Simultaneous Multilateral Search*

Sergei Glebkin† Bart Zhou Yueshen‡ Ji Shen§

This version: January 7, 2021

* This paper benefits tremendously from discussions with Jean-Edouard Colliard, Jérôme Dugast, Uday Rajan, Ioanid Roşu, Sebastian Vogel, Zhuo Zhong, and Haoxiang Zhu. In addition, comments and feedback are greatly appreciated from participants at conferences and seminars at LSE and INSEAD. There are no competing financial interests that might be perceived to influence the analysis, the discussion, and/or the results of this article.

† INSEAD; glebkin@insead.edu; Boulevard de Constance, Fontainebleau 773000, France.

‡ INSEAD; b@yueshen.me; 1 Ayer Rajah Avenue, Singapore 138676.

§ Peking University; jishen@gsm.pku.edu.cn; No. 38 Xueyuan Road, Haidian District, Beijing 100871, China.

Simultaneous Multilateral Search

Abstract

This paper studies simultaneous multilateral search (SMS) in over-the-counter (OTC) markets: when searching, a customer contacts several dealers and then trades with the one offering the best quote. Search intensity (how frequently one can search) and search capacity (how many potential counterparties one can contact) affect market qualities differently. Contrasting SMS to bilateral bargaining (BB), the model shows that customers might favor BB over SMS when in distress. Such preference for BB might be inefficient for welfare and suggests an intrinsic hindrance in the adoption of request-for-quote type of electronic trading in OTC markets.

Keywords: request-for-quote, over-the-counter market, search, bargaining

(There are no competing financial interests that might be perceived to influence the analysis, the discussion, and/or the results of this article.)

1 Introduction

Search is a key feature in over-the-counter (OTC) markets. Duffie, Gârleanu, and Pedersen (2005, hereafter DGP) pioneered the theoretical study of OTC markets in a framework of random matching and *bilateral bargaining* (BB): Investors search for counterparties and are randomly matched over time. Upon successful matching between a buyer and a seller, the pair engage in Nash bargaining and split the trading gain according to their endowed bargaining power.

However, investors' interaction is not always bilateral. For example, in recent years, there is a rise of electronic trading in OTC markets, mainly in the form of Request-for-Quote (RFQ). In such marketplaces, where many corporate bonds and derivatives are traded, customers contact multiple dealers for quotes and then trade with the one offering the best price. Hendershott and Madhavan (2015) report that more than 10% of trades in the $8tn corporate bond market is completed via RFQ. O'Hara and Zhou (2020) document a continued growth of RFQ-based trading of corporate bonds, but the growth has been sluggish, with the highest trading volume share below 14% in their sample. See also Bessembinder, Spatt, and Venkataraman (2020) for an extensive review on OTC market structure.

This paper develops a theoretical model, tailoring to the above one-to-many searching. Specifically, a customer is allowed to query *multiple* dealers *at the same time*, hence the name "Simultaneous Multilateral Search" (SMS). The objective is two-fold. First, the model aims at understanding the equilibrium features of SMS: How do dealers quote when contacted? How does the quality of the SMS technology affect welfare? Second, contrasting SMS to BB, the paper studies how customers choose to search: Do they favor SMS over BB? Which is more efficient in terms of welfare? How to understand the sluggish growth of SMS-type of electronic trading (O'Hara and Zhou, 2020)? What are the policy and market design implications?

Section 2 sets up the model following Hugonnier, Lester, and Weill (2020, hereafter HLW). A continuum of customers trade an asset through a continuum of dealers. All agents have inventory constraints and can hold either zero or one unit of asset. The customers are subject to stochastic valuation shocks. Those who hold the asset but have low valuation want to sell, while those without the asset but with high valuation want to buy. They actively search for dealers according to independent Poisson processes with intensity ρ. Upon searching, each customer randomly contacts n dealers, who make take-it-or-leave-it offers to the searching customer. In addition, with probability q, the customer can commit to a reserve price, i.e., indicating a worst price she is willing to accept from the dealers. Effectively, a searching customer runs a first-price auction, probabilistically with a reserve price, among n randomly selected dealers.[1]

Importantly, the n randomly contacted dealers might not be the right counterparty for the searching customer, and they may be unable to quote. For example, for a searching buyer, not all n contacted dealers will have the asset to sell—their inventories might be empty. Section 3 characterizes the equilibrium, where the (expected) response rate to a searching buyer (high-valuation non-owner) is the proportion of dealers who have inventory; and vice versa. Such an endogenous response rate is a unique feature and yields novel results in the random matching framework à la DGP and HLW. For example, if the equilibrium response rate is high (in expectation), competition among the contacted dealers becomes fierce, allowing the customers to acquire a larger share of the trading gain. In this sense, SMS endogenizes the bargaining powers, which are by and large exogenous in existing search models.

The search and matching of SMS is characterized by two parameters, the intensity ρ (how frequently one can search) and the capacity n (how many potential counterparties one can contact). Perhaps surprisingly, Section 3.4 finds that the two have contrasting implications for various

[1] The purpose of introducing the parameter q is two-fold. First, it allows the model to capture various forms of SMS in terms of how likely customers can "negotiate" trading prices with dealers. When $q = 0$, the platform represents a typical RFQ platform like MarketAxess, where customers can only receive quotes from dealers but not set prices (O'Hara and Zhou, 2020). Instead, when trading is less formally organized, q can be larger. For example, a bid-wanted-in-competition (BWIC) auction to sell collateralized loan obligations (CLOs) is conducted by emails, through which the selling customers might communicate their indicative reserve prices with dealers (Hendershott et al., 2020). Second, it allows SMS to nest BB as a special case of $n = 1$, under which the customers effectively have Nash bargaining power q and the dealers $(1 - q)$. See Section 3.2 for details.

equilibrium objects. For example, a higher ρ always pushes the equilibrium asset allocation toward the Walrasian level, reducing the sizes of both the buyer- and the seller-customers, thus improving welfare. In contrast, larger n can drive up the size of the short-side customers, i.e., away from the efficient Walrasian level, possibly *reducing welfare*.

The key mechanism is a "bottleneck" effect, which arises from how the search capacity n *asymmetrically* affects the matching on the two sides of the market. To see this, suppose 90% of the dealers have inventory and the other 10% do not. Let us examine what happens when the capacity increases from $n = 2$ to $n = 3$: For a customer-seller, the matching rate with a no-inventory dealer increases from $1 - 0.9^2 = 19\%$ to $1 - 0.9^3 = 27.1\%$. Such an improvement in matching significantly adds to the asset inflow to dealers from customer-sellers. However, the matching rate for customer-buyers and dealers with inventories only improves by 0.9%, from $1 - 0.1^2 = 99\%$ to $1 - 0.1^3 = 99.9\%$. The negligible increase of the outflow is not at all enough to balance the significant rise in the inflow. That is, the asset flow is "stuck" at the dealers, creating a bottleneck, and more buyers are left unmatched.[2] Such an increase in customer-buyers leads to a surge in unrealized trading gains and may reduce welfare. To emphasize, this bottleneck effect is unique to the search capacity n. It always arises except in the knife-edge case where the proportions of dealers with and without inventory are exactly equal. In contrast, the search intensity ρ does not create any asymmetry in matching and always improves welfare.

Section 4 studies how customers choose between BB and SMS. Given that SMS offers faster (electronic) connection and more connection with dealers, one might wonder whether customers still use BB at all. The analysis reveals a downside of SMS when customers have low chance to commit to their reserve prices, i.e., when q is low in SMS. On MarketAxess for example, a customer always receive take-it-or-leave-it offers from dealers, effectively $q = 0$. In this case, the

[2] It is the increase of the unmatched customer-buyers that balances the inflow and the outflow in equilibrium. More precisely, the asset inflow to (outflow from) the dealer sector is the product of (i) the population *size* of customer-sellers (-buyers) and (ii) the dealer-seller (-buyer) matching *rate*. Whereas the inflow increases via the significantly higher matching *rate* (the intensive margin), the outflow increases via the increment in the larger customer-buyer population *size* (the extensive margin). This echoes the asymmetric effect of the search capacity n.

customer's expected trading gain is completely determined by the price competition among the contacted dealers. If such competition is insufficient, little trading gain is left for the customer, because any matched counterparty dealer will charge a monopoly price, knowing that she is likely the only counterparty that the customer finds (out of the n). In contrast, in BB, a customer always has some chances to secure the full trading gain, as long as q in BB is positive.

In equilibrium, the customers do not always use SMS. In particular, when the asset is very imbalanced in supply, one side of the market strictly prefers BB over SMS. Consider the case of excess supply for example. The large number of customer-sellers flood the dealer sector with the asset, making most of the dealers full in inventory. Consequently, the remaining customer-sellers find it very difficult to find dealer counterparty and, even if they do use SMS, any matched dealer will knowingly charge a monopoly price. Instead, resorting to BB, a customer-seller can still negotiate a reasonable price with the dealer. This prediction echos the empirical finding in O'Hara and Zhou (2020) that when corporate bonds are under fire sell (i.e., in excess supply), the electronic trading volume share drops. Arguably, such an intrinsic tradeoff between SMS and BB could have hindered the adoption of electronic OTC trading in corporate bond markets.

The customers' endogenous choices between BB and SMS also have welfare and market design implications. The analysis shows that when the search intensity ρ is high, a social planner strictly prefers SMS over BB, simply because SMS offers better matching, which creates large trading gains. Unlike the planner, who does not care about the split of trading gains, the customers might shy away from SMS because the trading gain split there is inferior, compared to that in BB. Such inefficiency in technology adoption can be reduced by policies and market designs that incentivizes customers to use SMS. In the model, this can be achieved by setting a large enough q in SMS, e.g., by allowing customers to commit to their reserve prices in RFQ platforms.

The analysis, however, caveats that such policy and design fixes might not always work, depending on the intrinsic characteristics of the asset traded. For example, when the search intensity ρ is low, having all investors using SMS is not efficient, because of the bottleneck effect—

the dealer sector might be inefficiently holding too many assets. Such a distinction between "fast" and "slow" moving assets is realistic. While corporate bonds on SMS trade in a few minutes (Hendershott and Madhavan, 2015), auctions of collateralized loan obligations (CLOs) can take a day or two (Hendershott et al., 2020). Asset-specific design and policies should be considered, as opposed to market-wide, blanket recommendations.

Contribution and related literature

The paper contributes to four strands of the literature. First, adding to the search models of OTC markets, this paper introduces the possibility for investors to search for *multiple* potential counterparties *at the same time*. In contrast, previous search models focus on BB as in, for example Duffie, Gârleanu, and Pedersen (2005, 2007), Weill (2007), Vayanos and Weill (2008), Lagos and Rocheteau (2009), Lagos, Rocheteau, and Weill (2011), Üslü (2019), Hugonnier, Lester, and Weill (2020). A noteworthy consequence is that in SMS, the competition among uncertain number of quoting investors generates price dispersion. Several other works also feature price dispersion but with different underlying mechanisms. In Hugonnier, Lester, and Weill (2020) and Shen, Wei, and Yan (2018), investors' heterogeneous valuations drives price dispersion. Vayanos and Wang (2007) show that investors with different horizons form a "clientele" equilibrium, where assets of the same fundamentals are priced differently. Dealers of different inventory levels may quote prices differently as in Yang and Zeng (2018), who show that dealers' coordination leads to multiple equilibria with high and low liquidity. In Zhang (2018), dealers offer different price menus, contingent on customers' history, to screen customers of unobservable but persistent types. Arefeva (2017) studies a housing market in which each seller runs an auction among potential buyers, similar to SMS but with an exogenous influx of buyers.[3] The nature of the price dispersion in the current

[3] Price dispersion has also been often associated with the structure of dealer networks. Li and Schurhoff (2019) show that central dealers charge much higher markups than do peripheral ones in the municipal bond market; see also Maggio, Kermani, and Song (2017). Hollifield, Neklyudov, and Spatt (2017) turn to the pricing of securitizations and, in contrast, find a centrality discount for core dealers. On the theory side, Colliard, Foucault, and Hoffmann (2020) study the distribution of inter-dealer prices on an exogenous network and generate predictions regarding the

model is different. It is due to the strategic behavior of quoters, not to the heterogeneity among them, and such strategic behavior is endogenously affected by search frictions. A unique consequence is that the search friction shapes both the response rate dispersion and the price dispersion.

Second, this paper contributes to the theory literature on electronic OTC markets. Vogel (2019) studies a hybrid OTC market where investors can trade in both the traditional voice market (modeled after Duffie, Dworczak, and Zhu, 2017) and the electronic RFQ platform. Liu, Vogel, and Zhang (2017) compare the the electronic RFQ protocol in an OTC market with a centralized exchange market. Both papers model the RFQ trading similarly to the current paper, in which the searching agent reaches out to a finite number of potential counterparties who respond with uncertainty. The key difference is that in these two papers the RFQ response rates are exogenous, whereas they are endogenous in this paper and depend on both search intensity and search capacity. Importantly, such an endogenous response rate drives the results of asset allocation and efficiency, as well as the comparison between SMS and BB. Riggs et al. (2019) study the RFQ trading in Swap Exchange Facilitites. Their model share with this paper a same prediction that RFQ response rate decreases in n, the number of potential counterparties (i.e., dealers in their model). They explain this phenomenon through winner's curse: winning the RFQ from customer against more competitor dealers implies a worse interdealer price later on. This adverse inference reduces the dealers incentive to bid in the RFQ. The mechanism in this paper is different: a larger n makes matching more efficient, reducing the number of traders who will respond to RFQ, i.e., those unmatched traders with opposite trading needs. In a different line, Saar et al. (2019) compare dealers' market making (directly liquidity provision) and matchmaking (searching on customers' behalf for counterparties) and study the effects of bank dealers' balancesheet costs.

Third, there is a growing literature comparing centralized versus decentralized trading (Pagano,

connectedness of core and peripheral dealers. Neklyudov (2019) shows that dealers' heterogeneous search technology creates a centrality discount but inter-dealer trades might result in a centrality discount. Zhong (2014) analyzes the endogenous network formation of dealers and find that order sizes are, in addition to the network structure, important in determining prices. Compared to the above, a key message of this paper is that even when agents are homogeneous and in the absence of a specific (dealer) network, search frictions alone can generate price dispersion.

1989; Chowdhry and Nanda, 1991) in various aspects. Babus and Parlatore (2017) study the endogenous formation of fragmented markets due to investors' strategic behavior. Glode and Opp (2019) compare the efficiency of OTC and limit-order markets in a setting where investors endogenously acquire expertise. Lee and Wang (2019) study uninformed and informed investors' venue choice through an adverse selection channel. Dugast, Üslü, and Weill (2019) examine banks' choice among centralized trading, OTC trading, or both, in a setting where the banks differ in their risky asset endowment and in their capacity of OTC trading. This paper instead compares the conventional voice trading versus the relatively new electronic trading within the OTC setting.

Finally, this paper contributes to the auctions literature with uncertain number of bidders (see, e.g., the survey by Klemperer, 1999) and to the literature on pricing with heterogeneously informed consumers (e.g., Butters, 1977; Varian, 1980; and Burdett and Judd, 1983). Apart from the above literature speaking to OTC markets, applications of such "random pricing" mechanisms are also seen recently in exchange trading, such as Jovanovic and Menkveld (2015) and Yueshen (2017). The main insight from this paper is that such uncertainty about the number of quoters (bidders) can arise endogenously from the search process.

2 Model setup

Time is continuous. All random variables and stochastic processes are defined on a fixed probability space. We consider the trading of an asset that produces a constant flow of a consumption good. The asset is in fixed supply s.

Customers and dealers. There is a continuum of *c*ustomers with mass m_c and a continuum of *d*ealers with mass m_d. Both groups of agents are risk-neutral, discount the future utility at the same rate r, and can each hold either zero or one unit of the asset. An owner of the asset will be denoted by o and a non-owner n.

The two groups of agents differ in their preferences for the consumption good produced by the asset. Specifically, a customer owner derives instantaneous utility $y(t) \in \{y_h, y_l\}$ (*h*igh or *l*ow), which evolves stochastically according to a continuous time Markov chain:

$$\mathbb{P}(y(t + \mathrm{d}t) = y_h | \, y(t) = y_l) = \lambda_u \mathrm{d}t \text{ and } \mathbb{P}(y(t + \mathrm{d}t) = y_l | \, y(t) = y_h) = \lambda_d \mathrm{d}t,$$

where λ_d and λ_u are the switching intensities. A dealer owner instead receives instantaneous utility y_d, where y_d is constant—dealers are not subject to preference shocks.

In summary, there are four types of customers, $\{ho, hn, lo, ln\}$, and two types of dealers, $\{do, dn\}$. Their population measures at any time t are denoted by $m_\sigma(t)$ for $\sigma \in \{ho, hn, lo, ln, do, dn\}$, with $m_{ho}(t) + m_{hn}(t) + m_{lo}(t) + m_{ln}(t) = m_c$ and $m_{do}(t) + m_{dn}(t) = m_d$.

Search and trading. The setup above follows Hugonnier, Lester, and Weill (2020) but with all dealers having the same preference. We generalize how customers interact with dealers by introducing a trading technology characterized by three parameters, $\{n, \rho, q\}$: At a Poisson process with intensity ρ, a customer can contact n dealers, selected from the whole dealer population at random. The Poisson processes for different customers are independent from one another.[4] Upon contacting the dealers:

- With probability q, the customer moves first, making a take-it-or-leave-it offer (TIOLIO) to all the contacted dealers, who then choose to accept the offer or walk away. If more than one dealer accepts, the customer randomly chooses one to trade with.
- With probability $1-q$, the n dealers move first, simultaneously making independent TIOLIOs to the customer, who then chooses the best quote or walk away.

A contacted dealer may be unable to accommodate the contacting customer due to the inventory constraint. For example, the customer might want to sell, while the contacted dealer might already hold one unit of the asset. Importantly, each dealer makes his decision independently, not knowing the types of the other $(n-1)$ contacted dealers.

[4] Customers can choose how many, possibly fewer than n, dealers to contact. Since there is no cost of contacting more, in equilibrium, customers will always choose to contact n dealers. With contact cost, investors in Riggs et al. (2019) choose an interior number of contacts. Such cost does not bring novel insights in the current model setting and, hence, is set to zero.

Parameter values and supports. We normalize the customer mass to $m_c = 1$ and require the dealer mass $m_d > 0$. We also require $s \in (0, 1 + m_d)$ so as to study asset allocation meaningfully. The customers' preference switching intensities are strictly positive, i.e., $\lambda_u > 0$ and $\lambda_d > 0$. We set $y_h > y_l$ so that some customers are "high" type and some "low." An additional constraint on y_d will be introduced later in Proposition 2 to ensure there is always positive gains from trade. The technology parameters have supports $n \in \mathbb{N}$ (the natural numbers), $\rho \in (0, \infty)$, and $q \in [0, 1]$.

Remarks. Several remarks about the model are in order.

Remark 1. The trading technology is general enough to encompass some most common protocols in OTC trading. For example, the case of $n = 1$ can be thought of customers reaching dealers by phone or email and determining the terms of trade via bilateral bargaining (BB), a setup frequently seen in the literature. The case of $n > 1$ captures technologies that allow a customer reach multiple dealers in one "click," hence the name simultaneous multilateral search (SMS). For example, this is the case for the RFQ protocol on electronic platforms (like MarketAxess and Swap Execution Facilities, SEFs); for auctions like bid/offer-wanted-in-competition (B/OWIC); and in housing markets where a seller can be in touch with possibly many buyers at the same time.

Remark 2. In practice, customers can choose how to get in touch with dealers. They can always dial up to call dealers (BB) but they can also click buttons on electronic platforms like RFQs (SMS). After exploring the equilibrium properties of one general technology in Section 3, we study how customers choose between "call" and "click" by introducing both technologies in Section 4.

Remark 3. The general trading technology is governed by three parameters:

- The search intensity ρ, inherited from DGP and HLW, implies that the technology connects a customer with dealers at exponential waiting times with expectation $1/\rho$. For example, auctions on MarketAxess vary in length from 5 to 20 minutes (Hendershott and Madhavan, 2015). Trading of collateralized loan obligations (CLOs) is typically organized through B/OWIC by email (Hendershott et al., 2020) and can take considerably longer time.

- The search capacity n, new to this paper, flexibly nests bilateral bargaining ($n = 1$) with SMS-like protocols that allow customers to contact multiple dealers. For example, on the MD2C platform operated by Bloomberg Fixed Income Trading, clients can select up to $n = 6$ quotes (Fermanian, Guéant, and Pu, 2017). On Bloomberg Swap Execution Facility (SEF), this upper bound is set to $n = 5$ (Riggs et al., 2019).
- The probability q reflects the customer's intrinsic "bargaining power," relative to the dealer(s), when using the trading technology. When $n = 1$, q is the Nash bargaining power parameter as in DGP and HLW. When $n > 1$, q can reflect the customers' ability to communicate, and commit to, their reserve prices.[5] On a typical RFQ platform such as MrketAxess, q is effectively zero as customers can only solicit quotes from dealers but cannot set reserve prices (O'Hara and Zhou, 2020). Instead, when trading is less formally organized, q can be larger. The BWIC to sell CLOs is conducted by email and it is possible that customers communicate through such emails their indicative reserve price. In housing markets, sellers often post indicative asks that are negotiable.

Remark 4. In reality, not only customers can search for dealers. Dealers can also take initiatives to reach customers. In this paper, we shut down such dealer-to-customer searches for two reasons. First, to our knowledge, most SMS-like trading protocols occur between a single customer and multiple dealers, not the other way. Hence, to study SMS, it is realistic to just focus on customer-to-dealer searches, like RFQs and B/OWICs. Second, allowing dealers to also search for customers essentially adds some baseline matching and trading to the economy. This will not affect the core, novel results about SMS ($n > 1$) of this paper.

[5] In our setting, a customer will always set reserve price equal to the reservation value of her counterparty, an outcome equivalent to the customer making a TIOLIO to the n dealers. Thus, q can equivalently be thought of as the customer's probability to set reserve price.

3 Stationary equilibrium

There are three sets of equilibrium objects: 1) the demographics $\{m_\sigma\}$; 2) the agents' pricing strategies (detailed below); and 3) their value functions $\{V_\sigma\}$. These objects in general can depend on time. This section looks for a stationary Markov perfect equilibrium, under which the objects are time-invariant constants. We also focus on symmetric pricing strategies; that is, agents of the same type use the same strategy when making TIOLIOs.

3.1 Demographics

There are in total six demographic variables, $\{m_{ho}, m_{ln}, m_{hn}, m_{lo}, m_{do}, m_{dn}\}$, one for each type of the agents. The following three conditions must hold in equilibrium by definition:

(1) market clearing: $m_{ho} + m_{lo} + m_{do} = s;$

(2) total customer mass: $m_{ho} + m_{ln} + m_{hn} + m_{lo} = 1;$

(3) total dealer mass: $m_{do} + m_{dn} = m_d.$

In a stationary equilibrium, the total measure of *h*igh type customers must be time-invariant; i.e., the net flow during any instance $\mathrm{d}t$ must be zero:

(4) net flow of high type customers: $(m_{lo} + m_{ln})\lambda_u \cancel{\mathrm{d}t} - (m_{ho} + m_{hn})\lambda_d \cancel{\mathrm{d}t} = 0,$

which also ensures that the net flow of low type customers is zero.

Two more equations are needed in order to pin down the six demographic variables. These two last conditions arise from trading. In equilibrium, only two types of customers want to trade with dealers: The *lo*-type wants to sell to *dn*-buyer, and the *hn*-type wants to buy from *do*-seller. The other two types, *ho* and *ln*, stand by and do not trade (which is a conjecture for now and we will later verify it after Proposition 2).

Consider the inflows to and the outflows from the the *lo*-sellers. In a short period of $\mathrm{d}t$, a measure of $m_{lo}\rho \mathrm{d}t$ of sellers will be searching, each having probability

$$\nu_{lo} := 1 - \left(1 - \frac{m_{dn}}{m_d}\right)^n$$

to find at least one dn-buyer (out of n) to trade.[6] Hence, there is an outflow of $\nu_{lo} m_{lo}\rho \mathrm{d}t$ due to the searching lo-sellers. In addition, due to preference shocks, there is an inflow of $m_{ho}\lambda_d \mathrm{d}t$ and an outflow of $m_{lo}\lambda_u \mathrm{d}t$. In a stationary equilibrium, the sum of the in/outflows above must be zero:

$$\text{(5)} \qquad \text{net flow of } lo\text{-sellers:} \qquad -\nu_{lo} m_{lo}\rho - m_{lo}\lambda_u + m_{ho}\lambda_d = 0.$$

Analogously, define

$$\nu_{hn} := 1 - \left(1 - \frac{m_{do}}{m_d}\right)^n$$

as the probability for a searching hn-buyer to find at least one do-seller. Then the zero net flow condition for hn-buyers becomes:

$$\text{(6)} \qquad \text{net flow of } hn\text{-buyers:} \qquad -\nu_{hn} m_{hn}\rho - m_{hn}\lambda_d + m_{ln}\lambda_u = 0,$$

which is the last equation needed to pin down the stationary demographics.

Lemma 1 (Stationary demographics). *The demographics equations* (1)-(6) *uniquely pin down the population sizes* $\{m_{ho}, m_{ln}, m_{hn}, m_{lo}\} \in (0,1)^4$ *and* $\{m_{do}, m_{dn}\} \in (0, m_d)^2$.

Note that the stationary equilibrium population sizes depend on both the search intensity ρ and the capacity n. In particular, the parameter n appears in the matching probabilities ν_{hn} and ν_{lo}, which is new compared to the the bilateral bargaining protocol often seen in the literature. We will see shortly that n has novel implications on various equilibrium objects.

A few additional observations are worth highlighting. First, in this economy, the hn-buyer-initiated trading volume amounts to $\nu_{hn} m_{hn}\rho$, while the lo-seller-initiated volume is $\nu_{lo} m_{lo}\rho$. They are also, respectively, the asset outflow from and into the dealer sector. Therefore, in a steady state

[6] The exact law of large numbers in Duffie, Qiao, and Sun (2019) is applied so that the fractions of the populations of each type are their expected values. See also Sun (2006) and Duffie and Sun (2007, 2012).

equilibrium, the trading volume t must satisfy:

$$t := v_{hn} m_{hn} \rho = v_{lo} m_{lo} \rho, \tag{7}$$

for otherwise the dealer-owner mass, m_{do}, will not be stable. Indeed, Equation (7) is guaranteed by the system (1)-(6); in particular, by (5) − (6) + (4).

Second, while the system (1)-(6) has only two zero-flow conditions (Equations 5 and 6), the stationarity of all other types of agents is also implied. Apart from the dealer stationarity (7) above, −(4) − (5) gives $v_{lo} m_{lo} \rho - m_{ln} \lambda_u + m_{hn} \lambda_d = 0$, ensuring that the net flow in and out of ln-bystanders is zero. Likewise, (4) − (6) gives $v_{hn} m_{hn} \rho - m_{ho} \lambda_d + m_{lo} \lambda_u = 0$, ensuring that the net flow of ho-bystanders is zero.

Finally, we derive some useful expressions for the customer masses. Equations (1), (2), and (4) together imply the stable fractions of the high-type and the low-type customers:

$$m_{ho} + m_{hn} = \frac{\lambda_u}{\lambda_d + \lambda_u} =: \eta \quad \text{and} \quad m_{lo} + m_{ln} = \frac{\lambda_u}{\lambda_d + \lambda_u} = 1 - \eta. \tag{8}$$

Then combining the market clearing condition (1) and the lo-seller net flow (5), we obtain

$$m_{lo} = (1 - \eta)(s - m_{do}) - \frac{t}{\lambda_u + \lambda_d}, \tag{9}$$

which intuitively says that the stationary mass of lo-sellers is a fraction $(1 - \eta)$ of the residual asset supply $(s - m_{do})$ available to customers, less a term $t/(\lambda_u + \lambda_d)$ due to their active trading. Combining (1) and (5) gives

$$m_{hn} = \eta \cdot (1 + m_{do} - s) - \frac{t}{\lambda_u + \lambda_d}. \tag{10}$$

Note that $1 + m_{do} - s$, which is the total mass of non-owner customers in this economy. That is, the stationary mass of hn-buyers is the high-type fraction η of all non-owner customers, less the same term due to trading. The above expressions are in fact generic in the search literature. For example, if, as in DGP, customers find each other at intensity ρ without dealers, then the equations (9) and (10) still hold with $m_{do} = 0$ and $t = 2\rho m_{hn} m_{lo}$.

3.2 Pricing strategies

This subsection studies the agents' pricing strategies. We first take the agents' value functions $\{V_\sigma\}$ as given to derive the potential gains from trade. (The value functions will be solved in the next subsection.) Specifically, an agent's reservation value R for the asset is her/his value function with the asset less without:

$$R_l := V_{lo} - V_{ln},\ R_h := V_{ho} - V_{hn},\ \text{and } R_d := V_{do} - V_{dn}.$$

Therefore, for a trade between an lo-seller and a dn-buyer to happen, the transaction price p must fall between

$$R_l \leq p \leq R_d; \tag{11}$$

and likewise, for a trade between an hn-buyer and a do-seller, the price must fall between

$$R_d \leq p \leq R_h. \tag{12}$$

For notation simplicity, denote the trading gains for the two kinds of trades respectively by

$$\Delta_{dl} := R_d - R_l \text{ and } \Delta_{hd} := R_h - R_d. \tag{13}$$

For now, we make the conjecture that there are positive trading gains: $0 \leq R_l \leq R_d \leq R_h$, which will be guaranteed by a condition on y_d (see Proposition 2).[7]

Consider now a customer contacts n dealers using the trading technology. First, there is probability q that the searching customer makes a TIOLIO to the dealers. In this case, it is optimal for the customer to quote a price at the dealers' reservation value, i.e., $p = R_d$.

Second, there is probability $1 - q$ that the n dealers independently quote to the customer. For concreteness, suppose the customer is an hn-buyer (the case of a lo-seller is symmetric and omitted).

[7] When this condition on y_d is not met, there might be no trade in this economy. For example, suppose $R_d > R_h$. Then there is no trade between do-dealers and hn-buyers and by the stationarity condition (7), there must be no trade between dn-dealers and lo-sellers, either. We therefore focus on the more interesting and empirically relevant case with trades.

In this case, a quoting dealer must be a *do*-seller and would love to capture the full surplus by setting $p \uparrow R_h$. However, he faces potential competition from the other $(n-1)$ dealers, as their asking quotes might be lower than his. Yet not all of the other $(n-1)$ dealers are necessarily also *do*-sellers. The quoting *do*-seller therefore engages in a price competition with *unknown number of competitors*.

Such price competition differs from the standard Bertrand price competition, in which every dealer-seller quotes his reservation price of R_d and the customer-buyer gets the full surplus Δ_{hd}. Instead, every dealer-seller has an incentive to charge a higher price, $R_d + \alpha\Delta_{hd}$ for some $\alpha \in [0, 1]$. (When $\alpha = 1$, $R_d + \alpha\Delta_{hd} = R_h$ which is the customer-buyer's reservation value.) This is because he might actually be the only *do*-seller among the n contacted dealers, in which case his quote is the only price available to the contacting *hn*-buyer. As long as $\alpha \leq 1$, the buyer will accept it[8] and the dealer can pocket the difference $\alpha\Delta_{hd}$ as his profit. In a Nash equilibrium, however, the fraction α cannot be deterministic, as the undercutting argument of Bertrand competition will lead to $\alpha \downarrow 0$. Yet, it would be strictly better off to quote some $\alpha > 0$ as all the potential competitors were to quote $\alpha \downarrow 0$. The heuristic discussion above is formalized in the proof and summarized by the following proposition.

Proposition 1 (Dealers' equilibrium quoting). *Suppose a customer contacts n (≥ 1) dealer(s). With probability $1-q$, each dealer independently makes a TIOLIO. Within symmetric strategies, there is a unique mixed-strategy equilibrium for the dealers. Define*

$$F(x; \mu, n) := \frac{1}{\mu} - \left(\frac{1}{\mu} - 1\right) x^{-\frac{1}{n-1}}, \quad \text{with support } (1-\mu)^{n-1} \leq x \leq 1,$$

for some $\mu \in (0, 1)$ and $n \in \mathbb{N}$. Then,

- *a do-seller asks at $R_l + \alpha\Delta_{hd}$, where α is random with c.d.f. $F(\alpha; m_{do}/m_d, n)$; and*
- *a dn-buyer bids at $R_h - \beta\Delta_{dl}$, where β is random with c.d.f. $F(\beta; m_{dn}/m_d, n)$.*

Note that when $n = 1$, $F(\cdot)$ becomes a degenerate c.d.f. with a single probability mass at the

[8] To see this, note that by accepting an offer $p = R_d + \alpha\Delta_{hd}$, the customer-buyer becomes *ho*-bystander and gets a continuation value of $V_{ho} - p$. If instead he rejects the offer, his value remains as V_{hn}. This customer-buyer will accept the offer as long as $V_{ho} - p \geq V_{hn}$, a condition equivalent to $\alpha \leq 1$.

maximum support $x = 1$.

The proposition above implies that a quoting *do*-seller expects a trading price of $R_l + \bar{\alpha}\Delta_{hd}$ and a quoting *dn*-buyer expects $R_h - \bar{\beta}\Delta_{dl}$, where

$$\bar{\alpha} := \mathbb{E}[\alpha] = \left(1 - \frac{m_{do}}{m_d}\right)^{n-1} \quad \text{and} \quad \bar{\beta} := \mathbb{E}[\beta] = \left(1 - \frac{m_{dn}}{m_d}\right)^{n-1}. \tag{14}$$

To see this, consider a quoting *do*-seller and note that under the mixed-strategy equilibrium, he must be indifferent across all possible $\alpha \in [0, 1]$. In particular, the only situation for quoting $\alpha = 1$ to "win" is that there are no other competing *do*-sellers; that is, with probability $(1 - m_{dn}/m_d)^{n-1}$. Therefore, when contacted, a quoting *do*-seller expects a profit of $\bar{\alpha}\Delta$, where $\bar{\alpha}$ can be interpreted as his expected trading gain share. Likewise, a quoting *dn*-buyer expects $\bar{\beta}\Delta$.

Proposition 1 characterizes a contacted dealer's quoting strategy. From a contacting customer's perspective, however, the expected trading price has a different distribution, because she can pick the best quote and because there might not be a quote if none of the contacted dealers are of the matching type. Consider a contacting *hn*-buyer for example. He contacts n dealers knowing that the number of counterparties he will actually find, N_{do}, is random and follows a binomial distribution with n draws and success rate $\frac{m_{do}}{m_d}$, which is the expected response rate. Each of these N_{do} dealers then quotes a random price according to $F\left(\alpha; \frac{m_{do}}{m_d}, n\right)$, following Proposition 1. (The *hn*-buyer can safely ignore the other $n - N_{do}$ dealers' quotes, as they both want to buy.) The contacting *hn*-buyer then picks the lowest ask among the N_{do} available quotes. Conditional on the realization $N_{do} \geq 1$, the c.d.f. of this minimum ask is $1 - (1 - F(\alpha; \cdot))^{N_{do}-1}$. (When $N_{do} = 0$, the *hn*-buyer finds no ask quote and there is no trade.) Averaging across all possible $N_{do} \in \{1, ..., n\}$, the corollary below gives the expectation of this minimum ask quote.

Corollary 1 (Trading prices). *Define* $G(\mu, n) := \frac{n\mu\cdot(1-\mu)^{n-1}}{1-(1-\mu)^n}$ *for some* $\mu \in (0, 1)$ *and* $n \in \mathbb{N}$. *Then, with probability q, a searching customer sets the price equal to the dealers' reservation value* R_d*; and with probability* $1 - q$,

- *an hn-buyer expects ask quotes from do-dealer(s) with probability* $\left(1 - \left(1 - \frac{m_{do}}{m_d}\right)^n\right)$ *and* *the expected best ask, is* $R_d + \bar{A}\Delta_{hd}$*, where* $\bar{A} = G\Big(\frac{m_{do}}{m_d}, n\Big)$*; and*

- *an lo-seller expects bid quotes from dn-dealer(s) with probability* $\Big(1 - \Big(1 - \frac{m_{dn}}{m_d}\Big)^n\Big)$ *and the expected best bid, is* $R_d - \bar{B}\Delta_{dl}$*, where* $\bar{B} = G\Big(\frac{m_{dn}}{m_d}, n\Big)$*.*

The above expected quotes, $R_d + \bar{A}\Delta_{hd}$ *and* $R_d - \bar{B}\Delta_{dl}$*, are also the average trading prices in buyer- and seller-initiated trades, respectively. The average trading price across all trades is*

$$\bar{p} = R_d + \frac{1-q}{2}\big(\bar{A}\Delta_{hd} - \bar{B}\Delta_{dl}\big). \tag{15}$$

Note that when $n = 1$*,* $\bar{A} = \bar{B} = 1$ *for all* $\mu \in (0, 1)$*.*

Several features of the equilibrium pricing above are worth highlighting.

Splitting the surplus. Corollary 1 shows how the trading gains are split between one contacting customer and n *potential* counterparty dealers. Recall that with probability q, the customer is able to capture the full trading gain. Therefore, conditional on finding at least one dealer of her matching type, an *hn*-buyer expects a profit of

$$q\Delta_{hd} + (1-q)(R_h - (R_d + \bar{A}\Delta_{hd})) = (q + (1-q)(1-\bar{A}))\Delta_{hd}, \tag{16}$$

while an *lo*-seller expects

$$q\Delta_{dl} + (1-q)((R_d - \bar{A}\Delta_{hd}) - R_d) = (q + (1-q)(1-\bar{B}))\Delta_{dl}. \tag{17}$$

Thus, a contacting *hn*-buyer expects a fraction of $(q + (1-q)(1-\bar{A}))$, and the rest $(1-q)\bar{A}$ is expected by the N_{do} contacted *do*-sellers. The split of the trading gains depends on the search capacity n varies. When $n = 1$, the *do*-seller becomes a monopolist who sets $\bar{A} = 1$, effectively extracting all the surplus Δ_{hd}. When $n \uparrow \infty$, the *do*-seller is effectively price-competing with infinitely many others and all the surplus is attributed to the contacting buyer, as in a Bertrand competition with $\bar{A} = 0$. That is, a larger search capacity n "adjusts" the split of the trading gains from dealers to customers ($\bar{A}$ is decreasing in n, taking m_{do} as given).[9]

[9] Proposition 1 offers another way to decompose the trading gains, between a pair of *matched* customer and dealer. For example, between a contacting *hn*-buyer and a matched quoting *do*-seller, the former gets $(1-\bar{\alpha})\Delta_{hd}$ and the

Endogenous bargaining power. In the example above, the fractions $(q+(1-q)(1-\bar{A}))$ vs. $(1-q)\bar{A}$ are reminiscent of the bargaining power parameters in a Nash bargaining game like in DGP and in HLW. There are three key differences. First, these fractions are *endogenous* in the current model, depending on the equilibrium population sizes of counterparties, reflected in the endogenous $\bar{A}$. As a special case, when $n = 1$, $\bar{A} = 0$ and the split of trading gain defaults to the Nash bargaining case where the customer has the exogenous bargaining power q and the dealer has $1 - q$. On the other hand, under the RFQ-like protocols, customers only solicit quotes ($q = 0$) and they only rely on the n (> 1) dealers' price competition to extract trading gains. Second, when $n > 1$, a customer's bargaining power is *one-to-many*, as she contacts multiple potential counterparties. In DGP and HLW for example, the bargaining power parameters are always one-to-one (bilateral). Third, not only the agent type (customer vs. dealer), but also the direction of the trade (buying vs. selling), matters. For example, a *do*-seller gets a fraction of $(1-q)(1-\bar{A})$, while a *dn*-seller gets $(1-q)(1-\bar{B})$. In contrast, the exogenous bargaining power parameters, like q, are typically not directional.

Price dispersion. Proposition 1 and Corollary 1 imply that there is price dispersion in equilibrium, in the form of *random* markups or markdowns. Such dispersion is due to the unknown number of competitors, an intrinsic feature in SMS: The contacted dealers' types are unknown to each other. In the current stylized model, such types boil down to the dealers' inventory holdings (*do* vs. *dn*). In real-world trading, agents' other characteristics (like risk-aversion, patience, wealth, relationship with customers, etc.) can enrich their possible types. As long as such a friction remains, price dispersion will be a robust feature in equilibrium. Empirical evidence supports this equilibrium result. For example, Hendershott and Madhavan (2015) document a significant dispersion in dealers' responding quotes.

latter gets $\bar{\alpha}\Delta_{hd}$. Recall that N_{do} is a Binomial random variable of n draws and success rate m_{do}/m_d. Then indeed, $\bar{A} = \mathbb{E}[N_{do}|\, N_{do} \geq 1]\bar{\alpha}$, where $\mathbb{E}[N_{do}|\, N_{do} \geq 1] = (nm_{do}/m_d)/(1-(1-m_{do}/m_d)^n)$.

3.3 Value functions

We study the stationary equilibrium value functions $\{V_\sigma\}$ in this subsection. Consider first an ho-bystander (a customer not trading). Over a short period $\mathrm{d}t$, the ho-bystander gets a flow utility $y_h\mathrm{d}t$ from holding the asset; plus, with intensity $\lambda_d\mathrm{d}t$, she switches to lo-type and her value changes by $V_{lo}-V_{ho}$, minus the depreciation of $rV_{ho}\mathrm{d}t$ due to discounting. Hence, the Hamilton-Jacobi-Bellman (HJB) equation is

$$0 = y_h + \lambda_d \cdot (V_{lo} - V_{ho}) - rV_{ho}. \tag{18}$$

Similarly, an ln-bystander has HJB equation

$$0 = \lambda_u \cdot (V_{hn} - V_{ln}) - rV_{ln}. \tag{19}$$

Consider next an lo-seller. Just like before, over $\mathrm{d}t$ units of time, her value increases by $y_l\mathrm{d}t$ due to the asset holding. It may also change by $V_{ho} - V_{lo}$ with intensity $\lambda_u\mathrm{d}t$ due to a preference shock. The value also reduces by $rV_{lo}\mathrm{d}t$ due to discounting. Apart from these three, there is trading, from which she expects an instantaneous trading gain of $\zeta_{lo}\Delta_{dl}\mathrm{d}t$, with coefficient

$$\zeta_{lo} := \rho \nu_{lo} \cdot (q + (1 - q)(1 - \bar{B}))$$

representing an lo-seller's "expected trading gain intensity:" She searches at intensity ρ, finds at least one counterparty dealer (dn-type) with probability ν_{lo}, and, by Equation (16), expects a gain of $(q + (1 - q)(1 - \bar{B}))\Delta_{dl}$ in such a case. Therefore, the HJB equation for an lo-seller is

$$0 = y_l + \lambda_u \cdot (V_{ho} - V_{lo}) - rV_{lo} + \zeta_{lo}\Delta_{dl}. \tag{20}$$

Similarly, an hn-buyer has

$$0 = \lambda_d \cdot (V_{ln} - V_{hn}) - rV_{hn} + \zeta_{hn}\Delta_{hd}, \tag{21}$$

where the expected trading gain intensity is

$$\zeta_{hn} := \rho v_{hn} \cdot (q + (1-q)(1-\bar{A})).$$

Finally, consider the dealers. A *do*-seller's HJB equation has the similar structure as before:

$$0 = y_d - rV_{do} + \zeta_{do}\Delta_{hd}. \tag{22}$$

Note that there is no type-switching term because dealers do not receive preference shocks. To find a *do*-seller's trading gain intensity ζ_{do}, note that the total trading gain from all *hn*-buyer initiated trades amounts to $m_{hn}\rho v_{hn}\Delta_{hd}$. Since each *hn*-buyer expects $\zeta_{hn}\Delta_{hd}$, a *do*-seller gets the per capita remainder; that is,

$$\zeta_{do} := \frac{m_{hn}\rho v_{hn} - m_{hn}\zeta_{hn}}{m_{do}} = \frac{m_{hn}\rho v_{hn}}{m_{do}}(1-q)\bar{A}.$$

Similarly, a *dn*-buyer has

$$0 = -rV_{dn} + \zeta_{dn}\Delta_{dl} \tag{23}$$

with trading gain intensity

$$\zeta_{dn} := \frac{m_{lo}\rho v_{lo} - m_{lo}\zeta_{lo}}{m_{dn}} = \frac{m_{lo}\rho v_{lo}}{m_{dn}}(1-q)\bar{B}.$$

Recall from Equation (13) that both trading gains Δ_{hd} and Δ_{dl} are linear combinations of the value functions $\{V_\sigma\}$. Thus, the equations (18)-(23) constitute a linear system with six equations and six unknowns. The proposition below solves the system in terms of the total trading gain and the reservation prices.

Proposition 2 (Equilibrium value functions). *Define the thresholds $\overline{y}_d$ and $\underline{y}_d$ as*

$$\overline{y}_d := y_h - (y_h - y_l)\frac{\lambda_d}{\lambda_d + \lambda_u + r} \quad \text{and} \quad \underline{y}_d := y_l + (y_h - y_l)\frac{\lambda_u}{\lambda_d + \lambda_u + r}.$$

When $\underline{y}_d \le y_d \le \overline{y}_d$, the reservation values satisfy $R_l < R_d < R_h$ and there exists a unique stationary equilibrium, where the value functions are the solution to the linear equation systems

(18)-(23).

The proposition highlights that an assumption is needed that $y_d \in (\overline{y}_d, \underline{y}_d)$, which effectively guarantees positive trading gains, i.e., $\Delta_{hd} = R_h - R_d > 0$ and $\Delta_{dl} = R_d - R_l > 0$. (Similar conditions are also seen in the literature; e.g., Proposition 3 of HLW.) When $y_d \notin (\overline{y}_d, \underline{y}_d)$, intuitively, the dealers are no longer "intermediaries" between buyers and sellers and might be unwilling to trade with the buyers or the sellers and the economy enters a steady state without trading. To rule out such an uninteresting scenario, in the rest of the analysis, we focus on the case of strictly positive trading gains by assuming that $y_d \in (\overline{y}_d, \underline{y}_d)$ always holds.

Finally, we verify the earlier conjecture that indeed *ho*- and *ln*-customers are bystanders:

Lemma 2 (Bystanders do stand by). *When $\Delta_{hd} > 0$ and $\Delta_{dl} > 0$, both ho- and ln-customers stay out of trading.*

Proof. If one did switch to trading, her expected trading price p would fall between the reservation values. For example, if an *ho*-customer were to sell, she would get a price between $R_l = V_{lo} - V_{ln} \leq p \leq V_{do} - V_{dn} = R_d$ and continue with V_{hn}. Given the strictly positive trading gains, we have $R_d < R_h = V_{ho} - V_{hn}$, implying $V_{ho} > V_{hn} + p$, and the *ho*-customer never wants to sell. The same holds for an *lo*-customer. They are really bystanders. □

3.4 Equilibrium properties

In this subsection, we study the equilibrium properties of SMS. We are particularly interested in the contrast of the two search parameters, the intensity ρ and the capacity n—how fast customers can find dealers vs. how many dealers can be reached in one "click."

We focus on the case where the asset is in excess supply, formally defined below:

Lemma 3 (The long and the short sides.). *The hn-buyers are on the short-side of the market, i.e., $m_{hn} < m_{lo}$, if and only if*

$$s > \eta + \frac{1}{2} m_d, \text{ where } \eta := \frac{\lambda_u}{\lambda_u + \lambda_d}. \tag{24}$$

Intuitively, the threshold on the right-hand side of (24) is the "intrinsic demand" for the asset: The fraction η is the population size of the steady-state high-type customers, who are natural holders of the asset. In addition, since the dealers are homogeneous, half of them are also natural holders of the asset. When the asset supply s is in excess of such intrinsic demand, the hn-buyers are on the short-side and the lo-sellers on the long-side. (The case of excess demand, $s \leq \eta + \frac{1}{2} m_d$, is symmetric and is omitted for brevity.)

3.4.1 Dealer sizes and matching rates

Figure 1 illustrates how the dealer sector is affected by the two search parameters. The contour graphs have the search intensity ρ on the vertical axis and the capacity n on the horizontal axis, both in log scale. Panel (a) and (b) focus on the population sizes of the dealers. It can be seen that as either ρ or n increases, the size of do-sellers increases, while that of dn-buyers reduces. Note that the isoquants in the two panels complement to $m_{do} + m_{dn} = m_d$, which we choose to be $m_d = 0.1$ for the numerical illustration.

Note that when either n or ρ is sufficiently large, the dealer masses converge to $m_{do} \to 0.1$ and $m_{dn} \to 0$, which turn out to be the values in the Walrasian equilibrium: Under the chosen parameter values, there are $\eta = \lambda_u/(\lambda_u + \lambda_d) = 0.5$ units of h-type of customers. Given the supply of $s = 0.6$, the Walrasian allocation is that all h-type customers hold the asset and the rest $s - \eta = 0.1$ units of the asset go to the dealer sector (because $y_h > y_d > y_l$), i.e., $m_{do} \to 0.1$ and $m_{dn} \to 0$. That is, both ρ and n share the same effect of pushing the dealer masses towards the Walrasian equilibrium.

Panel (c) and (d) illustrate the effects on the matching rates, $\nu_{hn} = 1 - \left(1 - \frac{m_{do}}{m_d}\right)^n$ and $\nu_{lo} = 1 - \left(\frac{m_{dn}}{m_d}\right)^n$, which are the respective probabilities for a searching hn-buyer and a searching lo-seller

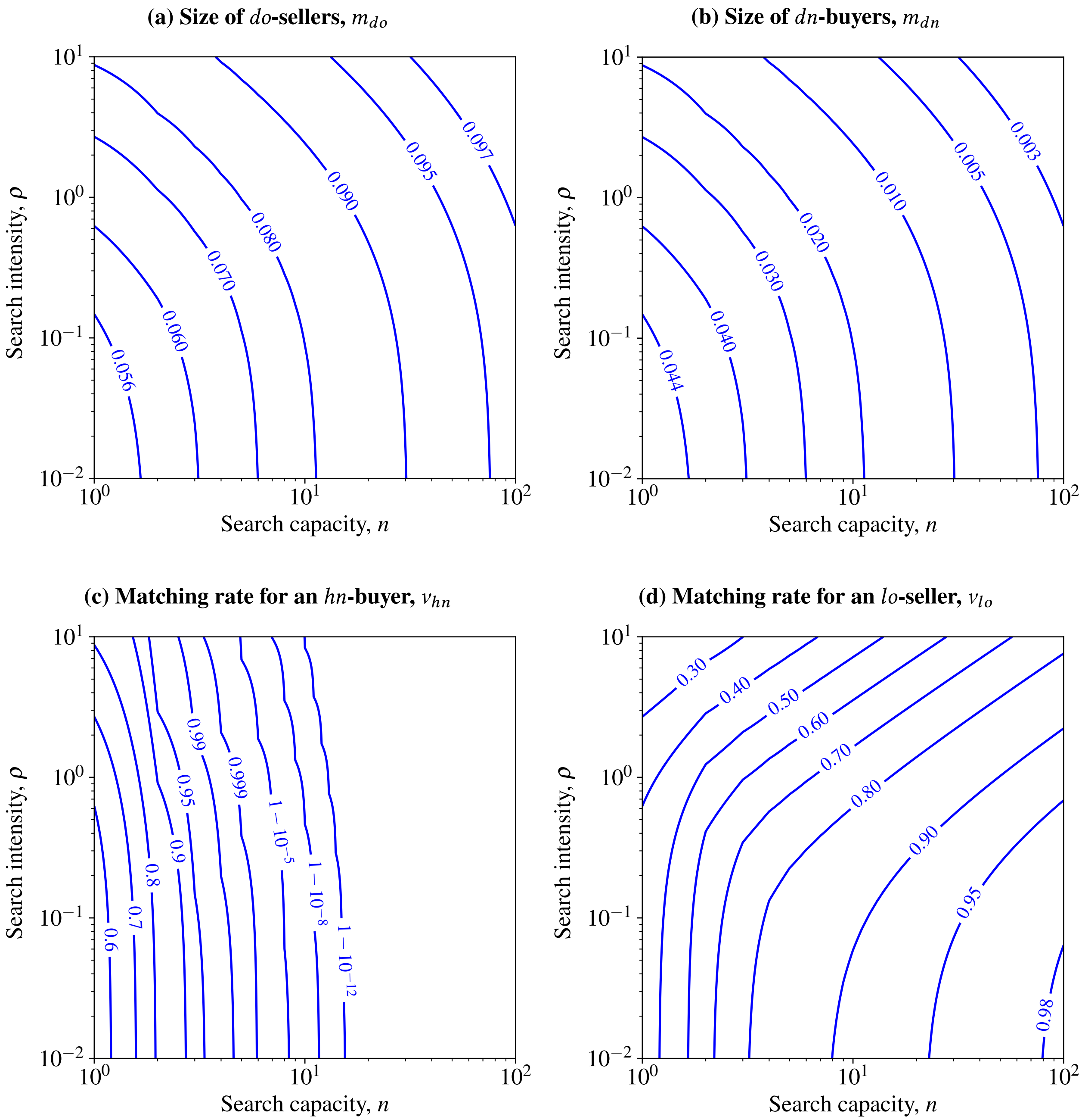

Figure 1: Dealer sizes and matching rates. This figure plots how the search intensity ρ and the search capacity n affect the dealer sizes and the customers' matching rates with dealers. Panel (a) and (b) plot the sizes of dealers. Panel (c) and (d) plot the probability for a customer to find a matching dealer. Other than ρ and n, the parameters are set at $s = 0.6$, $m_d = \lambda_d = \lambda_u = r = 0.1$, $y_h = 10.0$, $y_d = 3.5$, and $y_l = 0.0$. (The customers' intrinsic bargaining power q is irrelevant here as the equilibrium demographics do not depend on it; see Equations 1-6 and Lemma 1.)

to find at least one counterparty dealer. As seen in Panels (a) and (b), a higher search intensity ρ increases m_{do} and decreases m_{dn}, which makes it easier for hn-buyers (the short side) to find a counterparty dealer, but more difficult for lo-sellers (the long side). That is, along any vertical cut, a higher ρ increases v_{hn} in Panel (c) but decreases in v_{lo} in Panel (d).

This effect, however, is *not* shared by the search capacity n: Along any horizontal cut in Panel (c) or (d), both v_{hn} and v_{lo} increase with n. This is because a higher n allows the searching customer to reach more potential counterparties, as reflected in the exponent n in the expressions of v_{hn} and v_{lo} above. This effect of n dominates the decreasing m_{dn} in v_{lo}, resulting in the increasing trend seen in Panel (d). The follow proposition formally summarizes the patterns seen in the figure.

> **Proposition 3 (Search technology and the customer-dealer matching).** *When there is excess supply, both the search intensity ρ and the capacity n increase m_{do} and reduce m_{dn}, but their effects on customers' matching rates are different: a higher ρ increases v_{hn} but decreases v_{lo}, while a larger n increases both v_{hn} and v_{lo}.*

This contrast between ρ and n is the key to understand the effect on asset allocation and welfare, which we examine below.

3.4.2 Customer sizes, trading volume, and welfare

Figure 2(a) shows that the trading volume t (Equation 7) increases with both the search intensity ρ and the capacity n. This is intuitive: both search technology parameters improve the matching efficiency between customers and dealers.

Instead, Figure 2(b) and (c) shows there is stark contrast between how ρ and n affect the customer population sizes. While a higher intensity ρ always decreases the mass of the unmatched, trading customers—helping the efficient allocation, the effect of a larger capacity n differs for the short and the long side of the market. Specifically, while n monotonically reduces m_{lo} (the long side), it can increase m_{hn} (the short side) when sufficiently large. For example, a horizontal cut of $\rho \approx 0.02$ in

Panel (b) shows that a sufficiently large n increases m_{hn}. That is, a larger search capacity n might exacerbate inefficient allocation.

> **Proposition 4 (Search technology and customer sizes).** *The trading volume $t = \rho m_{lo} v_{lo} = \rho m_{hn} v_{hn}$ increases in both n and ρ. The search intensity ρ always reduces both m_{hn} and m_{lo}. The search capacity n always reduces the long-side customer mass but has ambiguous effect on the short-side customer mass. In particular, when n is sufficiently large, the short-side customer mass increases with n.*

The bottleneck effect. Intuitively, a larger search capacity n helps matching: As seen in Figure 1(c) and (d), both the probabilities of v_{lo} and v_{hn} of finding at least one dealer counterparty increase with n. However, the magnitudes of the increases are far from equal: along any horizontal cut, the increment in v_{lo} is much more substantial than that in v_{hn}. That is, all else equal, an increase in n matches many more lo-dn pairs than hn-do pairs. This is because the hn-buyers are on the short-side of the market and there are many more do-dealers to be easily found (than dn-dealers for the long-side lo-sellers); see Figure 1(a) and (b).

Note that the lo-dn trades let the asset flow into the dealer sector from lo-sellers, while the hn-do trades let the asset flow out of the dealers to hn-buyers. Such asymmetric effects of n—the substantially larger increment in the inflow and than in the outflow—clog the asset flow at the dealers, creating a "bottleneck."

In other words, the bottleneck effectively takes in the asset from lo-sellers but not giving it out to hn-buyers. Therefore, the size of hn-buyers, m_{hn}, increases, while the size of lo-sellers, m_{lo}, reduces. Hence, there are two pairs of asymmetric effects: In terms of matching probability, v_{lo} increases much faster than v_{hn}. In terms of population sizes, m_{lo} shrinks whereas m_{hn} surges. These two pairs together ensure the stationarity of dealers in the new equilibrium, with $\rho v_{lo} m_{lo} = \rho v_{hn} m_{hn}$ (Equation 7). Two comments regarding this bottleneck are worth emphasizing.

- The bottleneck effect arises only with the search capacity n, but *not* with the intensity ρ. This is because ρ scales up the inflow $\rho v_{lo} m_{lo}$ and the outflow $\rho v_{hn} m_{hn}$ together and there is no

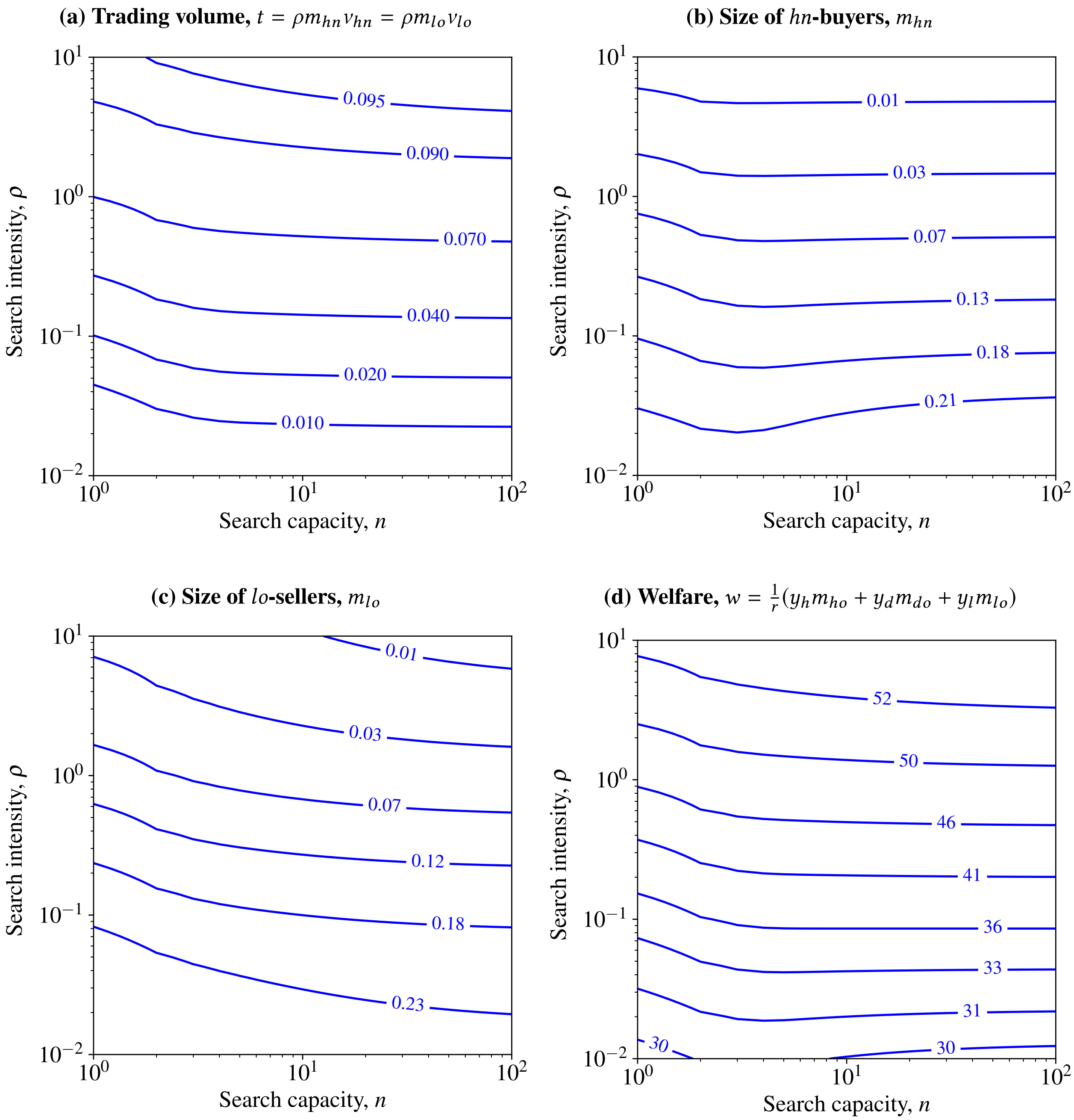

Figure 2: Customer sizes, trading volume, and welfare. This figure plots how the search intensity ρ and the search capacity n affect customer sizes and welfare. Panel (a) plots the trading volume. Panel (b) and (c) plot the sizes of hn-buyers and of lo-sellers, respectively. Panel (c) plots welfare. Other than ρ and n, the parameters are set at $s = 0.6$, $m_d = \lambda_d = \lambda_u = r = 0.1$, $y_h = 10.0$, $y_d = 3.5$, and $y_l = 0.0$. (The customers' intrinsic bargaining power q is irrelevant here as the equilibrium demographics do not depend on it; see Equations 1-6 and Lemma 1. Neither is q relevant for welfare, as it only affects the split of the trading gain but not the total size of it.)

asymmetry.

- The short-side customer mass—m_{hn} in Figure 2(b)—increases with n only when n is sufficiently large. This is because the matching probability v_{hn} is strongly concave in n. When n is still small, its incremental effect on v_{hn} is still substantial. For example, in Figure 1(c), it can be seen that the incremental effect of a small n easily pushes v_{hn} up by 10% (from 0.6 to 0.7) but quickly diminishes as n becomes large. That is, when n is small, its increase in the asset outflow from dealers to hn-buyers is still significant, thus preventing the bottleneck.

Welfare implications. The inefficient allocation due to the bottleneck can hurt welfare, which is the present value of all asset-owners' utility flows:

$$w := \frac{1}{r}(y_h m_{ho} + y_d m_{do} + y_l m_{lo}).$$

Note that welfare is only determined by the demographics (Section 3.1). Unsurprisingly, this is because the pricing strategies (Section 3.2) only affect the and the split of trading gains (Section 3.3), but not the size of the total "pie." Using Equations (9) and (10), recalling also that $m_{ho} + m_{hn} = \eta$, one can rewrite the above as

$$w = \left(\frac{y_d}{r} m_{do} + \frac{\hat{y}}{r}(s - m_{do})\right) + \frac{(y_h - y_l)}{r}\frac{t}{\lambda_u + \lambda_d}, \tag{25}$$

where

$$\hat{y} := \eta y_h + (1 - \eta) y_l$$

can be interpreted as an average customer's instantaneous utility flow for the asset, because in a steady state there is always η fraction of h-type and $1 - \eta$ fraction of l-type customers.

Expression (25) highlights that welfare is composed of: (i) the steady-state asset allocation—m_{do} units to the dealers and the rest $(s - m_{do})$ to the customers; and (ii) the gains from trade $\frac{y_h - y_l}{r}$, passing the asset from l-type to h-type. Effect (ii) is always positive and it scales with the trading volume t, which is increasing in the search capacity n (Figure 1c). Effect (i) is ambiguous: As n increases, a "swelling" bottleneck of m_{do} captures the asset from the customers, who on average value the asset at $\hat{y}$, at the dealers, who value the asset at y_d. A welfare loss occurs whenever $y_d < \hat{y}$; and when such loss dominates the trading benefit (ii), the bottleneck destroys welfare.

Proposition 5 (Search technology and welfare). *A higher search intensity ρ always improves welfare. A larger search capacity n improves welfare only when ρ is sufficiently high. In particular, when ρ is low enough and when $y_d < \hat{y} \in (\underline{y}_d, \overline{y}_d)$, a larger n reduces welfare.*

Figure 2(d) illustrates such welfare losses. For example, along a horizontal cut at $\rho \approx 0.02$, welfare decreases with n. As the proposition suggests, the loss is particularly salient when the trading is sparse; e.g., comparing $\rho \approx 0.02$ with $\rho \approx 0.01$. This is because when trading intensity ρ is very low, the trading benefit (ii) above is effectively shut down with $t = \rho m_{hn} v_{hn} = \rho m_{lo} v_{lo} \downarrow 0$. Note that unlike the search capacity n, the intensity ρ always improves welfare. The reason is precisely because, as discussed earlier, there is no bottleneck effect from ρ.

3.4.3 The asset price

Figure 3(a) plots the contour graph of the asset's average trading price, $\bar{p}$, as given in Corollary 1. When the search intensity ρ increases, $\bar{p}$ monotonically reduces, eventually converges to the Walrasian equilibrium level (which is $y_l/r = 0$ under the current parametrization, as the hn-buyers are on the short side). Intuitively, this is due to "the matching effect" of ρ: As customers can initiate trades more frequently, the asset is allocated more efficiently and its price approaches the efficient level. This monotonic convergence to Walrasian price is inherited from, e.g., DGP and HLW.

Yet, a larger search capacity n can push the asset price in the *inefficient* direction. For example, along a horizontal cut at $\rho \approx 8$, the price $\bar{p}$ is initially decreasing but eventually *increasing*, away from the Walrasian level. To understand why, note that the trading price $\bar{p}$ is largely driven by the dealers' reservation value R_d, as shown in Panel (d), which has almost the same pattern as seen in (c). So the key is to understand how n affects R_d.

A larger search capacity n has a novel "competition effect:" Competing with more potential competitors, dealers' trading gains are lower and so are their value functions. This competition

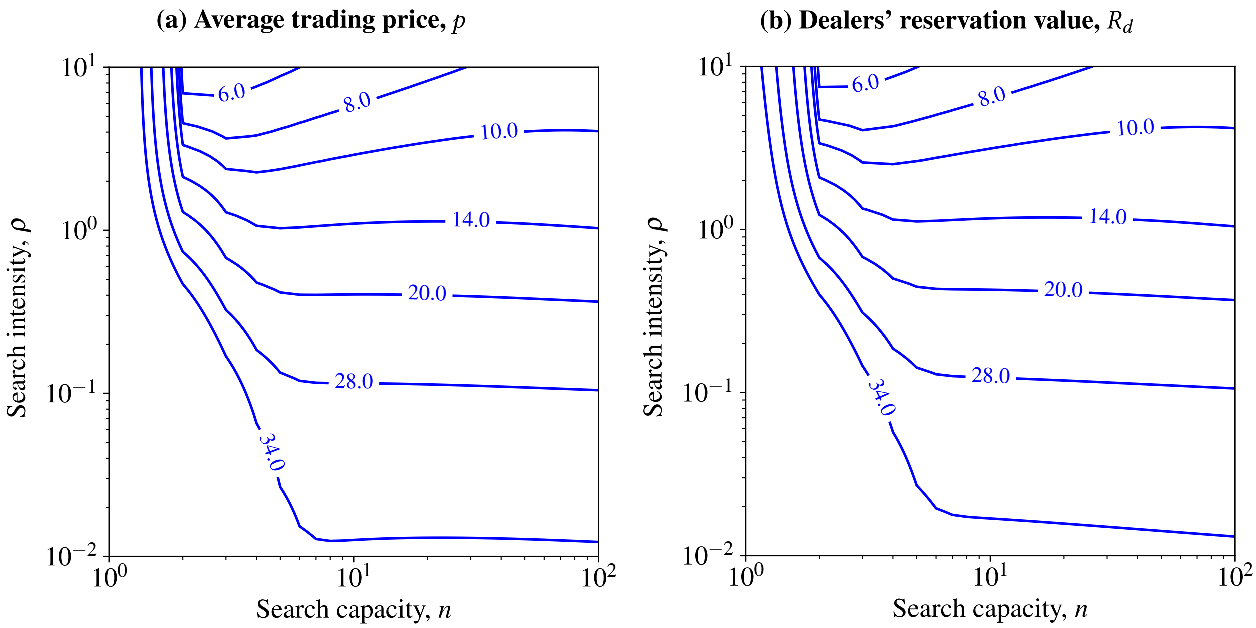

Figure 3: The asset price and reservation values. This figure plots how the search intensity ρ and the search capacity n affect the asset price and agents' reservation values. Panel (a) plots the average trading price of the asset. Panel (b) plots the dealers' reservation value. Other than ρ and n, the parameters are set at $s = 0.6$, $m_d = \lambda_d = \lambda_u = r = 0.1$, $y_h = 10.0$, $y_d = 3.5$, and $y_l = 0.0$. The customers' chance to commit to their reserve prices, q, is set to 0.0 to reflect a realistic RFQ setting.

effect, however, has different magnitudes of impact on *do*-sellers and on *dn*-buyers. From Figure 1(a) and (b), it can be seen that for a moderately large n, there are many more *do*-sellers, who face the short-side customers, than *dn*-buyers, who face the long-side. That is, the competition for *do*-sellers is already close to perfect competition. The marginal increase in the competition thus hurts V_{dn} much more than V_{do}. Since $R_d := V_{do} - V_{dn}$, a larger n therefore tends to raise the dealer's reservation value, which in turn pushes up the average trading price of the asset, against the Walrasian equilibrium price.[10]

[10] This competition effect of n on the asset price is always against the Walrasian level. In the case of excess demand, i.e., when the *hn*-buyers are on the long side, a larger n hurts V_{do} more than V_{dn}. The dealers' reservation value R_d will therefore decrease, again moving away from the Walrasian equilibrium price.

4 SMS versus BB: How to search

In real-world trading, investors can choose their trading technologies. For example, while bilateral bargaining is still the dominant form of trading in corporate bonds, electronic platforms with RFQ protocols have been on the rise (O'Hara and Zhou, 2020). We consider investors' choice of "Click or Call" (Hendershott and Madhavan, 2015) in this section.

Specifically, in the framework set up in Section 2, we introduce two trading technologies, BB and SMS, which differ in their parameters $\{n^k, \rho^k, q^k\}$, $k \in \{\text{BB}, \text{SMS}\}$. (Some realistic parameter restrictions are imposed below.) Each customer can choose, at any point of time, which technology to use to contact dealers and to trade, if she wants to. All dealers can be reached either by BB and by SMS. The other model ingredients remain the same as in Section 2.

The objective is threefold. First, Section 4.1 analyzes how customers choose between the two technologies. Second, we ask, can SMS-like electronic trading (e.g., RFQ) completely replace traditional bilateral bargaining? The answer is no. Section 4.2 shows that in stress periods (e.g., after a fire sale), BB is used more often than SMS. Third, Section 4.3 draws implications on welfare, policy, and market design.

Parameter constraints: Motivated by "calls" (BB) and "clicks" (SMS), we assume

$$n^{\text{BB}} = 1, \quad n^{\text{SMS}} > 2, \text{ and } \rho^{\text{BB}} \leq \rho^{\text{SMS}}. \tag{26}$$

In a bilateral call, a customer bargains with one dealer, hence $n^{\text{BB}} = 1$. By clicking, a typical real-world RFQ protocol connects the customer to multiple dealers, at least three in most of the applications (see Remark 3), hence $n^{\text{SMS}} > 2$.[11] Earlier research has shown that electronic platforms like MarketAxess can "provide considerable time savings relative to ... bilateral negotiations" (Hendershott and Madhavan, 2015); and can "improve the speed of execution" (O'Hara and Zhou, 2020), motivating that $\rho^{\text{BB}} \leq \rho^{\text{SMS}}$.

[11] Excluding the special case of $n = 2$ allows us to reduce the number of cases to consider when characterising the equilibrium in Proposition 6 and helps to streamline the exposition.

The probabilities to set prices in respective technology, q^{BB} and q^{SMS}, also play an important role. In most of the applications (e.g., MarketAxess), a customer using RFQ is always on the receiving end of dealers' TIOLIOs, suggesting $q^{\text{SMS}} = 0$. On the other hand, in bilateral calls, there is always room for negotiation and it is natural to expect $q^{\text{BB}} > 0$. We do not impose such constraints here but proceed to study how general values of q^{SMS} and q^{BB} affect the customers' technology choices.

4.1 Choosing between SMS and BB

As in Section 3, the analysis focuses on a stationary equilibrium. We proceed in four steps: (i) We first examine customers' optimal technology choices taking their value functions as given. We then take the technology choices as given and study (ii) the demographics and (iii) the value functions. Finally, (iv) we establish the equilibrium by finding technology choices, demographics, and value functions consistent with each other.

4.1.1 Technology choices

Recall from Section 2 that depending on her inventory holding and her preference, a customer can be categorized into four types, $\sigma \in \{ho, ln, hn, lo\}$. Now the type-$\sigma$ customers can be further split into subtypes σ-BB and σ-SMS. We distinguish these two subtypes by superscripting the relevant variables with their chosen technology $k \in \{\text{BB}, \text{SMS}\}$. For example, their masses satisfy $m_\sigma^{\text{BB}} + m_\sigma^{\text{SMS}} = m_\sigma$ and they have possibly different value functions V_σ^{BB} and V_σ^{SMS}, respectively.

The analysis can be simplified in two ways. First, note that in a stationary equilibrium, the value functions are time-invariant. That is, if a type-σ customer prefers one technology over the other at some point of time t, her technology choice will persist until her type changes (due either to a preference shock or to trading). Hence, without loss of generality, we can focus on a type-σ customer's technology choice at the moment she just becomes type-σ. Second, both ho and ln customers will be bystanders in equilibrium and do not trade—a result following positive trading gains (Lemma 2). Therefore, there is no need to distinguish ln^{SMS} vs. ln^{BB} or ho^{SMS} vs. ho^{BB} and we can focus only on the technology choices of the trading customers, hn and lo.

Denote by $\theta_\sigma \in [0,1]$ the probability of a newborn type-σ customer choosing SMS (hence choosing BB with probability $1-\theta_\sigma$), where $\sigma \in \{hn, lo\}$. Then

$$\theta_\sigma \begin{cases} = \mathbb{1}_{\{V_\sigma^{\text{SMS}} > V_\sigma^{\text{BB}}\}}, & \text{if } V_\sigma^{\text{SMS}} \neq V_\sigma^{\text{BB}}; \\ \in [0,1], & \text{if } V_\sigma^{\text{SMS}} = V_\sigma^{\text{BB}}. \end{cases} \tag{27}$$

We shall focus on *symmetric* equilibria, where all customers of type σ choose the same θ_σ. Where convenient, we will also occasionally write $\theta_\sigma^{\text{SMS}} = 1 - \theta_\sigma^{\text{BB}} := \theta_\sigma$.

4.1.2 Demographics

There are six customer population sizes: $\{m_{ho}, m_{ln}, m_{hn}^{\text{SMS}}, m_{hn}^{\text{BB}}, m_{lo}^{\text{SMS}}, m_{lo}^{\text{BB}}\}$; and in addition, there are two types of dealers, $\{m_{do}, m_{dn}\}$. For notation simplicity, write

$$m_{hn} = m_{hn}^{\text{SMS}} + m_{hn}^{\text{BB}}; \quad \text{and} \quad m_{lo} = m_{lo}^{\text{SMS}} + m_{lo}^{\text{BB}}.$$

Then the four (aggregate) customer masses, $\{m_{ho}, m_{ln}, m_{hn}, m_{lo}\}$, must satisfy the conditions (1)-(4) in Section 3.1. The other four conditions are analogous to the stationarity conditions (5) and (6):

$$\text{net flow of } lo\text{-sellers using SMS:} \quad -\nu_{lo}^{\text{SMS}} m_{lo}^{\text{SMS}} \rho^{\text{SMS}} - \lambda_u m_{lo}^{\text{SMS}} + \theta_{lo}\lambda_d m_{ho} = 0 \tag{28}$$

$$\text{net flow of } lo\text{-sellers using BB:} \quad -\nu_{lo}^{\text{BB}} m_{lo}^{\text{BB}} \rho^{\text{BB}} - \lambda_u m_{lo}^{\text{BB}} + (1-\theta_{lo})\lambda_d m_{ho} = 0 \tag{29}$$

$$\text{net flow of } hn\text{-buyers using SMS:} \quad -\nu_{hn}^{\text{SMS}} m_{hn}^{\text{SMS}} \rho^{\text{SMS}} - \lambda_d m_{hn}^{\text{SMS}} + \theta_{hn}\lambda_u m_{ln} = 0 \tag{30}$$

$$\text{net flow of } hn\text{-buyers using BB:} \quad -\nu_{hn}^{\text{BB}} m_{hn}^{\text{BB}} \rho^{\text{BB}} - \lambda_d m_{hn}^{\text{BB}} + (1-\theta_{hn})\lambda_u m_{ln} = 0 \tag{31}$$

where $\nu_{lo}^k = 1 - \left(1 - \frac{m_{dn}}{m_d}\right)^{n^k}$ and $\nu_{hn}^k = 1 - \left(1 - \frac{m_{do}}{m_d}\right)^{n^k}$ are the probabilities for a customer to find at least one counterparty dealer using technology $k \in \{\text{BB}, \text{SMS}\}$. Compared to Equations (5) and (6) in Section 3.1, the key differences are (i) that every variable here is technology-dependent and superscripted with $k \in \{\text{BB}, \text{SMS}\}$; and (ii) that only a fraction of θ_σ of the newborn σ-customer use SMS, while the rest $(1-\theta_\sigma)$ use BB, where $\sigma \in \{hn, lo\}$.

The conditions (1)-(4) and (28)-(31) exactly pin down the eight demographic variables:

> **Lemma 4 (Stationary demographics with technology choice).** *Given the customers' technology choices $\{\theta_{lo}, \theta_{hn}\} \in [0,1]^2$, Equations (1)-(4) and (28)-(31) uniquely pin down the demographics $\{m_{ho}, m_{ln}, m_{hn}^{SMS}, m_{hn}^{BB}, m_{lo}^{SMS}, m_{lo}^{BB}\} \in [0,1]^6$ and $\{m_{do}, m_{dn}\} \in (0, m_d)^2$.*

The resulting expressions are similar to those implied by Lemma 1. In particular, (28) + (29) − (30) − (31) + (4) gives the trading volume expression

$$t := \sum_k v_{lo}^k m_{lo}^k \rho^k = \sum_k v_{hn}^k m_{hn}^k \rho^k, \tag{32}$$

an analogue to Equation (7) in Section 3, ensuring the stationarity of both dealer types.[12] The h- and l-type customer stationarity (8) also holds the same, and so do the expressions for the total size of trading customers $m_{lo} = \sum_k m_{lo}^k$ and $m_{hn} = \sum_k m_{hn}^k$ as in Equations (9) and (10), respectively.

4.1.3 Value functions

Given the technology choices $\{\theta_\sigma\}$, hence also the demographics, the value functions for all six agent types can be derived analogously to those in Equations (18)-(23). For example, the value functions of an ho-bystander and an ln-bystander must satisfy the HJB equations

$$y_h + \lambda_d \cdot \left(\max\left[V_{lo}^{\text{SMS}}, V_{lo}^{\text{BB}}\right] - V_{ho}\right) - rV_{ho} = 0; \tag{33}$$

$$\lambda_u \cdot \left(\max\left[V_{ho}^{\text{SMS}}, V_{ho}^{\text{BB}}\right] - V_{ln}\right) - rV_{ln} = 0. \tag{34}$$

Compared with Equations (18) and (19), the only difference is that upon a preference shock, a newborn trading customer can choose which technology to use, hence the term of $\max\left[V_\sigma^{\text{SMS}}, V_\sigma^{\text{BB}}\right]$ in the above HJBs ($\sigma \in \{lo, hn\}$).

[12] The stationarity of all other types of agents are also ensured: For example, −(4) − (28) − (29) gives $-\lambda_u m_{ln} + \sum_k \left(v_{lo}^k m_{lo}^k \rho^k + \lambda_d m_{hn}^k\right) = 0$, which ensures the stationarity of ln-bystanders. Likewise, (4) − (30) − (31) gives $-\lambda_d m_{ho} + \sum_k \left(v_{hn}^k m_{hn}^k \rho^k + \lambda_u m_{ln}^k\right) = 0$, which ensures the stationarity of ho-bystanders.

The HJB equations for the trading agents are also similar to before:

$$\text{(35)} \quad \text{HJB of } lo\text{-sellers using technology } k: \quad y_l + \lambda_u \cdot (V_{ho} - V_{lo}^k) - rV_{lo}^k + \zeta_{lo}^k \Delta_{dl}^k = 0;$$

$$\text{(36)} \quad \text{HJB of } hn\text{-buyers using technology } k: \quad \lambda_d \cdot (V_{ln} - V_{hn}^k) - rV_{hn}^k + \zeta_{hn}^k \Delta_{hd}^k = 0;$$

$$\text{(37)} \quad \text{HJB of } do\text{-dealers:} \quad y_d - rV_{do} + \sum_k \zeta_{do}^k \Delta_{hd}^k = 0;$$

$$\text{(38)} \quad \text{HJB of } dn\text{-dealers:} \quad -rV_{dn} + \sum_k \zeta_{dn}^k \Delta_{dl}^k = 0.$$

Compared to Equations (20)-(23), the only difference is that the trading gains $\{\Delta_{hd}, \Delta_{dl}\}$ and the trading gain intensities $\{\zeta_{lo}, \zeta_{hn}, \zeta_{do}, \zeta_{dn}\}$ are technology specific, superscripted with $k \in \{\text{BB}, \text{SMS}\}$. For completeness, we derive these expressions below.

Using technology k, an lo-seller's reservation value is $R_l^k := V_{lo}^k - V_{ln}$, and that for an hn-buyer is $R_h^k := V_{ho} - V_{hn}^k$. A dealer's reservation value is the same $R_d := V_{do} - V_{dn}$ as before. Then, depending the customer's technology k, the trading gain between an hn-buyer and a do-seller is $\Delta_{hd}^k := R_h^k - R_d$ and that between a dn-buyer and an lo-seller is $\Delta_{dl}^k := R_d - R_l^k$. By Corollary 1, the dealers' respective average ask and bid are:

$$\bar{A}^k = \frac{n^k \frac{m_{do}}{m_d}\left(1 - \frac{m_{do}}{m_d}\right)^{n^k-1}}{1 - \left(1 - \frac{m_{do}}{m_d}\right)^{n^k}} \quad \text{and} \quad \bar{B}^k = \frac{n^k \frac{m_{dn}}{m_d}\left(1 - \frac{m_{dn}}{m_d}\right)^{n^k-1}}{1 - \left(1 - \frac{m_{dn}}{m_d}\right)^{n^k}}.$$

Thus, an hn-buyer expects $\zeta_{hn}^k \Delta_{hd}^k$, while a do-dealer expects $\zeta_{do}^k \Delta_{hd}^k$, where the respective trading gain intensities are

$$\zeta_{hn}^k = \rho^k v_{hn}^k \cdot \left(q^k + (1 - q^k)(1 - \bar{A}^k)\right) \text{ and } \zeta_{do}^k = \frac{m_{hn}^k \rho^k v_{hn}^k}{m_{do}}(1 - q^k)\bar{A}^k.$$

Analogously, an lo-seller expects $\zeta_{lo}^k \Delta_{dl}^k$, while a dn-dealer expects $\zeta_{dn} \Delta_{dl}^k$, with intensities

$$\zeta_{lo}^k = \rho^k v_{lo}^k \cdot \left(q^k + (1 - q^k)(1 - \bar{B}^k)\right) \text{ and } \zeta_{dn}^k = \frac{m_{hn}^k \rho^k v_{lo}^k}{m_{dn}}(1 - q^k)\bar{B}^k.$$

Corollary 2 (Positive trading gains). *When $\overline{y}_d \geq y_d \geq \underline{y}_d$ as defined in Proposition (2), there is strictly positive gains from trade, i.e., $R_d \in (R_l^k, R_h^k)$, for both $k \in \{BB, SMS\}$.*

The corollary ensures that the same condition for y_d as before is sufficient to guarantee positive trading gains regardless of the equilibrium technology choices $\{\theta_\sigma\}$. Note that the positive trading gains also ensures that both the $ho-$ and ln-customers do stay out of trading, following Lemma 2.

4.1.4 Equilibrium

A steady-state equilibrium is characterized by the following four sets of objects:

(i) trading customers' technology choices $\{\theta_{lo}, \theta_{hn}\}$;

(ii) agent demographics $\{m_{ho}, m_{ln}, m_{hn}^{\text{SMS}}, m_{hn}^{\text{BB}}, m_{lo}^{\text{SMS}}, m_{lo}^{\text{BB}}, m_{do}, m_{dn}\}$;

(iii) dealers' pricing $\{\bar{A}^{\text{BB}}, \bar{B}^{\text{BB}}, \bar{A}^{\text{SMS}}, \bar{B}^{\text{SMS}}\}$; and

(iv) agents' value functions $\{V_{ho}, V_{ln}, V_{hn}^{\text{SMS}}, V_{hn}^{\text{BB}}, V_{lo}^{\text{SMS}}, V_{lo}^{\text{BB}}, V_{do}, V_{dn}\}$.

As a quick recap, we started by fixing the technology choices at the end of Section 4.1.1 and then solved the demographics in Section 4.1.2. The dealers' pricing strategies remain the same as in Proposition 1. The value functions are pinned down by Equations (33)-(38) in Section 4.1.3.

To establish an equilibrium, we now need to examine when the technology choices $\{\theta_{hn}, \theta_{lo}\}$ are consistent with the value functions $\{V_\sigma\}$ according to Equation (27)—a fixed-point problem. This is a seemingly daunting task, because the value functions $\{V_\sigma\}$ are chained to the technology choices $\{\theta_\sigma\}$ via many layers of endogenous variables: the trading gain intensities ζs, the dealers' pricing $\bar{A}$ and $\bar{B}$, and the many demographic variables $\{m_\sigma\}$. Below we walk through these layers and show that solving for the equilibrium $\{\theta_{hn}, \theta_{lo}\}$ ultimately boils down to comparing the sizes of dealers, m_{do} (and $m_{dn} = m_d - m_{do}$) with some threshold:

Lemma 5 (Three key endogenous variables: $\{\theta_{lo}, \theta_{hn}, m_{do}\}$). *If the technologies satisfy*

$$\rho^{SMS} q^{SMS} n^{SMS} < \rho^{BB} q^{BB} n^{BB}, \tag{39}$$

then Equation (27) can be equivalently written as

$$(40)\qquad \theta_{hn}\begin{cases} = \mathbb{1}_{\{m_{do}>m_d\mu^*\}}, & \text{if } m_{do}\neq m_d\mu^* \\ \in[0,1], & \text{if } m_{do}= m_d\mu^*\end{cases} \text{ and } \theta_{lo}\begin{cases} = \mathbb{1}_{\{m_{dn}>m_d\mu^*\}}, & \text{if } m_{dn}\neq m_d\mu^* \\ \in[0,1], & \text{if } m_{dn}= m_d\mu^*\end{cases},$$

where $\mu^* \in \left(0, \frac{1}{2}\right)$ *uniquely solves* $z^{SMS}(\mu) = z^{BB}(\mu)$*, with* $z^k(\cdot)$ *defined in (42) below for* $k \in \{SMS, BB\}$*. If instead* $\rho^{SMS}q^{SMS}n^{SMS} \geq \rho^{BB}q^{BB}n^{BB}$*, then* $\theta_{hn} = \theta_{lo} = 1$*.*

Below we show the key steps behind the Lemma. We begin by comparing a trading customer's value functions under the two technologies, V_σ^{SMS} and V_σ^{BB}. Consider, for example, an *lo*-seller. Substitute $\Delta_{dl}^k = R_d - R_l^k = R_d - (V_{lo}^k - V_{ln})$ in her value function (35) to get

$$y_l + \lambda_u \cdot (V_{ho} - V_{lo}^k) - rV_{lo}^k + \Big(R_d - (V_{lo}^k - V_{ln})\Big)\zeta_{lo}^k = 0 \implies V_{lo}^k = \frac{y_l + \lambda_u V_{ho} + (R_d + V_{ln})\zeta_{lo}^k}{r + \lambda_u + \zeta_{lo}^k}.$$

The above highlights that the only difference between V_{lo}^{SMS} and V_{lo}^{BB} is the respective trading gain intensities ζ_{lo}^k. Clearly, V_{lo}^k is strictly increasing in ζ_{lo}^k. Likewise, V_{hn}^k is also strictly increasing in ζ_{hn}^k. Hence, the technology choice (27) can be equivalently written as, for $\sigma \in \{lo, hn\}$:

$$(41)\qquad \theta_\sigma\begin{cases} = \mathbb{1}_{\{\zeta_\sigma^{\text{SMS}} > \zeta_\sigma^{\text{BB}}\}}, & \text{if } \zeta_\sigma^{\text{SMS}} \neq \zeta_\sigma^{\text{BB}}; \\ \in [0,1], & \text{if } \zeta_\sigma^{\text{SMS}} = \zeta_\sigma^{\text{BB}}.\end{cases}$$

To ease the notations, define on the support of $\mu \in (0, 1)$

$$(42)\qquad z^k(\mu) := \Big(1 - (1-\mu)^{n^k-1}\Big(1 - \mu + (1-q^k)n^k\mu\Big)\Big)\rho^k.$$

Then the trading gain intensities of using technology k can be written as $\zeta_{hn}^k = z^k\left(\frac{m_{do}}{m_d}\right)$ and $\zeta_{lo}^k = z^k\left(\frac{m_{dn}}{m_d}\right)$. Note that $z^k(\cdot)$ is parametrized only by the exogenous technology parameters $\{\rho^k, n^k, q^k\}$. Hence, the technology choices $\{\theta_\sigma\}$ boil down to whether and how the functions $z^{\text{SMS}}(\mu)$ and $z^{\text{BB}}(\mu)$ cross each other.

Lemma 5 has characterized such crossing: Under the condition (39), $z^{\text{SMS}}(\mu)$ crosses $z^{\text{BB}}(\mu)$ from below once at $\mu^* \in \left(0, \frac{1}{2}\right)$. That is, a customer prefers BB over SMS when $\mu < \mu^*$. This might come as a surprise, given that the condition (26) guarantees that SMS not only helps reach dealers faster but also induces more competitive quotes. Why would a customer still prefer BB?

To see the potential advantage of BB, consider for example an *hn*-buyer looking for *do*-sellers but m_{do} is very low. In this case, if we write $\mu = m_{do}/m_d$, then, using technology-k, the *hn*-buyer customer finds one counterparty dealer with probability approximately $n^k\mu$—one and only one success from n^k Bernoulli draws at rate μ. (As μ is small, the event of finding multiple dealers is negligibly unlikely.) It follows that a successfully contacted dealer in this case knows that she is almost surely a monopolist and will always quote an ask as high as possible, i.e., $\bar{A} \uparrow 1$, leaving no trading gains to the *hn*-buyer. Hence, the customer gets non-zero trading gains only if she can make a TIOLIO, i.e., with probability q^k. Taken together, for small μ, the customers' trading gain intensity is $z^k(\mu) \approx \rho^k \cdot (n^k\mu) \cdot q^k$. Comparing BB with SMS in this case yields:

$$\lim_{\mu \downarrow 0} \frac{z^{\mathrm{BB}}(\mu)}{z^{\mathrm{SMS}}(\mu)} = \frac{\rho^{\mathrm{BB}} n^{\mathrm{BB}} q^{\mathrm{BB}}}{\rho^{\mathrm{SMS}} n^{\mathrm{SMS}} q^{\mathrm{SMS}}}.$$

The condition (39), therefore, ensures that for sufficiently small μ, i.e., for relatively few counterparty dealers, BB has an advantage over SMS. In real-world trading, the condition (39) seems to hold because customers using SMS, like RFQ protocols, do not have many opportunities, if at all, to set reserve prices. That is, q^{SMS} is observed to be sufficiently low in the real world.[13]

We are now ready to state the equilibrium.

Proposition 6 (Steady state equilibrium with technology choices). *A unique stationary equilibrium exists depending on the asset supply s: There exist thresholds* $0 < s_{hn,0} < s_{hn,1} \leq s_{lo,1} < s_{lo,0} < 1 + m_d$ *so that*

[13] Complementing the condition (39), the condition (26) in turn ensures that SMS is preferred when there are sufficiently many dealer counterparties.That is, $\lim_{\mu \uparrow 1} \left(z^{\mathrm{BB}}(\mu)/z^{\mathrm{SMS}}(\mu) \right) = \rho^{\mathrm{BB}} q^{\mathrm{BB}} / \rho^{\mathrm{SMS}} \leq 1$. It is interesting to note that only q^{BB} appears but not q^{SMS} in the limit of $\mu \uparrow 1$. With $n^{\mathrm{SMS}} > 1$ and $\mu \uparrow 1$, the multiple counterparty dealers in SMS almost always engage in Bertrand competition, and the customer always gets the full trading gain, regardless of q^{SMS}. On the contrary, with $n^{\mathrm{BB}} = 1$, a customer using BB meets only one counterparty dealer, who will always set the monopolist price, leaving surplus to the customer only with probability q^{BB}.

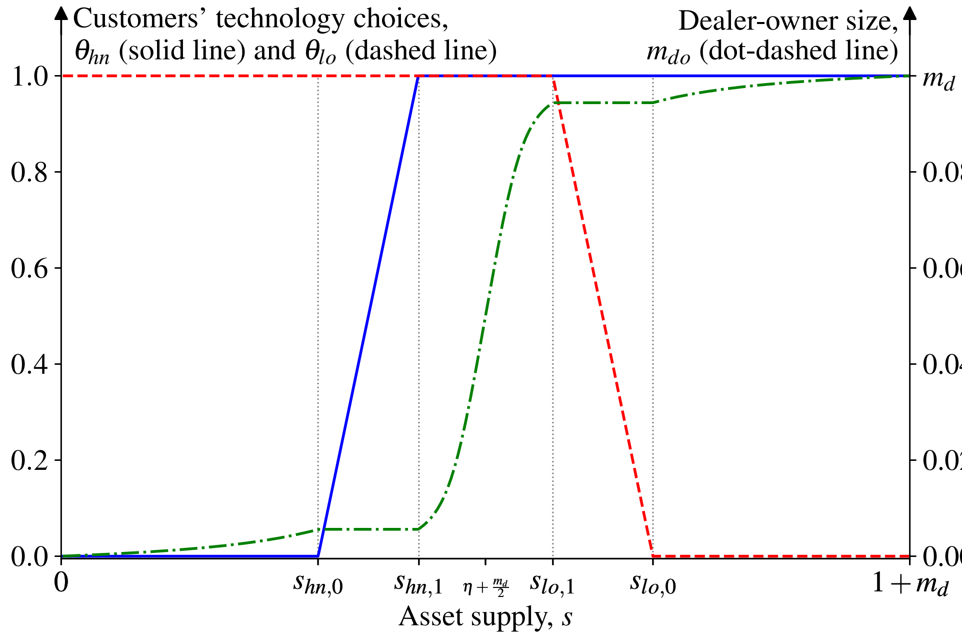

Figure 4: Equilibrium technology choice plotted against asset supply. This figure plots customers' technology choices against the asset supply s in equilibrium. The hn-buyers' choice θ_{hn} is plotted in the solid line, while the lo-sellers' choice θ_{lo} is plotted in the dashed line (the left axis). The dot-dashed line plots the population size of do-seller dealers (the right axis). The technology parameters are set at $(\rho^{BB}, n^{BB}, q^{BB}) = (3.0, 1, 0.5)$ and $(\rho^{SMS}, n^{SMS}, q^{SMS}) = (3.0, 5, 0.0)$. The other parameters are $\lambda_d = \lambda_u = 0.1$, $r = 0.1$, $m_d = 0.1$, $y_h = 10.0$, $y_d = 3.5$, and $y_l = 0.0$.

	(a) hn-buyers' proba -bility to use SMS	*(b) lo-sellers' proba -bility to use SMS*	*(c) asset holding by dealers*
(1) $0 < s \leq s_{hn,0}$	$\theta_{hn} = 0$	$\theta_{lo} = 1$	$g(0, 1, m_{do}) = 0$
(2) $s_{hn,0} \leq s \leq s_{hn,1}$	$g(\theta_{hn}, 1, \mu^* m_d) = 0$	$\theta_{lo} = 1$	$m_{do} = \mu^* m_d$
(3) $s_{hn,1} < s < s_{lo,1}$	$\theta_{hn} = 1$	$\theta_{lo} = 1$	$g(1, 1, m_{do}) = 0$
(4) $s_{lo,1} \leq s \leq s_{lo,0}$	$\theta_{hn} = 1$	$g(1, \theta_{lo}, (1-\mu^*)m_d) = 0$	$m_{do} = (1-\mu^*)m_d$
(5) $s_{lo,0} < s < 1 + m_d$	$\theta_{hn} = 1$	$\theta_{lo} = 0$	$g(1, 0, m_{do}) = 0$

where $g(x_1, x_2, x_3) = s$ *uniquely solves* θ_{hn}*,* θ_{lo}*, and* m_{do} *in column (a), (b), and (c), respectively.*

The function $g(\cdot)$ *and the the thresholds* $\{s_{hn,0}, s_{hn,1}, s_{lo,1}, s_{lo,0}\}$ *are given in the proof.*

Figure 4 illustrates the equilibrium by plotting the technology choices θ_{hn} (solid) and θ_{lo}

(dashed) on the left-axis and the dealer-owner population size m_{do} (dotted) on the right-axis. The four thresholds of $\{s_{hn,0}, s_{hn,1}, s_{lo,1}, s_{lo,0}\}$ cut the support of $s \in (0, 1 + m_d)$ into five regions in the horizontal axis. Consider the blue solid line, i.e., θ_{hn}, for example. When the asset supply s is extremely low, SMS is very unattractive for the hn-buyers, because they know it is very difficult to find a counterparty do-dealer (the green dot-dashed line), and even if they do, they are going to be charged with a monopoly price of $\bar{A} \uparrow 1$. When s is sufficiently high, there are sufficiently many do-dealers, whose price competition makes SMS sufficiently attractive with high trading gain intensity ζ_{hn}^{SMS} for hn-buyers. As such, the blue solid line flattens at $\theta_{hn} = 1$ for $s > s_{hn,1}$. In between, we see θ_{hn} monotonically increases for $s_{hn,0} \leq s \leq s_{hn,1}$. Such a mixed-strategy is supported by the constant $m_{do} = \mu^* m_d$ in the region—the hn-buyers are indifferent between BB and SMS. The pattern for the red dashed line, i.e., θ_{lo}, is exactly the opposite, as lo-sellers seek dn-dealers, whose mass is $m_{dn} = m_d - m_{do}$.

4.2 Stress periods

O'Hara and Zhou (2020) show that after downgrade, a corporate bond's electronic (SMS) volume share falls relative to voice trading (BB). The analysis developed above provides a theoretical framework to study investors' endogenous technology choice when under such stress.

One consequence of a corporate bond downgrade is that many previously buy-and-hold long-term investors now no longer wish to hold such bonds. Ambrose, Cai, and Helwege (2008) and Ellul, Jotikasthria, and Lundblad (2011) document such fire sales by insurance companies. In the context of our model, we interpret such fire selling in two different ways, (i) an exogenous increase in the total supply s of the asset and / or (ii) an exogenous increase in the customers' intensity of drawing low preference λ_d. Effectively, (i) is a supply shock and (ii) is a demand shock.[14] To fit the fire-selling interpretation, we also assume that the asset is in excess supply as defined in Lemma 3.

[14] As we will show shortly, such a supply shock and a demand shock are essentially equivalent. A third alternative, reducing λ_u, is also equivalent and omitted for brevity.

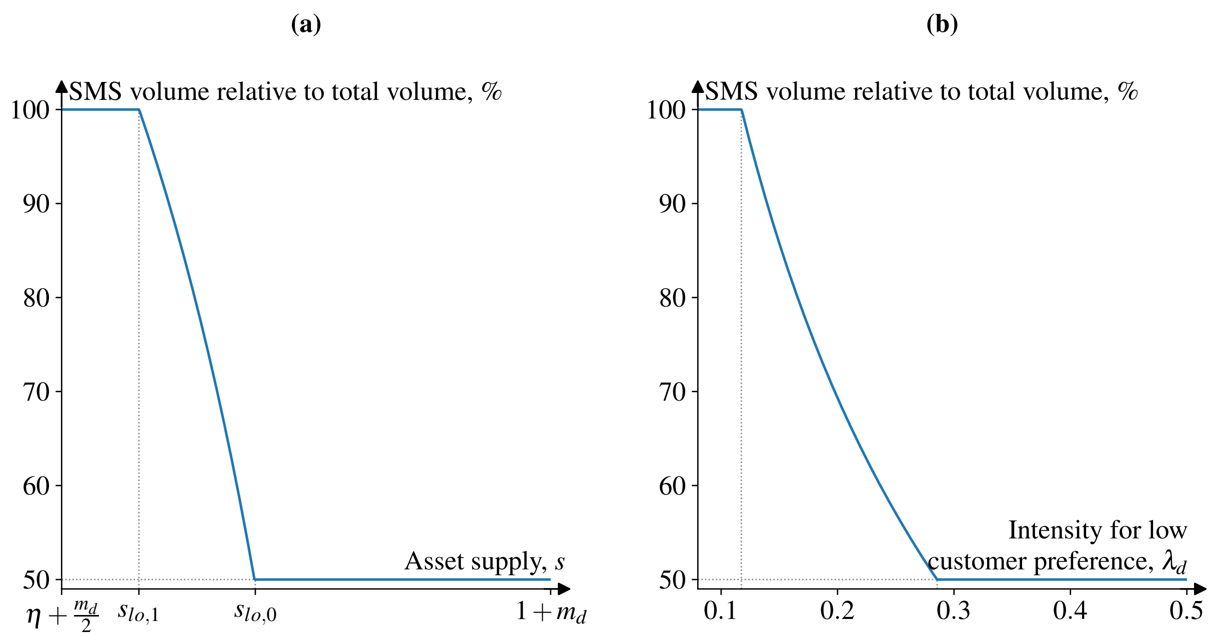

Figure 5: Usage of SMS in a stationary equilibrium after surges in supply. This figure plots the usage of SMS (in a stationary equilibrium)—SMS volume relative to total volume—when the asset supply s surges in Panel (a) and when the customers' low-valuation preference shock intensity λ_d increases in Panel (b). The technology parameters are set at $(\rho^{\text{BB}}, n^{\text{BB}}, q^{\text{BB}}) = (3.0, 1, 0.5)$ and $(\rho^{\text{SMS}}, n^{\text{SMS}}, q^{\text{SMS}}) = (3.0, 5, 0.0)$. The other parameters are $\lambda_d = 0.1$ (for Panel a), $\lambda_u = 0.1$, $r = 0.1$, $m_d = 0.1$, $s = 0.6$ (for Panel b), $y_h = 10.0$, $y_d = 3.5$, and $y_l = 0.0$.

Figure 5(a) and (b) below illustrate how the SMS volume share responds to shocks in s and λ_d, respectively. The SMS volume share is defined as:

$$\text{(43)} \qquad \frac{\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}} + \rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}}{\left(\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}} + \rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}\right) + \left(\rho^{\text{BB}} m_{lo}^{\text{BB}} v_{lo}^{\text{BB}} + \rho^{\text{BB}} m_{hn}^{\text{BB}} v_{hn}^{\text{BB}}\right)}.$$

In Panel (a), the volume ratio is initially flat at 100% because both lo-sellers and hn-buyers always use SMS ($\theta_{hn} = \theta_{lo} = 1.0$). As the supply s rises higher (between $s_{lo,1}$ and $s_{lo,0}$), lo-sellers start to use less SMS, resulting in the decreasing segment. As s increases further, there are no more lo-sellers using SMS—all of them use BB, while all hn-buyers use SMS. That is, $m_{lo}^{\text{SMS}} = m_{hn}^{\text{BB}} = 0$.

In this case, the SMS volume ratio above reduces to

$$\frac{\rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}}{\rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}} + \rho^{\text{BB}} m_{lo}^{\text{BB}} v_{lo}^{\text{BB}}} = \frac{t}{2t} = 50\%,$$

where the second equality follows the trading volume expressions (32). Overall, the SMS volume ratio drops with the decline of the SMS usage θ_{lo}, as seen before in Figure 4. The same pattern is observed from Panel (b), where we increase the customers' negative preference shock intensity λ_d, effectively reducing the demand for the asset. The proposition below summarizes the result formally.

Proposition 7 (SMS usage under stress). *In a steady state equilibrium, the usage of SMS decreases as either the asset's excess supply (demand) surges. That is, all else equal, for* $s > s_{hn,1}$ *(*$s < s_{lo,0}$*), the ratio defined in* (43) *weakly decreases when* s *increases (decreases) or when* λ_d *increases (decreases).*

The proposition also gives the mirroring result: When the asset's excess demand exacerbates ($s < \eta + \frac{m_d}{2}$), SMS usage also drops.

The key intuition for the decrease of the SMS volume share can be understood from the worsening pricing for the *lo*-sellers. As the asset supply s increases after the fire sell, there are fewer and fewer *dn*-dealers (see the dot-dashed line in Figure 4 and note that $m_{dn} = m_d - m_{do}$). Facing less competition, therefore, the *dn*-buyers will charge worse and worse prices to the *lo*-sellers in SMS. Expecting such worsening prices from SMS, the *lo*-sellers then avoid using SMS and switch to BB. In particular, our model yields an additional prediction regarding prices in SMS and in BB under a fire sell:

Corollary 3 (Prices in SMS vs. in BB under fire sell). *When there is excess supply, an lo-seller's expected trading price using SMS worsens relative to using BB. That is, the ratio* $(\bar{B}^{SMS}/\bar{B}^{BB})$ *is weakly increasing in* s *and in* λ_d*, where* $\bar{B}^k$ *reduces the lo-seller's expected selling price* $R_d - \bar{B}^k \Delta_{dl}$ *as in Equation* (17).

Therefore, one way to empirically test our theory is to compare the trading prices in BB and in

SMS when the asset is under fire sell and examine if the price in SMS is worse than that in BB.

It is worth emphasizing that only the *stationary* equilibrium is studied. Hence, the above results should be read as comparisons of steady states before and after corporate bond downgrades.

4.3 Efficiency and welfare

In this subsection, we ask whether the market's equilibrium technology choices are socially optimal: Given the technologies $\{n^k, \rho^k, q^k\}$, $k \in \{\text{BB}, \text{SMS}\}$, how will a social planner assign BB and SMS to the customers? When, if at all, will the market's equilibrium choices $\{\theta_{lo}, \theta_{hn}\}$ coincide with the planner's $\{\theta^*_{lo}, \theta^*_{hn}\}$? What are the implications for policies and market design?

It turns out that the answers critically depend on the characteristics of the asset. Among others, how quickly can customers find dealers, i.e., $\{\rho^{\text{BB}}, \rho^{\text{SMS}}\}$, matters a lot. We discuss the high-ρ and the low-ρ cases separately below.

4.3.1 The case of high search intensity

Proposition 8 (A social planner's technology choices). *When the search intensity* $\rho := \min[\rho^{SMS}, \rho^{BB}]$ *is sufficiently high, welfare* w *is monotone increasing in SMS usages by both types of customers and the social planner chooses* $\theta^*_{lo} = \theta^*_{hn} = 1$.

The intuition largely follows Proposition 5. When the search intensity is high, Proposition 5 shows that welfare is monotone increasing in n. As such, by assigning both $\theta^*_{lo} = \theta^*_{hn} = 1$, the planner essentially chooses n^{SMS} over n^{BB} to maximize welfare.

However, the market's technology choices do not always coincide with the planner's. This is because a searching customer cares not only about the probability of finding a counterparty dealer but also about the endogenous split of the trading gain with the dealer. Figure 6 sketches such possible discrepancies. The solid line and the dashed line plot, respectively, the market's choices of θ_{hn} and θ_{lo} against the asset supply s. Note that the patterns for the θs are qualitatively the same

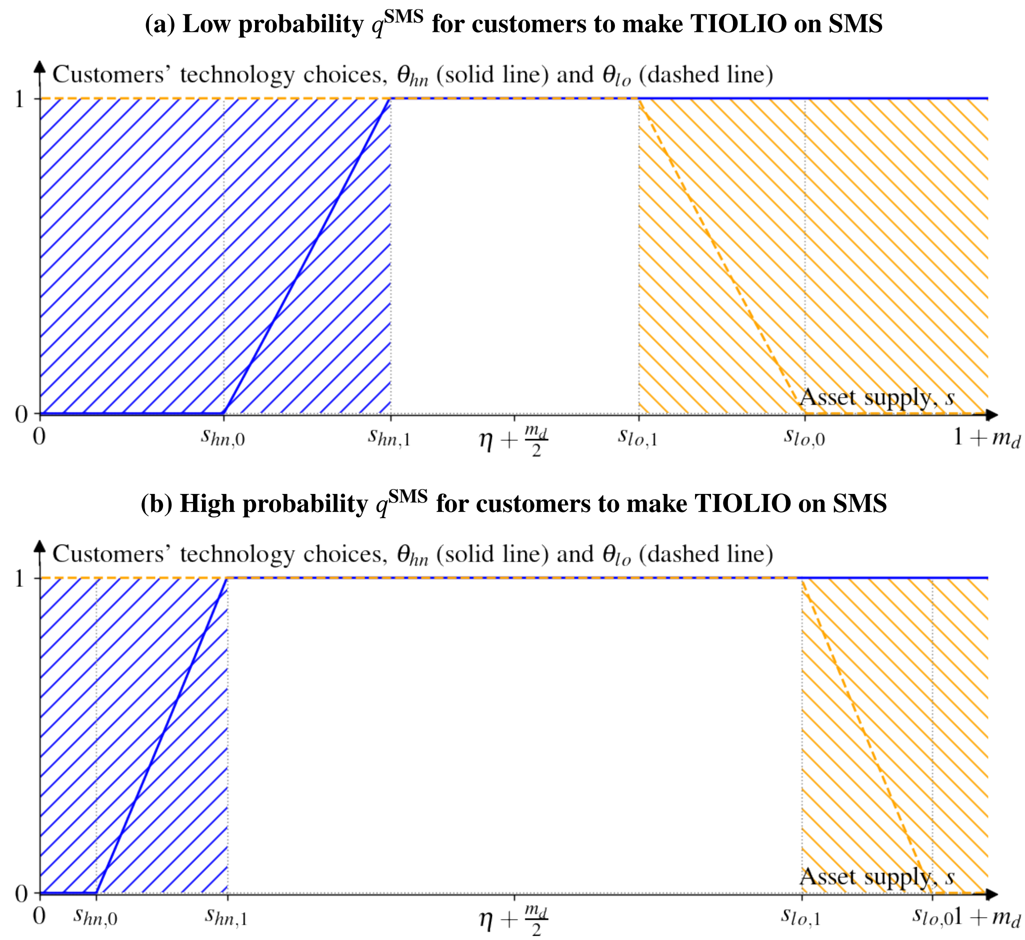

Figure 6: Market's technology choices vs. a social planner's under high search intensity. This figure sketches the inefficiency due to the difference between the market's equilibrium technology choices and a social planner's when the search intensity $\rho := \min[\rho^{\text{BB}}, \rho^{\text{SMS}}]$ is high. The solid (blue) line and the dashed (orange) line are θ_{hn} and θ_{lo}, respectively, the hn-buyers' and the lo-sellers' equilibrium probabilities of using SMS. The "//"(blue) and "\\" (orange) shaded areas indicate, respectively, where θ_{hn} and θ_{lo} differ from the planner's corresponding choices θ^*_{hn} and θ^*_{lo}. Panel (a) shows the patterns for low q^{SMS}, while Panel (b) shows for a higher q^{SMS}.

as seen in Figure 4. The shaded areas indicate that there is inefficiency in the market's technology choices. For example, when the excess supply s is relatively extreme $s > s_{lo,1}$, as in the case of fire sell in Section 4.2, the dealer sector becomes overloaded (m_{do} too large), giving lo-sellers a hard time finding dn-dealers. They then become unwilling to use SMS (θ_{lo} decreases with s) because in

SMS their trading gains is too low. The same holds when the excess demand is relatively extreme, i.e., when $s < s_{hn,1}$.

Since the planner wants to encourage SMS usage, a simple, welfare-improving market design mandate seems to readily follow: Let customers indicate their reservation values, i.e., make TIOLIOs when searching dealers via SMS. In our model, such a design translates to an increase in q^{SMS} and by Lemma **??**, this raises the customers' trading gain intensities $\zeta_{\sigma}^{\text{SMS}}$ and therefore induces higher θ_{hn}^{SMS} and θ_{lo}^{SMS} (by Equation 41). Indeed, this is what we find by contrasting Figure 6(a) with 6(b), low q^{SMS} vs. high q^{SMS}. The shaded area of the market's inefficient technology adoption is reduced. (Note that a change in q^k does not affect welfare as only the split of trading gain between customers and dealers is affected, not the size of the pie.)

In practice, however, customers are almost always on the receiving end of TIOLIOs on electronic platforms; i.e., q^{SMS} tends to be zero. We argue that one reason behind such an inefficient design is the dealers' incentive to participate. Recall from Section 4.1.3 that

$$\zeta_{do}^{\text{SMS}} \propto (1 - q^{\text{SMS}}) \text{ and } \zeta_{dn}^{\text{SMS}} \propto (1 - q^{\text{SMS}}).$$

That is, a higher q^{SMS} tilts the split of trading gains away from dealers to customers. Therefore, to the extent that the dealers have certain influence on the design of trading protocols on the electronic platforms, they would avoid choosing a high q^{SMS}, or perhaps not at all let customers make TIOLIOS.[15]

4.3.2 The case of low search intensity

The case of low search intensity $\rho = \max[\rho^{\text{SMS}}, \rho^{\text{BB}}]$ is more nuanced. The planner's choices in addition depend on the comparison between dealers' instantaneous utility y_d and an average customer's $\hat{y} := \eta y_h + (1 - \eta) y_l$:

[15] Even if the dealers are independent of the trading protocol design, the platform operator will have to incentivize dealers' endogenous participation, without which the platform will not run. The dealers' endogenous participation in SMS is not modeled in the current paper.

Proposition 9 (A social planner's technology choices). *When the search intensity* $\rho := \max[\rho^{SMS}, \rho^{BB}]$ *is sufficiently low, the social planner chooses* $\theta^*_{lo} = 1 - \theta^*_{hn} = \mathbb{1}_{\{y_d > \hat{y}\}}$ *to maximize welfare.*

To see why, recall from Equation (25) that there are two components in welfare w, (i) the steady state allocation among agents and (ii) the trading gains. The latter component is proportional to the trading volume t, which diminishes when $\rho = \max[\rho^{\text{BB}}, \rho^{\text{SMS}}]$ is sufficiently low. That is, trading is no longer the top priority for the planner—the asset allocation is. Effectively, only component (i) remains:

$$w \approx \frac{\hat{y}}{r}(s - m_{do}) + \frac{y_d}{r} m_{do}.$$

The planner, therefore, wants to maximize (minimize) m_{do}, i.e., to shift as much asset holding as possible to dealers (customers), if and only if $y_d > \hat{y}$ ($y_d < \hat{y}$). To do so, the planner will polarize $\{\theta^*_{lo}, \theta^*_{hn}\}$ because they affect m_{do} in opposite directions: If more *lo*-sellers use SMS, *dn*-dealers get to buy more often and become *do*-dealers; but if more *hn*-sellers use SMS, more *do*-dealers get to sell their assets and become *dn*-dealers. As a result, $\theta^*_{lo} = \mathbb{1}_{\{y_d > \hat{y}\}}$ and $\theta^*_{hn} = 1 - \theta^*_{lo}$.

Figure 7(a) sketches the case of $y_d < \hat{y}$. In this case, the planner wants to allocate the asset to the customers as much as possible, thus assigning $\theta^*_{hn} = 1$ and $\theta^*_{lo} = 0$. This is clearly against the *lo*-sellers' incentive, as they want to sell the asset to the dealers. As a result, the market's technology choices are efficient (coinciding with the planner's) only when the asset is in extreme supply, i.e., when $s > s_{lo,0}$. Panel (a) flips Panel (b) with $y_d > \hat{y}$.

The patterns shown in Figure 7 warns that the intuition regarding welfare and market design obtained from the high-ρ case does not carry through when the matching of the asset is intrinsically slow. For example, compared to corporate bonds, whose matching on MarketAxess take only a few minutes (Hendershott and Madhavan, 2015), collateralized loan obligations (CLOs) trade much more slowly, taking days as the B/OWIC run through emails require considerably longer time to organize (Hendershott et al. (2020)). For such "slow" assets, the planner always wants

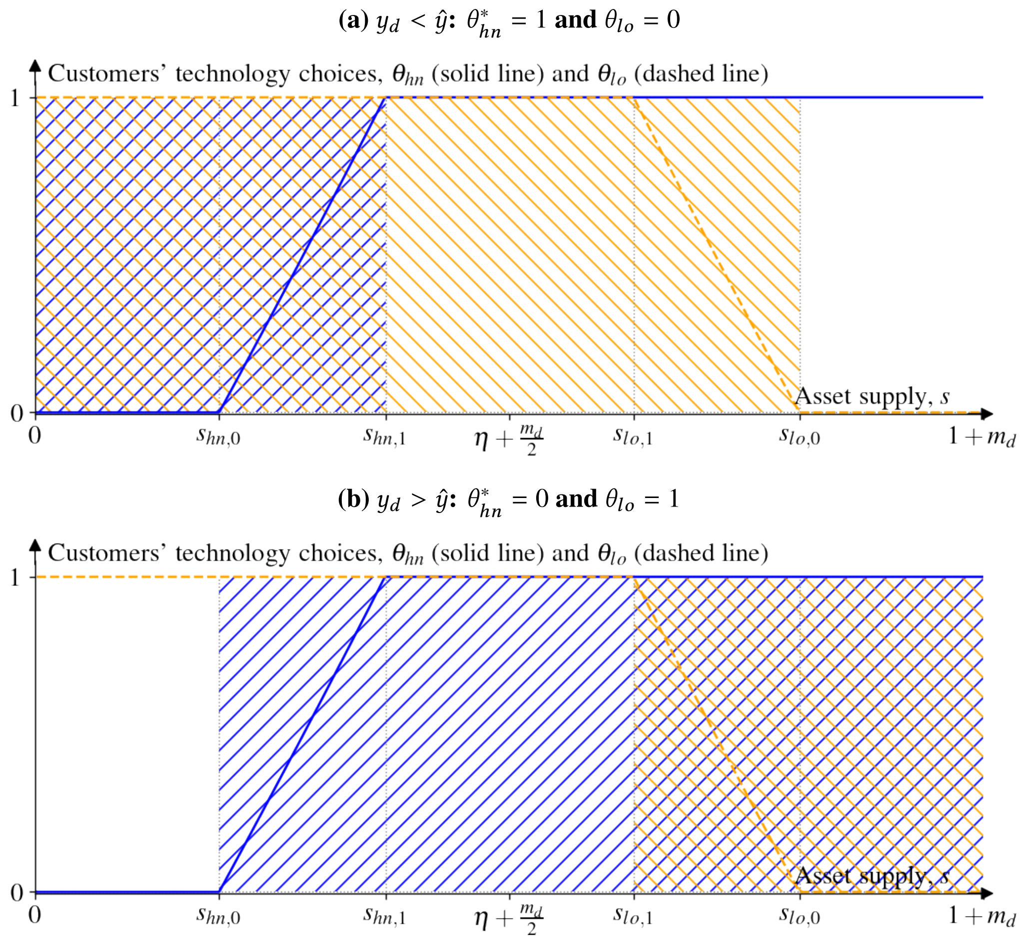

Figure 7: Market's technology choices vs. a social planner's under low search intensity. This figure sketches the inefficiency due to the difference between the market's equilibrium technology choices and a social planner's when the search intensity $\rho := \min[\rho^{\text{BB}}, \rho^{\text{SMS}}]$ is low. The solid (blue) line and the dashed (orange) line are θ_{hn} and θ_{lo}, respectively, hn-buyers' and lo-sellers' equilibrium probabilities of using SMS. The "//"(blue) and "\\" (orange) shaded areas indicate, respectively, where θ_{hn} and θ_{lo} deviate from the planner's corresponding choices θ^*_{hn} and θ^*_{lo}. Panel (a) shows the pattern for the case of $y_d < \hat{y}$, in which case $\theta^*_{hn} = 1$ and $\theta^*_{lo} = 0$, and Panel (b) the opposite, in which case $\theta^*_{hn} = 0$ and $\theta^*_{lo} = 1$.

some customers to use BB to prevent the asset from being held inefficiently in the wrong hands. In particular, allowing customers to make TIOLIOs on SMS (high q^{SMS}) no longer induces the socially optimal technology choices.

Along the same line, our analysis also cautions of regulations that might affect the search intensity ρ. With the increasing scrutiny from regulators after Dodd-Frank, dealers' compliance burden arguably has slowed down their responses to RFQs in SMS. Such changes in search intensity might hurt welfare, as the previously efficient market technology choices under high ρ might no longer be so under a low ρ.

5 Conclusion

This paper studies "simultaneous multilateral searching" (SMS), which has been popularized in practice recently through trading protocols like "Request-for-Quote" (RFQ) in OTC markets. The idea is that a searching customer can reach out to multiple dealers simultaneously, solicit quotes from them, and then trade with the one offering the best quote. This search mechanism differs from the conventional "bilateral bargaining" (BB), in which a searching customer spends effort negotiating terms with a single dealer.

A steady state equilibrium is characterized in a standard framework of the search literature (Hugonnier, Lester, and Weill, 2020). The key insight revealed is that the split of the trading gain between a searching and a quoting investor is an endogenous equilibrium outcome, as opposed to the exogenous split (à la Nash) in the literature assuming BB. In addition, two search parameters, the intensity and the capacity, are analyzed in terms of their, sometimes contrasting, implications for market quality. A novel bottleneck effect arises from, and only from, the search capacity is shown to hinder the efficient asset allocation and might possibly hurt welfare.

Allowing customers to endogenously choose between SMS and BB, the model finds an intrinsic hindrance in the adoption of SMS and further suggests potential inefficiency in terms of asset allocation. The model suggests channels through which both regulation (e.g., complexity of compliance) and market design (RFQ protocols) can affect customers' search preferences and, ultimately, the asset allocation efficiency.

Appendix

A Collection of proofs

Lemma 1 and 4

Proof. The proof considers the general case of Lemma 4 with arbitrary θ_{hn} and θ_{lo}. Lemma 1 is then just a special case of $\theta_{hn} = \theta_{lo} = 1$. The idea is to first express all other unknowns as monotone functions of m_{do}. The existence and the uniqueness then follow as long as the solution to m_{do} exists and is unique. To begin with, add up (28) and (29) to get

$$\text{(A.1)} \qquad \text{all } lo\text{-seller stationarity:} \qquad -\lambda_u m_{lo} + \lambda_d m_{ho} - \rho m_{lo} v_{lo} = 0,$$

where $m_{lo} := \sum_k m_{lo}^k$ is the total lo-seller mass, $\rho := \max\left[\rho^{\text{SMS}}, \rho^{\text{BB}}\right]$, $v_{lo} := \frac{1}{\rho m_{lo}} \sum_k \rho^k m_{lo}^k v_{lo}^k$ is the (weighted) average matching rate for an lo-seller, and $m_{ho} := \sum_k m_{ho}^k$ is the total ho-bystander mass. Similarly, adding up (30) and (31) yields

$$\text{(A.2)} \qquad \text{all } hn\text{-buyer stationarity:} \qquad -\lambda_d m_{hn} + \lambda_u m_{ln} - \rho m_{hn} v_{hn} = 0,$$

where $m_{hn} := \sum_k m_{hn}^k$, $v_{hn} := \frac{1}{\rho m_{hn}} \sum_k \rho^k m_{hn}^k v_{hn}^k$, and $m_{ln} := \sum_k m_{ln}^k$. Taking $\{v_{lo}, v_{hn}\}$ as given, the equations (8), (A.1), and (A.2) form a linear system of the four masses $\{m_{ho}, m_{ln}, m_{hn}, m_{lo}\}$, which have the unique solution of

$$\text{(A.3)} \qquad \begin{aligned} m_{ho} &= \eta \frac{\lambda_u v_{hn} + \rho v_{lo} v_{hn}}{\lambda_u v_{hn} + \lambda_d v_{lo} + \rho v_{hn} v_{lo}}; \quad m_{ln} = (1-\eta) \frac{\lambda_d v_{lo} + \rho v_{lo} v_{hn}}{\lambda_u v_{hn} + \lambda_d v_{lo} + \rho v_{hn} v_{lo}}; \\ m_{hn} &= (1-\eta) \frac{\lambda_u v_{lo}}{\lambda_u v_{hn} + \lambda_d v_{lo} + \rho v_{hn} v_{lo}}; \quad m_{lo} = \eta \frac{\lambda_d v_{hn}}{\lambda_u v_{hn} + \lambda_d v_{lo} + \rho v_{hn} v_{lo}}. \end{aligned}$$

Plug in the expressions of m_{ho} and $m_{lo} = \sum_k m_{lo}^k$ into the market clearing condition (1) to get

$$\text{(A.4)} \qquad \eta \frac{(\lambda_u + \lambda_d) v_{hn} + \rho v_{lo} v_{hn}}{\lambda_u v_{hn} + \lambda_d v_{lo} + \rho v_{lo} v_{hn}} + m_{do} - s = 0.$$

This is an equation with unknowns $\{m_{do}, v_{hn}, v_{lo}\}$. It remains to express v_{hn} and v_{lo} as (monotone) functions of m_{do}.

Consider v_{lo} for example. Note that (28) and (29) imply that

$$\text{(A.5)} \qquad m_{lo}^k = \frac{\lambda_d m_{ho} \theta_{lo}^k}{\lambda_u + \rho^k v_{lo}^k}$$

where $\theta_{lo}^{\text{BB}} := 1 - \theta_{lo}$ and $\theta_{lo}^{\text{SMS}} := \theta_{lo}$. Hence, from the earlier definition,

$$\nu_{lo} = \frac{\sum_k \rho^k m_{lo}^k \nu_{lo}^k}{\rho m_{lo}} = \frac{\sum_k \rho^k m_{lo}^k \nu_{lo}^k}{\rho \sum_k m_{lo}^k} = \frac{\sum_k \frac{\rho^k \theta_{lo}^k \nu_{lo}^k}{\lambda_u + \rho^k \nu_{lo}^k}}{\rho \sum_k \frac{\theta_{lo}^k}{\lambda_u + \rho^k \nu_{lo}^k}}, \tag{A.6}$$

which is monotone increasing in both ν_{lo}^k for $k \in \{BB, SMS\}$. Recall from the definition $\nu_{lo}^k := \rho^k \cdot \left(1 - \left(1 - \frac{m_{dn}}{m_d}\right)^{n^k}\right) = \rho^k \cdot \left(1 - \left(\frac{m_{do}}{m_d}\right)^{n^k}\right)$ that both ν_{lo}^k are monotone decreasing in m_{do}. Therefore, so is ν_{lo}. In the same way, both ν_{hn}^k are monotone increasing in m_{do} and so is ν_{hn}.

Now return to Equation (A.4). Since both ν_{lo} and ν_{hn} can be expressed as a unique function in m_{do}, (A.4) is an equation of a single unknown m_{do}. To prove the existence of the solution, consider the limits of the support of $m_{do} \in [0, m_d]$. As $m_{do} \downarrow 0$, both $\nu_{lo}^k \uparrow 1$ while both $\nu_{hn}^k \downarrow 0$, and as a result, $\nu_{lo} \uparrow 1$ and $\nu_{hn} \downarrow 0$. The left-hand side of (A.4), therefore, reaches $-s < 0$. Reversely, as $m_{do} \uparrow m_d$, $\nu_{lo} \downarrow 0$ and $\nu_{hn} \uparrow 1$, the left-hand side of (A.4) reaches $1 + m_d - s > 0$ (as it is assumed that $0 < s < 1 + m_d$). Therefore, by continuity, the solution to m_{do} always exists.

To prove uniqueness, examine the derivative of the left-hand side of (A.4) with respect to m_{do}:

$$-\eta\lambda_d \frac{(\lambda_u + \lambda_d + \rho\nu_{hn})\nu_{hn}}{(\lambda_u \nu_{hn} + \lambda_d \nu_{lo} + \rho \nu_{hn}\nu_{lo})^2} \frac{\partial \nu_{lo}}{\partial m_{do}} + \eta\lambda_d \frac{(\lambda_u + \lambda_d + \rho\nu_{lo})\nu_{lo}}{(\lambda_u \nu_{hn} + \lambda_d \nu_{lo} + \rho \nu_{hn}\nu_{lo})^2} \frac{\partial \nu_{hn}}{\partial m_{do}} + 1 > 0, \tag{A.7}$$

where the inequality holds because ν_{lo} decreases, while ν_{hn} increases, in m_{do}. That is, the left-hand side of (A.4) is strictly monotone increasing in m_{do}. Hence, there exists one and only one m_{do} that solves (A.4). Therefore, the demographics equation system always has a unique solution. □

Lemma 2

Proof. See the proof on p. 21, immediately after the lemma. □

Lemma 3

Proof. Calculate the difference between m_{hn} and m_{lo} using the expressions (10) and (9) to get

$$m_{hn} - m_{lo} = \eta + m_{do} - s = \eta\lambda_d \cdot \frac{\nu_{lo} - \nu_{hn}}{\lambda_u \nu_{hn} + \lambda_d \nu_{lo} + \rho \nu_{hn}\nu_{lo}}, \tag{A.8}$$

where the last equality follows Equation (A.4). Therefore, sign$[m_{hn} - m_{lo}]$ = sign$[\nu_{lo} - \nu_{hn}]$. Recall that $\nu_{lo} = 1 - (1 - m_{dn}/m_d)^n$ and $\nu_{hn} = 1 - (1 - m_{do}/m_d)^n$, from which it follows that $\nu_{lo} > \nu_{hn}$ if and only if $m_{dn} > m_{do}$. Given that $m_{dn} + m_{do} = m_d$, therefore, $m_{hn} > m_{lo}$ if and only if $m_{do} < m_d/2$. Use again $m_{hn} - m_{lo} = \eta + m_{do} - s$, which is negative if and only if $s > \eta + m_{do} > \eta + m_d/2$. □

Lemma 5

Proof. Consider **first** the case of $\rho^{\text{SMS}}q^{\text{SMS}}n^{\text{SMS}} < \rho^{\text{BB}}q^{\text{BB}}n^{\text{BB}}$. The proof first establishes the single-crossing of $z^{\text{SMS}}(\mu)$ and $z^{\text{BB}}(\mu)$ at some $\mu^* \in (0,1)$. The general idea is to characterize the shapes of $z^{\text{BB}}(\mu)$ and $z^{\text{SMS}}(\mu)$. In particular, it will be shown that z^{BB} is linearly increasing in μ, while z^{SMS} is sigmoid-shaped in μ, starting below z^{BB} for sufficiently small μ; and the two satisfy $z^{\text{BB}}(0) = z^{\text{SMS}}(0) = 0$ and $z^{\text{BB}}(1) < z^{\text{SMS}}(1)$. Therefore, there is always one and only one intersection point $\mu^* \in (0,1)$.

Consider z^{BB} first. With $n^{\text{BB}} = 1$, $z^{\text{BB}} = q^{\text{BB}}\rho^{\text{BB}}\mu$, which is linearly increasing from 0 at $\mu = 0$ to $q^{\text{BB}}\rho^{\text{BB}}$ at $\mu = 1$. Next, consider z^{SMS}. For notation simplicity, the superscripts SMS on n, ρ, and q are omitted when there is no confusion. With $n = n^{\text{SMS}} > 1$, $z^{\text{SMS}} = \big(1-(1-\mu)^{n-1}(1-\mu+(1-q)n\mu)\big)\rho$, whose first-order derivative with respect to μ is $\frac{\partial z^{\text{SMS}}}{\partial \mu} = -n\rho(1-\mu)^{n-2}(\mu(1-n)+q(\mu n-1))$, which is positive. To see why, note that the bracketed term, $\mu(1-n)+q(\mu n-1)$ is linear in μ and is negative for both $\mu = 0$ and $\mu = 1$ and so it is negative for all μ. Thus, $z^{\text{SMS}}(\mu)$ is strictly monotone increasing on $\mu \in (0,1)$. Its second-order derivative with respect to μ is $\frac{\partial^2 z^{\text{SMS}}}{\partial \mu^2} = (n-1)n\rho(1-\mu)^{n-3}(\mu+\mu(-n)+q(\mu n-2)+1)$, which is positive if and only if $\mu < \frac{1-2q}{n-1-nq}$. Note that $\frac{1-2q}{n-1-nq} > 0$, because $\rho^{\text{SMS}}q^{\text{SMS}}n^{\text{SMS}} < \rho^{\text{BB}}q^{\text{BB}}n^{\text{BB}}$ implies $q = q^{\text{SMS}} < 1/n^{\text{SMS}} \le 1/2$. Summarizing the above, $z^{\text{SMS}}(\cdot)$ is sigmoid-shaped on $\mu \in (0,1)$: it is monotone increasing, initially convex, but eventually concave.

Now note that in the lower end, $z^{\text{SMS}}|_{\mu\downarrow 0} = z^{\text{BB}}|_{\mu\downarrow 0} = 0$. Further, the slope of $z(\cdot)$ satisfies $\lim_{\mu\downarrow 0}\frac{\mathrm{d}z}{\mathrm{d}\mu} = n\rho q$. Therefore, the assumption $\rho^{\text{SMS}}q^{\text{SMS}}n^{\text{SMS}} < \rho^{\text{BB}}q^{\text{BB}}n^{\text{BB}}$ ensures that for μ sufficiently small, $z^{\text{SMS}} < z^{\text{BB}}$. On the upper end of $\mu \uparrow 1$, $z^{\text{SMS}} \to \rho^{\text{SMS}} \ge \rho^{\text{BB}} \ge q^{\text{BB}}\rho^{\text{BB}}$, where the first inequality follows (26) and the second follows $q^{\text{BB}} \in [0,1]$. That is, the sigmoid-shaped z^{SMS} exceeds z^{BB} eventually. Therefore, there exists a unique $\mu^* \in (0,1)$ at which $z^{\text{SMS}}(\mu^*) = z^{\text{BB}}(\mu^*)$.

To establish that $\mu^* < \frac{1}{2}$, note that $z^{\text{SMS}}(\mu)$ is monotone increasing in n^{SMS} and in q^{SMS}. Therefore, fixing $z^{\text{BB}}(\mu) = q^{\text{BB}}\rho^{B}B\mu$, the intersection μ^* must be higher as n^{SMS} and q^{SMS} reduce. Likewise, fixing $z^{\text{SMS}}(\mu)$, μ^* must be higher when the product of $q^{\text{BB}}\rho^{B}B$ increases. Since $q^{\text{BB}} \in [0,1]$ and $\rho^{\text{BB}} \le \rho^{\text{SMS}}$, the maximum of this product is $q^{\text{BB}}\rho^{\text{BB}} \le \rho^{\text{SMS}}$. Therefore, the maximum μ^* is the solution to $z^{\text{SMS}}(\mu; n^{\text{SMS}} = 3, q^{\text{SMS}} = 0) - \rho^{\text{SMS}}\mu = 0$. Solving this equation gives the unique interior solution of $\mu^* = \frac{1}{2}$.

Next, following the discussion right after the lemma, it is clear that V_σ^k is monotone increasing in ζ_σ^k, where $k \in \{\text{BB}, \text{SMS}\}$ and $\sigma \in \{lo, hn\}$. Hence, comparing the value functions is equivalent to comparing the trading gain intensities $\{\zeta_\sigma^k\}$; i.e., the technology choice (27) is equivalent to (41). With the single-crossing property established above, it then follows that the comparison of the $\{\zeta_\sigma^k\}$

is equivalent to (40).

Finally, consider the case of $\rho^{\text{SMS}} q^{\text{SMS}} n^{\text{SMS}} \geq \rho^{\text{BB}} q^{\text{BB}} n^{\text{BB}}$. The only change is that the slope of $z^k(\mu)$ at the lower end now is higher for SMS than for BB. Thus, the only intersection possible is at $\mu = 0$, i.e., $z^{\text{SMS}} > z^{\text{BB}}$ for all $\mu \in (0, 1)$, i.e., SMS is always preferred and, hence, $\theta_{hn} = \theta_{lo} = 1$. □

Lemma 6

> **Lemma 6.** *Write the left-hand side of Equation* (A.4) *as a function of* $f(\theta_{lo}, \theta_{hn}, m_{do}, s)$. *Then (1)* $\frac{\partial f}{\partial m_{do}} > 0$, *(2)* $\frac{\partial f}{\partial \theta_{lo}} < 0$, *(3)* $\frac{\partial f}{\partial \theta_{hn}} > 0$, *and (4)* $\frac{\partial f}{\partial s} < 0$. *In particular, (5)* $m_{do} \downarrow 0$ *when* $s \downarrow 0$ *and* $m_{do} \uparrow 1 + m_d$ *when* $s \uparrow 1 + m_d$ *regardless of* θ_{lo} *and* θ_{hn}.

Proof. (1) $\frac{\partial f}{\partial m_{do}}$ has been evaluated in (A.7) in the proof of Lemma 4. (2) Note that θ_{lo} affects $f(\cdot)$ only through v_{lo}, which is given by (A.6). Carefully simplifying, it can be found that

$$\frac{\partial v_{lo}}{\partial \theta_{lo}} = \frac{(v_{lo}^{\text{SMS}} - v_{lo}^{\text{BB}})(\lambda_u + v_{lo}^{\text{SMS}})(\lambda_u + v_{lo}^{\text{BB}})}{(\lambda_u + (1 - \theta_{lo}) v_{lo}^{\text{SMS}} + \theta_{lo} v_{lo}^{\text{BB}})^2} > 0$$

where the inequality holds because $v_{lo}^{\text{SMS}} > v_{lo}^{\text{BB}}$ always holds (with $\rho^{\text{SMS}} \geq \rho^{\text{BB}}$ and $n^{\text{SMS}} > n^{\text{BB}} = 1$). The partial derivative of $f(\cdot)$ with respect to v_{lo} is $\frac{\partial f}{\partial v_{lo}} = -\frac{(\lambda_d + \lambda_u + v_{hn})\lambda_d v_{hn}}{(\lambda_u v_{hn} + (\lambda_d + v_{hn}) v_{lo})^2} < 0$. Therefore, by chain rule, $\frac{\partial f}{\partial \theta_{lo}} < 0$. (3) can be proved similarly by showing that $\frac{\partial v_{hn}}{\partial \theta_{hn}} > 0$ and that $\frac{\partial f}{\partial v_{hn}} > 0$. The details are omitted for brevity. (4) is straightforward as $\frac{\partial f}{\partial s} = -1$. (5) By implicit function theorem, $f(\cdot) = 0$ implies that m_{do} strictly increases in s; see (1) and (4) above. The limit values as $s \downarrow 0$ or $s \uparrow 1$ can then be easily verified, regardless of θ_{lo} and θ_{hn}. □

Proposition 1

Proof. The proof only focuses on a contacted *do*-seller's symmetric quoting strategy. The same analysis applies to *dn*-buyers and is omitted. Consider first the trivial case of $n = 1$. A contacted *do*-seller then knows that he is the only one quoting. It is then trivial that with probability $(1 - q)$, he will quote the highest possible ask price, i.e., the *hn*-buyer's reservation value $R_h = R_d + \Delta_{hd}$. This can be viewed as a degenerate mixed strategy with c.d.f. $F(\alpha)$ converging to a unity probability mass at $\alpha = 1$ as stated in the proposition.

Next consider $n \geq 2$. Given the reservation values, it suffices to restrict the ask quote within $[R_d, R_h]$. Without loss of generality, a *do*-seller's strategy can be written as $R_d + \alpha \Delta_{hd}$ by choosing $\alpha \in [0, 1]$. Suppose α has a c.d.f. $F(\alpha)$ with possible realizations $[0, 1]$ (some of which might have zero probability mass). The following four steps pin down the specific form of $F(\cdot)$ so that it sustains a symmetric equilibrium.

Step 1: There are no probability masses in the support of $F(\cdot)$. If at $\alpha^* \in (0, 1]$ there is some non-zero probability mass, any *do*-seller has an incentive to deviate to quoting with the same probability mass but at a markup level infinitesimally smaller than α^*. This way, he converts the strictly positive probability of tying with others at α^* to winning over others. (The undercut costs no expected revenue as it is infinitesimally small.) If at $\alpha^* = 0$ there is non-zero probability mass, again, any *do*-seller will deviate, this time to a markup slightly above zero. This is because allocating probability mass at zero markup brings zero expected profit. Deviating to a slightly positive markup, therefore, brings strictly positive expected profit. Taken together, there cannot be any probability mass in $\alpha \in [0, 1]$. Note that any symmetric-strategy equilibria are ruled out.
Step 2: The support of $F(\cdot)$ is connected. The support is not connected if there is $(\alpha_1, \alpha_2) \subset [0, 1]$ on which there is zero probability assigned and there is probability density on α_1. If this is the case, then any *do*-seller will deviate by moving the probability density on α_1 to any $\alpha \in (\alpha_1, \alpha_2)$. Such a deviation is strictly more profitable because doing so does not affect the probability of winning (if one wins at bidding α_1, he also wins at any $\alpha > \alpha_1$) and because $\alpha > \alpha_1$ is selling at a higher price.
Step 3: The upper bound of the support of $F(\cdot)$ is 1. The logic follows Step 2. Suppose the upper bound is $\alpha^* < 1$. Then, allocating the probability density at α^* to 1 is a profitable deviation: It does not affect the probability of winning and upon winning sells at a higher price.
Step 4: Deriving the c.d.f. $F(\cdot)$. Suppose all other *do*-sellers, when contacted, quote according to some same distribution $F(\cdot)$. Consider a specific seller called i. Quoting $R_d + \alpha\Delta_{hd}$, i gets to trade with the searching buyer if, and only if, such a quote is the best that the buyer receives. The buyer examines all quotes received. For each of the $n - 1$ contacts, with probability $1 - \frac{m_{do}}{m_d}$ the dealer is not a *do*-seller and in this case i's quote beats the no-quote. With probability $\frac{m_{do}}{m_d}$, the contacted investor is indeed another *lo*-seller, who quotes with markup α'. Then, only with probability $\mathbb{P}(\alpha < \alpha') = 1 - F(\alpha)$ will i's quote win. Taken together, for each of the $n - 1$ potential competitor, i wins with probability $\left(1 - \frac{m_{do}}{m_d}\right) + \frac{m_{do}}{m_d}(1 - F(\alpha))$, and he needs to win all these $n - 1$ times to capture the trading gain of $\alpha\Delta_{hd}$. That is,i expects a profit of $\left(1 - \frac{m_{do}}{m_d}F(\alpha)\right)^{n-1}\alpha\Delta_{hd}$. In particular, at the highest possible markup $\alpha = 1$, the above expected profit simplifies to $\left(1 - \frac{m_{do}}{m_d}\right)^{n-1}\Delta_{hd}$, because $F(1) = 1$. In a mixed-strategy equilibrium, i must be indifferent of quoting any markup in the support. Equating the two expressions above and solving for $F(\cdot)$, one obtains the c.d.f. stated in the proposition. It can then be easily solved that the lower bound of the support must be at $\left(1 - \frac{m_{do}}{m_d}\right)^{n-1}$, where $F(\cdot)$ reaches zero. This completes the proof. □

Proposition 2

Proof. Note that the trading gain is $\Delta = R_{hn} - R_{lo} = (V_{ho} - V_{hn}) - (V_{lo} - V_{ln})$, a linear combination of the four unknown value functions. The four equations (18)-(21), therefore, is a linear equation system that uniquely pins down the four unknowns.

It only remains to prove that the trading gains are strictly positive when $\underline{y}_d \leq y_d \leq \overline{y}_d$. Difference Equation (18) and (21) to get $0 = y_h - rR_h - \zeta_{hn}\Delta_{hd} - \lambda_d \cdot (R_h - R_l)$. Similarly, difference Equation (20) and (19) to get $0 = y_l - rR_l + \zeta_{lo}\Delta_{dl} + \lambda_u \cdot (R_h - R_l)$. Finally, difference the two dealers' HJB equations, (22) and (23), to get $y_d - rR_d + \zeta_{do}\Delta_{hd} - \zeta_{dn}\Delta_{dl}$. Note that $\Delta_{hd} = R_h - R_d$ and $\Delta_{dl} = R_d - R_l$. Therefore, taking the $\{\zeta\}$ as given, the above form a 3-equation-3-unknown linear system, from which the reservation values $\{R_h, R_d, R_l\}$ can be uniquely solved. The resulting expressions are complicated and omitted here, but it is straightforward verify that they are all monotone increasing in y_d. (Note that the trading gain intensities $\{\zeta\}$ are independent of y_d.) Therefore, one can find the upper and the lower thresholds by solving $\overline{y}_d^{\prime}$ explicitly from $R_h = R_d$ and $\underline{y}_d^{\prime}$ from $R_d = R_l$:

$$\overline{y}_d^{\prime} := y_l + (y_h - y_l)\frac{\zeta_{dn} + \zeta_{lo} + \lambda_u + r}{\zeta_{lo} + \lambda_d + \lambda_u + r} \text{ and } \underline{y}_d^{\prime} := y_h - (y_h - y_l)\frac{\zeta_{do} + \zeta_{hn} + \lambda_d + r}{\zeta_{hn} + \lambda_d + \lambda_u + r}.$$

The above thresholds are still endogenous of $\{\zeta\}$. To obtain the thresholds composed of exogenous parameters, note that $\overline{y}_d^{\prime}$ is increasing in both ζ_{dn} and ζ_{lo}, that $\underline{y}_d^{\prime}$ is decreasing in both ζ_{do} and ζ_{hn}, and that $\{\zeta\} \geq 0$. Therefore,

$$\begin{aligned}
\overline{y}_d^{\prime} &\geq y_l + (y_h - y_l)\frac{\lambda_u + r}{\lambda_d + \lambda_u + r} = y_h - (y_h - y_l)\frac{\lambda_d}{\lambda_d + \lambda_u + r} =: \overline{y}_d; \\
\underline{y}_d^{\prime} &\leq y_h - (y_h - y_l)\frac{\lambda_d + r}{\lambda_d + \lambda_u + r} = y_l + (y_h - y_l)\frac{\lambda_u}{\lambda_d + \lambda_u + r} =: \underline{y}_d.
\end{aligned}$$

Clearly, $\overline{y}_d > \underline{y}_d$. As such, $\underline{y}_d \leq y_d \leq \overline{y}_d$ is sufficient to ensure $R_l < R_d < R_h$. □

Proposition 3

Proof. The key equation is (A.4) in the proof of Lemma 1. Define the left-hand side as $f(m_{do}, \rho, n)$. Recall that Equation (A.7) has shown that $\frac{\partial f}{\partial m_{do}} > 0$. In addition, simple calculus gives $\text{sign}\left[\frac{\partial f}{\partial \rho}\right] = \text{sign}[v_{lo} - v_{hn}]$. Since excess supply is assumed, i.e., $m_{hn} < m_{lo}$, Equation (A.8) gives $v_{lo} < v_{hn}$. Hence, $\frac{\mathrm{d}m_{do}}{\mathrm{d}\rho} = -\frac{\partial f}{\partial \rho} / \frac{\partial f}{\partial m_{do}} > 0$, i.e., a higher ρ increases m_{do} and, because $m_{dn} = m_d - m_{do}$, decreases m_{dn}. It then also follows that a higher ρ increases $v_{hn} = 1 - (1 - m_{do}/m_d)^n$ but decreases $v_{lo} = 1 - (1 - m_{dn}/m_d)^n$.

Consider the effect of a larger n next. Given the excess supply, $v_{lo} < v_{hn}$ as established above. From the definition of v_{hn} and v_{lo}, therefore, $m_{do} > m_{dn}$ and $\mu := m_{do}/m_d > 1/2$. Taking μ as given, $\frac{\partial v_{hn}}{\partial n} = -\ln(1-\mu)(1-\mu)^n > 0$ and $\frac{\partial v_{lo}}{\partial n} = -\ln\mu \cdot \mu^n > 0$. Further, $\frac{\partial v_{hn}}{\partial n} / \frac{\partial v_{lo}}{\partial n} = \frac{\ln(1-\mu)}{\ln \mu}\left(\frac{1-\mu}{\mu}\right)^n$, which is a function monotone decreasing in μ and equals 1 if and only if $\mu = 1/2$, $\forall n \geq 2$. Therefore, $\frac{\partial v_{lo}}{\partial n} > \frac{\partial v_{hn}}{\partial n}$ (> 0). Simple calculus gives $\frac{\partial f}{\partial v_{hn}} v_{hn} = -\frac{\partial f}{\partial v_{lo}} v_{lo} > 0$. Therefore, by chain rule, $\frac{\partial f}{\partial n} = \frac{\partial f}{\partial v_{hn}}\frac{\partial v_{hn}}{\partial n} + \frac{\partial f}{\partial v_{lo}}\frac{\partial v_{lo}}{\partial n} = \frac{\partial f}{\partial v_{hn}}\left(\frac{\partial v_{hn}}{\partial n} - \frac{v_{hn}}{v_{lo}}\frac{\partial v_{lo}}{\partial n}\right) < \frac{\partial f}{\partial v_{hn}}\frac{\partial v_{hn}}{\partial n}\left(1 - \frac{v_{hn}}{v_{lo}}\right) < 0$. Hence, $\frac{\mathrm{d}m_{do}}{\mathrm{d}n} = -\frac{\partial f}{\partial n} / \frac{\partial f}{\partial m_{do}} > 0$, i.e., a higher n increases m_{do} and, hence, decreases m_{dn}.

Finally, $\frac{\mathrm{d}v_{hn}}{\mathrm{d}n} = \frac{\partial v_{hn}}{\partial n} + \frac{\partial v_{hn}}{\partial \mu}\frac{\partial \mu}{\partial m_{do}}\frac{\partial m_{do}}{\partial n} > 0$ (note that $\frac{\partial v_{hn}}{\partial \mu} > 0$). Take total derivative on Equation (A.4) with respect to v_{lo}, v_{hn}, and m_{do}: $\frac{\partial f}{\partial v_{lo}}\frac{\mathrm{d}v_{lo}}{\mathrm{d}n} + \frac{\partial f}{\partial v_{hn}}\frac{\mathrm{d}v_{hn}}{\mathrm{d}n} + \frac{\partial f}{\partial m_{do}}\frac{\mathrm{d}m_{do}}{\mathrm{d}n} = 0$. Therefore, $\frac{\mathrm{d}v_{lo}}{\mathrm{d}n} = -\left(\frac{\partial f}{\partial v_{hn}}\frac{\mathrm{d}v_{hn}}{\mathrm{d}n} + \frac{\partial f}{\partial m_{do}}\frac{\mathrm{d}m_{do}}{\mathrm{d}n}\right) / \frac{\partial f}{\partial v_{lo}} > 0$, noting that $\frac{\partial f}{\partial v_{lo}} < 0$. □

Proposition 4

Proof. The effects of ρ and n are proved separately below. For concreteness, assume that the asset is in excess supply. (The case of excess demand is symmetric and omitted.)

A higher search intensity ρ: The trading volume can be written as $t = \rho m_{hn} v_{hn}$ (Equation 7). Equation (10) gives another link between t and m_{hn}. Combining the two gives

$$t = \frac{(1 + m_{do} - s)\lambda_u \rho}{(\lambda_d + \lambda_u)v_{hn}^{-1} + \rho},$$

which is increasing in ρ and in m_{do} (note that v_{hn} is also increasing in m_{do}). Proposition 3 has shown that a higher ρ increases m_{do} (given excess supply). Therefore, the volume increases with ρ. It is then also clear from (9) that m_{lo} decreases. Finally, $m_{hn} = \frac{v_{lo}}{v_{hn}} m_{lo}$ by (7). The ratio $\frac{v_{lo}}{v_{hn}} = \frac{1-\mu^n}{1-(1-\mu)^n}$ with $\mu := m_{do}/m_d > 1/2$ given the excess supply. Simply computing the derivative with respect to μ can show that the ratio decreases with μ. That is, a higher ρ, increasing m_{do} and μ, results in a lower m_{hn} as well.

A larger search capacity n: Proposition 3 has shown that a larger n also increases m_{do} (given excess supply). Note that since $v_{hn} = 1 - (1 - m_{do}/m_d)^n$, $\frac{\partial v_{hn}}{\partial m_{do}} > 0$ and $\frac{\partial v_{hn}}{\partial n} > 0$. From the same expression of t above, therefore, n also increases trading volume. Again, from Equation (9), it is clear that m_{lo}, the long-side, then decreases with n.

The effect on $m_{hn} = \frac{v_{lo}}{v_{hn}} m_{lo}$, the short-side, is more complicated, because now n also affects the ratio $\frac{v_{lo}}{v_{hn}}$. To prove the statement, instead, it is easier to turn to the following equivalent expression:

$$\text{(A.9)} \qquad m_{hn}(m_{do}, n) := \frac{t}{\rho v_{hn}} = \frac{(1 + m_{do} - s)\lambda_u}{\lambda_d + \lambda_u + \rho\left(1 - \left(1 - \frac{m_{do}}{m_d}\right)^n\right)},$$

where the second equality follows Equation (10). It is straightforward to find that $\lim_{n\to\infty} \frac{\partial m_{hn}}{\partial n} =$

0; and $\lim_{n\to\infty} \frac{\partial m_{hn}}{\partial m_{do}} = \frac{\lambda_u}{\lambda_d+\lambda_u+\rho} > 0$. Recall from Proposition 3 that $\frac{\mathrm{d}m_{do}}{\mathrm{d}n} > 0$. Therefore, $\lim_{n\to\infty} \frac{\mathrm{d}m_{hn}}{\mathrm{d}n} = \lim_{n\to\infty}\left(\frac{\partial m_{hn}}{\partial n} + \frac{\partial m_{hn}}{\partial m_{do}}\frac{\mathrm{d}m_{do}}{\mathrm{d}n}\right) \geq 0$. □

Proposition 5

Proof. The proof considers the changes in ρ and in n separately. Only the case of excess supply, i.e., $s > \eta + m_d/2$, is analyzed (and the case of excess demand is analogous and is omitted).

When ρ increases: Recall welfare is $w = (y_h m_{ho} + y_d m_{do} + y_l m_{lo})/r$. By market clearing (1), substitute $m_{lo} = s - m_{ho} - m_{do}$ in the above welfare expression to get $w = (y_l s + (y_h - y_l)m_{ho} + (y_d - y_l)m_{do})/r$. By Proposition 3, m_{do} increases with ρ. By Proposition 4, m_{hn} and m_{lo} decrease with ρ. That is, $m_{ho} = \eta - m_{hn}$ increases with ρ. Note that $y_d \in [\underline{y}_d, \overline{y}_d]$ is assumed to ensure positive trading gains (Proposition 2) and that $y_l < \underline{y}_d < \overline{y}_d < y_h$. It then follows that $y_d \in (y_l, y_h)$. Therefore, welfare is increasing with ρ.

When n increases: Welfare can be written as $w = (y_l s + (y_d - y_l)m_{do} + (y_h - y_l)(\eta - m_{hn}))/r$. The effect of n goes through m_{do} and m_{hn}, which are linked through the trading volume definition of $t = \rho m_{hn} v_{hn}$. In the proof of Proposition 4, it has been shown that m_{hn} can be written as a function of m_{do} and n; see Equation (A.9). Applying the chain rule yields

$$\frac{\mathrm{d}m_{hn}}{\mathrm{d}n} = \frac{\partial m_{hn}}{\partial n} + \frac{\partial m_{hn}}{\partial m_{do}}\frac{\mathrm{d}m_{do}}{\mathrm{d}n}. \tag{A.10}$$

Combining the above, one can see that

$$\frac{\mathrm{d}w}{\mathrm{d}n} = \frac{1}{r}\left((y_d - y_l)\frac{\mathrm{d}m_{do}}{\mathrm{d}n} - (y_h - y_l)\frac{\mathrm{d}m_{hn}}{\mathrm{d}n}\right) = \frac{1}{r}\left(\left((y_d - y_l) - (y_h - y_l)\frac{\partial m_{hn}}{\partial m_{do}}\right)\frac{\mathrm{d}m_{do}}{\mathrm{d}n} - (y_h - y_l)\frac{\partial m_{hn}}{\partial n}\right).$$

Therefore, three derivatives of $\frac{\partial m_{hn}}{\partial n}$, $\frac{\partial m_{hn}}{\partial m_{do}}$, and $\frac{\mathrm{d}m_{do}}{\mathrm{d}n}$ need to be evaluated under $\rho \to 0$ and under $\rho \to \infty$.

Consider first the case of $\rho \to 0$. Directly computing the first partial derivative yields

$$\frac{\partial m_{hn}}{\partial n} = \frac{\lambda_u(1 + m_{do} - s)\left(1 - \frac{m_{do}}{m_d}\right)^n \rho \log\left(1 - \frac{m_{do}}{m_d}\right)}{\left(\lambda_d + \lambda_u + \rho\left(1 - \left(1 - \frac{m_{do}}{m_d}\right)^n\right)\right)^2}, \tag{A.11}$$

from which it follows that $\lim_{\rho\to 0} \frac{\partial m_{hn}}{\partial n} = 0$. Also, $\lim_{\rho\to 0} \frac{\partial m_{hn}}{\partial m_{do}} = \frac{\lambda_u}{\lambda_d+\lambda_u} = \eta$. Hence, $\lim_{\rho\to 0} \frac{\mathrm{d}m_{hn}}{\mathrm{d}n} = \eta \lim_{\rho\to 0} \frac{\mathrm{d}m_{do}}{\mathrm{d}n}$. Therefore, $\lim_{\rho\to 0} \frac{\mathrm{d}w}{\mathrm{d}n} = \frac{1}{r}\left((y_d - y_l)\lim_{\rho\to 0} \frac{\mathrm{d}m_{do}}{\mathrm{d}n} - (y_h - y_l)\eta \lim_{\rho\to 0} \frac{\mathrm{d}m_{do}}{\mathrm{d}n}\right) = \frac{1}{r}(y_d - \hat{y})\lim_{\rho\to 0} \frac{\mathrm{d}m_{do}}{\mathrm{d}n}/r$, where $\hat{y} := \eta y_h + (1-\eta)y_l$. Note that $\lim_{\rho\to 0} \frac{\mathrm{d}m_{do}}{\mathrm{d}n} > 0$ because (i) from (A.4), $\lim_{\rho\to 0} m_{do} \in (0, m_d)$; and (ii) given the excess supply, m_{do} increases in n (Proposition 3). Therefore,

$\text{sign}\left[\lim_{\rho\to 0}\frac{\mathrm{d}w}{\mathrm{d}n}\right] = \text{sign}[y_d - \hat{y}]$, proving the statement.

It remains to prove that $\hat{y}_d \in (\underline{y}_d, \overline{y}_d)$ for this low ρ case. As $\rho \to 0$, the upper and the lower bounds for $y_d, \overline{y}_d$ and $\underline{y}_d$, as given in Proposition 2, converge to, respectively, $\overline{y}_d' = y_h - (y_h - y_l)\frac{\lambda_d}{\lambda_d+\lambda_u+r}$ and $\underline{y}_d' = y_l + (y_h - y_l)\frac{\lambda_u}{\lambda_d+\lambda_u+r}$, because both ζ_{hn} and ζ_{lo} are proportional to ρ and converge to zero. It is then easy to verify that $\overline{y}_d' - \hat{y} > 0 > \underline{y}_d' - \hat{y}$.

Next, consider the case of $\rho \to \infty$. Note that $\frac{\mathrm{d}m_{do}}{\mathrm{d}n} > 0$ (Proposition 3). Then signing $\frac{\mathrm{d}w}{\mathrm{d}n}$ in this case is equivalent to

$$\text{sign}\left[\lim_{\rho\to\infty}\frac{\mathrm{d}w}{\mathrm{d}n}\right] = \text{sign}\left[\frac{y_d - y_l}{y_h - y_l} - \lim_{\rho\to\infty}\frac{\partial m_{hn}}{\partial m_{do}} - \lim_{\rho\to\infty}\left(\frac{\partial m_{hn}}{\partial n}/\frac{\mathrm{d}m_{do}}{\mathrm{d}n}\right)\right].$$

From (A.9) it can be seen that $\lim_{\rho\to\infty} m_{hn} = 0$, a constant irrespective of n. Therefore, $\lim_{\rho\to\infty}\frac{\partial m_{hn}}{\partial m_{do}} = 0$. Further, (A.11) has shown that $\frac{\partial m_{hn}}{\partial n} \le 0$ (because of the $\log(\cdot)$ term). Thus, $\lim_{\rho\to\infty}\left(\frac{\partial m_{hn}}{\partial n}/\frac{\mathrm{d}m_{do}}{\mathrm{d}n}\right) < 0$. This proves $\text{sign}\left[\lim_{\rho\to\infty}\frac{\mathrm{d}w}{\mathrm{d}n}\right] > 0$.

□

Proposition 6

Proof. To begin with, note that both v_{lo} and v_{hn} are only functions of θ_{lo}^k and θ_{hn}^k, respectively; see, e.g., Equation (A.6). Equation (A.4) can then be written as $g(\theta_{lo}, \theta_{hn}, m_{do}) = s$. Define the four thresholds $\{s_{hn,0}, s_{hn,1}, s_{lo,1}, s_{lo,0}\}$ to be the respective unique solution to $g(\cdot) = s$ for $\{\theta_{lo}, \theta_{hn}, m_{do}\} \in \{\{1, 0, \mu^* m_d\}, \{1, 1, \mu^* m_d\}, \{1, 1, (1-\mu^*)m_d\}, \{0, 1, (1-\mu^*)m_d\}\}$. It is easy to see the four thresholds indeed exist according to this definition. In particular, the monotonicity shown in Lemma 6 guarantees the sorting of these thresholds. To complete the proof, for each region of s, the stated values of $\{\theta_{lo}, \theta_{hn}, m_{do}\}$ are first verified to indeed sustain an equilibrium and then shown to be unique in that region.

Region 1: $0 < s < s_{hn,0}$**.** With $\{\theta_{lo}, \theta_{hn}\} = \{1, 0\}$, m_{do} is uniquely pinned down by Equation (A.4). Since $s < s_{hn,0}$, Lemma 6 implies that $m_{do} < \mu^* m_d$. Hence, by Lemma 5, $\zeta_{hn}^{\text{SMS}} < \zeta_{hn}^{\text{BB}}$ but $\zeta_{lo}^{\text{SMS}} > \zeta_{lo}^{\text{BB}}$ and, indeed, $\{\theta_{lo}, \theta_{hn}\} = \{1, 0\}$ sustains an equilibrium.

There are three possible deviations. First, suppose instead $\{\theta_{lo}, \theta_{hn}\} \in (0, 1) \times (0, 1)$. This would require both *hn*-buyers and *lo*-sellers be indifferent between the two technologies. That is, $m_{do}/m_d = m_{dn}/m_d = \mu^*$ must hold, implying $\mu^* = 1/2$ (because $m_{do} + m_{dn} = m_d$), which is ruled out because Lemma 5 has shown that $0 < \mu^* < 1/2$. Second, suppose $\theta_{lo} = \theta_{hn} = 0$. But by Lemma 6, this reduction in θ_{lo} would only reduce m_{do} (for a fixed s) and increase m_{dn}, making *lo*-sellers prefer SMS more, hence inconsistent with the required $\zeta_{lo}^{\text{SMS}} < \zeta_{lo}^{\text{BB}}$. Third, suppose $\theta_{lo} = \theta_{hn} = 1$. Likewise, this increase in θ_{hn} would decrease m_{do}, inconsistent with *hn*-buyers' switch from BB to SMS as a lower m_{do} would only strengthen $\zeta_{hn}^{\text{SMS}} < \zeta_{hn}^{\text{BB}}$. Since none of these alternative values of θ_{lo} and θ_{hn} can sustain the equilibrium, in this range of s, the only possible equilibrium is $\{\theta_{lo}, \theta_{hn}\} = \{1, 0\}$.

Region 2: $s_{hn,0} \leq s \leq s_{hn,1}$**.** With $\{\theta_{lo}, m_{do}\} = \{1, \mu^* m_d\}$ in this region, $g(\cdot) = s$ uniquely solves $\theta_{hn} \in [0, 1]$. This is indeed an equilibrium because at $m_{do} = \mu^* m_d$, hn-buyers are indifferent between SMS and BB and, hence, any $\theta_{hn} \in [0, 1]$ is admissible. On the other hand, $m_{dn} = m_d - m_{do} > \mu^* m_d$ because $\mu^* < 1/2$. Therefore, $\zeta_{lo}^{\text{SMS}} > \zeta_{lo}^{\text{BB}}$ by Lemma 5 and $\theta_{lo} = 1$ is sustained.

To rule out other equilibria, consider alternative values. Suppose $m_{do} > \mu^* m_d$, implying $\theta_{hn} = 1$. Recall that $s = s_{hn,1}$ is the unique solution to $g(\cdot) = s$ when $\theta_{lo} = \theta_{hn} = 1$ and $m_{do} = \mu^* m_d$. The monotonicity in Lemma 6 would then require $s > s_{hn,1}$, out of this region. Suppose instead $m_{do} < \mu^* m_d$, implying $\theta_{hn} = 0$. Then similarly, the monotonicity in Lemma 6 would require $s < s_{hn,0}$, again out of this region. Finally, suppose $m_{do} = \mu^* m_d$ but $\theta_{lo} < 1$. This immediately contradicts with $\zeta_{lo}^{\text{SMS}} > \zeta_{lo}^{\text{BB}}$ as implied by $m_{dn} = m_d - m_{do} > \mu^* m_d$.

Region 3: $s_{hn,1} < s < s_{lo,1}$**.** When $\theta_{lo} = \theta_{hn} = 1$, $s_{hn,1} < s < s_{lo,1}$ ensures that m_{do} as solved from $g(\cdot) = s$ satisfies $\mu^* m_d < m_{do} < (1 - \mu^*) m_d$; and, hence, $m_{dn} = m_d - m_{do} > \mu^* m_d$. That is, $\zeta^{\text{SMS}} > \zeta^{\text{BB}}$ for both hn and lo, which indeed guarantee that $\theta_{lo} = \theta_{hn} = 1$ as an equilibrium.

Again, consider other values for $\{\theta_{lo}, \theta_{hn}\}$. First, $\{\theta_{lo}, \theta_{hn}\} \in (0, 1)^2$ cannot be an equilibrium for the same reason as explained in Region 1. Second, suppose $\{\theta_{lo}, \theta_{hn}\} = \{1, 0\}$. By Lemma 6, this reduction in θ_{hn} would result in an increase in m_{do}, but such an increase would only make SMS more attractive for hn-buyers, contradicting the reduction of θ_{hn}. Third, suppose $\{\theta_{lo}, \theta_{hn}\} = \{0, 1\}$. Then similarly by Lemma 6, this reduction in θ_{lo} would result in a decrease in m_{do} or an increase in m_{dn}, but such an increase would only make SMS more attractive for lo-sellers, contradicting the reduction of θ_{lo}.

Region 4: $s_{lo,1} \leq s \leq s_{lo,0}$**.** This region mirrors Region 2 and the proof is omitted for brevity.

Region 5: $s_{lo,0} < s < 1 + m_d$**.** This region mirrors Region 1 and the proof is omitted for brevity. □

Proposition 7

Proof. We consider the case $s > s_{hn,1}$ and prove that the ratio defined in (43) weakly decreases in s. The volume ratio in this region can be written as

$$VS = \frac{\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}} + \rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}}{\left(\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}} + \rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}\right) + \left(\rho^{\text{BB}} m_{lo}^{\text{BB}} v_{lo}^{\text{BB}} + \rho^{\text{BB}} m_{hn}^{\text{BB}} v_{hn}^{\text{BB}}\right)} = \frac{1}{2} + \frac{1}{2} \frac{m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}}}{m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}}.$$

This is because in the considered region, $\theta_{hn} = 1$. Then the dealer stationarity (32) reduces to

$$\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}} + \rho^{\text{BB}} m_{lo}^{\text{BB}} v_{lo}^{\text{BB}} = \rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}. \tag{A.12}$$

We consider three cases next:

- $s < s_{lo,1}$. In this case, $\theta_{lo} = 1$, which means that the dealer stationarity condition (32) writes as $\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}} = \rho^{\text{SMS}} m_{hn}^{\text{SMS}} v_{hn}^{\text{SMS}}$ implying $VS = 1$.
- $s > s_{lo,0}$. In this case, $\theta_{lo} = 0$, implying $VS = 1/2$.
- $s_{lo,1} \le s \le s_{lo,0}$. In this case, m_{do} is a constant, invariant of s, and so both v_{lo}^{BB} and v_{lo}^{SMS} are constants as well. Then $\text{sign}\frac{\text{d}VS}{\text{d}s} = \text{sign}\frac{\text{d}}{\text{d}s}\left(m_{lo}^{\text{SMS}}/m_{hn}^{\text{SMS}}\right)$. Using again (A.12),

$$\frac{m_{lo}^{\text{SMS}}}{m_{hn}^{\text{SMS}}} = \frac{\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{hn}^{\text{SMS}}}{\rho^{\text{SMS}} m_{lo}^{\text{SMS}} v_{lo}^{\text{SMS}} + \rho^{\text{BB}} m_{lo}^{\text{BB}} v_{lo}^{\text{BB}}} = \frac{\rho^{\text{SMS}} v_{hn}^{\text{SMS}}}{\rho^{\text{SMS}} v_{lo}^{\text{SMS}} + \rho^{\text{BB}} v_{lo}^{\text{BB}} \left(\frac{m_{lo}^{\text{BB}}}{m_{lo}^{\text{SMS}}}\right)}$$

  Hence, $\text{sign}\frac{\text{d}VS}{\text{d}s} = \text{sign}\frac{\text{d}}{\text{d}s}\left(m_{lo}^{\text{SMS}}/m_{hn}^{\text{SMS}}\right) = \text{sign}\frac{\text{d}}{\text{d}s}\left(m_{lo}^{\text{SMS}}/m_{lo}^{\text{BB}}\right)$. Using the stationarity conditions (28) and (29),

$$\frac{m_{lo}^{\text{SMS}}}{m_{lo}^{\text{BB}}} = \frac{\lambda_u + \rho^{\text{BB}} v_{lo}^{\text{BB}}}{\lambda_u + \rho^{\text{SMS}} v_{lo}^{\text{SMS}}} \frac{\theta_{lo}}{1 - \theta_{lo}},$$

  increasing in θ_{lo}, which is the only variable endogenous of s. Proposition 6 has shown that in this range, θ_{lo} decreases with s. Therefore, by chain rule, $\text{sign}\frac{\text{d}V}{\text{d}s} < 0$.

Combining the three cases completes the proof for the claims regarding s. To prove the claims regarding λ_d, note that from Equation (A.4), cateris paribus, the left-hand side is monotone increasing in λ_d (the excess supply implies $v_{hn} > v_{lo}$; see Equation (A.8)) but decreasing in s. Therefore, increases in s are equivalent to those in λ_d. Hence, all results about s above also hold for λ_d. □

Proposition 8 and 9

Proof. Welfare can be written as $w = \frac{1}{r}(y_l s + (y_d - y_l)m_{do} + (y_h - y_l)(\eta - m_{hn}))$. Consider a small change in either $\theta \in \{\theta_{hn}, \theta_{lo}\}$. We then have

$$\text{sign}\left[\frac{\text{d}w}{\text{d}\theta}\right] = \text{sign}\left[(y_d - y_l)\frac{\text{d}m_{do}}{\text{d}\theta} - (y_h - y_l)\frac{\text{d}m_{hn}}{\text{d}\theta}\right].$$

Moreover, following $m_{ho} + m_{hn} = \eta$ and using the expressions (9) and (10), we have

$$\frac{\text{d}m_{hn}}{\text{d}\theta} = -\frac{\text{d}m_{hn}}{\text{d}\theta} = \eta\frac{\text{d}m_{do}}{\text{d}\theta} - \frac{1}{\lambda_u + \lambda_d}\frac{\text{d}t}{\text{d}\theta}. \tag{A.13}$$

Combining the above two, we get

$$\text{(A.14)} \qquad \operatorname{sign}\left[\frac{\mathrm{d}w}{\mathrm{d}\theta}\right] = \operatorname{sign}\left[(y_d - \hat{y})\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta} + \frac{y_h - y_l}{\lambda_u + \lambda_d}\frac{\mathrm{d}t}{\mathrm{d}\theta}\right],$$

where $\hat{y} := \eta y_h + (1-\eta)y_l$. The derivative of $\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}$ can be signed by the implicit function theorem using the results from Lemma 6: $\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta_{lo}} > 0$ and $\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta_{hn}} < 0$. To see how volume t changes with respect to θ, recall from Equations (9) and (10) and use $t = \rho m_{lo} v_{lo} = \rho m_{hn} v_{hn}$ to get

$$\text{(A.15)} \qquad t = \frac{\lambda_d \rho (s - m_{do})}{\frac{\lambda_d + \lambda_u}{v_{lo}} + \rho} \text{ and } t = \frac{\rho \lambda_u (1 + m_{do} - s)}{\frac{\lambda_d + \lambda_u}{v_{hn}} + \rho}.$$

Note that θ_{hn} in the first expression only affects t through m_{do}. Therefore, t is increasing in θ_{lo}. Likewise, θ_{lo} affects t in the second expression only through m_{do}. Hence, t is also increasing in θ_{hn}. That is, $\frac{\mathrm{d}t}{\mathrm{d}\theta} > 0$ for either $\theta \in \{\theta_{lo}, \theta_{hn}\}$.

The case of sufficiently high ρ, i.e., $\rho := \min[\rho^{\mathbf{BB}}, \rho^{\mathbf{SMS}}] \to \infty$**:** Since $\frac{\mathrm{d}t}{\mathrm{d}\theta} > 0$,

$$\operatorname{sign}\left[\lim_{\rho\to\infty}\frac{\mathrm{d}w}{\mathrm{d}\theta}\right] = \operatorname{sign}\left[(y_d - \hat{y})\lim_{\rho\to\infty}\left(\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}/\frac{\mathrm{d}t}{\mathrm{d}\theta}\right) + \frac{y_h - y_l}{\lambda_u + \lambda_d}\right].$$

Hence, one needs to find $\lim_{\rho\to\infty}\left(\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}/\frac{\mathrm{d}t}{\mathrm{d}\theta}\right)$.

Consider first $\theta = \theta_{lo}$. Then differentiate the second expression of t in (A.15) with respect to $\theta = \theta_{lo}$, noting that v_{hn} is not affected by θ_{lo}, to get $\left(\frac{\lambda_d+\lambda_u}{v_{hn}} + \rho\right)\frac{\mathrm{d}t}{\mathrm{d}\theta} = \rho\lambda_u\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}$. Hence, $\lim_{\rho\to\infty}\left(\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}/\frac{\mathrm{d}t}{\mathrm{d}\theta}\right) = \frac{1}{\lambda_u}$. (Note that $v_{hn}^k = 1 - (1 - m_{do}/m_d)^{n^k}$ is always nonzero, because $m_{do} > m_d/2$ in the case of excess supply.) Then $\operatorname{sign}\left[\lim_{\rho\to\infty}\frac{\mathrm{d}w}{\mathrm{d}\theta}\right] = \operatorname{sign}\left[\frac{y_d-\hat{y}}{\lambda_u} + \frac{y_h-y_l}{\lambda_u+\lambda_d}\right] = \operatorname{sign}[y_d - \hat{y} - (y_h - y_l)\eta] = \operatorname{sign}[y_d - y_l] > 0$. (Recall that $y_d \in (\overline{y}_d', \overline{y}_d') \subset (y_l, y_h)$ by Corollary 2).

Consider $\theta = \theta_{hn}$. Then differentiate the first expression of t in (A.15) with respect to $\theta = \theta_{hn}$, noting that v_{lo} is not affected by θ_{hn}, to get $\left(\frac{\lambda_d+\lambda_u}{v_{lo}} + \rho\right)\frac{\mathrm{d}t}{\mathrm{d}\theta} = -\rho\lambda_d\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}$. Note that $v_{lo}^k = 1 - (m_{do}/m_d)^{n^k}$. As $\rho \to \infty$, the limit of m_{do} may be binding at m_d, resulting in $v_{lo} \to 0$. If it is not binding, i.e., if $\lim_{\rho\to\infty} m_{do} < m_d$, then $v_{lo}^k > 0$ and $\lim_{\rho\to\infty}\left(\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}/\frac{\mathrm{d}t}{\mathrm{d}\theta}\right) = -\frac{1}{\lambda_d}$. If it is binding, i.e., $m_{do} \to m_d$ and $v_{lo} \to 0$, then $\lim_{\rho\to\infty}\left(\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}/\frac{\mathrm{d}t}{\mathrm{d}\theta}\right) = -\frac{1}{\lambda_d}\left(1 + \lim_{\rho\to\infty}\frac{\lambda_d+\lambda_u}{\rho v_{lo}}\right)$. Note from (A.1) that $\rho v_{lo} = \lambda_d\frac{m_{ho}}{m_{lo}} - \lambda_u$. In this case, since $m_{do} \to m_d$, no dealers can intermediate *lo*-sellers. The stationarity of *lo*-seller population size then requires $m_{ho}\lambda_d = m_{lo}\lambda_u$ in this limit. Then, $\lim_{\rho\to\infty}(\rho v_{lo}) = 0$ and again the same result of $\lim_{\rho\to\infty}\left(\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}/f d_t\theta\right) = -\frac{1}{\lambda}$ holds. Therefore, $\operatorname{sign}\left[\lim_{\rho\to\infty}\frac{\mathrm{d}w}{\mathrm{d}\theta}\right] = \operatorname{sign}\left[-\frac{y_d-\hat{y}}{\lambda_d} + \frac{y_h-y_l}{\lambda_u+\lambda_d}\right] = \operatorname{sign}[-y_d + \hat{y} - (y_h - y_l)(1-\eta)] = \operatorname{sign}[y_h - y_d] > 0$.

The case of sufficiently low ρ, i.e., $\rho := \max\{\rho^{\mathbf{BB}}, \rho^{\mathbf{SMS}}\} \to 0$**:** For either $\theta \in \{\theta_{lo}, \theta_{hn}\}$, directly calculating $\frac{\mathrm{d}t}{\mathrm{d}\theta}$ from (A.15) and taking the limit yield $\lim_{\rho\to 0}\frac{\mathrm{d}t}{\mathrm{d}\theta} = 0$. Yet, $\lim_{\rho\to 0}\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta} \neq 0$, which follows by taking the limit in the calculations of Lemma 6. Hence, $\lim_{\rho\to 0}\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta_{lo}} > 0$ and $\lim_{\rho\to 0}\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta_{hn}} < 0$ remain. Therefore, $\lim_{\rho\to 0}\operatorname{sign}\left[\frac{\mathrm{d}w}{\mathrm{d}\theta}\right] = \operatorname{sign}\left[(y_d - \hat{y})\lim_{\rho\to 0}\frac{\mathrm{d}m_{do}}{\mathrm{d}\theta}\right]$, proving the statement made in the proposition. □

Corollary 1

Proof. Consider a searching *hn*-buyer, for example. He contacts n investors but knows that the number of counterparties he will actually find, N, is a random variable that follows a binomial distribution with n draws and success rate μ_{lo}. Each of these N counterparties then quotes a random price according to $F(\alpha;\mu_{lo},n)$, stated in Proposition 1. The searching buyer chooses the lowest ask (the lowest markup) across the N available quotes. The c.d.f. of this minimum markup is $1-(1-F(\alpha;\cdot))^{N-1}$ for $N \geq 1$. Since the probability of $N \geq 1$ is $(1-(1-\mu_{lo})^n)$, one obtains the the conditional c.d.f., as stated in the corollary. The same applies to a searching *lo*-seller. □

Corollary 2

Proof. In equilibrium, the trading customers either have a strict preference for one of the technology or are indifferent. Consider *lo*-sellers, for example. If the preference is strict, then only one of the two HJBs in (35) is relevant; and if indifference, then the two HJBs reduce to the same one. The same holds for *hn*-buyers in their two HJBs (36). Likewise, the max[·] operator in Equations (35) and (36) can be dropped in equilibrium. Hence, defining $V_{lo} = \max_k[\{V_{lo}^k\}]$ and $V_{hn} = \max_k[\{V_{hn}^k\}]$, the HJB equations (33)-(38) can be reduced to the exactly the same set of (18)-(23) as if there is only one technology. Therefore, solving the same equation system, Proposition 2 holds. □

Corollary 3

Proof. Since $n^{\mathrm{BB}} = 1 < n^{\mathrm{SMS}}$, below the notation n, without the superscript, indicates n^{SMS}. Corollary 1 gives $\bar{B}^k$. In particular, for BB, $\bar{B}^{\mathrm{BB}} = 1$, and for SMS, $\bar{B}^{\mathrm{SMS}} = \frac{n\cdot(1-\mu)\mu^{n-1}}{1-\mu^k}$, where $\mu := m_{do}/m_d$. Then $\bar{B}^{\mathrm{SMS}}/\bar{B}^{\mathrm{BB}} = \bar{B}^{\mathrm{SMS}}$. By Lemma 6, m_{do} is weakly increasing with s and hence so does μ, thus proving the claim. To prove the claims regarding λ_d, note that from Equation (A.4), cateris paribus, the left-hand side is monotone increasing in λ_d (the excess supply implies $\nu_{hn} > \nu_{lo}$; see Equation (A.8)) but decreasing in s. Hence, all results about s hold for λ_d. □

Results in Section ??

Proof. 1. Consider the c.d.f. $G(x)$ given in Corollary 1. Fixing any $x \in [0,1]$, it is easy to verify that $\partial G/\partial\mu \geq 0$; that is, the cumulative density at any x is increasing with μ. Therefore, $G(\cdot;\mu_i)$ first-order stochastically dominates $G(\cdot;\mu_j)$ when $\mu_i < \mu_j$. Likewise, one can treat n as if it has a continuous support $n \in [1\infty)$ and easily verify $\partial G/\partial n \geq 0$. Therefore, $G(\cdot;n_i)$ first-order stochastically dominates $G(\cdot;n_j)$ when $n_i > n_j$.

2. This result immediately follows the first-order stochastic dominance.

3. This result is self-evident.

4. The trading price dispersion (in fractions of total trading gain Δ) can be evaluated as $\sqrt{\text{var}[X]}$, where X follows the c.d.f. $G(\cdot)$ in Corollary 1. Evaluating the variance yields

$$\text{var}[X] = \frac{(1-\mu)^{n-2}\big((1-(1-\mu)^n)^2 + (-2+\mu+(1-\mu)^n(2-(n-1)^2\mu))\mu\big)}{(1-(1-\mu)^n)^2}\frac{n}{n-2}.$$

It is easy to see that $\text{var}[X] = 0$ for $\mu \in \{0,1\}$. It can be further verified that $\partial\text{var}[X]/\partial\mu = 0$ has a unique solution in terms of $\mu \in (0,1)$. Since $\text{var}[X] \geq 0$, therefore, the price dispersion must be quasi-concave in μ on the support of $[0,1]$.

5. Consider the nonparametric skewness, i.e., $(\mathbb{E}[X] - \text{median}[X])/\sqrt{\text{var}[X]})$, where X follows the c.d.f. $G(x)$ given in Corollary 1. The median can be calculated as the solution of $G(x) = 0.5$. In particular, $\text{median}[X] = \left(\frac{1}{2} + \frac{1}{2(1-\mu)^n}\right)^{-\frac{n-1}{n}} < \mathbb{E}[X] = \frac{n\mu\cdot(1-\mu)^{n-1}}{1-(1-\mu)^n}$; that is, the skewness is positive. Furthermore, the price for a searching *hn*-buyer is $R_{lo} + A\Delta$ but that for a searching *lo*-seller is $R_{hn} - B\Delta$, where A and B are positively skewed. Therefore, the *hn*-buyer's trading prices (with markups) are positively skewed but *lo*-sellers' trading prices (with markdowns) are negatively skewed. □

Following 200716b: proofs

Summary of the environment

We first re-state the endogenous variables and the conditions pinning them down. In 200716b, the endogenous variables are: the probabilities of choosing SMS technology for new buyers and sellers,

$$\{\theta_{lo}, \theta_{hn}\}$$

8 demographic variables

$$\{m_{ho}, m_{ln}, m_{hn}^{\text{BB}}, m_{hn}^{\text{SMS}}, m_{lo}^{\text{BB}}, m_{lo}^{\text{SMS}}, m_{do}, m_{dn}\};$$

and 8 value functions

$$\{V_{ho}, V_{ln}, V_{hn}^{\text{BB}}, V_{hn}^{\text{SMS}}, V_{lo}^{\text{BB}}, V_{lo}^{\text{SMS}}, V_{do}, V_{dn}\}.$$

Conditions for the demographics. Fixing the tech-choices $\{\theta_t\}$, the demographics are determined by:

(A.16) SMS lo-seller stationarity: $-\lambda_u m_{lo}^{\text{SMS}} + \lambda_d \theta_{lo} m_{ho} - \nu_{lo}^{\text{SMS}} m_{lo}^{\text{SMS}} = 0$

(A.17) BB lo-seller stationarity: $-\lambda_u m_{lo}^{\text{BB}} + \lambda_d (1-\theta_{lo}) m_{ho} - \nu_{lo}^{\text{BB}} m_{lo}^{\text{BB}} = 0$

(A.18) SMS hn-buyer stationarity: $-\lambda_d m_{hn}^{\text{SMS}} + \lambda_u \theta_{hn} m_{ln} - \nu_{hn}^{\text{SMS}} m_{hn}^{\text{SMS}} = 0$

(A.19) BB hn-buyer stationarity: $-\lambda_d m_{hn}^{\text{BB}} + \lambda_u (1-\theta_{hn}) m_{ln} - \nu_{hn}^{\text{BB}} m_{hn}^{\text{BB}} = 0$

(A.20) ln-bystander stationarity: $-\lambda_u m_{ln} + \lambda_d (m_{hn}^{\text{SMS}} + m_{hn}^{\text{BB}}) + \nu_{lo}^{\text{SMS}} m_{lo}^{\text{SMS}} + \nu_{lo}^{\text{BB}} m_{lo}^{\text{BB}} = 0$

The above 6 equations are about flows. The 3 conditions below are about stocks.

(A.21) market clearing: $m_{ho} + \sum_k \left(m_{lo}^k\right) + m_{do} = s$

(A.22) total customer mass: $m_{ho} + m_{ln} + \sum_k \left(m_{hn}^k + m_{lo}^k\right) = 1$

(A.23) total dealer mass: $m_{do} + m_{dn} = m_d$

The above conditions also ensure all other necessary stationarity of the system. For example, (A.18) + (A.19) + (A.20) implies

(A.24) dealer stationarity: $\sum_k \left(\nu_{lo}^k m_{lo}^k - \nu_{hn}^k m_{hn}^k\right) = 0.$

Also, (A.16) +(A.17)+(A.24) gives

(A.25) ho-bystander stationarity: $-\lambda_d m_{ho} + \lambda_u (m_{lo}^{\text{SMS}} + m_{lo}^{\text{BB}}) + \nu_{hn}^{\text{SMS}} m_{hn}^{\text{SMS}} + \nu_{hn}^{\text{BB}} m_{hn}^{\text{BB}} = 0.$

Then (A.18) + (A.19) + (A.25) + (A.24) gives $-\lambda_d \sum_k \left(m_{ho}^k + m_{hn}^k\right) + \lambda_u \sum_k \left(m_{ln}^k + m_{lo}^k\right) = 0,$

which, together with (A.22), implies both the high-type and the low-type stationarity:

(A.26) high type stationarity: $$m_{ho} + \sum_k \left(m^k_{hn}\right) = \frac{\lambda_u}{\lambda_u + \lambda_d} = \eta,$$

(A.27) low type stationarity: $$m_{ln} + \sum_k \left(m^k_{lo}\right) = \frac{\lambda_d}{\lambda_u + \lambda_d} = 1 - \eta.$$

Condition for the value functions. The HJB equations are:

$$\begin{aligned}
ho: &\quad 1 + \lambda_d \cdot \left(\theta_{lo} V^{\text{SMS}}_{lo} + (1-\theta_{lo}) V^{\text{BB}}_{lo} - V_{ho}\right) - rV_{ho} = 0 \\
ln: &\quad \lambda_u \cdot \left(\theta^{\text{SMS}}_{hn} V^{\text{SMS}}_{hn} + (1-\theta_{hn}) V^{\text{BB}}_{hn} - V_{ln}\right) - rV_{ln} = 0 \\
hn \text{ using technology } k: &\quad \lambda_d \cdot (V^k_{ln} - V^k_{hn}) - rV^k_{hn} + \zeta^k_{hn}\Delta^k_{hn} = 0 \\
lo \text{ using technology } k: &\quad (1-\delta) + \lambda_u \cdot (V^k_{ho} - V^k_{lo}) - rV^k_{lo} + \zeta^k_{lo}\Delta^k_{lo} = 0 \\
do: &\quad -rV_{do} + \sum_k \zeta^k_{do}\Delta^k_{hn} = 0 \\
dn: &\quad -rV_{dn} + \sum_k \zeta^k_{dn}\Delta^k_{lo} = 0
\end{aligned}$$

The equilibrium requires that the value functions and the technology choices of the two trading types, hn and lo, solve the following mixed-complementarity problem:

(A.28) $$\begin{cases} 0 < \theta^k_t < 1, & \text{if } V^k_t = V^{-k}_t \\ \theta^k_t = 0, & \text{if } V^k_t < V^{-k}_t \\ \theta^k_t = 1, & \text{if } V^k_t > V^{-k}_t \end{cases}$$

subject to $\sum_k \theta^k_t = 1$, for $t \in \{hn, lo\}$ and $k \in \{\text{BB}, \text{SMS}\}$.

Define

(A.29) $$\zeta(\mu; \rho, q, n) := \left(1 - (1-\mu)^{n-1}(1 - \mu + (1-q)n\mu)\right)\rho, \quad \text{with } \zeta(1; \rho, q, n = 1) := q\rho.$$

Then $\zeta^k_{hn} = \zeta\left(\frac{m_{do}}{m_d}; \rho^k, q^k, n^k\right)$ and $\zeta^k_{lo} = \zeta\left(\frac{m_{dn}}{m_d}; \rho^k, q^k, n^k\right)$.

Lemma 7. *The following is true:* $\zeta(\mu; \rho, q, n) = \mu\rho q n + o(\mu)$.

Proof. Follows by applying Taylor's theorem to $\zeta(\mu; \cdot)$ at $\mu = 0$. □

Discussion: when choosing techs short side cares about the product of tech parameters. Why SMS might be preferable? Because q is small.

Symmetric technology choice equilibrium. The analysis below considers on an equilibrium where all trading customers of the same type use the same technology choice strategy, hence the name "symmetric technology choice equilibrium."

Such an equilibrium is in general characterized by the four technology choices, the ten masses, and the ten value functions. For simplicity, instead, an equilibrium will be referred to as a tuple of $\{m_{do}, \theta_{lo}, \theta_{hn}\} \in [0, m_d] \times [0,1]^2$, pinned down by the three equations of (A.4), (A.28) for $t \in \{lo, hn\}$ and $k = \text{SMS}$. Once these three variables are fixed, the rest uniquely follow.

Results and proofs

Lemma 8. *An hn-buyer (lo-seller) prefers technology-k if and only if $\zeta^k_{hn} \geq \zeta^{-k}_{hn}$ ($\zeta^k_{lo} \geq \zeta^{-k}_{lo}$).*

Proof. Consider hn-buyers for example. Recall that $\Delta^k_{hn} = R^k_h - R_d = (V^k_{ho} - V^k_{hn}) - R_d$. Recall also that $V^k_{ho} = V_{ho}$ and $V^k_{ln} = V_{ln}$ for both k. Hence, the HJB for hn-buyers can be written as:

$$\lambda_d \cdot \left(V_{ln} - V^k_{hn}\right) - rV^k_{hn} + \zeta^k_{hn} \cdot \left(V_{ho} - R_d - V^k_{hn}\right) = 0$$

from which it gives

$$V^k_{hn} = \frac{\lambda_d V_{ln} + \zeta^k_{hn} \cdot (V_{ho} - R_d)}{r + \lambda_d + \zeta^k_{hn}}.$$

The deriative of V^k_{hn} with respect to ζ^k_{hn} is

$$\frac{(V_{ho} - R_d)r + (V_{ho} - V_{ln} - R_d)\lambda_d}{(r + \lambda_d + \zeta^k_{hn})^2}.$$

Note that $\Delta^k_{hn} = V_{ho} - R_d - V^k_{hn} > 0$, implying $V_{ho} - R_d > 0$. Also, $V_{ho} - V_{ln} - R_d = (V_{ho} - V^k_{hn}) - R_d + (V^k_{hn} - V_{ln}) = R^k_h - R_d + (V^k_{hn} - V_{ln}) > 0$. Hence, this derivatie is strictly positive; i.e., V^k_{hn} increases in ζ^k_{hn}. Note that V^k_{hn} has the same form as a function of ζ^k_{hn} for both k. Therefore, $V^k_{hn} \geq V^{-k}_{hn}$ if and only if $\zeta^k_{hn} \geq \zeta^{-k}_{hn}$. The same analysis applies to lo-sellers and is omitted here. □

Lemma 9. *Suppose the two technologies are characterized by $\rho^{SMS} \geq \rho^{BB}$ (> 0), and $n^{SMS} > n^{BB} = 1$. there exists a unique $\mu^* \in (0,1)$ such that*

$$\text{sign}\Big(\zeta(\mu; \rho^{SMS}, q^{SMS}, n^{SMS}) - \zeta(\mu; \rho^{BB}, q^{BB}, n^{BB})\Big) = \text{sign}(\mu - \mu^*).$$

If $\rho^{SMS} q^{SMS} n^{SMS} \geq \rho^{BB} q^{BB} n^{BB}$, then $\zeta(\mu; \rho^{SMS}, q^{SMS}, n^{SMS}) > \zeta(\mu; \rho^{BB}, q^{BB}, n^{BB})$ for all $\mu \in (0,1)$.

Proof. Consider the case $\rho^{\text{SMS}} q^{\text{SMS}} n^{\text{SMS}} < \rho^{\text{BB}} q^{\text{BB}} n^{\text{BB}}$. The general idea is to characterize the shapes of $\zeta^{\text{BB}}(\mu;\cdot)$ and $\zeta^{\text{SMS}}(\mu;\cdot)$. In particular, it will be shown that ζ^{BB} is linearly increasing in μ, while ζ^{SMS} is sigmoid-shaped in μ, starts below ζ^{BB} for small μ; and the two satisfy $\zeta^{\text{BB}}(0) = \zeta^{\text{SMS}}(0) = 0$ and $\zeta^{\text{BB}}(1) < \zeta^{\text{SMS}}(1)$. Therefore, there is always one and only one intersection point $\mu^* \in (0,1)$.

Consider ζ^{BB} first. With $n = 1$, $\zeta^{\text{BB}} = q\rho\mu$, which is linearly increasing from 0 at $\mu = 0$ to $q\rho$ at $\mu = 1$. In particular, $\zeta^{\text{BB}}(0) = 0$ and $\zeta^{BB}(1) = q^{\text{BB}}\rho^{\text{BB}}$.

Next, consider ζ^{SMS}. With $n > 1$, $\zeta^{\text{SMS}} = \big(1 - (1-\mu)^{n-1}(1 - \mu + (1-q)n\mu)\big)\rho$, whose first-order derivative with respect to μ is

$$\frac{\partial \zeta}{\partial \mu} = -n\rho(1-\mu)^{n-2}(\mu(1-n) + q(\mu n - 1)) > 0.$$

To see that the derivative is indeed positive, note that the bracketed term, $\mu(1-n) + q(\mu n - 1)$ is linear in μ and is negative for both $\mu = 0$ and $\mu = 1$ and so it is negative for all μ. Thus, $\zeta(\mu;\cdot)$ is strictly monotone increasing on $\mu \in [0,1]$. Its second-order derivative with respect to μ is

$$\frac{\partial^2 \zeta}{\partial \mu^2} = (n-1)n\rho(1-\mu)^{n-3}(\mu + \mu(-n) + q(\mu n - 2) + 1),$$

which is positive if and only if $\mu < \frac{1-2q}{n-1-nq}$.[16] Summarizing the above, ζ^{SMS} is sigmoid-shaped on $\mu \in [0,1]$: it is monotone increasing, initially convex, but eventually concave.

Now note that $\zeta^{\text{SMS}}|_{\mu \downarrow 0} = \zeta^{\text{BB}}|_{\mu \downarrow 0} = 0$ and that the assumption $\rho^{\text{SMS}} q^{\text{SMS}} n^{\text{SMS}} < \rho^{\text{BB}} q^{\text{BB}} n^{\text{BB}}$ and Lemma 7 imply that for small μ ζ^{SMS} is below ζ^{BB}. Finally, note that at $\mu \uparrow 1$, $\zeta^{\text{SMS}} \to \rho^{\text{SMS}} \geq \rho^{\text{BB}} \geq q^{\text{BB}}\rho^{\text{BB}}$. That is, the sigmoid-shaped ζ^{SMS} exceeds ζ^{BB} eventually. Therefore, there exists a unique $\mu^* \in (0,1)$ at which $\zeta_t^{\text{SMS}}(\mu^*) = \zeta_t^{\text{BB}}(\mu^*)$.

In the case $\rho^{\text{SMS}} q^{\text{SMS}} n^{\text{SMS}} \geq \rho^{\text{BB}} q^{\text{BB}} n^{\text{BB}}$ it can be shown that ζ^{SMS} is either sigmoid-shaped or convex, starts above ζ^{BB} for small μ and also exceeds it at $\mu = 1$. Thus, the only intersection possible is at $\mu = 0$.

□

Lemma 10. *For any integer $n > 2$ there are no solutions to $\rho^{SMS}\big(1 - 0.5^n(1 + n(1 - q^{SMS}))\big) - 0.5q^{BB}\rho^{BB} < 0$.*

Proof. Denote $f(n) = \rho^{\text{SMS}}\big(1 - 0.5^n(1 + n(1 - q^{\text{SMS}}))\big) - 0.5q^{\text{BB}}\rho^{\text{BB}}$.

[16] Note that $\frac{1-2q}{n-1-nq} > 0$, since $\rho^{\text{SMS}} q^{\text{SMS}} n^{\text{SMS}} < \rho^{\text{BB}} q^{\text{BB}} n^{\text{BB}}$ implies $q^{\text{SMS}} < 1/n^{\text{SMS}} \leq 1/2$. It can also be shown that $\frac{1-2q}{n-1-nq} < 1$.

Step 1. $f(n)$ increases in n. Indeed, computing the derivative yields

$$f'(n) = -\rho^{\text{SMS}}2^{-n}(\log(2)(n(q^{\text{SMS}} - 1) - 1) - q^{\text{SMS}} + 1) > 0.$$

(To see that the derivative is positive, one can verify it is positive for both $q^{\text{SMS}} = 0$ and $q^{\text{SMS}} = 1$. Given that $f'(\cdot)$ is linear in q, we have that $f'(\cdot) < 0 \, \forall q \in [0,1]$)

Step 2. $f(3) \geq 0$.

We have that

$$\begin{aligned} f(3) &= -0.5q^{\text{BB}}\rho^{\text{BB}} + \left(0.5 + 0.375q^{\text{SMS}}\right)\rho^{\text{SMS}} \\ &\geq -0.5q^{\text{BB}}\rho^{\text{BB}} + 0.5\rho^{\text{SMS}} \\ &\geq 0. \end{aligned}$$

□

Proposition A.1 (Equilibrium characterization). *Fix the technology parameters $\{\rho, q, n\}$ for both BB and SMS. Suppose the implied threshold μ^* from Lemma* **??** *satisfies $0 < \mu^* < 1/2$, . There exist thresholds $0 < s_1 < s_2 < s_2' < s_1' < 1 + m_d$ for the asset supply s, so that a symmetric technology choice equilibrium $\{m_{do}, \theta_{lo}, \theta_{hn}\}$ is given by*

- *if $0 < s < s_1$, then $\theta_{lo} = 1$, $\theta_{hn} = 0$, and m_{do} is solved uniquely by Equation* (A.4)*;*
- *if $s_1 \leq s \leq s_2$, then $m_{do} = \mu^* m_d$, $\theta_{lo} = 1$, and θ_{hn} is solved uniquely by Equation* (A.4)*;*
- *if $s_2 < s < s_2'$, then $\theta_{lo} = 1$, $\theta_{hn} = 1$, and m_{do} is solved uniquely by Equation* (A.4)*;*
- *if $s_2' \leq s \leq s_1'$, then $m_{do} = (1-\mu^*)m_d$, $\theta_{hn} = 1$, and θ_{lo} is solved uniquely by Equation* (A.4)*; and*
- *if $s_1' < s < 1 + m_d$, then $\theta_{lo} = 0$ and $\theta_{hn} = 1$, and m_{do} is solved uniquely by Equation* (A.4)*.*

The thresholds are the respective unique solution to Equation (A.4) *for $\{s, \theta_{lo}, \theta_{hn}, m_{do}\} \in \big\{\{s_1, 1, 0, \mu^* m_d\}, \{s_2, 1, 1, \mu^* m_d\}, \{s_2', 1, 1, (1-\mu^*)m_d\}, \{s_1', 0, 1, (1-\mu^*)m_d\}\big\}$.*

Proof. □

B Testable implications about price dispersion

References

Ambrose, Brent W., Nianyun (Kelly) Cai, and Jean Helwege. 2008. "Forced Selling of Fallen Angels." *The Journal of Fixed Income* 18 (1):72–85.

Arefeva, Alina. 2017. "How Auctions Amplify House-Price Fluctuations." Working paper.

Babus, Ana and Cecilia Parlatore. 2017. "Strategic Fragmented Markets." Working paper.

Bessembinder, Hendrik, Chester Spatt, and Kumar Venkataraman. 2020. "A Survey of the Microstructure of Fixed-Income Markets." *Journal of Financial and Quantitative Analysis* 55 (1):1–45.

Burdett, Kenneth and Kenneth L. Judd. 1983. "Equilibrium Price Dispersion." *Econometrica* 51 (4):955–969.

Butters, Gerard R. 1977. "Equilibrium Distributions of Sales and Advertising Prices." *The Review of Economic Studies* 44 (3):465–491.

Chowdhry, Bhagwan and Vikram Nanda. 1991. "Multimarket Trading and Market Liquidity." *Review of Financial Studies* 4 (3):483–511.

Colliard, Jean-Edouard, Thierry Foucault, and Peter Hoffmann. 2020. "Inventory Management, Dealers' Connections, and Prices in OTC Markets." Working paper.

Duffie, Darrell, Piotr Dworczak, and Haoxiang Zhu. 2017. "Benchmarks in Search Markets." *The Journal of Finance* 72 (5):1983–2044.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen. 2005. "Over-the-Counter Markets." *Econometrica* 73 (6):1815–1847.

———. 2007. "Valuation in Over-the-Counter Markets." *The Review of Financial Studies* 20 (6):1865–1900.

Duffie, Darrell, Lei Qiao, and Yeneng Sun. 2019. "Continuous-time Random Matching." Working paper.

Duffie, Darrell and Yeneng Sun. 2007. "Existence of Independent Random Matching." *The Annals of Applied Probability* 17 (1):386–419.

———. 2012. "The exact law of large numbers for independent random matching." *Journal of Economic Theory* 147:1105–1139.

Dugast, Jérôme, Semih Üslü, and Pierre-Olivier Weill. 2019. "A Theory of Participation in OTC and Centralized Markets." Working paper.

Ellul, Andrew, Chotibhak Jotikasthria, and Christian T. Lundblad. 2011. "Regulatory pressure and fire sales in the corporate bond market." *Journal of Financial Economics* 101 (3):596–620.

Fermanian, Jean-David, Olivier Guéant, and Jiang Pu. 2017. "The behavior of dealers and clients on the European corporate bond market: the case of Multi-Dealer-to-Client platforms." Working paper. URL https://arxiv.org/abs/1511.07773.

Glode, Vincent and Christian C. Opp. 2019. "Over-the-Counter versus Limit-Order Markets: The Role of Traders' Expertise." *The Review of Financial Studies* Forthcoming.

Hendershott, Terrence, Dan Li, Dmitry Livdan, and Norman Schürhoff. 2020. "Fake Liquidity?" Working paper.

Hendershott, Terrence and Ananth Madhavan. 2015. "Click or Call? Auction versus Search in the Over-the-Counter Market." *The Journal of Finance* 70 (1):419–447.

Hollifield, Burton, Artem Neklyudov, and Chester Spatt. 2017. "Bid-Ask Spreads, Trading Networks, and the Pricing of Securitizations." *The Review of Financial Studies* 30 (10):3048–3085.

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill. 2020. "Frictional Intermediation in Over-the-Counter Markets." *The Review of Economics Studies* 87 (3):1432–1469.

Jovanovic, Boyan and Albert J. Menkveld. 2015. "Dispersion and Skewness of Bid Prices." Working paper.

Klemperer, Paul. 1999. "Auction theory: A guide to the literature." *Journal of Economic Surveys* 13 (3):227–286.

Lagos, Ricardo and Guillaume Rocheteau. 2009. "Liquidity in Asset Markets with Search Frictions." *Econometrica* 77 (2):403–426.

Lagos, Ricardo, Guillaume Rocheteau, and Pierre-Olivier Weill. 2011. "Crises and Liquidity in Over-the-Counter Markets." *Journal of Economic Theory* 146 (6):2169–2205.

Lee, Tommy and Chaojun Wang. 2019. "Why Trade Over-the-Counter? When Investors Want Price Discrimination." Working paper.

Li, Dan and Norman Schurhoff. 2019. "Dealer Networks." *The Journal of Finance* 74 (1):91–144.

Liu, Ying, Sebastian Vogel, and Yuan Zhang. 2017. "Electronic Trading in OTC Markets vs. Centralized Exchange." Working paper.

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

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18 (2):155–162.

Neklyudov, Artem. 2019. "Bid-Ask Spreads and the Over-the-Counter Interdealer Markets: Core and Peripheral Dealers." *Review of Economic Dynamics* 33:57–84.

O'Hara, Maureen and Xing Zhou. 2020. "The Electronic Evolution of Corporate Bond Dealers." *Journal of Financial Economics* Forthcoming.

Pagano, Marco. 1989. "Trading Volume and Asset Liquidity." *The Quarterly Journal of Economics* 104 (2):255–274.

Riggs, Lynn, Esen Onur, David Reiffen, and Haoxiang Zhu. 2019. "Swap Trading after Dodd-Frank: Evidence from Index CDS." *Journal of Financial Economics* Forthcoming.

Saar, Gideon, Jian Sun, Ron Yang, and Haoxiang Zhu. 2019. "From Market Making to Matchmaking: Does Bank Regulation Harm Market Liquidity?" Working paper.

Shen, Ji, Bin Wei, and Hongjun Yan. 2018. "Financial Intermediation Chains in an OTC Market." Working paper.

Sun, Yeneng. 2006. "The exact law of large numbers via Fubini extension and characterization of insurable risks." *Journal of Economic Theory* 126:31–69.

Üslü, Semih. 2019. "Pricing and Liquidity in Decentralized Asset Markets." *Econometrica* Forthcoming.

Varian, Hal R. 1980. "A Model of Sales." *American Economic Review* 70 (4):651–659.

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

Vayanos, Dimitri and Pierre-Olivier Weill. 2008. "A Search-Based Theory of the On-the-Run Phenomenon." *The Journal of Finance* 63 (3):1361–1398.

Vogel, Sebastian. 2019. "When to Introduce Electronic Trading Platforms in Over-the-Counter Markets?" Working paper.

Weill, Pierre-Olivier. 2007. "Leaning against the Wind." *Review of Economic Studies* 74:1329–1354.

Yang, Ming and Yao Zeng. 2018. "The Coordination of Intermediation." Working paper.

Yueshen, Bart Zhou. 2017. "Uncertain Market Making." Working paper.

Zhang, Shengxing. 2018. "Liquidity Misallocation in an Over-the-Counter Market." *Journal of Economic Theory* 174:16–56.

Zhong, Zhuo. 2014. "The Risk Sharing Benefit versus the Collateral Cost: The Formation of the Inter-Dealer Network in Over-the-Counter Trading." Working paper.

List of Figures