Verringerung von asymmetrischen Informationen konsistent ist.

Abstract in English

In the chapter "When to Introduce Electronic Trading Platforms in Over-the-Counter Markets?" An equilibrium in a market is determined in which traders have the choice between using an electronic platform with a request-for-quote protocol or calling a dealer directly. The main takeaway is that platforms increase the bargaining power of traders with search costs, thereby increasing their market participation and potentially also welfare. Another friction that affects trading in OTC markets are informational frictions. market participants may learn from trading activity on platforms and trade on that information in other trading venues. The interplay between an electronic trading platform in the dealer-to-customer market and the interdealer market is analysed in the chapter "Electronic Trading in OTC Markets vs. Centralized Exchange." In that chapter, a two-tiered OTC market is compared with a fully centralized market, showing that informed traders may prefer the two-tiered structure in order to benefit from their information. This ability of decentralized OTC markets to deal with asymmetric information is also a major benefit of OTC markets compared to exchange markets in welfare considerations. In the chapter "Informed Traders and Dealers in the FX Forward Market," it is analyzed whether such informational frictions exist, whether dealers are able to mitigate those by charging different markups to different clients and whether we behavior consistent with the alleviation of informational frictions.

Contents

Introduction

Over the last 20 years, over-the-counter (OTC) markets have undergone a significant transformation. Where, traditionally, the only option for obtaining a price was to call a dealer, one now has a wide menu of different options of how to ask for a price. In particular, electronic trading platforms offer many diverse trading protocols like request-for-quote, request-for-market, streaming quotes and, to a limited extend, some form of central-limit-order book trading.

Given those diverse options, it is not obvious for participants in financial markets how to use those trading protocols optimally. Related to that question, it is unclear what the effects of introducing new trading protocols are on market equilibria or when financial regulators should mandate certain forms of trading.

This thesis addresses those questions. In the chapter "When to Introduce Electronic Trading Platforms in Over-the-Counter Markets?" I determine an equilibrium in a market in which traders have the choice between using an electronic platform with a request-for-quote protocol or calling a dealer directly. The main takeaway is that platforms increase the bargaining power of traders with search costs, thereby increasing their market participation and potentially also welfare. Another friction that affects trading in OTC markets are informational frictions. market participants may learn from trading activity on platforms and trade on that information in other trading venues. The interplay between an electronic trading platform in the dealer-to-customer market and the interdealer market is analysed in the chapter "Electronic Trading in OTC Markets vs. Centralized Exchange." In that chapter, a two-tiered OTC market is compared with a fully centralized market, showing that informed traders may prefer the two-tiered structure in order to benefit from their information. This ability of decentralized OTC markets to deal with asymmetric information is also a major benefit of OTC markets compared to exchange markets in welfare considerations. In the chapter "Informed Traders and Dealers in the FX Forward Market," it is analyzed whether such informational frictions exist, whether dealers are able to mitigate those by charging different markups to different clients and whether we behavior consistent with the alleviation of informational frictions.

To conclude, new trading protocols in OTC markets have the potential to increase efficiency. Different traders may optimally choose different trading protocols to suit there needs. One should however be cautious when trying to centralize markets, as decentralized markets may have their merits.

Chapter 1

When to Introduce Electronic Trading Platforms in Over-the-Counter Markets?

Electronic trading platforms are currently transforming the way investors trade in over-the-counter (OTC) markets. They also played a central role in an EU investigation[1] and a class action in the U.S. resulting in a $1.87bn settlement in 2015.[2] In this class action, a number of buy-side investors accused 13 dealer banks, along with Markit and the ISDA, of colluding to keep competition from electronic platform providers out of the market. Besides a large fee, the resulting settlement included promises by the defendants to promote electronic trading in the future. The settlement was welcomed by investors and regulators as a significant step toward market efficiency. The rather reluctant introduction of electronic trading in the CDS market stands in sharp contrast to trading practices in the bond market. The first electronic trading platform for bonds has been established in 1988[3] and new bond trading platforms easily gain traction.[4]

What do these disputed trading platforms do? In general, electronic trading platforms are meant to facilitate trading in OTC markets by increasing competition among dealers and leading to higher price transparency. A trader who wants to buy corporate bonds (Hendershott and Madhavan (2015a)), interest-rate swaps (Benos et al. (2018)) or foreign exchange securities (Hau et al. (2017) and Bjonnes et al. (2017)),

[1] For details regarding the investigation that started in 2011, see White and Bodoni (2015) and White (2016).
[2] See Drucker and Voris (2015) and McLannahan and Rennison (2015) for details.
[3] See MTS's timeline: `http://www.mtsmarkets.com/about-us/company-timeline`.
[4] See Marsh and Detrixhe (2015).

most of which are not traded on exchanges, usually has the choice between (i) calling a single dealer to ask for a price at which this dealer is willing to trade a certain quantity of the bond and (ii) submitting a so-called request-for-quote (RFQ) via an electronic trading platform to multiple dealers at once. A dealer who has received an RFQ may or may not reply by quoting a price at which the dealer is willing to trade a certain quantity of the bond. If some of the contacted dealers reply to an RFQ, the trader can pick the most attractive quote and trade the bond at this price. Thus, submitting an RFQ on an electronic trading platform means performing a first-price auction among the contacted dealers.

A number of questions arise from the anecdotes on the CDS and bond markets mentioned above. Why have electronic trading platforms been used in the bond market for such a long time, while dealers were very reluctant to introduce them in the CDS market? How do electronic trading platforms affect the ability of the market to match buyers and sellers of financial assets? How do dealers' quoting strategies change when there is a platform introduced in an OTC market? Is there room for a regulator to improve a given market structure?

To address these questions, I develop a model of a hybrid market (HM) in which traders have the opportunity to buy an asset in a bilateral market or on an electronic trading platform. In the bilateral market, each dealer has to be contacted separately, whereas a trading platform allows to contact all dealers at once. The assumptions on dealers and traders are as in Duffie et al. (2017), who model a pure bilateral market (PBM). There is a continuum of traders who want to buy one unit of an asset. A fraction of these traders find search costless and the other traders have positive search costs. Traders with search costs are called "slow" and traders without search costs are called "fast." A finite number of dealers can provide the asset at some cost.

One can assess the effects of introducing an electronic trading platform in an OTC market by comparing the equilibrium in an HM (modeled in this paper) with the equilibrium in a PBM (modeled in Duffie et al. (2017)). It turns out that introducing an electronic trading platform increases slow traders' market participation and makes slow and fast traders better off. However, the dealers' markups may decline sufficiently to make an HM undesirable for them. Introducing a trading platform increases competition among dealers and trading on a platform may be associated with less search effort. Thus, entering the market is more attractive for slow traders in the HM. Since the dealers' profits are mainly generated by trading with slow traders, dealers have incentives to introduce electronic trading if the slow traders' market participation is very low. The latter is the case if (i) the slow traders' search costs are very high, if (ii) most traders are slow and the lack of competition in the PBM results in high markups or (iii) if the number of dealers is large, which results in dealers charging high average markups and making slow traders to stay out of the market. If slow traders already participate actively in the market or if competition on the platform is too intense, dealers prefer the PBM.

If one views institutional investors as fast traders and retail investors as slow traders, then these model implications are consistent with the anecdotes about the CDS and bond markets mentioned above. There may be relatively more retail investors interested in trading bonds than in trading CDSs, which are used by institutional investors to hedge certain risks. The model in this paper suggests that it may be profitable for dealers to introduce a platform in bond markets while it may not be profitable for dealers to do so in CDS markets. This may explain why electronic trading platforms have been used in bond markets for many years, while they have been introduced in CDS markets only recently and under pressure from regulators. That the presence of institutional investors is a critical determinant of an endogenous market structure is in line with Biais and Green (2007). The latter suggest that the migration of bond trading from exchanges to the OTC market was related to the growing importance of institutional investors in bond markets. Analogously, my model suggests that bond trading migrated back to more centralized OTC markets due to a more dominant role of retail investors in the bond markets compared to the CDS markets.

A regulator, who wants to maximize the aggregate benefits of trade between traders and dealers net of the slow traders' search effort, faces a slightly different tradeoff: On one hand, an HM is generally associated with higher market participation by slow traders compared to the PBM. On the other hand, searching for a quote is more complicated in the HM than in the PBM and, even conditionally on entering the market, slow traders may exert a higher search effort in the HM than in the PBM, if the cost of using the electronic trading platform is high. If, however, the cost of using the electronic trading platform is very low, introducing electronic trading actually lowers the slow traders' aggregate search effort and mandating electronic trading always increases welfare by reducing aggregate search effort.

The policy recommendations that can be derived from the model are as follows. If dealers prefer the HM to the PBM, the HM is always efficient. The latter is arguably the case in the bond market, judging by the dealers' lack of opposition against electronic trading compared to the CDS market. If dealers prefer the PBM to the HM, a regulator may still want to mandate electronic trading if the searching on the electronic trading platform is sufficiently easy for slow traders. Search-theoretic arguments will imply simple estimators for the key model parameters determining the policy recommendation based on summary statistics of data on transactions or quotes.

To the best of my knowledge, this is the first paper providing an equilibrium model of a hybrid OTC market comprising a traditional (bilateral) voice market and electronic trading via RFQs, the common market structure for many assets like bonds, interest-rate swaps or foreign exchange securities.

The first contribution of the paper consists in the derivation of equilibrium search strategies of the traders and equilibrium quoting strategies of the dealers in an HM, drawing on results and techniques of Duffie et al. (2017). The latter develop a model of bilateral dealer markets in which traders have to contact dealers sequentially to obtain prices for an asset. They combine random-pricing strategies from consumer search models[5] with insights on optimal search by Weitzman (1979) and show that a benchmark can improve both dealer profits and welfare. In this paper, I am making similar assumptions on traders and dealers as those in Duffie et al. (2017), in particular that dealers' quotes do not change when a trader returns after having contacted other dealers. Zhu (2012a) models a bilateral OTC market under the assumption that a dealer will renegotiate an offer if a trader comes back to him after having visited other dealers, showing that dealers give less attractive quotes in equilibrium when contacted repeatedly since other dealers likely gave unattractive quotes as well in this case. These search-for-quotes models of bilateral OTC markets in the style of Duffie et al. (2017) or Zhu (2012a) have to be distinguished from intensity-driven search models as for instance Duffie et al. (2005), Weill (2007a), Lagos and Rocheteau (2009a), Gârleanu (2009), Lagos et al. (2011a), Feldhütter (2012), Philippon and Pagnotta (2011), Lester et al. (2015a) or Glebkin (2016), in which buyers and sellers trade with each other after waiting a random time. In search-for quotes models, a buyer exactly knows where to find the asset, but reaching out to the seller takes effort. Search-for-quotes models also allow bargaining and market entry to be inefficient, whereas in intensity-driven search models it is generally simply assumed that bargaining is efficient. Indeed, this paper argues that market entry in OTC markets is generally inefficient, but can be improved by electronic trading.

The model presented in this paper extends the search-for quotes model of Duffie et al. (2017) by introducing an electronic trading platform as a second trading venue in addition to the bilateral voice market. On the trading platform, dealers have different incentives to compete than in the bilateral market. On the platform, each dealer only responds with a certain probability less than one. In that sense, this paper relates to the literature on auctions with entry of bidders as in McAfee and McMillan (1987), Levin and Smith (1994) or Menezes and Monteiro (2000), in which equilibrium strategies for bidders in an auction are derived in case sellers do not know what types of other bidders or how many of them are competing. Jovanovic and Menkveld (2014) and Yueshen (2017) use techniques from this literature to model competition between market makers in limit order books. This paper differs from all of these models in that I study the

[5]See for example Varian (1980b), Burdett and Judd (1983a), Stahl (1989a) and Janssen et al. (2005, 2011). Dennert (1993)uses similar techniques to study market makers' quotes in limit order books.

interaction between an auction market with random entry of dealers and a bilateral dealer market in which each trader can only contact one dealer simultaneously. Due to the traders' endogenous choice of trading venue, interactions between trading venues affect what kinds of traders the dealers face in each market and the level of their reservation prices.

Thus, this paper also relates to a large strand of literature on interactions between trading venues. While Glosten (1994) states condition under which an electronic limit order book does not invite competition from third-party dealers, Seppi (1990a) and Lee and Wang (2018) study the interaction between an exchange and an upstairs or, respectively, and OTC market, arguing that investors that are likely to be informed mainly trade on the exchange in equilibrium while investors that are likely to be uninformed trade off the exchange. Similarly, Zhu (2014) argues that, if the volatility of the asset is not too high, informed traders trade on the exchange and only uninformed traders trade in a dark pool. As opposed to Seppi (1990a), Grossman (1992) argues that intermediaries in an upstairs market mitigate through their knowledge on future order flow the negative effects on liquidity of the fact that traders do not continuously participate on the exchange. Hendershott and Mendelson (2000) study electronic crossing networks and exchanges with market markers and argue that introducing crossing networks increases market liquidity, but crowds out high-net-gain traders. Parlour and Seppi (2003) argue that order preferencing plays a key role in determining how much volume is traded in a pure limit order market and a hybrid market structure also including specialists. Vayanos and Weill (2008) show how the different repo rates may arise for liquid and illiquid bonds and how this explains the on-the run phenomenon, namely that just-issued bonds have lower yields than previously issued bonds. This paper differers from all of the above, since I focus on an OTC market and study the interaction between two OTC trading venues for the same asset. OTC markets function very differently from exchanges. Whereas on exchanges, investors can either trade with each other or through an intermediary that quotes the same spread for all investors, trading in OTC markets requires search effort by investors and is associated with bargaining between buyers and sellers.

Hendershott and Madhavan (2015a) perform an empirical analysis of the hybrid OTC market structure comprising a bilateral voice market and an electronic trading platform. Whereas they propose a statistical model to estimate trading costs, controlling for the endogenous choice of trading venue, this paper provides an equilibrium model, in which trading decisions and the choices of trading venues are determined by primitive parameters. Hendershott and Madhavan (2015a) suggest that a bond is more likely to be traded electronically if the trader wants to sell or if the trade happens at the end of the month. Interpreting these trades as urgent trades that are made by traders with high costs of waiting, the model implies that these trades should indeed more likely happen on electronic trading platforms. Benos et al. (2018) provide evidence on the effect of electronic trading on price dispersion. Their claim that electronic trading lowers price dispersion is largely consistent with the model in this paper, which implies precise equilibrium price distributions for each trading venue and allows conditions under which the decrease in price dispersion is large to be derived. Bjonnes et al. (2017) and Hau et al. (2017) study how markups differ for different types traders and show that sophisticated traders get better prices, consistent with the implications of the model in this paper. The results of Hau et al. (2017) suggest that especially unsophisticated traders benefit from electronic trading, consistent with the model in this paper. Other empirical papers on electronic OTC markets that are different from the HM considered in this paper include Collin-Dufresne et al. (2018) and Riggs et al. (2018) who study the index CDS markets in which electronic trading is mandatory. In these markets, dealers have access to a central limit order book and trade with customers mostly through RFQs.[6]

As a second contribution, the model provides conditions under which a regulator can improve a given market structure by introducing electronic trading platforms. In this way, this paper is related to the literature comparing alternative market structures, usually centralized to decentralized ones. While Pagano (1989) argues that a single exchange welfare dominates a market with multiple exchanges, Malamud and Rostek (2017) show that a fragmented exchange market may be more efficient if traders have different risk aversions. Biais (1993) argues that spreads are, albeit equal in expectation, more variable in decentralized markets than in centralized markets and De Frutos and Manzano (2002) and Yin (2005b) use similar models to argue that decentralized markets may be more efficient. Biais et al. (1998) and Viswanathan and Wang (2002) compare different trading protocols on exchanges and Duffie and Zhu (2011) and Acharya and Bisin (2014) study how counterparty risk in OTC markets may be affected by clearing through central counterparties (CCPs). Babus and Parlatore (2016) study how fragmented markets may arise in equilibrium even through a centralized exchange market would be more efficient. An exchange market is more efficient than a hybrid market with exchange and OTC trading in Lee and Wang (2018), if and only if there is sufficiently much uninformed trading. Glode and Opp (2017b) argue that a bilateral OTC market may be more efficient than an exchange market because of different incentives for sellers of an asset to acquire information. This paper differs again from all of these papers, because it focuses on a market structure that has not been studied in any of the papers mentioned above. Moreover, electronic trading in OTC markets is fundamentally different from an exchange market: In Lee and Wang (2018) dealers charge higher spreads on the exchange than in the OTC market. Empirical evidence as for instance in Hau et al. (2017) suggests that unsophisti-

[6]Dealers have access to central limit order books also in other OTC markets. For instance, Östberg and Richter (2018) and Schneider et al. (2018) study the centralized interdealer market for European bonds.

cated traders receive more favorable prices in the electronic market. Consistent with this empirical evidence, this paper allows dealers to quote lower average markups in equilibrium in the electronic market than in the bilateral voice market, meaning that slow traders will receive better prices on the platform. In general, OTC markets are viewed as exhibiting price dispersion, which means that different traders pay different prices for the same asset due to dealers' markups. Other related models comparing centralized to fragmented markets either do not feature price dispersion in the OTC market in equilibrium (Lee and Wang (2018)) or do not model dealers' markups at all since traders are assumed to be able to trade directly with each other (Babus and Parlatore (2016)). In this paper, dealers quote according to endogenously determined nondegenerate probability distributions.[7] Whereas Glode and Opp (2017b) and Lee and Wang (2018) consider information asymmetries as the dominant friction in their models, the results in this paper are driven by search costs. This can by motivated by empirical evidence suggesting that most trading in OTC markets is not of speculative nature.[8]

The rest of the paper is organized as follows. Section 1.1 describes the PBM and reviews results from Duffie et al. (2017). Section 3.6 introduces an electronic trading platform into the model setup from Section 1.1 and provides first results on traders' and dealers' equilibrium strategies. In Section 3.6.2, an equilibrium is established and characterized. Section 1.4 discusses results on market design. In Section 1.5, policy implications and empirical predictions are derived. Section 1.6 evaluates the appropriateness of important model assumptions and discusses limitations of the model. Section 1.6 concludes. Appendix 1.8.1 contains details on the derivation of the HM equilibrium. Proofs are relegated to Appendix 1.8.2.

1.1 Benchmark: The Pure Bilateral Market

This section describes a PBM as modeled in Duffie et al. (2017) with no uncertainty about the dealers' cost. For expositional purposes, I will briefly review the main results regarding the equilibrium in the PBM below. Afterwards, I will present a model of the HM in which there is an electronic trading platform as an additional trading venue.

There are $\mathbb{N} \ni N > 1$ risk-neutral dealers and a continuum of risk-neutral traders with measure 1. Traders want to buy an asset to which they attribute value $v \in \mathbb{R}$. Dealers can provide the asset at cost

[7]Empirically, bargaining plays a major part in determining prices in OTC markets. Costs related to bargaining have been extensively studied for the bond markets. See for instance Edwards et al. (2007) and Goldstein et al. (2007). Harris and Piwowar (2006) compare costs of trading municipal bonds to costs of trading equities on exchanges. Green et al. (2007) suggest that dealers possess substantial bargaining power.

[8]For instance, Hilscher et al. (2015) argue that informed traders are primarily active in the equity market as opposed to the CDS market. Han and Nikolaou (2016) argue that due to the low risk of repo transactions, trades are plausibly motivated by liquidity needs and Oehmke and Zawadowski (2016) argue that speculative activity is very low in the bond market.

$c \in \mathbb{R}$. A fraction $\mu \in (0,1)$ of traders is fast and does not find search costly. The other traders are slow and have to pay cost $s_b > 0$ when contacting a dealer in the bilateral market. Traders are called "fast" or "slow," because the cost associated with search can be interpreted as the cost of waiting. It is assumed that v, c, μ and s_b are common knowledge. In the following it will also be assumed that $v > c + s_b$, which means that market entry by slow traders is efficient.

In a PBM, traders have to contact each of the N dealers separately to obtain a quote. A dealer does not observe whether the trader who contacts him is fast or slow. After each contact with a dealer, a trader has the option to buy the asset at one of the quotes obtained up to that point, leave the market without buying the asset or continue to search.

Duffie et al. (2017) derive the following equilibrium. Fast traders will always contact all dealers in the PBM. Slow traders will enter the market with a probability $\gamma^{PBM} \in (0,1]$ and follow a reservation price strategy with reservation price r^{PBM}. Each dealer independently draws a price p from a continuous distribution H^{PBM} without atoms or gaps and finite support $[\underline{p}^{PBM}, r^{PBM}] \subset \mathbb{R}$. On its support, this distribution is given by

$$H^{PBM}(p) = 1 - \left(\frac{\gamma^{PBM}(1-\mu)(r^{PBM} - p)}{N\mu(p-c)} \right)^{1/(N-1)}.$$

The distribution H^{PBM} is determined such that a dealer is indifferent between quoting any price in the support of H^{PBM}. The slow traders' reservation price satisfies

$$r^{PBM} = \int_{\underline{p}^{PBM}}^{r^{PBM}} p \, \mathrm{d}H^{PBM}(p) + s_b.$$

The reservation price in the PBM satisfies an indifference condition. When a slow trader is offered the reservation price, he is indifferent between accepting the offer and continuing to search in the PBM. The reservation price r^{PBM} is a strictly monotone increasing function in γ^{PBM} with $\lim_{\gamma^{PBM} \to 0} r^{PBM}(\gamma^{PBM}) = c + s_b$.

Finally, the slow traders' probability of market entry is given by

$$\gamma^{PBM} = \begin{cases} 1, & \text{if } r^{PBM}(1) < v, \\ x, & \text{if } r^{PBM}(1) \geq v, \end{cases}$$

where x is uniquely determined such that $r^{PBM}(x) = v$ and satisfies $x \in (0,1]$. Thus, when a slow trader expects a positive profit of entering the PBM even if all other slow traders enter with probability one, he enters with probability 1 as well. Otherwise, there is a unique equilibrium value $\gamma^{PBM} \in (0, 1]$ such that expected payoffs of entering the PBM are zero.

1.2 A Model of the Hybrid Market

In this section I present the setup for the model of the hybrid market structure. The setup builds on the model of Duffie et al. (2017) that was partially reviewed in Section 1.1. In the model of the HM, the assumptions on traders and dealers are the same as those laid out in Section 1.1. In the HM, however, traders have the choice between two different trading venues:

- **Bilateral market:** As in section 1.1, a trader may contact each of the N dealers individually and ask for a quote. After each contact with a dealer, a trader may buy the asset at one of the prices obtained so far, leave the market or continue searching in the bilateral market or on the electronic trading platform (see next bullet point). Continuing to search in the bilateral market is only feasible if there are dealers who have not yet been contacted individually by the trader. Similarly, continuing to search on the platform is only feasible if the trader has not yet submitted an RFQ though the platform, e.g. because the cost of returning to the platform is prohibitively high for reasons that are not modeled in this paper, like reputation concerns or information leakage. It is also assumed that each dealer can be contacted bilaterally at most once by every trader. Slow traders have to pay cost s_b each time they contact a dealer.

- **Electronic trading platform:** Traders can also submit a so-called “request-for-quote” (RFQ) via an electronic trading platform to all N dealers at once. Every time a trader submits an RFQ on the electronic trading platform, each dealer can decide to respond to this RFQ and provide a quote at which he is willing to sell the asset. Thus, the trader gets a number $n \in \mathbb{N}$ of quotes, with $0 \leq n \leq N$. It is quite possible that $n = 0$, as Hendershott and Madhavan (2015a) show empirically. Each dealer independently responds to an RFQ with probability $\eta \in (0, 1)$, which is an exogenous parameter of the model. After having received the n quotes, a trader can either decide to buy the asset at the most attractive quote he has received so far, continue searching in the bilateral market or go out of the market without buying the asset. Continuing to search in the bilateral market is only feasible if there are dealers left in the bilateral market who have not yet been contacted by the trader. Slow traders have to pay a cost $s_p \geq 0$ when submitting an RFQ. Fast traders find it costless to use the platform and submit an RFQ. Each trader is allowed to submit at most one RFQ via the electronic trading platform.

Notice that the slow traders' search costs in the bilateral market s_b may be different from their search costs on the platform s_p. In order for a nontrivial equilibrium to exist, I assume that $s_b > 0$, i.e. it takes at least some strictly positive effort for slow traders to contact a dealer in the bilateral market. However, $s_p = 0$ is allowed in this setup, taking into account that electronic trading platforms are meant to facilitate the access to dealers in OTC markets. Each dealer has a separate trading desk for the bilateral market and for the platform. The objective of each trading desk is to maximize its own profit.[9]

Traders can visit the N dealers in the bilateral market and use the platform in any order. Potentially, they can contact all available dealers and also go to the platform once to request quotes. Thus, there are potentially $N + 1$ periods in which traders can search for quotes. After that they can either choose the most attractive quote that has been offered and buy the asset wherever they found this offer or they can choose not to buy the asset if the quotes offered by the dealers were too expensive. In the event that a trader has received several lowest quotes and he decides to buy the asset, he is equally likely to choose any one of those lowest quotes. Traders can terminate their search early or stay out of the market completely. Terminating the search early or staying out of the market may be optimal for slow traders who find search costly. The dealers' RFQ response rate η is exogenously given and can represent the cost of paying attention as in Yueshen (2017). Risk management considerations might also prevent dealers from responding to every RFQ. In particular, operational risk may impose an exogenous restriction that makes dealers respond at a lower rate in the electronic market than in the bilateral voice market.[10] One may thus interpret η as a restriction on the RFQ response rate that banks impose on their electronic trading desks out of (to other parameters unrelated) fear that algorithms "go rogue."

Since dealers are identical, dealers' quoting strategies are assumed to be symmetric on both the platform and in the bilateral market. As in the PBM described in Section 1.1, a dealer does not observe the trader's

[9]Some readers may find the assumption that trading desks within one dealer are competing with each other strange. However, anecdotal evidence suggests that competition between different trading departments within a dealer bank is not uncommon. For instance, Rodgers (2017) describes how letting prop traders compete with the options traders helped Deutsche Bank implement automatic pricing of currency options, which required options traders to update inputs for an algorithm regularly: "The beauty of the scheme was that I knew that the prop guys would scour the screens all day long looking for discrepancies but, if they found them and dealt, the money that they'd make (and the options guys would lose as a consequence) would stay within the FX department [...]." In any case, the model assumption that all trading desks compete with each other will play a very minor role. Much more important are the interactions between the two trading venues as a whole.

[10]Relatedly, Basak and Buffa (2017) provide several anecdotes of trading failures of financial firms and study operational risk in the context of an investment problem. Rodgers (2017) writes, when describing the adoption of electronic trading in the foreign exchange market, on this particular source of risk: "Indeed, within Deutsche Bank and its rivals (and, I'm told, algorithmic funds) this is the main focus of internal controls - how ensure that systems don't 'go rogue' and accumulate huge positions that could cause financial harm."

type and does not know how many dealers the trader has contacted before. Therefore, dealers do not observe any information on which they can condition their quotes. The prices each dealer quotes in the bilateral market can therefore be assumed to be independent and identically distributed. Similarly, the prices each dealer quotes on the electronic trading platform can be assumed to be independent and identically distributed. Notice, however, that dealers' quoting strategies in the bilateral market may be different from those they use on the electronic trading platform.

A dealer would never want to quote a price below his cost c, since doing so may result in losses when a trader buys the asset from that dealer. Quoting below c is clearly never optimal. Without loss of generality it can therefore be assumed that a dealer quotes according to probability distributions with support in $[c, \infty)$. This means the following:[11]

1. In the bilateral market, dealers quote prices drawn independently from a distribution function $H : \mathbb{R} \to [0, 1]$ with $\text{support}(H) \subset [c, \infty)$.

2. On the electronic trading platform, dealers quote independently from other dealers according to a distribution function $G : \mathbb{R} \to [0, 1]$ with $\text{support}(G) \subset [c, \infty)$.

Since the dealers' strategies are symmetric, H and G are the same for all dealers. Traders' decisions about market entry and exit and the dealers' quotes are assumed to be essentially pairwise independent. This makes the exact law of large numbers by Sun (2006a) applicable when referring to prices and quantities of traders in the market.

This section proceeds as follows. First, the traders' optimal search strategies conditional on the dealers' quoting strategies are derived. A search strategy specifies where to start the search, when to accept an offer, when to leave the market as well as when and where to continue to search. Second, the dealers' optimal quoting strategies are determined given that the traders use specific search strategies.

Overall, the following discussion is meant to convey the central ideas in the derivation of the equilibrium. For the sake of conciseness and clarity, technical details have been relegated to Appendix 1.8.1.

1.2.1 Traders

The equilibrium search strategies are different for slow and fast traders. Since fast traders can costlessly canvass the entire market, a fast trader will always take advantage of this ability. Obtaining another quote can potentially result in a lower price the trader has to pay for the asset, while the other offers received

[11] In the following, "support(D)" refers to the support of the distribution D.

before remain valid. The exact order in which a fast trader contacts dealers or goes to the platform does not influence his expected payoff. A fast trader will choose the lowest offer received from the N dealers in the bilateral market and from the n dealers who responded to the RFQ on the platform.

Slow traders, on the other hand, have to carefully consider whether entering the market or continuing to search is worth the cost s_b or s_p, respectively. Slow traders also have to determine the exact order in which they go to the platform or contact dealers in the bilateral market. This paper will focus on the case in which slow traders follow the following reservation price strategy.

Definition 1. *A trader's reservation price strategy with reservation price $r \in \mathbb{R}$ and probability of continued search $\gamma \in (0, 1]$ is defined by the following actions:*

- *The trader almost surely enters the market by submitting an RFQ to the dealers on the trading platform.*
- *The trader buys the asset at the most attractive quote as soon as he has received a quote less than or equal to r.*
- *The trader keeps searching for quotes with some probability $\gamma \in (0, 1]$ in the bilateral market if the trader did not receive a quote less than or equal to r on the platform or from a dealer in the bilateral market. Searching in the bilateral market means that the trader contacts with equal probability one of the dealers that the trader has not yet contacted.*

In the following, it will be determined when following a reservation price strategy in the sense of Definition 1 is optimal for slow traders if dealers in turn follow optimal quoting strategies in response to the slow traders' reservation price strategy.

1.2.2 Dealers

This section derives the dealers optimal quoting strategies given that slow traders follow a reservation price strategy as described in Definition 1. For now, the slow traders' reservation price r and the probability of continued search γ will be taken as given. In Section 3.6.2, r and γ will be endogenized. In any equilibrium, it must be the case that $r > c$, which can be seen as follows. If a slow trader can obtain the asset (on the platform or in the bilateral market) at a price between c and $c + s_b$, the slow trader optimally buys the asset at this price, since continuing to search in the bilateral market[12] will result in the search effort $s_b > 0$ and a price for the asset greater than or equal to the dealers' cost c. Thus the slow trader is better of by buying

[12] According to Definition 1, only searching in the bilateral market is still a feasible option.

the asset immediately contradicting the optimality of $r \leq c$. The following discussion therefore assume that $r > c$.

The fraction μ of fast traders plays two important roles. First, a positive fraction of fast traders eliminates the possibility of a situation in which all dealers in the bilateral market charge the monopoly price r. Such a situation is also known as the Diamond (1971) paradox. Second, fast traders connect the two trading venues, since they search for the best price in the overall market. Thus, a dealer making a decision on either trading venue has to consider what is happening on the other trading venue.

If slow traders use a reservation price strategy with reservation price $r > c$ and probability of continued search $\gamma > 0$ as described in Definition 1, standard search-theoretic arguments like those in Varian (1980b) and Duffie et al. (2017) can be used to derive some properties of the dealers' strategies (see Lemma 3 in Appendix 1.8.1). Dealers will not quote prices greater than the slow traders' reservation price r, because they will not be able to sell the asset by doing so. It now follows that slow traders will buy the asset as soon as they receive a quote. This means that one can calculate the masses of slow traders that a dealer faces in the bilateral market or on the platform, respectively. By Definition 1, the mass of slow traders k_p that a dealer on the platform faces is given by

$$k_p := 1 - \mu. \tag{1.1}$$

Slow traders will only continue to search in the bilateral market if they did not get a quote on the platform, which happens with probability $(1-\eta)^N$. In this case, a slow trader chooses each dealer in the bilateral market with equal probability. Thus, the mass of slow traders that a dealer in the bilateral market faces is given by

$$k_b := (1-\eta)^N \gamma (1-\mu) \frac{1}{N}. \tag{1.2}$$

It turns out that there cannot be any atoms in the distributions G and H according to which dealers quote on the two trading venues. Intuitively, a dealer would always try to avoid ties with other dealers if there were atoms in G or H at prices above their cost c. An atom a price equal to the dealers' cost c can be ruled out since dealers would not earn a profit in this case, while a positive profit can be achieved by quoting r all the time and sell to the strictly positive mass k_p or k_b, respectively, of slow traders in the market.

Dealers are only willing to quote random prices if the expected payoffs from quoting these prices are the same. In the proof of Lemma 4 in the appendix it is shown that the suprema of the supports of G and

H are both equal to r, i.e. trading desks always want to make use of their ability to charge high markups to slow traders. Thus, the profit from quoting any price in the support of G (H) on the platform (in the bilateral market) must thus be equal to the profit that results from quoting r, the supremum of support G and support H, respectively. Formally, this means that for every p in the support of H it has to hold that

$$(p-c)\left[k_b+\mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right]=(r-c)k_b. \tag{1.3}$$

Analogously,

$$(p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p+\mu(1-H(p))^N\right]=(1-\eta)^{N-1}(r-c)k_p \tag{1.4}$$

has to hold for every p in the support of G.

The intuition behind the condition in equation (1.3) is similar to the intuition in the case of a pure bilateral market as in Duffie et al. (2017). One option for the dealer is to quote the slow traders' reservation price r and get only slow traders as customers, since the fast traders almost surely get a better offer elsewhere. The expected profit in this case is expressed by the right-hand side of equation (1.3). Another option is to lower the price in order to be able to sell to slow traders and a positive measure of fast traders. The expected profit in this case is expressed by the left-hand side of equation (1.3). There is one difference between (1.3) and an analogous condition for a pure bilateral market (without coexisting platform). In a hybrid market, a dealer also has to take the probability $(1-\eta G(p))^N$ of not being undercut by dealers on the platform into account.

Equation (1.4) expresses a similar trade-off for the dealers on the platform. A dealer on the platform can only sell, if he is not undercut by other dealers on the platform. This event happens with probability $(1-\eta G(p))^{N-1}$. Then, the dealer has to be indifferent between selling only to slow traders at price r and being able to sell also to a positive measure of fast traders, but at a lower price.

1.3 Equilibrium

Having characterized the traders' and dealers' strategies, an equilibrium can be constructed. This means that the slow traders' reservation price r and their probability of continued search γ needs to be endogenized. Also, it needs to be determined when following a reservation price strategy in the sense of Definition 1 is indeed optimal for slow traders.

The slow traders' optimal behavior is determined as follows. The reservation price r must be such that a slow trader prefers buy the asset at a price below r to continuing the search. Since slow traders start their search on the platform and at most one dealer is contacted in the bilateral market, the search will always be continued in the bilateral market. Similarly, a slow trader prefers continuing the search in the bilateral market to buying the asset at a price which is greater than r. If being offered a price equal to the reservation price r, a slow trader is indifferent between taking the offer and continuing the search in the bilateral market:

$$\underbrace{v - r}_{\text{payoff from taking offer immediately}} = \underbrace{v - \mathbb{E}\left(\min(r, p_b)\right) - s_b,}_{\text{expected payoff from searching}} \tag{1.5}$$

where p_b denotes the price a dealer in the bilateral market quotes when contacted, a random variable with distribution H. A trader always receives the value of the asset v when buying the asset. When deciding to search instead of taking the offered price equal to r, a slow trader almost surely receives a better price, since dealers do not give quotes greater than r. The expected price a trader has to pay after contacting a new dealer is therefore lower than r. However, the slow trader has to exert the search effort s_b. The reservation price r makes (1.5) hold, i.e.

$$r = s_b + \int_c^r p \,\mathrm{d}H(p). \tag{1.6}$$

Slow traders will always continue to search in the bilateral market after not having received a quote on the platform, i.e. choose $\gamma = 1$, if it is profitable to do so. Searching in the bilateral market is profitable for slow traders if the right-hand side of (1.5) is positive. Now, (1.5) implies that slow traders choose $\gamma = 1$ whenever $v > r$. Analogously, slow traders are willing to choose any probability of continued search in $[0, 1]$, if $r = v$. If the reservation price r as defined in (1.6) is greater than v, slow traders do not want to enter the bilateral market. However, if only fast traders are in the bilateral market, the resulting Betrand competition would make dealers quote their cost c, in turn implying $r = c + s_b < v$. Thus, there is no equilibrium in which slow traders continue searching in the bilateral market with probability $\gamma = 0$. Either $r > 1$ and $\gamma = 1$ or $r = v$ and $\gamma \in (0, 1]$. The following definition summarizes the above reasoning and defines the notion of the equilibrium in the HM considered in this paper.

Definition 2. *The equilibrium in the HM is defined by the following conditions:*

- *Fast traders visit every dealer in the bilateral market and on the platform and buy from the dealer that offered the best quote.*

- *Slow traders find it optimal to follow a reservation price strategy according to Definition 1. The reservation price r solves (1.5). The slow traders' choice of γ satisfies*

$$\gamma = \begin{cases} 1, & \text{if } r < v \text{ or} \\ x \in (0,1], & \text{such that } r = v. \end{cases}$$

- *The dealers quote according to continuous distributions that satisfy (1.3) and (1.4) on their respective suppports. The suprema of the supports of these distributions are equal to r.*

Trading desks have different incentives to compete with each other depending on whether they operate in the bilateral market or on the platform. In the bilateral market, there is only one reason why dealers compete with each other: the presence of fast traders who can compare quotes of different dealers. On the platform, there is, besides the presence of fast traders, another source of competition: Since a dealer on the platform can only sell if no other dealer provides a better quote, the RFQ trading protocol incentivizes dealers to undercut each other. However, since slow traders start their search on the trading platform and often do not even enter the bilateral market, the relative amount of fast traders compared to slow traders asking for quotes is lower on the platform. Thus, although dealers on the platform compete with each other through the RFQ trading protocol, they still have a relatively high incentive to extract high rents from slow traders who will buy the asset at the first price they receive. This reasoning is summarized in Table 1.1.

Table 1.1: **Incentives for dealers to quote low prices to attract fast or slow traders.**

trader type	Platform	Bilateral market
slow traders	**low incentives:** RFQ incentivizes to compete, but slow traders always buy	**no incentives:** slow traders always buy
fast traders	**moderate incentives:** RFQ incentivizes to compete, competition with bilateral market	**high incentives:** competition with other trading desks, few slow traders in the market

Looking at Table 1.1, competition should be higher in the bilateral market if there are many fast traders, i.e. if μ is high: In this case, competition is moderate on the platform and high in the bilateral market. If the fraction of fast traders μ is low, competition is low on the platform as opposed to basically nonexistent in the bilateral market. Besides establishing the existence of an equilibrium in the sense of Definition 2, the following proposition confirms the intuitive reasoning on competition, if an inverse measure of competition is given by a trading desk's profits per quote. This means that competition is said to be higher if a trading desk's profits divided by the mass of its quotes sent to customers are lower.

Proposition 1. *For any combination of the exogenous parameters N, v, c, s_b, s_p, μ and η, there is a positive threshold $\overline{s} > 0$ such that a unique equilibrium in the HM as characterized in Definition 2 exists if and only if $s_p \leq \overline{s}$. Moreover, there exists a threshold $\overline{\mu} > 0$ that depends only on N and η such that the following statements are true for any equilibrium:*

(i) If $\mu > \overline{\mu}$, profits per quote are higher on the trading platform than in the bilateral market.

(ii) If $\mu \leq \overline{\mu}$, profits per quote are higher on the trading platform than in the bilateral market if and only if the search cost in the bilateral market s_b is sufficiently high.

A high search cost s_b of searching in the bilateral market intensifies competition for fast traders in the bilateral market, since many slow traders will stay out of the bilateral market if searching there is costly. Therefore, part (ii) of Proposition 1 states that competition on the platform is higher on the the platform if and only if μ and s_b are small. Figure 1.1 illustrates for which specific parameter choices competition must be higher in the bilateral market and for which parameter choices competition can be higher on the platform. Profits per quote can be higher in the bilateral market only for very low values of μ. This result shows that a common intuitive reasoning that competition in electronic markets is automatically higher might be too simplistic. Since slow traders endogenously choose trading venues, competition is for most parameter choices lower in the market with otherwise more incentives to compete.

In Definition 2 it has been assumed that slow traders start their search on the trading platform. Doing so is only optimal for slow traders if the cost of searching on the platform is not too high. For this reason, Proposition 1 states that an equilibrium as described in Definition 2 exists if and only if the cost of searching for a quote on the platform is not too high. In particular, an equilibrium of the type described in Definition 2 always exists if $s_p = 0$. If s_p is very high, any equilibrium, in case one exists, must be different from the type of equilibrium described in Definition 2.

In the following, some properties of the equilibrium in the HM are established. A statement of the next proposition will use the notation

$$\underline{q} := \inf\left[\text{support}(G)\right] \qquad \text{and} \qquad \underline{b} := \inf\left[\text{support}(H)\right] \tag{1.7}$$

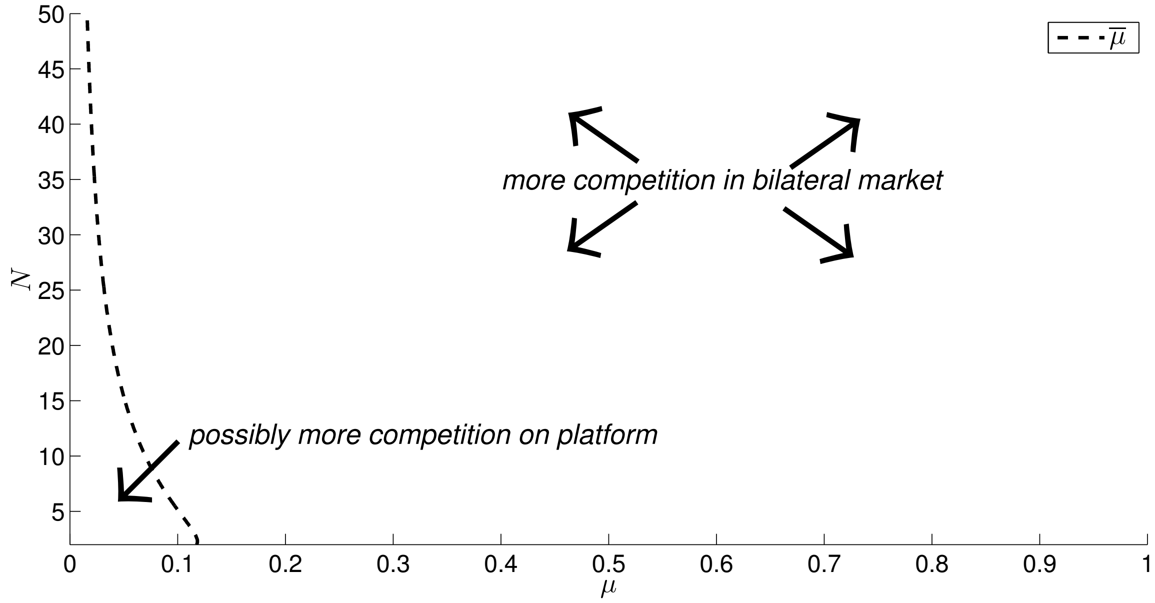


Figure 1.1: **Parameter space.** This figure shows the threshold $\overline{\mu}$ referred to in Proposition 1 with $\eta = 0.2$ and various combinations of the number of dealers N and the fraction of fast traders μ. Given that s_p is sufficiently low so that an equilibrium exists, profits per quote are higher on the platform for all (μ, N) pairs to the right of the dashed line. For all (μ, η) pairs to the left of the dashed line, profits per quote are higher in the bilateral market if and only if s_b is sufficiently small.

to link profits per quote to the smallest markup the trading desks are willing to quote. In the fifth statement of the following proposition, q refers to the random variable that represents the best quote on the platform conditional on at least one response by one dealer to an RFQ. Furthermore, p_b refers to the random variable that represents the price a dealer quotes when contacted in the bilateral market.

Proposition 2. *For any equilibrium in the HM that is described by Definition 2, the following holds:*

(i) ***On the choice of trading venue:*** *Fast traders are more likely to trade in the bilateral market than slow traders are. Slow traders are more likely to trade on the platform than fast traders are.*

(ii) ***On profits per quote:*** *A trading desk's profits per quote are higher on the platform than in the bilateral market if and only if $\underline{b} < \underline{q}$.*

(iii) ***On average markups for slow traders:*** *The expected price $\mathbb{E}(q)$ a slow trader has to pay on the platform conditional on at least one response by a dealer satisfies the following inequality:*

$$\mathbb{E}(q) \leq \mathbb{E}(p_b) + s_b - \frac{s_p}{1-(1-\eta)^N}. \tag{1.8}$$

The first statement of Proposition 2 describes how the hybrid market separates fast and slow traders to some extent: fast traders execute a larger proportion of their trades in the bilateral market than slow traders do. The reason for this behavior is that fast traders make use of their ability to costlessly search the entire market for the best quote, whereas slow traders often terminate their search on the platform.

The second statement of Proposition 2 states the following. Whether profits per quote are higher on the trading platform or in the bilateral market can also be determined based on the supports of the quoting strategies G and H. While it has been derived in Section 3.6 that the suprema of the quoting strategies' supports are equal to the slow traders' reservation price r, nothing has been said about the respective lower bounds. The result in the second statement of Proposition 2 has a simple intuition: Suppose $\underline{b} < \underline{q}$. Then, a dealer on the platform can decide to quote $\underline{b}$ at each RFQ response and expect the same profit per quote as a dealer in the bilateral market expect when quoting $\underline{b}$, since such a dealer would expect to sell to all traders requesting quotes. However, because $\underline{b} < \underline{q}$, the dealer on the platform optimally never quotes $\underline{b}$ since this would lower the dealer's profit. Thus, the profits per quote must be higher on the platform. An analogous reasoning applies to the case in which $\underline{b} \geq \underline{q}$.

The intuition for the third statement of Proposition 2 is as follows. If slow traders find it optimal to start their search on the platform, the expected quote is better than the expected quote in the bilateral market, adjusting for search costs. The search cost on the platform is adjusted to account for the possibility that there is no response from dealers.

The following proposition compares the HM equilibria to equilibria in the PBM from Duffie et al. (2017).

Proposition 3. *Let the exogenous parameters N, v, c, s_b, s_p, η and μ be fixed. Moreover, let $\overline{s}$ be defined as in Proposition 1 and let $s_p < \overline{s}$ such that a unique equilibrium in the HM that is described by Definition 2 exists (see Proposition 1). Then, the following is true for the equilibrium in the HM and the equilibrium in the PBM described in Section 1.1:*

(i) ***On reservation prices:*** *The slow slow traders' reservation price in the PBM is at least as large as the reservation price of the slow traders in the HM, i.e. $r^{PBM} \geq r$.*

(ii) ***On slow traders' probability of bilateral search:*** *The equilibrium values of γ and γ^{PBM} satisfy the following.*

- *If $\gamma^{PBM} = 1$, then $\gamma = 1$.*

- *If* $\gamma^{PBM} < 1$, *then*

$$\gamma = \begin{cases} 1, & \text{if } r < v, \\ x \in (0,1], & \text{otherwise,} \end{cases}$$

where x *is determined such that* $r = v$ *and satisfies* $x > \gamma^{PBM}$.

(iii) ***On total turnover:*** *Slow traders trade at least as much in the HM as they do in the PBM. If the market participation* γ^{PBM} *in the PBM is less than 1, slow traders trade strictly more in the HM compared to the PBM. Thus, since fast traders always trade, total turnover in the HM is at least as high as total turnover in the PBM and total turnover in the HM is strictly greater than total turnover in the PBM if* $\gamma^{PBM} < 1$. *Moreover, turnover is strictly higher in the HM or* $r > r^{PBM}$.

(iv) ***On turnover in the bilateral market:*** *Slow traders are less likely to request a quote in the bilateral market of the HM than they are in the PBM, i.e.* $\gamma(1-\eta)^N < \gamma^{PBM}$ *holds. Also, fast traders are less likely to request a quote in the bilateral market of the HM than they are in the PBM. Thus, turnover in the bilateral market of the HM is lower than in the PBM.*

(v) ***On profits per quote in the bilateral market:*** *The infimum of the quoting strategy's support in the HM is lower than infimum of the quoting strategy's support in the PBM:*

$$\underline{b} < \underline{p}^{PBM}. \tag{1.9}$$

This implies that profits per quote in the bilateral market of the HM are lower than profits per quote in the PBM.

(vi) ***On average markups:*** *The expected price a fast trader is paying for the asset is lower in the HM than in the PBM. The expected price a slow trader has to pay in the bilateral market of the HM is not greater than the expected price a slow trader has to pay in the PBM.*

The intuition for statement (i) in Proposition 3 is that more competition in the bilateral market of the HM, compared to the PBM, can never make continuing to search less attractive in the HM than in the PBM. Thus, the introduction of an HM never increases the slow traders' reservation price, since an increase in the reservation price makes slow traders more likely to accept given offers and stop searching. The same intuition applies for statement (ii): If entering bilateral market with a certain probability was worth doing in the PBM, it should also be worth doing in the HM, where dealers compete more. A consequence of the slow traders' higher probability of continuing to search in the bilateral market is that total turnover in the HM is greater than or equal to the turnover in the PBM, which is the claim in statement (iii) in Proposition 3. However, as described in statement (iv), slow traders' market participation does not increase sufficiently to also make turnover in the bilateral market of the HM larger than the turnover in the PBM. That total turnover in the HM is larger than in the PBM is due to the relatively large volume traded on the platform.

Statement (v) in Proposition 3 is reminiscent of statement (ii) in Proposition 2 linking the profits per quote to the lower bounds of the quoting strategies' supports. Competition, inversely measured by profits per quote, is more intense in the bilateral market of the HM than in the PBM, since trading desks in the bilateral market of the HM compete with trading desks on the platform. These differences in the amount of competition are reflected in the relative magnitudes of the lower bounds of the quoting strategies' supports. The same intuition that was given for statement (ii) in Proposition 2 also applies here: a dealer in the PBM could obtain the same profits per quote as a dealer in the bilateral market of the HM by quoting $\underline{b}$ all the time, but optimally choses not to do so.

Higher competition in the bilateral market of the HM enables fast traders to obtain on average a better price for the asset in the HM compared to the PBM, as claimed in statement (vi) in Proposition 3. Without further knowledge on the parameters of the model, it cannot be assessed whether slow traders also obtain on average a lower price in the HM than in the PBM, since slow traders execute a large portion of their trades on the trading platform. Theoretically, it is possible that average prices on the platform are much higher than in the PBM or the bilateral market of the HM and slow traders nevertheless prefer trading on the platform if the search cost s_p on the platform is very low compared to the search cost s_b in the bilateral market. However, conditional on trading in the bilateral market, slow traders will on average not pay more for the asset in the HM than in the PBM. Figure 1.2 illustrates most of the properties of HM equilibria derived so far in a concrete numerical example.

The slow traders' market participation in the PBM, γ^{PBM}, is less than 1 in both equilibria considered in Figure 1.2 and determined such that $r^{PBM} = 1$. Introducing a trading platform lowers the reservation price such that $\gamma = 1$ in both equilibria, consistent with statements (i) and (ii) in Proposition 3.

In Panel A of Figure 1.2, approximately 0.3% of fast traders and 71% of slow traders buy the asset on the platform. On the other hand, approximately 99.7% of fast traders and 29% of slow traders buy the asset in the bilateral market. Thus, fast traders mainly trade in the bilateral market, while slow traders mainly trade on the platform. In Panel B of Figure 1.2, approximately 88.6% of fast traders and 89.3% of slow traders buy the asset on the platform. On the other hand, approximately 11.4% of fast traders and 10.7%

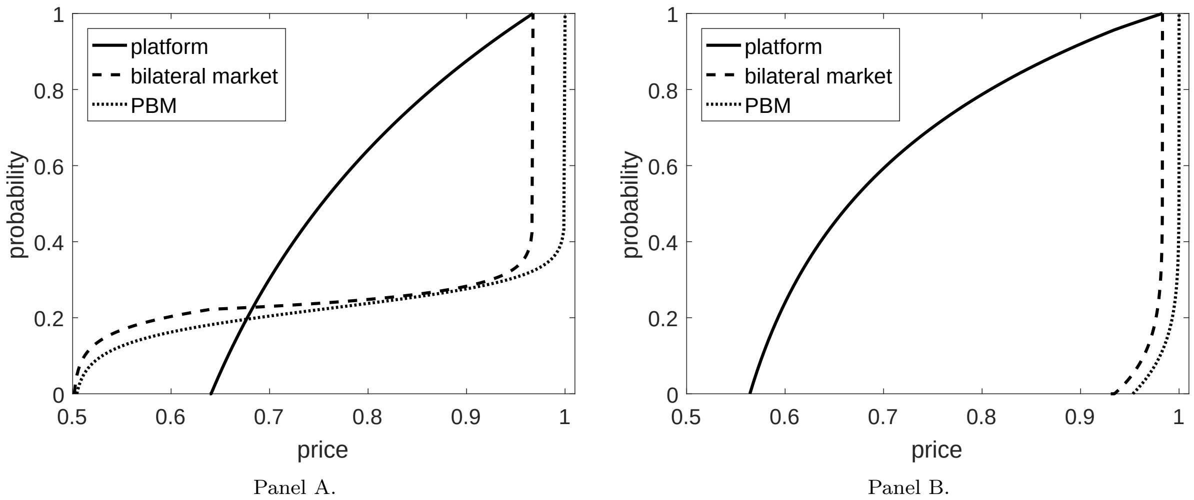


Figure 1.2: **Numerical example.** The distributions in Panel A correspond to HM and PBM equilibria for the following exogenous parameters: $N = 20$, $v = 1$, $c = 0.5$, $s_b = 0.11$, $s_p = 0$, $\mu = 0.8$ and $\eta = 0.06$. Endogenous parameters are $\gamma = 1$, $\gamma^{PBM} \approx 0.692$, $r \approx 0.968$ and $r^{PBM} = 1$. The distributions in Panel B correspond to HM and PBM equilibria for the following exogenous parameters: $N = 10$, $v = 1$, $c = 0.5$, $s_b = 0.005$, $s_p = 0$, $\mu = 0.01$ and $\eta = 0.2$. Endogenous parameters are $\gamma = 1$, $\gamma^{PBM} \approx 0.975$, $r \approx 0.983$ and $r^{PBM} = 1$.

of slow traders buy the asset in the bilateral market. Thus, both fast and slow traders mainly trade on the platform. However, fast traders execute a higher fraction of their trades in the bilateral market than slow traders do, as stated in statement (i) in Proposition 2. Profits per quote are higher on the platform than in the bilateral market of the HM in Panel A of Figure 1.2 and higher in the bilateral market of the HM than on the platform in Panel B. This fact can be observed from the lower bounds of the supports of the respective quoting strategies, as claimed in statement (ii) of Proposition 2.

1.4 Endogenous Market Design

Now that the equilibrium in the hybrid market is characterized, one can deal with the question of when to introduce electronic trading platforms in OTC markets. When trying to find an answer to this question, different perspectives can be taken:

It turns out that both fast and slow traders are better off in in the HM equilibrium, described in Definition 2, than in the corresponding equilibrium in the PBM. However, dealers may be better or worse off in the HM than in the PBM, depending on whether higher turnover in the HM makes up for potentially lower average markups. Moreover, aggregate welfare may be higher or lower in the HM compared to the PBM, depending on whether more realized benefits of trade between dealers and traders make up for potentially higher search effort by slow traders.

1.4.1 Welfare

Revisiting the numerical example corresponding to Figure 1.2 in Section 3.6.2, one finds that it would be inefficient from a social point of view, if an HM were not introduced. This has two reasons. First, there is a higher turnover in the HM, which has an increasing effect on welfare, since each trade is beneficial. Second, since there is no search cost on the platform, search effort by slow traders is reduced, which also has an increasing effect on welfare.

Let total welfare Ω be defined as the sum of the benefits of trade and the negative search effort exerted by slow traders, i.e.

$$\begin{aligned} \Omega \;:=\; & \underbrace{(1-\mu)\left[\left(1-(1-\eta)^N\right)(v-c)+\gamma(1-\eta)^N(v-c)\right]}_{\text{benefits from trades with slow traders}} \\ & -\underbrace{(1-\mu)(s_p+\gamma(1-\eta)^N s_b)}_{\text{total search effort}} \\ & + \underbrace{\mu(v-c)}_{\text{benefits from trades with fast traders}}. \end{aligned} \tag{1.10}$$

The first term in (1.10) refers to the net benefits of trade generated by slow traders. Slow traders always pay the cost of going to the platform. Then they obtain the asset with probability $1-(1-\eta)^N$ or visit the bilateral market. The second term refers to the total search effort exerted by slow traders who always are always paying the cost of searching on the platform s_p and may also pay the cost s_b in case they decide to also search in the bilateral market. Fast traders always generate net benefits $v-c$ per trade which is expressed in the third term in (1.10).

In an analogous way, the welfare in the PBM Ω^{PBM} can be expressed as the sum of the benefits of trade generated by trades with slow traders $(1-\mu)\gamma^{PBM}(v-c-s_b)$, an adjustment for the search effort and and the benefits of trade generated by trades with fast traders $\mu(v-c)$.

$$\Omega^{PBM} := (1-\mu)\gamma^{PBM}(v-c-s_b)+\mu(v-c). \tag{1.11}$$

When determining whether to mandate a hybrid market structure instead of letting market participants trade in a pure bilateral market, a regulator should consider the following effects that the introduction of a platform has on welfare:

- **market participation effect:** As statement (ii) in Proposition 3 states, market participation in the HM is never lower than in the PBM and strictly higher whenever market participation γ^{PBM} in the PBM is less than 1. Since traders value the asset more than dealers do, higher market participation results in more realized benefits of trade.

- **market fragmentation effect:** Slow traders do not find a satisfactory quote on the platform if none of the contacted dealers respond to the RFQ. Since dealers independently respond with some probability $\eta < 1$,the trader will not find a satisfactory quote on the platform with strictly positive probability $(1-\eta)^N$. A slow trader then keeps searching in the bilateral market with strictly positive probability γ and exerts search effort s_b. Thus, some slow traders exert some search effort twice in the HM whereas they only exert effort once in the PBM. This market fragmentation effect may increase welfare, if the cost of searching in the electronic market s_p is low, but decrease welfare if s_p is high.

A practical decision rule can be found in the following proposition, which says that an HM is never inefficient if dealers prefer the HM. The claim follows from the fact that slow traders are better off in the HM than in the PBM. Since fast traders always get better prices in the HM, which has been stated in Proposition 3, all market participants are better off in the HM if dealers are also better off in the HM.

Proposition 4. *Let the exogenous parameters be fixed such that $s_p \leq \overline{s}$, with $\overline{s}$ defined as in Proposition 1. Then an equilibrium in the HM exists. If dealers prefer the HM to the PBM, the HM is efficient, because slow traders are always better off in the HM.*

Moreover, an HM equilibrium exists and the HM is efficient if one of the following conditions holds, where $\overline{s}$ is again defined as in Proposition 1.

1. $N \to \infty$, *holding other parameters constant with* $s_p < v - c$,
2. $\mu \to 0$, *holding other parameters constant with* $s_p < \overline{s}$,
3. $s_b \to v - c$, *holding other parameters constant with* $s_p < \overline{s}$,
4. $\eta \to 1$, *holding other parameters constant with* $s_p < \overline{s}$,
5. $s_p < \min\left[(1-(1-\eta)^N)s_b, \overline{s}\right]$.

Either an HM equilibrium as described in Definition 2 does not exist or a PBM is efficient if one of the following conditions holds.

6. $s_p > 0$ *and* $s_b \to 0$,

7. $s_p > (1-(1-\eta)^N)s_b$ *and* $\mu \to 1$,

8. $\eta \to 0$ *and* $s_p > 0$.

Sometimes, dealers seem to be opposed to an introduction of electronic trading platforms in OTC markets, as illustrated in the anecdote on the CDS market mentioned in the introduction. In those cases, an HM may still be the efficient market structure, which is optimally mandated by the regulator. Proposition 4 also mentions conditions under which a regulator optimally mandates a hybrid market structure irrespective of what dealers prefer. Given that the search cost on the platform is not too high (this condition ensures the existence of an equilibrium and that the market fragmentation effect does not excessively increase search effort of slow traders), the HM is efficient if $N \to \infty$, $\mu \to 0$, or $s_b \to v - c$. This is because all of the three latter conditions result in $\gamma^{PBM} \to 0$ and the introduction of an electronic trading platform leads to a jump in the slow traders' market participation. The economic intuition for this result goes as follows. As the number of dealers N becomes large, dealers have a low chance of offering the best quote in the entire market. Thus, in equilibrium, dealers focus to a large extent on extracting high rents from slow traders as $N \to \infty$ and many slow traders will not enter the market, since doing so is not attractive in this case. Similarly, competition among dealers is low as the fraction of fast traders μ is small and many slow traders stay out of the market in equilibrium. If the search cost of slow traders s_b is large, dealers can only sell to slow traders if their markups are sufficiently low. This can only be the case in equilibrium if many slow traders do not enter the market and thus increase competition through the higher ratio of fast to slow traders asking for quotes.

If $\eta \to 1$, as required in condition 4 in Proposition 4, slow traders buy the asset on the platform that goes to 1. Now welfare in the HM is higher than in the PBM if the search costs on the platform are sufficiently low. Moreover, the platform can always be used to decrease search effort by slow traders if search costs on the platform are sufficiently low. This is why the HM is efficient under condition 6.

The PBM is efficient if slow traders already fully participate in the market and a platform would only increase search effort, i.e. if s_p is sufficiently high. Slow traders will fully participate fully in the PBM, if searching is easy ($s_b \to 0$) or if fast traders ensure that competition leads to favorable markups ($\mu \to 1$). The introduction of a trading platform would only increase search effort for any positive search cost s_p if the chance of obtaining the asset after an RFQ is low enough, as stated in condition 8. in Proposition 4.

Note that by Proposition 3 and Proposition 4, traders are always better off in the HM. A regulator who is only concerned about investor welfare may therefore introduce an HM even if $\Omega < \Omega^{PBM}$. In this case however, traders gain less from the introduction of an HM than dealers lose due to lower markups. The reason for this is that slow traders' search effort in the HM outweighs any additional benefits of trade due to higher market participation.

Proposition 4 states that an HM is efficient whenever the dealers prefer the HM to the PBM. If the converse would also be true, a regulator who wants to maximize welfare could just leave the decision whether to mandate a hybrid market structure or keep a bilateral market up to the dealers. However, as will be discussed in the following subsection, dealers may sometimes prefer the PBM even though an HM is efficient. In these cases, a regulator could improve welfare by mandating electronic trading.

1.4.2 The dealers' perspective, the need for regulation and other effects of introducing a platform

Dealers prefer the HM to the PBM, if their joint profits are higher in the HM than in the PBM.[13] Since there are N dealers in total, the indifference conditions (1.3) and (1.4) give the following collective profit Π_{HM} in the HM.

$$\Pi = \underbrace{N(r-c)k_b}_{\text{profits from bilateral market}} + \underbrace{N\eta(r-c)k_p(1-\eta)^{N-1}}_{\text{profits from platform}}. \tag{1.12}$$

The first term on the left-hand side of (1.12) expresses the profit the trading desks in the bilateral market makes by always quoting the slow traders' reservation price r. This value is equal to the trading desks' actual profit since trading desks are indifferent between any price they are willing to quote. The second term on the left-hand side of (1.12) expresses the profit a trading desks on the platform make by always quoting r, conditional on quoting at all. Analogously to the argument for the bilateral market, this value must be equal to the trading desks' actual profit.

Also in the PBM, dealers are indifferent between quoting r^{PBM} and quoting any other price in the support of H^{PBM}. The dealers' aggregate profit in the PBM is thus given by

$$\Pi^{PBM} = (r^{PBM} - c)\gamma^{PBM}(1-\mu). \tag{1.13}$$

With these expressions for the dealer profits under the two market structures, one can revisit the numerical

[13]Since dealers are identical, joint dealer profits are higher in the HM if and only if individual dealer profits are higher in the HM.

example considered in Figure 1.2. In Panel A, dealers make a profit of approximately 0.069 in the PBM compared to 0.062 in the HM. In Panel B, dealers make a profit of approximately 0.482 in the PBM compared to 0.180 in the HM. Thus, despite a higher turnover in the HM, dealers earn less in in the HM in both cases, because transaction prices are lower on average. In general, there are two main effects the introduction of an electronic trading platform has on dealer profits:

- **Volume effect:** This effect is analogous to the market participation effect discussed in the previous subsection. The introduction of an electronic trading platform always weakly increases the slow traders' market participation, which results in a higher trading volume. This volume effect always positively affects dealer profits.

- **Price effect:** Due to the increased competition in the HM, average markups that dealers charge their clients may decline. This effect potentially harms dealer profits.

Thus, dealer profits are highest in the PBM if and only if a negative price effect dominates the weakly positive volume effect of the introduction of the trading platform. The formal conditions under which this is the case are stated in the next proposition.

Proposition 5. *An HM equilibrium exists and the HM is preferred by the dealers if one of the following conditions holds, where $\overline{s}$ is defined as in Proposition 1.*

1. $N \to \infty$ *and* $\eta < \frac{s_b}{v-c}$*, holding other parameters constant with* $s_p < v - c$*,*

2. $\mu \to 0$*, holding other parameters constant with* $s_p < \overline{s}$*,*

3. $s_b \to v - c$*, holding other parameters constant with* $s_p < \overline{s}$*,*

Either an HM equilibrium as described in Definition 2 does not exist or a PBM is preferred by the dealers if one of the following conditions holds.

4. $s_b \to 0$*,*

5. $\mu \to 1$*,*

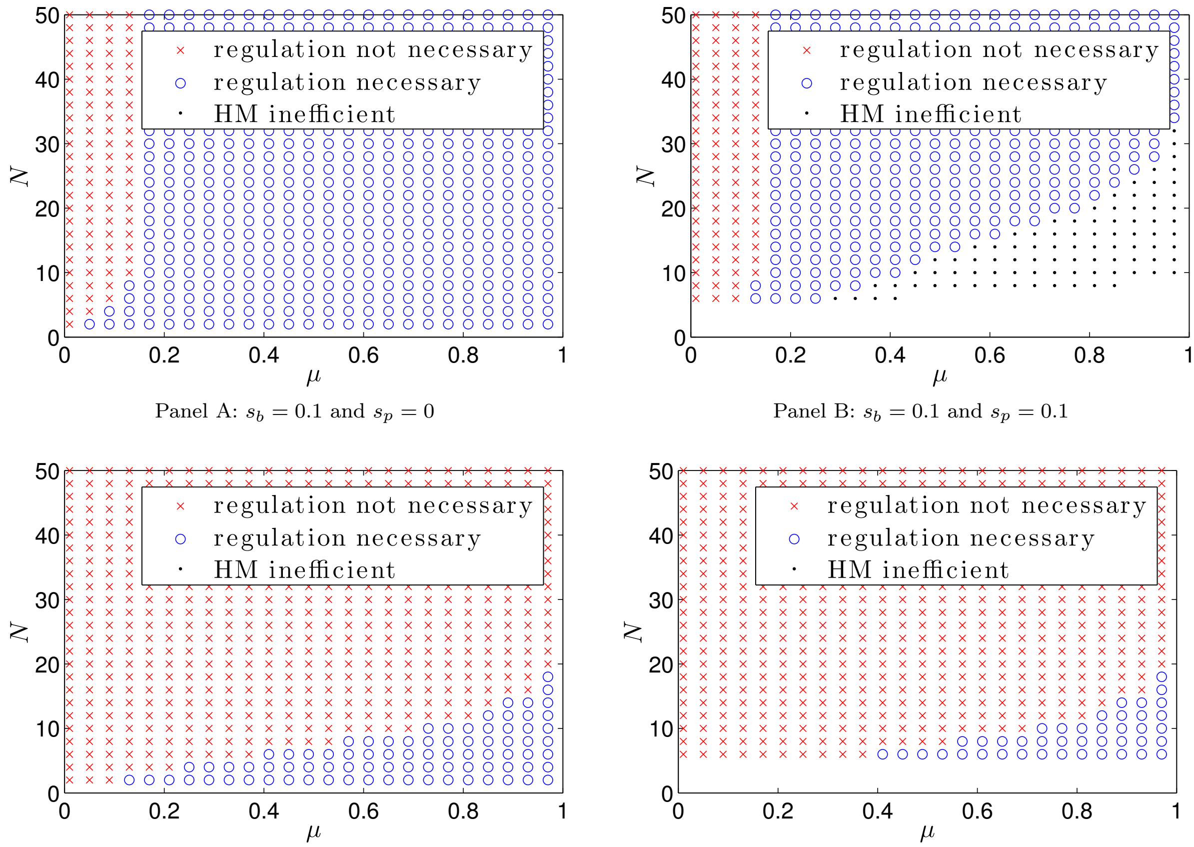
6. $N \to \infty$ *and* $\eta > \frac{s_b}{v-c}$*,*

The intuition behind the conditions in Proposition 5 is analogous to that behind the conditions in Proposition 4. In general, dealers prefer the HM if the introduction of a platform increases turnover significantly, i.e. if the slow traders market participation in the PBM γ^{PBM} is very low. The latter will be the case if the cost of searching in the bilateral market s_b is very high, the fraction fast traders μ is very low or the number of dealers N is very large, as described in the discussion after Proposition 4. However, if the number of dealers is very large, competition on the platform is also very high. Thus, if N is very high, dealers only prefer the HM, if the dealers response rate η is below a certain threshold to ensure that competition is not too high. If η is too high and N is large, dealers prefer the PBM, because competition on the platform would be too intense. Dealers also prefer the PBM, if slow traders already participate actively in the market. This is the case if s_b is low or μ is high, as discussed in the text following Proposition 4.

A platform can be seen as a device that allows dealers to commit to lower markups in order to make more slow traders enter the market. Since slow traders have to make their market entry decision before dealers decide which price to quote, the slow traders' search cost is already a sunk cost when the dealer has to make a decision. This may incentivize the dealer to quote markups ex post that are larger than the markups that would maximize dealer profits ex ante. Duffie et al. (2017) argue that benchmarks may alleviate this problem by reducing information asymmetries with respect to the dealers' costs of providing the asset. This paper takes a different perspective: a platform increases bargaining power of the dealers' customers. Introducing a platform may be beneficial even if information with respect to dealers' costs is perfectly symmetric, as in this paper.

Figure 1.3 illustrates the effect of the different parameters on the need for regulation. Consistent with Propositions 4 and 5, dealers efficiently want to use electronic trading protocols if μ is small, which can be observed in all four panels. In Panel C and D it is observable that dealers prefer the HM as the number of dealers N becomes large. This is the case because the RFQ response rate η is sufficiently low. In the cases for which the regulator's and the dealers' preferences are aligned, a regulator could leave the decision whether to introduce electronic trading platforms up to the dealers For lower values of N and larger values of μ, a dealer may want to mandate electronic trading unless the costs of searching in the electronic market are too high, as shown in Panel B. Comparing Panels A and B to Panels C and D, one can observe that dealers are more willing to introduce a platform if searching in the bilateral markets is costly, i.e. s_b is high. Comparing Panels A and C to Panels B and D, one can observe that an HM is inefficient only if searching in the electronic market is relatively expensive, compared to searching in the bilateral market. An equilibrium exists for most parameter choices considered in all four panels.

Whereas Propositions 2 and 3 stated some general properties of the HM and comparisons to the PBM, the aim of the following proposition is to asses under which conditions changes in certain trading patterns are large or small when a trading platform is introduced into a PBM.



Panel A: $s_b = 0.1$ and $s_p = 0$

Panel B: $s_b = 0.1$ and $s_p = 0.1$

Panel C: $s_b = 0.2$ and $s_p = 0$

Panel D: $s_b = 0.2$ and $s_p = 0.1$

Figure 1.3: **When to regulate markets?** The red crosses indicate where the welfare-maximizing regulator's preferences are aligned with the dealers' preferences. The blue circles indicate the cases in which the regulator needs to mandate the electronic trading and the black dots indicate when the HM is inefficient. Extended white areas indicate cases in which the equilibrium described in Definition 2 does not exist. In all four panels, the parameters not mentioned in the figure are chosen as follows: $\eta = 0.2$, $v = 1$ and $c = 0.5$.

Proposition 6. *Let $s_p = 0$.*[14] *Then an equilibrium in the HM as described in Definition 2 exists and, holding other parameters constant, the following statements are true:*

(i) ***On the increase in turnover:*** *Consider the cases in which $N \to \infty$, $s_b \to v - c$ or $\mu \to 0$. Then, compared to the PBM, turnover in the HM experiences a jump, whose size is specified in the proof of the proposition. However, turnover is the same in HM and PBM if $\mu \to 1$, $s_b \to 0$ or $\eta \to 0$.*

(ii) ***On volume traded on the platform:*** *There are three extreme cases:*

- *As $\eta \to 0$, almost all traders who decide to trade at all trade in the bilateral market.*

[14]This condition is overly restrictive and made for expositional reasons to avoid keeping track of the precise threshold for s_b that ensures the existence of the equilibrium.

- *As $N \to \infty$ or $\eta \to 1$, almost all slow traders trade on the platform. As $s_b \to v - c$ a fraction of $1 - (1 - \eta)^N$ slow traders trade on the platform and almost no slow traders trade in the bilateral market. Almost all fast traders trade in the bilateral market in all three cases.*
- *As $\mu \to 0$, almost all traders who decide to trade trade on the platform*

(iii) ***On average markups for slow traders across trading venues in the HM:***

- *$s_b \to v - c$, expected markups are higher on the platform.*
- *As $N \to \infty$ or $\eta \to 1$, expected markups are higher in the bilateral market.*
- *As $s_b \to 0$, there is no difference in expected markups across trading venues.*

(iv) ***On average transaction prices across trading venues in the HM:***

- *$s_b \to v - c$, transaction prices are higher on the platform.*
- *As $s_b \to 0$, $N \to \infty$ or $\eta \to 1$, transaction prices in the HM are in the interval $[c, c + \varepsilon]$ with probability 1, where $\varepsilon > 0$ is arbitrary.*

(v) ***On average markups for slow traders in the bilateral market:*** *As $\eta \to 1$, expected markups in the bilateral market will jump from some positive number to zero after a platform has been introduced. As $\eta \to 0$ or $s_b \to v - c$, average markups for slow traders remain unchanged in the bilateral market.*

Proposition 6 states that there is a significant increase in turnover after the introduction of a platform if slow traders' market participation in the PBM is low, i.e. $\gamma^{PBM} \approx 0$. The latter is the case if the number of dealers is large ($N \to \infty$), search costs are large ($s_b \to v - c$) or almost all traders are slow ($\mu \to 0$), as has been discussed in the text following Proposition 4. If market entry in the PBM is large because search is easy ($s_b \to 0$) or competition is high because of many fast traders ($\mu \to 1$), a platform does not yield more turnover. Straightforwardly, a platform also does not increase turnover if dealers do not reply to RFQs ($\eta \to 0$), in which case almost all traders who trade at all trade in the bilateral market and average markups compared to the bilateral market do not change.

There is an almost complete separation between fast and slow traders if the number of dealers is large ($N \to \infty$) or dealers reply actively to RFQs ($\eta \to 1$). Then, slow traders execute trades on the platform, while fast traders trade in the bilateral market. With many dealers or high response rates, slow traders very likely get a response to an RFQ and trade on the platform, not continuing to search in the bilateral market. Fast traders however benefit from the competition they can establish in a bilateral market, which is more intense if the number of dealers is large or if dealers on the platform are competitive ($\eta \to 1$). If slow traders find bilateral search very costly ($s_b \to v - c$), they trade on the platform if they can, which is the case with probability $1 - (1-\eta)^N$ and mostly stay out of the bilateral market, while fast traders benefit from the lack of slow traders in the bilateral market that leads to intense competition. If almost all traders are slow ($\mu \to 0$), almost no trader will enter the bilateral market due to the lack of competition in that trading venue.

That slow traders trade more on the platform, as in the cases mentioned in statement (ii) of Proposition 6, is not necessarily related to more attractive prices on the trading platform. For instance, if the main benefit of the platform lies in the reduction of search costs ($s_b \to v - c$), slow traders are better off trading on the platform even though average markups are lower in the bilateral market. As a result, transaction prices are higher on the platform than in the bilateral market. However, if the main benefit of the platform consists in the improvement of the slow traders' bargaining power, e.g. if on average many dealers respond to an RFQ ($N \to \infty$ or $\eta \to 1$), a high share of trades of slow traders is executed on the platform corresponds to lower average markups on the platform. As slow traders become fast ($s_b \to 0$), they make dealers compete more and make differences in markups across trading venues disappear. Indeed, expected markups go to zero in this case, implying that also the variance of transaction prices vanishes. Dispersion in transaction prices also vanishes if the number of dealers N becomes large. Then, slow traders benefit from competition on the platform and fast traders benefit from many opportunities to search in the bilateral market. In general, prices in the bilateral market are also affected by the introduction of the platform. With extreme dealer response rates ($\eta \to 1$), markups go to zero after a platform has been introduced. If dealers are very inactive on the platform ($\eta \to 0$) or search costs s_b are very high, markups do not change. In the latter case, dealers keep charging prices that will make slow traders just indifferent between entering and not entering the market even after a platform has been introduced.

1.4.3 Competition and the role of quoting activity in HM and PBM

Some of the key results of this paper are driven by the fact that the introduction of a trading platform increases competition among the dealers. Since an HM has twice as many trading desks as a PBM, some readers may wonder whether the economic mechanism in this paper can simply be expressed as "more competitors lead to more competition."

Besides yielding the non-trivial result that dealers may actually benefit from more competition (see Proposition 5), the model presented in this paper makes features specific to the RFQ trading protocol explicit. The key characteristics of an HM do not result from increased quoting activity. Instead, the way in which quotes are are sent to different traders determines key properties of the HM and the PBM. In order to show this, I am considering a modified setup in this section.

In the following, I am assuming that a dealer cannot simultaneously quote on the platform and in the bilateral market after a platform has been introduced. Thus, the number of active trading desks will not increase after a platform has been introduced. It will even turn out that the total number of quotes that are provided in the market may be lower in the HM than in the PBM. This suggests that the way dealers compete on the platform is the driving force that leads to an increase in traded volume: In the PBM, a lot of quoting activity is concentrated toward the fast traders, whereas slow traders trade with the first dealer they meet (if they trade at all). However, once a platform is introduced, even slow traders have a chance of obtaining multiple quotes.

Suppose that there are $2N$ dealers in a PBM, with $\mathbb{N} \ni N \geq 2$ as before. If a platform is introduced in this market, a dealer can no longer operate both on the platform and the bilateral markets. Instead, each dealer is exogenously assigned to one of the two trading venues (e.g. by a regulator). Specifically, consider the case in which N dealers quote on the platform and N dealers quote in the bilateral market. Thus, a PBM with $2N$ dealers is compared to an HM with N dealers as previously modeled in this paper.

In the HM, there are ηN quotes on the platform in expectation. Additionally, there are $N\mu$ quotes to fast traders in the bilateral market and $(1-\eta)^N\gamma(1-\mu)$ quotes to slow traders in the bilateral market.

In the PBM, one has $2N\mu$ quotes to fast traders and $\gamma^{PBM}(1-\mu)$ quotes to slow traders. To sum up, one has

$$\text{Quotes} = \begin{cases} N\left(\eta + \mu + k_b\right) & \text{in HM} \\ 2N\mu + \gamma^{PBM}(1-\mu) & \text{in PBM} \end{cases}.$$

The following proposition illustrates that dealers may still want to establish an HM in this modified setup. The introduction of an HM by dealers will increase social welfare in the cases considered below. The intuition for the dealers' and the regulator's preferences for the HM in those cases is the same as in the discussion of Propositions 4 and 5: If $N \to \infty$, $s_b \to v - c$ or $\mu \to 0$, the slow traders market entry is so low that the introduction of a trading platform is desirable from both perspectives. The proof of Proposition 7 has been relegated to Internet Appendix **??**.

Proposition 7. *Under conditions 1-3 in Proposition 5, an equilibrium in the HM as described in Definition*

2 exists, dealers prefer the HM also in the modified setup considered in this section and the HM is the efficient market structure. Moreover, the following is true:

- *Under condition 1 ($N \to \infty$) or under condition 3 ($s_b \to v - c$) in Proposition 5, quoting activity is lower in the HM if and only if $\mu > \eta$.*
- *Under condition 1 ($N \to \infty$) in Proposition 5, the expected best quote on the platform conditional on at least one response is lower than the expected quote in the bilateral market of the HM or PBM: $\mathbb{E}(q) < \mathbb{E}(p_b) = \mathbb{E}(p_b^{PBM})$.*

Most importantly, the first bullet point in Proposition 7 shows that the benefits of the HM do not depend on increased quoting activity. In particular, if the fraction of fast traders μ is large enough relative to the dealers' RFQ response rate η, quoting activity is lower in the HM. The reason for this result is that fast traders can only contact N dealers in the bilateral market of the HM instead of $2N$ dealers in the PBM. Thus, unless the dealers' quoting activity on the platform makes up for this decline in bilateral quoting activity, overall quoting activity in the HM is lower.

As $N \to \infty$, markups on the platform are always lower than markups in the PBM. Thus, the latter effect is also not driven by quoting activity. The intuition for this result is that a large number of dealers N results in intense competition on the platform. For reasons discussed after Proposition 4, however, dealers charge on average high markups in a pure bilateral market as N becomes large.

The reason why dealers can increase the slow traders' market participation in the cases stated in Proposition 7 lies in the way quoting activity is directed towards different traders in the HM and PBM. In the PBM, slow traders receive at most one quote in equilibrium and most of the quoting activity is directed toward fast traders. In the HM, however, even slow traders can receive multiple quotes on the trading platform, which substantially increases their bargaining power and, consequently, their market participation.

1.5 Policy Implications and Empirical Predictions

The theoretical results in Sections 3.6.2 and 1.4 can be used by regulators to improve a given OTC market structure. Furthermore, several testable implications can be derived to either explain empirical observations or evaluate the mechanisms of the model.

1.5.1 Estimating parameters

In order to answer questions about real markets, it is important to know how the parameters of the model can be estimated. Clearly, the number of dealers N, as well as their RFQ response rate can be directly observed or estimated from data that platform operators like MarketAxess, MTS or Tradeweb are willing to share.[15] The reservation price r of slow traders is simply the maximum price which dealers quote in equilibrium in the entire market. This endogenous parameter can also be estimated. Regarding the search cost in the bilateral market s_b, the following inequality is implied by (1.6):

$$s_b = r - \mathbb{E}(p_b).$$

Using the estimator for r suggested above, the last equation means that the search cost s_b is the difference between the maximum and the average of all quotes given in the bilateral market. One may look at transactions prices of slow traders instead of quotes, since both have the same distributions conditional on the trading venue. Using (1.8) in Proposition 2, one can establish an upper bound for the search costs on the platform:

$$s_p \leq \left(1 - (1-\eta)^N\right) \left(\mathbb{E}(p_b) - \mathbb{E}(q) + s_b\right).$$

This upper bound can be estimated from data on quotes, since the average of quotes given in the bilateral market $\mathbb{E}(p_b)$, the average of quotes given on the platform $\mathbb{E}(q)$ and the search cost s_b can be estimated.

When categorizing traders into fast and slow traders, one may use one of at least four possible approaches. The first two approaches use model-independent measures, the last two approaches use implications of the model. First, the identity of the trader can be used. One may distinguish institutional investors from retail investors, since the latter arguably find it more costly to contact several dealers. Second, one can categorize each trade separately, using some measure of demand for immediacy. E.g. a trader has arguably higher costs of waiting (search costs) if he has to execute a fire sale. Third, the ratio of collected quotes to executed trades can be used, which is higher for fast traders in the model. Fourth, trades executed on the platform comprise, according to the model, relatively more trades by slow traders.

RFQ data from platforms will reveal the general trading interest of all traders interested in a given asset in the economy. Even if a trader does not buy the asset, according to the equilibrium in this paper, a trader will at least submit an RFQ via the trading platform. The difference between trading interest and actual

[15] See Hendershott and Madhavan (2015a) for an analysis of such data.

trades may give insights on market participation.

The parameters v and c, i.e. the value of the asset for the traders and the dealers' cost of providing the asset, are the hardest to estimate directly. In general, their estimation would perhaps require a structural approach. However, most of the policy implications do not depend on these parameters.

1.5.2 Policy implications

Proposition 4 specifies conditions under which a regulator could improve welfare by mandating electronic trading. In particular, electronic trading is always efficient if dealers themselves earn more in a hybrid market structure than in a PBM. If dealers voluntarily introduce electronic trading platforms, a regulator does not need to think about inhibiting electronic trading. In general, and for reasons discussed in the text following Proposition 4, electronic trading is efficient if the number of dealers is large, if searching for quotes in the electronic market is very easy (either because searching does not take a lot of effort or because dealers are quoting very actively in the electronic market), if many dealers are slow or if the slow traders' cost of searching in the bilateral market is high. The latter two conditions mean that electronic trading is efficient if most traders are retail investors or have a high demand for immediacy.

Continuing to interpret Proposition 4, the cost of market fragmentation would outweigh any other welfare benefits of electronic trading if searching in the electronic market is too difficult relative to searching in the bilateral market. The latter may be the case, if dealers rarely quote electronically, if fast traders make dealers compete heavily in the bilateral market or if traders have a low demand for immediacy.

Since traders are always better off in the HM, a regulator that is only interested in investor welfare should always mandate electronic trading. Then, Proposition 5 suggests when to expect this to be necessary, namely when the PBM is already very competitive because many fast traders or low search costs. Dealers also prefer a PBM to an overly competitive HM, in which dealers respond very actively to RFQs, as the last condition in Proposition 5 states. Thus, the reason why dealers allegedly inhibited electronic trading in the CDS market and not in bond markets, as discussed in the introduction of the paper, may lie in the fact that traders in the CDS market are in general very sophisticated and the number of dealers is relatively small, whereas in bond markets, there are many dealers and traders are more diverse.

1.5.3 Testable implications

The testable implications can be grouped into two categories: (i) implications on trading patterns in the HM and (ii) comparisons between the HM and the corresponding PBM equilibrium.

Implications on trading patterns in the HM: Statement (i) in Proposition 2 states that fast traders trade relatively more in the bilateral market and slow traders trade relatively more on the platform. Moreover, statement (ii) in Proposition 6 gives conditions under which this separations between fast and slow traders becomes extreme: as the bilateral market becomes relatively unattractive due to a large number of dealers, competitive quoting on the platform or high search costs s_b or low competition that is generated by fast traders transactions on the platform are made by slow traders and transactions in the bilateral market are made by fast traders. Hendershott and Madhavan (2015a) find that bonds are more likely to be traded electronically, if the trader is selling or trading at the end of the month suggests that urgent trades are more likely to be executed via RFQs. In other words, if the trader has higher costs of waiting (search costs), the bond is more likely traded on a platform.

Statement (ii) in Proposition 2 implies that the dealer profits per quote are lowest in the trading venue in which dealers are willing to quote the lowest prices. Thus, profits per quote should be lowest in the trading venue in which the lowest quote is observed. This result in particular shows that it may not be sufficient to look at average markups to determine the amount of competition. For instance, as search costs are high, the number of dealers are large or dealers respond actively to RFQs, profits per quote are higher on the platform.[16] However, average markups that a slow trader receives upon one dealer contact are higher on the platform than in the bilateral market if search costs in the bilateral market are large and lower on the platform than in the bilateral market if the number of dealers is large or if dealers respond actively to RFQs, as statement (iii) in Proposition 6 states. In particular, empirical results in Hau et al. (2017) suggest that markups for unsophisticated traders in fact decrease strongly with electronic trading. Statement (iv) in Proposition 6 states that average transaction prices are higher on the platform if search costs in the bilateral market are large. Dispersion of transaction prices in the HM becomes very small if the number of dealers is large, if dealers respond actively to RFQs or if the search cost in the bilateral market is very small.

Comparisons between the HM and the corresponding PBM equilibrium: Statement (iii) of Proposition 3 states that turnover increases or the slow traders' reservation price decreases if a platform is introduces into a PBM.[17] Proposition 6 states when these effects are large or small: If the number of dealers or the search cost in the bilateral market are large or if the fraction of fast traders is small, the increase in turnover is large. If the fraction of fast traders is large or if search costs or the dealers' RFQ response rate is low, the increase in turnover is low. Holding search costs constant, a decrease in the

[16] For $s_b \to v - c$, this follows directly from Proposition 1. For $N \to \infty$ or $\eta \to 1$, this follows from the fact that $\overline{\mu}$ as defined in the proof of Proposition 1 goes to 0.

[17] Both may happen at the same time.

reservation price is equivalent to a decrease in average markups for slow traders in the bilateral market (an implication of the model that could be tested).[18] This means that, according to statement (v) in Proposition 6, the decrease in the slow traders' reservation price is large if the dealers respond actively to RFQs and small if the search costs in the bilateral market are large or dealers respond rarely to RFQs.

Statements (iv) and (v) of Proposition 3 state that turnover and profits per quote in the bilateral market decreases when a platform is introduced.

Dispersion of transaction prices becomes very small in the HM as the search costs in the bilateral market become small, the number of dealers becomes large or if dealers respond actively to RFQs, as statement (iv) in Proposition 6 states. Benos et al. (2018) find that price dispersion indeed decreases after the introduction of electronic trading in a PBM. Statement (iv) of Proposition 6 also states that transaction prices are higher on the platform if the search costs in the bilateral market are large.

1.6 Discussion

This section discusses some of the critical model assumptions and limitations of the model.

The model has four crucial assumptions that warrant further discussion. First, there are slow traders with search costs and fast traders without search costs. This is perhaps the most unproblematic assumption, since it is generally accepted that there are many institutional investors with sophisticated traders who find it not very costly to either call a trader at a bank or submit an RFQ on the platform. Retail investors, private high-net-worth individuals, however may find these things much more costly. On the one hand, it may take more time for them to do these things, causing costs associated with the delay of executing the trade. On the other hand, many dealers may require their clients to set up margin accounts with them. While institutional investors may already have those, setting up these accounts is costly for retail investors.

Second, it is assumed that dealers do not observe whether they face a fast or slow trader. This assumption is plausibly satisfied in the case of anonymous RFQs in which the name of the counterparty is not revealed. In the case of name give-up RFQs or bilateral trading, this assumption may be justified to the extent that being fast or slow is not a fixed attribute. Even a sophisticated institutional investor may be under pressure to execute a certain trade and therefore has high costs of waiting to search for the best quote.

Third, it is assumed that dealer A's quote remains valid even if a trader visits other dealers before coming back to dealer A to buy the asset. This assumption is at odds with the claim that dealers' quotes are "only

[18]This follows from (1.6) and the expression for r^{PBM} in Section 1.1.

as good as the breath is warm" (Bessembinder and Maxwell, 2008). However, given that RFQ auctions last several minutes and many bilateral transactions are made in email conversations, at least the most sophisticated institutional investors should be able to compare multiple quotes similar to the way described in the model. The assumption does not matter in the case of slow investors, since they buy the asset when they receive their first quote in equilibrium. A formal proof that the equilibrium in this paper will not change if dealers will increase their price when contacted repeatedly can be found in Appendix 1.9.

Fourth, the dealers' RFQ response rate $\eta < 1$ is exogenously given. Given that dealers give up a profit by not responding, it may be surprising that dealers do not respond all the time. However, it is an empirical fact (Hendershott and Madhavan, 2015a) that the dealers' response rate is very low and that a binomial distribution of responses is a reasonable approximation. There are many reasons why dealers do not respond all the time including costs of paying attention or risk management concerns. One may think of dealers as artificially limiting the RFQ response rate due to their concern that algorithms "go rogue" (Rodgers (2017)). It is not clear ex ante how this concern may be related in any way to the primitive model parameters, so that the assumption that η is exogenous seems to be a plausible prior.[19] The number of queried dealers at MarketAxess is generally quite high: Hendershott and Madhavan (2015a) show averages above 25 dealers per RFQ. In that light, the model assumption that all dealers are contacted by submitting an RFQ seems a reasonable simplification.

The most important limitations of the model are as follows. There is no uncertainty about the traders' value of the asset and the dealers' costs of providing the asset. Furthermore, information about the asset is symmetric. As Hendershott and Madhavan (2015a) argue, information leakage may incentivize investors to trade bilaterally as opposed to trading on an electronic trading platform. Information leakage is not modeled in this paper. Arguably, information asymmetries may vary among different asset classes. U.S. Treasury bonds are perhaps an asset class with relatively small information asymmetries, while information asymmetries are stronger for junk bonds.[20] Thus, the model presented in this paper can be thought of as referring to markets like the market for U.S. Treasuries. Alternatively, the model may still apply to other asset classes, given that search costs are more important in determining the dealers' quotes than information asymmetries are.

[19]There may still be opportunities for future research in this direction: in a different context, Basak and Buffa (2017) shows how operational risk may affect model selection for forecasting investment returns.

[20]Even though Hilscher et al. (2015) argue that informed traders are active in the equity market and not so much in the CDS market, Oehmke and Zawadowski (2016) argue that speculative activity is higher in the CDS market than in the corporate bond market.

1.7 Conclusions

This paper provides an equilibrium model in which the interaction between a bilateral market and an electronic trading platform is analyzed. It may explain why dealers prefer an HM in some cases and a PBM in other cases. The market participation of slow traders in the PBM plays a crucial role. If slow traders participate actively in the market, dealers do not want to introduce an HM to avoid excessive competition. The slow traders' market participation depends on the fraction of fast traders in the market. With many fast traders of institutional investors in the market, dealers' markups are very competitive and slow traders find it profitable to enter the market. With very few institutional investors in the market, the slow traders' bargaining power can be increased by introducing a trading platform on which traders can contact many dealers at once. It has been shown that this result is not simply driven by more quotes in the HM. Instead, it matters toward whom the quoting activity is directed. In a PBM, slow traders always buy from the first dealer they contact so that there is no chance of receiving multiple quotes at once. In an HM, substantially more quotes are sent to slow traders, which makes the HM more attractive to them.

The model provides several policy implications. Since many OTC markets are currently under scrutiny by regulators, these implications can potentially be used to make more informed decisions when regulating these markets.

The empirical implications can potentially explain several empirical findings in previous empirical studies. Moreover, new predictions have been derived that can be used to validate the model mechanism.

1.8 Appendices

1.8.1 Derivation of the HM Equilibrium

This appendix discusses the missing details that were neglected in the main text. First, the traders' strategies are considered, then the dealers' quoting strategies are derived and finally, an equilibrium is established.

Traders' search strategies: The derivation of the slow traders' search strategy builds on the insights of Weitzman (1979). The key insight from the latter is that one can assign so-called reservation prices to each trading venue. Then, an optimal search strategy can be constructed solely based on these reservation prices and quotes received while searching.

Let p_b denote the (random) quote that a trader receives when contacting a dealer in the bilateral market and let q denote the (random) best quote (of all responses) that a trader receives when submitting an RFQ on the trading platform, conditional on receiving at least one response. Then the reservation prices r_b and r_p for the bilateral market and for the trading platform, respectively, are defined by

$$r_b := \mathbb{E}(\min(p_b, r_b)) + s_b \tag{1.14}$$

and

$$r_p := (1 - (1 - \eta)^N) \cdot \mathbb{E}(\min(q, r_p)) + (1 - \eta)^N r_p + s_p. \tag{1.15}$$

One can interpret r_b as follows. Assume a slow trader has the opportunity to buy the asset at the price r_b. Then he would be just indifferent between buying the asset at that price and contacting a dealer in the bilateral market to obtain an expected price improvement $r_b - \mathbb{E}(\min(p_b, r_b))$ in exchange for the search cost s_b. An analogous interpretation holds for r_p. When defining r_p one additionally has to take into account that a trader may not get a response at all when submitting an RFQ on the electronic trading platform. The latter event happens with probability $(1 - \eta)^N$. The first term in (1.15) refers to the expected price improvement if the trader searches on the platform and at least one dealer responds to an RFQ. The second term in (1.15) refers to the event in which there is be no price improvement since no dealer responded to the RFQ. Together with the search cost s_p, these terms have to add up to r_p, if the trader is indifferent between buying the asset at r_p or searching on the platform.

The random variable q is distributed according to the distribution function F defined by

$$F(x) := \frac{1 - (1 - \eta G(x))^N}{1 - (1 - \eta)^N}.$$

Since support$(G) \subset [c, \infty)$, one obtains support$(F) \subset [c, \infty)$. One can now solve the slow traders' search problem as in Weitzman (1979).

Lemma 1. *The reservation prices r_b and r_p as defined in equations (1.14) and (1.15) exist and r_b is unique. If $s_p > 0$, the reservation price r_p is unique as well. It is optimal for slow traders to start their search on the platform if*

$$v \geq r_b \geq r_p. \tag{1.16}$$

holds. Furthermore, if condition (1.16) holds and $r_b < v$, there is the following optimal continuation rule: If a trader received no quote or a lowest quote greater than r_b on the platform, the trader will continue to search in the bilateral market until he finds a quote less than or equal to r_b.

If (1.16) holds, $r_b = v$ and a slow trader did not receive a quote less than or equal to r_b on the platform, then the slow trader is indifferent between continuing to search in the bilateral market and terminating the search. In this case, continuing to search with any probability $\gamma \in [0, 1]$ is optimal for slow traders.

The following statement gives a lower bound for r_b.

Lemma 2. *It must be the case that $r_b \geq c + s$, where r_b is defined as in (1.14). The strict inequality $r_b > c + s$ holds if the probability of the event $\{p_b > c\}$ is positive.*

Lemma 2 states that r_b is always strictly greater than c. This fact will be used below when determining the dealers' optimal quoting strategies. The inequality $r > c$ will imply that dealers make a positive profit on both the platform and in the bilateral market.

Dealers' quoting strategies:

The following Lemma states some general properties of the dealers' quoting strategies.

Lemma 3. *Let the slow traders start their search on the platform and let them use a reservation price strategy with reservation price r. Let the slow traders' probability γ of continuing the search in the bilateral market after not having received a satisfactory offer on the platform be positive.*

Then, neither G nor H can have any atoms. Neither in the bilateral market nor on the platform, a dealer ever quotes a price greater than r or less than or equal to c. A dealer on the platform faces a mass k_p of slow traders given by (1.1). A dealer in the bilateral market faces a mass k_b of slow traders given by (1.2).

The next Lemma derives conditions that G and H have to satisfy.

Lemma 4. *Let slow traders start their search on the platform and use the reservation price r. Equation (1.3) has to hold for all prices p in the support of H. Equation (1.4) has to hold for all prices p in the support of G. The suprema of the supports of both G and H are equal to the slow traders' reservation price r.*

The following lemma states that unique distributions H and G which satisfy the properties mentioned in Lemma 4 exist for any given reservation price strategy of the slow traders.

Lemma 5. *Holding the slow traders' choices of γ and $r > c$ fixed, there are unique monotone increasing functions $G, H : [c, r] \to [0, 1]$ with densities h and g, such that the following holds:*

- *(1.4) is satisfied for all p with $g(p) > 0$.*
- *(1.3) is satisfied for all p with $h(p) > 0$.*
- $\sup \operatorname{support} H = \sup \operatorname{support} G = r$.
- *If $p \in \mathbb{R} \setminus \operatorname{support} G$, one has*

$$(p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p + \mu(1-H(p))^N\right] \leq (1-\eta)^{N-1}(r-c)k_p, \qquad (1.17)$$

  i.e. dealers on the platform cannot increase their profit by choosing a quoting strategy that differs from G.

- *If $p \in \mathbb{R} \setminus \operatorname{support} H$, one has*

$$(p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] \leq (r-c)k_b., \qquad (1.18)$$

  i.e. dealers in the bilateral market cannot increase their profit by choosing a quoting strategy that differs from H.

The next Lemma ensures that that slow traders' choices of r and γ are well-defined.

Lemma 6. *Holding γ fixed, there are unique functions H and G and a reservation price r, such that*

$$r = s_b + \int_c^r p \,\mathrm{d}H(p) \qquad (1.19)$$

holds and the functions G and H satisfy the conditions mentioned in Lemma 4.

Moreover, the slow traders' reservation price r is continuous and strictly monotone increasing in γ. One has $\lim_{\gamma\to 0} r = c + s_b$, if an equilibrium exists.

Lemma 6 states that the slow traders' reservation price for a fixed γ is well-defined. This result is different from the result in Lemma 1 that the solution to equation (1.14) well-defined. The distribution H is fixed in (1.14), while H depends on r in equation (1.19). The fact that the reservation price that solves (1.19) is strictly monotone increasing in γ implies that any equilibrium value of γ is unique. If $\gamma = 1$ and $r < v$, the equilibrium value of γ is unique by the construction of the equilibrium. If $r = v$ for a $\gamma \le 1$, one would have $r \neq v$ for any other $\gamma \le 1$.

The next lemma describes how the slow traders' choice of γ must behave as the search cost in the bilateral market s_b varies.

Lemma 7. *The slow traders' equilibrium choice of γ is monotone decreasing inn s_b and strictly monotone decreasing if $\gamma > 1$. Moreover, one has $\lim_{s_b\to 0}\gamma = 1$ and $\lim_{s_b\to v-c}\gamma = 0$.*

Defining $\underline{q} := \inf \operatorname{support}(G)$ and $\underline{b} := \inf \operatorname{support}(H)$, one obtains the following result.

Lemma 8. *For any given reservation price strategy of the slow traders with $\gamma \in (0, 1]$ and $r \in (c, v]$ define*

$$\mu^* := \frac{\gamma(1-\eta) - \gamma(1-\eta)^N}{N - \gamma(1-\eta)^N}.$$

If $\mu > \mu^*$, the following equivalent statements are true: (i) $\underline{q} > \underline{b}$ and profits per quote are higher for trading desks on the platform than for trading desks in the bilateral market.

If $\mu \le \mu^*$, the following equivalent statements are true: (i) $\underline{q} \le \underline{b}$ and profits per quote of trading desks on the platform are smaller than or equal to the profits of trading desks in the bilateral market.

The next lemma gives a sufficient condition for the inequality $r_p < r_b$ to hold.

Lemma 9. *There is an upper bound for the expected best quote on the platform conditional on at least one response to the RFQ. This upper bound is given by*

$$\mathbb{E}(q) := \int_{\underline{q}^p}^{r} p \, \mathrm{d}F(p) \le c + (r-c)(1-\eta)^{N-1} \frac{N\eta}{1-(1-\eta)^N}. \tag{1.20}$$

The inequality $r_p \le r_b$ holds if

$$s_p \le (1-(1-\eta)^N)(r - \mathbb{E}(q)). \tag{1.21}$$

1.8.2 Proofs

This appendix contains all proofs that have been omitted in the main text.

Proof of Lemma 1. Recalling that support H, support $G \subset [c, \infty)$, one can rewrite equations (1.14) and (1.15) the following way:

$$s_b = \int_c^{r_b} (r_b - x)\,\mathrm{d}H(x) =: \varphi_b(r_b),$$

$$s_p = (1 - (1-\eta)^N) \int_c^{r_p} (r_p - x)\,\mathrm{d}F(x) =: \varphi_p(r_p).$$

The integrals in the definitions of φ_b and φ_p indeed exist, since

$$\int_c^{r_b} |r_b - x|\,\mathrm{d}H(x) \leq |r_b - c|$$

and

$$(1 - (1-\eta)^N) \int_c^{r_p} |r_p - x|\,\mathrm{d}F(x) \leq (1 - (1-\eta)^N)|r_b - c|.$$

The functions $\varphi_p, \varphi_b : \mathbb{R} \to \mathbb{R}$ are continuous, monotone increasing and strictly monotone increasing in all $r_b, r_p \in \mathbb{R}$, such that $\varphi_b(r_b), \varphi_p(r_p) > 0$. Furthermore, $\lim_{x\to\infty} \varphi_i(x) = \infty$ and $\varphi_i(c) = 0$ for $i \in \{p, b\}$. Thus, there exist some $r_b, r_p \in \mathbb{R}$ such that $\varphi_b(r_b) = s_b$ and $\varphi_p(r_p) = s_p$ for $s_b > 0$ and $s_p \geq 0$. The Weitzman price r_b must be unique due to strict monotonicity of φ_b in any candidate r_b. If $s_p > 0$, also r_p is unique due strict monotonicity of φ_p in any candidate r_p.

The remaining part of the proof consists of two main steps. First, it is shown that given that the platform has been contacted, the stated continuation rule is optimal. Then it is shown that prioritizing the platform is indeed the optimal decision.

Step 1: Optimality of the continuation rule. Suppose the trader received no quote on the platform at all. Then slow traders at least weakly prefer to continue to search to not searching in the bilateral market, since $r_b \leq v$ and the definition of r_b implies

$$\mathbb{P}(p_b \leq v)\mathbb{E}(v - p_b | p_b \leq v) - s_b = v - \mathbb{E}(\min(p_b, v)) - s_b \geq 0.$$

The last equation means that the trader would get a non-negative payoff from searching in the bilateral market, since the expected payoff from being able to buy the asset for a price less than or equal to v weakly outweighs the search cost s. Thus, it is always at least as good to continue to search as to terminate the search. If $r_b = v$, continuing to search with any probability $\gamma \in [0,1]$ is optimal. If $r_b < v$, continuing to search almost surely is the unique optimal strategy.

Analogously, suppose the trader received a quote $p > r_b$ as a response to an RFQ on the platform. Then it is also at least as good to continue to search in the bilateral market as to terminate the search, since the definition of r_b implies

$$\begin{aligned}
\mathbb{P}(p_b \leq r_b)\mathbb{E}(v - p_b|p_b \leq r_b) - s_b &\geq \mathbb{P}(p_b \leq r_b)\mathbb{E}(r_b - p_b|p_b \leq r_b) - s_b \\
&= r_b - \mathbb{E}(\min(p_b, r_b)) - s_b \\
&= 0.
\end{aligned}$$

This means that the expected payoff from being able to buy the asset at a price less than or equal to r_b weakly outweighs the search cost s_b. If $r_b = v$ one obtains again that the slow trader is indifferent between searching and not searching. Thus, continuing to search with any probability $\gamma \in [0,1]$ is optimal. If $r_b < v$, continuing to search almost surely is the unique optimal strategy.

On the other hand, if the trader has received a quote $p \leq r_b \leq v$ as a response to an RFQ on the platform, the definition of r_b implies $p - \mathbb{E}(\min(p_b, p)) - s_b \leq 0$, i.e. search costs are weakly larger than the potential benefits of obtaining a price better than p. Thus, terminating the search is always weakly preferred to searching in the bilateral market. Here it is assumed that a slow trader will terminate the search even in the case $p = r_b$. It will turn out, however, that this is without loss of generality, since p will have a continuous distribution in any case and a zero-probability event will not influence a dealer's behavior.

Step 2: Optimality of prioritizing the platform. Here, r_p will stand for any solution to (1.15). It remains to show that prioritizing the platform is indeed the optimal strategy. This will be shown by induction over the number m of dealers left to contact in the bilateral market. If $m = 0$ (for instance, because all dealers have been contacted), one can verify analogously to the arguments above that it is optimal to go to the platform if no quote less than or equal to r_p has been received until that point, since $r_p < v$. Suppose that $m \geq 0$ and let it be optimal to go to the platform if no offer less than r_p has been received. It has to be shown that it is optimal to go to the platform if $m + 1$ dealers remain uncontacted in the bilateral market and no quote less than r_p has been received so far. This implies that the trader should start his search on the platform, when N dealers are uncontacted in the bilateral market.

Let y be the best current quote at which the trader can buy the asset. If $r_b > y > r_p$, search costs will be greater than the expected price improvement in the bilateral market. On the other hand, the price improvement on the platform will be greater than the search costs and it is therefore optimal to go to the platform.

Now, let $y \geq r_b$ or assume that no quote has been received so far. Both going to the bilateral market and going to the platform is at least as good as not to search. In this case, it is not trivial that prioritizing the platform is optimal.

Let B denote the expected payoff a trader gets if he goes to the bilateral market first and receives the quote p_b. Now, three cases are possible.

Case 1: $p_b \leq r_p$: It is clearly optimal to buy the asset at the price p_b, since search costs dominate any expected price improvements from searching.

Case 2: $r_p < p_b \leq r_b$: The inductive hypothesis states that it is optimal to continue to search on the platform and receive the quote q with probability $\omega := 1 - (1 - \eta)^N$. If $q \leq r_p$, the trader buys the asset at price q. If $q > r_b$ or no quote has been received, the trader will buy the asset at price p_b. If $r_p < q \leq r_b$, the trader will buy the asset at price $\min(p_b, q)$. Continuing to search is not profitable in either case.

Case 3: $p_b > r_b$: The inductive hypothesis states that it is optimal to continue to search on the platform and receive the quote q with probability ω. If $q < r_b$, the trader buys the asset at price q. Let $q > p_b$ or assume no quote has been received on the platform. Then the trader will continue to search in the bilateral market (with probability γ).

Notice that only case 1 and case 3 are possible if $r_b = r_p$. The following text will also consider events that require $r_b > r_p$ and that therefore occur with zero probability if $r_b = r_p$.

Let A be the expected payoff the trader gets if he starts to search on the platform and receives the quote q with probability ω. There are again three different possible cases.

Case 1: $q \leq r_b$: The optimal continuation rule prescribes to buy the asset at the price q.

In the next two cases, the trader optimally continues to search in the bilateral market and receives the offer p_b.

Case 2: There is no $q \leq r_b$ and one has $p_b \leq r_b$: The optimal continuation rule prescribes to buy the asset at the price p_b.

Case 3: There is no $q \leq r_b$ and one has $p_b > r_b$: The optimal continuation rule prescribes to keep searching in the bilateral market (with probability γ).

If both $p_b > r_b$ and no quote $q \leq r_b$ has been received, the trader will keep searching in the bilateral market (with probability γ) after having visited both the platform and one dealer. The probability of this outcome and the value of future search opportunities are independent of the order in which one visited the trading venues. The continuation value in this event multiplied by the probability of occurrence of the event is equal to some number X (the exact value is not important since it will cancel out later).

One can now calculate the payoffs according to the above reasoning

$$\begin{aligned} A &= -s_p + \omega \mathbb{P}(q \leq r_p) \cdot \mathbb{E}(v - q | q \leq r_p) + \omega \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{E}(v - q | r_p < q \leq r_b) \\ &\quad + (1 - \omega \mathbb{P}(q \leq r_b)) \cdot (-s_b + \mathbb{P}(p_b \leq r_b) \cdot \mathbb{E}(v - p_b | p_b \leq r_b)) + X. \end{aligned}$$

Rewriting equations (1.14) and (1.15) and using the expressions $s_p = \omega \mathbb{P}(q \leq r_p) \cdot \mathbb{E}(r_p - q | q \leq r_p)$ and $s_b = \mathbb{P}(p_b \leq r_p) \cdot \mathbb{E}(r_p - p_b | p_b \leq r_p)$, one gets

$$\begin{aligned} A &= -s_p + \underbrace{\omega \cdot \mathbb{P}(q \leq r_p) \cdot \mathbb{E}(r_p - q | q \leq r_p)}_{=s_p} + \omega \cdot \mathbb{P}(q \leq r_p) \cdot (v - r_p) \\ &\quad + \omega \cdot \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{E}(v - q | r_p < q \leq r_b) \\ &\quad + (1 - \omega \cdot \mathbb{P}(q \leq r_b)) \cdot (-s_b + \underbrace{\mathbb{P}(p_b \leq r_b) \cdot \mathbb{E}(r_b - p_b | p_b \leq r_b)}_{=s_b} + \mathbb{P}(p_b \leq r_b) \cdot (v - r_b)) + X \\ &= \omega \cdot \mathbb{P}(q \leq r_p) \cdot (v - r_p) + \omega \cdot \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{E}(v - q | r_p < q \leq r_b) \\ &\quad + (1 - \omega \cdot \mathbb{P}(q \leq r_b)) \cdot \mathbb{P}(p_b \leq r_b) \cdot (v - r_b) + X. \end{aligned}$$

Analogously, one gets

$$\begin{aligned}
B &= -s_b + \mathbb{P}(p_b \leq r_p) \cdot \mathbb{E}(v - p_b | p_b \leq r_p) \\
&+ \mathbb{P}(r_p < p_b \leq r_b) \cdot \Big\{ -s_p + \omega \cdot \mathbb{P}(q \leq r_p) \cdot \mathbb{E}(v - q | q \leq r_p) \\
&+ \omega \cdot \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{E}(v - \min(q, p_b) | r_p < q, p_b \leq r_b) \\
&+ (1 - \omega \cdot \mathbb{P}(q \leq r_b)) \cdot \mathbb{E}(v - p_b) | r_p < p_b \leq r_b) \Big\} \\
&+ \mathbb{P}(r_b < p_b) \cdot \Big\{ -s_p + \omega \cdot \mathbb{P}(q \leq r_p) \cdot \mathbb{E}(v - q | q \leq r_p) \\
&+ \omega \cdot \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{E}(v - q | r_p < q \leq r_b) \Big\} + X
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{P}(p_b \leq r_b) \cdot (v - r_b) + \mathbb{P}(r_p < p_b) \cdot \omega \cdot \mathbb{P}(q \leq r_p) \cdot (v - r_p) \\
&+ \mathbb{P}(r_p < p_b \leq r_b) \cdot \omega \cdot \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{E}(v - \min(q, p_b) | r_p < q, p_b \leq r_b) \\
&- \mathbb{P}(r_p < p_b \leq r_b) \cdot \omega \cdot \mathbb{P}(q \leq r_b) \cdot \mathbb{E}(v - p_b | r_p < p_b \leq r_b) \\
&+ \mathbb{P}(r_b < p_b) \cdot \omega \cdot \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{E}(v - q | r_p < q \leq r_b) + X.
\end{aligned}$$

After some further manipulations one gets

$$\begin{aligned}
A - B &= (r_b - r_p) \cdot \mathbb{P}(p_b \leq r_p) \cdot \omega \cdot \mathbb{P}(q \leq r_p) \\
&+ (r_b - \mathbb{E}(p_b | r_p < p_b \leq r_b)) \cdot \mathbb{P}(r_p < p_b \leq r_b) \cdot \omega \cdot \mathbb{P}(q \leq r_p) \\
&+ (r_b - \mathbb{E}(q | r_p < q \leq r_b)) \cdot \omega \cdot \mathbb{P}(r_p < q \leq r_b) \cdot \mathbb{P}(p_b \leq r_p) \\
&+ \Big\{ \mathbb{E}(\min(q, p_b) | r_p < q, p_b \leq r_b) - \mathbb{E}(q | r_p < q \leq r_b) - \mathbb{E}(p_b | r_p < p_b \leq r_b) \\
&+ r_b \Big\} \cdot \mathbb{P}(r_p < p_b \leq r_b) \cdot \omega \cdot \mathbb{P}(r_p < q \leq r_b) \\
&\geq 0,
\end{aligned}$$

since $r_b \geq r_p$ and

$$\begin{aligned}
\mathbb{E}(\min(q, p_b) | r_p < q, p_b \leq r_b) &= r_b + \mathbb{E}(\min(q - r_b, p_b - r_b) | r_p < q, p_b \leq r_b) \\
&\geq r_b + \mathbb{E}(\min(q - r_b + p_b - r_b) | r_p < q, p_b \leq r_b) \\
&= \mathbb{E}(q | r_p < q \leq r_b) + \mathbb{E}(p_b | r_p < p_b \leq r_b) - r_b.
\end{aligned}$$

Thus, contacting the platform first is optimal.

□

Proof of Lemma 2. Rewriting (1.14) gives

$$s_b = \mathbb{E}(\max(r_b - p_b, 0)).$$

The right-hand side of the last equation is continuous and monotone increasing in r_b. One has $\mathbb{E}(\max(c - p_b, 0)) = 0$, since no dealer quotes below his cost. Thus, $r_b > c$ for $s > 0$. Equation (1.14) now gives

$$r_b := \mathbb{E}(\min(p_b, r_b)) + s_b \geq c + s_b.$$

The last inequality is strict, if prices are greater than c with positive probability. □

Proof of Lemma 3. First, there is no demand from slow traders, if a dealer quotes a price higher than r. Quoting prices above r in order to sell to fast traders cannot be an equilibrium strategy in the bilateral market due to the resulting Bertrand competition. Since dealers in the bilateral market do not give quotes above r, dealers on the platform cannot sell to fast traders by quoting above r. Thus, no dealer quotes above r with positive probability.

On the other hand, a dealer's profit is zero or negative if he quotes a price at or below his cost c. Now, let the dealer quote a price p with $c < p \leq r$. Lemma 2 ensures the existence of such a p. Now the dealer gets a positive profit on both the platform and in the bilateral market. This can be seen as follows. With probability $(1-\eta)^{N-1}$ the dealer is alone on the platform when responding to an RFQ. If a slow trader requested the quote, the dealer would be able able to sell the asset at price r due to the slow traders' reservation price strategy. Thus, the dealer's profits are at least $(1-\eta)^{N-1}(1-\mu)(p-c) > 0$ on the platform in the event that the dealer charges p. The same logic can be applied to the bilateral market in which case the dealer would get at least $(1-\eta)^N\gamma(1-\mu)(p-c)/N > 0$, since slow traders will not get a satisfactory offer on the platform with probability $(1-\eta)^N$. Then they continue to search with probability $\gamma > 0$ in the bilateral market where they choose a particular dealer with probability $1/N$. Thus, the equilibrium quotes are greater than c and less than or equal to r.

$k_p = 1 - \mu$ follows from the slow traders' strategy of starting the search on the platform. Since dealers never quote a price greater than r, the slow traders always buy the asset when they receive a quote on the platform. The only case in which they continue their search in the bilateral market is when they do not get a quote on the platform. Then they continue to search with probability $\gamma > 0$ and buy the asset from the first dealer they contact in the bilateral market. The exact law of large numbers implies $k_b = (1-\eta)^N \gamma(1-\mu)/N$.

It can furthermore be shown that there cannot be atoms in the distribution G of prices quoted on the platform, no matter how the distribution H of prices in the bilateral market looks like. Suppose there is a price p with $r \geq p > c$ that is quoted on the platform with probability $\rho > 0$. A single dealer can now profitably deviate from this strategy as follows. Since the number of prices charged with positive probability must be countable, one can find for each $\delta > 0$ an ε_δ, such that $\delta \geq \varepsilon_\delta > 0$ and the price $p - \varepsilon_\delta$ is charged with probability zero. A dealer can now charge price $p - \varepsilon_\delta$ with probability ρ and charge price p with probability zero. Using the fact that $\lim_{\delta \to 0} G(p - \varepsilon_\delta) = G(p) - \rho$, one can express the difference Δ_p in profits between quoting $p - \varepsilon_\delta$ and quoting p for small δ as follows. The quoting dealer only makes a positive profit if no other dealer on the platform quotes a lower price. If no other dealer quotes a lower price, there might be $j = 0, 1, ..., N-1$ dealers who quote price p. The calculation below considers the cases in which j dealers quote price p on the platform separately.

I allow H to have atoms. To simplify the algebra, I introduce the function $D : \mathbb{Z} \times \mathbb{R} \to \mathbb{R}$. I define $D(j,p)$ as the expected quantity of fast traders the dealer quoting on the platform faces when he quotes p, j other dealers on the platform also quote p and no other dealer on the platform quotes a lower price. $D(j,p)$ cannot be increasing in p. As the $j+1$ dealers raise their quote, no dealer in the bilateral market becomes more likely to quote a higher price than the dealers on the platform do. D is decreasing in j, since sharing of the total demand reduces the demand for each individual dealer if j increases.

If H has no atom at price p then one has

$$D(j,p) := \frac{\mu}{j+1}(1 - H(p))^N.$$

The last expression says that the $j+1$ dealers who quote p on the platform share the demand by fast traders equally, if all dealers on in the bilateral market quote a higher price.

If H has an atom at price p, then price p is quoted with some probability $\rho > 0$ in the bilateral market and D is given by

$$D(j,p) := \sum_{k=0}^{N} \binom{N}{k} (1 - H(p))^{N-k} \rho^k \frac{\mu}{j+1+k}.$$

If k dealers in the bilateral market quote p and all other dealers in the bilateral market quote a higher price, the dealers quoting p will share the demand equally.

Now it is shown that there would be a profitable deviation if the equilibrium choice of G had atoms. The first two terms in the first equation below look at the event in which the quoting dealer has the lowest quote on the platform and no other dealer quotes p, which happens with probabilities $(1-\eta G(p-\varepsilon_\delta)-\eta\rho)^{N-1}$ and $(1-\eta G(p))^{N-1}$, respectively. The last two terms look at the events in which $j>0$ other dealers quote p. If the quoting dealer quotes p as well, he will only get a fraction of the traders' demand in expectation. If the quoting dealer quotes $p-\varepsilon_\delta$, he will face a jump in demand.

$$\begin{aligned}
\Delta_p \;=\;& (1-\eta G(p-\varepsilon_\delta)-\eta\rho)^{N-1}(p-\varepsilon_\delta-c)(k_p+D(0,p-\varepsilon_\delta)) \\
& -(1-\eta G(p))^{N-1}(p-c)(k_p+D(0,p)) \\
& +\sum_{j=1}^{N-1}\binom{N-1}{j}(1-\eta G(p-\varepsilon_\delta)-\eta\rho)^{N-1-j}(\eta\rho)^j(p-\varepsilon_\delta-c)(k_p+D(0,p-\varepsilon_\delta)) \\
& -\sum_{j=1}^{N-1}\binom{N-1}{j}(1-\eta G(p))^{N-1-j}(\eta\rho)^j(p-c)\left(\frac{k_p}{j+1}+D(j,p)\right) \\
\geq\;& (1-\eta G(p-\varepsilon_\delta)-\eta\rho)^{N-1}(p-\varepsilon_\delta-c)(k_p+D(0,p)) \\
& -(1-\eta G(p))^{N-1}(p-c)(k_p+D(0,p)) \\
& +\sum_{j=1}^{N-1}\binom{N-1}{j}(1-\eta G(p-\varepsilon_\delta)-\eta\rho)^{N-1-j}(\eta\rho)^j(p-\varepsilon_\delta-c)(k_p+D(j,p)) \\
& -\sum_{j=1}^{N-1}\binom{N-1}{j}(1-\eta G(p))^{N-1-j}(\eta\rho)^j(p-c)\left(\frac{k_p}{j+1}+D(j,p)\right) \\
\to\;& \sum_{j=1}^{N-1}\binom{N-1}{j}(p-c)\frac{jk_p}{j+1}(1-\eta G(p))^{N-1-j}(\eta\rho)^j>0 \quad \text{as } \delta\to 0.
\end{aligned}$$

Thus, the proposed deviation is profitable. In equilibrium, G cannot have any atoms. The calculation shows that the increase in profits is possible because ties can be avoided and the dealer gets the full demand of the trader, when he would have had to split the demand in expectation with other dealers.

Analogously, one can show that there cannot be any atoms in the distribution of prices in the bilateral market. The difference Δ_b in profits between quoting $p-\varepsilon_\delta$ and quoting p in this case is

$$
\begin{aligned}
\Delta_b \;=\; & (1-H(p-\varepsilon_\delta)-\rho)^{N-1}(p-\varepsilon_\delta-c)(k_b+(1-\eta G(p-\varepsilon_\delta))^N\mu) \\
& -(1-H(p))^{N-1}(p-c)(k_b+(1-\eta G(p)^N)\mu) \\
& +(1-(1-H(p-\varepsilon_\delta))^{N-1})(p-\varepsilon_\delta-c)k_b \\
& -(1-(1-H(p)+\rho)^{N-1})(p-c)k_b \\
& +\sum_{j=1}^{N-1}\binom{N-1}{j}(1-H(p-\varepsilon_\delta)-\rho)^{N-1-j}\rho^j(p-\varepsilon_\delta-c)(k_p+(1-\eta G(p-\varepsilon_\delta))^N\mu) \\
& -\sum_{j=1}^{N-1}\binom{N-1}{j}(1-H(p))^{N-1-j}\rho^j(p-c)(k_p+(1-\eta G(p))^N\frac{\mu}{j+1})
\end{aligned}
$$

The first two terms look at the events in which the quoting dealer has the lowest quote in the bilateral market and no other dealer quotes p. These events happen with probabilities $(1-H(p-\varepsilon_\delta)-\rho)^{N-1}$ and $(1-H(p))^{N-1}$, respectively. The next two terms look at the events in which the quoting dealer does not have the lowest quote, i.e. at least one other dealer has a strictly lower quote. In this case, the dealer can only sell to slow traders. The last two terms look at the events in which the quoting dealer has the lowest quote and $j>0$ other dealers quote p. This means that there are ties if the dealer quotes p as well and the demand by fast traders will be split in expectation equally among those $j+1$ dealers. Since G is continuous, the event that the lowest quote on the platform is equal to p has probability zero. Thus, demand does not have to be split with dealers on the platform. A dealer always gets the full demand of a slow trader in the bilateral market. One has $\lim_{\delta\to 0} H(p-\varepsilon_\delta)=H(p)-\rho$. Thus, as $\delta\to 0$, the first four terms in the previous equation vanish and the difference between the last two terms is positive, since $N\geq 2$:

$$
\lim_{\delta\to 0}\Delta_b=\sum_{j=1}^{N-1}\binom{N-1}{j}(1-H(p))^{N-1-j}\rho^j(p-c)(1-\eta G(p))^N\frac{j\mu}{j+1}>0.
$$

Thus, the proposed deviation is profitable. Again, the profitable deviation exists because ties with other dealers can be avoided. In equilibrium, H cannot have any atoms. □

Proof of Lemma 4. If a dealer is in the bilateral market, then for every p in the support of H, the expected profit

$$
\Pi_b(p)=(p-c)\left[k_b+\mu(1-H(p))^{N-1}\sum_{j=0}^{N-1}\binom{N-1}{j}(1-\eta)^{N-1-j}\eta^j(1-G(p))^j\right]
$$

must be constant. Using the binomial formula $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$, the previous equation simplifies to

$$\Pi_b(p) = (p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right].$$

Suppose there is a $\overline{p} < r$ such that $H(\overline{p}) = 1$. Then a dealer could profitably deviate from this strategy by quoting for instance r with positive probability. The difference in expected profits would be

$$\Pi_b(r) - \Pi_b(\overline{p}) = (r-\overline{p})k_b > 0.$$

Thus, $H(\overline{p}) < 1$ for $\overline{p} < r$. By Lemma 3, no dealer quotes a price above r. Therefore, r must be the supremum of the support of H. Lemma 3 also states that H cannot have any atoms. Dealers are only willing to quote according to a continuous distribution that has a support with supremum r, if profits for prices in the support of H are equal to $\Pi_b(r)$. When quoting a price equal to r, a dealer only sells to slow traders, since other dealers almost surely quote a lower price if H is a continuous distribution. The indifference condition determining H is therefore

$$(p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] = (r-c)k_b.$$

This is equation (1.3). If a dealer is on the platform, then for every p in the support of H, the expected profit

$$\begin{aligned}\Pi_p(p) \quad = \quad & (p-c)\,[\; k_p \sum_{j=0}^{N-1} \binom{N-1}{j} (1-\eta)^{N-1-j} \eta^j (1-G(p))^j \\ & +\mu(1-H(p))^N \sum_{j=0}^{N-1} \binom{N-1}{j} (1-\eta)^{N-1-j} \eta^j (1-G(p))^j \;]\end{aligned}$$

must be constant. Using the binomial formula, this equation can again be simplified to

$$\Pi_p(p) = (p-c)\left[k_p\left(1-\eta G(p)\right)^{N-1} + \mu(1-H(p))^N\left(1-\eta G(p)\right)^{N-1}\right].$$

Suppose there is a $\overline{p} < r$ such that $G(\overline{p}) = 1$. Then a dealer could profitably deviate from this strategy by quoting for instance r with positive probability. This can be seen as follows. From (1.3) and $G(p) = 1$

for $p \geq \overline{p}$ it follows that

$$(r-c)k_b = (r-c)k_p\frac{\gamma}{N}(1-\eta)^N \geq (\overline{p}-c)(k_b + (1-\eta)^N\mu(1-H(\overline{p}))^{N-1}),$$

where the last inequality is an equality whenever $\overline{p}$ is in the support of H. Further algebraic manipulation gives

$$\begin{aligned}
\Pi_p(r) &= (r-c)k_p(1-\eta)^{N-1} \\
&\geq \frac{N}{\gamma(1-\eta)}(\overline{p}-c)(k_b + (1-\eta)^N\mu(1-H(\overline{p}))^{N-1}) \\
&= \frac{N}{\gamma(1-\eta)}(\overline{p}-c)(k_p\frac{\gamma}{N}(1-\eta)^N + (1-\eta)^N\mu(1-H(\overline{p}))^{N-1}) \\
&= (\overline{p}-c)(k_p(1-\eta)^{N-1} + (1-\eta)^{N-1}\frac{N}{\gamma(1-H(\overline{p}))}\mu(1-H(\overline{p}))^N) \\
&> (\overline{p}-c)(1-\eta)^{N-1}(k_p + \mu(1-H(\overline{p}))^N) \\
&= \Pi_p(\overline{p}).
\end{aligned}$$

The strict inequality in the above calculation holds, since $H(\overline{p}) < 1$ for $\overline{p} < r$, $0 < \gamma \leq 1$ and $N > 1$. It is therefore more profitable to quote r than to quote $\overline{p}$. Thus, it must be the case that $G(\overline{p}) < 1$ for $\overline{p} < r$. The supremum of the support of G is therefore given by r, since no dealer quotes above r. The equilibrium profit that results from quoting prices in the support of G on the platform is determined by the profit that results from quoting r:

$$(p-c)\left[k_p\left(1-\eta G(p)\right)^{N-1} + \mu(1-H(p))^N\left(1-\eta G(p)\right)^{N-1}\right] = (1-\eta)^{N-1}(r-c)k_p.$$

This is equation (1.4), which has to hold for prices in the support of G. □

Proof of Lemma 5. Let

$$H(p) = \int_c^r h(x)\,\mathrm{d}x, \qquad G(p) = \int_c^r g(x)\,\mathrm{d}x,$$

where $H(c) = G(c) = 0$ and

$$h(p) := \begin{cases} 0, & \text{if } (p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] < (r-c)k_b, \\ h_1(p), & \text{if } (p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] = (r-c)k_b \\ & \text{and } (p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p + \mu(1-H(p))^N\right] < (1-\eta)^{N-1}(r-c)k_p, \\ h_2(p), & \text{if } (p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] = (r-c)k_b \\ & \text{and } (p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p + \mu(1-H(p))^N\right] = (1-\eta)^{N-1}(r-c)k_p, \end{cases}$$

$$g(p) := \begin{cases} 0, & \text{if } (p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p + \mu(1-H(p))^N\right] < (1-\eta)^{N-1}(r-c)k_p, \\ g_1(p), & \text{if } (p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p + \mu(1-H(p))^N\right] = (1-\eta)^{N-1}(r-c)k_p \\ & \text{and } (p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] < (r-c)k_b, \\ g_2(p), & \text{if } (p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p + \mu(1-H(p))^N\right] = (1-\eta)^{N-1}(r-c)k_p \\ & \text{and } (p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] = (r-c)k_b, \end{cases}$$

Note that $g(p)$ and $h(p)$ are not defined whenever

$$(p-c)\left[k_b + \mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right] > (r-c)k_b$$

or

$$(p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p + \mu(1-H(p))^N\right] > (1-\eta)^{N-1}(r-c)k_p.$$

However, g and h will be defined in such a way that the last two inequalities never hold. The reason for this is as follows. Since dealers maximize their profits, by choosing r as the supremum of the support of their quoting strategies (Lemma 3) any profit that is achievable by quoting prices different from r must be lower than the profit from quoting r.

Let

$$
\begin{aligned}
\pi_H(h(p), g(p)) \quad &:= \quad \frac{\partial}{\partial p}\left[(p-c)\left(k_b + \mu(1-H(p))^{N-1}(1-\eta G(p))^N\right)\right] \\
&= \quad k_b + \mu(1-H(p))^{N-1}(1-\eta G(p))^N \\
&\quad\quad -(p-c)\mu(N-1)(1-H(p))^{N-2}(1-\eta G(p))^N h(p) \\
&\quad\quad -(p-c)\mu N(1-H(p))^{N-1}(1-\eta G(p))^{N-1}\eta g(p) \qquad (1.22)
\end{aligned}
$$

and

$$
\begin{aligned}
\pi_G(h(p), g(p)) \quad &:= \quad \frac{\partial}{\partial p}\left[(p-c)\big(1-\eta G(p)\big)^{N-1}\left(k_p + \mu(1-H(p))^N\right)\right] \\
&= \quad \big(1-\eta G(p)\big)^{N-1}\left(k_p + \mu(1-H(p))^N\right) \\
&\quad\quad -(p-c)\mu N(1-H(p))^{N-1}(1-\eta G(p))^{N-1}h(p) \\
&\quad\quad -(p-c)(N-1)\left(k_p + \mu(1-H(p))^N\right)(1-\eta G(p))^{N-2}\eta g(p). \qquad (1.23)
\end{aligned}
$$

Now $h_1(p)$ can be defined as the unique solution to $\pi_H(h_1(p), 0) = 0$, i.e. the value of the density of H at p that keeps the profit of quoting p for dealers in the bilateral market constant if p is increased marginally:

$$
h_1(p) = \frac{k_b + \mu(1-H(p))^{N-1}(1-\eta G(p))^N}{(p-c)\mu(N-1)(1-H(p))^{N-2}(1-\eta G(p))^N}.
$$

Analogously, $g_1(p)$ can be defined by $\pi_G(0, g_1(p)) = 0$, i.e.

$$
g_1(p) = \frac{\big(1-\eta G(p)\big)}{(p-c)(N-1)\eta}.
$$

When defining the densities of G and H for all p at which both quoting on the platform and quoting on the bilateral market is as profitable as quoting r, one has to distinguish three cases:

Case 1: $\pi_H(0, g_1(p)) > 0$ and $\pi_G(h_1(p), 0) \leq 0$. In this case, profits in the bilateral market from quoting p increase as p increases if dealers in the bilateral market stop quoting for prices above p while dealers on the platform choose their optimal response. However, quoting on the platform does not become more profitable if dealers on the platform stop quoting and dealers in the bilateral market respond optimally. Thus, in equilibrium, dealers on the platform indeed stop quoting, i.e.

$$g_2(p) = 0,$$

and dealers in the bilateral market respond optimally, i.e.

$$h_2 = h_1(p).$$

Case 2: $\pi_H(0, g_1(p)) \leq 0$ *and* $\pi_G(h_1(p), 0) > 0$. In this case, profits on the platform from quoting p increase as p increases if dealers on the platform stop quoting for prices above p while dealers in the bilateral market choose their optimal response. However, quoting in the bilateral market does not become more profitable if dealers in the bilateral market stop quoting and dealers on the platform respond optimally. Thus, in equilibrium, dealers in the bilateral market indeed stop quoting, i.e.

$$h_2(p) = 0,$$

and dealers in the bilateral market respond optimally, i.e.

$$g_2 = h_1(p).$$

Case 3: $\pi_H(0, g_1(p)) > 0$ *and* $\pi_G(h_1(p), 0) > 0$. In this case, profits in the bilateral market from quoting p increase as p increases if dealers in the bilateral market stop quoting for prices above p while dealers on the platform choose their optimal response. At the same time, profits on the platform from quoting p increase as p increases if dealers on the platform stop quoting for prices above p while dealers in the bilateral market choose their optimal response. In the following, it will be shown that there are unique solutions $h_2(p) > 0$ and $g_2(p) > 0$ that solve $\pi_H(h_2(p), g_2(p)) = 0$ and $\pi_G(h_2(p), g_2(p)) = 0$, i.e. both dealers on the platform and dealers in the bilateral market keep quoting and their profit from doing so stays constant. Let $X := 1 - H(p)$ and $Y := 1 - \eta G(p)$ and

$$J := (p - c) \begin{pmatrix} \mu(N-1)X^{N-2}Y^N & \mu N X^{N-1} Y^{N-1} \eta \\ \mu N X^{N-1} Y^{N-1} & (N-1)Y^{N-2}(k_p + \mu X^N)\eta \end{pmatrix}.$$

Using (1.22) and (1.23), the equations $\pi_H(h_2(p), g_2(p)) = 0$ and $\pi_G(h_2(p), g_2(p)) = 0$ can be rewritten as

$$J\begin{pmatrix} h_2(p) \\ g_2(p) \end{pmatrix} = \begin{pmatrix} k_b + \mu X^{N-1}Y^N \\ Y^{N-1}\left(k_p + \mu X^N\right) \end{pmatrix} \tag{1.24}$$

This system of equations has a unique solution whenever the determinant of J is nonzero. Since

$$\det J = (p-c)^2 X^{N-2} Y^{2N-2} \mu\eta \left((N-1)^2(k_p + \mu X^N) - \mu N^2 X^N\right),$$

if follows for $p \in (c, r)$, which implies $Y > 1 - \eta$ and $X > 0$ (see Lemma 3), that $\det J > 0$ exactly if

$$(N-1)^2(k_p + \mu X^N) - \mu N^2 X^N > 0. \tag{1.25}$$

Thus, if the last inequality holds, there is a unique solution to (1.24). Since $\pi_H(0, g_1(p)) > 0$, one has

$$k_b + \mu X^{N-1}Y^N - (p-c)\mu N X^{N-1} Y^{N-1} \eta g_1(p) > 0.$$

Plugging in the expression for $g_1(p)$ from above gives

$$k_b + \mu X^{N-1}Y^N - \frac{N}{N-1}\mu X^{N-1}Y^N > 0. \tag{1.26}$$

From $\pi_G(h_1(p), 0) > 0$, one gets

$$Y^{N-1}\left(k_p + \mu X^N\right) - (p-c)\mu N X^{N-1} Y^{N-1} h_1(p) > 0.$$

Plugging in the expression for $h_1(p)$ from above gives

$$Y^{N-1}\left(k_p + \mu X^N\right) - \frac{NX}{(N-1)Y}(k_b + \mu X^{N-1}Y^N) > 0. \tag{1.27}$$

Now, (1.27) and (1.26) imply

$$\frac{Y^N(N-1)}{NX}\left(k_p + \mu X^N\right) > k_b + \mu X^{N-1}Y^N > \frac{N}{N-1}\mu X^{N-1}Y^N.$$

Since

$$\frac{Y^N(N-1)}{NX}\left(k_p + \mu X^N\right) > \frac{N}{N-1}\mu X^{N-1}Y^N$$

is equivalent to (1.25), the system of equations (1.24) has a unique solution. This solution is given by

$$\begin{aligned} & J^{-1}\begin{pmatrix} k_b+\mu X^{N-1}Y^N \\ Y^{N-1}\left(k_p+\mu X^N\right) \end{pmatrix} \\ = \quad & \frac{p-c}{\det J}\begin{pmatrix} Y^{N-2}\eta(k_p+\mu X^N)\left[(N-1)(k_b+\mu X^{N-1}Y^N)-\mu N X^{N-1}Y^N)\right] \\ Y^{N-1}X^{N-2}\mu\left[(N-1)(k_p+\mu X^N)Y^N-NX(k_b+\mu X^{N-1}Y^N)\right] \end{pmatrix} \end{aligned}$$

For $p>r$, the condition $\pi_H(0,g_1(p))>0$ is equivalent to

$$(N-1)(k_b+\mu X^{N-1}Y^N)-\mu N X^{N-1}Y^N)>0,$$

as can be seen in (1.26). Thus, $h_2(p)>0$ for $p>r$.

The condition $\pi_G(h_1(p),0)>0$ is equivalent to

$$(N-1)(k_p+\mu X^N)Y^N-NX(k_b+\mu X^{N-1}Y^N)>0,$$

as can be seen in (1.27). Thus, $g_2(p)>0$ for $p>r$.

By construction, (1.3) holds whenever $h(p)>0$ and (1.4) holds whenever $g(p)>0$. Furthermore one has $\sup\operatorname{support} H=\sup\operatorname{support} G=r$. This can be seen as follows.

First, it cannot be the case that $G(p)\geq 1$ while $H(p)\leq 1$ for $p<r$, since that would imply that

$$(p-c)\left(1-\eta G(p)\right)^{N-1}\left[k_p+\mu(1-H(p))^N\right]<(1-\eta)^{N-1}(r-c)k_p, \tag{1.28}$$

which in turn implies that $g(p)>0$ for a set of prices of positive measure, with the last inequality holding for those prices. This has been ruled out in the definition of G.

Similarly, it cannot be the case that $H(p)\geq 1$ while $G(p)\leq 1$for $p<r$, since that would imply that

$$(p-c)\left[k_b+\mu(1-H(p))^{N-1}\left(1-\eta G(p)\right)^N\right]<(r-c)k_b, \tag{1.29}$$

which in turn implies that $h(p)>0$ for a set of prices of positive measure, with the last inequality holding for those prices. This has been ruled out in the definition of H. That both $G(p)=1$ and $H(p)=1$ for $p<r$ can be ruled out analogously.

Third, as $p \to r$ from below, it must be the case that $H(p) \to 1$. if $H(p) > \overline{H} > 1$ as $p \to r$, the inequality in (1.28) would be reversed, which is ruled out by the Definition of G and H. Thus, for every $p' < r$ there must be a $p > p'$, such that (1.3) holds and $h(p) > 0$. Thus, p can be chosen large enough such that X is sufficiently small to ensure $h(p'') > 0$ for all p'' with $r > p'' \geq p$.

An analogous argument implies that there is a $p < r$ such that (1.4) holds and that $h(p'') > 0$ for all p'' with $r > p'' \geq p$. Thus, $\sup \text{support}\, H \geq r$ and $\sup \text{support}\, G \geq r$.

The last argument implies that (1.3) and (1.4) hold if $p < r$ is above some threshold. Then $H \to 1$ as $p \to r$ implies $G \to 1$ as $p \to r$ (from (1.4)). Thus, $\sup \text{support}\, H \leq r$ and $\sup \text{support}\, G \leq r$.

Thus, it must be the case that $\sup \text{support}\, H = \sup \text{support}\, G = r$.

By construction of G, (1.17) holds for all $p \in [c, r]$. For $p \in \mathbb{R} \setminus [c, r]$, (1.17) holds as well, since the dealer either does not trade at all or sells at loss. Thus, (1.17) holds especially for all $p \in \mathbb{R} \setminus \text{support}\, G$.

By construction of H, (1.18) holds for all $p \in [c, r]$. For $p \in \mathbb{R} \setminus [c, r]$, (1.18) holds as well, since the dealer either does not trade at all or sells at loss. Thus, (1.18) holds especially for all $p \in \mathbb{R} \setminus \text{support}\, H$.

The distribution functions G and H according to which dealers optimally quote are uniquely determined. If dealers would choose densities that almost everywhere equal to g and h, (1.3) would not hold anymore on the support of H or (1.4) would not hold anymore on the support of G.

□

Proof of Lemma 6. For a fixed γ and a fixed r, Lemma 5 states that there are two unique functions G and H with the desired properties.

That at least one solution r to equation (1.19) exists for a fixed γ can be shown as follows.

Define

$$z := \left(\frac{(r-p)\gamma(1-\eta)^N(1-\mu)}{N\mu(p-c)(1-\eta G(p))^N} \right)^{1/(N-1)}.$$

Using, (1.3), one can check that $z = 1 - H(p)$. Now,

$$p = \frac{r\gamma(1-\mu)(1-\eta)^N + cz^{N-1}N\mu(1-\eta G(p))^N}{N\mu(1-\eta G(p))^N z^{N-1} + \gamma(1-\mu)(1-\eta)^N}. \tag{1.30}$$

Substituting into (1.19) gives

$$r = s_b + \int_0^1 \frac{r\gamma(1-\mu)(1-\eta)^N + cz^{N-1}N\mu(1-\eta G(p(z)))^N}{N\mu(1-\eta G(p(z)))^N z^{N-1} + \gamma(1-\mu)(1-\eta)^N} \,\mathrm{d}z, \tag{1.31}$$

Defining

$$\alpha(\gamma, r) := \int_0^1 \left(\frac{N\mu(1-\eta G(p(z)))^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1 \right)^{-1} \mathrm{d}z \tag{1.32}$$

(G still depends on r), one can rewrite (1.31) as

$$r = c + \frac{s_b}{1-\alpha(\gamma, r)}. \tag{1.33}$$

Therefore, a solution to (1.19) exists if there is a solution r to

$$(1-\alpha(\gamma, r))(r-c) = s_b \tag{1.34}$$

for some fixed γ. The left-hand side of the previous equation is continuous and goes to zero as $r \to c$, since $\alpha(\gamma, r) \in (0, 1)$ for any $r > c$ and $\gamma \in (0, 1]$. As $r \to \infty$, the left-hand side of the previous equation goes to infinity, since

$$\alpha(\gamma, r) \leq \overline{\alpha}(\gamma) := \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma(1-\mu)} + 1 \right)^{-1} \mathrm{d}z < 1$$

for any $\gamma \in (0, 1]$. Thus, since $s_b > 0$, there is at least one solution r to (1.19) for a fixed γ.

In order to show that the solution r to equation (1.19) is unique, it is shown that

$$\frac{\partial}{\partial r}\alpha(\gamma, r) = 0. \tag{1.35}$$

Then, the left-hand side of (1.34) is strictly monotone increasing in r and any solution that solves (1.34) must be unique. For $p \in \text{support}(H) \setminus \text{support}(G)$,

$$\frac{\partial}{\partial r} \left(\frac{N\mu(1-\eta G)^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1 \right)^{-1} = 0 \tag{1.36}$$

is clear, since G is constant for those prices.

For $p \in \text{support}(H) \cap \text{support}(G)$, one has to take into account that the price p changes for any fixed value of $X = 1 - H$ if r changes, so that (1.3) (1.4) hold. In order to take those joint dynamics into account, the implicit function theorem is used. The notation $Y = 1 - \eta G(p(z))$ is used and Y and p are viewed as implicit functions of r. Then one can calculate the derivative of Y with respect to r for a given value of X.

Substituting $1 - H = X$ and $1 - \eta G = Y$ in (1.3) and (1.4) and calculating the derivatives of the left-hand sides the resulting equations with respect to p and Y (the right hand sides do not depend on p or Y) gives the Jacobian M.

$$M := \begin{pmatrix} \mu Y^N X^{N-1} + k_b & N(p-c)X^{N-1}Y^{N-1}\mu \\ k_p Y^{N-1} + \mu X^N Y^{N-1} & (N-1)(p-c)Y^{N-2}\left(\mu X^N + k_p\right) \end{pmatrix}.$$

Its determinant is given by

$$\det M = (p-c)Y^{N-2}(k_p + \mu X^N)\left((N-1)k_b - \mu X^{N-1}Y^N\right).$$

If $r > p > c$, one has $0 < X \leq 1$. Thus, $\det M > 0$, if

$$(N-1)k_b - \mu X^{N-1}Y^N > 0.$$

The last inequality is implied by (1.26), which always holds if $p \in \text{support}(H) \cap \text{support}(G)$.

The implicit function therefore exists.

Analogously to the Jacobian M, the derivatives of the left-hand sides minus the right-hand sides of equations (1.3) and (1.4) with respect to r are give by f_r defined as

$$f_r := \begin{pmatrix} -k_b \\ -(1-\eta)^{N-1}k_p \end{pmatrix}.$$

Then, the derivatives of p and Y with respect to r are given by $-M^{-1}f_r$. The second row of $-M^{-1}f_r$ expresses the derivative of Y with respect to r:

$$\frac{\partial Y}{\partial r} = \frac{1}{\det M}\left[k_p X^{N-1}Y^N(1-\eta)^{N-1}\mu + k_b\left(k_p(1-\eta)^{N-1} - Y^{N-1}(k_p + X^N\mu)\right)\right]. \tag{1.37}$$

Substituting $X = 1 - H$ and $Y = 1 - \eta G$ in (1.3) and rearranging terms gives

$$X^{N-1}Y^N = \frac{r-p}{p-c}\frac{k_b}{\mu} = \left(\frac{r-c}{p-c} - 1\right)\frac{k_b}{\mu}.$$

Analogously, (1.4) implies

$$Y^{N-1}(k_p + X^N\mu) = \frac{r-c}{p-c}k_p(1-\eta)^N.$$

Thus, (1.37) is equivalent to

$$\begin{aligned}\frac{\partial Y}{\partial r} &= k_p(1-\eta)^{N-1}k_b\frac{r-c}{p-c} - k_p(1-\eta)^{N-1}k_b \\ &\quad +k_bk_p(1-\eta)^{N-1} - \frac{r-c}{p-c}k_bk_p(1-\eta)^{N-1} \\ &= 0.\end{aligned}$$

Since $\det M > 0$, one gets $\frac{\partial Y}{\partial r} = 0$, i.e. $\frac{\partial G}{\partial r} = 0$ for $p < r$, which in turn implies

$$\frac{\partial}{\partial r}\left(\frac{N\mu Y(X)^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1\right)^{-1} = 0. \tag{1.38}$$

Now (1.36) and (1.38) imply (1.35). Thus, uniqueness of the solution to (1.34) is shown.

It will now be shown that the solution r to (1.19) is strictly monotone increasing in γ. This is done by showing that

$$\frac{\partial}{\partial \gamma}\alpha(\gamma, r) > 0. \tag{1.39}$$

If $p \in \text{support}(H) \setminus \text{support}(G)$, setting $Y(p) = \overline{Y}$ in equation (1.32) constant, where as above $X = 1 - H$ and $Y = 1 - \eta G$, one can see that

$$\frac{\partial}{\partial \gamma}\left(\frac{N\mu\overline{Y}^N X^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1\right)^{-1} > 0. \tag{1.40}$$

For the case in which $p \in \text{support}(H) \cap \text{support}(G)$, the implicit function theorem is used. Recall that $k_b = \frac{\gamma}{N}(1-\eta)^N(1-\mu)$ and $k_p = 1-\mu$. Now define

$$f_g := \begin{pmatrix} \frac{(p-r)(1-\eta)^N(1-\mu)}{N} \\ 0 \end{pmatrix}$$

as the derivatives of the left-hand sides minus the right-hand sides of equations (1.3) and (1.4) with respect to γ. Viewing p and Y (as defined above) as implicit functions of γ, one can express their respective derivatives with respect to γ as $-M^{-1} \cdot f_g$, where M is defined as above. This gives the following derivative of Y with respect to γ:

$$\frac{\partial Y}{\partial \gamma} = -\frac{1}{\det M}\frac{(1-\mu)(1-\eta)^n Y^{N-1}(r-p)\left(k_p + \mu X^N\right)}{N} < 0,$$

since $\det M > 0$ for $p < r$. Thus,

$$\frac{\partial}{\partial \gamma}\left(\frac{N\mu Y(X)^N X^{N-1}}{\gamma(1-\mu)(1-\eta)^N}+1\right)^{-1} > 0. \tag{1.41}$$

Now, (1.40) and (1.41) imply (1.39). It follows from (1.34) that the reservation price r must be increasing in γ.

The continuity of the solution r to (1.19) in γ follows from equation (1.33), since α is a continuous function because it is an integral over a finite interval with a continuous and bounded integrand.

Now it will be shown that $r \to c + s_b$ as $\gamma \to 0$.

Since $G(p) \in [1-\eta, 1]$, one has $\alpha(\gamma) \geq \underline{\alpha}(\gamma) := \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma(1-\mu)(1-\eta)^N}+1\right)^{-1} \mathrm{d}z$. This gives

$$\lim_{\gamma\to 0} r \geq c + \lim_{\gamma \to 0} \frac{s_b}{1-\underline{\alpha}(\gamma)} = s_b + c, \tag{1.42}$$

since $\lim_{\gamma\to 0}\underline{\alpha}(\gamma) = 0$. On the other hand, one has $\alpha(\gamma) \leq \overline{\alpha}(\gamma) := \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma(1-\mu)}+1\right)^{-1} \mathrm{d}z$. This gives

$$\lim_{\gamma\to 0} r \leq c + \lim_{\gamma \to 0} \frac{s_b}{1-\overline{\alpha}(\gamma)} = s_b + c, \tag{1.43}$$

since $\lim_{\gamma\to 0}\overline{\alpha}(\gamma) = 0$. Taking (1.43) and (1.43) together, one gets $\lim_{\gamma\to 0} r = c + s_b$.

□

Proof of Lemma 7. Recall from the proof of Lemma 6 that the reservation price r has to satisfy (1.34). Since the term $\alpha(\gamma, r)$, as defined in (1.32), lies in $(0, 1)$, (1.34) can only hold if r is greater than v as $s_b \to v - c$, if $\gamma > 0$ is held fixed. Thus, in order to make $r = v$ hold, it must be the case that $\gamma \to 0$.

Moreover, it must be the case that γ is strictly monotone decreasing in s_b whenever $\gamma < 1$. Let $\gamma = \gamma' < 1$ such that $r = v$ ($r = v$ must be the case if $\gamma < 1$) holds for some $s_b > 0$. As shown in the proof of (1.35) in Lemma 6, the term $\alpha(\gamma, r)$, as defined in (1.32), does in fact not depend on r, holding other parameters fixed.

Thus, if

$$(1 - \alpha(\gamma'', r''))(r'' - c) = s_b + \varepsilon,$$

for some $\varepsilon \in (0, v - c - s_b)$, is supposed to hold, it must be the case that $r'' = v$ and $\gamma'' < \gamma'$, since the term $\alpha(\gamma'', r'')$ is strictly monotone increasing in γ'', as shown in the proof of (1.39) in Lemma 6.

As stated in (1.39), $\alpha(\gamma, r)$ is strictly monotone decreasing in γ. However, since $\alpha(\gamma, r) < 1$ always holds, also $\alpha(1,v) < 1$ and

$$(1-\alpha(1,v))(v-c) > s_b,$$

if s_b is sufficiently small. Thus, (1.35) can only hold, if $\gamma = 1$ and $r < v$ if $s_b \to 0$. This shows $\lim_{s_b \to 0} \gamma = 1$.

□

Proof of Lemma 8. Part 1: $\mu > \mu^* \implies \underline{q} > \underline{b}$.

Assume that $\underline{q} \leq \underline{b}$. Choosing $p = \underline{q}$ and multiplying (1.4) by

$$\frac{k_b(r-c)}{(1-\eta)^N(r-c)k_p} = \frac{\gamma(1-\eta)}{N}$$

gives

$$\frac{\gamma(1-\eta)}{N}(\underline{q}-c)\left(k_p+\mu\right) = k_b(r-c). \tag{1.44}$$

Moreover, (1.18) in Lemma 5 implies that

$$\begin{aligned}(r-c)k_b &\geq (\underline{q}-c)\left[k_b + \mu(1-H(\underline{q}))^{N-1}\left(1-\eta G(\underline{q})\right)^N\right]\\ &= (\underline{q}-c)\left[k_b+\mu\right].\end{aligned} \tag{1.45}$$

Together, (1.44) and (1.45) imply that

$$\frac{\gamma(1-\eta)}{N}(\underline{q}-c)\left(k_p+\mu\right) \geq (\underline{q}-c)\left[k_b+\mu\right].$$

Using $k_p = 1-\mu$ and $k_b = (1-\eta)^N \frac{\gamma}{N}(1-\mu)$, the last inequality implies

$$\mu \leq \mu^*.$$

Part 2: $\mu < \mu^* \implies \underline{q} < \underline{b}$.

Assume that $\underline{q} \geq \underline{b}$. Choosing $p = \underline{b}$ and multiplying (1.3) by

$$\frac{(1-\eta)^N(r-c)k_p}{k_b(r-c)} = \frac{N}{\gamma(1-\eta)}$$

gives

$$\frac{N}{\gamma(1-\eta)}(\underline{b}-c)\left(k_b+\mu\right) = (1-\eta)^{N-1}k_p(r-c). \tag{1.46}$$

Moreover, (1.17) in Lemma 5 implies that

$$\begin{aligned}(r-c)k_p(1-\eta)^{N-1} &\geq (\underline{b}-c)\left[k_p+\mu(1-\eta G(\underline{b}))^{N-1}\left(1-H(\underline{b})\right)^N\right] \\ &= (\underline{b}-c)\left[k_p+\mu\right]. \end{aligned} \tag{1.47}$$

Together, (1.46) and (1.47) imply that

$$\frac{N}{\gamma(1-\eta)}(\underline{b}-c)\left(k_b+\mu\right) \geq (\underline{b}-c)\left[k_p+\mu\right].$$

Using $k_p = 1-\mu$ and $k_b = (1-\eta)^N\frac{\gamma}{N}(1-\mu)$, the last inequality implies

$$\mu \geq \mu^*.$$

Part 3: $\mu = \mu^* \implies \underline{q} = \underline{b}$. This follows directly from Part 1 and Part 2.

Part 4: If profits per quote are higher on the platform than in the bilateral market, $\underline{q} > \underline{b}$*.*

Suppose it were the case that profits on the platform were higher on the platform than in the bilateral market and that $\underline{q} \leq \underline{b}$, i.e. $H(\underline{q}) = 0$.

A dealer in the bilateral market gives a total number of $k_b + \mu$ quotes and by (1.3) earns a profit of $(\underline{b}-c)(k_b+\mu(1-\eta G(\underline{b}))^N)$. Since a dealer on the platform gives a total number of $\eta(k_p+\mu)$ quotes and by the indifference condition (1.4) earns a profit of $(\underline{q}-c)(k_p+\mu(1-H(\underline{q})))^N$ whenever he is quoting (which he does with probability η), (1.18) in Lemma 5 gives

$$
\begin{array}{rcl}
& (\underline{q}-c) & \\
= & \eta(\underline{q}-c)\frac{(k_p+\mu)}{\eta(k_p+\mu)} & \text{platform per quote profits since } H(\underline{q})=0 \\
> & (\underline{b}-c)\frac{k_b+\mu(1-\eta G(\underline{b}))^N}{k_b+\mu} & \text{higher per quote profits on platform} \\
\geq & (\underline{q}-c)\frac{k_b+\mu(1-\eta G(\underline{q}))^N}{k_b+\mu} & \text{due to (1.18)} \\
= & (\underline{q}-c), & \text{since } G(\underline{q})=0.
\end{array}
$$

The last calculation implies $\underline{q} > \underline{q}$, a contradiction. Thus, the original assumption $\underline{q} \leq \underline{b}$ must be wrong and $\underline{q} > \underline{b}$ must hold.

Part 5: If profits per quote are higher in the bilateral market than on the platform, $\underline{q} < \underline{b}$.

Suppose it were the case that profits per quote in the bilateral market were higher than on the platform and that $\underline{q} \geq \underline{b}$, i.e. $G(\underline{b}) = 0$.

Analogously to Part 4 above, (1.17) in Lemma 5 gives

$$
\begin{array}{rcl}
& (\underline{b}-c) & \\
= & (\underline{b}-c)\frac{(k_b+\mu)}{(k_b+\mu)} & \text{bilateral market per quote profits since } G(\underline{b})=0 \\
> & (\underline{q}-c)\frac{k_p+(1-H(\underline{q})^N)\mu}{k_p+\mu} & \text{higher per quote profits in bilateral market} \\
\geq & (\underline{b}-c)\frac{k_p+(1-H(\underline{b})^{N-1})\mu}{k_b+\mu} & \text{due to (1.17)} \\
= & (\underline{b}-c), & \text{since } H(\underline{b})=0.
\end{array}
$$

The last calculation implies $\underline{b} > \underline{b}$, a contradiction. Thus, the original assumption $\underline{q} \geq \underline{b}$ must be wrong and $\underline{q} < \underline{b}$ must hold.

Part 6: If profits per quote are equal on both trading venues, $\underline{b} = \underline{q}$. This follows from Part 4 and Part 5.

Part 7: Proof of first statement in the lemma. Suppose $\mu > \mu^*$. Then It must be the case that $\underline{q} > \underline{b}$, as shown in Part 1. Now $\underline{q} > \underline{b}$ implies that profits per quote are higher on the platform than in the bilateral market. If the latter were not the case, Part 5 and Part 6 would imply that $\underline{q} \leq \underline{b}$, which contradicts $\underline{q} > \underline{b}$. On the other hand, Part 4 states that $\underline{q} > \underline{b}$ is implied by higher profits per quote on the trading platform, which proves the equivalence claimed in the first statement.

Part 8: Proof of second statement in the lemma. Suppose $\mu \leq \mu^*$. Then It must be the case that $\underline{q} \leq \underline{b}$, as shown in Part 2 and Part 3. Now $\underline{q} \leq \underline{b}$ implies that profits per quote are weakly higher in the bilateral market than on the platform. If the latter were not the case, Part 4 would imply that $\underline{q} > \underline{b}$, which contradicts $\underline{q} \leq \underline{b}$. On the other hand, Part 5 and Part 6 state that $\underline{q} \leq \underline{b}$ is implied by weakly higher profits per quote in the bilateral market, which proves the equivalence claimed in the second statement.

□

Proof of Lemma 9. Let φ_p be the function defined in the proof of Lemma 1. Then $r_b > r_p \Leftrightarrow \varphi_p(r) > s_p$, since $r_b = r$, $\varphi_p(r_p) = s_p$ and φ_p is strictly monotone increasing. Thus, $r_b > r_p$ is equivalent to

$$s_p < \varphi_p(r) = (1 - (1-\eta)^N) \int_{-\infty}^{r} (r - x)\, \mathrm{d}F(x) = (1 - (1-\eta)^N)(r - \mathbb{E}(q)),$$

since the dealers do not give quotes above r on the platform.

I now derive the upper bound expressed in (1.20). Conditional on responding, each dealer on the platform quotes according to the distribution G characterized in Lemma 4. Since each dealer only responds with probability η, the lowest quote conditional on at least one response is distributed according to the distribution F defined by

$$F(p) := \frac{1 - (1 - \eta G(p))^N}{1 - (1-\eta)^N}.$$

Rewriting equation (1.4), one gets for $p \in [\underline{q}, r]$ that

$$\begin{aligned} G(p) &= \frac{1}{\eta} - \frac{1}{\eta}\left(\frac{(1-\eta)^{N-1} \cdot (r-c)}{(p-c)} - \frac{\mu}{k_p}(1 - H(p))^N (1 - \eta G(p))^{N-1}\right)^{1/(N-1)} \\ &\geq \max\left[0, \frac{1}{\eta} - \frac{1}{\eta}\left(\frac{(1-\eta)^{N-1} \cdot (r-c)}{(p-c)}\right)^{1/(N-1)}\right] \\ &=: \underline{G}(p) \end{aligned}$$

and therefore

$$
\begin{aligned}
F(p) &\geq \underline{F}(p) \\
&:= \frac{1-(1-\eta\underline{G}(p))^N}{1-(1-\eta)^N} \\
&= \max\left[0, \frac{1-\left(\frac{(1-\eta)^{N-1}\cdot(r-c)}{(p-c)}\right)^{N/(N-1)}}{1-(1-\eta)^N}\right].
\end{aligned}
$$

It follows that $\int_{\underline{q}}^{r} p\,\mathrm{d}F(p) \leq \int_{\underline{q}}^{r} p\,\mathrm{d}\underline{F}(p)$ by first-order stochastic dominance. Performing the change of variables $p = c + \frac{(r-c)(1-\eta)^{N-1}}{(1-(1-(1-\eta)^N)z)^{(N-1)/N}}$, one gets

$$
\begin{aligned}
\int_{\underline{q}}^{r} p\,\mathrm{d}\underline{F}(p) &= \int_0^1 \left(c + \frac{(r-c)(1-\eta)^{N-1}}{(1-(1-(1-\eta)^N)z)^{(N-1)/N}}\right)\mathrm{d}z \\
&= c + (r-c)(1-\eta)^{N-1}\frac{N\eta}{1-(1-\eta)^N}.
\end{aligned}
$$

It follows that $\int_{\underline{q}}^{r} p\,\mathrm{d}F(p) \leq c + (r-c)(1-\eta)^{N-1}\frac{N\eta}{1-(1-\eta)^N}$. □

Proof of Proposition 1. The proof of this proposition consists of three steps. First, the existence of a unique equilibrium is shown. Then proofs of the statements in two bullet points are given.

Step 1: Existence of a unique equilibrium.

Holding the slow traders' reservation price strategy as described in Definition 1 with some $\gamma \in (0,1]$ and $r \in (c, v]$ fixed, Lemma 5 states that there are unique quoting strategies H and G that satisfy (1.3) and (1.4). Moreover, (1.17) and (1.17) state that dealers cannot improve their profit by deviating from H and G.

Furthermore, Lemma 6 ensures that slow traders can choose either $\gamma = 1$ (in case this choice of gamma results in a reservation price r that solves (1.6) with $r \leq 1$) or some unique value $\gamma < 1$ such that the reservation price r that solves (1.6) is equal to 1.

It remains to show that following a reservation price strategy as described in Definition 1 is optimal for slow traders. Lemma 1 states, that such a reservation price strategy optimal for slow traders if (1.16) holds. Since by their respective definitions in (1.14) and (1.6), one has $r = r_b$ and slow traders choose γ such that $r \leq v$, the first inequality in 1 holds. Now it just needs to be verified that $r_p < r_b$ holds, where r_b and r_p are defined as in (1.14) and (1.15), respectively, if and only if s_p is below a certain unique threshold. Since

φ_p as defined in the proof of Lemma 1 is strictly monotone increasing if $\varphi_p(r_p) > 0$, $\lim_{x\to\infty} \varphi_p(x) = \infty$ and $\varphi_p(0) = 0$, there is a unique threshold $\overline{s}$ such that $r_p \leq r_b$ if and only if $s_p \leq \overline{s}$. If $r_p > r_b$, the slow trader gets a payoff of $v - r$, if he starts the search in the bilateral market and accepts the first offer. This follows from (1.6) and the fact that dealers do not quote above r. However, following the reservation price strategy in Definition 1 only gives the payoff $(1-(1-\eta)^N)(v - \mathbb{E}(q) + (1-\eta)^N(v-r) - s_p < v - r$, where the last inequality follows from (1.15) and $\mathbb{E}(\min(q, r_p)) = \mathbb{E}(q)$ for $r < r_p$. Thus, following a reservation price strategy as in Definition 1 would not be optimal.

Step 2: Proof of the statement in the first bullet point.

Let

$$\overline{\mu} := \frac{(1-\eta) - \gamma(1-\eta)^N}{N - (1-\eta)^N}. \tag{1.48}$$

If $\mu > \overline{\mu}$, then $\mu > \mu^*$, where μ^* is defined as in Lemma 8, since μ^* is monotone decreasing in γ which is not greater than 1. Lemma 8 now states that profits per quote must be greater on the platform than in the bilateral market in any equilibrium with $\mu > \overline{\mu}$.

Step 3: Proof of the statement in the second bullet point.

$\mu \leq \overline{\mu}$, one has $\mu > \mu^*$ if and only if γ is sufficiently low. Due to the monotonicity of γ in s_b and the limits stated in Lemma 7, this means $\mu > \mu^*$ if and only if s_b is sufficiently high and the result now follows from the statements in Lemma 8.

□

Proof of Proposition 2. The statements (i) to (iii) are proved separately.

Proof of statement (i): In the HM, slow traders will always buy the asset on the platform if they receive a price. Therefore, slow traders buy the asset on the platform with probability $1-(1-\eta)^N$ and buy the asset in the bilateral market with probability $\gamma(1-\eta)^N$. Fast traders, however, will buy the asset wherever they find the best quote in the whole market. Thus, even if there is a response on the platform, there is a positive probability that they find a better price in the bilateral market, since the supports of G and H overlap. Fast traders will therefore buy the asset with probability $\Psi < 1-(1-\eta)^N$ on the platform and buy the asset with probability $1 - \Psi > \gamma(1-\eta)^N$ in the bilateral market, since they will almost surely buy the asset somewhere in the HM.

Proof of statement (ii): This equivalence is stated in Lemma 8 in Appendix 1.8.1.

Proof of statement (iii): Equation (1.6) states $r = \mathbb{E}(p_b) + s_b$. Rewriting equation (1.15) and using

$r_p \leq r$, which has to hold in equilibrium, gives (with F being the distribution function of q)

$$\begin{aligned} s_p &= (1-(1-\eta)^N)\int_{\underline{q}}^{r_p}(r_p - p)\,\mathrm{d}F(p) \\ &\leq (1-(1-\eta)^N)\int_{\underline{q}}^{r}(r - p)\,\mathrm{d}F(p) \\ &= (1-(1-\eta)^N)(\mathbb{E}(p_b)+s_b-\mathbb{E}(q)). \end{aligned}$$

The claim now follows from $s_p \leq (1-(1-\eta)^N)(\mathbb{E}(p_b)+s_b-\mathbb{E}(q))$.

□

Proof of Proposition 3. The statements (i) and (ii) are proved jointly. Statements (ii) to (vi) are proved separately.

Proof of statements (i) and (ii): Using a change of variables with

$$p(z) := \frac{r\gamma(1-\mu)(1-\eta)^N + cz^{N-1}N\mu(1-\eta G(p))^N}{N\mu(1-\eta G(p))^N z^{N-1} + \gamma(1-\mu)(1-\eta)^N},$$

one can express the reservation price in the HM as

$$r = s_b + \int_{\underline{b}}^{r} p\,\mathrm{d}H(p) = c + \frac{s_b}{1-\alpha(\gamma, r)},$$

where

$$\alpha(\gamma, r) := \int_0^1 \left(\frac{N\mu(1-\eta G(p(z)))^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1\right)^{-1} \mathrm{d}z.$$

One has

$$\alpha(\gamma, r) < \overline{\alpha}(\gamma) := \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma(1-\mu)} + 1\right)^{-1} \mathrm{d}z,$$

since $G(p) < 1$ for all $p < r$. It is shown in Duffie et al. (2017) that the reservation price in the PBM satisfies $r^{PBM} = c + \frac{s_b}{1-\overline{\alpha}(\gamma^{PBM})}$. If slow traders in the HM choose $\gamma = \gamma^{PBM}$, one gets

$$r = c + \frac{s_b}{1-\alpha(\gamma^{PBM}, r)} < c + \frac{s_b}{1-\overline{\alpha}(\gamma^{PBM})} = r^{PBM}. \tag{1.49}$$

Thus, holding the probability of entry of slow traders fixed, the reservation price of slow traders is lower in the HM.

Let $\gamma^{PBM} = 1$. Then

$$v \geq r^{PBM} > r \tag{1.50}$$

for any choice of γ that slow traders in the HM make. This follows from the monotonicity of r in γ shown in Lemma 6 and (1.49). By the characterization of the equilibria in the HM in Definition 2 it follows that $\gamma = 1$ is the unique equilibrium probability of continued search of slow traders in the HM.

Now, let $\gamma^{PBM} < 1$. Suppose that it were the case that $\gamma \leq \gamma^{PBM}$. Then Lemma 6 and (1.49) give

$$v = r^{PBM} > r.$$

The last result contradicts the notion of the equilibrium characterized in Definition 2, according to which $\gamma = 1$ if $r < 1$. It follows that $\gamma \leq \gamma^{PBM}$ is not a possible in equilibrium. Thus $\gamma > \gamma^{PBM}$ must hold.

This proof also showed that $r^{PBM} \geq r$ always holds: If $\gamma^{PBM} = 1$, $r^{PBM} \geq r$ due to (1.50) and if $\gamma^{PBM} < 1$, $r^{PBM} \geq r$ due to $r^{PBM} = 1$.

Proof of statement (iii):

The turnover generated by fast traders is the same in HM and PBM, since fast traders always buy the asset.

In the HM, slow traders will buy the asset on the platform with probability $1 - (1 - \eta)^N$ and in the bilateral market with probability $\gamma(1 - \eta)^N$. In the PBM, slow traders will buy the asset with probability γ^{PBM}. Since, by statement (ii), $\gamma \geq \gamma^{PBM}$, one has

$$(1 - (1 - \eta)^N) + \gamma(1 - \eta)^N \geq \gamma^{PBM},$$

where the last inequality is strict and turnover in the HM is higher, if $\gamma^{PBM} < 1$. Turnover in the HM is not strictly higher in the HM than in the PBM if and only if $\gamma^{PBM} = 1$. In this case, one must have $\gamma = 1$, due to statement (i). Now (1.49) implies $r < r^{PBM}$. This proves the last sentence of statement (iii).

Proof of statement (iv):

In the PBM, fast traders buy the asset with probability 1. In the HM, however, there is a positive chance that the lowest quote on the platform is lower than any price a dealer in the bilateral market quotes. This is the case, because the supports of G and H have the same supremum and G and H are continuous.

Suppose that slow traders would not trade less in the the bilateral market of the HM than in the PBM. This would mean $\gamma(1-\eta)^N \geq \gamma^{PBM}$. It follows that

$$\begin{aligned}\int_0^1 \left(\frac{N\mu(1-\eta G(p(z)))^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N}+1\right)^{-1} \mathrm{d}z \;\; &> \;\; \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma(1-\mu)(1-\eta)^N}+1\right)^{-1} \mathrm{d}z \\ &\geq \;\; \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma^{PBM}(1-\mu)}+1\right)^{-1} \mathrm{d}z.\end{aligned}$$

Defining r and r^{PBM} as in the proof of statement (i) and (ii) in Proposition 3, one now gets $r^{PBM} < r$. If $\gamma^{PBM} = 1$, one clearly has a contradiction, since γ cannot be greater than one. If $\gamma^{PBM} < 1$, it must be the case that $r^{PBM} = v$. But then the above results imply $r > v$. This cannot hold in equilibrium since slow traders would make a negative expected profit in the bilateral market. This contradiction proves that slow traders trade less in the bilateral market of the HM than in the PBM.

Proof of statement (v):

First, it is shown that

$$\gamma < \gamma^{PBM} \frac{(1-\eta G(\underline{b}))^N}{(1-\eta)^N}, \tag{1.51}$$

where $\underline{b}$ is defined as in (1.7). Suppose that the last inequality does not hold. Then calculating the reservation prices in the HM and PBM as in (1.49) in the proof of Proposition 3 gives

$$r > c + \frac{s_b}{1-\int_0^1 \left(\frac{N\mu(1-\eta G(\underline{b}))^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N}+1\right)^{-1} \mathrm{d}z} \geq c + \frac{s_b}{1-\int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma^{PBM}(1-\mu)}+1\right)^{-1} \mathrm{d}z} = v.$$

This is a contradiction to statement (i) in Propsition 3, which says $r \leq r^{PBM}$.

The expression for H^{PBM} in Section 1.1 gives

$$\underline{p}^{PBM} = \frac{N\mu c + \gamma^{PBM}(1-\mu)r^{PBM}}{N\mu + \gamma^{PBM}(1-\mu)}. \tag{1.52}$$

Since (1.3) has to hold for $p = \underline{b}$, one has

$$\underline{b} = c + \frac{(r-c)k_b}{k_b + \mu(1-\eta G(\underline{b}))^N} = c + \frac{(r-c)k_b}{k_b + \mu(1-\eta G(\underline{b}))^N}.$$

Now (1.51), (1.52) and $r \leq r^{PBM}$ (by statement (i)) imply that $\underline{b} < \underline{p}^{PBM}$:

$$
\begin{aligned}
\underline{b} &< c + \frac{(r-c)\gamma^{PBM}(1-\mu)}{\gamma^{PBM}(1-\mu)+\mu N} \\
&\leq \frac{r^{PBM}(1-\mu)\gamma^{PBM} + cN\mu}{(1-\mu)\gamma^{PBM}+N\mu} \\
&= \underline{p}^{PBM}.
\end{aligned}
$$

It is now shown that $\underline{b} < \underline{p}^{PBM}$ implies lower profits per quote in the bilateral market of the HM. In the PBM, a dealer's profit is equal to $(\underline{p}^{PBM} - c)(\mu + \frac{\gamma^{PBM}}{N}(1-\mu))$, since a dealer is indifferent, between any price in the support of H^{PBM}. Since a dealer is contacted by $\mu + \frac{\gamma^{PBM}}{N}(1-\mu)$ traders, the profits per quote in the PBM are given by $\underline{p}^{PBM} - c$.

In the HM, the indifference condition (1.3) implies that a dealer makes a profit equal to $(\underline{b}-c)(\mu+k_b(1-\eta G(\underline{b}))^N)$ in the bilateral market of the HM. Since a dealer receives a total of $\mu + k_b$ quotes and $G(\underline{b}) \geq 0$, if follows that the dealers profits per quote are lower than $\underline{b} - c$ and, due to $\underline{b} < \underline{p}^{PBM}$, lower than the profits per quote in the PBM.

Proof of statement (vi): First, consider slow traders. That the expected price in the bilateral market of the HM is lower than the expected price in the PBM, follows from the claim $r^{PBM} \geq r$ in statement (i) earlier in the proposition, since the slow traders' reservation price is, in equilibrium, equal to the sum of the expected price in the bilateral market and the search cost s_b.

Now consider fast traders. In the HM, fast traders choose the lowest price among all quotes on the platform and in the bilateral market. In the PBM, fast traders choose the lowest price among the quotes from all dealers. In the following, it will be shown that the lowest quote in the bilateral market of the HM is on average lower than the lowest quote in the PBM. Then it follows that the expected lowest quote in the whole HM is also lower than the expected lowest quote in the PBM.

Since $r \leq r^{PBM}$, the expected price quoted by one dealer in the bilateral market in the HM is not higher than the expected price a dealer quotes in the PBM. Using the expressions for r and r^{PBM} from the proof of statements (i) and (ii) above gives

$$
\frac{s_b}{1-\alpha(\gamma, r)} \leq \frac{s_b}{1-\overline{\alpha}(\gamma^{PBM})} \Leftrightarrow \alpha(\gamma, r) \leq \overline{\alpha}(\gamma^{PBM}).
$$

Using the respective definitions of $\alpha(\gamma, r)$ and $\overline{\alpha}(\gamma^{PBM})$ gives

$$\int_0^1 \left(\frac{N\mu(1-\eta G(p(z)))^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1 \right)^{-1} dz \leq \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma^{PBM}(1-\mu)} + 1 \right)^{-1} \mathrm{d}z. \tag{1.53}$$

Let $\varphi_{HM} : \mathbb{R} \to [0,1]$ denote the function defined by

$$\varphi_{HM}(p) := 1 - (1 - H(p))^N$$

and $\varphi_{PBM} : \mathbb{R} \to [0,1]$, defined by

$$\varphi_{PBM}(p) := 1 - (1 - H^{PBM}(p))^N.$$

Performing the substitution $z = \left(\frac{(r-p)\gamma(1-\eta)^N(1-\mu)}{(r-c)N\mu(1-\eta G(p(z)))^N} \right)^{1/(N-1)}$, one can write the expected price a fast trader gets in the bilateral market in the HM as $\int p \, \mathrm{d}\varphi_{HM}(p) = r\beta + c(1-\beta)$, where

$$\beta := \int_0^1 \left(\frac{N\mu(1-\eta G(p(z)))^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1 \right)^{-1} N z^{N-1} \, \mathrm{d}z.$$

Performing analogously the substitution $z = \left(\frac{(r-p)\gamma^{PBM}(1-\mu)}{(r-c)N\mu} \right)^{1/(N-1)}$, one can write the expected price a fast trader gets in the PBM as $\int p \, \mathrm{d}\varphi_{PBM}(p) = r^{PBM}\overline{\beta} + c(1-\overline{\beta})$, where

$$\overline{\beta} := \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma^{PBM}(1-\mu)} + 1 \right)^{-1} N z^{N-1} \, \mathrm{d}z.$$

Since $r^{PBM} \geq r$, it follows that fast traders get a better price in the HM than in the PBM if $\beta < \overline{\beta}$. The latter is indeed the case as can be seen as follows. Let Φ be the function defined by

$$\Phi(z) := \left(\frac{N\mu(1-\eta G(p(z)))^N z^{N-1}}{\gamma(1-\mu)(1-\eta)^N} + 1 \right)^{-1} - \left(\frac{N\mu z^{N-1}}{\gamma^{PBM}(1-\mu)} + 1 \right)^{-1}.$$

Then equation (1.53) implies $\int_0^1 \Phi(z) \, \mathrm{d}z \leq 0$. The case $\Phi(x) = 0$ for all $x \in (0,1)$ is not possible, since G is monotone increasing. Thus, $\Phi(x) < 0$ for some $x \in (0,1)$. If $\Phi(x) \leq 0$ for all $x \in (0,1)$, then clearly $\int_0^1 \Phi(z) N z^{N-1} \, \mathrm{d}z < 0$ and therefore $\beta < \overline{\beta}$ and the claim follows. Let $\Phi(x) > 0$ for some $x \in (0,1)$. Since $G \circ p$ is strictly monotone decreasing in z, there is a unique $x^* \in (0,1)$ such that $\Phi(x) < 0$ for $0 \leq x < x^*$ and $\Phi(x) < 0$ for $x^* < x \leq 1$. Then

$$
\begin{aligned}
\beta - \overline{\beta} &= \int_0^1 \Phi(z) N z^{N-1}\, \mathrm{d}z \\
&< \int_0^1 \Phi(z) N (x^*)^{N-1}\, \mathrm{d}z \\
&\leq 0
\end{aligned}
$$

Thus, $\beta < \overline{\beta}$ holds in any case. As discussed above, it follows that fast traders can always expect a better price in the HM than in the PBM.

□

Proof of Proposition 4. Proposition Proposition 1 states that the HM equilibrium exists if $s_p \leq \bar{s}$. In the PBM, the slow traders' payoff is given by $\gamma^{PBM}(v - r^{PBM})$, since v is their valuation of the asset and $r^{PBM} = \mathbb{E}(p_b^{PBM}) + s_b$ holds in equilibrium. In the HM, a slow trader always pays the cost s_p and receives in expectation $(1-(1-\eta)^N)(v - \mathbb{E}(q))$ from visiting the platform and $\gamma(1-\eta)^N(v-r)$ in expectation from continuing to search in the bilateral market. Let r_p be defined as in (1.15). One now gets

$$
\begin{aligned}
\gamma^{PBM}(v - r^{PBM}) &\leq (1-(1-\eta)^N)(v - r^{PBM}) + \gamma^{PBM}(1-\eta)^N(v - r^{PBM}) \\
&\leq (1-(1-\eta)^N)(v - r_p) + \gamma(1-\eta)^N(v-r) \\
&\leq (1-(1-\eta)^N)\left(v - \mathbb{E}(q) - \frac{s_p}{1-(1-\eta)^N}\right) + \gamma(1-\eta)^N(v-r) \\
&= (1-(1-\eta)^N)(v - \mathbb{E}(q)) + \gamma(1-\eta)^N(v-r) - s_p,
\end{aligned}
$$

where the second line follows from (1.16), the fact that $r = r_p$ (which can be seen from (1.14), (1.6) and $\sup \operatorname{support} H = r$), as well as $r \leq r^{PBM}$ and $\gamma \geq \gamma^{PBM}$ (see Proposition 3).

By the argument given in the main text, it follows that an HM is always efficient, if dealers want to introduce an HM.

Derivation of conditions under which the HM is more efficient:

Equations (1.10) and (1.11) imply

$$
\Omega - \Omega^{PBM} = (1-\mu)\left[\left(1-(1-\eta)^N\right)(v-c) - \left(\gamma^{PBM} - \gamma(1-\eta)^N\right)(v - c - s_b) - s_p\right]. \tag{1.54}
$$

Given that $s_p < (1-(1-\eta)^N)(v-c)$, $\Omega - \Omega^{PBM} > 0$ in (1.54) as $\gamma^{PBM} \to 0$, $s_b \to v-c$ or $\eta \to 1$.

Since $\mathbb{E}(q)$ in (1.21) is greater than c, the proof of Proposition (1) shows that $\overline{s} \leq (1-(1-\eta)^N)(v-c)$. As shown in Duffie et al. (2017), one has $\gamma^{PBM} \to 0$ as $\mu \to 0$, $N \to \infty$ or $s_b \to v-c$. Since $s_p < (1-(1-\eta)^N)(v-c) \Leftrightarrow s_p < v-c$ as $N \to \infty$, conditions 1.-4. imply that $\Omega - \Omega^{PBM} > 0$. By Proposition (1), conditions 2.-4. imply the existence of an equilibrium. If $N \to \infty$, (1.20) implies that $\mathbb{E}(q) \to c$. Since also $r = v$ holds due to $\gamma < 1$ if $N \to \infty$, (1.21) and the proof of Proposition 1 imply $\overline{s} \to (1-(1-\eta)^N)(v-c) \to (v-c)$ as $N \to \infty$. Thus, condition 1. also implies the existence of an equilibrium.

That condition 5. implies the existence of an equilibrium and the efficiency of the HM follows also from (1.54), $\gamma^{PBM} - \gamma(1-\eta)^N \leq \gamma^{PBM}(1-(1-\eta)^N) \leq 1-(1-\eta)^N$ and Proposition 1.

Derivation of conditions under which the PBM is more efficient: Let $s_p > (1-(1-\eta)^N)s_b$. If also $\gamma^{PBM} \to 1$, then $\Delta_\Omega \to (1-\mu)\left((1-(1-\eta)^N)s_b - s_p\right) > 0$.

As shown in Duffie et al. (2017), one has $\gamma^{PBM} \to 1$ if $s_b \to 0$ or $\mu \to 1$. Thus, the PBM must be efficient, if the HM equilibrium exists at all, in case conditions 6. or 7. hold.

Note that $\gamma \geq \gamma^{PBM} \geq \gamma(1-\eta)^N$ from Proposition 3 implies $\gamma \to \gamma^{PBM}$ as $\eta \to 0$. Let $s_p > 0$. Now it follows from (1.54) that $\Omega - \Omega^{PBM} \to -s_p < 0$ as $\eta \to 0$. Thus, the PBM must be efficient, if the HM equilibrium exists at all, in case conditions 8. holds.

□

Proof of Proposition 5. First a sufficient and a necessary condition for higher dealer profits in the HM are derived. Since dealers prefer the market structure that gives them the greatest aggregate profit, (1.12) and (1.13) imply that dealers strictly prefer the HM exactly if

$$(r-c)\eta N(1-\eta)^{N-1} > (r^{PBM}-c)\gamma^{PBM} - (1-\eta)^N\gamma(r-c). \tag{1.55}$$

Since $r^{PBM} \geq r$, (1.55) implies

$$\gamma^{PBM} - (1-\eta)^N\gamma < \eta N(1-\eta)^{N-1}.$$

Now $\gamma \leq 1$ implies that

$$\gamma^{PBM} < \eta N(1-\eta)^N + (1-\eta)^N < 1 \tag{1.56}$$

must necessarily hold if dealers strictly prefer the HM.

Suppose, on the other hand, that

$$\gamma^{PBM} < \frac{s_b}{r^{PBM} - c - s_b(1-\eta)^N} \eta N (1-\eta)^{N-1}. \tag{1.57}$$

From (1.14) and the fact that all quotes from dealers are above c, it follows that $r - c > s_b$. Together with $\gamma \geq \gamma^{PBM}$, it follows from (1.57) that

$$\begin{aligned} (r^{PBM} - c)\gamma^{PBM} &< s_b \left(\gamma^{PBM}(1-\eta)^N + (1-\eta)^{N-1}\eta N \right) \\ &\leq (r-c)\left(\gamma(1-\eta)^N + (1-\eta)^{N-1}\eta N \right). \end{aligned}$$

This shows the sufficiency of (1.57) for (1.55).

Derivation of conditions under which dealers prefer the HM:

As $s_b \to v - c$ or $\mu \to 0$, $\gamma^{PBM} \to 0$ as already stated in the proof of Proposition 4, implying $r^{PBM} = v$. Thus, (1.57) holds if $s_b \to v - c$ or $\mu \to 0$. If $N \to \infty$, (1.57) holds if and only if $\limsup_{N\to\infty} \frac{\gamma^{PBM}}{N(1-\eta)^{N-1}} < \frac{s_b}{r^{PBM} - c - s_b(1-\eta)^N}\eta =: M$. Assume the latter were not the case if $\eta < \frac{s_b}{v-c}$. Then, using the definition of $\overline{\alpha}$ from the proof of Proposition 3, one would get

$$\begin{aligned} \limsup_{N\to\infty} \overline{\alpha} &:= \limsup_{N\to\infty} \int_0^1 \left(\frac{N\mu z^{N-1}}{\gamma^{PBM}(1-\mu)} + 1 \right)^{-1} \mathrm{d}z \\ &\geq \limsup_{N\to\infty} \int_0^1 \left(\frac{\mu \left(\frac{z}{1-\eta}\right)^{N-1}}{(1-\mu)M} + 1 \right)^{-1} \mathrm{d}z \\ &= \int_0^{1-\eta} 1 \, \mathrm{d}z + \int_{1-\eta}^1 0 \, \mathrm{d}z \\ &= 1 - \eta. \end{aligned}$$

Since, as stated in the proof of Proposition 3, $r^{PBM} = c + \frac{s_b}{1-\overline{\alpha}}$, the result of the previous calculation would imply that $\limsup_{N\to\infty} r^{PBM} \geq c + \frac{s_b}{\eta}$. Now $\eta < \frac{s_b}{v-c}$ would imply $\limsup_{N\to\infty} r^{PBM} > v$, which cannot be the case. Thus, dealers prefer the HM if condition 3. holds. The existence of the HM equilibrium follows from $s_b < \overline{s}$. As shown in the proof of Proposition 4, one has $\overline{s} \to v - c$ if $N \to \infty$. One can analogously show that (1.56) does not hold if $N \to \infty$ and $\eta > \frac{s_b}{v-c}$. Since $\gamma^{PBM} \to 0$ if $s_b \to 0$ or $\mu \to 1$, (1.56) does not hold if conditions 4. or 5. hold. Thus, dealers would prefer the PBM if an HM equilibrium exists at all in these cases. □

Proof of Proposition 6. Proof of statement (i): As stated in the proof of Proposition 4, each of the conditions $N \to \infty$, $s_b \to v - c$ and $\mu \to 0$ implies that $\gamma^{PBM} \to 0$. Thus, only fast traders buy the asset in the PBM under these conditions. As $N \to \infty$, $s_b \to v - c$ or $\mu \to 0$, it must also be the case that $\gamma \to 0$. This follows from the expression of r in (1.33) in Proposition 6 and the fact that α as defined in (1.32) goes to 1 as $N \to \infty$ or $\mu \to 0$.

Since slow traders buy the asset on the platform with probability $1-(1-\eta)^N$, there is a jump in turnover of size $(1-(1-\eta)^N)(1-\mu)$ if $N \to \infty$ or $s_b \to v-c$. Analogously, turnover jumps by $1-(1-\eta)^N$ as $\mu \to 0$.

As $\mu \to 1$ or $s_b \to 0$, one has $\gamma^{PBM} = 1$, as already stated in the proof of Proposition 4. Now statement (ii) of Proposition 3 implies $\gamma = 1$. Thus, in both the HM and PBM, all traders fully participate in the market and turnover remains unchanged.

If $\eta \to 0$, slow traders never receive a quote on the platform and the relation $\gamma \geq \gamma^{PBM} > \gamma(1-\eta)^N$ from Proposition 3 implies $\gamma = \gamma^{PBM}$ as $\eta \to 0$. Thus, market participation remains unchanged.

Proof of statement (ii): As $\eta \to 0$ and the probability of obtaining a quote on the platform goes to zero, all traders who trade at all have to trade in the bilateral market with a probability that goes to 1.

As stated above, each of the conditions $N \to \infty$, $s_b \to v-c$ or $\mu \to 0$ implies $\gamma \to 0$ and slow traders will continue to search with a probability that goes to zero after not having obtained a quote on the platform. Then a fraction $(1-(1-\eta)^N)$ of slow traders trades at all and does so on the platform. As $\eta \to 1$ slow almost all traders receive a quote on the platform and do not continue to search. This shows the choices of trading venue of slow traders in the second and third bullet point. Since $\mu \to 1$ means that almost all traders are slow, the proof of the third bullet point is already complete. To complete the proof of the second bullet point, it remains to show that almost all fast traders will trade in the bilateral market if $N \to \infty$, $s_b \to v-c$ or $\eta \to 1$.

This is done as follows. First, $s_b \to v-c$ implies $\gamma \to 0$. It follows that $\mu > \mu^*$, where μ^* is defined as in Lemma 8, as $N \to \infty$, $s_b \to v-c$ or $\eta \to 1$. Thus, Proposition 1 and statement (iii) in Proposition 2 imply that $\underline{b}$, the infimum of support H, is smaller than $\underline{q}$, the infimum of support H.

Since (1.4) holds for all $p \in \operatorname{support} G$, it must also hold for $\underline{q} := \inf \operatorname{support} G$, i.e.

$$(\underline{q} - c)\left(1 - \eta G(\underline{q})\right)^{N-1}\left(k_p + \mu\left(1 - H(\underline{q})\right)^N\right) = (r-c)k_p(1-\eta)^{N-1}.$$

Since $G(\underline{q}) = 0$, one gets

$$\underline{q} = c + \frac{(1-\eta)^N (r-c)k_p}{k_p + \mu(1-H(\underline{q}))^N} \geq c + (1-\eta)^N (r-c)k_p,$$

where the last equation holds because of $H \geq 0$.

Similarly, the indifference condition (1.3) and the fact that the dealers' quoting strategy is optimal implies

$$(\underline{q} - c)\left(1 - \eta G(\underline{q})\right)^{N-1} \left(k_b + \mu \left(1 - \eta G(\underline{q})\right)^N \left(1 - H(\underline{q})\right)^{N-1}\right) \leq (r-c)k_b.$$

The last equations implies together with the lower bound for $\underline{q}$ that

$$\begin{aligned} \left(1 - H(\underline{q})\right)^{N-1} &\leq \frac{r - \underline{q}}{\underline{q} - c} \frac{k_b}{\mu} \\ &\leq \left(1 - (1-\eta)^{N-1} k_p\right) \frac{(1-\eta)\gamma}{N\mu}. \end{aligned}$$

Since $\frac{(1-\eta)\gamma}{N\mu} \to 0$ as $N \to \infty$, $\gamma \to 0$ or $\eta \to 1$, it follows that $\left(1 - H(\underline{q})\right)^N$ goes to zero as $N \to \infty$, $s_b \to v - c$ or $\eta \to 1$. This means that fast traders receive a quote in the bilateral market that is lower than $\underline{q}$ with probability 1. Therefore, almost all fast traders trade in the bilateral market.

Proof of statement (iii):

As $s_b \to v - c$, one has $\gamma \to 0$ and therefore $r = v$. Now (1.6) implies that $\mathbb{E}(p_b) = v - s_b \to c$. On the other hand, $\underline{q}$ remains stricly greater than c.

As $N \to \infty$, $\gamma \to 0$ implies analogously that $\mathbb{E}(p_b) = v - s_b > c$. On the other hand, (1.20)in Lemma 9 implies that $\limsup_{N\to\infty} \mathbb{E}(q) \leq c$. Since also $\mathbb{E}(q) \geq c$, one must have $\lim_{N\to\infty} \mathbb{E}(q) = c$. Thus, expected markups are lower on the platform.

As $\eta \to 1$, it analogously follows that $\lim_{N\to\infty} \mathbb{E}(q) = c$. Since the expected markup in the bilateral market stays strictly positive, expected markups are lower on the platform.

As $s_b \to 0$, one has $r \to c$. This follows from (1.33) in Proposition 6. Thus, average markups must go to zero on both the platform and the bilateral market.

Proof of statement (iv):

As shown in the proof of statement (ii), fast only fast traders trade in the bilateral market as $s_b \to v - c$. Moreover, in the proof of statement (iii) it is shown that markups for slow traders go to zero as $s_b \to v - c$. Since fast traders cannot trade at worse prices than slow traders would do, transaction prices go to c in the bilateral market as $s_b \to v - c$. Slow traders will trade at a price strictly above c on the platform.

As shown in the proof of statement (iii), $r \to c$ as $s_b \to 0$ and all quotes given must be in an arbitrarily small interval $[c, c+\varepsilon], \varepsilon > 0$.

The expression for $\underline{q}$ in the proof of statement (ii) implies $\underline{q} \to c$ as $\eta \to 1$. For any other $p \in \text{support}\, G$ the indifference condition (1.4) implies

$$G(p) = \frac{1}{\eta} - \frac{1-\eta}{\eta}\left(\frac{(r-c)k_p}{p-c}\right) \to 1$$

as $\eta \to 1$. In particular, this implies that $G(p) \to 1$ for any $p \in [c, c+\varepsilon], \varepsilon > 0$. Thus, the price at which slow traders buy is arbitrarily close to c. Since fast traders do not buy at worse prices, fast traders also pay a price arbitrarily close to c.

Proof of statement (v): As stated above, markups in the bilateral market of the HM go to zero as $\eta \to 0$. Since markups in the PBM are strictly positive, the first claim follows. If $\eta \to 0$, there is no trade on the platform and trading decisions in the bilateral market of the HM are made as in the PBM. As $s_b \to v - c$, $\gamma \to 0$, as claimed in Lemma 7, implying $\gamma^{PBM} \to 0$. Thus, $r = rPBM = v$ and, by (1.6) and the definition of r^{PBM} in Section 1.1, the expected price a slow trader has to pay in the bilateral market is given by $v - s_b$ in both the HM and PBM.

□

Proof of Proposition 7. As $N \to \infty$, $s_b \to v - c$ or $\mu \to 0$, it must be the case that $\gamma \to 0$. This is shown as in the proof of Proposition 6. Thus, (1.33) implies that r goes to infinity or $N \to \infty$, $s_b \to v - c$ or $\mu \to 0$. The only way to make $r = v$ hold is to let $\gamma \to 0$. Analogously, by (1.49) in the proof of Proposition 3, also γ^{PBM} has to go to zero as $N \to \infty$, $s_b \to v - c$ or $\mu \to 0$. By the construction of the equilibrium, $\gamma, \gamma^{PBM} \to 0$ as $N \to \infty$ implies that $r, r^{PBM} < v$ only for finitely many $N \in \mathbb{N}$, i.e. $r = r^{PBM} = v$ if N is large enough. Analogously, $r = r^{PBM} = v$ if s_b is large enough or μ is small enough. That there are now $2N$ dealers in the pure bilateral market does not change anything in this argument.

The joint dealer profits in the PBM are still given by (1.13). Thus, dealers still strictly prefer the HM exactly if (1.55) holds. Since $\gamma, \gamma^{PBM} \to 0$ as $s_b \to v - c$ or $\mu \to 0$ and $r = r^{PBM} = v$ if N or s_b are large enough, if follows that dealers prefer the HM if N or s_b are large enough. That (1.55) holds also if $N \to \infty$ and $\eta < \frac{s_b}{v-c}$, can be verified analogously to the argument given in Proposition 5: The inequality in (1.55) holds as $N \to \infty$ if and only if

$$(v-c)\eta N(1-\eta)^{N-1} > (v-c)\gamma^{PBM} - (1-\eta)\gamma$$

as $N \to \infty$, since $r = r^{PBM} = v$ if N is large enough. Thus, a sufficient condition for (1.55) to hold is that $\limsup_{N\to\infty} \frac{\gamma^{PBM}}{N(1-\eta)^{N-1}} < \eta$. Assume the latter were not the case if $\eta < \frac{s_b}{v-c}$. Then, using the definition of $\overline{\alpha}$ from the proof of Proposition 3 and recalling that there are $2N$ dealers in the PBM, one would get

$$
\begin{aligned}
\limsup_{N\to\infty} \overline{\alpha} &:= \limsup_{N\to\infty} \int_0^1 \left(\frac{2N\mu z^{2N-1}}{\gamma^{PBM}(1-\mu)} + 1 \right)^{-1} \mathrm{d}z \\
&\geq \limsup_{N\to\infty} \int_0^1 \left(\frac{2\mu \frac{z^{2N-1}}{(1-\eta)^{N-1}}}{(1-\mu)\eta} + 1 \right)^{-1} \mathrm{d}z \\
&= \limsup_{N\to\infty} \int_0^{(1-\eta)^{(N-1)/(2N-1)}} 1 \, \mathrm{d}z + \int_{(1-\eta)^{(N-1)/(2N-1)}}^1 0 \, \mathrm{d}z \\
&= (1-\eta)^{1/2} \\
&> 1-\eta.
\end{aligned}
$$

Since, as stated in the proof of Proposition 3, $r^{PBM} = c + \frac{s_b}{1-\overline{\alpha}}$, the result of the previous calculation would imply that $\limsup_{N\to\infty} r^{PBM} \geq c + \frac{s_b}{\eta}$. Now $\eta < \frac{s_b}{v-c}$ would imply $\limsup_{N\to\infty} r^{PBM} > v$, which cannot be the case. Thus, dealers must prefer the HM if condition 3 in Proposition 5 holds.

The existence of the equilibrium in the HM under the conditions stated in the proposition follows from Proposition 1, since $\overline{s} \to v - c$ as $N \to \infty$, as mentioned in the proof of Proposition 4. Mathematically and under the given assumptions, it does not make a difference whether the $2N$ trading desk in the market belong to N dealers or $2N$ dealers.

As in the proof of Proposition 4, (1.54), $\gamma^{PBM} \to 0$ and the fact that $s_p < \overline{s} \implies s_p < (1-(1-\eta)^N)(v-c)$ implies that an HM is the efficient market structure. Proposition 4 is applicable since welfare in the PBM does not depend on N beyond its dependence on r^{PBM}, γ^{PBM}.

Proof of first bullet point: As $\gamma, \gamma^{PBM} \to 0$, the difference in quotes between HM and PBM reduces to

$$
N(\eta - \mu) < 0,
$$

if and only if one has $\eta < \mu$.

Proof of second bullet point:

In the case that $N \to \infty$, (1.20) in Lemma 9 implies that $\limsup_{N\to\infty} \mathbb{E}(q) \leq c$. Since all prices are above c, one must have $\lim \mathbb{E}(q) = c$. Since $r = r^{PBM} = v$ becuase of $\gamma, \gamma^{PBM} < 1$ if N is large enough, (1.6) gives $\mathbb{E}(p_b) = \mathbb{E}(p_b^{PBM}) = v - s_b > c$ as $N \to \infty$.

□

1.9 Proof that the Ability to Renegotiate with Slow Traders Does not Change the Equilibrium

In this appendix, alternative assumptions on the dealers' behavior will be considered: In particular, if a slow trader returns to a dealer A, who has offered a quote p, after having asked another dealer B for a quote, dealer A will renegotiate. This means that dealer A will offer a new quote $p + m(p)$, where m is some nonnegative random variable that may depend on p. Suppose, as has been done in Section 3.6, that slow traders use a reservation price strategy in the sense of Definition 1. All arguments used in Section 3.6 to show that dealers quote according to continuous distribution functions whose supports have supremum r now apply. As argued in Section 3.6.2, a slow trader optimally accepts any price which cannot be improved in expectation by continuing to search. Due to renegotiation, (1.5) becomes

$$v - r = v - \mathbb{E}\left(\min(r + m(r), p_b)\right) - s_b = v - \mathbb{E}(p_b) - s_b,$$

where the second equation holds because of $p_b \leq r \leq r + m(r)$ a.s. If offered a price $p < r$, it is clearly better to accept that offer than continuing to search, in which case the trader would get

$$v - \mathbb{E}\left(\min(p + m(p), p_b)\right) - s_b \leq v - \mathbb{E}\left(\min(p, p_b)\right) - s_b < v - p,$$

where the last inequality follows from the uniqueness of any solution to (1.14), as shown in Lemma 1 and the fact that φ_b, as defined in the proof of Lemma 1, is strictly monotone increasing. Analogously, accepting any price above r is suboptimal, since a higher payoff could be achieved by searching.

Thus, even in this new modified setup, the slow traders' reservation price must be given by (1.6), if a reservation price strategy in the sense of Definition 1 is optimal for slow traders in the first place. In this case, The rest of the equilibrium derivation is identical to the derivation of the equilibrium described in the main text. As in the main text, using a reservation price strategy will be optimal for slow traders.

In the equilibrium considered in this appendix, slow traders get a payoff of

$$(1 - (1 - \eta)^N)\mathbb{E}(q) + \gamma(1 - \eta)^N \left(\mathbb{E}(p_b) - s_b\right) - s_b,$$

if they start to search on the platform and a payoff of

$$\mathbb{E}(p_b) - s_b$$

if they start to search in the bilateral market. As before, q denotes the lowest quote on the platform conditional on at least one response and p_b denotes the price a dealer in the bilateral market quotes when contacted. These payoffs are identical to those in the main text. Thus, it is optimal to start searching on the platform in the modified setup with renegotiation if and only if it is optimal to start searching on the platform in the original setup. The last step implies that there is an equilibrium with the modified setup whenever there is an equilibrium with the original setup and the equilibrium behavior of traders and dealers is identical.

Chapter 2

Electronic Trading in OTC Markets vs. Centralized Exchange (joint work with Ying Liu and Yuan Zhang)

2.1 Introduction

Trading in over-the-counter (OTC) markets is traditionally done over the phone, i.e. an investor who wants to trade an asset has to call a dealer and negotiate the price bilaterally. A recent trend in OTC markets is the growing electronification. Instead of calling dealer by dealer separately, an investor can use electronic trading platforms to send a request-for-quote (RFQ) to many dealers at once to obtain quotes at which the dealers are willing to trade. Some estimates suggest that in 2015, more than 40% of OTC-traded credit default swaps and more than 60% of OTC-traded interest rate swaps were traded electronically.[1]

Electronic trading platforms can potentially increase the connectedness between market participants and thereby make OTC markets more exchange-like. However, there remains a fundamental difference between centralized exchanges and electronic trading platforms in OTC markets. Whereas exchanges can be viewed as all-to-all platforms, electronic trading platforms in OTC markets are one-to-many platforms. On exchanges, each market participant can trade through a central limit order book with all other market participants. On electronic trading platforms, the RFQ trading protocol prescribes that only one investor can initiate a trade

[1] See for instance Stafford (2016) for a brief overview of recent developments in OTC markets.

at a time and choose one dealer to trade with. Therefore, electronic trading still incorporates many of the features of traditional bilateral trading in OTC markets.

The contribution of this paper is twofold. First, we model the trading process on trading platforms via an RFQ protocol. In our model, an investor who has some information about the assets payoff can choose a quantity to trade on the platform. In equilibrium, this quantity is informative about the asset's payoff. Our model therefore provides a theoretical foundation of information leakage on electronic trading platforms that is examined in empirical studies such as Hendershott and Madhavan (2015b) or Hagströmer and Menkveld (2016). Increasing the number of dealers who are contacted by an RFQ has three competing effects on trading costs: If an RFQ is sent to more dealers, (i) competition among dealers lowers the expected markup the investor has to pay, (ii) the investor is more likely to receive a quote in the first place, since each dealers' response is uncertain and (iii) information leakage about the asset's fundamental value increases the dealers' cost of providing the asset, which results in worse prices for the investor. If dealers respond very frequently to each RFQ, the cost of information leakage dominates the benefits of more competition and contacting only few dealers maximizes the investor's payoff. Only if the dealers' RFQ response rate is sufficiently low, an RFQ has to be sent to a certain minimum number of dealers in order for an equilibrium to exist in the first place. In an off-equilibrium analysis, we deal with the price impact an investor faces on the platform. The presence of adverse selection makes the permanent price impact on the trading platform larger than the permanent price impact in the interdealer market. This result is consistent with the findings of Collin-Dufresne et al. (2017).

Second, we determine conditions under which investors are better off trading on a centralized exchange among themselves and when they are better off in the two-tiered market structure with an electronic trading platform and an interdealer market. In our model, all investors are equally informed about the asset's fundamental value and benefits from trade in the centralized market only arise due to private values of obtaining the asset (e.g. hedging benefits). Since dealers are less informed about the asset's value, investors can also benefit from their information about the asset in the OTC market structure. The dealers are willing to trade with the more informed investor, because they expect to be able to partially offset the trade at a favorable price in the interdealer market. If private values of obtaining the asset are small, investors are better off in the OTC market structure where they can benefit from information asymmetries between them and the dealers. On the other hand, if the total mass of investors is large, information about the asset's fundamental value quickly leaks into the interdealer market. In this case, the price investors have to pay on the platform is approximately the sum of the fundamental value and a markup. Then, investors are better off in the centralized exchange where they can avoid the dealers' markups and uncertainty about transactions. Only if competition among dealers is very high, investors will prefer to trade in the OTC markets. In this case, markups are very low, a trade is very likely and dealers efficiently intermediate trades between their customers. Additionally, investors can benefit from their information advantage over dealers in the OTC market. These results extend previous research on the comparison between OTC markets and exchanges in terms of investor welfare (Babus and Kondor, 2016; Glode and Opp, 2017a). In this strand of literature, our study is the first one to specifically look at electronic trading platforms.

The paper proceeds as follows. Section 2.2 relates our paper to previous research. In Section 2.3, we explain the basic setup that is studied in Section 2.4. In Section 2.5, we slightly modify this setup to accommodate a continuum of investors and compare the two-tiered market structure to a centralized market. Concluding remarks are presented in Section 2.6. All proofs are in Appendix A.

2.2 Related Literature

Collin-Dufresne et al. (2017) empirically study the two-tiered index CDS market in the US. In the market for the most liquid index CDSs, the Dodd-Frank Act required trading via swap execution facilities (SEFs). As a result, investors trade with dealers almost exclusively via RFQs on electronic trading platforms. Dealers, on the other hand, trade among themselves via a continuous limit order book.[2] This market structure very closely corresponds to the setup we assume in our paper. The results of Collin-Dufresne et al. (2017) suggest that the permanent price impact in the D2C segment, i.e. when the investor trades on the platform, is higher than the permanent price impact in the interdealer market. These results justify our assumption that investors have some information about the asset that dealers do not have and are consistent with our result that there is information leakage from the trading platform to the dealers. Hendershott and Madhavan (2015b) empirically study what kind of bonds are traded over the phone and which bonds are traded on an electronic trading platform. Controlling for endogenous venue selection, they examine the trading costs on these two trading venues. Hagströmer and Menkveld (2016) estimate information flows between dealers and provide further empirical evidence for information leakage on trading platforms in OTC markets. Bjønnes et al. (2008) and Bjønnes et al. (2016) argue that dealers in the foreign exchange market learn from their clients' order flow and exploit this information in the interdealer market.

Babus and Parlatore (2017) and Glode and Opp (2017a) theoretically study investor welfare in OTC

[2]Block trades are exempt from the requirement to be traded on SEFs. However, most trades in the interdealer market are executed in the continuous limit order book, which also allows for mid-market matching and workup.

markets and centralized markets. Our model is different from those studies, since we specifically assume an RFQ trading protocol in the OTC market. Moreover, the information structure in our model differs from that in Babus and Parlatore (2017), since we have a common value of the asset for both investors and dealers. Compared to Glode and Opp (2017a) we allow the investor to trade continuous quantities of the asset in the OTC market. Malamud and Rostek (2014) show that decentralized exchange markets may be more efficient than centralized ones. Lester et al. (2017) show in a search-theoretic model that competition in fragmented markets may decrease welfare.

In modeling the information leakage on trading platforms, our paper relates to a large strand of literature that models how information is shared between economic agents. Notable papers in this strand of literature include Duffie and Manso (2007), Duffie et al. (2009, 2014), Andrei and Cujean (2017) and Babus and Kondor (2016). Traditionally, OTC markets are modeled as pure search markets as for instance in Duffie et al. (2005), Weill (2007b), Lagos and Rocheteau (2009b), Gârleanu (2009), Lagos et al. (2011b), Feldhütter (2005), Pagnotta and Philippon (2011) or Lester et al. (2015b). Zhu (2012b) and Duffie et al. (2016) explicitly model dealer markets. Our paper differs from all of those those papers since we consider an electronic trading platform.

Our assumption that dealers' responses on trading platforms are uncertain has been used by Jovanovic and Menkveld (2015) and Zhou (2017) to model the behavior of market makers in central limit order books to derive similar random-pricing strategies.[3]

We also draw on the techniques of noisy rational-expectations models of Grossman and Stiglitz (1980), Hellwig (1980) and Diamond and Verrecchia (1981). These models assume that agents behave competitively. Kyle (1989) shows that those models can be extended to allow for strategic traders that take their price impact into account. However, few closed-form solutions are available in this case. Since the competitive case is generally viewed as a reasonable approximation to the strategic case in large markets (Vives, 2010), we will model a competitive dealer market.

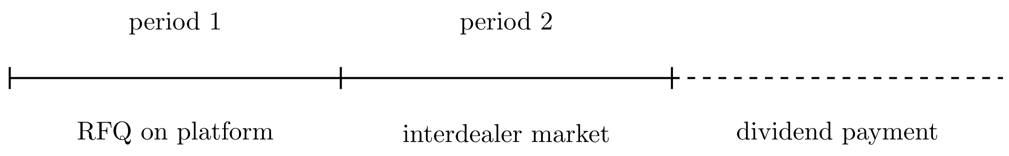
As Pagano and Röell (1996) argue, auction markets are in many ways more transparent than bilateral dealer markets. Naik et al. (1999) show that increased post-trade transparency has an ambiguous effect on dealers risk-sharing ability in two-tiered markets. Other papers which study the effects of transparency include De Frutos and Manzano (2002) and Yin (2005a). In this paper, however, we do not consider any specific disclosure policies that are enforced by regulators. In our model, information is disseminated through the different trading mechanisms.

[3]Random-pricing strategies in turn have their origin in the consumer search literature. See for instance Varian (1980a), Burdett and Judd (1983b), Stahl (1989b) and Janssen et al. (2005, 2011).

2.3 Model

There are two periods and two types of agents. In the first period, an investor can contact a number of dealers via an RFQ trading protocol on an electronic trading platform to buy or sell a quantity of an asset. In the second period, dealers trade with each other in a central limit order book. After period 2, the dividend is paid. This is illustrated in Figure 2.1.

Figure 2.1: Timeline



The asset pays an uncertain dividend $D = \theta + \varepsilon$ after the second period, where θ and ε are both independent and normally distributed random variables with zero mean and variances $\sigma_\varepsilon^2 > 0$ and $\sigma_\theta^2 > 0$, respectively. The informed investor knows the realization of θ already in the beginning of period 1. The investor also receives a private benefit $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$, with $\sigma_\delta^2 \geq 0$ for holding one unit of the asset. This private benefit is realized in the beginning of period 1 and is independent from all other random variables. Only the investor can observe δ.

There are N, $\mathbb{N} \ni N \geq 2$ dealers. On the trading platform, the investor can specify the quantity x of the asset he wants to trade. The investor also selects M dealers, with $\mathbb{N} \ni M \leq N$ from which he wants to obtain prices at which they are willing to offer quantity x of the asset. The dealers respond independently with probability $q \in (0, 1]$ to the RFQ. That dealers do not necessarily respond may reflect the cost of paying attention. We will throughout this paper assume that the number of contacted dealers M is exogenously given, i.e. the trading protocol specifies that the investor has to contact exactly M dealers. This is a slight simplification of RFQ protocols in real-world markets where investors can often freely choose a number of dealers to contact.

The dealers are ex ante identical and hold zero initial inventory in the beginning of period 1. In period 2, the aggregate supply of the asset in the interdealer market is noisy. We denote the aggregate supply of the asset in the interdealer market by W. This aggregate supply is normally distributed: $W \sim \mathcal{N}(0, \sigma_W^2)$ and $\sigma_W^2 > 0$. A noisy aggregate supply is necessary in order to prevent uninformed dealers from observing the information of informed dealers. One can interpret noise in the aggregate supply as demand from noise traders or inventory shocks to dealers' portfolios, even though there is a slight difference between inventory shocks and noisy aggregate supply. Both dealers and the investor have mean-variance preferences. There is no discounting and each agent's utility is linear in the payments made when trading the asset. Let $\overline{\omega}_k$ denote dealer k's final inventory in the end of period 2 and let Z_k denote the sum of all payments made or received by dealer k from trading the asset. Then dealer k's utility in the end of period 2 with final inventory $\overline{\omega}_k$ is given by

$$U_d(\overline{\omega}_k, Z) = \overline{\omega}_k \cdot \mathbb{E}(D|\mathcal{I}_k) - \frac{\gamma_d}{2} \cdot \overline{\omega}_k^2 \cdot \mathbb{V}(D|\mathcal{I}_k) - Z_k, \tag{2.1}$$

where $\gamma_d > 0$ is the dealers' risk-aversion parameter. The expectation and the variance in equation (2.1) are taken with respect to each dealer k's specific information set $\mathcal{I}_k$, which will be determined later. Equation (2.1) says that dealers care linearly about the mean of their expected dividend payment in the end of period 1 and sum they have to pay in both period 1 and 2. They also have to pay an inventory cost which is increasing in the expected variance of the dividend payment. This inventory cost depends on the risk-aversion parameter $\gamma_d > 0$. It is clear that equation (2.1) can be derived from a first-order condition of an exponential utility function. We specifically do not assume exponential utility because an exponential utility function and the functional form specified in equation (2.1) have different implications for the equilibrium on the platform. On the platform, a dealer has to take into account the possibility of being undercut by another dealer when giving quotes to the investor. The model becomes more tractable, if the dealers' utility is linear in the payments made when trading. We will assume that the dealers follow symmetric strategies on the platform and symmetric and linear strategies in the interdealer market.

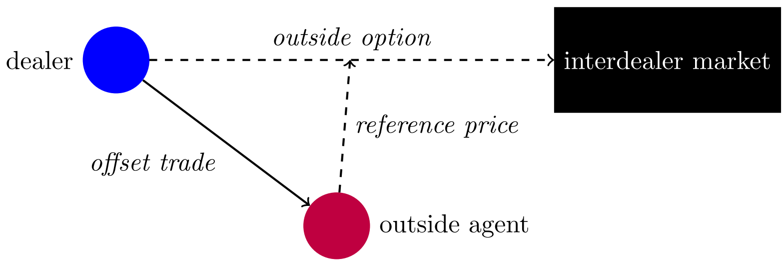
Similar to the dealers, the investor has mean-variance preferences. The investor, however, receives the private benefit δ per unit of the asset held. If the investor buys a quantity $x_1 \in \mathbb{R}$ on the platform at price p_1, the investor's utility is given by

$$U_I(x_1, p_1) = x_1 \cdot (\theta + \delta) - \frac{\gamma_I}{2} \cdot x_1^2 \cdot \sigma_\varepsilon^2 - p_1 x_1, \tag{2.2}$$

where $\gamma_I > 0$ is the investor's risk-aversion parameter. Comparing (2.1) and (2.2), note that the investor's expectation of the dividend payment and its variance is given by θ and σ_ε, respectively. On the other hand, dealers potentially learn about the dividend from the other agents and thus have a less trivial information set $\mathcal{I}_k$ for each dealer k. Also, dealers trade with each other, which results in a more complex final inventory $\overline{\omega}_k$ and more complex total payments Z_k for each dealer k.

When dealing with the case of one investor in Section 2.4, we need to make a technical assumption in order to keep the model tractable. In Section 2.4, we will assume the presence of an "outside agent". If the investor contacts $M < N, M > 0$ dealers on the platform, these M dealers will learn from the investor about the the realization of θ. Thus, there will be informed and uninformed dealers in the interdealer market. In the interdealer market, the uninformed dealers may then make inferences about the dividend level from the observed market price. It will turn out that this price is affected by both the dealers' inventories and the informed dealers' expectation of the dividend payment. In order to keep this inference problem tractable, we make the dealers' inventory independent of the expected dividend level. To this end, we assume that a dealer who traded on the platform with the investor offsets this trade with the outside agent, who does not participate in the interdealer market. The outside agent does not behave strategically. The price at which the dealer offsets his trade with the investor is such that the dealer is indifferent between trading with the outside agent and going directly to the interdealer market. This way, we keep the dealers' inventories independent of the dividend level and still keep the key economic trade-offs that the dealers and the investor face in our model. This setup is summarized in Figure 2.2. After the trader has offset his trade with the outside agent, all dealers start to trade in the interdealer market. A version of our model without the outside agent will be studied in Section 2.5.

Figure 2.2: The outside agent



2.4 Equilibrium with one investor

The equilibrium is determined by backward induction. The first step is to establish the equilibrium in the interdealer market. We will assume and later verify that the investor reveals a noisy signal about the dividend level θ to the dealers he contacts. After an equilibrium in the interdealer market has been established, dealers on the platform can anticipate their expected final payoff conditional on the quantity they trade on the platform. This payoff will ultimately be a key determinant of the expected price for the asset on the platform which is derived by standard auction-theoretic arguments. Using the derived quoting strategies of the dealers and assuming that the quantity the investor wants to trade is linear in $\theta + \delta$, an equilibrium on the trading platform can be constructed.

2.4.1 The equilibrium in the interdealer market

The equilibrium in the interdealer market considered in this paper is a rational expectations equilibrium in linear demand schedules as first studied by Grossman and Stiglitz (1980). This means that dealers behave competitively. Even though not completely realistic, this assumption can be viewed as a rather good approximation in the case of large interdealer markets.

Let x_k denote the quantity of the asset that dealer k buys in the interdealer market. Since we assume that a dealer who trades on the platform offsets his trade with a outside agent, the final inventory $\overline{\omega}_k$ of dealer k is equal to the traded quantity in the interdealer market: $\overline{\omega}_k = q_k$ for all $k \in \{1, ..., N\}$.

Since $M \leq N$ dealers have been contacted on the platform, there will be M informed dealers, who observe $\theta + \delta$ from the investor's demand. The other $N - M$ dealers are uninformed and will use the market price to make inferences about the dividend level. In the following, we will represent all dealers by the set $\{1, ..., N\}$ and say that dealer k is informed if $k \leq M$. Conversely, we say that dealer k is uninformed if $k > M$.

Let p_2 denote the price for the asset in the interdealer market. Differentiating the dealer's utility (2.1) with respect to q_k and using $\frac{\partial Z_k}{\partial q_k} = p_2$ gives the first-order condition

$$\mathbb{E}(D|\mathcal{I}_k) - \gamma_d \overline{\omega}_k \mathbb{V}(D|\mathcal{I}_k) - p_2 = 0. \tag{2.3}$$

Since $\overline{\omega}_k = q_k$, the second order condition is $-\gamma_d \mathbb{V}(D|\mathcal{I}_k) < 0$. The second order condition always holds, since $\gamma_d > 0$ and $\mathbb{V}(D|\mathcal{I}_k) \geq \sigma_\varepsilon^2$. If dealer k receives the signal $s_d := \theta + \delta$, one obtains by standard Bayesian updating that

$$\xi := \mathbb{E}(D|s_d) = \frac{\sigma_\theta^2 s_d}{\sigma_\theta^2 + \sigma_\delta^2}. \tag{2.4}$$

Similarly, one obtains

$$\tau_\xi := \frac{1}{\sigma_\xi^2} := \frac{1}{\mathbb{V}(D|s_d)} = \frac{1}{\frac{\sigma_\theta^2 \sigma_\delta^2}{\sigma_\theta^2 + \sigma_\delta^2} + \sigma_\varepsilon^2}, \tag{2.5}$$

where we defined τ_ξ and σ_ξ^2 as the precision and the variance of the dividend payment based on the informed dealers' information that includes the signal s_d.

The first order condition (3.5) now implies the following demand schedule:

$$q_k = \frac{\tau_\xi(\xi - p_2)}{\gamma_d} \qquad \text{for } k \leq M. \tag{2.6}$$

If dealer k is uninformed, his demand is assumed to be of the form

$$q_k = \frac{\mathbb{E}(D|p_2) - p_2}{\gamma_d \mathbb{V}(D|p_2)} \qquad \text{for } k > M. \tag{2.7}$$

Equation (2.7) takes into account that uninformed dealers can only learn about the conditional distribution of D by observing the market price p_2. We will use the standard approach to conjecture a price that is linear in ξ and the aggregate supply of the asset W:

$$p_2 = a\xi + bW, \tag{2.8}$$

with $a, b \in \mathbb{R}$. Then, uninformed dealers can use the normal projection theorem to calculate $\mathbb{E}(D|p_2)$ and $\mathbb{V}(D|p_2)$.

In equilibrium, also the market clearing condition

$$\sum_{k=1}^{N} q_k = W \tag{2.9}$$

has to be satisfied. Using (2.6) and (2.7) in (2.9) determines the market clearing price. Matching of coefficients in the obtained expression for the market clearing price with the coefficients in the conjectured expression (2.8) then gives the rational expectations equilibrium price function. This price function in turn determines the uninformed dealers' equilibrium demand schedules.

The following Proposition confirms the existence of an equilibrium in the interdealer market and states the corresponding expressions for equilibrium price.

Proposition 8. *There is always a rational expectations equilibrium such that the market clearing price is given by (2.8). Define*

$$\rho : \;=\; \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}, \tag{2.10}$$

$$\tau_u : \;=\; \frac{1}{Var(D|p_2)} = \frac{1}{\sigma_\theta^2 + \sigma_\epsilon^2 - \psi\rho\sigma_\theta^2}, \tag{2.11}$$

$$\psi : \;=\; \frac{a^2\rho\sigma_\theta^2}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2} = \frac{\rho}{\rho + \frac{\gamma_d^2\sigma_W^2}{M^2\tau_\xi^2\sigma_\theta^2}} \tag{2.12}$$

Then a and b are given by

$$a = \frac{M\tau_\xi + (N-M)\psi\tau_u}{M\tau_\xi + (N-M)\tau_u} \tag{2.13}$$

and

$$b = -\frac{\gamma_d}{M\tau_\xi} a, \tag{2.14}$$

One has $a > 0$ if $M > 0$. One also has $a \leq 1$ with strict inequality if $M < N$.

The fact that $a < 1$ for $M < N$ means that the price in the interdealer market is inefficient in the sense that the price does not fully reflect the informed dealers' information. In the absence of private benefits for the investor ($\sigma_\delta^2 = 0$), dealers are only willing to trade with the investor because of this informational inefficiency in the interdealer market.

2.4.2 The equilibrium on the trading platform

The equilibrium on the trading platform is derived as follows. We will assume that dealers who are contacted by the investor can observe $s_d = \theta + \delta$ and therefore form a conditional expectation of θ given by ξ as defined in (2.4). We will then use Proposition 8 and the optimal demand schedules (2.6) to determine the lowest price at which a dealer is willing to sell (or the highest price at which he is willing to buy) a given quantity of the asset. The dealers then infer from the investor's utility function the maximum markup they can charge. In equilibrium, dealers will charge a random markup on the platform. The expectation of this price can be used to determine the investor's equilibrium strategy that reveals s_d.

Assume an investor submitted an RFQ to M dealers on the platform to buy x units of the asset (if $x < 0$, the investor wants to sell). If a dealer is contacted on the platform, but does not trade the asset, he will observe s_d and will therefore be informed in the interdealer market, expecting a dividend level of ξ. Let $V_{d,1} : \mathbb{R}^2 \to \mathbb{R}$ denote the function that maps the expectation ξ and dealer k's traded quantity to dealer k's expected utility that he will get after period 2. The dealer will anticipate that the price p_2 is a linear function of ξ and W as stated in Proposition 8. Now, the optimal demand schedule (2.6) and the dealers utility function (2.1) imply the following payoff from not trading (i.e. from trading quantity 0):

$$V_{k,1}(\theta, 0) := \mathbb{E}_k \left[q_k(D - p_2) - \frac{\gamma_d}{2}\sigma_\xi^2 q_k^2 \right] = \frac{\xi^2(1-a)^2 + b^2\sigma_W^2}{2\gamma_d\sigma_\xi^2}.$$

We now consider the case in which dealer k sells quantity x to the investor[4] and goes directly to the interdealer market, while other dealers think that dealer k already offset his trade with the outside agent. Now, dealer k has the initial inventory $-x$ in the beginning of period 2. However, only dealer k knows that.

The following result states the expected price in the interdealer market for dealer k and dealer k's optimal demand.

Lemma 10. *Assume dealer k traded quantity $x \neq 0$ with the investor on the platform and directly goes to the interdealer market. Let the other dealers believe, dealer k offset his trade before going to the interdealer market. Then, according to dealer k's information, the price in the interdealer market is given by*

$$p_2 = a\xi - bx + bW$$

and his optimal demand schedule is given by

$$q_k = \frac{\xi - p_2}{\gamma_d \sigma_\xi^2} + x,$$

where a and b are defined as in Proposition 8.

Using Lemma 10, one can calculate dealer k's expected utility if he goes directly to the interdealer market holding a quantity $-x \neq 0$ and having expectation about the dividend payment ξ.

$$V_{k,1}(\xi, x) := \mathbb{E}_k\left[D(q_k - x) - p_2 q_k - \frac{\gamma_d}{2}\sigma_\xi^2 (q_k - x)^2\right] = \frac{\xi^2(1-a)^2 + 2(1-a)b\xi x + b^2\sigma_W^2 + b^2x^2}{2\gamma_d\sigma_\xi^2} - (a\xi - bx)x.$$

Comparing $V_{k,1}(\xi, x)$ and $V_{k,1}(\xi, 0)$ one can observe that the dealer expects a different return from holding a final inventory due to a different expected price. The second term in $V_{k,1}(\xi, x)$ represents the additional payment a dealer has to make to offset his inventory x in the interdealer market. We define

$$p_c(x) := \frac{V_{k,1}(\xi, 0) - V_{k,1}(\xi, x)}{x} \qquad k \le M \tag{2.15}$$

as the break-even price for any contacted dealer k. A dealer who charges $p_c(x)$ per quantity of the asset and sells x units to the investor, does not change his final utility. The payment from the investor exactly matches the difference in utility due to different inventory holdings. Analogously, we define

[4] If $x < 0$ the dealer is buying from the investor.

$$p_v(x) := \theta + \delta - \frac{\gamma_I}{2} x \sigma_\varepsilon^2 = \frac{\xi(\sigma_\theta^2 + \sigma_\delta^2)}{\sigma_\theta^2} - \frac{\gamma_I}{2} x \sigma_\varepsilon^2 \tag{2.16}$$

as the price at which the investor is indifferent between trading and not trading the asset. As one can immediately verify, equation (2.2) implies $U_I(x, p_v(x)) = 0$. One can interpret $p_c(x)$ as the cost for each contacted dealer of supplying x units of the asset. Analogously, $p_v(x)$ is the investor's value of acquiring x units of the asset. The investor can only trade a certain quantity $x > 0$ with a dealer if $p_v(x) \geq p_c(x)$. Analogously, it has to hold that $p_c(x) \geq p_v(x)$ if $x < 0$.

In the following, we assume that dealers follow symmetric strategies when giving a quote to the investor. This approach is standard, since dealers are ex-ante identical. In the appendix we show that standard search-theoretic arguments imply that the price a dealer quotes on the platform for a certain quantity x has to be a continuous random variable if $p_v(x) \neq p_c(x)$. Let $F_x : \mathbb{R} \to [0, 1]$ denote the distribution of the price a dealer quotes on the platform conditional on the quantity x that the investor wants to trade. If $x > 0$ and $p_v(x) > p_c(x)$, then $p_v(x)$ will turn out to be the supremum of the support of F_x. That quoting a higher price than $p_v(x)$ cannot be optimal follows from $U_I(x, p) < 0$ for $p > p_v(x)$ and $x > 0$. The investor would not be willing to buy the asset at such a price since doing so would make him worse off. Analogously, $p_v(x)$ is the infimum of the support of the distribution of quoted prices if $x < 0$. The investor would not be willing to sell the asset at a lower price.

Dealers are only willing to quote random prices if the expected profit they make is the same for any price in the support of F_x. If $p_v(x)$ is in the support of F_x, this indifference condition means that

$$x(p - p_c(x)) \sum_{j=0}^{M-1} \binom{M-1}{j} (1-q)^{M-1-j} q^j (1 - F_x(p))^j = (1-q)^{M-1} (p_v(x) - p_c(x)) x \tag{2.17}$$

has to hold for all $p \in \text{support}(F_x)$. The left-hand side of equation (2.17) describes the expected profit a dealer makes by quoting any $p \in \text{support}(F_x)$. The payment $x(p - p_c(x))$ in excess of the indifference level $x p_c(x)$ is weighted by the probability that the dealer has the best quote among all dealers that respond to the RFQ. Since the response of a dealer is uncertain and occurs with probability $q < 1$, one has to consider the cases in which $j = 0, ..., M - 1$ other dealers respond. The right hand side describes the expected profit for a dealer that quotes $p_v(x)$. Since F_x is continuous, this dealer will only sell the asset if no other dealer responds to the RFQ. This happens with probability $(1-q)^{M-1}$. In this case, the dealer's utility will increase by $x(p_v(x) - p_c(x)) > 0$.

The following result gives the closed-form expression for the distribution function F_x that solves (2.17) for any x with $x(p_v(x) - p_c(x)) > 0$. The last inequality is a necessary condition for the existence of strictly positive benefits of trade between dealers on the platform and the investor. In the statement of Lemma 11, we will leave implicit that p_c and p_v depend on x. In Section 2.5, we will study a version of the model in which the dealers' cost p_c does not depend on x. Since Lemma 11 holds irrespective of what variables p_v and p_c depend on, we will state it without reference to any of those variables.

Lemma 11. *Let p_c be the dealers' cost of providing a certain quantity $x \in \mathbb{R} \setminus \{0\}$ of the asset and let p_v denote the investor's value of acquiring x units of the asset. Let the investor submit an RFQ to $M \geq 2$ dealers on the platform to trade quantity x with $x(p_v - p_c) > 0$. Let $q < 1$. Assume that dealers who get contacted know θ.*

If a dealer responds to an RFQ, he will charge a random price that is distributed according to the distribution function F_x. This function is defined by

$$F_x(p) := \frac{1}{q} - \frac{1-q}{q}\left(\frac{p_v - p_c}{p - p_c}\right)^{1/(M-1)}. \tag{2.18}$$

If $x > 0$, the support of F_x is given by $[\overline{p}_x, p_v]$, where $\overline{p}_x$ is determined by $F_x(\overline{p}_x) = 0$ and satisfies $\overline{p}_x > p_c$.

If $x < 0$, the support of F_x is given by $[p_v, \overline{p}_x]$, where $\overline{p}_x$ is again determined by $F_x(\overline{p}_x) = 0$ and satisfies $\overline{p}_x < p_c$.

In this case, the expected price the investor has to pay for the asset conditional on at least one response to the RFQ is given by

$$P(x) := \mathbb{E}(p_1 \mid x, \text{at least one response}) = \int_{\text{support}(F_x)} p \mathrm{d}G_x(p) = p_c + \kappa(p_v - p_c), \tag{2.19}$$

where the distribution $G_x : \mathbb{R} \to [0,1]$ is defined by

$$G_x(p) := \frac{1 - (1 - qF_x(p))^M}{1 - (1-q)^M}$$

and

$$\kappa := \frac{Mq(1-q)^{M-1}}{1-(1-q)^M} < 1 \in [0, 1) \tag{2.20}$$

If $q = 1$, Bertrand competition implies that dealers have to set a price that equal to their cost p_c. Thus, the above expression for $P(x)$ holds for all $q \in (0, 1]$.

Equation (2.19) states that the expected price the investor receives on the platform is equal to the dealers cost p_c plus a fraction of the total gains from trade $p_v - p_c$. The fraction of this surplus that the investor has to pay is equal to κ, defined as in (3.29). Thus, κ can be viewed as the endogenously determined bargaining power of the dealers. By taking derivatives, it can be shown that κ is decreasing in M and q, which is consistent with economic intuition. As M becomes larger, competition among the dealers for the business of the investor increases. This competition is also higher, if the presence of other dealers on the platform becomes more likely.

Note that the results in Lemma 11 required the assumptions that $x(p_v - p_c) > 0$ and that contacted dealers observe $\theta + \delta$. In the remaining part of this section we will derive an optimal strategy of the investor that allows both assumptions to hold in equilibrium. We will restrict the possible strategies of the investor to strategies that are linear in the sum $\theta + \delta$. This means that the quantity the investor wants to trade is a (positive) multiple of $\theta + \delta$. It is obvious that dealers then can infer $\theta + \delta$ from the quantity the investor wants to trade. However, it is a nontrivial result that the investor finds it indeed optimal to reveal $\theta + \delta$ and the associated information about θ through his choice of the quantity x. The reason why such an equilibrium is possible, even as the private value δ becomes negligible, lies in the fact that the parameter a as defined in Proposition 8 is generally less than one. If the investor reveals a given value of $\theta + \delta$ to the dealers, the dealers expect a dividend payment equal to ξ as defined in (2.4). The price for the asset in the interdealer market will be $\xi a < \xi$ in expectation. This price in the interdealer market determines the cost for dealers of providing the asset, which according to Lemma 11 determines the expected price the investor receives on the platform. If $a < 1$, the quotes the investor gets on the platform are less sensitive to θ than the investor's utility. This makes an equilibrium possible in which the investor partially reveals his information θ to the dealers.

We now conjecture that the investor's demand for the asset on the platform is given by

$$x = \alpha(\theta + \delta), \tag{2.21}$$

for some $\alpha \in \mathbb{R}$. In the appendix we show that the expected price $P(x)$ from Proposition 11 is linear in x and ξ:

$$P(x) = \beta_1 \xi + \beta_2 x, \tag{2.22}$$

with $\beta_1, \beta_2 \in \mathbb{R}$. From the investor's conjectured strategy (2.21), the contacted dealers infer $\theta + \delta = \frac{x}{\alpha}$.

Using (2.2), (2.22) and (2.4), the investor's problem therefore becomes

$$\max_{x\in\mathbb{R}}\left[(\theta+\delta)x - x^2\frac{\gamma_I}{2}\sigma_\varepsilon^2 - x\left(\beta_1\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\frac{x}{\alpha}+\beta_2 x\right)\right]. \tag{2.23}$$

Note that (2.23) considers the investor's expected payoff conditional on at least one response to the RFQ. Since the probability of this event is exogenous and always gives a zero payoff, it can be neglected. The first-order condition for (2.23) implies the investor's optimal demand schedule

$$x = (\theta+\delta)\frac{\alpha(\sigma_\theta^2+\sigma_\delta^2)}{2\alpha\beta_2(\sigma_\theta^2+\sigma_\delta^2)+\alpha\gamma_I\sigma_\varepsilon^2(\sigma_\theta^2+\sigma_\delta^2)+2\beta_1\sigma_\theta^2}. \tag{2.24}$$

Therefore, the investor's optimal demand is indeed linear in $\theta+\delta$. Matching the coefficient in (2.24) with the conjectured strategy (2.21) gives

$$\alpha = \frac{\sigma_\theta^2+\sigma_\delta^2-2\beta_1\sigma_\theta^2}{(2\beta_2+\gamma_I\sigma_\varepsilon^2)(\sigma_\theta^2+\sigma_\delta^2)}. \tag{2.25}$$

The following proposition summarizes these results and states formal conditions under which the equilibrium exists.

Proposition 9. *The expected price on the platform $P(x)$ from Lemma 11 is linear in ξ and x, as stated in (2.22). Let $M \geq 2$. If*

$$\kappa < \frac{1}{2}, \tag{2.26}$$

with κ as in (3.29), there is a threshold $\overline{a} > 0$, such that the equilibrium on the platform described below exists if and only if $a < \overline{a}$. The last condition holds as $N \to \infty$ and $\sigma_W^2 \to \infty$ or as $\sigma_\delta \to \infty$. The inequality in (2.26) will always hold for all $M \geq 2, \sigma_\theta^2, \sigma_\delta^2 > 0$ if $q \to 1$. If (2.26) does not hold, the equilibrium does not exist.

The equilibrium is characterized as follows. The investor submits a demand x as determined in equations (2.21) and (2.25). The dealers quote independently with probability q according to the distribution function F_x in (2.18).

One furthermore has $0 < \beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$, $\beta_2 > -\frac{\gamma_I}{2}\sigma_\varepsilon^2$ and $\alpha > 0$ in each such equilibrium.

With the RFQ trading protocol, an equilibrium with linear strategies described in Proposition 9 is possible even though a linear equilibrium in double auctions and two strategic traders would not exist due to correlated values as Du and Zhu (2017) show. With the RFQ trading protocol, only the investor has the option to avoid price impact by reducing his demand. The dealers have to take the traded quantity as given and can merely charge a markup in addition to their cost of providing the asset.

We will illustrate the results derived so far with an example.

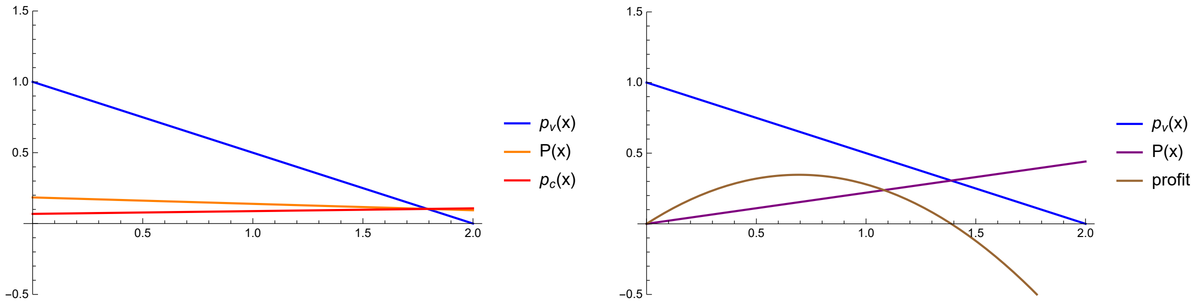
2.4.3 A brief example

For illustrative purposes we fix the exogenous parameters as follows: $N = 100$, $M = 10$, $\sigma_\varepsilon = 1$, $\sigma_W = N$, $\sigma_\theta = 1$, $\gamma_d = 1$, $\gamma_I = 1$, $q = 0.3$. To illustrate the economic mechanism of our model, we first consider the case in which $\theta + \delta = 1$, which corresponds to a realization one standard deviation above the mean. Afterwards we consider the case when $\theta + \delta = -1$. It is sufficient to only consider the sum of the common value and the investor's private, since both the investor's demand and the dealers' inferences depend only on this sum.

In Figure 2.3, $\theta + \delta$ has the high realization. In Panel (a) we plot the price $p_v(x)$ that the investor is willing to pay for x units of the asset. If the absolute value of x is small, this price is approximately equal to $\theta + \delta$, since the cost of bearing risk is small. The price $p_v(x)$ is linearly decreasing in x because of the quadratic cost of bearing risk. We also plot the dealer's cost $p_c(x)$ of providing x units of the asset, if they believe the dividend payment is normally distributed with mean ξ and precision σ_ξ^2, as defined in (2.4) and (2.5). One can see that this cost is slightly increasing in x, which represents the difficulty of offsetting the trade in the interdealer market or with the outside agent, respectively. The average price the investor can expect conditional on at least one response to the RFQ, $P(x)$, is between the other two curves.

In Panel (b) of Figure 2.3 we keep the investor's reservation price $p_v(x)$, but now look at the average price dealers quote when they infer the expected dividend level ξ from the investors demand x, one can see that this price increases faster in x than the cost $p_c(x)$ in Panel (a). We also plot the profit the investor gets for demanding a certain quantity x. This profit is the solution to problem (2.23) weighted by the probability of at least one response to the RFQ. We see that the optimum is approximately at $x = 0.69$. This also turns out to be the value of α. Thus, Panel (b) illustrates that the investor has indeed no incentive to deviate from the equilibrium strategy determined in the last section.

Figure 2.3: High realization of $\theta + \delta$

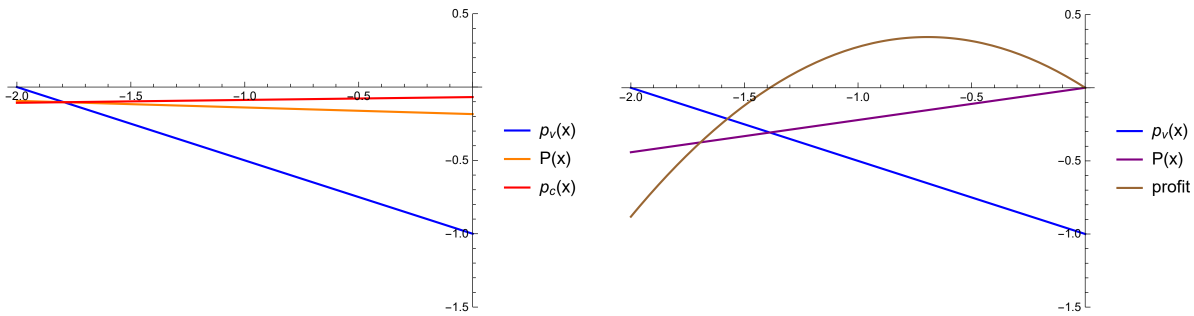


a: Expected prices and reservation prices if ξ is held fixed.

b: Dealers infer ξ from the investor's demand

In Figure 2.4, we consider the low realization of $\theta + \delta$. Comparing Panel a to Panel a in Figure 2.3, we observe that all curves have been shifted downwards by a constant. The curve of $p_v(x)$ has been shifted downwards more than the curve of $p_c(x)$. This has two reasons. First, the dealers' expectation ξ is a weighted average between $\theta + \delta$ and zero, as (2.4) shows. Second, the dealers do not find it as costly to hold a bad asset as the investor does. The dealers expect to be able to resell the asset again at a favorable price, since there are many uninformed dealers in the interdealer market. Panel (b) of Figure 2.4 shows a similar picture as Panel (b) of Figure 2.3. In Figure 2.4, however, the investor sells the asset at a negative expected price. The investor finds it profitable to do so, since $p_v(x)$ indicates that he would be willing to sell the asset at an even lower price due to the negative expected dividend. The equilibrium strategies have not changed in Figure 2.3 and Figure 2.4. Therefore, the optimal demand in Figure 2.4 is the negative of the optimal demand in Figure 2.3, since the respective realizations of $\theta + \delta$ have the same absolute value in both cases.

Figure 2.4: Low realization of $\theta + \delta$



a: Expected prices and reservation prices if ξ is held fixed.

b: Dealers infer ξ from the investor's demand.

2.4.4 Competition vs. information leakage

In this section we take a closer look at the equilibrium described in Proposition 9. Specifically, we take a look how the investor's profits from trading on the platform are affected by varying the number of dealers who are contacted on the trading platform.

We define π_I as the investor's ex-ante expected payoff in the equilibrium described in Proposition 9. By the investor's utility function (2.2), his equilibrium strategy (2.25) and (2.22), one has

$$\pi_I = \mathbb{E}\left[(1-(1-q)^M)(\theta+\delta)^2\frac{1}{2}\alpha\right]. \tag{2.27}$$

Equation (2.27) takes into account that the investor does not receive any quote with probability $(1-q)^M$ and that dealers infer ξ from the investor's demand.

Our first goal is to study the role of M, the number of recipients of each RFQ. Increasing M has three major effects that determine the investor's profit:

- As is evident from (2.27) a higher M increases the probability of a trade $1-(1-q)^M$, whenever $q<1$. Holding everything else equal, this increases expected profits.
- A higher M increases the fraction of informed dealers in the interdealer market. One can verify that a as defined in (3.6) is strictly increasing in M for $M<N$. This makes prices in the interdealer market more informative and it therefore becomes more difficult to offset any inventory that was acquired on the platform.
- A higher M decreases κ, as mentioned in the discussion after Lemma 11. Therefore, the bargaining power of the investor increases, which has a positive effect on his profit.

Considering these three bullet points, the investor's profit should be maximal for $M=2$, if $q=1$. If $q=1$, one has $\kappa=0$ and $1-(1-q)^M=1$, i.e. the investor's bargaining power is maximal and a trade happens with probability 1. Then the first and third bullet point above become irrelevant and increasing M is only associated with the cost of information leakage, discussed in the second bullet point. The following proposition formally confirms that the conclusion of the above heuristic reasoning is indeed true. Since we focus on the cost of information leakage we assume for better algebraic tractability that there are no private benefits, i.e. $\sigma_\delta=0$.

Proposition 10. *Let $2\le M$ and $q=1$ and $\sigma_\delta=0$. The equilibrium described in Proposition 9 exists if and only if a is below a certain threshold $\overline{a}$, with $\overline{a}<\frac{1}{2}$. In this equilibrium, one has $\beta_1,\beta_2,\alpha>0$.*

Furthermore, the equilibrium exists for any other choice of the number M' of dealers to contact with $2 \leq M' < M$. If $M = 2$, the payoff for the investor is higher than in any other possible equilibrium with $M > 2$.

When $q < 1$, the investor has incentive to contact more dealers, i.e. $M \geq 2$. Because when $q \neq 1$, the first and third effects turn out to be relevant: increasing M will improve the probability of trading , as well as the bargaining power of the investor. But at the same time, the cost of information leakage is also increased (second bullet point).

The following proposition states that M sometimes has to be larger than a certain threshold in order for an equilibrium to exist in the first place. If q is relatively small, the bargaining power κ of dealers may be so high that investors do not want to incur any price impact they have on the trading platform. Increasing M lowers this bargaining power. Under the condition that prices in the interdealer market remain sufficiently uninformative, an equilibrium exists for a sufficiently large M. On the other hand, there is a clear upper bound on the possible number of dealers that are contacted on the platform for which an equilibrium exists. In particular, if more than half of the dealers are contacted and there is strong asymmetric information about the asset's payoff ($\sigma_\delta = 0$), an equilibrium cannot exist, because information leakage on the platform is too strong.

Proposition 11. *Let $\sigma_\delta^2 = 0$. If $M > \frac{1}{2}N$, there is no equilibrium on the trading platform as described in Proposition 9.*

If $q < \frac{1+\sqrt{1-\frac{2\left(a(2)^2-5a(2)+2\right)}{a(2)^2-5a(2)+4}}}{2}$, there is no such equilibrium with $M = 2$. If furthermore $a < \bar{a}$, for an $\bar{a} \in (0, \frac{1}{2})$, then there is such an equilibrium with $M \geq 3$.

2.4.5 Price impact

In this section we want to relate our theoretical results to the empirical findings of Collin-Dufresne et al. (2017). In particular, we want to study the price impact that an investor faces on the trading platform and the price impact that dealers face in the interdealer market. The total price impact a trader faces can be decomposed as

$$\text{price impact} = \text{permanent impact} + \text{transitory impact}.$$

Collin-Dufresne et al. (2017) find that price impact in the D2C segment is higher than in the D2D segment. This difference is largely due to a difference in the permanent price impact.

We now want to derive the price impact and find analogues in our model that correspond to a permanent component and a transitory component. As commonly argued in theoretical studies (Sannikov and Skrzypacz, 2016; Kyle et al., 2017), the study of price impact is an off-equilibrium analysis. We will therefore assume an equilibrium as described in Proposition 9 and examine how the price a trader faces changes if the demanded quantity changes.

Equation (2.19) in Lemma 11 directly provides an expression of the expected price an investor receives on the platform. If the investor changes his demanded quantity x, then p_v and p_c in (2.19) and consequently the expected price for this quantity will change. Since the model presented in this paper is static, we have to find a decomposition of this price impact that would correspond to a decomposition into a permanent and a transitory component in a dynamic model. In empirical studies in Market Microstructure, it is generally assumed that the transitory component reflects a markup of the dealers, whereas the permanent component reflects the cost of the dealers of providing the asset due to future price changes. In our following analysis, we adopt this interpretation. We say that the price impact is permanent, if it was caused by a change in the dealers' cost of providing the asset.[5] Therefore, we define

$$PI := \frac{\partial}{\partial x} p_c(x) \tag{2.28}$$

as the permanent price impact of the investor, because (2.28) reflects the change in the price that is due to an increase in the dealers' cost of trading the asset.

In the following, we will consider p_c as defined in (2.15). Due to adverse selection, we also need to take into account that the dealers form their expectation ξ about the dividend payment based on (2.4) and (2.21). The following proposition contains some statements about the price impact on the platform and in the interdealer market.

Proposition 12. *The (permanent) price impact an informed dealer faces in the interdealer market is given by $-b$, where b is defined as in Proposition 8. Without adverse selection (dealers do not update their belief ξ), one has*

$$PI < -b,$$

i.e. the permanent impact on the trading platform is smaller then the permanent impact the dealers face in

[5]This assumes that changes in the dealers' cost have no transitory component. Transitory changes in the dealers' cost may arise due to inventory holding costs or order processing costs. In our model, dealers can immediately offset their inventory and the interdealer market is competitive. Order processing is costless. Therefore, such transitory components of dealers' costs are not present in our model.

the interdealer market. In the presence of adverse selection and $\rho = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2} > 1/4$*, one has*

$$PI > -b.$$

The dealers' permanent price impact $-b$ derived in Proposition 12 is due to a change in the uninformed dealers' belief about the dividend payment and a permanent change in the aggregate inventory held by other dealers in the interdealer market. Proposition 12 shows that the permanent price impact on the trading platform higher than the permanent impact in the interdealer market if and only if investors know more about the asset than dealers do. If there is no adverse selection and dealers do not update their belief about the asset's payoff, the dealers' cost of providing the asset changes by a lower rate than the price in the interdealer market would when trading the same quantity. This result is due to the dealers' optimal portfolio choice in period 2. A dealer could always offset the investor's demand in the interdealer market with price impact $-b$. If the investor however changed his demanded quantity, the dealer would, due to risk sharing considerations, in general not offset the total amount of this quantity in the interdealer market. Due to optimality of the dealer's portfolio choice, the dealer must be able to provide the quantity at a lower price than the one he would pay for this quantity in the interdealer market.

In the presence of adverse selection, however, the permanent price impact on the trading platform is higher than the permanent price impact the dealers face in the interdealer market. This makes our model (which assumes information asymmetries) consistent with the findings of Collin-Dufresne et al. (2017) that the permanent price impact is higher on the trading platform than in the interdealer market.

2.5 Centralized trading vs. electronic trading via RFQs

This section develops the market-design implications of our model. Our final goal is to characterize situations in which investors are better off trading in a centralized market and when an OTC market can improve their utility. In order to do this, we extend our previous model from to the case in which there is a continuum of investors of measure μ who all know the realization of θ in the beginning of period 1.[6] The investors' utility function is still give by (2.2). The risk-aversion parameter γ_I is the same for all investors. The investors receive a private benefit δ_i, where as before $\delta_i \sim \mathcal{N}(0, \sigma_\delta^2)$. The private benefits for different investors are essentially pairwise independent for different investors. This assumptions lets us apply the exact law of large

[6]Formally, let $(\Omega, \mathcal{F})$ denote the measurable space of investors. Then there is a bijective measurable map $\Phi : \Omega \to [0, \mu]$ and the measure of any set of investors $F \in \mathcal{F}$ is equal to the Lebesgue measure of the set $\Phi(F)$.

numbers of Sun (2006b). The model assumptions about the dealers are as in Section 2.3, except that we do not assume the presence of an outside agent in this section. Before we establish an equilibrium in the OTC market, we quickly describe how the investors would trade in a centralized market.

2.5.1 The centralized-market benchmark

Investors trade through double auctions in the centralized market. In these double auctions, each investor specifies a demand schedule, i.e. conditional on each price $p \in \mathbb{R}$ the investor specifies a quantity he wants to trade. The equilibrium price in the centralized market will be the market-clearing price. The market clearing price will be the unique price for which the investors' aggregate demand is equal to the aggregate supply of the asset (zero). The specification of the investor's utility function (2.2) gives the following maximization problem for each investor for each $p \in \mathbb{R}$:

$$\max_{x_i \in \mathbb{R}} \left[x_i(\theta + \delta_i - p) - \frac{\gamma_I \sigma_\epsilon^2}{2} x_i^2 \right],$$

where x_i denotes the quantity the investor demands given the price p on the exchange.

The sufficient first-order condition for the above optimization problem gives

$$x_i = \frac{\theta + \delta_i - p}{\gamma_I \sigma_\epsilon^2},$$

To determine the market-clearing price, we substitute each investor's demand schedule x_i into the market clearing condition, $\int x_i \mathrm{di} = 0$.[7] We get

$$0 = \mu \frac{\theta}{\gamma_I \sigma_\epsilon^2} - \mu \frac{p}{\gamma_I \sigma_\epsilon^2} \Leftrightarrow p = \theta,$$

where we have used the fact that $\int \delta_i \mathrm{di} = 0$ almost surely by the exact law of large numbers.

Using each investor's optimal demand schedule, the utility function (2.2) and the fact that the market clearing price is given by θ, we can define each investor's ex-ante payoff:

$$\pi_i^c := \mathbb{E}\left(\frac{1}{2} \frac{(\theta + \delta_i - p)^2}{\gamma_I \sigma_\epsilon^2} \right) = \frac{1}{2} \frac{\sigma_\delta^2}{\gamma_I \sigma_\epsilon^2}. \tag{2.29}$$

Equation (2.29) states that the centralized market realizes all the gains from trade that arise due to dispersed private values. When all investors have the same valuation of the asset ($\sigma_\delta^2 = 0$), no trade happens

[7] The notation di means that we integrate with respect to the measure on set of investors defined in Footnote 6.

and the investor's profits become zero. Each investor's profit decreases if the cost of bearing risk increases.

2.5.2 Electronic trading with a continuum of investors

The model with a continuum of investors is very similar to the model with one investor. It will turn out that a continuum of investors allows us to derive an equilibrium without the assumption of an outside agent. We will let the mass of investors have a measure $\mu \in (0, \infty)$. In period 1, all investors submit RFQs to M dealers. Afterwards, dealers trade in the interdealer market. All investors contact the same M dealers at the same time. The dealers then independently respond with a probability q to each RFQ. As before, we will determine the equilibrium in this model by backward induction.

Since there is no outside agent anymore in this section, uninformed dealers in the interdealer market take into account that the aggregate supply of the asset is correlated with the investors' information about the dividend level θ. We will conjecture that each investor demands a quantity x_i on the trading platform, where

$$x_i = \alpha_1 \theta + \alpha_2 \delta_i, \tag{2.30}$$

for some $\alpha_1, \alpha_2 \in \mathbb{R}$. As in Section 2.4, it will turn out that an investor always trades the asset if he receives a quote on the trading platform. Since each dealer responds independently with probability q to each RFQ, the an investor is able to trade the asset with probability $\mathbb{P}(trade) = 1 - (1-q)^M$. By the exact law of large numbers and (2.30), the investors' aggregate demand traded on the platform given by

$$X^{agg} := \int \mathbb{P}(trade)(\alpha_1 \theta + \alpha_2 \delta_i) \mathrm{d}i = (1 - (1-q)^M) \int (\alpha_1 \theta + \alpha_2 \delta_i) \mathrm{d}i = (1 - (1-q)^M) \mu \alpha_1 \theta, \tag{2.31}$$

where the last equality holds almost surely. By symmetry, each dealer gets an equal fraction of this aggregate demand. We define $X_k := -\frac{X^{agg}}{M}$ as the inventory of each dealer $k \leq M$ who gets contacted on the trading platform. From the dealers' utility function (2.1), one obtains the optimal demand schedule q_k for each dealer $k \leq M$:

$$q_k = \frac{\theta - p_2}{\gamma_d \sigma_\varepsilon^2} + X_K. \tag{2.32}$$

Notice that X_k is a multiple of θ, this will simplify the inference problem that the uninformed dealers face in the interdealer market. Analogously to Section 2.4, we conjecture that the market-clearing price in the interdealer market is given by

$$p_2 = a\theta + bW, \tag{2.33}$$

where W is the noise in the aggregate supply of the asset. The uninformed dealers use the normal projection theorem obtain the distribution of the dividend payment conditional on the market-clearing price p_2. The dealers' utility function (2.1) now gives the optimal demand

$$q_k = \frac{\mathbb{E}(D|p_2) - p_2}{\mathbb{V}(D|p_2)} \tag{2.34}$$

for the uniformed dealers who do not get contacted on the trading platform. Analogously to Proposition 8, we now state the equilibrium in the interdealer market, conditional on the investors' trading strategy (2.30).

Proposition 13. *For any given α_1, there is a rational expectations equilibrium such that the market clearing price is given by (2.33). Define*

$$\varphi : \;=\; (1-(1-q)^M)\mu, \tag{2.35}$$

$$\tau_u : \;=\; \frac{1}{Var(D|p_2)} = \frac{1}{\sigma_\theta^2 + \sigma_\epsilon^2 - \psi\sigma_\theta^2}, \tag{2.36}$$

$$\psi : \;=\; \frac{a^2\sigma_\theta^2}{a^2\sigma_\theta^2 + b^2\sigma_W^2} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \left(\frac{\gamma_d}{M\tau_\epsilon + \gamma_d\varphi\alpha_1}\right)^2 \sigma_W^2} \tag{2.37}$$

Then a and b are given by

$$a = \frac{M\tau_\epsilon + \gamma_d\alpha_1\varphi + (N-M)\psi\tau_u}{M\tau_\epsilon + (N-M)\tau_u} \tag{2.38}$$

and

$$b = -\frac{\gamma_d}{M\tau_\epsilon + \gamma_d\varphi\alpha_1} a, \tag{2.39}$$

One has $a > 0$. One also has $a \leq 1$ with a strict inequality if $M < N$.

Lemma 11 gives the dealers optimal quoting strategy for any aggregate quantity X_k that dealer k trades with the investors and any demand x_i they face from an individual investor i. There is only a slight difference between the case in Section 2.4 and the setup considered here. Whereas the aggregate demand a dealer faced was equal to the demand by the single investor in Section 2.4, the quantities x_i and X_k are different here. Using Lemma 11 and taking account of this difference gives the expected price $P(x_i)$, investor i gets for his demand x_i conditional on at least one response to the RFQ:

$$P(x_i) = p_c(X_k) + (p_v(x_i) - p_c(X_k)\frac{Mq(1-q)^{M-1}}{1-(1-q)^M}. \tag{2.40}$$

We already determined in (2.30) which form each investor's demand x_i takes. We also know the quantity X_k given these individual demand schedules. We now determine the values of $p_v(x_i)$ and $p_c(X_k)$, so that we can use (2.40) to determine the expected price that each investor faces for his demand.

We define this dealer's value function that maps his inventory after period 1 X_k to expected utility as

$$V_{k,1}(\theta, X_k) := \mathbb{E}_k\left[D(q_k - X_k) - p_2 q_k - \frac{\gamma_d}{2}\sigma_\varepsilon^2(q_k - X_k)^2\right], \tag{2.41}$$

where we used the dealer's utility function (2.1).

Having obtained the dealer's utility $V_{k,1}(\theta, X_k)$ when holding X_k units of the asset, we define the dealer's break even price $p_c(X)$ such the payment compensates for the marginal cost of holding an additional marginal unit of the asset. The resulting expression is stated in the following Lemma.

Lemma 12. *Conditional on the equilibrium inventory of dealer $k \leq M$, the dealer's equilibrium break-even price for for the asset is given by*

$$p_c(X_k) := -\frac{\partial}{\partial X_k}V_{k,1}(\xi, X) = a\theta - b\frac{(1-a)\theta}{\gamma_d\sigma_\varepsilon^2} + b\frac{\varphi}{M}\alpha_1\theta, \tag{2.42}$$

where a, b, φ are defined as in Proposition 13.

Given that a dealer inferred the realization of θ from the investors' demand, a dealer can infer the private value δ_i of investor i from this investor's individual demand and (2.30). Given the investor's demand for x_i units of the asset, a dealer can infer the maximum price the investor is willing to pay for these x_i units by using (2.2):

$$p_v(x_i) := \theta + \delta_i - \frac{\gamma_I}{2}x\sigma_\varepsilon^2. \tag{2.43}$$

Using (2.31), (2.42) and (2.42), one can rewrite (2.40) as

$$P(x_i) = \beta_1\theta + \beta_2 x_i, \tag{2.44}$$

for some $\beta_1, \beta_2 \in \mathbb{R}$ stated in the appendix. We now determine the optimal amount x_i that an investor wants to demand given that the expected price he faces on the platform is given by (2.44). The maximization problem of investor i is given by

$$\max_{x\in\mathbb{R}} \left[(\theta + \delta_i)x_i - x^2\frac{\gamma_I}{2}\sigma_\varepsilon^2 - x_i\left(\beta_1\theta + \beta_2 x_i\right)\right]. \tag{2.45}$$

The expression in (2.45) considers the investor's expected payoff conditional on at least one response to the RFQ, since the investor's payoff is maximized when his payoff conditional on at least one response is maximized. The first-order condition to the problem in (2.45) gives the investor's optimal demand schedule

$$x = \theta\frac{1-\beta_1}{2\beta_2 + \gamma_I\sigma_\varepsilon^2} + \delta\frac{1}{2\beta_2 + \gamma_I\sigma_\varepsilon^2}. \tag{2.46}$$

One can immediately determine α_1 and α_2 from (2.30) by looking at (2.46):

$$\alpha_1 = \frac{1-\beta_1}{2\beta_2 + \gamma_I\sigma_\varepsilon^2}, \tag{2.47}$$

$$\alpha_2 = \frac{1}{2\beta_2 + \gamma_I\sigma_\varepsilon^2}. \tag{2.48}$$

We are now ready to establish the existence of an equilibrium.

Proposition 14. *The expected price on the platform $P(x_i)$ that an investor gets on the platform for his demand x_i is given by (2.44) for some $\beta_1, \beta_2 \in \mathbb{R}$. Let $M \geq 2$. There is an equilibrium on the platform described below if and only if $M < N$ and*

$$\kappa = \frac{Mq(1-q)^{M-1}}{1-(1-q)^M} < \frac{1}{2}.$$

The equilibrium is characterized as follows. The investor submits a demand x_i as determined in equations (2.30) with $\alpha_1, \alpha_2 \in \mathbb{R}$, with $0 < \alpha_1 < 1$ and $\alpha_1 \leq \alpha_2$. The dealers quote independently with probability q according to the distribution function F in (2.18) with $p_c(X_k)$ and $p_v(x_i)$ given by (2.42) and (2.43).

2.5.3 Market design

In this section we will use the results derived in Section 2.5.1 and Section 2.5.2 and study when investors prefer the centralized market and when they prefer the OTC market with an electronic trading platform. Proposition 14 states that there cannot be an equilibrium on the electronic trading platform if $\kappa \geq \frac{1}{2}$ or $N = M$. In this case, there is only an equilibrium in the centralized market. Therefore, we restrict our further discussion to the case in which $\kappa < \frac{1}{2}$ and $M < N$. The following claim follows from (2.29) and Proposition 14.

Proposition 15. *Let $0 < \kappa < \frac{1}{2}$ and $2 \leq M < N$. As $\sigma_\delta^2 \to 0$, investors prefer to trade in on the trading platform in the OTC market. As $\sigma_\delta^2 \to \infty$, investors prefer to trade in the centralized market.*

If $\sigma_\delta^2 \to 0$, equation (2.29) implies that investors' gains from trading in the centralized market go to zero. However, due to information asymmetries between dealers and investors, investors can still benefit from trading in the OTC market.

Suppose on the other hand, that $\sigma_\delta^2 > 0$ and the mass of investors μ becomes very large. Then holding everything else constant, the investors' demand will be very sensitive to variations in θ. In this case, an equilibrium is only possible if α_1, the coefficient in the investors' demand on θ is very small and investors will mainly trade based on their private value of holding the asset. If markups in the interdealer market are positive, investors will therefore prefer to trade in the centralized market instead. The following proposition proves this statement formally.

Proposition 16. *Let $0 < \kappa < \frac{1}{2}$, $2 \leq M < N$ and $\sigma_\delta^2 > 0$. As $\mu \to \infty$, investors prefer to trade in the centralized market.*

The proof of Proposition 16 shows that $\alpha_1 \to 0$ as $\mu \to 0$. According to (2.47), this is equivalent to $\beta_1 \to 1$, holding everything else equal and noting that by (2.75), β_2 is unaffected by μ. Thus, (2.44) implies that the expected price an investor receives on the platform when $\mu \to \infty$ is approximately the sum of the common value θ of the dividend payment and a markup. In this case, the investors' gains from trade are derived mostly from their private values.

So far, we assumed that $\kappa > 0$, which lead to positive expected markups for the dealers when quoting on the trading platform. In the following we consider the case in which $q \to 1$, which leads to $\kappa \to 0$. If $\kappa \to 0$, these markups become negligible and dealers efficiently intermediate trades between their customers as if these customers were trading in a centralized market. Furthermore, the probability of not receiving a quote goes to zero as $q \to 1$. Thus, all the gains from trade that could be realized in the centralized market would also be realized in the OTC market. However, investors can still benefit from information asymmetries between them and the dealers in the OTC market. As $q \to 1$, investors therefore prefer to trade in the OTC market. This claim is formally proved in the next proposition.

Proposition 17. *Let $2 \leq M < N$. As $q \to 1$, investors prefer to trade on the trading platform.*

2.6 Discussion and concluding remarks

Electronic trading platforms play a central role in today's OTC markets. The implications of our model are consistent with recent empirical research that studies OTC markets with electronic trading platforms. One important feature of our model is information leakage which is studied in Hendershott and Madhavan (2015b) and Hagströmer and Menkveld (2016). We also showed that information asymmetries between dealers and investors are a sufficient and necessary condition to generate the price impact patterns observed in Collin-Dufresne et al. (2017). Therefore, the first part of this paper can be viewed as a theoretical foundation of several empirical findings in recent research. The model can also be used to evaluate the impact of recent financial regulation on investors' trading profits. The Dodd-Frank Act mandates that the most liquid index CDS in the US are trades on electronic platforms. An RFQ furthermore should be sent to at least three dealers.[8] We show that increasing the number of contacted dealers may decrease investor's profits if the cost of information-leakage is high. On the other hand, the number of contacted dealers has to be sufficiently high in order for an equilibrium to exist, if competition among dealers on the platform (in terms of response rates) is low.

In the second part of the paper, we considered a hypothetical scenario in which there is either a centralized exchange or an OTC market and studied the respective implications on investor welfare. Some of our results are consistent with the recent theoretical literature in the area of market design. That investor welfare is generally higher on exchanges if the investors associate strong private values with holding the asset, can be viewed as an analogue to the result of Babus and Parlatore (2017) that there is only a centralized-market equilibrium if the investors' values of holding the asset are sufficiently independent. We also emphasize the role of information asymmetries that becomes important in OTC markets. In this respect our paper is related to Glode and Opp (2017a). However, the specific trading protocol on electronic trading platforms features some aspects that are not present in other models of OTC markets. As the RFQ response rate q of dealers becomes high, our model shows that electronic trading platforms indeed become similar to exchanges,

[8] See Collin-Dufresne et al. (2017) for an overview of the regulatory changes in the US CDS market.

in the sense that dealers efficiently intermediate the demand from their customers. This result justifies the common opinion that electronic trading platforms represent a natural compromise between exchanges and OTC markets.[9]

To conclude, we want to make some general remarks on our model assumptions. As every theoretical model, also the one presented in this paper is build on some simplifying assumptions trading-off analytical tractability against appropriate representation of the real world. The fact that all investors are equally informed about the asset's payoff is certainly not completely realistic, but should capture the general information asymmetry between investors and dealers that in many markets seem to exist. To justify the way we model trading in the interdealer market, we want to refer to the event that made both the academic world and international regulatory authorities focus so much on OTC markets in the first place: the recent financial crisis. Arguably, demand for certain credit derivatives originated from informed hedge funds who wanted to bet against a credit bubble in the US credit market. Some investment banks may have learned about the value of certain securities from this informed demand and may have tried to use this knowledge against other less informed investment banks or other clients (which may be represented by noise traders in our model).

This example also suggests to interpret the welfare results derived from our model with a slight grain of salt. In this paper, we exclusively focused on investor welfare. While this approach may be viewed as standard in market design, it does not take into account financial stability considerations that may be important when determining the optimal level of transparency in the market. If losses to dealers or noise traders are large, the financial system may very well be affected in ways that cannot be captured in the model presented here. While the trade-off between the efficient allocation of assets and financial stability is a common theme in banking, examining the trade-off between investor welfare and financial stability in OTC markets may be a theme for future research.

2.7 Appendix A

This appendix contains all proofs that have been omitted in the main text.

Proof of Proposition 8. By the conjecture (2.8), the market clearing price p_2 is jointly normally distributed with θ. By the definition of ξ in (2.4), one has

[9]See Stafford (2016).

$$\mathrm{Cov}(D, p_2) = \mathrm{Cov}\left(\theta + \varepsilon, a\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}(\theta + \delta)\right) = a\rho\sigma_\theta^2. \tag{2.49}$$

Furthermore, one has

$$\mathbb{V}(\xi) = \mathbb{V}\left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}(\theta + \delta)\right) = \rho\sigma_\theta^2. \tag{2.50}$$

Now (2.49), (2.50) and the normal projection theorem give

$$\mathbb{E}(D|p_2) = \frac{a\rho\sigma_\theta^2}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2}p_2 = \frac{\psi}{a}(a\xi + bW), \tag{2.51}$$

$$\mathbb{V}(D|p_2) = \frac{1}{\tau_u} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \frac{a^2\rho^2\sigma_\theta^4}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \psi\rho\sigma_\theta^2. \tag{2.52}$$

Plugging (2.51) and (2.52) into (2.7), using the result with (2.6) in the market-clearing condition (2.9):

$$\frac{M\tau_\xi(\xi - p_2)}{\gamma_d} + \frac{(N-M)\tau_u(\psi\xi + \frac{\psi b}{a}W - p_2)}{\gamma_d} = W$$

solving for p_2 and matching coefficients with (2.8) yields

$$M\tau_\xi + (N-M)\psi\tau_u = [M\tau_\xi + (N-M)\tau_u]a \tag{2.53}$$
$$(N-M)\tau_u\frac{\psi b}{a} - \gamma_d = [M\tau_\xi + (N-M)\tau_u]b. \tag{2.54}$$

Substituting $\psi = \frac{a^2\rho\sigma_\theta^2}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2}$ into equations (2.53) and (2.54) and solve for a and b gives the expressions in (3.6) and (3.7).

It is immediately clear from (3.6) that $a > 0$ if $M > 0$, since both numerator and denominator are always positive in this case. Since $\psi > 0$ it follows also that $a \leq 1$, with an equality only if $N = M$.

□

Proof of Lemma 10. The dealer's optimal demand schedule follows directly from the first-order condition (3.5) by substituting $\overline{\omega}_k = q_k - x$. The demand schedules of other informed dealers do not change, since they do not make inferences from the price in the interdealer market. The dealers who have not been contacted by the investor perform inferences as described in Proposition 8. One can now conjecture $p_2 = a\xi + b(W - x)$. Thus using dealer k's demand schedule and demand schedules (2.6) and (2.7) for the other dealers in the market clearing condition and following the exact procedure described in the proof of

Proposition 8 determines a and b as in 8.

□

Proof of Lemma 11. Let F_x denote the dealers' optimal quoting strategy. This means dealers quote a price p_0 that is a random variable with the distribution function F_x.

Let $x > 0$. Then $x(p_v - p_c) > 0$ implies $p_v > p_c$. If the dealers' optimal strategy were such that there is a $p^* \in (p_v, p_c)$ such that dealers quote a price $p \leq p^*$ with a probability of 1, then a dealer could profitably deviate from this strategy by quoting p_v. This would contradict optimality. On the other hand, quoting a prices greater than p_v with any positive probability cannot be optimal, since the investor would not buy the asset at that price. Thus, one obtains $\sup \text{support}(F_x) = p_v$.

Now we show that F_x must be continuous, i.e. there cannot be any atoms in the distribution of p_0. Clearly, quoting a price less than or equal to p_c with any positive probability cannot be optimal, since a dealer would not make any positive profit by doing so, whereas he would make a positive expected profit by quoting p_v. Now, suppose there is a price p' with $p_v \geq p' > p_c$ that is quoted with probability $\rho > 0$ by all dealers.

Then a single dealer could again profitably deviate from this strategy which contradicts optimality. The profitable deviation is constructed as follows. Since the number of prices charged with positive probability must be countable, one can find for each $\delta > 0$ an ε_δ, such that $\delta \geq \varepsilon_\delta > 0$ and the price $p' - \varepsilon_\delta$ is charged with probability zero by all dealers. The deviating dealer can now charge price $p' - \varepsilon_\delta$ with probability ρ and charge price p' with probability zero. Using the fact that $\lim_{\delta \to 0} F_x(p' - \varepsilon_\delta) = F_x(p') - \rho$, one can express the difference Δ in profits between the original strategy and the proposed deviation as follows. A dealer quoting p' only makes a positive profit if no other dealer on the platform quotes a lower price. If no other dealer quotes a lower price, there might be $j = 0, 1, ..., M - 1$ dealers who quote p' as well. In the latter case, each of the $j + 1$ is equally likely to be chosen by the investor for trading the asset. The calculation below considers the cases in which j dealers quote price p on the platform separately.

$$\begin{aligned}\Delta &= (1-qF_x(p'-\varepsilon_\delta)-q\rho)^{M-1}(p'-\varepsilon_\delta-p_c)x \\ &-(1-qF_x(p'))^{M-1}(p-c)x \\ &+\sum_{j=1}^{M-1}\binom{M-1}{j}(1-qF_x(p'-\varepsilon_\delta)-q\rho)^{M-1-j}(q\rho)^j(p'-\varepsilon_\delta-p_c)x \\ &-\sum_{j=1}^{M-1}\binom{M-1}{j}(1-qF_x(p'))^{M-1-j}(q\rho)^j(p'-p_c)\frac{x}{j+1}.\end{aligned}$$

The first two lines in the above expression compare expected profits from quoting $p'-\varepsilon_\delta$ and expected profits from quoting p' in the event that all other dealers quote a price above p'. Since $\lim_{\delta\to 0} F_x(p'-\varepsilon_\delta) = F_x(p')-\rho$, the difference in these two lines goes to zero as δ goes to zero. The last two lines compare the respective profits in the cases in which $j>0$ other dealers quote p'. Since $M\geq 2$, the deviating dealer can get a jump in expected trading volume in this case, since he can avoid ties with other dealers. Therefore one obtains

$$\Delta \to \sum_{j=1}^{M-1}\binom{M-1}{j}(p'-p_c)\frac{jx}{j+1}(1-qF_x(p'))^{M-1-j}(q\rho)^j > 0 \quad \text{as } \delta\to 0.$$

Thus, the proposed deviation is profitable for a small δ. In equilibrium, F_x cannot have any atoms.

If $x<0$, one verifies analogously to the case of $x>0$, that $\inf \text{support}(F_x) = p_v$ must hold for any optimal strategy. That the distribution cannot have any atoms follows analogously as well.

The dealers are only willing to randomize over prices if they earn the same profit in expectation with each price in the support of F_x. This profit must be equal to the profit in which the dealer quotes $p_v(x)$. This gives the indifference condition expressed in (2.17).

Using the binomial formula $(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$, (2.17) simplifies to

$$(p-p_c)x(1-qF_x(p))^{M-1} = (1-q)^{M-1}(p_v(x)-p_c)x,$$

which can be solved for F_x. The solution is given by (2.18).

Using (2.18) and solving $F_x(\overline{p}_x)=0$ for $\overline{p}_x$ gives

$$\overline{p}_x = p_c + (p_v-p_c)(1-q)^{M-1}.$$

Since $x(p_v - p_c) > 0$, one obtains $\overline{p}_x > p_c$ for $x > 0$ and $\overline{p}_x < p_c$ for $x < 0$.

The event that at least one dealer is on the platform happens with probability $1 - (1-q)^M$, since all dealers respond independently with probability q. The unconditional probability that no dealer quotes above $p \in \text{support}(F_x)$ can be expressed by $(1 - qF_s(p))^M$. Therefore, the conditional distribution G_x has to satisfy $G(p)(1-(1-q)^M) = 1 - (1 - qF_x(p))^M$. Performing a change of variables $p = p_c + \frac{(p_v - p_c)(1-q)^{M-1}}{(1-(1-(1-q)^M)u)^{(M-1)/M}}$, one can calculate

$$\int_{\text{support}(F_x)} p\mathrm{d}G_x(p) = \int_0^1 \left[p_c + \frac{(p_v - p_c)(1-q)^{M-1}}{(1-(1-(1-q)^M)u)^{(M-1)/M}} \right] \mathrm{d}u = p_c + \frac{(p_v - p_c)(1-q)^{M-1}}{1-(1-q)^M} Mq.$$

The claim that $0 \leq \kappa < 1$, can be shown as follows. That $0 \leq \kappa$ is immediately clear from the definition (3.29). The other inequality can be seen as follows.

- κ as a function of q is strictly decreasing in q for all $q \in (0, 1]$, since

  $$\frac{\partial \kappa}{\partial q} = \frac{-M(1-q)^{M-2}\left[(1-q)^M + Mq - 1\right]}{(1-(1-q)^M)^2} < 0$$

  for $q \in (0, 1]$.

- By L'Hospital's rule, one has

  $$\lim_{q \to 0} \kappa = \frac{\lim_{q\to 0}(M(1-q)^{M-1} - M(M-1)q(1-q)^{M-2})}{\lim_{q \to 0} M(1-q)^M} = \frac{M}{M} = 1.$$

The last two bullet points imply $\kappa < 1$ for all $q \in (0, 1]$.

This proves all statements in the lemma.

□

Proof of Proposition 9. Claim 1: The expected price on the platform is linear in ξ and x. Define

$$\beta_1 := \left[1 - \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\right]\left[a - \frac{(1-a)b}{\gamma_d \sigma_\xi^2}\right] + \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\left(1 + \frac{\sigma_\delta^2}{\sigma_\theta^2}\right) \tag{2.55}$$

and

$$\beta_2 := \left[1 - \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\right]\left[-b - \frac{b^2}{2\gamma_d \sigma_\xi^2}\right] + \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\left(-\frac{\gamma_I \sigma_\epsilon^2}{2}\right). \tag{2.56}$$

Using the definitions of $p_c(x)$ and $p_v(x)$, it follows by direct computation that $P(x)$ as defined in Lemma 11 is given by (2.22).

Claim 2: Let $\frac{M(1-q)^{M-1}}{1-(1-q)^M} < \frac{1}{2}$. An equilibrium exists if and only if $a < \overline{a}$ for some $\overline{a} \in \mathbb{R}$.

To verify the existence of the described equilibrium, there are several things to check. The strategy (2.25) is well-defined if

$$2\beta_2 + \gamma_I \sigma_\varepsilon^2 \neq 0. \tag{2.57}$$

Furthermore, the investor's second-order condition from the maximization problem (2.23) requires

$$-\gamma_I \sigma_\varepsilon^2 - \left(2 \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2} \frac{\beta_1}{\alpha} + 2\beta_2 \right) < 0. \tag{2.58}$$

In order to apply Lemma 11 we also need to verify that

$$x(p_v(x) - p_c(x)) > 0 \tag{2.59}$$

for holds for any $x \neq 0$ demanded by the investor in the proposed equilibrium.

If (2.57), (2.58) and (2.59) hold, one can use Lemma 11 to see that there exist optimal strategies for the dealers that yield $P(x)$ as the expected price on the platform conditional on at least one response. As demonstrated in the text, the stated strategy for the investor (2.25) indeed solves the first order condition (2.24), given that dealers rationally infer $\theta + \delta$ from the investor's demand. Thus, both dealers and the investor behave optimally given the strategies of the others and an equilibrium is established.

The strategy of the proof of this claim is as follows. We will assume that the average price is given by the expression in Lemma 11. We then show that the investor's strategy is well-defined so that the first-order and second-order conditions of the maximization problem (2.23) are satisfied. We that verify that in this case

In order to prove our claim, we first note that (2.59) is satisfied in this case, so that dealers indeed find it optimal to quote as described in Lemma 11.

For the following proof, it is worth noting that Lemma 11 states that

$$0 \leq \kappa < 1 \tag{2.60}$$

for all $q \in (0, 1]$ and $M \geq 2$.

„⇒": *Proof that equilibrium exists under the stated conditions.*

Let now $\kappa < \frac{1}{2}$.

We rewrite (2.55) and (2.56) as using $b = -\frac{\gamma_d \sigma_\xi^2}{M} a$:

$$\beta_1 = (1-\kappa)\left(a + \frac{(1-a)a}{M}\right) + \kappa \frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2} > 0, \tag{2.61}$$

$$\beta_2 = (1-\kappa)\left(\frac{\gamma_d \sigma_\xi^2}{M} a - \frac{a^2 \gamma_d \sigma_\xi^2}{2M^2}\right) - \kappa\left(\frac{\gamma_I \sigma_\epsilon^2}{2}\right) > -\frac{\gamma_I \sigma_\epsilon^2}{2}, \tag{2.62}$$

where the inequalities follow from (2.60) and $1 \leq a > 0$.

Define

$$\Psi := \frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2} \frac{\frac{1}{2} - \kappa}{1 - \kappa} > 0$$

and define $\overline{a}$ as the smaller solution to the quadratic equation

$$a + \frac{(1-a)a}{M} = \Psi,$$

if there is a real solution to the equation. Set $\overline{a} = 1$ otherwise. Then it follows from (2.61) that

$$\beta_1 < \frac{1}{2} \frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2},$$

if $a < \overline{a}$.

Thus, all that remains to show is that there is an equilibrium if the last inequality involving β_1 holds. As described above it is sufficient to check that (2.57), (2.58) and (2.59) hold. It is immediately clear from (2.62) that (2.57) always holds for any set of parameters.

Regarding (2.58), note that using (2.25), (2.62) and the assumption on β_1 imply $\alpha > 0$. Using (2.61), one therefore obtains

$$-\gamma_I \sigma_\varepsilon^2 - \left(2 \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2} \frac{\beta_1}{\alpha} + 2\beta_2\right) < 2 \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2} \frac{\beta_1}{\alpha} + 2\beta_2 < 0.$$

Thus, the investor's second-order condition holds if $\beta_1 < \frac{1}{2} \frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2}$.

Lastly, we check that (2.59) holds which justifies the use of Lemma 11 for determining the expected price on the platform. Note that by optimality of the investor's choice of x and $\alpha > 0$, it follows that the investor makes a positive profit if $x \neq 0$ This can be seen, since the investor could always make a zero profit by not trading, but instead chooses a different x. By the convexity of the maximization problem (2.23), the optimal quantity is uniquely determined and therefore must give a positive profit. This implies

$$x(p_v(x) - p_c(x)) \geq x(p_v(x) - p_c(x))(1-\kappa) = x(p_v(x) - P(x)) > 0.$$

Therefore (2.59) indeed holds and Lemma 11 can be used to determine the dealer's quoting strategies on the platform.

Since (2.57), (2.58) and (2.59) indeed hold, the equilibrium exists.

„$\Leftarrow$": *Proof that equilibrium does not exist if $a \geq \overline{a}$.*

The definition of $\overline{a}$ and β_1 imply that $\beta_1 \geq \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$ if $a \geq \overline{a}$. If the last inequality is an equality, it follows that $\alpha = 0$. This means, the investor does not trade and the quoting strategies of the dealers are not defined. Let the inequality be strict. Note that by (2.61), $\kappa \in [0,1]$ and $a \in [0,1]$, one has $\beta_1 \leq a + (1-a) = 1$. This in turn implies

$$\frac{1}{1 - 2\beta_1 \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}} > -1.$$

One now obtains

$$\begin{aligned}
& -\gamma_I \sigma_\varepsilon^2 - \left(2\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\frac{\beta_1}{\alpha} + 2\beta_2\right) \\
= \quad & -\gamma_I \sigma_\varepsilon^2 + 2\beta_2 - 2\frac{2\beta_2 + \gamma_I\sigma_\varepsilon^2}{1-2\beta_1}\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\beta_1 \\
> \quad & -\gamma_I \sigma_\varepsilon^2 + 2\beta_2 + 2(2\beta_2 + \gamma_I\sigma_\varepsilon^2)\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\beta_1 \\
> \quad & -\gamma_I \sigma_\varepsilon^2 + 2\beta_2 + (2\beta_2 + \gamma_I\sigma_\varepsilon^2) = 0.
\end{aligned}$$

Therefore, the second-order condition for the investor's maximization problem (2.23) is not satisfied. Thus, the investor's strategy is clearly not optimal and the described equilibrium does not exist.

Claim 3: The equilibrium does not exist if $\kappa \geq \frac{1}{2}$

In this case, $a \geq 0$ and (2.61) imply $\beta_1 \geq \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$. The prove that the equilibrium doe not exist is identical to the proof in Claim 2.

Claim 4: $a \to 0$ as $N \to \infty$ and $\sigma_W \to \infty$.

By equation (3.6), one can see that

$$\begin{aligned}\lim_{\sigma_W\to\infty}\lim_{N\to\infty} a &= \lim_{\sigma_W\to\infty}\lim_{N\to\infty} \frac{\frac{M\tau_\xi}{N-M}+\psi\tau_u}{\frac{M\tau_\xi}{N-M}+\tau_u}\\ &= \lim_{\sigma_W\to\infty}\psi\\ &= 0.\end{aligned}$$

Claim 5: An equilibrium exist if $\kappa < \frac{1}{2}$ and $\sigma_\delta \to \infty$. As $\sigma_\delta \to \infty$, one has $\Psi \to \infty$. This means $a + \frac{(1-a)a}{2M} < \Psi$ for all $a \in \mathbb{R}$ and in particular for all $a \in (0,1]$ As shown in the proof of Claim 2, this implies $\beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$ and the equilibrium exists.

□

Proof of Proposition 10. We divided this proof into several steps. The first step is an auxiliary result that will be used later in the proof.

Step 1: $\frac{\partial a}{\partial M} \geq \frac{1}{M}a(1-a)$.

Since $\sigma_\delta = 0$, we get $\sigma_\xi = \sigma_\varepsilon$ and $\rho = \sigma_\theta^2$. We now rewrite a as defined in (3.6) as

$$a = 1 - \frac{\gamma_d^2\sigma_\varepsilon^4\sigma_W^2(N-M)}{M^2N\sigma_\theta^2 + \gamma_d^2M\sigma_\varepsilon^2\sigma_\theta^2\sigma_W^2 + \gamma_d^2N\sigma_\varepsilon^4\sigma_W^2} \tag{2.63}$$

Using the expression in (2.63), one obtains by direct calculation and simplifying terms that

$$\begin{aligned}\frac{\partial a}{\partial M} - \frac{1}{M}a(1-a) &= \frac{\gamma_d^2M\sigma_\varepsilon^4\sigma_W^2\left(\gamma_d^2\sigma_\varepsilon^2\sigma_W^2\left(\sigma_\varepsilon^2+\sigma_\theta^2\right)+N^2\sigma_\theta^2\right)}{\left(M^2N\sigma_\theta^2 + \gamma_d^2M\sigma_\varepsilon^2\sigma_\theta^2\sigma_W^2 + \gamma_d^2N\sigma_\varepsilon^4\sigma_W^2\right)^2}\\ &\geq 0.\end{aligned}$$

This proves the first step

Step 2: The equilibrium exists if and only if a is below a certain threshold.

This result follows directly from Proposition 9 by noting that κ as defined (2.60) is equal to zero if $q = 1$. Furthermore, since $q = 1$, one has $\Psi = \frac{1}{2}$, where Ψ is defined in the proof of Proposition 8. Defining $\overline{a}$ as in the proof of Proposition 8, one gets that $\overline{a}$ is the smaller real solution to

$$a + \frac{(1-a)a}{M} = \frac{1}{2},$$

which is always greater than zero and less than $\frac{1}{2}$.

Step 3: There is an equilibrium for all $M' < M$.

In the proof of Proposition 9 it was established that the described equilibrium exists if and only if $\beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$. If an equilibrium exists when M dealers get contacted, it consequently must be the case that $\beta_1 < \frac{1}{2}$. If furthermore, $\beta_1 < \frac{1}{2}$ for all $M' < M$, the result follows. The last claim will be shown next. If $q = 1$, one has

$$\begin{aligned}
\frac{\partial \beta_1}{\partial M} &= \frac{(1-2a)\frac{\partial a}{\partial M}}{M} + \frac{\partial a}{\partial M} - \frac{(1-a)a}{M^2} \\
&= \frac{M+1-2a}{M}\frac{\partial a}{\partial M} - \frac{(1-a)a}{M^2} \\
&\geq \frac{M+1-2a}{M}\frac{1}{M}a(1-a) - \frac{(1-a)a}{M^2} \\
&> (3(1-a)+a^2)\frac{(1-a)a}{M^2} \\
&> 0.
\end{aligned}$$

The third line follows from Step 1. Therefore, one has $0 \leq \beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$ for all $M' < M$. Note that event though M represents an integer in the model, β_1 can be interpreted as a function in $C^1(\mathbb{R})$.

Step 4: The investor's payoff is highest if $M' = 2$ compared to all other $M'' \leq M$.

Since we know that a nonzero-trade equilibrium exists for all $M' < M$, we can calculate the investor's equilibrium payoff as defined by (2.27).

Using the expressions for β_1 and β_2 stated in Step 2 and using the definition of α from (2.25), one gets

$$\pi_I = \frac{\sigma_\theta^2 M\left(2a^2 - 2a(M+1) + M\right)}{2\sigma_\varepsilon^2\left(2a\gamma_d M + \gamma_I M^2 - a^2\gamma_d\right)}.$$

In equilibrium, one has $\pi_I > 0$. Since $a < \overline{a}$, the numerator in the above expression for π_I is positive. Therefore, the denominator must be positive as well. Interpreting π_I as a function in $C^1(\mathbb{R})$, one can show that π_I is strictly decreasing in M by showing that $\ln(\pi_I)$ is strictly decreasing in M. It then follows that the lowest possible M', i.e. $M' = 2$ is profit maximizing among all possible values less than M.

$$\begin{aligned}
\frac{\partial}{\partial M}\ln \pi_I &= \frac{M\left(4a\frac{\partial a}{\partial M} - 2(M+1)\frac{\partial a}{\partial M} - 2a + 1\right) + \left(2a^2 - 2(M+1)a + M\right)}{M\left(2a^2 - 2(M+1)a + M\right)} \\
&\quad - \frac{2\gamma_d M\frac{\partial a}{\partial M} - 2\gamma_d a\frac{\partial a}{\partial M} + 2\gamma_d a + 2\gamma_I M}{2\gamma_d M a - \gamma_d a^2 + \gamma_I M^2}
\end{aligned}$$

Collecting terms gives

$$\begin{aligned}\frac{\partial}{\partial M}\ln\pi_I &= \frac{1}{M}+\frac{1}{2a^2-2(M+1)a+M}-\frac{2a}{2a^2-2(M+1)a+M}\\ &-\frac{2\gamma_I M}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}-\frac{2\gamma_d a}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}+\\ &+\Big(\frac{4a}{2a^2-2(M+1)a+M}-\frac{2(M+1)}{2a^2-2(M+1)a+M}\\ &\frac{2\gamma_d a}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}-\frac{2\gamma_d M}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}\Big)\frac{\partial a}{\partial M}\end{aligned}$$

Since $a<\bar{a}$ one has $a^2-a(M+1)+M/2>0$. One can now see that the term in front of $\frac{\partial a}{\partial M}$ is negative. Therefore, one can obtain an upper bound for the $\frac{\partial a}{\partial M}\ln\pi_I$ by plugging in the result from Step 1 for $\frac{\partial a}{\partial M}$. Simplifying gives

$$\frac{\partial}{\partial M}\leq\frac{-2a\left(-2a^2(\gamma_d-\gamma_I M)+Ma(\gamma_d-\gamma_I(M+2))+\gamma_d a^3+\gamma_I M^2\right)}{\left(2a^2-2(M+1)a+M\right)\left(2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2\right)}.$$

The denominator is positive due to $a<\bar{a}$. Simplifying the numerator gives

$$-2\gamma_I a\left(2Ma^2-(M+2)Ma+M^2\right)-2\gamma_d a\left(a^3-2a^2+Ma\right)<0.$$

Therefore one has $\frac{\partial}{\partial M}\ln\pi_I<0$ and the claim follows.

Step 5: $M=2$ is profit-maximizing among all possible values.

Assume there would be an $M'>2$ such that $M=M'$ gives a higher profit than $M=2$ in equilibrium. By Step 3, it must be the case that $a<\bar{a}$ for $M=M'$. Now it follows by Step 3 that having $M=2$ gives a higher profit for the investor than having $M=M'$. Thus, contacting only 2 dealers is indeed profit-maximizing.

□

Proof of Proposition 11. It is shown in proposition 10 that equilibrium exists when $a<\bar{a}<\frac{1}{2}$. We now replace a as defined in equation (3.6): $\frac{M^2N\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2 M+\gamma_d^2\sigma_\epsilon^4\sigma_W^2 M}{M^2N\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2 M+\gamma_d^2\sigma_\epsilon^4\sigma_W^2 N}<\frac{1}{2}$. Equivalently,

$$M<\frac{1}{2}\frac{M^2N\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2 M+\gamma_d^2\sigma_\epsilon^4\sigma_W^2 N}{MN\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2+\gamma_d^2\sigma_\epsilon^4\sigma_W^2}<\frac{1}{2}N.$$

Therefore, if $M>\frac{1}{2}N$, one has $a\geq\bar{a}$, which implies that the equilibrium does not exist from proposition 10.

In the following, we show that the equilibrium existence condition $a<\bar{a}$ is equivalent to $M\in(\bar{M}_1,\bar{M}_2)$, where $\bar{M}_1$ and $\bar{M}_2$ are roots to the equation $a(M) = \bar{a}(M)$.

First, $a(M)$ is an increasing function of M and $\lim_{M\to N} a(M) = 1$. In terms of $\bar{a}(M, q)$, one can calculate the two derivatives

$$\frac{\partial \bar{a}}{\partial M} = 1 - \frac{2M + \frac{\kappa(1-\kappa) + \frac{M\partial\kappa}{\partial M}}{(1-\kappa)^2}}{\sqrt{4M^2 + 1 + 4M(1 - 2\Psi)}},$$

where $\frac{\partial \kappa}{\partial M} = \kappa \left[\frac{1}{M} + \frac{\ln(1-q)}{1-(1-q)^M}\right] < 0$, so $\frac{\partial \bar{a}}{\partial M} > 0$.

And

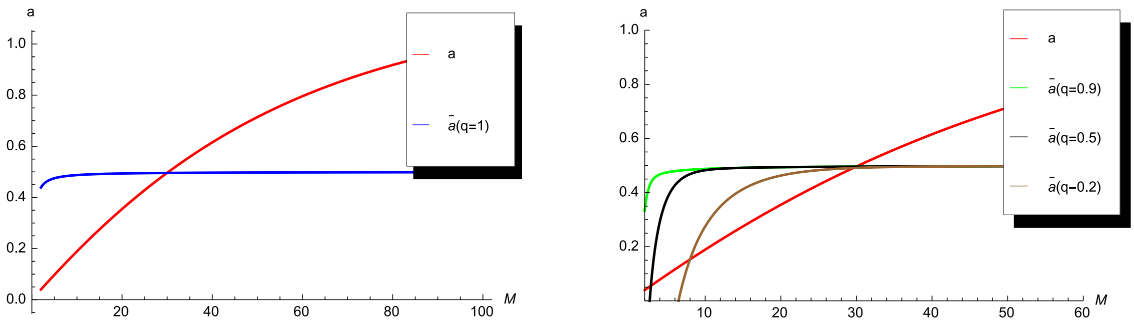
$$\frac{\partial \bar{a}}{\partial q} = -\frac{\frac{M}{(1-\kappa)^2}\frac{\partial \kappa}{\partial q}}{\sqrt{4M^2 + 1 + 4M[\frac{\kappa}{1-\kappa} - \frac{\sigma_\delta^2}{\sigma_\theta^2}]}},$$

where $\frac{\partial \kappa}{\partial q} = \frac{M(1-q)^{M-2}(1-Mq-(1-q)^M)}{(1-(1-q)^M)^2} < 0$, so $\frac{\partial \bar{a}}{\partial q} > 0$. Thus, $\bar{a}$ is an increasing function of both M and q.

Moreover, comparing the the value of $a(M)$ and $\bar{a}(M, q)$ at the limits, one gets

$$\begin{aligned} \lim_{M\to N} a(M) = 1 \quad &> \quad \frac{1}{2} > \lim_{M\to N} \bar{a}(M, q), \\ \lim_{M\to 2} a(M) \quad &< \quad \lim_{M\to 2, q\to 1} \bar{a}(M, q), \\ \lim_{M\to 2} a(M) > 0 \quad &> \quad \lim_{M\to 2, q\to 0} \bar{a}(M, q). \end{aligned}$$

Figure 2.5: Illustration



(a) (b)

So there are maximum two roots to the equation $a(M) = \bar{a}(M, q)$ for $M \in [2, N]$. As has been shown and demonstrated by figure (2.5) that there exists at least one root when $q = 1$, since $\bar{a}(M, q)$ decreases when q

decreases, the larger root $\bar{M}_2$ also decreases. Note that $\frac{\partial^2 \bar{a}(M,q)}{\partial q^2} < 0$, implies that the concavity of $\bar{a}(M,q)$ becomes larger , so the smaller root $\bar{M}_1$ increases when q decreases. More specifically,

(1) When $q = 1$, $\bar{M}_1 < 0$ and $\bar{M}_2 > 2$.

(2) When $q \in (\frac{1+\sqrt{1-\frac{2(a(2)^2-5a(2)+2)}{a(2)^2-5a(2)+4}}}{2}, 1]$, $\bar{M}_1 < 2$ and $\bar{M}_2 > 2$.

(3) When $q = \frac{1+\sqrt{1-\frac{2(a(2)^2-5a(2)+2)}{a(2)^2-5a(2)+4}}}{2}$, $\bar{M}_1 = 2$ and $\bar{M}_2 > 2$.

(4) When $q \in (\underline{q}, \frac{1+\sqrt{1-\frac{2(a(2)^2-5a(2)+2)}{a(2)^2-5a(2)+4}}}{2})$, $\bar{M}_1 > 2$ and $\bar{M}_2 > \bar{M}_1 > 2$.

(5) When $q = \underline{q}$, $\bar{M}_1 = N$, where $\underline{q}$ is the solution to the equation $\frac{Nq(1-q)^{N-1}}{1-(1-q)^N} = \frac{1}{2}$.

(6) When $q \in [0, \underline{q})$, there is no solution to $a(M) = \bar{a}(M)$ and $a(M) > \bar{a}(M)$.

The existence of equilibrium is summarized in figure (2.6).

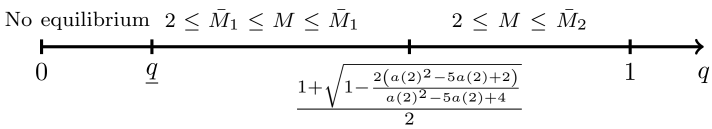


Figure 2.6: The existence of equilibrium

The above results show that when $q < \frac{1+\sqrt{1-\frac{2(a(2)^2-5a(2)+2)}{a(2)^2-5a(2)+4}}}{2}$, the equilibrium exists when $\bar{M}_1 < M < \bar{M}_2$. But $\bar{M}_1 \geq 2$, so there is no equilibrium when $M = 2$. Moreover, the minimum value of M such that the equilibrium exist is 3. Overall, we can conclude that when $q < \frac{1+\sqrt{1-\frac{2(a(2)^2-5a(2)+2)}{a(2)^2-5a(2)+4}}}{2}$, there exists an equilibrium with $M \geq 3$.

□

Proof of Proposition 12. Using the optimal demand schedule (2.6) for $M-1$ informed dealers and demand schedule (2.7) for the uninformed dealers, where the the conditional beliefs of uninformed dealers are formed as described in the proof of Proposition 8, the market clearing condition

$$q_k + \sum_{l \leq M, l \neq k} q_l + \sum_{lM+1}^{N} = W$$

can be rearranged as

$$p_2 = b\left(W - \frac{(M-1)\xi}{\gamma_d \sigma_\xi^2} - q_k\right).$$

therefore, one has $\frac{\partial}{\partial q_k} p_2 = -b$, whenever dealer k is informed.

If dealers do not update their belief about ξ when the investor changes his demanded quantity x, taking the derivative of (2.15) w.r.t. x gives the following permanent price impact

$$\frac{\partial}{\partial x} p_c = -\frac{b^2}{2\gamma_d \sigma_\xi^2} - b < -b.$$

Using (2.15) and taking into account that the dealers form their expectation ξ about the dividend payment based on (2.4) and (2.21), one obtains

$$\begin{aligned}
\frac{\partial}{\partial x} p_c &= \frac{a\rho}{\alpha} - b - \frac{2(1-a)b\frac{\rho}{\alpha} + b^2}{2\gamma_d \sigma_\xi^2} \\
&= \frac{\rho\tilde{\beta}_1}{\alpha} + \tilde{\beta}_2 \\
&\geq \frac{\rho\tilde{\beta}_1}{\tilde{\alpha}} + \tilde{\beta}_2,
\end{aligned}$$

where $\tilde{\alpha}$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ denote the value of α, β_1 and β_2 when $q = 1$, respectively. The inequality holds since $\alpha = \tilde{\alpha} - \frac{\kappa}{1-\kappa} \frac{1}{2\tilde{\beta}_2 + \gamma_I \sigma_\epsilon^2}$. Replacing $\tilde{\alpha}$ by (2.25), one has

$$\begin{aligned}
\frac{\partial}{\partial x} p_c &\geq \frac{\rho\tilde{\beta}_1}{1 - 2\rho\tilde{\beta}_1}(2\tilde{\beta}_2 + \gamma_I \sigma_\epsilon^2) + \tilde{\beta}_2 \\
&> \frac{\rho\tilde{\beta}_1}{1 - 2\rho\tilde{\beta}_1} 2\tilde{\beta}_2 + \tilde{\beta}_2 \\
&= \frac{\tilde{\beta}_2}{1 - 2\rho\tilde{\beta}_1} \\
&= \frac{-b - \frac{b^2}{2\gamma_d \sigma_\xi^2}}{1 - 2\rho\left[a - \frac{(1-a)b}{\gamma_d \sigma_\xi^2}\right]} \\
&= \frac{-b\left(1 - \frac{a}{2M}\right)}{1 - 2\rho a\left(1 - \frac{1-a}{M}\right)} \\
&> -b.
\end{aligned}$$

The last inequality holds since $1 - \frac{a}{2M} > 1 - 2\rho a\left(1 - \frac{1-a}{M}\right)$, that is equivalent to $a > 0 > 1 - M + \frac{1}{4\rho}$.

□

Proof of Proposition 13. By the conjecture (2.33), the market clearing price p_2 is jointly normally distributed with θ. One has

$$\text{Cov}(D, p_2) = \text{Cov}(\theta + \varepsilon, a\theta) = a\sigma_\theta^2. \tag{2.64}$$

Now (2.64), (2.37) and the normal projection theorem give

$$\mathbb{E}(D|p_2) = \frac{a\sigma_\theta^2}{a^2\sigma_\theta^2 + b^2\sigma_W^2} p_2 = \frac{\psi}{a}(a\theta + bW), \tag{2.65}$$

$$\mathbb{V}(D|p_2) = \frac{1}{\tau_u} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \frac{a^2\sigma_\theta^4}{a^2\sigma_\theta^2 + b^2\sigma_W^2} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \psi\sigma_\theta^2. \tag{2.66}$$

Plugging (2.65) and (2.66) into (2.34), using the result with (2.32) in the market-clearing condition (2.9):

$$\frac{M\tau_\epsilon(\theta - p_2)}{\gamma_d} + MX_k + \frac{(N-M)\tau_u(\psi\theta + \frac{\psi b}{a}W - p_2)}{\gamma_d} = W$$

solving for p_2 and matching coefficients with (2.33) yields

$$\begin{aligned} M\tau_\epsilon + (N-M)\psi\tau_u + \gamma_d\varphi\alpha_1 &= [M\tau_\xi + (N-M)\tau_u]a & (2.67) \\ (N-M)\tau_u\frac{\psi b}{a} - \gamma_d &= [M\tau_\xi + (N-M)\tau_u]b. & (2.68) \end{aligned}$$

Solving for a and b gives the expressions in (3.8) and (3.9).

It is immediately clear from (3.8) that $a > 0$, both numerator and denominator are always positive. Since $\psi > 0$ it follows also that $a \leq 1$ with a strict inequality only if $N = M$.

□

Proof of Lemma 12. To show the second equality in (2.42), we note that the equilibrium price in the inter-dealer market depends on the aggregate inventory by market clearing. If market clearing holds, then

$$\sum_{l=1}^{M} q_l + \sum_{k=M+1}^{N} q_l + W,$$

where the demand schedules are defined as in (2.32) and (2.34). Using these definitions, the normal projection theorem to determine the conditional expectations gives and solving the previous equation for p_2 gives

$$p_2 = \frac{W - \sum_{l=1}^{M} X_l - \frac{\theta M}{\gamma_d\sigma_\varepsilon^2}}{\frac{a\sigma_\theta^2(N-M)}{\gamma_d\left(a^2\sigma_\theta^2+b^2\sigma_W^2\right)\left(-\frac{a^2\sigma_\theta^4}{a^2\sigma_\theta^2+b^2\sigma_W^2}+\sigma_\varepsilon^2+\sigma_\theta^2\right)} - \frac{N-M}{\gamma_d\left(-\frac{a^2\sigma_\theta^4}{a^2\sigma_\theta^2+b^2\sigma_W^2}+\sigma_\varepsilon^2+\sigma_\theta^2\right)} - \frac{M}{\gamma_d\sigma_\varepsilon^2}}.$$

Using the definition of a and b in Proposition 13, some algebra yields that the denominator on the right-hand side of the previous equation is equal to $\frac{1}{b}$. Therefore, it follows that

$$\frac{\partial}{\partial X_k} p_2 = -b. \tag{2.69}$$

Using $\mathbb{E}(p_2) = a\theta$, one can now calculate

$$\begin{aligned} & \frac{\partial}{\partial X_k} \mathbb{E}_k \left[D(q_k - X_k) - p_2 q_k - \frac{\gamma_d}{2} \sigma_\varepsilon^2 (q_k - X_k)^2 \right] \\ = \quad & a\theta - b \frac{(1-a)\theta}{\gamma_d \sigma_\varepsilon^2} + b X_k. \end{aligned}$$

Since in equilibrium, one has $X_k = \frac{\varphi}{M}\alpha_1\theta$, the result follows.

□

Proof of Proposition 14. Step 1: expressions of β_1 and β_2

Substituting equation $p_c(x_i)$ and $p_v(x)$ into the price $P(x)$ formula gives

$$\beta_1 = \kappa \left(1 - \frac{\alpha_1}{\alpha_2} \right) + (1-\kappa) \left[a + \frac{b(1-a)}{\gamma_d \sigma_\epsilon^2} - \frac{b\varphi\alpha_1}{M} \right] \tag{2.70}$$

$$\beta_2 = \kappa \left(\frac{1}{\alpha_2} - \frac{\gamma_I \sigma_\epsilon^2}{2} \right). \tag{2.71}$$

Step 2: Solving α_1, α_2, β_1 and β_2

Combing the equations (2.47), (2.48), (2.70) and (2.71) and solving α_1, α_2, β_1 and β_2 leads to the following:

$$\alpha_1 = \frac{1 - a - \frac{b(1-a)}{\gamma_d \sigma_\epsilon^2}}{1 - \frac{b\varphi}{M} \frac{1-2\kappa}{(1-\kappa)\gamma_I \sigma_\epsilon^2}} \frac{1-2\kappa}{(1-\kappa)\gamma_I \sigma_\epsilon^2}, \tag{2.72}$$

$$\alpha_2 = \frac{1-2\kappa}{(1-\kappa)\gamma_I \sigma_\epsilon^2}, \tag{2.73}$$

$$\beta_1 = \frac{a + \frac{b(1-a)}{\gamma_d \sigma_\epsilon^2} - \frac{b\varphi}{M} \frac{1-2\kappa}{(1-\kappa)\gamma_I \sigma_\epsilon^2}}{1 - \frac{b\varphi}{M} \frac{1-2\kappa}{(1-\kappa)\gamma_I \sigma_\epsilon^2}}, \tag{2.74}$$

$$\beta_2 = \frac{\kappa}{2(1-2\kappa)} \gamma_I \sigma_\epsilon^2. \tag{2.75}$$

Note that when $\kappa < \frac{1}{2}$, one gets $\alpha_2 > 0$ and $\beta_2 > 0$.

Step 3: Show the existence of $\alpha_1 \in (0, \alpha_2]$ and $a \in (0, 1]$

We first show that α_1 is a decreasing function of a. Secondly, show that a is an increasing function of α_1, then prove that the two curves insect at $\{\alpha_1 \times a : (0, \alpha_2] \times (0, 1]\}$. Replacing b by equation (3.9) into formula (2.72) and derive the expression of α_1 as a function of a:

$$\alpha_1 = \frac{1}{2\gamma_d\varphi}\left[-M\tau_\epsilon - \gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right) + \sqrt{[M\tau_\epsilon + \gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2 + 4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}\right] \tag{2.76}$$

Once $a \leq 1$, one could derive that $\alpha_1 > 0$ and further α_1 is monotonically decreasing on a. Since

$$\begin{aligned}\frac{\partial \alpha_1}{\partial a} &= \frac{\frac{\alpha_2}{2}(\frac{1}{M}+1)\left[M\tau_\epsilon + \gamma_d\varphi\alpha_2(\frac{a}{M}+a-1) - \sqrt{[M\tau_\epsilon + \gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2 + 4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}\right]}{\sqrt{[M\tau_\epsilon + \gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2 + 4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}} \\ &+ \frac{\alpha_2 M\tau_\epsilon(\frac{-2a}{M}+\frac{1}{M}-1)}{\sqrt{[M\tau_\epsilon + \gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2 + 4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}} \\ &< 0.\end{aligned}$$

Moreover, one has

$$\begin{aligned}\lim_{a\to 0}\alpha_1 &= \frac{1}{2\gamma_d\varphi}\left[-M\tau_\epsilon + \gamma_d\varphi\alpha_2 + \sqrt{(M\tau_\epsilon - \gamma_d\varphi\alpha_2)^2 + 4\gamma_d\varphi\alpha_2 M\tau_\epsilon}\right] = \alpha_2, \\ \lim_{a\to 1}\alpha_1 &= \frac{1}{2\gamma_d\varphi}\left[-M\tau_\epsilon - \gamma_d\varphi\alpha_2 + \sqrt{(M\tau_\epsilon + \gamma_d\varphi\alpha_2)^2}\right] = 0.\end{aligned}$$

Since α_1 is monotonically decreasing on a, one gets $\alpha_1 \in [0, \alpha_2)$.

In terms of a, one can rewrite a as a function of α_1 by substituting ψ and τ_u by equations (2.37) and (2.36), and rearranging:

$$a = \frac{(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2(N\tau_\epsilon + \gamma_d\varphi\alpha_1) + \gamma_d^2\sigma_W^2(\tau_\theta + \tau_\epsilon)(M\tau_\epsilon + \gamma_d\varphi\alpha_1)}{(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2 N\tau_\epsilon + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)}$$

Next, one can compute the derivatives of a in terms of α_1 as

$$\begin{aligned}\frac{\partial a}{\partial \alpha_1} &= \frac{(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^4 N\tau_\epsilon + 2\gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2(M\tau_\epsilon + N\tau_\theta) + \gamma_d^4\sigma_W^4\tau_\epsilon(\tau_\theta + \tau_\epsilon)(M\tau_\epsilon + N\tau_\theta)}{\left[(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2 N\tau_\epsilon + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)\right]^2} \\ &+ \frac{\gamma_d^2\sigma_W^2(N-M)\tau_\epsilon^2(M\tau_\epsilon + \gamma_d\varphi\alpha_1)(M\tau_\epsilon + 2N\tau_\theta - \gamma_d\varphi\alpha_1)}{\left[(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2 N\tau_\epsilon + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)\right]^2} \\ &> 0.\end{aligned}$$

Moreover, the values at the two bounds:

$$\begin{aligned}\lim_{\alpha_1 \to 0} a &= \frac{M^2 N \tau_\epsilon^3 + M\gamma_d^2 \sigma_W^2 \tau_\epsilon(\tau_\theta + \tau_\epsilon)}{M^2 N\tau_\epsilon^3 + \gamma_d^2 \sigma_W^2 \tau_\epsilon (M\tau_\epsilon + N\tau_\theta)} < 1, \\ \lim_{\alpha_1 \to +\infty} a &= +\infty,\end{aligned}$$

where the inequality above inequality follows from $M < N$. So, one gets that a is a monotonically increasing function of α_1 and $a \in (\frac{M^2 N\tau_\epsilon^3 + M\gamma_d^2\sigma_W^2 \tau_\epsilon(\tau_\theta+\tau_\epsilon)}{M^2 N\tau_\epsilon^3+\gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon+N\tau_\theta)}, +\infty)$

Since $\alpha_1(a)$ is monotonically decreasing on a and $\alpha_1 \in [0, \alpha_2)$, $a(\alpha_1)$ is monotonically increasing on α_1 and $a \in (\frac{M^2 N\tau_\epsilon^3 + M\gamma_d^2\sigma_W^2 \tau_\epsilon(\tau_\theta+\tau_\epsilon)}{M^2 N\tau_\epsilon^3+\gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon+N\tau_\theta)}, +\infty)$, by the fixed point theorem, there exists one unique solution (α_1^*, a^*) to the problem

$$\begin{cases} a(\alpha_1) = a \\ \alpha_1(a) = \alpha_1 \end{cases}$$

as demonstrated in figure(2.7).

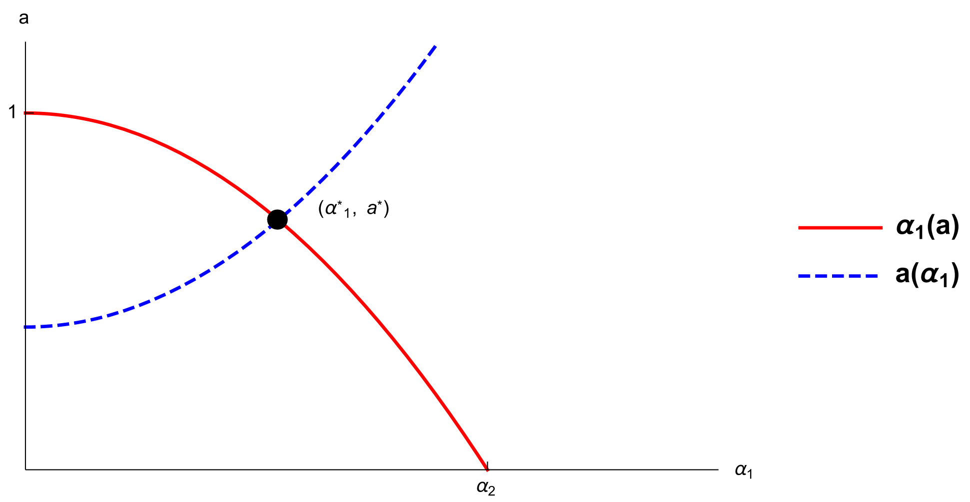


Figure 2.7: The curves of $\alpha_1(a)$ and $a(\alpha_1)$

Step 4: prove that $0 < \alpha_1^* \leq \alpha_2$ *and* $0 < a^* < 1$

First, it's obvious that $\alpha_1^* > 0$, we only need to prove that $\alpha_1^* \leq \alpha_2$. Suppose that $\alpha_1^* > \alpha_2$, then $a(\alpha_1^*) > a(\alpha_2)$, since a is an increasing function of α_1. Note that when $\alpha_1 = \alpha_2$, we have $\beta_1 = 1$ by equation(2.47), which implies that $a + \frac{b(1-a)}{\gamma_d \sigma_\epsilon^2} = 1$. Further, we get that either $a = 1$, or $a \neq 1$ and $b = \gamma_d \sigma_\epsilon^2$.

But neither of the solution is consistent with the properties of α_1 and a function. On one hand, if $a = 1$, then $a(\alpha_1^*) > a(\alpha_2) = 1$, which is contrary to $a^* \leq 1$. On the other hand, if $b = \gamma_d \sigma_\epsilon^2$, then $a < 0$ by the equation (3.9), which is also contrary to $a > 0$. So $\alpha_1^* \leq \alpha_2$.

Next, we prove that $a^* < 1$. Assuming $a^* \geq 1$, then we should have $\alpha_1(a^*) \leq \alpha_1(1) = 0$, which is contrary to $\alpha_1^* > 0$. So $a^* < 1$. As in the equilibrium, $0 < \alpha_1 < \alpha_2$, one could get that $0 < \beta_1 < 1$ since $\alpha_1 = (1 - \beta_1)\alpha_2$.

Last, we verify that the second order condition of the maximization problem (2.45) is satisfied when $\kappa < \frac{1}{2}$:

$$-(2\beta_2 + \gamma_I \sigma_\epsilon^2) = -\left(\frac{\kappa}{1 - 2\kappa}\gamma_I \sigma_\epsilon^2 + \gamma_I \sigma_\epsilon^2\right) = -\frac{1 - \kappa}{1 - 2\kappa}\gamma_I \sigma_\epsilon^2 < 0$$

The fact that Lemma 11 is applicable in order to derive the dealers' quoting strategies is proved as in the proof of Proposition 9.

Step 5: Show the equilibrium does not exist when $\kappa \geq \frac{1}{2}$

First, when $\kappa = \frac{1}{2}$, one has $\alpha_1 = 0$, $\alpha_2 = 0$ and $\beta_2 = +\infty$ from equation (2.72), (2.73) and (2.71). That is, the investor does not trade, and the price is not defined since $P(x_i) = \infty$. Thus, the equilibrium does not exist.

Second, when $\kappa > \frac{1}{2}$, one has $\beta_2 < 0$, and the second order condition of the optimization problem (2.45):

$$-(2\beta_2 + \gamma_I \sigma_\epsilon^2) = -\left(\frac{\kappa}{1 - 2\kappa}\gamma_I \sigma_\epsilon^2 + \gamma_I \sigma_\epsilon^2\right) = -\frac{1 - \kappa}{1 - 2\kappa}\gamma_I \sigma_\epsilon^2 > 0.$$

This means that the investor's maximization problem does not have a solution and the equilibrium does not exist.

□

Proof of Proposition 15. Claim 1: investors prefer to trade on the platform as $\sigma_\delta^2 \to 0$.

Equation (2.29) implies that each investor's ex-ante profits go to zero in the centralized market if $\sigma_\delta \to 0$. Proposition 14 states that an equilibrium exists if $\kappa < \frac{1}{2}$. All that is left to show is that expected profits for each investor remain strictly positive as $\sigma_\delta^2 \to 0$. This can be seen as follows. From the definition of the investors' utility (2.2), the dealers expected quotes conditional on a response on the platform (2.44) and the investors' equilibrium strategy (2.30) one obtains the following expression for the expected profit π_i of an investor trying to trade on the platform:

$$\pi_i = (1-(1-q)^M)\mathbb{E}\left[(\alpha_1\theta+\alpha_2\delta_i)\left(\theta+\delta-\beta_1\theta-(\alpha_1\theta+\alpha_2\delta_i)\left(\frac{\gamma_I\sigma_\varepsilon^2}{2}+\beta_2\right)\right)\right]. \tag{2.77}$$

As $\sigma_\delta^2 \to 0$, one obtains from (2.77) that

$$\begin{aligned}
\pi_i &\to (1-(1-q)^M)\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\alpha_1\left(\frac{\gamma_i\sigma_\varepsilon^2}{2}+\beta_2\right)\right] \\
&= (1-(1-q)^M)\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\alpha_1\frac{1}{2\alpha_2}\right] \\
&= (1-(1-q)^M)\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\frac{1-\beta_1}{2}\right] \\
&> 0,
\end{aligned}$$

where the second line follows from the expressions for β_1 and α_2 in (2.72) and (2.75). The third line follows from the expressions for α_1, α_2 in (2.47) and (2.48). The inequality follows from $\alpha_1 > 0$ and (2.74), which implies $\beta_1 < 1$, since $b < 0$ and $a < 1$ by the proof of Proposition 14. This proves the first claim.

Claim 2: investors prefer to trade in the centralized market as $\sigma_\delta^2 \to \infty$*.*

Computing the expectation in (2.77) gives

$$\pi_i = (1-(1-q)^M)\left\{\underbrace{\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\alpha_1\left(\frac{\gamma_i\sigma_\varepsilon^2}{2}+\beta_2\right)\right]}_{A}+\underbrace{\alpha_2\sigma_\delta^2\left[1-\alpha_2\left(\frac{\gamma_i\sigma_\varepsilon^2}{2}+\beta_2\right)\right]}_{B}\right\} \tag{2.78}$$

In the(2.78), A is not affected by σ_δ^2. Using the expressions for β_1 and α_2 in (2.72) and (2.75), one gets

$$B = \frac{1-2\kappa}{1-\kappa}\frac{\sigma_\delta^2}{2\gamma_I\sigma_\varepsilon^2} = \frac{1-2\kappa}{1-\kappa}\pi_i^c, \tag{2.79}$$

where π_i^c is the expected profit of the investor in the centralized market as defined in (2.29). It trivially follows that $\pi_i^c \to \infty$ as $\sigma_\delta^2 \to \infty$. Therefore, it follows that

$$\lim_{\sigma_\delta^2\to\infty}\frac{\pi_i}{\pi_i^c} = \lim_{\sigma_\delta^2\to\infty}(1-(1-q)^M)\frac{A+\frac{1-2\kappa}{1-\kappa}\pi_i^c}{\pi_i^c} = (1-(1-q)^M)\frac{1-2\kappa}{1-\kappa} < 1,$$

because of our assumption $\kappa > 0$. Therefore, investors will have a higher expected payoff in the centralized market as $\sigma_\delta^2 \to \infty$.

$\square$

Proof of Proposition 16. We will show that the term denoted by A in (2.78) goes to zero as $\mu \to \infty$. Then it follows from (2.79) and $\kappa > 0$ that $\pi_i < \pi_i^c$ as $\mu \to \infty$, with $\pi_i < \pi_i^c$ defined as in (2.77) and (2.29).

In order to show $A \to 0$ as $mu \to \infty$, it is sufficient to show that $\alpha_1 \to 0$ as $mu \to \infty$, since β_2 is by (2.75) unaffected by μ and β_1 is by (2.74) between zero and one.

We show In order to show $A \to 0$ as $mu \to \infty$ as follows. Define the function $a(\cdot)$ as in the proof of Proposition 14. It has been shown in the proof of Proposition 14 that $\alpha_1 > 0$ for any $\mu > 0$ must hold in equilibrium. For any fixed $\alpha_1 > 0$, one has $a(\alpha_1) \to \infty$ for $\mu \to \infty$. The equilibrium condition $a(\alpha_1) = a < 1$ can only hold if $\alpha_1 \to 0$ for $\mu \to \infty$ (since $a(\cdot)$ is monotone increasing with $\lim \alpha_1 a(\alpha_1) \in (0,1)$). This proves the claim.

$\square$

Proof of Proposition 17. Using (2.78), (2.79) and (2.29), one gets

$$\lim_{q \to 1} (\pi_i - \pi_i^c) = \lim_{q \to 1} A,$$

where A is defined as in (2.78). We proceed as in the proof of Proposition 15:

$$\begin{aligned}
\lim_{q \to 1} A &= \lim_{q \to 1} \alpha_1 \sigma_\theta^2 \left[(1 - \beta_1) - \alpha_1 \left(\frac{\gamma_i \sigma_\varepsilon^2}{2} + \beta_2 \right) \right] \\
&= \lim_{q \to 1} \alpha_1 \sigma_\theta^2 \left[(1 - \beta_1) - \alpha_1 \frac{1}{2\alpha_2} \right] \\
&= \lim_{q \to 1} \alpha_1 \sigma_\theta^2 \left[(1 - \beta_1) - \frac{1 - \beta_1}{2} \right] \\
&> 0,
\end{aligned}$$

where the second line follows from the expressions for β_1 and α_2 in (2.72) and (2.75). The third line follows from the expressions for α_1, α_2 in (2.47) and (2.48). The inequality follows from $\alpha_1 > 0$ and (2.74), which implies $\beta_1 < 1$ as $q \to 1$, since $b < 0$ and $a < 1$ by the proof of Proposition 14.

$\square$

Chapter 3

Informed Traders and Dealers in the FX Forward Market (joint work with Pierre Collin-Dufresne and Peter Hoffmann)

3.1 Introduction

The two-tiered market structure, where clients' trades are intermediated by dealers who can trade amongst each other in an interdealer market, remains prevalent in many OTC markets including fixed income, credit, and foreign exchange. In theory, such a structure may naturally arise if clients are differentially informed, and can signal their type to dealers who can price-discriminate (Seppi (1990b), Lee and Wang (2019)). More recently, Glode and Opp (2016, 2019) argue that, in the presence of asymmetric information, intermediation chains may be needed to generate efficient trading behavior. A property of these intermediation chains is that trades are expected to occur between counterparties that are similarly informed. In practice however, it is less clear how these efficient intermediation chains can arise[1] and whether we can empirically observe that trading relationships are, at least to some extent, determined by informedness of the different traders.

[1] Glode and Opp (2016, 2019) consider the case in which the trading network is held fixed. If this would not be the case, informed traders had incentives to choose less informed counterparties.

In this paper we use data on foreign exchange transactions made available through the European Market Infrastructure Regulation (EMIR) to shed new light on the functioning of one of the largest OTC markets: the euro-dollar forward exchange rate market. This market is largely two-tiered in that most client transactions occur with a limited set of dealers and there is a very active interdealer market. Since the data set contains all the individual transactions in the EU with information on trader identities, we can ask the following questions. Is there evidence that clients are differentially informed? Is there evidence that dealers are differentially informed? Are markups charged by dealers related to client and/or dealer informedness? Are the client-to-dealer (and dealer-to-dealer) trading networks affected by these differences in informedness?

To answer the first question, we measure price impact of individual clients' trades at different horizons (1-minute, 30-minutes, 1-day). A positive price impact implies that clients tend to buy (sell) from the dealer when benchmark rates increase (decrease) subsequent to their trade. Such situation should arise if clients are on average better informed than dealers about future exchange rate changes. This is the standard adverse selection mechanism presented in the traditional microstructure literature (e.g., Kyle (1985) and Glosten and Milgrom (1985)). Of course, it is perhaps less intuitive to think that clients have private information about future exchange rate fundamentals, which we typically think of as reflecting macro-economic risks. However, as we show in a simple theoretical model, clients information may pertain to their individual order flow, which may be correlated with total order flow, which in the short run may affect the change in exchange rates (see also Evans and Lyons (2002, 2005)). Alternatively, some investors may also be better at interpreting public news, say about macroeconomic fundamentals, and thus effectively also have private information about systematic sources of risk (e.g., Kim and Verrecchia (1991)). Empirically, we find that on average clients' price impact is highly statistically significant and positive at a 1-minute horizon. At a longer (1-day) horizon it remains highly statistically significant and positive for hedge funds on average. However, there is considerable cross-sectional and time-series variation across traders. When we look at the individual trader level, we find significant persistence in price impact. Breaking down the sample into subperiods, we find that traders that tend to have a higher price impact in the first subperiod tend to remain in the high-price impact group in subsequent periods. This suggests that some (groups of) traders are consistently better informed, in that their trades seem to, on average, correctly anticipate future exchange rate changes.

At some level, these findings are consistent with the original findings of Evans and Lyons (2002), who documented, using 4 months of data in 1996, that it was possible to predict future exchange rate changes based on aggregate interdealer order flow. Since, as we show in our theoretical model, one would expect the interdealer order flow to be driven by their clients' order flow, it seems natural to anticipate that there should be some information in at least some of the clients' trades. We confirm this intuition by extending the Evans and Lyons (2002) study to the individual dealer level. Specifically, we investigate whether the aggregated clients' order imbalance observed by each dealer allows them to predict future exchange rate movements. We find strong evidence of predictability at the dealer level. That is, individual dealers could earn significant Sharpe ratios from trading based on their clients aggregated order flow. However, there is also substantial cross-sectional variation across dealers. We label the dealers with the highest predictive client order flow, the 'informed dealers,' and we study the characteristics of these dealers and whether we see specific patterns in the client-dealer trading network. Interestingly, we find that dealer informedness is not isomorphic to the standard centrality measures such as connectedness. We find that more informed dealers indeed use their information when giving quotes to traders. Markups are generally higher, the higher the price impact of the trade. The more informed the dealer, the stronger is this effect. We also find that traders are more likely to trade with informed dealers if they are informed themselves. Relatedly, for all traders that are not HFT whose investment horizon is arguably very short, traders are more likely to trade with informed dealers if volatility, a proxy for adverse selection, is high.

To interpret our empirical findings we develop a simple model of a two-tiered OTC market in which dealers intermediate trades between their customers and subsequently hedge their inventory risk in an interdealer market. In this model, order flow is informative about future price changes. Moreover, some dealers may forecast future price changes better than other dealers. Those dealers not only incorporate some of their information in their markups, but are also more likely to further attract informed traders due to an adverse selection problem that the less informed dealers face: If an informed trader asks an uninformed dealer for a bid and an ask price, this dealer likely has difficulties to respond as the willingness of the informed trader to buy or sell at a given ask or bid price will mean bad news for the uninformed dealer. On the other hand, the informed dealer is able to correctly price an asset and is able to make a market for informed traders.

By focusing on the informedness of traders, we extend a growing literature on the network structure of OTC markets.[2] While Wang (2017) develops an inventory-based model of the OTC market structure and Sambalaibat (2018) develops a model in which dealers specialize on the trading frequency of their clients, our findings suggest that dealers also specialize based on the informedness of their customers. In the light of theoretical and empirical results of Babus and Kondor (2018) and Kondor and Pintér (2019) it is surprising that the more informed market participants have fewer counterparties than their less informed counterparts. However, this result is not unreasonable, since informed counterparties are especially vulnerable to infor-

[2]Characteristic for OTC markers is a core-periphery structure, see for instance Abad et al. (2016), Li and Schürhoff (2019) or Neklyudov et al. (2017).

mation leakage that arises from being in contact with many counterparties.[3] While previous research on a two-tiered OTC market pointed out the benefits of trading with connected dealers, who can provide more immediacy (Di Maggio et al. (2017) and Li and Schürhoff (2019)), we show in this paper that it can sometimes be beneficial to trade with a less connected dealer: In the presence of strong information asymmetries, only informed dealers may be willing to provide quotes to informed traders. However, these dealers generally have fewer counterparties.

Also Bjønnes et al. (2017) and Ranaldo and Somogyi (2018) study informed trading in the FX market. The advantage of the EMIR dataset compared to the datasets used in these studies lies in the availability of the traders' and the dealers' identities. Thus, we can not only document the presence of informed trading, but also study the persistence of informedness and characteristics as well as the trading behavior of informed traders with different dealers.

In an influential paper, Evans and Lyons (2002) show that aggregate orderflow predicts future price changes in the FX market. Menkhoff et al. (2017) show that the order flows from different subsets of traders have different forecasting abilities. We connect to this strand of literature by showing that the order flow of different dealers has different forecasting abilities. Moreover, the informedness measures we obtain for traders and dealers allow us to study the traders' dealer choice problem.

3.2 Data and Summary Statistics

We use three different databases: The EMIR database contains information on derivatives transactions in which at least one counterparty is located in the EU. We use the full database to which the ESRB and ESMA have unique access. We focus on the FX forward market for the following reasons. First, it is one of the largest derivatives markets. As shown in Abad et al. (2016), the FX forward market is the second largest derivatives market in the EU. While the interest rate swaps (IRS) market is still larger in terms of notional volume, Abad et al. (2016) show that approximately 85% of notional volume of IRSs is being traded among G16 dealers and other banks. Nonfinancial firms only generate less than 1% of the notional volume traded in the IRS market. Traders in the FX forward market are more diverse. Less than 70% of notional volume is generated among G16 dealers and other banks.

The other two databases we use are the ORBIS database, which contains information on the different

[3] Hendershott and Madhavan (2015b) argue that information leakage is an important concern when evaluating whether to contact many dealers via an RFQ trading protocol or a single dealer in the voice market. Hagströmer and Menkveld (2019) measure information flow between dealers in the FX market and Liu et al. (2018) model the information leakage in a two-tiered OTC market, showing that informed traders may benefit from limiting the number of contacted dealers.

types of traders, and the Thomson Reuters Tick History (TRTH) from which we derive benchmark prices for the forward contracts. The following subsections describe the data in more detail.

3.2.1 EMIR and ORBIS data

We use the EMIR activity report and focusing on the last message submitted for each trade. Moreover, we look at the period from May 2018 to April 2019 and restrict our attention to transactions that happen between Monday and Friday as well as between 8am and 8pm UTC and exclude transactions with no reported price rate, or markups with an absolute value of more than 5 %. How we determine the markup for each trade is explained further below in this section. As in Abad et al. (2016) and Hau et al. (2019) we use the ORBIS database to assign a type to each trader. Possible types are FUND, BANK, G16, INSURANCE & PENSION, NON-FINANCIAL, CENTRAL BANK and EMPTY. Firms not covered by the ORBIS dataset are also classified as EMPTY. The EMIR database reports only the legal entities that were involved in a transaction. Many firms, especially the G16 dealers use many legal entities. The ORBIS dataset allows us to associate each legal entity with its parent company.

Table 3.1: **Averages of firm characteristics for D2C market.** This table shows the number of firms for each type in the sample, shows how large the notional volume in EUR per trade involving these firms is and looks at the trader characteristics for trades involving these firms. The trader characteristics include the number of monthly counterparties and average monthly trades conditional on trading in that month. The sample period ranges from May 2018 to April 2019. The last column shows the average maturity (in days) of the contracts that are traded by the firms of different types. Only trades between dealers and other firms have been considered when calculating the statistics. Numbers are rounded to the nearest integer or to the nearest hundred thousand.

trader type	# traders	avg. notional	CPs/month	trades/month	avg. maturity
CENTRAL BANK	45	35,013,323	5	41	24
EMPTY	14,215	13,395,139	3	144	42
FUND	11,055	10,155,767	4	1,190	38
GOVERNMENT	94	41,151,354	13	1,118	43
INSURANCE & PENSION	524	66,889,994	10	406	36
NON-FINANCIAL	6,739	9,458,687	7	1,349	59

Virtually no trades in the FX forward dealer-to-customer (D2C) market are cleared through CCPs. Of almost 3 million trades, we only have less than 500 trades involving CCPs. Table 3.1 considers all customers that are not CCPs and shows characteristics of D2C trades in our sample for each type of trader. The average notional of trades by insurers and pension funds is largest. The mean notional volumes of trades by funds and nonfinancial firms are considerably smaller (EUR 10.2 million and EUR 9.5 million, respectively). As traders, funds have on average 4 counterparties in a month conditional on trading in the first place. Firms classified as EMPTY, have even fewer counterparties with on average 3 trading partners per month conditional on trading in that month. Nonfinancial firms have more counterparties per month, with an average number of monthly trading partners of almost 7. Governments have the highest number of monthly counterparties. Similar comments apply to the average number of monthly trades conditional on trading in the first place. Strikingly, despite having the largest average notional volume per trade, insurance firms and pension funds are associated with a small number of monthly trades compared to funds or non-financial firms. One can also see in the last column of Table 3.1 that there is some disperion in the types of contracts traded across the different types of traders. While central banks rather trade contracts with a short maturity (24 days), funds, governments and nonfinancial firms trade contracts with longer maturities (38, 43 and 59 days, respectively). Figure 3.5 shows the distribution of maturities across all traded contracts.

Firms that act as dealers in the FX forward market are labelled either as banks or as G16 dealers. Table 3.2 shows the same statistics considered in Table 3.1 for the two types of dealers. The average notional of trades involving G16 dealers is roughly EUR 15 million which is considerably larger than the average notional of roughly EUR 7 million of trades between other banks and their clients. Consistent with the core-periphery structure described in Abad et al. (2016) and analogous findings for other OTC markets, G16 dealers have a lot more monthly counterparties and on average much more monthly trades than other banks. The average maturities of the contracts traded by the two different types of intermediaries are relatively similar.

Table 3.2: **Dealer characteristics in the D2C market.** This table shows the number of firms for each type in the sample, shows how large the notional volume in EUR per trade involving these firms is and looks at the dealer characteristics for trades involving these dealers. The dealer characteristics are the number of counterparties and the number of trades in the sample period from May 2018 to April 2019. The last column shows the average maturity of the contracts that are traded by the dealers of different types in days. Only trades between dealers and other firms have been considered when calculating the statistics. Numbers are rounded to the nearest integer or to the nearest hundred thousand.

trader type	# dealers	notional/trade	CPs	trades	avg. maturity
BANK	201	7,232,527	821	42,824	42
G16	16	14,672,750	4,553	208,195	44

This paper mostly focuses on the D2C market for two reasons. First, assuming D2C trades are client-initiated allows us to sign these trades. Since the EMIR dataset does not indicate which counterparty initiates the trade, it is harder to sign client-to-client (C2C) or dealer-to-dealer (D2D) trades. Second, the C2C market is not very active. While a limited number of firms generated a high number of trades in the sample period, notional volumes tend to be small. Thus, compared to the notional volume in the D2D and D2C markets, the C2C market is small. Table 3.15 in Appendix 3.8.1 shows how many trades were executed between the different groups of traders in the C2C market and Table 3.16 in Appendix 3.8.1 shows the corresponding average notional volumes.

The D2D market, on the other hand, is very large. Table 3.17 in Appendix 3.8.1 shows how many trades were executed between the different types of dealers in the D2D market and how much notional was exchanged on average.

Table 3.3 breaks down the trading done by different types of counterparties with the two sets of dealers. The most striking feature is the predominance of trading with G16 dealers as opposed to with smaller banks. In terms of notional volume G16 dealers execute around 80% of the volume. In terms of the number of trades we see more diversity. For instance, government entities execute 94% of their notional trades with G16 dealers, while that share falls to 62% for nonfinancial firms.

Table 3.3: **Who trades with whom in the D2C market?** This table shows how much notional volume in EUR a trader of each type trades on average with G16 dealers and other banks, respectively, as well as how many transactions happen on average between a trader of a given type and G16 dealers of other banks, respectively. Notional values are rounded to the nearest million. Numbers of trades are rounded to the nearest integer.

	notional volume		# trades	
	total	% traded with G16	total	% traded with G16
CENTRAL BANK	3,001mn	74%	87	85%
EMPTY	436mn	84%	33	77%
FUND	1,142mn	87%	112	79%
GOVERNMENT	8,900mn	85%	221	94%
INSURANCE & PENSION	8,302mn	89%	125	83%
NON-FINANCIAL	782mn	82%	80	62%

3.2.2 Benchmark rates

We use data from the Thomson Reuters Tick History database in order to calculate benchmark forward rates. We follow the same procedure to compute a benchmark for the spot rate and the forward adjustment separately. Specifically:

1. For each second, the best bid and ask prices among all dealers are determined
2. In case there are no observations in a second, the benchmark price from the previous second is used. However, a given price can only be carried forward for 30 consecutive seconds.
3. In each second, the benchmark is the average of the best bid and ask.

The final benchmark forward rate for a given tenor in a given second is the sum of the benchmarks for spot rate and forward adjustment for a specific tenor. The tenor can be overnight, 1 week, 2 weeks, 3 weeks, 1 month, 2 months, 3 months, 6 months, 9 months or 1 year. In order to obtain the benchmark rates for the forward contracts in the EMIR dataset, we use linear interpolation between the two nearest-maturity benchmark rates. We use the same procedure to calculate the benchmark rates 1, 5 or 30 minutes after each transaction.

3.2.3 Volatility

As exchange-rate volatility measure we use an exponentially weighted moving average of squared returns of the one-week forward exchange rate from one second to the next, i.e.

$$volatility_t^2 = 0.001 \times ret_t^2 + 0.999 \times volatility_{t-1}^2,$$

where ret_t refers to the one-second return (between t and $t+1$) on the one-week forward exchange rate. This measure captures short-lived fluctuations in volatility within a day. Using a different maturity forward rate (instead of one-week) will not significantly affect this measure as short-run fluctuations in the forward exchange rates are mostly driven by the spot exchange rate.

3.2.4 Price impact

The 1-minute price impact is defined as the 1-minute change in the benchmark rate times the direction of the trade (+1 if it is a client-buy and −1 if it is a sell). Analogously, we calculate the x-day permanent price impact as the difference between the benchmark rate at the time of a transaction and the end-of-day benchmark rate x-days later times the direction of the trade.[4]

Table 3.18 shows the average price impact of trades by the different groups of market participants along with the corresponding standard deviations for the 1-minute and 1-day horizons. We see that 1-minute price-impact tends to be positive and highly statistically significant for all trader groups except for central banks. A positive price impact implies that clients buy (sell) on average when benchmark rates increase (decrease) subsequent to their trade. Such situation should arise if clients are on average better informed than dealers about future exchange rate movements, that is if dealers face adverse selection. This is the mechanism presented in the traditional microstructure literature (e.g., Kyle (1985) and Glosten and Milgrom (1985)). Of course, it is perhaps less intuitive to think that clients have private information about future exchange rate movements, which we typically think of as reflecting macro-economic risks. However, as we show in the model section, clients' information may pertain to their individual order flow, which may be correlated to total order flow, which in the short run may affect the change in exchange rates (see also Evans and Lyons (2002, 2005)). Alternatively, some investors may also be better at interpreting public news, say about macroeconomic fundamentals, and thus effectively also have private information about systematic sources of risk (e.g., Kim and Verrecchia (1991)). Interestingly, we see that, at a 1-day horizon, price impact remains

[4] For that calculation we hold interpolation weights fixed and for each tenor, use the last quoted price before 8pm UCT. We further ignore week-ends that is treat the data as if Mondays follow Fridays.

positive and highly statistically significant on average for all traders except for Insurance & Pension trader types who display negative price impact, which effectively implies that at the longer horizon their trades on average tend to loose money, as one might explain if their trading were motivated by hedging motives for example.

Table 3.19 shows the price impact of the C2D trades aggregated at the dealer level for G16 and Banks separately. We find strong evidence that dealers face adverse selection both at the 1-minute and 1-day horizon, as price impact is positive and highly statistically significant in all cases.

Of course, if dealers expect to incur a price impact cost on their client trades, it would be natural for them to charge an ex-ante premium, a 'markup,' to account for this risk. We next explain how we compute markups on C2D trades.

3.2.5 Markups

We define a trade's markup as the difference between the transaction rate and the benchmark rate times the direction of the trade. Table 3.20 shows the average markups for the different trader types and their respective standard deviations. We see that markups tend to be positive and statistically significant for all trader types. There seems to be an interesting positive relation between price impacts and markups, in that trader types that have higher price impact typically tend to be charged higher markups. For example, central banks face the smallest markup and hedge funds the highest. However, the relation is not monotone as non-financial traders face high markups (even higher than funds on average) and there seems to be substantial cross-sectional variation in markups. Table 3.21 shows the markups aggregated at the dealer type. G16 dealers charge on average significantly smaller markups then non G16 banks, but there is a lot of variation across dealers. Figure 3.6 in Appendix 3.8.1 shows the time series of daily average markups across all trades. The distribution of markups does not seem to exhibit any trends. In the next sections, we take a closer look at the determinants of price impact and markups, and at the relation between both.

3.3 Informed Clients

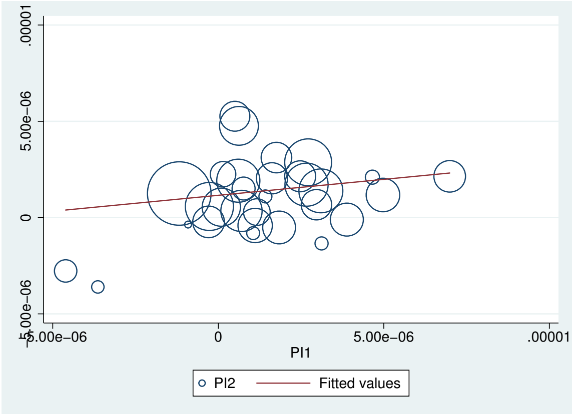
As shown in Tables 3.18 and 3.19, there is evidence that some groups of traders have significant positive price impact both at the 1-day and 1-minute horizons, which suggests that some traders have better information about future exchange rate changes. However, there is also substantial cross-sectional variation in measured price impact across traders and over time. In this section, we investigate if there are persistent differences in price impact across traders. That is, if we can find evidence that some (groups of) traders are consistently better than others at predicting future exchange rates, in the sense that they earn consistently significantly higher trading profits. To be more specific, suppose that the price impact of trader i at time t is given by

$$PI_{it} = \mu_i + \varepsilon_{it},$$

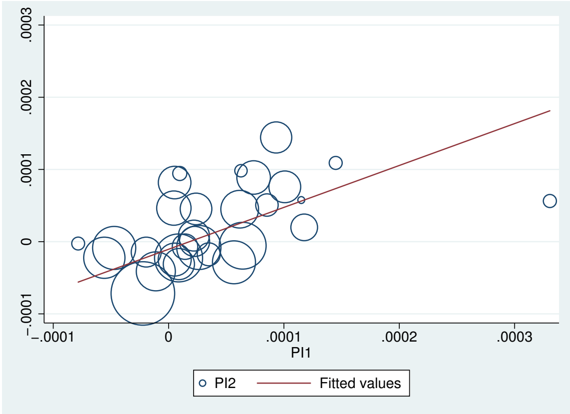
where ε_{it} is iid with finite variance and zero expectation and $\mu_i \in \mathbb{R}$. We would like to test if there is dispersion in μ_i across traders and specifically, whether some (groups of) traders have significantly higher $\mu_i > \mu_j$, say. The more transactions we observe for a given trader, the better our estimate of μ_i. Analogously, forming groups of traders gives us a relatively precise estimate of a group's average μ_i. In order to still have a sufficient dispersion in those averages, our number of groups cannot be too low. We choose to form 30 groups to obtain a good trade off between minimizing the error variance while retaining enough dispersion in the groups' price impacts. Lastly, we would like to form different groups of traders according to characteristics that are correlated with the μ_i, but not with the error ε_{it}. This rules out sorting traders based on their realized price impact, since this measure is correlated with the error. Instead, we sort traders based on their number of trades, since this is likely uncorrelated with the error, but potentially correlated with skill μ_i.

To be specific, we proceed as follows. Considering only traders that traded in both halves of our sample period, we sort the traders based on the number of trades done in the first half of the sample. Then we keep adding the traders to a group until the total number of trades in that group exceeds 1/30 times the total number of trades in the first half of the sample period. We then start adding the next traders to a group until the total number of trades of group 1 and 2 exceeds 2/30 times the total number of trades in the first half of the sample period. We continue until we have sorted the traders into 30 groups. For each group and each half of our sample period, we calculate the average permanent price impact of all trades made by that group. In Panel A of Figure 3.1, the average 1-minute price impact in the second half of the sample (PI2) of each group is plotted against the corresponding average 1-minute price impact in the first half of the sample (PI1). In Panel B, the same is done using the 1-day price impact. Both Panels of Figure 3.1 suggest that price impact is persistent.

Different traders may have different investment horizons. In particular, there may be a set of traders whose goal it is to trade intraday on very short-lived signals. Even though these market participants might trade very profitably, their 1-day price impact may not be very persistent, since it is not their objective to trade on long-term price changes. On the other hand, we would expect strong persistence in 1-minute price impact if there are informed traders in this set. Figure 3.7 in Appendix 3.8.1 shows the distribution



Panel A (1 minute).



Panel B (1 day).

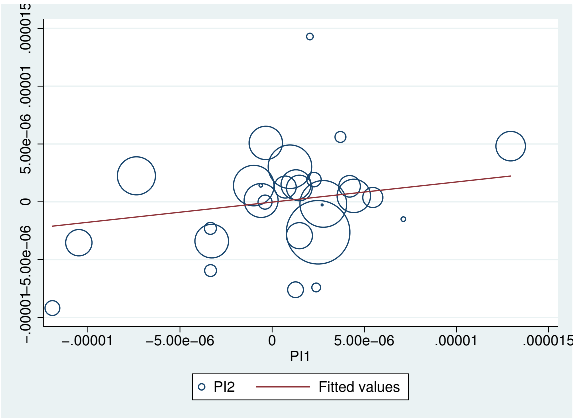
Figure 3.1: **Persistence of price impact.** In Panel A, a group's average 1-minute price impact in the second half of the sample is plotted against its 1-minute price impact in the first half of the sample. In Panel B, a group's average 1-day price impact in the second half of the sample is plotted against its 1-day price impact in the first half of the sample. Larger circles correspond to groups in which traders generate a higher notional volume (EUR) per trade in the first half of the sample. The red lines show the fitted values of linear regressions.

of the traders' average numbers of trades per day (conditional on trading on a day). Roughly 1% of traders do more than 10 trades per day if they trade at all. We classify these traders as HFT. To investigate whether there are differences between the price impact of the HFT traders and the lower-frequency traders, we compare the persistence of the 1-minute price impact of HFT traders in Panel A of Figure 3.2 with that of non-HFT traders in Panel B. We find strong evidence - indeed stronger compared to Panel A of Figure 3.1 - of persistence in 1-day price impact for HFT traders (Panel A), and hardly any evidence of persistence for non-HFT traders in Panel B.

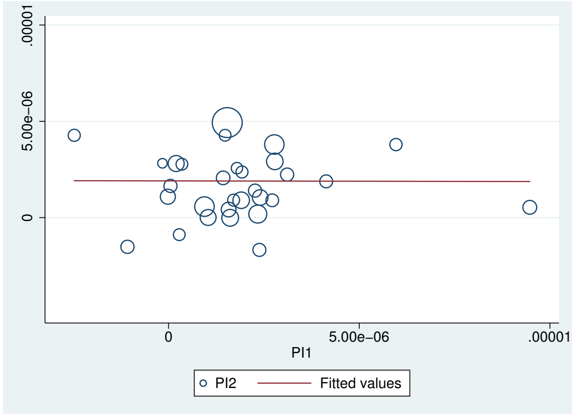
One can see in both panels that a high price impact is not related to high notional volume per trade, as it might be the case if the price impact were inventory-driven. Such an inventory-based price impact may arise as follows. If a dealer takes a large customer order, this is a private transaction between a dealer and its client. Other dealers will not change their quotes at the very time of the transaction. But shortly after the trade, the dealer who took the customer order may try to offset the inventory shock in the interdealer market, leading other dealers to change their quotes as well.

It seems that traders with high price impact typically have a lower average notional per trade. Similarly, Figure 3.9 in the appendix shows that groups with high price impact are also not groups with high total notional traded. Thus, price impact seems more likely to be information-based.[5]

[5]However, this information may very well be information about aggregate inventory or order flow.



Panel A (HFT).

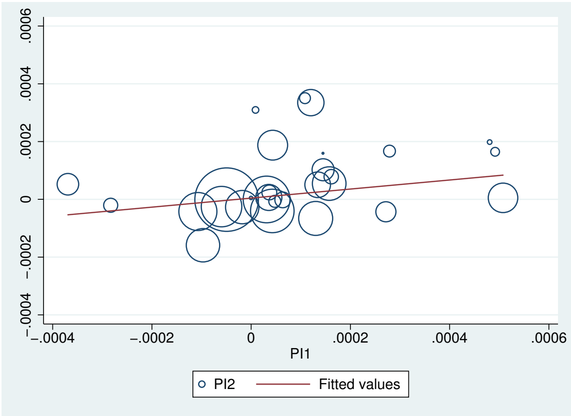


Panel B (non-HFT).

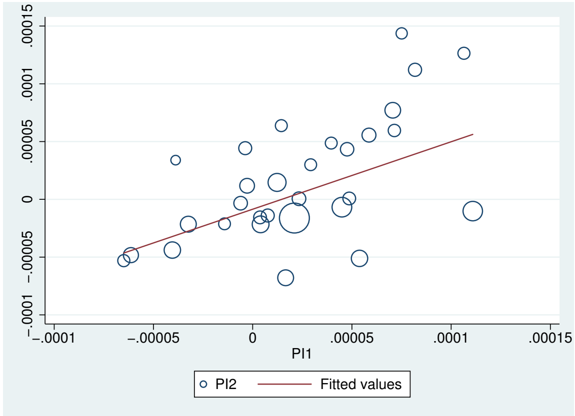
Figure 3.2: **Persistence of 1-minute price impact for HFT vs. non-HFT market participants.** In both panels, a group's average 1-minute price impact in the second half of the sample is plotted against its 1-minute price impact in the first half of the sample. In Panel A, only HFT market participants are considered and in Panel B, only non-HFT market participants are considered. Larger circles correspond to groups in which traders generate a higher notional volume (EUR) per trade in the first half of the sample. The red lines show the fitted values of linear regressions.

If instead we focus on the 1-day price impact, the opposite picture emerges. Figure 3.3 shows that the 1-day price impact is more persistent for non-HFT traders (Panel B) than for HFT traders (Panel A). Neither in Figure 3.2 nor in Figure 3.3, is it the case that high price impact is associated to high notional volume per trade. Figures 3.10 and 3.11 in the appendix show that, for the same groups of traders, high price impact is not related to high total notional volume traded. Again, this is consistent with information-driven price impact, but inconsistent with inventory-based explanation of persistently positive price impact.

In order to formally assess the persistence of the price impact in Figures 3.1 to 3.3, we regress a group's average price impact in the second half of the sample on its average price impact in the first half of the sample. The results for the various groups of traders and different horizons are shown in Table 3.4. One can see that estimates for the coefficient in front of the price impact in the first half of the sample are statistically significant except for the 1-minute price impact of non-HFT traders. Moreover the R^2 statistics from the regressions are generally large, especially for non-HFT traders' 1-day price impact and for HFT traders' 1-minute price impact.



Panel A (HFT).



Panel B (non-HFT).

Figure 3.3: **Persistence of 1-day price impact for HFT vs. non-HFT market participants.** In both panels, a group's average 1-day price impact in the second half of the sample is plotted against its 1-day price impact in the first half of the sample. In Panel A, only HFT market participants are considered and in B, only non-HFT market participants are considered. Larger circles correspond to groups in which traders generate a higher notional volume (EUR) per trade in the first half of the sample. The red lines show the fitted values of linear regressions.

Table 3.4: **Persistence of price impact for different traders and horizons.** This table regression coefficients and robust standard errors for the regression

$$PI2 = \beta_0 + \beta_1 PI1 + \varepsilon,$$

where $PI2$ is the price impact in the first half of the sample, $PI2$ is the price impact in the second half of the sample and ε is an error term. We use the average price impact generated by the groups of traders shown in Figures 3.1 to 3.3. Columns 1 and 2 refer to the groups from Panel A and B, respectively, of Figure 3.1. Columns 3 and 4 refer to the groups from Panel B of Figures 3.2 and 3.3, respectively. Columns 5 and 6 refer to the groups from Panel A of Figures 3.2 and 3.3, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	all	all	non-HFT	non-HFT	HFT	HFT
	1 min	1 day	1 min	1 day	1 min	1 day
PI1	0.33**	0.34**	-0.01	0.72***	0.37***	0.20**
	(0.14)	(0.14)	(0.16)	(0.21)	(0.13)	(0.09)
Constant	0.00	0.00	0.00***	-0.00	-0.00	0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	30	30	30	30	29	29
r2	0.17	0.24	0.00	0.38	0.15	0.10

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In order to characterize informed and uninformed traders, we look at the three groups with the highest price impact in the second half of the sample in Panel B of Figure 3.1 and call the traders in those groups 'high-PI traders.' Analogously, we call all traders in the groups from Panel B in Figure 3.1 with negative price impact in both halves of the sample 'neg-PI traders.'

Table 3.5 shows characteristics of neg-PI or high-PI traders depending on whether they belong to the HFT group or not. One can see that for both non-HFT and HFT market participants, informedness is negatively related to trading volume and number of counterparties. On the other hand, the relationship between the number of monthly trades is nonmonotone. For the less active non-HFT market participants, more monthly trades are associated with being uninformed. On the other hand, informed HFT market participants trade more often than their less informed neg-PI counterparts. For both non-HFT and HFT market participants, trading longer maturities is associated with being in the high-PI group, i.e. higher informedness, which may be consistent with these traders seeking the largest exposure.

Table 3.5: **Trader characteristics.** The following table shows properties of neg-PI and high-PI traders depending on whether they belong to the HFT group or not. Notional values are rounded to the nearest hundred thousand. The numbers of counterparties and average monthly trades are rounded to one decimal and the average maturity is rounded to the nearest integer.

	non-HFT		HFT	
	neg-PI	high-PI	neg-PI	high-PI
notional/trade (EUR)	21.4mn	9.6mn	11.9mn	1.4mn
counterparties	4.1	1.2	6.1	3.8
avg. monthly trades	102.9	3.0	706.0	1221.0
average maturity (days)	34	61	62	69

Looking only at high-PI and neg-PI traders, Table 3.6 examines which trader characteristics are associated with being informed. Table 3.6 shows the result of a linear probability regression model to explain a dummy variable that is one if the trader is a high-PI trader and zero otherwise. A trader's number of counterparties, traded notional and the number of monthly trades are negatively related to being a high-PI trader, as one can observe in columns 1 to 3. Also, HFT market participants are more likely to have a lower 1-day price impact, as shown in column 4. Notional volume is not significant anymore in column 5, when controlling for the number of trades and the number of counterparties. One can also see that the number of trades has different implications for the probability of being informed depending on whether the trader belongs to the HFT group or not. The number of counterparties is still negatively related to being informed even when controlling for other factors. This result seems to run against the findings of Kondor and Pintér (2019) that informed traders have more counterparties, but is consistent with information leakage examined empirically in Hendershott and Madhavan (2015b) and Hagströmer and Menkveld (2019) and modeled theoretically in Liu et al. (2018). The cost of information leakage is higher for informed traders, which may be the reason why they contact fewer dealers.

Table 3.6: **Probability of being an informed trader.** This table shows coefficient estimates B and robust standard errors for the regression

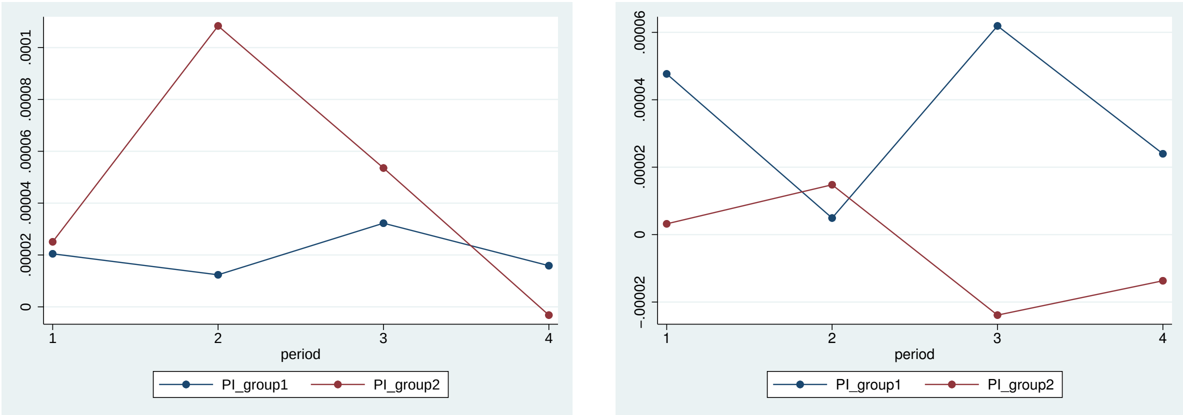
$$informed = BX + \varepsilon,$$

where $informed$ is a dummy variable that is equal to one if a trader belongs to the high-1-day-PI group and equal to zero otherwise. The vector X includes different trader characteristics specified in the table and ε is an error term. The sample includes all high-Pi and neg-PI traders from the groups in Figure 3.1. The averages of the numbers of trades have been divided by 10^3 and averages of the notionals per trade have been divided by 10^{11}.

	(1)	(2)	(3)	(4)	(5)
avg. monthly counterparties (CPs)	-84.70***				-12.07***
	(12.17)				(3.22)
avg. monthly trades		-0.19**			-7.87***
		(0.09)			(0.23)
avg. monthly trades × HFT dummy					7.91***
					(0.23)
avg. notional in EUR			-23.31***		4.82
			(7.95)		(5.99)
HFT dummy				-0.52***	-0.57***
				(0.08)	(0.08)
Constant	1.07***	0.96***	0.96***	0.96***	1.03***
	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
N	9628	9628	9628	9628	9628
r2	0.26	0.03	0.00	0.03	0.80

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To see how robust our findings about the persistence of the price impact of different traders, we break the sample into four different subperiods, and check whether traders who had a high price impact in the first period also have a high price impact in the following periods. To this end, we sort all traders who were active in all four quarters of our sample into two groups based on their number of trades in the first quarter and then compute their average price impact in each quarter of the sample. Figure 3.4 shows that the traders who had a high price impact in the first quarter generally also have a higher price impact in the subsequent quarters. Moreover, if one restricts attention to non-HFT market participants, traders who have a higher price impact in the first quarter also have a higher price impact in the last quarter.



Panel A (all traders). Panel B (non-HFT).

Figure 3.4: **Persistence across multiple subperiods.** We sort all traders who were active in all four quarters of our sample into two groups based on their number of trades in the first quarter and then plot their average 1-day price impact in each quarter of the sample. In Panel A, transactions by all traders were considered and in Panel B, only trades by non-HFT market participants were considered.

3.4 Informed Dealers

Evans and Lyons (2002) find that aggregate orderflow in the interdealer market predicts future price changes in the DM/USD spot market using four months of data from an interdealer trading system in 1996. Most of the variation in forward rates comes from variation in the spot rate. Moreover, order flow in the interdealer market is (in most models, like the one presented in this paper) generated by order flow from customers. Thus, one may hypothesize that customers' order flow predicts future price changes in the EUR/USD forward market. The model presented in Section 3.6 presents a mechanism that would generate such predictability and also suggests that the customer order flow of different dealers has different predictive power for future price changes. Similar to Evans and Lyons (2002), we look at the order imbalance, i.e. the difference between daily buy and sell orders received by a dealer. We define the volume imbalance as the difference between notional volume bought by customers and notional volume sold by customers on a given day. Table 3.7 describes order imbalance and volume imbalance for the different dealer types. One can see that both kinds of dealers have relatively small average order imbalances and small average volume imbalances. However, G16 dealers have much more volatile order imbalances and volume imbalances as indicated by their larger mean absolute order and volume imbalances.

Table 3.7: **Order imbalance (OI).** This table shows the average order imbalance, the average of the absolute values of order imbalances for each type of dealer as well as the respective values for the volume imbalance.

trader type	mean OI	mean absolute OI	mean volume imbalance	mean absolute volume imbalance
BANK	-3	8	-2.6mn	58.1mn
G16	-2	66	3.6mn	809.0mn

In order to measure how well each dealer can forecast future price changes using their customers' order flow, we regress changes in the benchmark price for the one-week forward exchange rate between the end of day t and the end of day $t+1$ on the sum of the order imbalances on the last five days on which the dealer has traded. The higher the resulting R^2 statistic, the more informed is the dealer. As we show in the model in section 3.6.3 the R^2 statistic should also be a determinant of a dealer's markups as it is related to a dealer's expectation of the price impact it will incur.

Table 3.8 shows summary statistics of the R^2 we obtain for every dealer in the sample, provided that enough data is available to perform the regression described above. We also transform the R^2 into an annual Sharpe ratio using the method in Table 1 in Cochrane (1999), assuming the true mean of price rate movements is zero and 250 trading days in a year.[6] One can see in Table 3.8 that even though the R^2 statistics obtained in *daily* forecasts seem small, annualized Sharpe ratios are actually substantial.

[6] Cochrane (1999) derives the formula

$$SR^{informed,\,annual} = \frac{\sqrt{(SR^{uninformed,daily})^2 + R^2}}{\sqrt{1-R^2}}\sqrt{days/year},$$

where $SR^{informed\,annual}$ is the annualized Sharpe ratio, a dealer that has a given R^2 when forecasing daily returns can achieve, given that an investor not forecasting returns can achieve a daily Sharpe ratio of $SR^{uninformed\,daily}$. Under the assumptions in the text, this formula becomes

$$SR^{informed,\,annual} = \frac{\sqrt{R^2}}{\sqrt{1-R^2}}\sqrt{250}.$$

Table 3.8: **Dealer informedness and Sharpe ratios.** This table shows summary statistics for the R^2 statistic from the regression of price changes on the dealers order imbalances described in the text, i.e.

$$\frac{rate_{t+1} - rate_t}{rate_t} = \beta_{0,i} + \beta_{1,i} sum\_OI_i + \varepsilon_{it},,$$

where ε_{it} is an error term, sum_OI_i refers to the sum of the order imbalances of dealer i for its last 5 trading days and $rate_t$ refers to the one-week forward exchange rate at day t.

Variable	Obs	Mean	Std. Dev.	Min	Max
R^2	142	0.022383	0.047305	.000002	0.27981
Sharpe ratio	142	1.72584	1.868606	0.020974	9.855478

In order to study informed and uninformed dealers, we look at trade-by-trade data and assign trades to two quantiles according to the informedness of the dealer trading. We create a dummy variable that is equal to 1, if the dealer informedness is above the median informedness across all trades. For all other trades, the dummy variable is equal to zero. We call dealers for which this dummy variable is one "informed."

Some of the dealers in our sample trade very infrequently. Especially those dealers with extreme Sharpe ratios only have a very limited number of days on which they trade in our sample. Figure 3.12 in Appendix 3.8.1 shows the distributions of the R^2 across dealers. Most dealers have an R^2 of less than 0.025. In order to avoid focussing on outliers when studying the characteristics of informed dealers, we focus on dealers with an R^2 of less than 2% and a notional trading volume of at least 0.5% of the entire market. As an additional robustness check we also look separately at dealers with a notional trading volume greater than 2.5% of the entire market.

In Table 3.9 one can see the characteristics of the high versus low R^2 dealers, depending on the fraction of total D2C volume in EUR they are responsible for. We see that more informed dealers tend to have smaller notionals per trade and have fewer counterparties in D2C and D2D markets.

Table 3.9: **Dealer characteristics.** The following table shows properties informed and uninformed dealers depending on how much notional volume they trade. Percentages are rounded to one decimal, numbers of counterparties are rounded to the nearest integer and notional volumes are rounded to the nearest hundred thousand.

	2.5% > volume > 0.5%		volume > 2.5%	
	uninformed	informed	uninformed	informed
% G16	66.6%	16.7%	100%	100%
dealer's avg. notional/trade	12.1mn	9.2mn	17.5mn	16mn
D2C counterparties	1399	1288	5106	3947
D2D counterparties	123	102	266	220
% of total notional D2C volume (in EUR)	5.2%	8.7%	50%	30.3%

Table 3.10 shows the result of a linear probability regression model , which predicts the informdness dummy based on various dealer characteristics. One can see that both traded notional and the number of trades are highly significant and negatively correlated with being informed. This is true for both subsets of dealers. In the regressions shown in Table 3.25 in Appendix 3.8.1 we use the dealer's R^2 as the left-hand side variable instead of the informedness dummy and get similar results.

Table 3.10: **Probability of being an informed dealer.** We run the regression

$$info\ dummy = BX + \varepsilon,$$

where $info\ dummy$ is the dummy variable described in the text which measures the informedness of a dealer. This table shows the coefficient estimates B for various explanatory variables X as well as robust standard errors. In columns 1 to 3 we focus on dealers executing more than 0.5% of the notional volume (EUR) of the entire D2C market. In columns 4-6 we focus on dealers executing more than 2.5% of the notional volume of the entire D2C market. The number of trades has been divided by 10^7 and the notional traded has been divided by 10^{10}.

	> 0.5% notional volume			> 2.5% notional volume		
	(1)	(2)	(3)	(4)	(5)	(6)
avg. notional traded	-202.75		-333.44*	-149.26		-807.10***
	(151.30)		(189.49)	(258.89)		(92.44)
# trades		-24.36**	-41.15***		-36.34***	-62.32***
		(8.60)	(12.92)		(10.81)	(7.64)
D2D counterparties			0.00			
			(0.00)			
D2C counterparties			0.00			
			(0.00)			
Constant	0.88***	0.90***	1.23***	0.80	1.18***	2.99***
	(0.24)	(0.18)	(0.29)	(0.50)	(0.28)	(0.28)
N	20	20	20	11	11	11
r2	0.06	0.20	0.30	0.02	0.43	0.87

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.8 in Appendix 3.8.1 shows that traders with higher price impact tend to pay higher markups. This suggests that dealers may price discriminate based on client identities or characteristics. Since there is evidence that clients' price impact is persistent, it would be natural to think that dealers set markups based on past client price impact. At the same time, since we have shown that dealers are differentially informed about future price changes based on their clients' aggregated order imbalance, it is also natural to think that dealers will use that information to set markups. To better understand the determinants of markups, we now perform a panel regression, where we explain client markups with various client, dealer, and trade characteristics.

To eliminate data errors, we exclude all trades with markups below -2% and above 3%. This effectively puts a 2% band around the traders' average markups shown in Figure 5 in Hau et al. (2019). In column 1 of Table 3.11, the coefficient for realized values of 1-day price impact is significant at the 5% level. Moreover, we find a significant interaction term with the informedness dummy, which suggests that informed dealers' markups react more strongly to the realized 1-day impact. This is consistent with the implications of our model in section 3.6.3, where dealers that have more informed order-flow are better at predicting future price changes and thus set markups that are more in line with future price changes.

Realized values of price impact measures may be significant in the regression from Table 3.11, because they are correlated with other trader characteristics. In order to control for those, we add fixed effects for each dealer-trader pair along with other time-varying control variables. The results of this regression are shown in column 2 of Table 3.11. The coefficients on the realized 1-day price impact and the interaction term in the first row are very similar to the estimates in column 1 and still significant.

In order to asses how much the results in columns 1 and 2 of Table 3.11 are driven by the dealers' connections in the D2D market, we replace the informed-dealer dummy by a connected-dealer dummy in column 3. To this end, we assign trades into two quantiles according to the number of the dealer's counterparties in the D2D market. The connected-dealer dummy is a dummy variable that is equal to 1, if the dealer's number of D2D counterparties is above the median across all trades. The results of this regression are shown in column 3 of Table 3.11. We find that more connected dealers respond less to changes in future prices, as measured by realized values of 1-day price impact. However, the coefficient in front of the interaction term is not even significant at the 10% level, whereas the coefficient in front of the interaction term in column 2 of Table 3.11 is significant at the 5% level. It thus seems that the informed dealer dummy better captures differences in the dealers' sensitivity with respect to future price changes.

In column 4 of Table 3.11, we study what may drive potential dealer and trader fixed effects on markups. One can see that traders that have on average a higher 1-day price impact also have to pay higher markups. The more counterparties a trader has, the lower the markups possibly due to increased bargaining power.

Table 3.11: **Markups and price impact.** We run the regression

$$markup_{it} = BX + \varepsilon,$$

where ε_{it} is an error term, X are explanatory variables specified in the table and $markup_{it}$ is the markup that trader i has to pay at time t. We report coefficients B and standard errors that are clustered at the dealer level. We excluded all trades with markups below -2% and above 3%. Order imbalance and its standard deviation have been divided by 10^6. Lastly, the logs of counterparties and trades have been divided by 1000.

	(1)	(2)	(3)	(4)
realized 1-day impact × informedness dummy info	0.0112***	0.0099**		0.0104**
	(0.0043)	(0.0043)		(0.0044)
realized 1-day impact × connectedness dummy			-0.0068	
			(0.0042)	
realized 1-day impact	0.0107***	0.0091***	0.0171***	0.0091***
	(0.0012)	(0.0015)	(0.0031)	(0.0015)
realized 1-minute impact		0.0949***	0.0892***	0.0837**
		(0.0330)	(0.0337)	(0.0338)
market conditions:				
volatility		0.7146*	0.7286*	1.0947*
		(0.3640)	(0.3804)	(0.6527)
Smart average 1-day impact group		0.0013	0.0019	0.0035
		(0.0061)	(0.0064)	(0.0073)
time-varying trader characteristics:				
log(traders' monthly counterparties)		-0.0139	-0.0145	-0.0901***
		(0.0212)	(0.0213)	(0.0215)
log(traders' monthly trades)		0.0269***	0.0276***	-0.0319
		(0.0086)	(0.0086)	(0.0238)

(To be continued)

Table 3.11-Continued.

	(1)	(2)	(3)	(4)
time-varying dealer characteristics:				
dealer's signed OI		0.0350	0.0343	0.0326
		(0.0323)	(0.0323)	(0.0339)
fixed trader characteristics:				
trader's average 1-day impact				0.1098***
				(0.0173)
high-PI dummy				0.0002***
				(0.0001)
neg-PI dummy				-0.0001*
				(0.0000)
HFT				-0.0000
				(0.0001)
fixed dealer characteristics:				
informedness dummy	0.0000			-0.0000
	(0.0001)			(0.0001)
connectedness dummy				0.0000
				(0.0001)
standard deviation of dealer's OI				-0.0983*
				(0.0499)
Constant	0.0002***	0.0000	0.0000	0.0004**
	(0.0000)	(0.0001)	(0.0001)	(0.0002)
dealer-client fixed effects	no	yes	yes	no
N	2770512	2742738	2684941	2684941

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To examine how trades by very active traders are different from those trades made by less active market participants, we focus only on HFT market participants in column 1 of Table 3.12. We can see that results related to 1-day price impact are less strong than for the whole sample, but the coefficient on 1-minute price impact is larger. On the other hand, the results associated with 1-day price impact are stronger when focussing on non-HFT market participants. The results are not much affected if we exclude neg-PI traders, or if we distinguish between low- or high-volatility periods. However, comparing columns 5 and 6 in Table 3.12, we see that traders with a high average price impact have to pay even higher markups in high-volatility periods.

Table 3.12: **Markups and price impact in different subsamples** We run the regression

$$markup_{it} = BX + \varepsilon,$$

where ε_{it} is an error term, X are explanatory variables specified in the table and $markup_{it}$ is the markup that trader i has to pay at time t. We report coefficients B and standard errors that are clustered at the dealer level. In column 1, we focus on HFT market participants, and in column 2 we focus on non-HFT market participants. In column 3 we exclude all traders labelled as neg-PI traders. In column 4 we focus only on those traders. In column 5 we focus on the trades for which volatility is below the median across all trades and in column 6 we focus on trades for which volatility is above the median across all trades. We excluded all trades with markups below -2% and above 3%. Order imbalance and its standard deviation have been divided by 10^6.

	(1) HFT	(2) non-HFT	(3) no neg-PI	(4) neg-PI	(5) low vol	(6) high vol
real. 1-day impact × info dummy	0.006	0.012**	0.009**	0.020***	0.011*	0.009***
	(0.004)	(0.005)	(0.004)	(0.007)	(0.006)	(0.003)
realized 1-day impact	0.010***	0.009***	0.009***	0.007***	0.009***	0.009***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
realized 1-minute impact	0.139***	0.054*	0.078***	0.116	0.089***	0.081
	(0.053)	(0.029)	(0.027)	(0.088)	(0.030)	(0.055)

(To be continued)

Table 3.12-Continued.

	(1)	(2)	(3)	(4)	(5)	(6)
	HFT	non-HFT	no low-PI	low-PI	low vol	high vol
market conditions:						
volatility	2.399**	0.348	1.197*	0.284	2.551***	-0.870
	(1.003)	(0.501)	(0.708)	(0.512)	(0.674)	(1.344)
Smart average 1-day impact	0.005	0.002	0.005	-0.004	-0.004	0.011**
	(0.010)	(0.008)	(0.007)	(0.014)	(0.012)	(0.005)
varying trader characteristics:						
log(traders' monthly counterparties)	-0.033**	-0.131***	-0.101***	-0.042**	-0.088***	-0.095***
	(0.016)	(0.035)	(0.023)	(0.019)	(0.022)	(0.023)
log(traders' monthly trades)	0.008	-0.068***	-0.030	-0.018	-0.049**	-0.016
	(0.023)	(0.022)	(0.025)	(0.017)	(0.021)	(0.025)
varying dealer characteristics:						
dealer's signed OI	0.050	0.019	0.037	0.006	0.025	0.040
	(0.043)	(0.026)	(0.036)	(0.024)	(0.025)	(0.043)
fixed trader characteristics:						
trader's average 1-day impact	0.043	0.117***	0.112***	-0.059	0.099***	0.121***
	(0.071)	(0.018)	(0.017)	(0.084)	(0.021)	(0.027)
high-PI dummy	0.000**	0.000**	0.000***	0.000	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(.)	(0.000)	(0.000)
neg-PI dummy	0.000	0.000	0.000	0.000	-0.000	-0.000**
	(0.000)	(0.000)	(.)	(.)	(0.000)	(0.000)
HFT	0.000	0.000	-0.000	-0.000	0.000	-0.000
	(.)	(.)	(0.000)	(0.000)	(0.000)	(0.000)

(To be continued)

Table 3.12-Continued.

	(1) HFT	(2) non-HFT	(3) no low-PI	(4) low-PI	(5) low vol	(6) high vol
fixed dealer characteristics:						
informedness dummy	0.000	-0.000	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
connectedness dummy	0.000***	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
standard deviation of dealer's OI	-0.040	-0.103	-0.114**	-0.017	-0.111**	-0.084*
	(0.040)	(0.069)	(0.055)	(0.048)	(0.054)	(0.048)
Constant	-0.000	0.001***	0.000**	0.000	0.000***	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	905843	1779098	2274265	410676	1347896	1337045

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.26 in Appendix 3.8.1 shows similar regressions as those from Table 3.12, but using deale and trader fixed effects. The coefficient estimates in Table 3.26 are very similar to those shown in Table 3.12.

Figure 3.13 in Appendix 3.8.1 illustrates the findings from the regression with a simple plot that shows how more informed dealers charge higher markups when the price impact is higher.

Appendix 3.8.3 discusses the possible errors-in-variables problem that may arise because both the measured price impact and the measured markup are affected by errors in the benchmark rate. Such errors may arise because of noisy quotes in the TRTH database or imprecise time-stamps in the EMIR database, which would lead to a mechanical positive correlation between markups and realized price impact. As we discuss in the appendix, this bias is unlikely to affect our main results however.

3.5 Endogenous Dealer Choice

The model presented in Section 3.6.4 suggests that informed traders are more likely to trade with dealers who have informative order flow. In order to test this hypothesis, we study how a trader's average 1-day price impact affects the probability of trading with an informed dealer.

As in sections 3.3 and 3.4, HFT market participants arguably behave differently from non-HFT market participants. For this reason, we consider these two groups separately. In Table 3.13 one can see that non-HFT traders are more likely to trade with informed traders if their 1-day price impact is higher on average. Even 1-minute price impact measures positively affect the probability of trading with an informed dealer, as can be seen in column 2. These statements remain true when controlling for other trader characteristics that are associated with being informed (see Section 3.3), like the number of counterparties, or the number of trades. On the other hand, 1-day price impact measures do not affect dealer choice for HFT market participants, since they likely have a much shorter investment horizon. However, for HFT average 1-minute impact still positively affects the probability of trading with an informed dealer. One can also see that non-HFT clients are more likely to trade with informed dealer if volatility is high, while HFT clients are more likely to trade with uninformed dealers if volatility is high.

The results shown in Table 3.13 suggest that traders are more likely to trade with informed dealers if they are more informed on average (i.e. have a higher average 1-day price impact), when adverse selection is higher on average (volatility is higher), or when they are particularly well informed (the realization of the 1-day price impact turns out to be high). To avoid the endogeneity problem that we find that informed traders choose informed dealers, because we define an informed dealer based on the informativeness of her clients' order flow, we now define dealer informedness as the informativeness of the order flow from non-financial companies. This means we perform the same steps as in Section 3.4, including the definition of a dummy variable, but use order flow only from non-financial customers to predict future price changes. We then look at all traders except for non-financial firms and examine the probability of trading with an informed dealer. The results are shown in Table 3.14. We see that non-HFT traders are still more likely to trade with dealers that are informed according to this new measure.

Table 3.13: **Probability of trading with informed dealers.** We run the regression

$$informed = BX + \varepsilon,$$

where ε_{it} is an error term, X are explanatory variables specified in the table and $informed$ is the informed-dummy discussed in Section 3.4. We report coefficients B and standard errors that are clustered at the trader level. The dataset includes D2C trades. We excluded all trades with markups below -2% and above 3% .

	excluding HFT			HFT only		
	(1)	(2)	(3)	(4)	(5)	(6)
avg. 1-day impact	10.83***	10.81***	8.98**	-36.34	-25.24	-3.60
	(3.77)	(3.77)	(3.75)	(71.99)	(70.88)	(67.83)
avg. 1-min impact		148.05*	153.02*		3531.22	4118.06*
		(81.34)	(81.03)		(2180.98)	(2092.95)
realized 1-day impact		0.02	0.03		0.10	0.21
		(0.05)	(0.05)		(0.12)	(0.13)
realized 1-min impact		3.82***	3.88***		-0.92	-1.85
		(0.98)	(0.98)		(1.15)	(1.18)
log(monthly counterparties)			-21.45**			58.11
			(9.58)			(38.63)
log(monthly trades)			-6.75			-87.91***
			(6.63)			(29.49)
volatility			428.48***			-1563.72**
			(131.95)			(750.82)
Constant	0.54***	0.54***	0.56***	0.38***	0.38***	1.01***
	(0.01)	(0.01)	(0.01)	(0.06)	(0.06)	(0.18)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.14: **Probability of trading with informed dealers without circularity.** We run the regression

$$informed^{*} = BX + \varepsilon,$$

where ε_{it} is an error term, X are explanatory variables specified in the table and $informed^{*}$ is the alternative info dummy discussed this section. We report coefficients B and standard errors that are clustered at the trader level. The dataset includes D2C trades. We excluded all trades with markups below -2% and above 3% as well as trades made by non-financial firms.

	excluding HFT			HFT only		
	(1)	(2)	(3)	(4)	(5)	(6)
avg. 1-day impact	10.92**	10.77**	10.01**	221.28**	219.26**	288.62***
	(4.81)	(4.80)	(4.83)	(101.24)	(99.80)	(98.29)
avg. 1-min impact		-53.34	-37.87		-846.62	-244.94
		(98.22)	(97.69)		(3456.66)	(3409.79)
realized 1-day impact		-0.02	-0.01		-0.03	0.09
		(0.05)	(0.05)		(0.19)	(0.20)
realized 1-min impact		1.41	1.37		2.10	1.17
		(1.12)	(1.12)		(3.44)	(3.33)
log(monthly counterparties)			-23.17**			120.16***
			(10.91)			(24.41)
log(monthly trades)			4.07			-118.95***
			(7.53)			(15.32)
volatility			864.03***			-932.14*
			(150.65)			(528.57)
Constant	0.50***	0.50***	0.45***	0.30***	0.30***	1.07***
	(0.01)	(0.01)	(0.02)	(0.06)	(0.06)	(0.13)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.22 in Appendix 3.8.1 shows the 1-day and 1-minute price impact that trades executed by informed and uniformed dealers have. One can see that both the 1-minute price impact and the 1-day price impact are higher for trades executed by informed dealers. One obtains a slightly different result when using other proxies for dealer informedness like connections in the D2C market or connections in the D2D market. Tables

3.23 and 3.24 in Appendix 3.8.1 show that more connected dealers face a higher 1-minute price impact, but a lower 1-day price impact. In Table 3.27, we run regressions similar to those in Table 3.13, except that we replace the informedness dummy with a connectedness dummy. In this case average price impact measures have no significant effect on dealer choice. Thus, the informedness of a dealer cannot simply be captured by the dealers D2D connections.

3.6 Model

The aim of this section is to lay out a mechanism how dealers learn from the orderflow of their customers and incorporate this information in their quotes. For the sake of clarity, other determinants of the dealers' quotes are not modeled explicitly. In section 3.6.4, the model is extended in order to account for informed trading and endogenous dealer choice.

3.6.1 Setup

There are $N \in \mathbb{N}$ dealers, with $N > 2$. The dealers are indexed by the set $\mathcal{D} := \{1, 2, ..., N\}$ and have two periods to trade one asset: In the first period, each dealer $i \in \mathcal{D}$ can trade with a a group of clients. The clients initiate the trade by specifying a quantity they want to trade. The dealer responds with a competitive quote at which all orders of the dealer's clients are executed. In the second period, the N dealers can trade among each other in a centralized market. Each dealer submits linear demand schedules and the price is determined by market clearing.

Let x_i denote the net amount of the asset that clients buy from dealer i.[7] These quantities are jointly normally distributed across dealers with covariance matrix Σ:

$$(x_1, ..., x_N)' \sim \mathcal{N}(0, \Sigma). \tag{3.1}$$

Moreover, let $p_{1,i}$ denote the price at which the transaction between dealer i and the clients happens, let τ_i denote the quantity that dealer i buys from other dealers in period 2 and let p_2 denote the price at which the dealers trade among themselves in period 2.[8]

Then, the utility of dealer $i \in \mathcal{D}$ in the end of period 2 is given by

[7]Here, $x_i < 0$ means that the clients sell to dealer $i \in \mathcal{D}$.
[8]Again, $\tau_i > 0, i \in \mathcal{D}$ means that dealer i buys the asset.

$$U_i(x_i, p_{1,i}, p_2) := -(-x_i + \tau_i)^2 \frac{\gamma}{2} + x_i p_{1,i} - p_2 \tau_i, \tag{3.2}$$

where $\gamma > 0$ determines the aversion of the dealer to holding inventory. The first term in (3.2) represents quadratic inventory holding costs that the dealer needs to pay after period 2. The remaining two terms in (3.2) represent the revenue generated by trading with the clients and the money paid in the interdealer market, respectively. We assume that in period 1, dealers quote a competitive price, p_1, to all their clients such that they are indifferent between trading or not trading in the first round. In period 2, when dealers trade among themselves, dealers maximize the expectation of the expression in (3.2).

3.6.2 Equilibrium

An equilibrium in this model consists of a pricing rule for each dealer that specifies a price for a given quantity that the clients demand as well as a demand schedule for each dealer such that the interdealer market clears in period 2. Moreover, dealers behave competitively when facing clients and maximize expected utility in the interdealer market.

We conjecture that dealer $i \in \mathcal{D}$ uses a linear demand schedule $\tau_i(x_i, p_2) : \mathbb{R}^2 \to \mathbb{R}$ with

$$\tau(x_i, p_2) := a p_2 + b x_i, \tag{3.3}$$

where $a, b \in \mathbb{R}$. Market clearing gives the price in the interdealer market as a function of the quantities traded between dealers and their clients:

$$p_2(x_1, ..., x_N) = \frac{-b}{aN} \sum_{i=1}^{N} x_i.$$

If agent $i \in \mathcal{D}$ is strategic in the interdealer market and conjectures strategies for all the other dealers $j \neq i$ to be as in (3.3), then market clearing implies a residual inverse demand function such that:

$$\lambda := \frac{\partial}{\partial \tau_i} p_2 = \frac{-1}{a(N-1)}. \tag{3.4}$$

Maximizing the expression in (3.2) for a given price p_2 with respect to τ_i and calculating the corresponding the first-order condition gives the necessary condition[9]

[9]By calculating the second derivative, one can see that this condition is also sufficient.

$$\tau_i = \frac{\gamma}{\gamma + \lambda} x_i - \frac{1}{\gamma + \lambda} p_2. \tag{3.5}$$

Now (3.3), (3.4) and (3.5) imply

$$a = -\frac{1}{\gamma - \frac{1}{a(N-1)}} \tag{3.6}$$

and

$$b = \frac{\gamma}{\gamma - \frac{1}{a(N-1)}}. \tag{3.7}$$

Comparing (3.6) and (3.7) gives

$$a = -\frac{b}{\gamma}. \tag{3.8}$$

Using the last result and (3.7) gives

$$b = \frac{N-2}{N-1}. \tag{3.9}$$

It follows that if we define the average client demand $\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$, then we have

$$p_2 = \gamma \bar{x} \tag{3.10}$$

$$\tau_i = \frac{N-2}{N-1}(x_i - \bar{x}) \tag{3.11}$$

We see that the dealer's trade in the inter-dealer market does not depend on her risk-aversion, but only on how different her clients' trades are from the average trade. Instead, the price level in the interdealer market is equal to the average inventory held by dealers times their risk-aversion. The price fully reveals the average inventory which is a sufficient statistic for all the dealers trades, given their own client demand.

To finish the characterization of the equilibrium, it remains to derive the prices that the dealers are offering their clients. Dealers behave competitively and achieve the same expected utility when trading with their clients as in the hypothetical scenario in which they do not trade with the clients (but nevertheless observe the clients' demand and thus make inferences about aggregate orderflow). Fixing a price $p_{1,i}$ and a quantity x_i, taking expectations of the utility defined in (3.2) as well as using the characterizations of the price p_2 and demand schedule τ in (3.3), (3.8) and (3.9), one gets

$$\mathbb{E}\left(U_i^d(x_i, p_{1,i}, p_2, \tau_i) \mid \text{clients demand } x_i\text{, trade happens}\right)$$
$$= \mathbb{E}\left(-(\tau_i - x_i)^2 \tfrac{\gamma}{2} + x_i p_{1,i} - p_2 \tau_i\right)$$

for all $i \in \mathcal{D}$. If dealer i does not trade, but all other dealers do trade with their clients, dealer i expects the market price

$$p_2^{notrade}(x_1, ..., x_N) = \frac{-b}{aN}(\sum_{j=1, j \neq i}^{N} x_j) = \gamma(\bar{x} - \frac{x_i}{N}), \tag{3.12}$$

since dealer i optimally uses the demand schedule

$$\tau^{notrade}(p_2^{notrade}) := ap_2^{notrade} = \frac{N-2}{N-1}(\frac{x_i}{N} - \bar{x}),$$

while all other dealers use the demand schedule given by (3.3). Market clearing now implies (3.12). Analogously to the case in which the dealer trades with the clients, one gets

$$\mathbb{E}\left(U_i^d(x_i, p_{1,i}, p_2, \tau_i) \mid \text{clients demand } x_i\text{, no trade happens}\right)$$
$$= \mathbb{E}\left(-(\tau_i^{notrade})^2 \tfrac{\gamma}{2} - p_2^{notrade} \tau_i^{notrade}\right)$$

Requiring that the utility that the dealer derives from trading and not trading, respectively, are the same and solving for $p_{1,i}$ gives

$$p_{1,i} = \mathbb{E}\left[\gamma \frac{\sum_{j=1}^{N} x_j (N-2) + x_i}{(N-1)N} \mid \text{clients demand } x_i\right]. \tag{3.13}$$

Using the distributional assumption on order flow (3.1) and the conditional expectation of a multivariate normal distribution gives

$$p_{1,i} = \gamma \frac{(N-2)\, \mathbf{1}' \Sigma_{i,i}^{-1} \Sigma_{*,i} + 1}{(N-1)N} x_i, \tag{3.14}$$

where $\mathbf{1}$ is an N-dimensional column vector and $\Sigma_{*,i}$ refers to the i-th column of the covariance matrix Σ. Using the expression $p_2 = -b/(aN) \sum_{j=1}^{N} x_j$, one has

$$\mathbb{E}\left[p_2 \mid \text{customer demands } x_i\right] = \frac{\gamma}{N}\mathbf{1}'\Sigma_{i,i}^{-1}\Sigma_{*,i}x_i. \tag{3.15}$$

Now, (3.14) and (3.15) imply

$$p_{1,i} = \frac{(N-2)}{(N-1)}\mathbb{E}\left[p_2 \mid \text{customer demands } x_i\right] + \gamma\frac{x_i}{(N-1)N}. \tag{3.16}$$

3.6.3 Client markups, dealer informedness, and price impact

One may define the markup as the difference between the quoted price (adjusted for the direction of the trade) and the quote of an uniformed dealer of zero (since $p_2 = 0$ in expectation):

$$m_i := (\chi_{buy} - \chi_{sell})\left(\frac{(N-2)}{(N-1)}\mathbb{E}\left[p_2 \mid \text{net orderflow} = x_i\right] + \gamma\frac{x_i}{N(N-1)}\right), \tag{3.17}$$

where $\chi_{buy} = 1$ if a client buys, $\chi_{buy} = 0$ if a client sells and $\chi_{sell} = 1 - \chi_{buy}$. As the number of dealers gets large, we get

$$m_i \to (\chi_{buy} - \chi_{sell})\mathbb{E}\left[p_2 \mid \text{net orderflow} = x_i\right]. \tag{3.18}$$

A regression of dealer i's markups on the realized second-period prices (adjusted for the direction of the trade), i.e.

$$m_i = \beta_0 + \beta(\chi_{buy} - \chi_{sell})p_2 + \varepsilon, \tag{3.19}$$

is similar to the regression

$$\mathbb{E}\left[p_2 \mid \text{net orderflow} = x_i\right] = \beta_0' + \beta' p_2 + \varepsilon'. \tag{3.20}$$

We get from (3.19) to (3.20) using the limit in (3.18) and dropping the sign of the trade $\chi_{buy} - \chi_{sell}$ from both variables in the regression (3.19). In (3.20) we regress dealer i's conditional expectations of p_2 on the realized values of p_2 and we know that

$$\beta' \xrightarrow{p} R_i^2$$

as the sample becomes large, where R_i^2 is the r-squared statistic from the regression

$$p_2 = \beta_0 + \beta_1 x_i + \varepsilon,$$

i.e. a regression of the price p_2 on dealer i's order imbalance x_i.

If we run the regression described in (3.19), where m_i is replaced by the empirical estimates of the dealer's markups, and $(\chi_{buy} - \chi_{sell})p_2$ refers to the price impact measures, we expect that $\hat{\beta}$ is larger if R_i^2 is larger. We can obtain an estimate of R_i^2 from regressions of future price changes on the dealers' order imbalances.[10]

3.6.4 Endogenous dealer choice

We add two features to the model to capture the endogenous client-dealer choice.

First, we introduce an "arbitrageur" who is endowed with superior information about futures prices. The arbitrageur has access to the interdealer market and can request a quote from one dealer in the first trading round. In the first period, the arbitrageur and other customers of the contacted dealer have to pay the same price.

Second, dealers will have a different reservation utility, when giving quotes to their customers and may refuse to offer a quote if no quote can be found at which they break even.

All other assumptions on dealers and their clients will remain the same as before.

The arbitrageur can send a request for quote (RFQ) to one dealer in the first period, specifying some (exogenously given) quantity $\alpha > 0$ that the arbitrageur wants to trade. The RFQ does not specify whether the arbitrageur wants to buy or sell. Thus, the dealer responds with a bid-ask spread. After the arbitrageur has communicated whether she wants to buy or sell, the dealer charges the same price to the arbitrageur and to the uninformed customers. The uninformed customers are, as before, happy to trade at any price. Conditional on having traded with a dealer in period 1, the arbitrageur offloads the entire quantity α in the interdealer market in period 2.

In the interdealer market, the optimal demand schedule of a dealer who only traded with uninformed customers is still determined by (3.5). If, additionally, dealer i traded with the arbitrageur, then her inventory in period 2 is either $x_i - \alpha$ or $x_i + \alpha$, depending on whether the informed trader bought or sold. Thus, the first order condition for characterizing the maximum of (3.2) using the different initial inventory implies

[10] Since $\mathbb{E}(p_2) = 0$, we get from the law of iterated expectations that $\beta_0' \xrightarrow{p} 0$ in large samples. Using this limit and mutiplying both sides in (3.20) by $\chi_{buy} - \chi_{sell}$ gives $\beta_0 \xrightarrow{p} 0$ and $\beta \xrightarrow{p} \beta'$ in large samples, since $\chi_{buy} - \chi_{sell}$ is uncorrelated with the error in (3.20) (due to symmetry of the normal distribution, signed order flow predicts signed price changes. The sign has no additional predictive power). We thus get $\beta \xrightarrow{p} R_i^2$ in large samples.

$$\tau_i^b := \frac{\gamma(x_i + \alpha) - p_2}{\gamma + \lambda} \tag{3.21}$$

if the arbitrageur bought in period 1 and

$$\tau_i^s := \frac{\gamma(x_i - \alpha) - p_2}{\gamma + \lambda} \tag{3.22}$$

if the arbitrageur sold.

Since the demand of the arbitrageur is price insensitive, the price impact λ is still given by (3.4), where a, and b are defined in (3.8) and (3.9). If the arbitrageur decides to trade with dealer i, the market clearing condition in the interdealer market becomes

$$0 = \begin{cases} \sum_{j\neq i}^{N} \tau_j + \tau_i^b - \alpha & \text{if arbitrageur bought in period 1,} \\ \sum_{j\neq i}^{N} \tau_j + \tau_i^s + \alpha & \text{if arbitrageur sold in period 1.} \end{cases} \tag{3.23}$$

In equation (3.23), the quantity α enters with a negative sign if the arbitrageur bought in period 1, since, in that case, the arbitrageur will sell the same quantity in period 2. Solving (3.23) for p_2, using , (3.4), (3.5), (3.8), (3.9), (3.21) and (3.22) gives

$$p_2^b := \frac{\gamma \sum_{j=1}^{N} x_j}{N} - \frac{\gamma\alpha}{N(N-2)} \tag{3.24}$$

if the arbitrageur bought in period and

$$p_2^s := \frac{\gamma \sum_{j=1}^{N} x_j}{N} + \frac{\gamma\alpha}{N(N-2)} \tag{3.25}$$

if the arbitrageur sold in period 1.

The prices a dealer quotes when contacted by the arbitrageur are determined as follows.

- *ask price:* At the ask, the dealer is indifferent between selling α to the arbitrageur and trading with the uninformed customers and not trading in period 1, assuming that the arbitrageur will the buy α from another dealer, that the uninformed customers will also trade with another dealer and that the price in the interdealer market is *higher* than the ask price.

- *bid price:* At the bid, the dealer is indifferent between buying α from the arbitrageur and trading with the uninformed customers and not trading in period 1, assuming that the arbitrageur will sell α to another dealer, that the uninformed customers will also trade with another dealer, and that the price in the interdealer market is *lower* than the bid price.

Formally, this means

$$0 = \mathbb{E}\left[\left(-(\tau_i^b - x_i - \alpha)^2 \cdot \gamma/2 + (x_i + \alpha) \times ask - p_2^b \tau_i\right) - \left(-(\tau^{notrade})^2 \cdot \gamma/2 - p2^b \tau^{notrade}\right) \mid x_i, p_2 > ask\right]$$

Solving the last equation for ask gives

$$ask = \frac{N}{N+1}\mathbb{E}\left[p_2^b \mid p_2 > ask, x_i\right] + \frac{\gamma x_i}{(N-1)^2}. \tag{3.26}$$

Analogously, one gets

$$bid = \frac{N}{N+1}\mathbb{E}\left[p_2^s \mid p_2 < bid, x_i\right] + \frac{\gamma x_i}{(N-1)^2}. \tag{3.27}$$

For the sake of clarity, we now neglect any influence of the dealer's uninformed orderflow x_i on markups, i.e. we consider the limiting case for $\sigma_i \to 0$ and $\alpha \to 0$. Then the arbitrageur, who knows p_2 will buy from the dealer if and only if

$$p_2 > \frac{N}{N+1}\mathbb{E}\left[p_2 \mid x_i, p_2 > ask\right],$$

where p_2 is again determined as in Section 3.6.2.

Analogously, the arbitrageur will sell the asset to the dealer if and only if

$$p_2 < \frac{N}{N+1}\mathbb{E}\left[p_2^s \mid x_i, p_2 < bid\right].$$

The ask price converges in probability to $\frac{N}{N+1}p_2$ if $(corr(x_i, p_2))^2 \to 1$. Thus, the probability that the arbitrageur buys if $p_2 > 0$ goes to one if $(corr(x_i, p_2))^2 \to 1$.

Analogously, one can show that the probability that the arbitrageur will sell to the dealer goes to one if $p_2 < 0$ and $(corr(x_i, p_2))^2 \to 1$. Since $p_2 \neq 0$ almost surely, the arbitrageur will trade with a probability that is arbitrarily close to 1 if contacting a sufficiently informed dealer.

On the other hand, consider the case in which $corr(x_i, p_2) = 0$. Then, it can be shown that the dealer will charge a fixed positive bid-ask spread with midpoint zero and the arbitrageur will not trade in the (positive-probability) event that p_2 falls into that spread. Proposition 18 formalizes and proves these statements.

3.6.5 Discussion

The model presented above describes a mechanism according to which dealers may include expectations about future prices because the dealers' reservation price, i.e. the price at which dealers are indifferent between trading and not trading, changes.

In actual markets, dealers may not necessarily quote their own reservation price, but may add a markup that depends on the clients bargaining power. The model does not analyse those additional markups. However, in empirical analyses it may be necessary to control for determinants of those markups in order to determine the proper effect of expected price changes on dealers' quotes.

3.7 Conclusion

Extending previous work on informed trading in FX markets, we document heterogeneity in the traders' informedness, as measured by their price impact. We also show that informedness is persistent, i.e. traders with a high price impact in one period are likely to have a high price impact in another period. Moreover, we present evidence suggesting that informed traders are more likely to trade with informed dealers. These findings are consistent with recent theory papers (Lee and Wang (2019)), Glode and Opp (2019)) that argue that non-anonymous OTC markets with a two-tiered market structure with "intermediation chains" may help alleviate asymmetric information frictions. Our findings are also consistent with Evans and Lyons (2002) and suggest that information frictions are also prevalent in a large market such as the Foreign-Exchange market, even though its underlying fundamentals are traditionally largely thought of as driven by systematic macro-economic risks. This has implications for the regulation of financial markets. If OTC markets with their two-tiered structure exist largely due to the rent-seeking behavior of a small set of large dealers who find ways to stymie competition, then regulators should mandate trading on exchanges, which would reduce excessive markups and, by lowering trading costs, lead to more efficient trade.[11] If, on the other hand, OTC markets help alleviate significant financial frictions, then it is less clear that all assets should be traded on anonymous centralized exchanges. The results in this paper offer some support in favor of this second view, and thus suggest that caution may be warranted when regulating financial market structure.

[11] Convincing evidence that channeling trades of smaller less sophisticated traders onto electronic trading platforms indeed leads to less price dispersion and more competitive prices is given in Hau et al. (2019). See also Duffie and Strulovici (2012).

3.8 Appendix

3.8.1 Additional description of the data

Figure 3.5: **Histogram of maturity dates.** This figure shows the frequency with which contracts of the maturities with values on the horizontal axis are traded. The sample period is May 2018 to April 2019. Maturity is expressed in days.

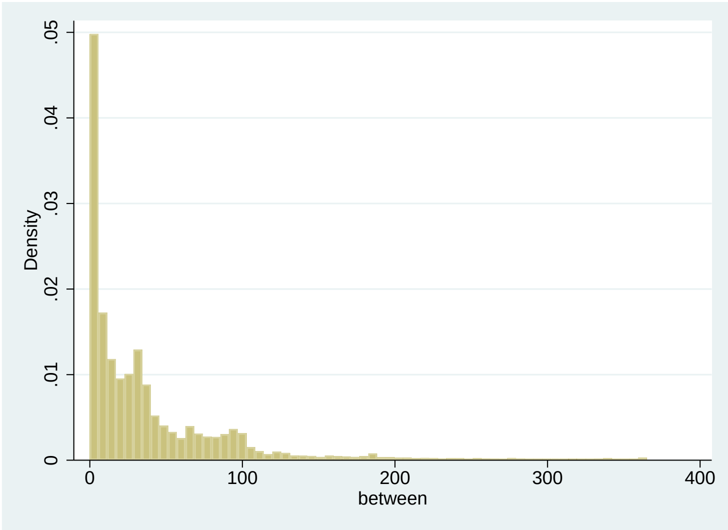


Table 3.15: **Trades in the C2C market.** This table shows the number of trades that are executed between the dealers' customers directly in the sample period from May 2018 to April 2019. Stars signal that values have been omitted due to confidentiality concerns.

	Non-financial	Fund	Ins. & Pen.	Central Bank	Government	Empty
Non-financial	7,115					
Fund	247,269	165,206				
Insurance & Pension	312	991	8			
Central Bank	0	0	0	0		
Government	279	509	7	0	0	
Empty	14,351	125,155	891	98	92	*

Table 3.16: **Average notional in the C2C market.** This table shows the average notional in EUR of trades that are executed between the dealers' customers directly in the sample period from May 2018 to April 2019.

	Non-financial	Fund	Ins. & Pen.	Central Bank	Government	Empty
Non-financial	892,935					
Fund	125,273	3,454,508				
Insurance & Pension	333,659	6,798,043	44,725,209			
Central Bank	0	0	0	0		
Government	11,874,660	4,830,630	10,092,969	0	0	
Empty	6,711,257	573,142	67,393,467	25,702,818	110,600,000	34,234

Table 3.17: **Average notional and number of trades in the D2D market.** This table shows the number of trades and the average notional in EUR of trades that are executed between the dealers' customers directly in the sample period from May 2018 to April 2019. Stars signal that values have been omitted due to confidentiality concerns.

	number of trades		average notional volume	
	G16	Bank	G16	Bank
G16	995,490		98,028,476	
Bank	383,859	*	90,904,190	40,811,433

Figure 3.6: **Markups.** This figure shows the average markups across all traders over the sample period.

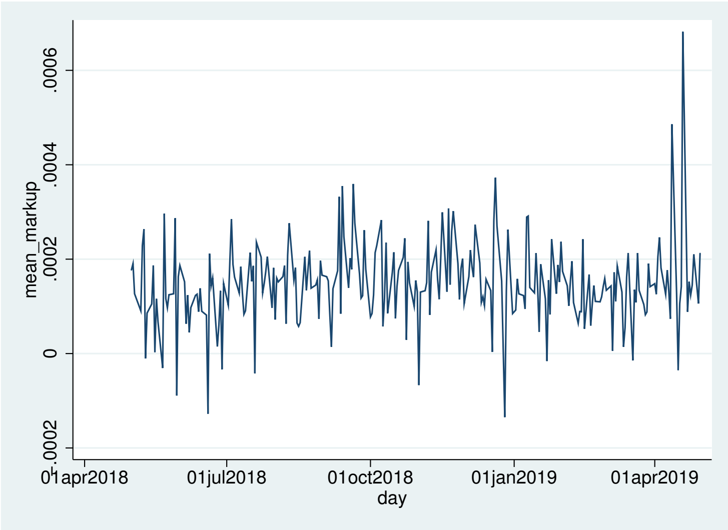
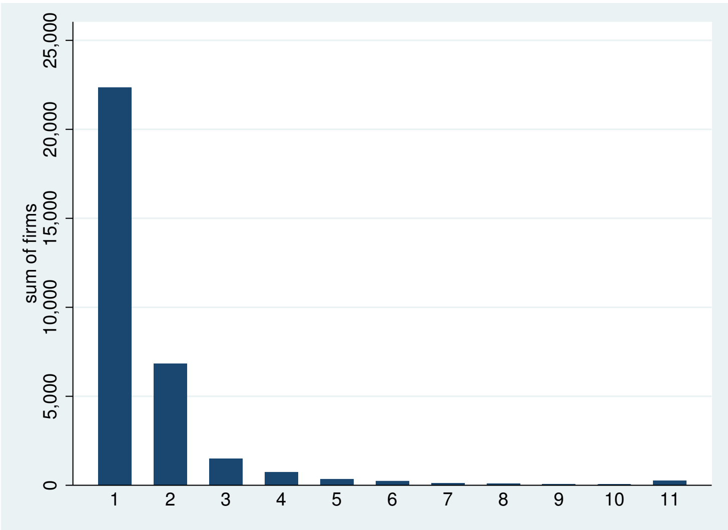
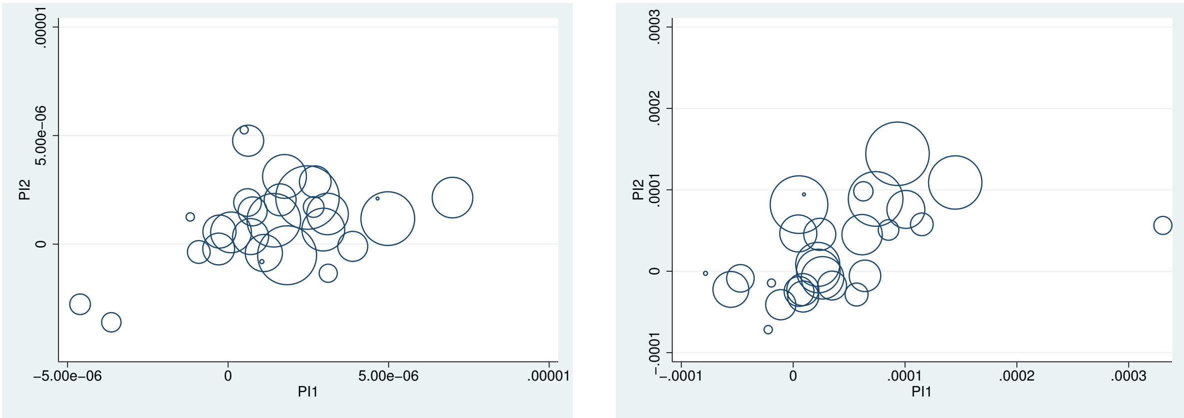


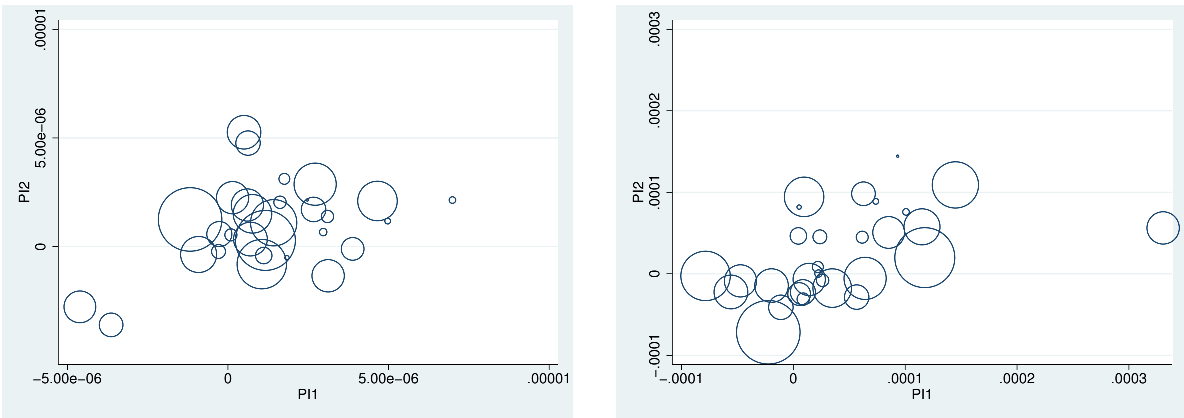
Figure 3.7: **Histogram of average trades per day.** This figure shows the distribution of the trader's average number of trades per day conditional on trading. All traders that trade 12 time or more on average per day when trading have been allocated to the last group, i.e. to the 11-trades group.





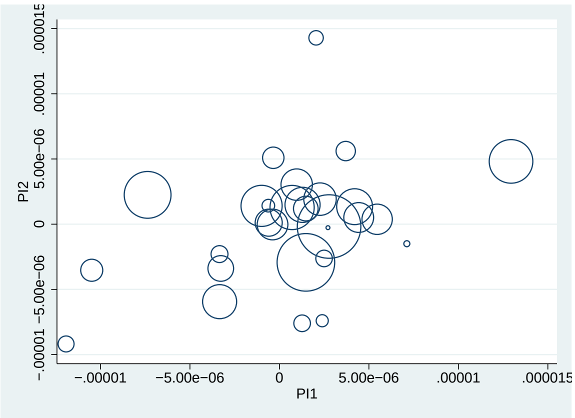
Panel A (1 minute). Panel B (1 day).

Figure 3.8: **Persistence of price impact and the size of markups.** In Panel A, a group's average 1-minute price impact in the second half of the sample is plotted against its 1-minute price impact in the first half of the sample. In Panel B, a group's average 1-day price impact in the second half of the sample is plotted against its 1-day price impact in the first half of the sample. Larger circles correspond to groups in which trades are associated with higher markups on average in the first half of the sample.

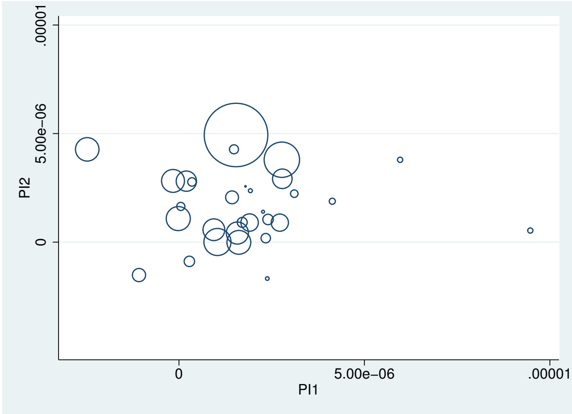


Panel A (1 minute). Panel B (1 day).

Figure 3.9: **Persistence of price impact and the size of total notional traded.** In both panels, a group's average 1-day price impact in the second half of the sample is plotted against its 1-day price impact in the first half of the sample. In Panel A, only HFT market participants are considered and in B, only non-HFT market participants are considered. Larger circles correspond to groups in which traders generate a higher total notional volume (EUR) in the first half of the sample.



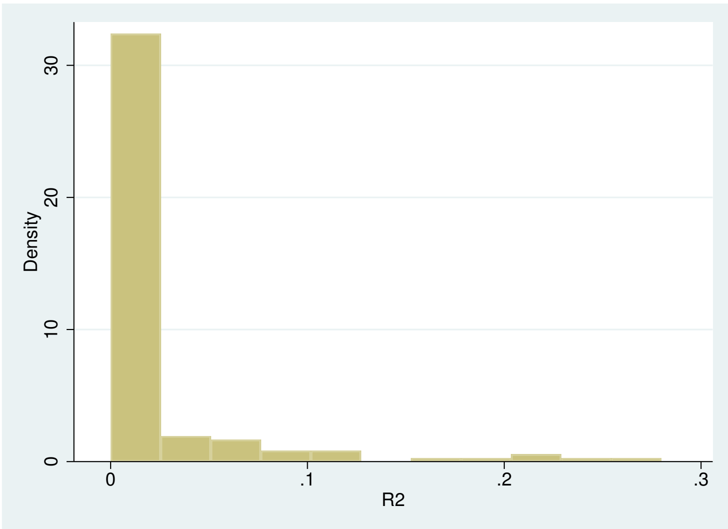
Panel A (HFT).

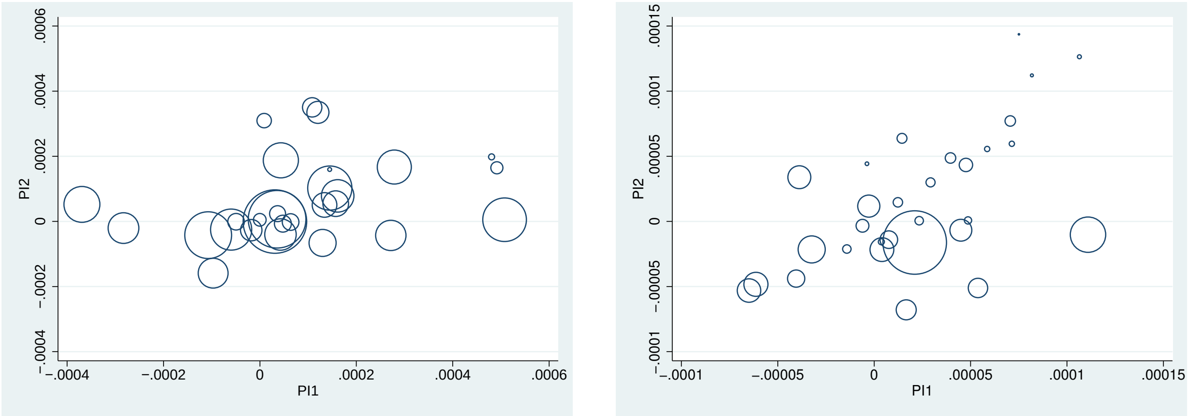


Panel B (non-HFT).

Figure 3.10: **Persistence of 1-minute price impact and the size of total notional traded for HFT and non-HFT market participants.** In both panels, a group's average 1-minute price impact in the second half of the sample is plotted against its 1-minute price impact in the first half of the sample. In Panel A, only HFT market participants are considered and in B, only non-HFT market participants are considered. Larger circles correspond to groups in which traders generate a higher total notional volume (EUR) in the first half of the sample.

Figure 3.12: **Histogram of the dealers' informedness.**



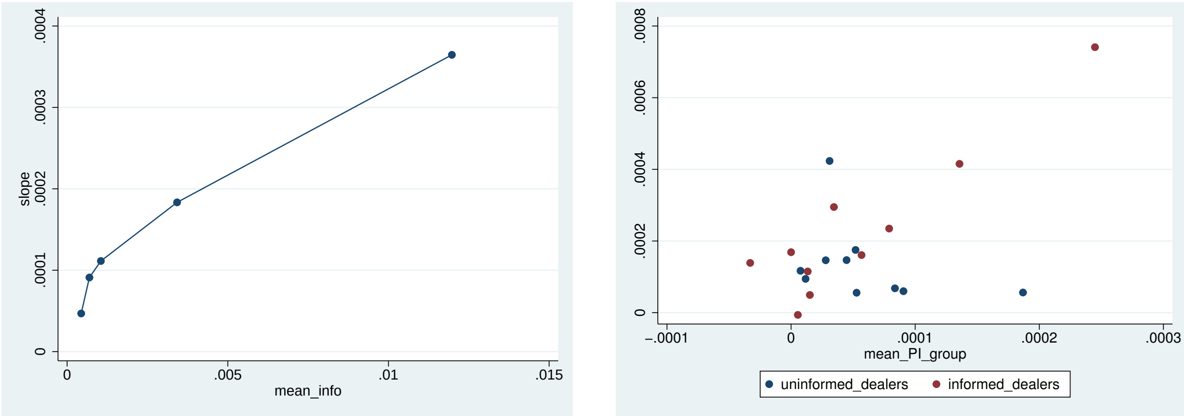


Panel A (HFT). Panel B (non-HFT).

Figure 3.11: **Persistence of 1-day price impact and the size of total notional traded for HFT and non-HFT market participants.** This figure a group's average 1-minute price impact in the second half of the sample against its 1-minute price impact in the first half of the sample. Larger dots correspond to groups in which traders generate a higher notional volume (EUR) per trade in the first half of the sample.

Table 3.18: **Permanent price impact of traders.** This table shows the average changes in the benchmark rates for the 1-minute and the 1-day horizon, the corresponding standard deviations as well as t-stats for each trader type. The standard deviations refer to single observations as opposed to the mean estimates.

	1-day impact			1-minute impact		
trader type	mean	std. dev.	t-stat	mean	std. dev.	t-stat
CENTRAL BANK	-0.0000133	0.0048531	-0.21	0.00000326	0.0002118	1.16
EMPTY	0.0000224	0.0046875	3.39	0.00000106	0.0002322	3.23
FUND	0.0000189	0.0046789	4.99	0.0000011	0.0002296	5.91
GOVERNMENT	-0.00000701	0.004862	-0.24	0.00000564	0.0002611	3.57
INSURANCE & PENSION	-0.0000763	0.0049035	-4.71	0.00000216	0.0002271	2.88
NON-FINANCIAL	0.0001174	0.0048899	19.03	0.00000133	0.0002408	4.38



Panel A. Panel B.

Figure 3.13: **Informedness and sensitivity with respect to price impact.** After having sorted trades first into quintiles based on dealer informedness and then having sorted trades within each quintile into two groups based on 1-day price impact, this figure shows the difference in the average markups of the group with a high 1-day price impact and the group with a low 1-day price impact for the 5 quintiles. The difference in markups is called slope. In Panel B, After having sorted trades first into terciles based on dealer informedness and then having sorted trades within each tercile into 10 groups based number of trades (holding the number roughly constant within each group), this figure shows average markups for each group an the average 1-day price impact of the trades of each group. Red dots (informed_dealers) refer to trades where the dealer informedness is in the highest tercile, while blue dots (uninformed_dealers) refer to trades where the dealer informedness is in the lowest tercile.

Table 3.19: **Price impact of dealers' clients.** This table shows the average clients price impact for the 1-minute and the 1-day horizon as well as the corresponding standard deviations aggregated for each dealer type.

	1-day impact			1-minute impact		
trader type	mean	std. dev.	t-stat	mean	std. dev.	t-stat
BANK	0.0000706	0.004668	12.54	0.0000011	0.0002506	3.74
G16	0.0000278	0.004762	8.45	0.0000013	0.0002267	8.03

Table 3.20: **Markups for traders.** This table shows the average markup for each trader type as well as the corresponding standard deviations. Stars indicate that values are not reported due to confidentiality concerns.

trader type	mean	standard deviation
CENTRAL BANK	*	0.0009116
EMPTY	0.0001678	0.0042364
FUND	0.0000754	0.0020724
GOVERNMENT	0.0001240	0.0014073
INSURANCE & PENSION	0.0000091	0.0015384
NON-FINANCIAL	0.0003284	0.0055631

Table 3.21: **Markups from dealers.** This table shows the average markup for each dealer type as well as the corresponding standard deviations.

trader type	mean	standard deviation
BANK	0.000317	0.004075
G16	0.000092	0.003378

Table 3.22: **Informedness and price impact.** This table shows the average price impact of clients whose dealers have informedness above the median compared to analogous statistics when dealers' informedness is below the median. Informedness is defined as in Section 3.4. Besides average 1-day and 1-minute price impact, the table shows the total number of trades executed by each group of dealers as well as the number of dealers that belongs to each group.

	1-day price impact	1-minute impact	trades	number of dealers
uninformed	0.000036600	0.000000538	1447095	34
informed	0.000041900	0.000001960	1359241	108

Table 3.23: **Connectedness in D2D market and price impact.** This table shows the average price impact of clients whose dealers have more connections in the D2D market than the median compared to analogous statistics when dealers have fewer connections than the median. Besides average 1-day and 1-minute price impact, the table shows the total number of trades executed by each group of dealers as well as the number of dealers that belongs to each group.

	1-day price impact	1-minute impact	trades	number of dealers
unconnected	.000039	0.0000013200	1,508,747	125
connected	.000044	0.0000009440	1,237,597	7

Table 3.24: **Connectedness in D2C market and price impact.** This table shows the average price impact of clients whose dealers have more connections in the D2C market than the median compared to analogous statistics when dealers have fewer connections than the median. Besides average 1-day and 1-minute price impact, the table shows the total number of trades executed by each group of dealers as well as the number of dealers that belongs to each group.

	1-day price impact	1-minute impact	trades	number of dealers
unconnected	0.000048	0.00000117	1,595,402	137
connected	.0000276	0.00000131	1,210,934	5

Table 3.25: **Probability of being an informed dealer.** We run the regression

$$R^2 = BX + \varepsilon,$$

where R^2 is the informedness measure described in the text. This table shows the coefficient estimates B for various explanatory variables X as well as robust standard errors. In columns 1 to 3 we focus on dealers executing more than 0.5% of the notional volume of the entire D2C market. In columns 4-6 we focus on dealers executing more than 2.5% of the notional volume (EUR) of the entire D2C market.

	> 0.5% notional volume			> 2.5% notional volume		
	(1)	(2)	(3)	(4)	(5)	(6)
total notional traded notional	-0.96		0.04	-0.60		-2.90***
	(1.05)		(1.60)	(0.91)		(0.65)
# trades dealer		-0.10**	-0.01		-0.12**	-0.22***
		(0.05)	(0.07)		(0.04)	(0.05)
D2D counterparties			-0.00			
			(0.00)			
D2C counterparties			-0.00			
			(0.00)			
Constant	0.00**	0.00***	0.01**	0.00*	0.00***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	20.00	20.00	20.00	11.00	11.00	11.00

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.26: **Markups and price impact in different subsamples with fixed effects** Using dealer-client fixed effects, we run the regression

$$markup_{it} = BX + \varepsilon,$$

where ε_{it} is an error term, X are explanatory variables specified in the table and $markup_{it}$ is the markup that trader i has to pay at time t. We report coefficients B and standard errors that are clustered at the dealer level. In column 1, we focus on HFT market participans, and in column 2 we focus on non-HFT market participants. In column 3 we exclude all traders labelled as neg-PI traders. In column 4 we focus only on those traders. In column 5 we focus on the trades for which volatility is below the median across all trades and in column 6 we focus on trades for which volatility is above the median across all trades. We excluded all trades with markups below -2% and above 3%. Order imbalance has been divided by 10^6.

	(1)	(2)	(3)	(4)	(5)	(6)
	HFT	non-HFT	no low-PI	low-PI	low vol	high vol
real. 1-day impact × info dummy	0.006	0.011**	0.008**	0.019***	0.008**	0.011*
	(0.004)	(0.005)	(0.004)	(0.007)	(0.003)	(0.006)
realized 1-day impact	0.009***	0.009***	0.009***	0.007***	0.008***	0.009***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
realized 1-minute impact min	0.150***	0.061**	0.083***	0.148*	0.103**	0.085***
	(0.049)	(0.029)	(0.026)	(0.075)	(0.051)	(0.029)
market conditions:						
volatility	1.126	0.444*	0.870**	0.107	0.553	0.685
	(0.699)	(0.235)	(0.409)	(0.434)	(0.723)	(0.545)
smart average 1-day impact	0.004	-0.000	0.003	-0.005	0.007	-0.004
	(0.009)	(0.007)	(0.006)	(0.012)	(0.005)	(0.011)

(To be continued)

Table 3.26-Continued.

	(1) HFT	(2) non-HFT	(3) no low-PI	(4) low-PI	(5) low vol	(6) high vol
varying trader characteristics:						
log(traders' monthly counterparties)	0.014	-0.027**	-0.005	-0.077***	0.002	-0.017
	(0.046)	(0.011)	(0.022)	(0.028)	(0.028)	(0.018)
log(trader's monthly trades)	0.032**	0.023**	0.032***	-0.008	0.011	0.040***
	(0.012)	(0.010)	(0.007)	(0.042)	(0.012)	(0.011)
varying dealer characteristics:						
dealer's signed OI	0.048	0.020	0.038	0.015	0.044	0.026
	(0.041)	(0.024)	(0.034)	(0.026)	(0.040)	(0.025)
Constant	-0.000	0.000***	0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	961931	1780807	2286548	456190	1384874	1357864

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.27: **Probability of trading with connected dealers.** We run the regression

$$connected = BX + \varepsilon,$$

where ε_{it} is an error term, X are explanatory variables specified in the table and *connected* is the the connectedness dummy discussed in Section 3.4. We report coefficients B and standard errors that are clustered at the trader level. The dataset includes D2C trades. We excluded all trades with markups below -2% and above 3%.

	excluding HFT			HFT only		
	(1)	(2)	(3)	(4)	(5)	(6)
avg. 1-day impact	-2.26	-2.11	-3.09	18.18	19.86	-11.90
	(3.57)	(3.56)	(3.41)	(73.43)	(72.51)	(70.97)
avg. 1-min impact		33.42	-29.67		320.21	-53.76
		(77.69)	(73.73)		(2227.70)	(2190.89)
realized 1-day impact		0.03	0.03		-0.09	-0.23*
		(0.05)	(0.05)		(0.13)	(0.12)
realized 1-min impact		-2.38**	-2.24**		-0.28	0.28
		(1.04)	(1.04)		(1.85)	(2.07)
log(counterparties)			68.15***			-69.73
			(9.05)			(50.16)
log(monthly trades)			-48.87***			115.78***
			(6.00)			(44.19)
volatility			-929.89***			65.94
			(120.59)			(597.69)
Constant	0.42***	0.42***	0.58***	0.51***	0.52***	-0.21
	(0.01)	(0.01)	(0.01)	(0.07)	(0.07)	(0.31)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.8.2 Technical details

This appendix contains technical results related to the model extension in section 3.6.4. First, it is shown that there are unique solutions to the fixed-point problems in (3.26) and (3.27). To this end, a result on truncated normal random variables is useful.

Definition 3. *A* truncated normal random variable *with parameters* μ *and* σ *and threshold* a *has the distribution of a normal random variable* Y *with mean* μ *and variance* σ^2 *conditional on* $Y > a$*. Thus,* X *has has the density*

$$\frac{\mathrm{d}\mathbb{P}(\{X < x\})}{\mathrm{d}x} = \begin{cases} \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{1-\Phi\left(\frac{x-\mu}{\sigma}\right)} & \text{if } x \geq a, \\ 0 & \text{if } x < a, \end{cases} \tag{3.28}$$

where ϕ *denotes the density of a standard normal random variable and* Φ *denotes the distribution function of a standard normal random variable.*

Lemma 13. *Let* X *be a truncated normal random variable with parameters* μ*,* σ *and threshold* a*. Let*

$$\kappa := \frac{a-\mu}{\sigma}. \tag{3.29}$$

Then,

$$\mathbb{E}(X) = \mu + \sigma\lambda\left(\kappa\right) \tag{3.30}$$

and

$$\mathbb{V}(X) = \sigma^2\left[1 - \lambda(\kappa)(\lambda(\kappa) - \kappa)\right], \tag{3.31}$$

where λ *denotes the hazard rate function of a standard normal random variable, i.e.*

$$\lambda(x) := \frac{\phi(x)}{1-\Phi(x)}. \tag{3.32}$$

Proof. Using the density in (3.28), the moment generating function of X is given by

$$
\begin{aligned}
M(t) &:= \mathbb{E}\left(e^{tX}\right) \\
&= \frac{1}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)} \frac{1}{\sigma\sqrt{2\pi}} \int_a^\infty e^{ts} e^{-\frac{(s-\mu)^2}{2\sigma^2}}\, \mathrm{d}s \\
&= \frac{1}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)} \frac{1}{\sigma\sqrt{2\pi}} e^{\mu t+\sigma^2 t^2/2} \int_a^\infty e^{-\frac{(s-\mu-\sigma^2 t)^2}{2\sigma^2}}\, \mathrm{d}s \\
&= e^{\mu t+\sigma^2 t^2/2} \frac{1-\Phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}
\end{aligned}
$$

The first two derivatives are given by

$$
M'(t) = (\mu+\sigma^2 t) e^{\mu t+\sigma^2 t^2/2} \frac{1-\Phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)} + e^{\mu t+\sigma^2 t^2/2} \frac{\sigma\phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}
$$

and (using $\phi'(x) = -x\phi(x)$)

$$
\begin{aligned}
M''(t) &= \left(\sigma^2 + (\mu+\sigma^2 t)^2\right) e^{\mu t+\sigma^2 t^2/2} \frac{1-\Phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)} + 2(\mu+\sigma^2 t) e^{\mu t+\sigma^2 t^2/2} \frac{\sigma\phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)} \\
&\quad +\sigma^2 \left(\frac{a-\mu}{\sigma} - \sigma t\right) e^{\mu t+\sigma^2 t^2/2} \frac{\phi\left(\frac{a-\mu}{\sigma}-\sigma t\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}.
\end{aligned}
$$

Evaluating those derivatives at zero gives

$$
\mathbb{E}(X) = M'(0) = \mu + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}
$$

and

$$
\mathbb{E}(X^2) = M''(0) = \mu^2 + \sigma^2 + 2\mu\sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)} + \sigma^2 \left(\frac{a-\mu}{\sigma}\right) \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}.
$$

Using (3.29), (3.32) and $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, one now gets (3.31).

□

The results in Lemma 13 can be used to show the existence and uniqueness of a bid ask spread that a dealer is willing to quote once contacted by an arbitrageur.

Lemma 14. *For any given customer demand x_i and a given covariance matrix Σ of the demanded quantities x_j, $j = 1, ..., N$, there are a unique ask $\in \mathbb{R}$ and a unique bid $\in \mathbb{R}$ such that (3.26) and (3.27) hold.*

Proof. This proof focuses on a solution to (3.26), since the corresponding statement for (3.27) is shown analogously. For a fixed demand of uninformed customers x_i, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(t) := \frac{N}{N+1}\mathbb{E}\left[p_2 | x_i, p_2 > t\right] + \frac{\gamma x_i}{(N-1)^2}$$

such that (3.26) can be rewritten as $f(ask) = ask$. We also have

$$\mathbb{E}\left[p_2 | x_i, p_2 > t\right] = \frac{\int_t^\infty s\rho(s)\,\mathrm{d}s}{\int_t^\infty \rho(s)\,\mathrm{d}s},$$

where ρ is the density of p_2 with respect to the Lebesgue measure of p_2 conditional on the realization of x_2. Due to (3.24), the joint normality and analogously to the reasoning in the text before (3.15), this means (by the normal projection theorem) that μ is the density of a normal random variable with mean

$$\overline{\mu} := \frac{\gamma}{N}\mathbf{1}'\Sigma_{i,i}^{-1}\Sigma_{*,i}x_i - \frac{\gamma\alpha}{N(N-2)} \tag{3.33}$$

and variance

$$\overline{\sigma}^2 := \frac{\gamma^2}{N^2}\left[\mathbf{1}'\Sigma\mathbf{1} - \frac{\left(\mathbf{1}'\Sigma_{*,i}\right)^2}{\Sigma_{i,i}}\right]. \tag{3.34}$$

By (3.30) in Lemma 13, we have

$$\mathbb{E}\left(p_2 | x_i, p_2 > K\right) = \overline{\mu} + \overline{\sigma}\lambda\left(\frac{K - \overline{\mu}}{\overline{\sigma}}\right). \tag{3.35}$$

By the Banach Fixed Point Theorem, there is a unique $ask \in \mathbb{R}$ with $ask = f(ask)$ if $|f'(x)| \leq M$ with $0 < M < 1$ for all $x \in \mathbb{R}$. The latter is shown in two steps:

1. The hazard rate function λ ist strictly increasing, i.e. $\lambda' > 0$.

2. The hazard rate function λ satisfies $\lambda' < 1$

Then, $\lambda'(x) \in (0, 1)$ implies $f'(x) \in (0, \frac{N}{N+1})$. Thus, the claim follows by the Banach Fixed-Point Theorem.

To prove the first step, differentiate the hazard rate function. Using the facts that $\phi'(x) = -x\phi(x)$ (which can be verified by simple computation) and $\Phi'(x) = \phi(x)$ (which holds by definition), one gets

$$\begin{aligned}\lambda'(x) &= \frac{\phi'(x)(1-\Phi(x)) + \Phi'(x)\phi(x)}{(1-\Phi'(x))^2} \\ &= -x\frac{\phi(x)}{1-\Phi(x)} + \left(\frac{\phi(x)}{1-\Phi(x)}\right)^2 \\ &= \lambda(x)(\lambda(x)-x). \end{aligned} \tag{3.36}$$

Applying the formula (3.30) for the expectation of truncated normal random variables to a standard normal random variable X, one gets

$$\lambda(x) = \mathbb{E}(X|X > x) > x.$$

Since by (3.32) one has $\lambda > 0$, it now follows from (3.36) that $\lambda'(x) > 0$ for all $x \in \mathbb{R}$.

To prove the second step, note that the expression on the left-hand side of (3.31) must be positive for any $\kappa \in \mathbb{R}$. This implies that

$$\lambda(x)(\lambda(x) - x) < 1$$

for all $x \in \mathbb{R}$. Now it follows with (3.36) that $\lambda'(x) < 1$ for $x \in \mathbb{R}$.

□

In order to characterize the bid ask spread for different levels of dealer informedness, the following two auxiliary results are useful.

Lemma 15. *The expression $\sigma\lambda\left(\frac{x}{\sigma}\right)$, where λ is given by (3.32), is strictly monotone increasing in σ for all $x \in \mathbb{R}$ and all $\sigma > 0$.*

Proof. Let $x \in \mathbb{R}$ and $0 < \sigma_1 < \sigma_2$. As shown in the proof of Lemma 14, one has $0 < \lambda(y) > y$ and $\lambda'(y) < 1$ for all $y \in \mathbb{R}$. From these two fact it follows that the line $l : \left[\, 0, \frac{x}{\sigma_1} \,\right) \to \mathbb{R}$ with

$$l(y) := y\frac{\lambda(x/\sigma_1)}{x/\sigma_1}$$

lies strictly below λ on its domain. Using $y = x/\sigma_2$, this implies

$$\frac{\sigma_1}{\sigma_2}\lambda(x/\sigma_1) < \lambda(x/\sigma_2).$$

Multiplying both sides by σ_2 gives

$$\sigma_1\lambda\left(\frac{x}{\sigma_1}\right) < \sigma_2\lambda\left(\frac{x}{\sigma_2}\right).$$

□

We can now prove the claim that the arbitrageur is almost always able to trade with the dealer if the dealer is sufficiently informed. Moreover, the following results states the resulting bid-ask spreads for both informed and uninformed dealers and provides an expression for the probability with which an arbitrageur is able to trade with a completely uninformed dealer given that the dealer's order flow from uninformed clients is negligible.

Proposition 18. *Let*

$$R^2 := \frac{\left(\mathbf{1}'\Sigma_{*,i}\right)^2}{\Sigma_{i,i}\mathbf{1}'\Sigma\mathbf{1}}.$$

be the squared correlation between dealer i's order flow from uninformed customers and the future price p_2*, let* $\Sigma_{i,i} \to 0$ *and let* $\alpha \to 0$*. Then,*

- *As* $R^2 \to 1$*, both the ask (bid) of dealer i converges in probability to* $\frac{N}{N+1}p_2$ *if* $p_2 > 0$ *(*$p_2 < 0$*) and to zero otherwise and the probability that the arbitrageur trades with the dealer goes to 1.*
- *If* $R^2 = 0$*, one has*

$$\frac{ask}{\gamma\sqrt{\mathbf{1}'\Sigma\mathbf{1}}/N} = S$$

and

$$\frac{bid}{\gamma\sqrt{\mathbf{1}'\Sigma\mathbf{1}}/N} = -S,$$

where S satisfies the fixed-point equation

$$S := \frac{N}{N+1}\lambda(S),$$

where λ is given by (3.32). In this case, the probability that the arbitrageur trades with the dealer is strictly below 1.

In the following, the first and second bullet point are proved separately. Due to symmetry, only the expressions for the limit of the ask price will be derived explicitly.

Step1: Proof of the first bullet point.

With the ask price given by (3.26) and using (3.33), (3.34) and (3.35) from the proof of Lemma 14, one has

$$\begin{aligned} ask &\xrightarrow{p} \frac{N}{N+1}\mathbb{E}\left[p_2 \mid p_2 > ask, x_i\right] \\ &= \frac{N}{N+1}\left(\overline{\mu} + \overline{\sigma}\lambda\left(\frac{ask - \overline{\mu}}{\overline{\sigma}}\right)\right) \end{aligned} \tag{3.37}$$

as $\Sigma_{i,i} \to 0$. This convergence can be shown using Chebyshev's inequality. In the following, the limit in (3.37) will be examined for $R^2 \to 1$. First, $R^2 \to 1$ and (3.34) imply

$$\overline{\sigma} \to 0. \tag{3.38}$$

Moreover, for $\overline{\sigma} > 0$, it must always be the case that $ask > \frac{N}{N+1}\overline{\mu}$. Suppose that is not the case. Then one would get the contradiction

$$\frac{N}{N+1}\overline{\mu} \geq ask = \frac{N}{N+1}\left(\overline{\mu} + \overline{\sigma}\lambda\left(\frac{ask - \overline{\mu}}{\overline{\sigma}}\right)\right) > \frac{N}{N+1}\overline{\mu}.$$

Since, for any fixed $\overline{\mu} > 0$, the ask prices are bounded from below and are, by Lemma 15, strictly increasing in $\overline{\sigma}$, the prices must converge as $\overline{\sigma} \to 0$, holding $\overline{\mu}$ fixed.

Note that $0 < \lambda'(x) < 1$ implies that

$$\max\{\lambda(x) - x | x \geq 0\} = \lambda(0) = \sqrt{\frac{2}{\pi}}. \tag{3.39}$$

Suppose now, that $\lim_{\overline{\sigma}\to 0} ask \geq \overline{\mu}$. Then, using (3.39), one would obtain the contradiction

$$\lim_{\overline{\sigma}\to 0} ask = \lim_{\overline{\sigma}\to 0} \frac{N}{N+1}\left(\overline{\mu} + \overline{\sigma}\lambda\left(\frac{ask - \overline{\mu}}{\overline{\sigma}}\right)\right) \leq \lim_{\overline{\sigma}\to 0} \frac{N}{N+1}\left(\overline{\mu} + \overline{\sigma}\sqrt{\frac{2}{\pi}} + (ask - \overline{\mu})\right) = \frac{N}{N+1}\lim_{\overline{\sigma}\to 0} ask.$$

Thus,

$$\lim_{\overline{\sigma}\to 0} ask < \mu. \tag{3.40}$$

Since for any $x < \overline{\mu}$ and a standard normal random variable X, one has

$$\lim_{\overline{\sigma}\to 0} \lambda\left(\frac{x-\overline{\mu}}{\overline{\sigma}}\right) = \lim_{b\to -\infty} \mathbb{E}(X|X>b) = 0,$$

it must be the case that

$$\lim_{\overline{\sigma}\to 0} ask = \lim_{\overline{\sigma}\to 0} \frac{N}{N+1}\left(\overline{\mu} + \overline{\sigma}\lambda\left(\frac{ask-\overline{\mu}}{\overline{\sigma}}\right)\right) = \frac{N}{N+1}\overline{\mu}. \tag{3.41}$$

Suppose now that $\overline{\mu} < 0$. Because $ask - \overline{\mu} > -\frac{1}{N+1}\mu$ for any $\overline{\sigma} > 0$ due to the above reasoning, one has

$$\lim_{\overline{\sigma}\to 0} \overline{\sigma}\lambda\left(\frac{ask-\overline{\mu}}{\overline{\sigma}}\right) \to ask - \mu$$

because of (3.39). If now follows that the ask price has to satisfy

$$\lim_{\overline{\sigma}\to 0} ask = \frac{N}{N+1}(\overline{\mu} + \lim_{\overline{\sigma}\to 0} ask - \overline{\mu}),$$

i.e. $\lim_{\overline{\sigma}\to 0} ask = 0$.

Since $\overline{\mu} \xrightarrow{p} p_2$ as $\overline{\sigma} \to 0$ and $\alpha \to 0$, where p_2 is determined as in Section 3.6.2, the probability that $ask < p_2$ if $p_2 > 0$ goes to 1 as $\overline{\sigma} \to 1$. Analogously, the probability that $bid > p_2$ if $p_2 < 0$ goes to 1 as $\overline{\sigma} \to 1$. To conclude, the probability that the arbitrageur trades with the dealer goes to 1 as $\overline{\sigma} \to 1$, since $p_2 > 0$ or $p_2 < 0$ with probability 1.

Step2: Proof of the second bullet point.

As above, the limit in (3.37) is examined for $R^2 = 0$. The last two conditions and (3.34) imply

$$\overline{\sigma} \to \infty. \tag{3.42}$$

Moreover, since $R^2 = 0$, one always has $\overline{\mu} = 0$ as can be seen from (3.33), since $\mathbf{1}'\Sigma_{i,i}^{-1}\Sigma_{*,i} = 0$ in this case. Thus, (3.37) becomes

$$ask \xrightarrow{p} \frac{N}{N+1}\left(\overline{\sigma}\lambda\left(\frac{ask-\overline{\mu}}{\overline{\sigma}}\right)\right) \tag{3.43}$$

Instead of looking at the behavior of the ask prices directly, we consider the expression

$$K := \frac{ask}{\overline{\sigma}}.$$

If the ask price satisfies (3.43), then K satisfies

$$K = \frac{N}{N+1}\lambda(K). \tag{3.44}$$

Since $0 < \lambda'(x) < 1$ for all $x \in \mathbb{R}$, a solution K to (3.44) exists and is unique, which can be shown using the Banach Fixed-Point Theorem. Moreover, $K > 0$, since $\lambda(x) > 0$ for all $x \in \mathbb{R}$.

The probability that $p_2 > ask$ is given by $1-\Phi(\frac{K}{\overline{\sigma}})$, where K satisfies (3.44) and Φ denotes the distribution function of a standard normal random variable. Since $K > 0$, this probability is less than $\frac{1}{2}$. Analogously, the probability that $p_2 < bid$ is less than $\frac{1}{2}$. Thus, the arbitrageur will trade with the uninformed dealer with a probability of less than 1. One obtains the expressions for the bid and ask prices in the statement of the proposition by using (3.34) to express $\overline{\sigma}$ in terms of the primitive parameters.

3.8.3 The effect of measurement errors in the FX benchmark price on regression coefficient estimates

While most data are generally affected by measurement errors, a potential measurement error in the benchmark price (e.g. due to noisy quotes from dealers in the TRTH database or imprecise timestamps in the EMIR database) may pose a special problem since the benchmark price is used for both calculating price impact and markups. This appendix has three goals:

1. It is explored how errors in the benchmark price of the transactions affect the coefficient estimates in a regression of markups on price impact.

2. An argument is presented that, given the empirical results in the main text, the estimate of the impact of 1-day price impact on markups has, under plausible conditions, an upward bias of no more than 1 % of the coefficient estimate.

3. It is shown that even if errors in the benchmark price are large, this effect does, under plausible conditions, not affect estimates on how differently informed dealers respond to markups.

A potential bias in coefficient estimates: To address the first of the above points, let t be the reported time of a transaction that is observed in the dataset and let T be the point in time a fixed horizon after the reported transaction (e.g. $T - t$ is equal to one minute, one day, etc). Consider now the following five random variables:

$$m_t^o := \text{ observed FX benchmark price at time } t,$$

$$m_t^* := \text{ actual FX benchmark price at time } t$$

$$m_T^o := \text{ observed FX benchmark price at time } T,$$

$$m_T^* := \text{ actual FX benchmark price at time } T,$$

$$P := \text{price paid for the contract,}$$

where the price of the contract is not affected by measurement errors. Moreover, let $sign(trade) := 1$ if the trader buys and $sign(trade) := -1$ if the trader sells. With the definitions given above, we can define the observed price impact, PI^o, the true price impact, PI^*, the observed markup, MU^o, and the true markup, MU^* as follows.

$$PI^o := sign(trade)(m_T^o - m_t^o),$$

$$PI^* := sign(trade)(m_T^* - m_t^*),$$

$$MU^o := sign(trade)(P - m_t^o),$$

$$MU^* := sign(trade)(P - m_t^*).$$

Using these definitions, we can express the observed variables in terms of the corresponding actual variables:

$$PI^o = PI^* + \underbrace{sign(trade)(m_t^* - m_t^o)}_{=:\varepsilon_1} + \underbrace{sign(trade)(m_T^o - m_T^*)}_{=:\varepsilon_2}, \tag{3.45}$$

$$MU^o = MU^* + \underbrace{sign(trade)(m_t^* - m_t^o)}_{=\varepsilon_1}, \tag{3.46}$$

where we also defined the error terms ε_1 and ε_2 that affect the observations of price impact and markup. Consider now the univariate regression model

$$MU^* = \beta_0 + \beta_1 PI^* + \varepsilon, \tag{3.47}$$

where ε is an error term and $\beta_1 = \frac{Cov(MU^*, PI^*)}{Var(PI^*)}$. We can estimate β_1 by replacing actual values by observed values, but then, by (3.45) and (3.46), our estimate $\hat{\beta}_1$ becomes (in a large sample)

$$\hat{\beta}_1 = \frac{Cov(MU^o, PI^o)}{Var(PI^o)} = \frac{Cov(MU^* + \varepsilon_1, PI^* + \varepsilon_1 + \varepsilon_2)}{Var(PI^* + \varepsilon_1 + \varepsilon_2)}, \tag{3.48}$$

which is not necessarily equal to the true β_1.

Quantifying the bias: In the following, an upper bound for the bias derived above is stated for the case in which errors are uncorrelated with the true markup. It is also assumed that errors are uncorrelated with each other as well as with the true price impact. Let $\hat{\beta}_1^{minute}$ and $\hat{\beta}_1^{day}$ denote the estimates of the regression coefficient β_1 from (3.47), where PI^o stands for the observed 1-minute price impact or for the 1-day price impact, respectively. Note that the error ε_1 that affects those estimates is the same for the case with the 1-day price impact as for the case with the 1-minute price impact. The coefficient estimates $\hat{\beta}_1$ for various horizons are shown in Table 3.28. As for the regressions in Section 3.4, trades by CCPs and central banks as well as trades with extreme markups have been excluded from the regressions shown in Table 3.28.

Table 3.28: **Markups and price impact.** This table shows coefficient estimates for an OLS regression of markups on realized values of the price impact for various horizons. Standard errors are clustered on the dealer level and shown in parentheses. Trades by central banks, CCPs as well as trades with markups smaller than 2% or greater than 3% have been excluded.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	all	non-HFT	HFT	all	all	all	all	all
1-min impact	0.081**	0.043	0.141***				0.085***	0.086**
	(0.032)	(0.029)	(0.049)				(0.032)	(0.034)
30-min impact				0.035***				0.026***
				(0.008)				(0.008)
1-day impact					0.016***		0.016***	0.006***
					(0.002)		(0.002)	(0.002)
5-day impact						0.013***		0.011***
						(0.002)		(0.001)
Constant	0.000***	0.000***	0.000**	0.000***	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The estimates

If $Cov(MU^*, PI^*) \geq 0$, we get from (3.48) and and the correlations assumed above that

$$Var(\varepsilon_1) \leq \hat{\beta}_1^{minute} Var(PI^o). \tag{3.49}$$

Looking at Table 3.28, we have $\hat{\beta}_1^{minute} \approx 0.1.$ and $\hat{\beta}_1^{day} \approx 0.01$. Using standard deviations of the observed 1-minute price impact, $PI^{o,minute}$, and 1-day priceimpact, $PI^{o,day}$ from Tables 3.18 and 3.19, we get $Var(PI^{o,minute}) \approx (0.0002)^2$ and $Var(PI^{o,day}) \approx (0.005)^2$. Using these approximations, we get from (3.49) that

$$Var(\varepsilon_1) \lessapprox 4 \cdot 10^{-10}.$$

We also get

$$Cov(PI^{o,day}, MU^o) = \hat{\beta}_1^{day} \cdot Var(PI^{o,day}) \approx 3 \cdot 10^{-7}.$$

and therefore, considering (3.48), that

$$\begin{aligned}
\frac{\hat{\beta}_1^{day}}{\beta_1^{day}} &= \frac{Cov(M^o, PI^{o,day})}{Cov(M^o, PI^{o,day}) - Var(\varepsilon_1)} \frac{Var(PI^{o,day}) - Var(\varepsilon_1) - Var(\varepsilon_2)}{Var(PI^{o,day})} \\
&\leq \frac{Cov(M^o, PI^{o,day})}{Cov(M^o, PI^{o,day}) - Var(\varepsilon_1)} \\
&\approx \frac{3 \cdot 10^{-7}}{3 \cdot 10^{-7} - 4 \cdot 10^{-10}}.
\end{aligned}$$

The potential upward bias can therefore only be in the order of magnitude of 0.1% of the original coefficient estimate.

One can see in Table 3.28 that the relationship between price impact and markups is very robust with respect to the horizon. Moreover, comparing model 1, model 3 and model 5, the coefficient estimates barely change, when including more variables. If the observations would be heavily affected by a common measurement error, we would expect that estimates would change more, since the independent variables would be heavily correlated. Only after including the 5-day price impact (model 6), the coefficient estimate for the 1-day price impact changes, which is plausibly explained by the correlation between 1-day and 5-day price changes.

Moreover, while 1-minute price impact is significantly related to the markups for non-HFT client, the same is not true for HFT clients. Unsless the trades of HFT clients have on average a different error, this suggests that the strong result for non-HFT clients is not driven by the error.

No upward bias for the estimate for the interaction term: Consider the linear regression model

$$MU^* = \beta_0 + \beta_1 PI^* + \varepsilon + \beta_2 PI^* \mathbf{1}_{informed} + \varepsilon, \tag{3.50}$$

whith the same interpretion as (3.47) and $\mathbf{1}_{informed} = 1$ if the trade happens with a dealer that is classified as informed and $\mathbf{1}_{informed} = 0$ otherwise. As the sample becomes large, one has

$$\hat{\beta}_2 = \hat{\beta}_1^{informed} - \hat{\beta}_1^{uninformed},$$

where $\hat{\beta}_1^{informed}$ and $\hat{\beta}_1^{uninformed}$ are the estimates of β_1 from running the regression in (3.47) for informed dealers only and uninformed dealers only, respectively. Using the results from above and assuming that $Var(\varepsilon_1)$ as well as the variance of the price impact do not change across informed and uninformed dealers, we have

$$\hat{\beta}_2 = \frac{Cov(MI^*, PI^*|\text{ informed}) - Cov(MI^*, PI^*|\text{ uninformed})}{Var(PI^o)}$$

as the sample becomes large, whereas the true coefficient satisfies

$$\beta_2 = \frac{Cov(MI^*, PI^*|\text{ informed}) - Cov(MI^*, PI^*|\text{ uninformed})}{Var(PI^*)}.$$

Under the assumptions stated above, $Var(PI^o) \geq Var(PI^*)$. Thus, the positive estimate of β_2 may only be biased towards zero but not upwards if the estimate is positive.

Bibliography

Abad, J., I. Aldasoro, C. Aymanns, M. D'Errico, L. Fache, Rousova, P. Hoffmann, S. Langfield, M. Neychev, and T. Roukny (2016). Shedding light on dark markets: First insights from the new eu-wide otc derivatives dataset. *ECB working paper*.

Acharya, V. and A. Bisin (2014). Counterparty risk externality: Centralized versus over-the-counter markets. *Journal of Economic Theory 149*, 153–182.

Andrei, D. and J. Cujean (2017). Information Percolation, Momentum, and Reversal. *Journal of Financial Economics 123*, 617–645.

Babus, A. and P. Kondor (2016). Trading and information diffusion in over-the-counter markets. *working paper*.

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

Babus, A. and C. Parlatore (2016). Strategic fragmented markets.

Babus, A. and C. Parlatore (2017). Strategic Fragmented Markets. *working paper*.

Basak, S. and A. M. Buffa (2017). A theory of model sophistication and operational risk. Technical report.

Benos, E., R. Payne, and M. Vasios (2018). Centralized trading, transparency, and interest rate swap market liquidity: Evidence from the implementation of the dodd-frank act. *Journal of Financial and Quantitative Analysis*.

Bessembinder, H. and W. Maxwell (2008). Transparency and the corporate bond market. *Journal of Economic Perspectives 22*(2), 217–234.

Biais, B. (1993). Price formation and equilibrium liquidity in fragmented and centralized markets. *The Journal of Finance 48*(1), 157–185.

Biais, B., T. Foucault, and F. Salanié (1998). Floors, dealer markets and limit order markets. *Journal of Financial Markets 1*(3-4), 253–284.

Biais, B. and R. C. Green (2007). The microstructure of the bond market in the 20th century. *Tepper School of Business*, 134.

Bjønnes, G., N. Kathitziotis, and C. Osler (2017). Bid-ask spreads in otc markets. *working paper.*

Bjønnes, G., N. Kathitziotis, and C. L. Osler (2016). Bid-ask spreads in otc markets. *working paper.*

Bjonnes, G., C. Osler, and N. Kathitziotis (2017). Bid-ask spreads in otc markets. Technical report.

Bjønnes, G., C. L. Osler, and D. Rime (2008). Asymmetric information in the interbank foreign exchange market. *working paper.*

Burdett, K. and K. L. Judd (1983a). Equilibrium price dispersion. *Econometrica: Journal of the Econometric Society*, 955–969.

Burdett, K. and K. L. Judd (1983b). Equilibrium Price Dispersion. *Econometrica 51*(4), 955–969.

Cochrane, J. H. (1999). Portfolio advice in a multifactor world. *Feceral Reserve Bank of Chicago - Economic Perspectives 23.*

Collin-Dufresne, P., B. Junge, and A. B. Trolle (2017). Market Structure and Transaction Costs of Index CDSs. *working paper.*

Collin-Dufresne, P., B. Junge, and A. B. Trolle (2018). Market structure and transaction costs of index cdss.

De Frutos, M. and C. Manzano (2002). Risk Aversion, Transparency, and Market Performance. *The Journal of Finance 57*(2), 959–984.

De Frutos, M. Á. and C. Manzano (2002). Risk aversion, transparency, and market performance. *The Journal of Finance 57*(2), 959–984.

Dennert, J. (1993). Price competition between market makers. *The Review of Economic Studies 60*(3), 735–751.

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

Diamond, D. W. and R. E. Verrecchia (1981). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics 9*(3), 221–235.

Diamond, P. A. (1971). A Model of Price Adjustment. *Journal of Economic Theory 3*, 156–168.

Drucker, J. and B. V. Voris (2015, September 11). Wall Street Banks to Settle CDS Lawsuit for $1.87 Billion. *Bloomberg*.

Du, S. and H. Zhu (2017). Bilateral Trading in Divisible Double Auctions. *Journal of Economic Theory 167*, 285–311.

Duffie, D., P. Dworczak, and H. Zhu (2016). Benchmarks in Search Markets. *The Journal of Finance*, forthcoming.

Duffie, D., P. Dworczak, and H. Zhu (2017). Benchmarks in search markets. *The Journal of Finance 72*(5), 1983–2044.

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

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

Duffie, D., S. Malamud, and G. Manso (2009). Information Percolation with Equilibrium Search Dynamics. *Econometrica 77*(5), 1513–1574.

Duffie, D., S. Malamud, and G. Manso (2014). Information Percolation with Equilibrium Search Dynamics. *Journal of Economic Theory 153*, 1–32.

Duffie, D. and G. Manso (2007). Information Percolation in Large Markets. *The American Economic Review 97*(2), 203–209.

Duffie, D. and B. Strulovici (2012). Valuation in Over-the-Counter Markets. *Econometrica 80*(6), 2469—-2509.

Duffie, D. and H. Zhu (2011). Does a central clearing counterparty reduce counterparty risk? *The Review of Asset Pricing Studies 1*(1), 74–95.

Edwards, A. K., L. E. Harris, and M. S. Piwowar (2007). Corporate Bond Market Transaction Costs and Transparency. *The Journal of Finance 62*(3), 1421–1451.

Evans, M. and R. Lyons (2002). Order flow and exchange rate dynamics. *Journal of Political Economy 110*, 170–180.

Evans, M. D. D. and R. K. Lyons (2005). Exchange rate fundamentals and order flow. *NBER working paper.*

Feldhütter, P. (2005). The Same Bond at Different Prices: Identifying Search Frictions and Selling Pressures. *The Review of Financial Studies 25*(4), 1155–1206.

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

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

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

Glebkin, S. (2016). Three essays in financial economics. *PhD thesis, London School of Economics.*

Glode, V. and C. Opp (2016). Asymmetric information and intermediation chains. *American Economic Review 106*, 2699–2721.

Glode, V. and C. Opp (2017a). Over-the-Counter vs. Limit-Order Markets: The Role of Traders' Expertise. *working paper.*

Glode, V. and C. Opp (2019). On the Efficiency of Long Intermediation Chains. *Journal of Financial Intermediation 38*, 11–18.

Glode, V. and C. C. Opp (2017b). Over-the-counter vs. limit-order markets: The role of traders' expertise.

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *The Journal of Finance 49*(4), 1127–1161.

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.

Goldstein, M. A., E. S. Hotchkiss, and E. R. Sirri (2007). Transparency and Liquidity: A Controlled Experiment on Corporate Bonds. *The Review of Financial Studies 20*(2), 235–273.

Green, R. C., B. Hollifield, and N. Schürhoff (2007). Financial Intermediation and the Costs of Trading in an Opaque Market. *The Review of Financial Studies 20*(2), 275–314.

Grossman, S. J. (1992). The informational role of upstairs and downstairs trading. *Journal of Business*, 509–528.

Grossman, S. J. and J. E. Stiglitz (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review 70*(3), 393–408.

Hagströmer, B. and A. J. Menkveld (2016). A Network Map of Information Percolation. *working paper.*

Hagströmer, B. and A. J. Menkveld (2019). Information revelation in decentralized markets. *Journal of Finance*, forthcoming.

Han, S. and K. Nikolaou (2016). Trading relationships in the otc market for secured claims: Evidence from triparty repos.

Harris, L. E. and M. S. Piwowar (2006). Secondary Trading Costs in the Municipal Bond Market. *The Journal of Finance, 61*(3), 1361–1397.

Hau, H., P. Hoffmann, S. Langfield, and Y. Timmer (2017). Discriminatory pricing of over-the-counter derivatives.

Hau, H., P. Hoffmann, S. Langfield, and Y. Timmer (2019). Discriminatory Pricing of Over-The-Counter Derivatives. *working paper.*

Hellwig, M. F. (1980). On the aggregation of information in competitive markets. *Journal of Economic Theory 22*(3), 477–498.

Hendershott, T. and A. Madhavan (2015a). Click or call? auction versus search in the over-the-counter market. *The Journal of Finance 70*(1), 419–447.

Hendershott, T. and A. Madhavan (2015b). Click or Call? Auction versus Search in the Over-the-Counter Market. *The Journal of Finance 70*(1), 419–447.

Hendershott, T. and H. Mendelson (2000). Crossing networks and dealer markets: Competition and performance. *The Journal of Finance 55*(5), 2071–2115.

Hilscher, J., J. M. Pollet, and M. Wilson (2015). Are credit default swaps a sideshow? evidence that information flows from equity to cds markets. *Journal of Financial and Quantitative Analysis 50*(3), 543–567.

Janssen, M., P. Pichler, and S. Weidenholzer (2011). Oligopolistic markets with sequential search and production cost uncertainty. *The RAND Journal of Economics 42*(3), 444–470.

Janssen, M. C., J. L. Moraga-González, and M. R. Wildenbeest (2005). Truly costly sequential search and oligopolistic pricing. *International Journal of Industrial Organization 23*(5-6), 451–466.

Janssen, M. C., J. L. Moraga-González, and M. R. Wildenbeest (2005). Truly costly sequential search and oligopolistic pricing. *International Journal of Industrial Organization 23*, 451–466.

Janssen, M. C., J. L. Moraga-González, and M. R. Wildenbeest (2011). Oligopolistic markets with sequential search and production cost uncertainty. *The RAND Journal of Economics 42*(3), 444–470.

Jovanovic, B. and A. J. Menkveld (2014). Dispersion and skewness of bid prices.

Jovanovic, B. and A. J. Menkveld (2015). Dispersion and Skewness of Bid Prices. *working paper.*

Kim, O. and R. E. Verrecchia (1991). Market reaction to anticipated announcements. *Econometrica 30*(2), 273–309.

Kondor, P. and G. Pintér (2019). Clients' connections: Measuring the role of private information in decentralised markets. *working paper.*

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1336.

Kyle, A. S. (1989). Informed Speculation with Imperfect Competition. *The Review of Economic Studies 56*(3), 317–355.

Kyle, A. S., A. A. Obizhaeva, and Y. Wang (2017). Smooth Trading with Overconfidence and Market Power . *The Review of Economic Studies*, 1–52.

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

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

Lagos, R., G. Rocheteau, and P.-O. Weill (2011a). Crises and liquidity in over-the-counter markets. *Journal of Economic Theory 146*(6), 2169–2205.

Lagos, R., G. Rocheteau, and P.-O. Weill (2011b). Crises and liquidity in over-the-counter markets. *Journal of Economic Theory 146*, 2169–2205.

Lee, T. and C. Wang (2018). Why trade over-the-counter? when investors want price discrimination.

Lee, T. and C. Wang (2019). Why trade over-the-counter? when investors want price discrimination. *working paper*.

Lester, B., G. Rocheteau, and P.-O. Weill (2015a). Competing for order flow in otc markets. *Journal of Money, Credit and Banking 47*(S2), 77–126.

Lester, B., G. Rocheteau, and P.-O. Weill (2015b). Competing for Order Flow in OTC Markets. *Journal of Money, Credit and Banking 47*(2), 77–126.

Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones (2017). Screening and adverse selection in frictional markets. *working paper*.

Levin, D. and J. L. Smith (1994). Equilibrium in auctions with entry. *The American Economic Review*, 585–599.

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

Liu, Y., S. Vogel, and Y. Zhang (2018). Electronic trading in otc markets vs. centralized exchange. *working paper*.

Malamud, S. and M. Rostek (2014). Decentralized Exchage,. *The American Economic Review 2017*, forthcoming.

Malamud, S. and M. Rostek (2017). Decentralized exchange. *American Economic Review 107*(11), 3320–62.

Marsh, A. and J. Detrixhe (2015, April 15). Fixed-Income Investors Have 99 Ways to Trade and One Big Problem. *Bloomberg*.

McAfee, R. P. and J. McMillan (1987). Auctions with entry. *Economics Letters 23*(4), 343–347.

McLannahan, B. and J. Rennison (2015, September 11). Wall Street banks to settle CDS lawsuit for $1.9bn. *Financial Times*.

Menezes, F. M. and P. K. Monteiro (2000). Auctions with endogenous participation. *Review of Economic Design 5*(1), 71–89.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf (2017). Information Flows in Foreign Exchange Markets: Dissecting Customer Currency Trades. *The Journal of Finance 71*(2), 601–634.

Naik, N. Y., A. Neuberger, and S. Viswanathan (1999). Trade Disclosure Regulation in Markets with Negotiated Trades. *The Review of Financial Studies 12*(4), 873–900.

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

Oehmke, M. and A. Zawadowski (2016). The anatomy of the cds market. *The Review of Financial Studies 30*(1), 80–119.

Östberg, P. and T. Richter (2018). The sovereign debt crisis: Rebalancing or freezes? *Swiss Finance Institute Research Paper* (17-32).

Pagano, M. (1989). Trading volume and asset liquidity. *The Quarterly Journal of Economics 104*(2), 255–274.

Pagano, M. and A. Röell (1996). Transparency and Liquidity: A Comparison of Auction and Dealer Markets with Informed Trading. *The Journal of Finance 51*(2), 579–611.

Pagnotta, E. and T. Philippon (2011). Competing on Speed. *working paper.*

Parlour, C. A. and D. J. Seppi (2003). Liquidity-based competition for order flow. *The Review of Financial Studies 16*(2), 301–343.

Philippon, T. and E. Pagnotta (2011). Competing on speed.

Ranaldo, A. and F. Somogyi (2018). Heterogeneous information content of global fx trading. *working paper.*

Riggs, L., E. Onur, D. Reiffen, and H. Zhu (2018). Swap trading after dodd-frank: Evidence from index cds.

Rodgers, K. (2017). *Why aren't they shouting?* London, UK: Random House UK.

Sambalaibat, B. (2018). Endogenous specialization and dealer networks. *working paper.*

Sannikov, Y. and A. Skrzypacz (2016). Dynamic trading: Price inertia and front-running. *working paper.*

Schneider, M., F. Lillo, and L. Pelizzon (2018). Modelling illiquidity spillovers with hawkes processes: an application to the sovereign bond market. *Quantitative Finance 18*(2), 283–293.

Seppi, D. J. (1990a). Equilibrium block trading and asymmetric information. *the Journal of Finance 45*(1), 73–94.

Seppi, D. J. (1990b). Equilibrium Block Trading and Asymmetric Information. *The Journal of Finance 45*(1), 73–94.

Stafford, P. (2016, October 5). ICAP's future reflects derivatives market in transition . *Financial Times*.

Stahl, D. O. (1989a). Oligopolistic pricing with sequential consumer search. *The American Economic Review*, 700–712.

Stahl, D. O. (1989b). Oligopolistic Pricing with Sequential Consumer Search. *The American Economic Review 79*(4), 700–712.

Sun, Y. (2006a). The exact law of large numbers via fubini extension and characterization of insurable risks. *Journal of Economic Theory 126*(1), 31–69.

Sun, Y. (2006b). The exact law of large numbers via Fubini extension and characterization of insurable risks. *Journal of Economic Theory 126*, 31–69.

Varian, H. R. (1980a). A Model of Sales. *The American Economic Review 70*(4), 651–659.

Varian, H. R. (1980b). A model of sales. *The American Economic Review 70*(4), 651–659.

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

Viswanathan, S. and J. J. Wang (2002). Market architecture: limit-order books versus dealership markets. *Journal of Financial Markets 5*(2), 127–167.

Vives, X. (2010). *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton, New Jersey: Princeton University Press.

Wang, C. (2017). Core-Periphery Trading Networks. *working paper*.

Weill, P.-O. (2007a). Leaning against the wind. *The Review of Economic Studies 74*(4), 1329–1354.

Weill, P.-O. (2007b). Leaning Against the Wind. *The Review of Economic Studies 74*, 1329–1354.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, 641–654.

White, A. (2016, April 28). Markit and ISDA Seek to Settle EU Antitrust Probe on CDS. *Bloomberg*.

White, A. and S. Bodoni (2015, December 4). Banks Escape Antitrust Charges as EU Shuts Down Swaps Probe. *Bloomberg*.

Yin, X. (2005a). A Comparison of Centralized and Fragmented Markets with Costly Search. *The Journal of Finance 60*(3), 1567–1590.

Yin, X. (2005b). A comparison of centralized and fragmented markets with costly search. *The Journal of Finance 60*(3), 1567–1590.

Yueshen, B. (2017). Uncertain market making.

Zhou, B. Y. (2017). Uncertain Market Making. *working paper*.

Zhu, H. (2012a). Finding a good price in opaque over-the-counter markets. *The Review of Financial Studies 25*(4), 1255–1285.

Zhu, H. (2012b). Finding a Good Price in Opaque Over-the-Counter Markets. *Review of Financial Studies 25*(4), 1255–1285.

Zhu, H. (2014). Do dark pools harm price discovery? *The Review of Financial Studies 27*(3), 747–789.

Sebastian Vogel

Contact Information

EPFL CDM SFI
EXTRA 127 (Extranef UNIL)
Quartier UNIL-Dorigny
CH-1015 Lausanne

Cell: +41765235857
E-mail: sebastian.vogel@epfl.ch
Homepage: sites.google.com/site/svogelresearch/

Research Interests

OTC Markets, Market Design, Market Microstructure, Benchmark Design, Dynamic Asset Pricing, Empirical Asset Pricing

Education

EPFL and Swiss Finance Institute, Lausanne, Switzerland
Ph.D. in Finance — *2015-2020 (expected)*
- Advisor: Prof. Pierre Collin-Dufresne

Stockholm School of Economics, Stockholm, Sweden
M.Sc. in Finance with Specialization in Investment Management — *2013-2015*

University of Hagen, Hagen, Germany
M.Sc. in Mathematics — *2017-2019*
B.Sc. in Mathematics — *2014-2017*

University of Halle-Wittenberg, Halle and der Saale, Germany
B.Sc. in Economics and Management — *2010-2013*
- Minor: Philosophy

Honors and Awards

EMIR Bridge Programme travel grant by the European Central Bank, 2019

AFA Doctoral Student Travel Grant, 2019

Swiss Finance Institute, Graduate Student Fellowship, 2015-2016

Deutschlandstipendium, 2012-2013

Academic Experience

European Central Bank, Frankfurt, Germany
EMIR Bridge Programme — *April 2019 - present*

École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Teaching Assistant — *2016 - present*
- Credit Risk (Master in Financial Engineering), 2016-2018
- Investments (Master in Financial Engineering), 2019

Stanford Graduate School of Business, Stanford, CA
Visiting Student Researcher — *2018*
- Host: Prof. Darrell Duffie

University of Halle-Wittenberg, Halle and der Saale, Germany
Research and Teaching Assistant in Theoretical Philosophy — *2012 - 2013*

Working papers

When to Introduce Electronic Trading Platforms in Over-the-Counter Markets?
(3rd-round revise and resubmit at the Journal of Finance)
I study a hybrid over-the-counter (OTC) market structure in which traders have the choice of obtaining an asset from dealers either in a bilateral market or on an electronic trading platform. In a hybrid market (HM), turnover is higher and traders are better off than in a pure bilateral market (PBM). I present sufficient conditions under which dealer profits are higher in the HM than in the PBM and vice versa. Regulators can improve welfare by mandating electronic trading if search costs on the platform are low. Whether search costs are sufficiently low could be tested, using the model implications.

Electronic Trading in OTC Markets vs. Centralized Exchange (with Ying Liu and Yuan Zhang)
We model a two-tiered market structure in which an investor can trade an asset on a trading platform with a set of dealers who in turn have access to an interdealer market. The investor's order is informative about the asset's payoff and dealers who were contacted by the investor use this information in the interdealer market. Increasing the number of contacted dealers lowers markups through competition but increases the dealers' costs of providing the asset through information leakage. We then compare a centralized market in which investors can trade among themselves in a central limit order book to a market in which investors have to use the electronic platform to trade the asset. With imperfect competition among dealers, investor welfare is higher in the centralized market if private values are strongly dispersed or if the mass of investors is large.

Informed Traders and Dealers in the FX Forward Market (with Pierre Collin-Dufresne and Peter Hoffmann)
There is strong heterogeneity in the permanent price impact of traders. Moreover, a trader's permanent price impact is persistent. A trade's ex-post permanent price impact is partially priced in dealers' markups, even when controlling for dealer-client fixed effects. This suggests that dealers are informed about the permanent price impact of their clients' trades. We present further evidence suggesting that dealers learn from their clients' order flow and use this knowledge when providing quotes. More informed customers are more likely to trade with informed dealers. We present a model explaining our empirical findings.

Conference Presentations

2019:
SGF Conference, Zurich, Switzerland; European Finance Association Doctoral Tutorial, Carcavelos, Portugal.

2018:
AFA PhD Poster Session, Philadelphia, PA; SGF Conference, Zurich, Switzerland; SFI Research Days, Gerzensee, Switzerland; China International Conference in Finance, Tianjin, China; European Finance Association, Warsaw, Poland; Northern Finance Association, Charlevoix, Quebec; Paris December International Finance Meeting, Paris, France

2017:
Quantitative Finance Workshop, Milan, Italy; SFI Research Days, Gerzensee, Sitzerland; Finance Theory Group Contributed Papers Session, St. Louis, Missouri; USI Summer School on Market Microstructure, Lugano, Switzerland; Northern Finance Association, Halifax, Nova Scotia; Paris December International Finance Meeting, Paris, France.

2016:
Market Microstructure: Confronting many viewpoints (poster session), Paris, France and *SFI PhD Workshop (poster session), Zurich, Switzerland.*

2015:
Seventh Annual Meeting of the Academy of Behavioral Finance, Philadelphia, PA.

Ce document a été imprimé au Centre d’impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.