From Market Making to Matchmaking: Does Bank Regulation Harm Market Liquidity?

Gideon Saar, Jian Sun, Ron Yang, and Haoxiang Zhu*

This version: February 2022

Abstract

Post-crisis bank regulations raised market-making costs for bank-affiliated dealers. We show that this can, somewhat surprisingly, improve overall investor welfare and reduce average transaction costs despite the increased cost of immediacy. Bank dealers in OTC markets optimize between two parallel trading mechanisms: market making and matchmaking. Bank regulations that increase market-making costs change the market structure by intensifying competitive pressure from non-bank dealers and incentivizing bank dealers to shift their business activities toward matchmaking. Thus, post-crisis bank regulations have the (unintended) benefit of replacing costly bank balance sheets with a more efficient form of financial intermediation.

Keywords: bank regulation, market making, matchmaking, financial crisis, corporate bonds, liquidity, over-the-counter markets, broker-dealers, Basel 2.5, Basel III, Volcker Rule, post-crisis regulation, market microstructure

JEL Classification: G01, G12, G21, G24, G28

*Gideon Saar is affiliated with the Johnson Graduate School of Management, Cornell University (gs25@cornell.edu). Jian Sun is from the MIT Sloan School of Management (jiansun@mit.edu). Ron Yang is affiliated with Harvard Business School (rnyang@g.harvard.edu). Haoxiang Zhu is on leave from the MIT Sloan School of Management and the NBER (zhuh@mit.edu) while serving as Director of the Division of Trading and Markets at the Securities and Exchange Commission. Haoxiang Zhu's work on this paper was completed prior to December 10, 2021. We thank Philip Bond, Briana Chang, Christopher Hrdlicka, Stacey Jacobsen, Pete Kyle, Jonah Platt, Dimitri Vayanos, Kumar Venkataraman, Yajun Wang, Yao Zeng, Zhuo Zhong, and Hao Zhou as well as seminar/conference participants at the American Finance Association 2020 meeting, the Chinese University of Hong Kong, the Chinese University of Hong Kong Shenzhen, the City University of Hong Kong, Cornell University, the Financial Industry Regulatory Authority, Manchester Business School, MIT, the PBC School of Finance, the University at Buffalo, the University of British Columbia, the University of Illinois, the University of Washington, Warwick Business School, the Wharton School, the 4th Sydney Market Microstructure Meeting, and the mini-conference on Regulatory Reform at the University of Wisconsin-Madison for helpful comments.

The aftermath of the financial crisis saw several regulatory initiatives designed to curtail risk-taking by banks, hindering proprietary trading and increasing the cost of market making. In part, these initiatives reflected a widespread belief in banking regulatory circles that the pre-crisis price of immediacy did not adequately incorporate the costs required to ensure that market makers are supported by sufficient capital and do not become a source of illiquidity contagion (see, for example, BIS Committee on the Global Financial System (2014, 2016)). While the regulations were meant to improve market-maker resilience, the post-crisis changes in the Basel framework (Basel 2.5 and Basel III) and the Volcker Rule have reduced banks' willingness to accommodate corporate bond trades on their balance sheets and in general have made their market-making operations more costly.[1] Some market observers have portrayed these regulatory initiatives in the light of a trade-off between market resilience (less severe contagion) in times of stress and liquidity during normal times. Reduced market liquidity during normal times, as the argument goes, may be a necessary compromise to achieve enhanced market resilience during stressful periods.

Against this backdrop, a growing body of empirical literature set out to investigate the impact of post-crisis regulations on liquidity, and the US corporate bond market has been the most commonly studied setting in this context given its large size and dealer-centric nature. Some papers indeed document an increase in the cost of immediacy (Bao, O'Hara, and Zhou (2018), Choi and Huh (2017), and Dick-Nielsen and Rossi (2018)), but, surprisingly, this literature finds on balance an improvement or at least no deterioration in the average transaction costs of corporate bond trades (Mizrach (2015), Adrian, Fleming, Shachar, and Vogt (2017), Anderson and Stulz (2017), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), and Trebbi and Xiao (2019)). These findings are accompanied by a shift in the execution modality of trades. The over-the-counter corporate bond market features two parallel trading mechanisms that correspond to the dual capacity of broker-dealers: market making, where the dealer provides immediacy to customers by taking bonds onto his balance sheet, and matchmaking, where the dealer searches for counterparties for his customers' orders. There is ample evidence that matchmaking has increased following

[1]For example, revisions of the Basel II framework ("Basel 2.5") increased inventory costs for bonds through the incremental risk capital (IRC) charge and the trading book's stressed VaR requirement. Basel III added the enhanced supplementary leverage ratio (e-SLR) that is widely viewed as a binding constraint on bank-affiliated dealers.

the implementation of post-crisis regulations (e.g., Bao, O'Hara, and Zhou (2018), Choi and Huh (2017), Schultz (2017)), and as a result the execution of large trades now tends to require more time (BIS Committee on the Global Financial System (2014, 2016)). This shift towards matchmaking was driven by bank-affiliated dealers that reduced the amount of capital they commit to market making (Bao, O'Hara, and Zhou (2018), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), and Choi and Huh (2017)), whereas non-bank dealers increased capital commitments to principal trading and decreased their matchmaking activity.

While the increase in the cost of immediacy is consistent with the regulation-induced higher cost of taking bonds onto banks' balance sheets, the decline in average transaction costs could suggest that a shift from market making to matchmaking benefits customers. Yet, transaction costs are an incomplete measure of overall customer welfare because of two less measurable but potentially important costs. First, the average time it takes to execute a transaction in the post-regulation era is likely longer. These execution delays may be costly to investors. Second, realized transaction costs capture only trades that were executed. If customers forgo transacting in response to the higher costs of immediacy (and their unwillingness to wait longer for execution), their welfare loss cannot be ascertained by analyzing executed trades. To assess the change in overall customer welfare, one needs a model that explicitly considers the trade-off between the cost of delay (matchmaking) and the cost of immediacy (market making), and captures customers' option to forgo trading altogether. This is the model we set out to investigate.

Our model features two representative intermediaries standing for two groups of dealers: bank-affiliated dealers and non-bank-affiliated dealers. A bank dealer offers customers a market-making service (incurring a balance sheet cost for taking on the bonds) and a matchmaking service (incurring a cost to help customers search for counterparties). To keep our benchmark model simple, a non-bank dealer offers only a market-making service, but in Section 4 we allow her to offer both services as we examine various extensions to ensure the robustness of our conclusions. The two dealers set prices for the services they offer to maximize profits. The balance sheet costs of the bank dealer and the non-bank dealer are different, reflecting, among other things, the costs imposed by bank regulations. Infinitesimal customers (buyers and sellers) arrive at a constant rate and wish to trade bonds. The customers are price takers and heterogeneous with regard to patience (or the value they attach to immediacy). Utilizing a market-making service allows them to trade immediately by paying a spread, while using a matchmaking service to search for a counterparty takes time and they incur both the cost of waiting for a match and a trading fee. Customers can choose any trading mechanism offered by the dealers, and they optimize over their choice regarding whether and how to trade.

Three key driving forces operate in the model. First, dealers wield market power over customers.[2] Second, the bank dealer offers two distinct trading mechanisms, setting prices in each mechanism to extract the most rents from customers. Third, potential competition from the non-bank dealer can help discipline the bank dealer's pricing, and therefore the extent of competition matters for utilization of the two trading mechanisms. The industrial organization aspect, manifested as competitive pressure from the non-bank dealer, and the market-microstructure perspective, which introduces substitution between the two trading mechanisms, join to deliver the richness of our implications.

We focus on how customer welfare and market outcomes change when the bank dealer's regulatory costs increase. When the balance sheet cost of the bank dealer is much lower than that of the non-bank dealer, he is unconstrained by competition and hence passes some of the cost increase to the market-making customers. Faced with a higher spread, some customers choose to shift from the market-making service to the matchmaking service. Average transaction costs increase initially but may start declining as more customers shift to matchmaking because the equilibrium matchmaking fee is lower than the market-making spread to compensate for the delay in executing trades. Even if average transaction costs decline, however, overall customer welfare unequivocally worsens because the bank dealer sets the matchmaking fee high to maximize profitability, and hence both the switching customers and those utilizing the market-making service are worse off. Unconstrained by competition, some portion of the increase in the balance sheet cost of the bank dealer in this equilibrium is passed on to customers, harming their welfare. This result is reminiscent of warnings

[2]Evidence supports the notion that dealers in the corporate bond market have market power. For example, it has long been established that per-share transaction costs in corporate bonds decline in trade size (e.g., Schultz (2001)), even though fixed costs do not appear to be very high.

made by some market observers that raising banks' costs would hurt investors in the corporate bond market.

When the bank dealer's balance sheet cost rises enough, however, his ability to pass on the increase in costs to customers is constrained by competitive pressure in market making from the non-bank dealer. Shrinking market-making profit margins incentivize the bank dealer to shift more of his business to the matchmaking mechanism by reducing the matchmaking fee. The lower fee also attracts customers who previously chose not to trade, resulting in higher volume. As a result, overall customer welfare, which takes into account not just transaction costs but also waiting costs and the welfare of customers who choose not to trade, unambiguously rises. While a shift from market making to matchmaking as the balance sheet costs of the bank dealer rise is not necessarily surprising, the reason behind the shift depends on the extent of competition between the bank and non-bank dealers. This dependence—a higher market-making spread in the unconstrained equilibrium versus a lower matchmaking fee in the constrained equilibrium—is the key to generating the more surprising implication of the model that increasing bank regulatory costs can improve customer welfare. The interaction of "internal" competition between the two trading mechanisms and "external" competition between the bank and non-bank dealers is necessary to generate these welfare gains.

We keep our main model parsimonious to show how and when an increase in bank regulatory costs can improve customer welfare in a transparent framework that highlights the key features of the environment that are necessary to deliver the results. We present several extensions of our model to demonstrate that these insights are robust to changing important dimensions of the analysis. Our first extension shows that allowing the non-bank dealer to compete in both market making and matchmaking does not change our conclusions. When the increase in bank regulatory costs shrinks the bank dealer's market-making profit margins, he responds by reducing his matchmaking fee much more aggressively than the non-bank dealer reduces hers. This results in the bank dealer gaining matchmaking market share relative to the non-bank dealer (which is consistent with empirical findings) to the benefit of the customers, whose overall welfare increases when competition between these two dealers intensifies.

With the second extension of our model we investigate the robustness of our results to replacing the Bertrand competition we use in the main model with Cournot quantity competition. An increase in regulatory costs causes the bank dealer to reduce his provision of market making while the non-bank dealer increases her market-making activity, exactly as documented in empirical studies. Overall customer welfare shows the same pattern as in our original model: it increases as bank regulatory costs rise beyond a certain threshold and competition between the two dealers intensifies. Lastly, we examine the importance of market power as a driver of the increase in welfare by studying an extension that features multiple bank and non-bank dealers. Moving away from the representative dealer framework does not eliminate our welfare result. Instead, the parameter range over which an increase in bank regulatory costs enhances customer welfare shrinks gradually as the number of competing bank dealers increases and the market power friction that drives our results eases.

Our model contributes to the debate over financial regulation by demonstrating how an increase in regulatory costs can serve as a catalyst for a healthy market-structure transition by removing obstacles to competition. Regulators often worry about the unintended consequences of their regulatory interventions. While increasing the costs of balance sheet financing was meant to enhance bank dealers' resilience in times of market stress, we believe that one (perhaps) "unintended" benefit of these bank regulations was to push bank dealers to enhance their matchmaking services. Our model shows that such a transition can potentially improve overall customer welfare during normal times and therefore materially changes the supposed trade-off between resilience in times of stress and day-to-day liquidity. Rather than dictating a particular market structure for trading securities, our work highlights the role that regulations can play in affecting the industrial organization of dealers in a manner that changes the market microstructure to the benefit of investors.

1 Market Making versus Matchmaking: Theory

Several recent theoretical papers recognize the importance of the dual mechanisms for trading bonds, namely market making and matchmaking, although each of these papers adopts a distinct approach to studying the two mechanisms. An, Song, and Zhang (2017) study intermediation chains by modeling interaction between one seller, a finite number of dealers, and an infinite number of buyers. Their model shows how an intermediary rat race gives rise to an inefficient amount of principal trading. Our paper does not feature an inter-dealer market or intermediation chains, but rather focuses on what happens to customer welfare and the market environment if the cost of market making increases for bank dealers. An and Zheng (2017) look at how the dual capacity of broker-dealers (principal and agency trading) gives rise to a conflict of interest, which results in dealers' holding too much inventory as a tool for extracting rent from customers. Unlike in our framework, customers in their model do not optimize and matchmaking is effortless and costless, leading An and Zheng to focus on inventory as a strategic variable. In contrast, we model a two-way market with balanced customer order flows and abstract away from inventory management—an approach that is orthogonal to that of An and Zheng (2017). Furthermore, we investigate the endogenous evolution of trading mechanisms by (i) having customers optimally choose whether and how to trade and (ii) having dealers optimize the pricing of their services.

Li and Li (2017) model a trade-off between inventory costs (in market making) and verification costs (in matchmaking). Moral hazard in matchmaking arises in their model when a dealer gains by providing worse executions for a customer. Because dealers have better information than customers, transparency influences the prevalence of market making over matchmaking.[3] Transparency plays no role in our model because we assume homogeneous common-value information. Instead, we emphasize competition from non-bank dealers and the role it plays in determining customer welfare and the extent of matchmaking.[4]

The paper closest to ours in objective is Cimon and Garriott (2019). In their model, market makers compete for quantity (Cournot) in separate buyer and seller markets and issue equity and debt to fund their operations. Market making is modeled as a more efficient form of trading than matchmaking, and therefore increased agency trading implies a higher price impact of trades. As a result, regulations that increase the cost of market making must hurt liquidity. In contrast, in our

[3] Li and Li also provide empirical results pertaining to the share of matchmaking around the financial crisis and how this share relates to transparency and volume.

[4] Chang and Zhang (2020) model endogenous network formation and discuss how network structure may change as a result of OTC reforms. Their framework also does not consider competition from non-bank dealers, which is central to our results.

model the equilibrium matchmaking fee is lower than the market-making spread, and the market power of dealers provides a role for bank regulations in enhancing competition and improving customer welfare.

Recently, Kargar, Lester, Lindsay, Liu, Weill, and Zuniga (2021) examine principal and agency trading in corporate bonds during the COVID-19 crisis. They construct a parsimonious model of market making versus matchmaking and use it to conduct structural estimation. They show that, as principal trading became more expensive during the COVID-19 crisis, customers substituted by shifting to agency trades. This made the decline in customers' surplus from immediacy much smaller than the increase in their expenditures on immediacy. The Fed's introduction of credit facilities impacted both customer demand and the costs and risks of balance-sheet financing, making principal trading less expensive to customers and increasing the utilization of dealer balance sheets to carry inventory.

Our model shares certain features that are examined in the vast industrial organization (IO) literature, in particular the branch that investigates multiproduct competition. To the best of our knowledge, however, no other paper captures all the important characteristics of bond markets that we wish to model. Mussa and Rosen (1978) characterize the optimal pricing strategy of a monopolist over a range of products that differ in quality. Katz (1984) analyzes competition between various multiproduct firms. He shows that because competition in one product spills over to another, endogenous specialization can arise. Johnson and Myatt (2003) consider duopoly competition between a multiproduct incumbent and a multiproduct entrant, where both face the same costs. The entrant in their model is assumed to focus on low-quality products, and the incumbent's equilibrium products are shown to be of weakly higher quality than the entrant's products. Nocke and Schutz (2018) show that, in a fairly general multiproduct setting, increasing competition leads to an expanded product offering because a firm worries less about cannibalizing its other products when facing more intense outside competition.

Our results are distinct from this line of the IO literature in at least two ways. First, we allow differentiated costs between the bank dealer and the non-bank dealer. This phenomenon of "same activity, different costs" is salient in financial markets that are regulated based on the types of entities involved. Our model predicts that the bank dealer expands into matchmaking when it still has a cost advantage, which matches the empirical fact that bank dealers change their business models even when they maintain overall dominance in liquidity provision. Second, we show that, under some conditions, customer welfare increases when production costs rise. This message is not present in the papers cited above.

Lastly, the increase in balance sheet costs in our model resembles an increase in taxes. Weyl and Fabinger (2013) characterize the pass-through of taxes to consumers when firms compete imperfectly in an oligopoly market of a single good, showing that, under some conditions, the pass-through can exceed one. In a more stylized setting with two goods (market making and matchmaking), we show that a higher tax (balance sheet cost) on market making can lead to a net *negative pass-through* to customers, manifested by lower quantity-weighted average transaction costs and higher customer welfare.

2 The Model

Time is continuous, $t \in [0, \infty)$. The traded asset has an expected fundamental value of v. All customers and dealers are risk neutral and have the same information about the fundamental value of the asset. The discount rate is $r > 0$.[5]

Customers and dealers. Infinitesimal buyers arrive in the market at rate μ; that is, the mass of buyers arriving during the time $(t, t + dt)$ is μdt. Each buyer wishes to buy one unit of the asset, and her private benefit (or "value") for trading immediately is an i.i.d. random variable $x \in [0, \infty)$, with cumulative distribution function G. Heterogeneity in this private value reflects the manner in which we model differences across customers in degree of patience.[6] Likewise, infinitesimal sellers arrive in the market at the same rate μ, and their private benefit for selling the asset immediately is also distributed according to G. A customer's need for immediacy is not observable by others, and the customer exits the market upon trading.

[5] We use the discount rate r to capture two effects: the rate at which customers and dealers discount future profits and the rate at which trading opportunities decay over time.

[6] The arrival process of buyers is time-invariant in the sense that the types of buyers arriving during each small time interval $(t, t + dt)$ are distributed according to G.

We make the familiar monotone-hazard-rate assumption with respect to the distribution of private value. In our context, this assumption simplifies the proofs by guaranteeing a unique equilibrium in some parameter ranges and helping to sign comparative statics when bank regulatory costs increase. While it is not entirely innocuous, the assumption of a non-decreasing hazard rate has been used extensively in the mechanism-design literature (see Fudenberg and Tirole (1991), Chapter 7). A non-decreasing hazard rate is equivalent to the log-concavity of the reliability function $1 - G(\cdot)$ and is satisfied by many common distributions, including uniform, normal, exponential, logistic, extreme value, Laplace, and, under some parametric restrictions, power, Weibull, Gamma, Chi-squared, and Beta (see Bagnoli and Bergstrom (2005)). For the sake of convenience, we state this assumption in terms of the inverse hazard function (or Mills ratio) of G,

$$\zeta(x) = \frac{1 - G(x)}{G'(x)},$$

and specify that $\zeta(x)$ is non-increasing in x. We stress that while customers' desire to trade in the model—motivated by risk-sharing, liquidity needs, and other non-informational reasons—is specified exogenously as is standard in many models, both the quantity of trading and its composition (market making versus matchmaking) arise endogenously.

Our model features two representative yet distinct strategic intermediaries, called dealers, who help customers trade the asset. One of the dealers is bank-affiliated and is subject to bank regulations, whereas the other dealer is unaffiliated with any bank and hence is not subject to bank regulations. The main friction in our model is that these dealers have market power when trading with customers. Each representative dealer optimally sets prices to maximize profits, subject to competition from the other representative dealer. In Section 4, we investigate how the extent of dealer market power impacts our main conclusions using an extension of the model with multiple bank and non-bank dealers.

Trading protocols: market making and matchmaking. Dealers provide liquidity in two ways: "market making" (as principal) or "matchmaking" (as agent). The market-making mechanism allows customers to trade immediately, while searching for a counterparty using the matchmaking mechanism takes time. Under the market-making protocol, a dealer immediately fills a customer's buy or sell order from his own balance sheet by incurring a balance sheet cost. In return, the bank (non-bank) dealer charges customers an endogenous per-unit spread of $S_B > 0$ ($S_{NB} > 0$), which is publicly observable. The bank (non-bank) dealer has a constant per-unit balance sheet cost c_B (c_{NB}) regardless of whether he (she) is accommodating a buy or a sell order. Given this specification of balance sheet costs and the risk-neutrality of dealers in the model, the level of inventory does not affect the spread.[7] As such, there is no loss of generality in assuming equal arrival rates for customers who wish to buy and customers who wish to sell, which simplifies the exposition of the model.

Conceptually, we can think about the bank dealer's balance sheet cost as comprised of three components:

$$c_B = c_{NB} \underbrace{-\mathit{ImplicitSubsidy} + \mathit{PostCrisisRegulatoryCosts}.}_{\text{only applicable to bank dealers}}$$

The first component, c_{NB}, reflects the costs and risks involved in running the market-making business of a generic non-bank dealer. The second and third components are applicable only to bank-affiliated dealers. The second component, *ImplicitSubsidy*, includes various advantages banks enjoy with respect to balance-sheet financing, including low-cost funding via deposits, relatively easy access to the central bank, and the too-big-to-fail subsidy that has been discussed extensively in regulatory circles.[8] This implicit subsidy reduces the capital costs of the bank dealer's trading book relative to that of the non-bank dealer's and hence could enable the bank dealer to offer cheaper liquidity. Post-crisis bank regulations, represented by the third component of c_B, were aimed at increasing the market-making costs of bank dealers to counteract this implicit subsidy. As post-crisis regulations were phrased in, we envision a gradual increase in c_B relative to c_{NB}, and this is also how we conduct comparative statics in the rest of the paper.

In contrast to market making that relies on balance-sheet financing, matchmaking relies on effort

[7] See An and Zheng (2017) for a model of the dual capacity of broker-dealers that focuses on the dealer's choice of inventory level.

[8] The Committee on the Global Financial System of the Bank for International Settlements writes in its report on fixed-income market liquidity that, in the pre-crisis era, "Underpriced liquidity services were predicated on expectations of an implicit public sector backstop for major financial institutions" (BIS Committee on the Global Financial System (2016)).

and technology. Under the matchmaking protocol, the bank dealer searches for a counterparty for the customer's order by incurring an exogenous search cost $I > 0$. The search process takes time, and the customer is matched with a counterparty at an exponentially distributed time τ with exogenous intensity $H \in [0, \infty)$. While the dealer searches, the customer incurs a delay cost because of the time discounting of the private benefit of trading. Given the exponential distribution of matching time τ, the effective discount factor is

$$\mathcal{H} \equiv E[e^{-r\tau}] = \int_{\tau=0}^{\infty} He^{-Hu}e^{-ru}du = \frac{H}{r+H}. \tag{1}$$

Higher $\mathcal{H}$ implies a shorter waiting time (with a lower cost of delay) for searching customers, and hence we refer to $\mathcal{H}$ as the *speed* of matchmaking. When a match is made, the bank dealer receives an endogenous fee f from both the buyer and the seller, and f is disclosed to customers before the search takes place.[9] Matchmaking consists of all dealer-facilitated trading that does not involve taking a trade onto the bank dealer's balance sheet, hence capturing both pure agency trades and riskless-principal trades. For parsimony of exposition, only the bank dealer operates a matchmaking mechanism in our baseline model presented in Section 3. In Section 4, we examine the robustness of our conclusions in an extension of the model that allows both dealers to provide matchmaking services.

Objective Functions. Customers choose between trading immediately with the bank dealer or the non-bank dealer (depending on the spread each dealer charges), searching for a counterparty using the bank dealer's matchmaking service, or not trading at all. From the customer's perspective, the bank and non-bank dealers' market-making services are identical. Therefore, a customer who opts to trade immediately will choose the market-making service that charges the lower spread,

[9] The assumption that both sides pay the same fee is likely without loss of generality from an ex-ante perspective. As Choi and Huh (2017) note, a dealer searching for a counterparty may need to offer better terms of trade to the counterparty to execute the trade. In other words, on a trade-by-trade basis, the side that initiates the search could pay more than the other side. Even if one side pays $f + \epsilon$ and the other side pays $f - \epsilon$ for the match, as long as customers are ex-ante similar and risk neutral, they pay the same expected fee f. One can view $2f$ as the net compensation earned by a dealer that executes both legs of an agency (or riskless principal) cross. We focus on modeling the total matchmaking fee paid by customers.

which we denote by

$$S = \min(S_B, S_{NB}). \tag{2}$$

Recall that x denotes the private benefit a customer obtains from trading immediately. The customer's profit from using a market-making service is $x - S$. Her expected profit from using the matchmaking mechanism offered by the bank dealer, which takes into account the expected waiting cost, is $(x - f)\mathcal{H}$. Her profit from leaving the market without trading is 0. Therefore, a customer prefers matchmaking to not trading if and only if $x \geq f$.

Let b be the value of the marginal customer who is indifferent between matchmaking and market making. The indifference condition is

$$(b - f)\mathcal{H} = b - S, \tag{3}$$

and we obtain

$$b = \frac{S - f\mathcal{H}}{1 - \mathcal{H}}. \tag{4}$$

The customer's optimization problem therefore results in a very simple behavior: do not trade if $x \in [0, f)$, choose matchmaking if $x \in [f, b]$, and choose market making with the dealer offering the lower spread if $x > b$.[10]

Our main objective in this paper is to analyze the impact of regulations on overall customer welfare. Given the two thresholds f and b, we can write the overall welfare of customers aggregated across the three ranges of x as:

$$\pi_c = \frac{2\mu}{r}\left[\underbrace{\int_{x=0}^{f} 0 \cdot dG(x)}_{\text{no trade}} + \underbrace{\int_{x=f}^{b} (x - f)\mathcal{H}dG(x)}_{\text{matchmaking}} + \underbrace{\int_{x=b}^{\infty} (x - S)dG(x)}_{\text{market making}}\right]. \tag{5}$$

The bank dealer's profit is comprised of two components: the matchmaking profit, which depends on the fee and the cost of searching, and the market-making profit (if the bank offers

[10] It is straightforward to show that any equilibrium in which the bank dealer operates the matchmaking service must satisfy $f < S$ (i.e., the matchmaking fee is lower than the market-making spread) because of the waiting costs associated with the search.

the lower spread), such that

$$\pi_B = \frac{2\mu}{r} \left[(\mathcal{H}f - I) \left(G(b) - G(f) \right) + (S - c_B)(1 - G(b)) \mathbb{I}_{S_B \leq S_{NB}} \right], \tag{6}$$

where $\mathbb{I}_{S_B \leq S_{NB}}$ is an indicator function that takes the value 1 if $S_B \leq S_{NB}$ and 0 otherwise.

The non-bank dealer's market-making profit can be expressed as:

$$\pi_{NB} = \frac{2\mu}{r} \left[(S - c_{NB})(1 - G(b)) \mathbb{I}_{S_{NB} < S_B} \right], \tag{7}$$

where $\mathbb{I}_{S_{NB} < S_B}$ is an indicator function that takes the value 1 if $S_{NB} < S_B$ and 0 otherwise.

Equilibrium Definition. An equilibrium consists of:

1. The bank dealer's choices of market-making spread S_B and matchmaking fee f,
2. The non-bank dealer's choice of market-making spread S_{NB}, and
3. Each arriving customer's choice between market making (with one of the dealers), matchmaking, and refraining from trading altogether,

such that dealers and customers maximize expected profits.

3 Equilibrium and Comparative Statics

The equilibrium structure in our model is comprised of four distinct regions depending on the degree of competition between the bank and non-bank dealers. In the first region, $c_B \leq \underline{c}$, the bank dealer's balance sheet cost is so low that his unconstrained monopoly spread is lower than the balance sheet cost of the non-bank dealer (and hence there is no competitive pressure from the non-bank dealer). In the second region, $\underline{c} < c_B \leq c_{NB}$, the bank dealer's spread is constrained by the non-bank dealer's balance sheet cost. In the third region, $c_{NB} < c_B \leq \overline{c}$, the non-bank dealer supplies market-making services, but his spread is constrained by the bank dealer's balance sheet cost. Lastly, in the fourth region, $c_B > \overline{c}$, the bank dealer's regulatory cost is so high that the non-bank dealer sets an unconstrained monopoly spread in the market-making business.

This equilibrium structure does not depend on a particular choice of distribution for the customers' private value (G), and the monotone-hazard-rate assumption enables us to fully characterize how overall customer welfare and market outcomes change in each region as c_B increases. All of the insights, however, can be demonstrated using a simple example in which the customers' private value follows a uniform distribution with cdf $G(x) = x/A$, where A is a positive constant that is set high enough to ensure that in equilibrium there are at least some impatient customers who demand market making services (i.e., $A > b$).[11] This uniform example has the advantage that it allows the intuition behind the results to come out more clearly. Furthermore, using the uniform distribution in our study of how customer welfare changes as c_B increases is essentially without loss of generality in the most relevant equilibrium regions (the unconstrained and constrained bank dealer equilibria) because the comparative statics are the same for all distributions that satisfy the monotone-hazard-rate assumption.[12] Hence, we present the uniform example in the paper and relegate the propositions and proofs for the general setting to the Internet Appendix.

3.1 Existence

When $c_B \leq c_{NB}$, the bank dealer is the more efficient provider of market-making services. Competition from the non-bank dealer may constrain the bank dealer's strategy by forcing $S \leq c_{NB}$, but the equilibrium is entirely about solving the bank dealers's problem of maximizing the expected profit from providing both market-making and matchmaking services:

$$\max_{0 \leq f \leq S \leq c_{NB}} \Pi_B\left(S, f; c_B\right) \equiv \frac{2\mu}{r}\left[\left(\mathcal{H}f - I\right)\left(\frac{S - \mathcal{H}f}{A\left(1-\mathcal{H}\right)} - \frac{f}{A}\right) + \left(S - c_B\right)\left(1 - \frac{S - \mathcal{H}f}{A\left(1-\mathcal{H}\right)}\right)\right]. \quad (8)$$

When $c_B > c_{NB}$, on the other hand, the non-bank dealer is the more efficient provider of market-making services. Thus, we have a form of specialization: market making is provided by the non-bank dealer and matchmaking is provided by the bank dealer. Specifically, given the bank dealer's choice

[11] The specific parametric restriction required to ensure $A > b$ in equilibrium is $A > c_{NB} + \frac{\mathcal{H}}{2(1-\mathcal{H})}\left(c_{NB} - \frac{I}{\mathcal{H}}\right)$.

[12] When $c_B > c_{NB}$, additional assumptions are made for a convex G in the general case. Details are provided in the Internet Appendix.

of f, the non-bank dealer's problem is

$$\pi_{NB}(c_B) = \max_{c_{NB} \le S \le c_B} \Pi_{NB}(S) = \frac{2\mu}{r}\left[(S - c_{NB})\left(1 - \frac{S - \mathcal{H}f}{A(1-\mathcal{H})}\right)\right].$$

The bank dealer's matchmaking business exerts competitive pressure on the non-bank dealer's price-setting behavior via the threshold customer type $b = \frac{S-\mathcal{H}f}{1-\mathcal{H}}$. Given the non-bank dealer's choice of S, the bank dealer sets a fee $f \le S$. Thus, the bank dealer solves the problem

$$\pi_B(c_B) = \max_{I/\mathcal{H} \le f \le S} \Pi_B(f) = \frac{2\mu}{r}(\mathcal{H}f - I)\left(\frac{S - \mathcal{H}f}{A(1-\mathcal{H})} - \frac{f}{A}\right).$$

The following proposition establishes the existence of an equilibrium.

Proposition 1. *There exists an equilibrium for any $c_B > I/\mathcal{H}$ such that,*[13]

- *If $c_B < \underline{c} = 2c_{NB} - A$, there is an* ***unconstrained bank dealer equilibrium*** *in which the bank dealer sets $S^* = \frac{1}{2}(c_B + A) < c_{NB}$ and $f^* = \frac{1}{2}\left(\frac{I}{\mathcal{H}} + A\right)$.*
- *If $\underline{c} = 2c_{NB} - A \le c_B \le c_{NB}$, there is a* ***constrained bank dealer equilibrium*** *in which the bank dealer sets $S^* = c_{NB}$ and $f^* = \frac{1}{2}\left(\frac{I}{\mathcal{H}} + 2c_{NB} - c_B\right)$.*
- *If $c_{NB} < c_B \le \bar{c} = \frac{2c_{NB}+I+2(1-\mathcal{H})A}{4-\mathcal{H}}$, there is a* ***constrained non-bank dealer equilibrium*** *in which the non-bank dealer sets $S^* = c_B$ and the bank dealer sets $f^* = \frac{1}{2}\left(\frac{I}{\mathcal{H}} + c_B\right)$.*
- *If $c_B > \bar{c}$, there is an* ***unconstrained non-bank dealer equilibrium*** *in which the non-bank dealer sets $S^* = \bar{c}$ and the bank dealer sets $f^* = \frac{c_{NB}\mathcal{H}+2I+(1-\mathcal{H})\mathcal{H}A}{(4-\mathcal{H})\mathcal{H}}$.*

We use the term "unconstrained" in the equilibrium definition to mean negligible competition in market making: if the dealer who operates the market-making business increases S^* slightly, the other dealer remains unable to compete in this business. In contrast, the term "constrained" is used to mean significant competition: the dealer who provides market-making services cannot increase S^* without losing this business to the other dealer.

[13]Our focus in this paper is on a market in which dealers provide both market-making and matchmaking services. The assumption $c_B > I/\mathcal{H}$ implies that matchmaking is a viable business for the bank dealer in equilibrium by setting a fee $f \in (I/\mathcal{H}, S)$.

3.2 What Happens when Bank Regulatory Costs Increase?

We are interested in understanding how an increase in the bank dealer's regulatory costs impacts customer welfare and market outcomes and collect all relevant comparative statics in Proposition 2.

Proposition 2. *As c_B increases through the four equilibrium regions,*

	Unconstrained Bank Dealer	*Constrained Bank Dealer*	*Constrained Non-Bank Dealer*	*Unconstrained Non-Bank Dealer*
Spread	↑	*flat*	↑	*flat*
Matchmaking fee	*flat*	↓	↑	*flat*
Avg transaction costs	*partial hump-shape*	↓	*ambiguous*	*flat*
Volume	*flat*	↑	↓	*flat*
Matchmaking (mkt shr)	↑	↑	↑	*flat*
Market making (mkt shr)	↓	↓	↓	*flat*
Overall customer welfare	↓	↑	↓	*flat*

Furthermore, customer welfare and all market outcomes transition continuously from one equilibrium region to the next.

Figure 1 provides a visual illustration of how customer welfare and market outcomes evolve throughout the four regions for a particular numerical example with $c_{NB} = 30$ basis points.[14] We plot the market-making spread, the matchmaking fee, and average transaction costs in Panel A, customers' trading and their choice of trading mechanisms in Panel B, and overall welfare in Panel C as functions of the bank dealer's balance sheet cost, c_B, holding all other model parameters constant. We distinguish between the four equilibrium regions in the figure using vertical dashed lines.

When the bank dealer's balance sheet costs are low enough, the optimal spread, S_B (depicted by the blue line in Panel A), is lower than the non-bank dealer's balance sheet cost, c_{NB}. In this

[14] For the numerical example, we set $A = 40$ basis points, $I = 0.10$ basis points, and $\mathcal{H} = 0.25$.

region, the bank dealer has an unconstrained monopoly on the provision of immediacy. As the bank dealer's balance sheet cost increases, he passes some of the increase to his market-making customers by increasing the spread. The matchmaking fee (the green line in Panel A) is determined by a basic tradeoff: a higher matchmaking fee increases compensation from each trade while decreasing the number of customers who choose to trade. Given a particular distribution of customers' private value (or patience), this tradeoff results in a unique fee that maximizes the bank dealer's expected profit from matchmaking and depends only on the distribution of private value (and hence does not change as bank regulatory costs increase).

As the spread increases in this region, some customers switch from the market-making mechanism to the matchmaking mechanism (the blue area shrinks and the green area expands in Panel B). The population of customers who refrain from trading (which depends on the magnitude of the matchmaking fee) does not change, however, and hence overall volume is unchanged. All trading customers are worse off in this equilibrium region when c_B increases, either because they pay a higher spread or because they are priced out of the market-making service and incur waiting costs when utilizing the matchmaking service. Furthermore, the population of customers who refrain from trading remains the same (because the matchmaking fee is unchanged), and hence overall customer welfare is lower in this region when bank regulatory costs increase.

As we transition from one equilibrium region to another, it is important to stress that the dealers' pricing strategy changes continuously between regions and, since overall customer welfare and all other market outcomes are continuous in the pricing strategy, there are no abrupt jumps in the transition to the constrained bank dealer equilibrium. As competition from the non-bank dealer in this region constrains the bank dealer's ability to pass the rising regulatory costs on to his market-making customers, the bank dealer seeks to extract higher profits from the matchmaking business. To this end, he increases overall trading volume (which is equivalent in our setting to increasing the fraction of customer types who trade) by reducing the matchmaking fee to attract customers with a low need for immediacy, which can be observed as the shrinking "no trade" (purple) area in Panel B.

Our key result is that overall customer welfare increases in this region of Panel C. The welfare gains come from three groups of customers. The first group refrains from trading when the matchmaking fee is high, but when c_B increases and the bank dealer lowers the matchmaking fee they begin trading and thus contribute to overall customer welfare. The second group consists of customers who trade via the matchmaking mechanism either way, but when c_B increases their welfare goes up because they pay a lower fee. The third group comprises customers who find it optimal to switch from market making to matchmaking when c_B increases because their expected utility considering both the lower fee and the expected waiting costs in the matchmaking service is higher.[15] Hence, all customers are either better off or no worse off as c_B rises, which means that equilibria with higher regulatory costs Pareto-dominate those with lower regulatory costs in this region.[16]

When c_B increases further such that it exceeds the non-bank dealer's balance sheet cost, c_{NB}, we transition into the third equilibrium region. In this region, the bank dealer does not provide market-making services.[17] This is arguably a less relevant region from an empirical standpoint given that Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018) estimate that bank dealers handle about 87% of principal trading even after the post-crisis regulatory reform. Still, the same paper also presents evidence that non-bank dealers increased their market share in principal trading from about 3% in the pre-crisis period to about 13% in the post-regulatory-reform period. Our model is simplified in that it has a bang-bang solution: either the bank dealer or the non-bank dealer captures all market-making clients.[18] The increase in principal trading by non-bank dealers

[15]The increase in c_B in this region of the equilibrium does not make customers utilizing the market-making mechanism worse off because the spread is constrained to equal the balance sheet cost of the non-bank dealer. Hence, all customers who switch from market making to matchmaking do so because the lower matchmaking fee makes them better off.

[16]To simplify the model, the matching rate of customers in the matchmaking mechanism is directly determined by the bank dealer's search technology. We believe that the increase in overall customer welfare when bank regulatory costs go up in the constrained bank dealer equilibrium would hold in a more general search specification (e.g., Duffie et al (2005)) in which the matching intensity depends on the mass of customers who choose matchmaking. Specifically, Proposition 2 shows that the market share of matchmaking monotonically increases in c_B. Under an alternative structure in which the size of the pool of searching customers impacts the matching rate, this increase in market share would result in a higher matchmaking speed $\mathcal{H}$ and make the matchmaking service even more beneficial to customers. In this case, the switch to matchmaking would likely be more pronounced as c_B goes up, further improving overall customer welfare.

[17]When $c_B = c_{NB}$, the only equilibrium spread possible is $S = c_B = c_{NB}$ and both bank and non-bank dealers make zero profits from market making. We include this boundary point in the second region, but it can be added to the third region ($c_B > c_{NB}$), or we could assume that customers who would like to use the market-making mechanism randomize between the bank and non-bank dealers.

[18]In Section 4 we investigate an extension of our model utilizing Cournot (rather than Bertrand) competition in

may suggest that we are getting closer to the point where $c_B = c_{NB}$. It is therefore important to investigate what would happen to overall customer welfare if bank regulatory costs were to increase beyond this point.

In the third region, it is the bank dealer's balance sheet costs that constrain the market-making strategy of the non-bank dealer, forcing her to set her spread equal to c_B. As c_B rises, the competitive pressure from the bank dealer eases, and the non-bank dealer can increase her market-making spread to extract more rents (the increasing red line in Panel A). This spread increase creates an opportunity for the bank dealer to increase his matchmaking fee as well.[19] As a result, more customers forgo trading and hence trading volume falls (observed as an increase in the purple area in Panel B). Most importantly, overall customer welfare in Panel C declines. The decline in welfare stems from the increase in the market-making spread and matchmaking fee for those customers who trade as well as the increase in the number of customers who refrain from trading because of the higher costs. As c_B rises beyond a certain point, the bank dealer no longer exerts any competitive pressure and we reach an unconstrained non-bank dealer equilibrium (the fourth region in the figure). In this region, which is somewhat extreme and hence of limited interest, further increases in the bank dealer's balance sheet cost no longer affect the equilibrium outcomes.

Our analysis provides several key results. First, the increase in regulatory costs causes customers to shift from market making to matchmaking. While this result is not necessarily surprising, our analysis reveals that this shift arises for different reasons in different regions. In the first and third regions, the higher market-making spread makes this service too expensive for some customers and they switch to matchmaking. In the second region, the reason is different: the bank dealer reduces the matchmaking fee to attract customers to the more profitable matchmaking business from the less profitable market-making business. Whether this shift, which has also been documented empirically, is driven by a "stick" (a higher spread) or a "carrot" (a lower matchmaking fee) turns out to be the key to understanding why overall customer welfare can either increase or decrease in the various

which both bank and non-bank dealers simultaneously provide market-making services in equilibrium.

[19] Proposition 2 shows that the market share of the matchmaking service increases in the constrained non-bank dealer equilibrium. This is the only result that is not general but rather depends on having a uniform distribution for customers' private value. For a general G, the change in the market share of the matchmaking service is ambiguous in this region.

equilibrium regions as regulatory costs rise.

Our second result is that customer welfare declines in the first region (the unconstrained bank dealer equilibrium), which fits with views expressed by some market participants according to which a higher bank dealer balance sheet cost would negatively impact market liquidity and customer welfare. Customers are worse off in our model, however, because the bank dealer prices the matchmaking service too high. While the high fee deters some customers from trading, it enhances market-making profitability because more customers choose to pay an even higher spread to avoid waiting for execution. Hence, the bank dealer's pricing strategy implies that the "internal" competition between market making and matchmaking is not enough to make customers better off. What is needed for an increase in customer welfare is interaction with another form of competition: the external competition between the bank and non-bank dealers.

Our third result, and the main insight our model generates, is that, when this external competition kicks in, an increase in regulatory costs increases overall customer welfare. The driving force behind the welfare improvement is that an increase in regulatory costs incentivizes the bank dealer to reduce the matchmaking fee, thereby both attracting new customers who were previously priced out by the high matchmaking fee and increasing utility for all existing customers who already chose matchmaking or switched to matchmaking from market making. Interaction between the "internal" competition and the "external" competition is critical in this region in delivering unambiguously higher customer welfare. A key insight that arises from our model is that counteracting the too-big-to-fail subsidy by increasing bank regulatory costs may not be welfare-improving without market discipline. Namely, competition from non-bank dealers who stand ready to offer market-making services is crucial to attaining the welfare-improvement result.[20]

Our fourth result is that the change in average transaction costs is not a sufficient statistic for the change in overall customer welfare. Average transaction costs are a weighted average of the market-making spread and the matchmaking fee, using the respective populations of the trading customers

[20]We emphasize that customer welfare increases in our model because the bank dealer wields market power and extracts rents from customers. In the Internet Appendix, we discuss a variation of the model in which a benevolent bank dealer maximizes customer welfare subject to breaking even on his liquidity-provision service. In that setting, an increase in the bank dealer's balance sheet cost always reduces overall customer welfare. In Section 4, we examine the robustness of our results to having imperfect competition from multiple bank dealers.

in each of these mechanisms as weights. When regulatory costs increase, the spread weakly increases but the fraction of customers who choose the market-making mechanism declines and hence the direction of the change in average transaction costs can go either way. When regulatory costs are low, most customers use the market-making mechanism. Increasing these costs causes some customers to switch to matchmaking, but the increase in the spread for a larger population of customers dominates the lower matchmaking fee for a smaller population of customers, resulting in an increase in average transaction costs. When the regulatory costs are high enough, however, the shift from market making to matchmaking can dominate and hence average transaction costs can start to decline at the same time that welfare worsens in the unconstrained bank dealer equilibrium (the first region). This divergence demonstrates the perils of thinking about customer welfare only in terms of the average cost of transacting.

Our parsimonious model delivers these key insights in a transparent environment that enables the intuition behind them to come through. A natural question is whether our specific choices drive the results or whether these insights are more general. In the next section we provide extensive robustness analyses to answer this question.

4 Robustness

Our analysis in Section 3 demonstrates that customer welfare could increase when bank regulatory costs rise. The basic model enables us to show what key elements are required to obtain this result: bank dealer market power, "internal" competition between the two businesses (or market structures) that the bank dealer operates, and "external" competition in market making from the non-bank dealer. Still, it is of interest to ask whether there are relevant features of the market that, when incorporated in the model, would counteract or vacate our main result. Therefore, in this section we examine the robustness of our main result by considering three extensions of the model. In each extension, we add a feature or change the model's structure to examine whether our main result remains intact.

The first extension allows the non-bank dealer to offer matchmaking services as well. As such, the non-bank dealer can exert competitive pressure on both businesses at the same time. The second extension considers Cournot competition between the bank and non-bank dealers to examine whether our results can be attributed to the particular nature of Bertrand competition. This extension also allows to us to study an equilibrium in which the market-making services of both the bank and non-bank dealers operate simultaneously. The third extension investigates in greater depth the role of market power. We introduce multiple bank and non-bank dealers and examine how the number of competing dealers impacts our main welfare result. As this welfare result arises in the (more empirically relevant) parameter region $c_B \leq c_{NB}$, we focus the three extensions on this parameter region.

4.1 Non-Bank Dealer Matchmaking Service

One objection to our parsimonious model could be that non-bank dealers in real-world bond markets do engage in matchmaking. Empirically, the quantity of their matchmaking was found to decrease following the post-crisis regulatory reform even as matchmaking by bank dealers has increased, which is why we did not focus on this aspect of the market in our main model.[21] Still, would the bank dealer in the model reduce his matchmaking fee if he were facing competition in the matchmaking service? Is the absence of any competition in matchmaking necessary for our result that welfare can improve when regulatory costs increase?

In this section, the only change we make to the main model is to let the non-bank dealer operate both market-making and matchmaking services. Specifically, by spending I on the search process for each customer, the bank dealer matches a customer with intensity H_B while the non-bank dealer matches a customer with intensity H_{NB}.[22] Given the substantial evidence that bank dealers have much larger customer networks, we assume that $H_B > H_{NB}$. In other words, by spending the same amount of money, the larger customer base of the bank dealer enables him to find a counterpart to his customer's order more quickly on average than the non-bank dealer can.[23]

[21] See Bao, O'Hara, and Zhou (2018) and Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018) for evidence on matchmaking by both bank and non-bank dealers.

[22] The corresponding speeds of matchmaking services are therefore $\mathcal{H}_B = \frac{H_B}{H_B+r}$ and $\mathcal{H}_{NB} = \frac{H_{NB}}{H_{NB}+r}$ for the bank and non-bank dealers, respectively.

[23] We assume that each customer trades through one matchmaking service. In other words, a customer cannot split an order between the matchmaking services offered by the bank and non-bank dealers and cannot employ both dealers to search simultaneously.

Assuming $H_B > H_{NB}$ does not imply that the non-bank dealer is ineffective in constraining the bank dealer's pricing strategy. In fact, to examine the robustness of our main result, we are specifically interested in equilibria in which both the bank dealer and the non-bank dealer offer matchmaking services. We therefore focus on these equilibria, which reflect the empirical findings that both types of dealers are engaged in matchmaking, to provide insights into what happens to overall customer welfare as c_B increases.

Denote the matchmaking fees of the bank and non-bank dealers by f_B and f_{NB}, respectively. To have a positive amount of matchmaking services provided by the non-bank dealer, we must have in equilibrium $f_B > f_{NB}$ (because customers of the non-bank dealer have to wait longer on average for execution). To have a positive amount of matchmaking services provided by the bank dealer, we must have in equilibrium $S_B > f_B$ because otherwise customers will switch to the instantaneous provision of liquidity in the market-making service. Therefore, given the three prices (S_B, f_B, f_{NB}), customers in these equilibria are sorted into four groups, as stated in the following lemma.

Lemma 1. *Define*

$$x_1 \equiv \frac{\mathcal{H}_B f_B - \mathcal{H}_{NB} f_{NB}}{\mathcal{H}_B - \mathcal{H}_{NB}}, \quad x_2 \equiv \frac{S_B - \mathcal{H}_B f_B}{1 - \mathcal{H}_B}. \tag{9}$$

If the bank dealer operates the market-making service and both dealers operate matchmaking services in equilibrium, customers with private value $x < f_{NB}$ forgo trading, customers with $x \in [f_{NB}, x_1)$ use the non-bank dealer's matchmaking service, customers with $x \in [x_1, x_2)$ use the bank dealer's matchmaking service, and customers with $x \geq x_2$ use the bank dealer's market-making service.

The cutoff value x_2 is entirely analogous to b in the main model. The cutoff value x_1 between the matchmaking services of the bank and non-bank dealers is derived from the indifference condition $\mathcal{H}_B(x_1 - f_B) = \mathcal{H}_{NB}(x_1 - f_{NB})$.

As is the case with the main model, there are two types of equilibria according to whether the non-bank dealer's market-making service constrains the bank dealer's strategy. The unconstrained market-making equilibrium is obtained by solving the three linear best response functions for f_{NB}^{br}, f_B^{br}, and S^{br}, whereas the constrained market-making equilibrium is obtained by solving the two best response functions for f_B^{br} and f_{NB}^{br} with $S = c_{NB}$. In either case, to ensure that we have equilibria in which all three liquidity-provision services are offered we need to impose the conditions $\frac{I}{\mathcal{H}_{NB}} < f_{NB} < f_B < S \leq c_{NB}$ and $x_2 < A$, which translate into the condition on A in equation (10).[24] The equilibrium is characterized in the following proposition.

Proposition 3. *Suppose that*

$$\max\left(\frac{c_{NB} - \frac{\mathcal{H}_B}{\mathcal{H}_{NB}}}{1-\mathcal{H}_B}, \frac{2I}{\mathcal{H}_{NB}}\right) < A < \frac{c_{NB}(4\mathcal{H}_B - \mathcal{H}_{NB}) - (1+2\mathcal{H}_B)I}{2\mathcal{H}_B - \mathcal{H}_{NB} - \mathcal{H}_B\mathcal{H}_{NB}}. \tag{10}$$

There exist constants c_1, c_2, and c_3 such that:

- *If $c_B \in (c_1, c_2]$, there is an **unconstrained market-making equilibrium** in which the bank dealer operates the market-making service and both dealers operate matchmaking services, with prices given by*

$$f^*_{NB} = \frac{I(2\mathcal{H}_B + \mathcal{H}_{NB}) + A\mathcal{H}_{NB}(\mathcal{H}_B - \mathcal{H}_{NB})}{4\mathcal{H}_{NB}(H_B - H_{NB})}, \tag{11}$$

$$f^*_B = \frac{3I + 2A(\mathcal{H}_B - \mathcal{H}_{NB})}{4\mathcal{H}_B - \mathcal{H}_{NB}}, \tag{12}$$

$$S^* = \frac{c_B}{2} + \frac{I\left(2\mathcal{H}_B + \mathcal{H}_{NB}\right) + A\left(\mathcal{H}_B - \mathcal{H}_{NB} + 3\mathcal{H}_B\left(1 - \mathcal{H}_{NB}\right)\right)}{8\mathcal{H}_B - 2\mathcal{H}_{NB}}. \tag{13}$$

- *If $c_B \in (c_2, c_3)$, there is a **constrained market-making equilibrium** in which the bank dealer operates the market-making service and both dealers operate matchmaking services, with prices given by*

$$f^*_{NB} = \frac{1}{2}\left(\frac{(4c_{NB} - 2c_B)(\mathcal{H}_B - \mathcal{H}_{NB}) + (3 - \mathcal{H}_B - 2\mathcal{H}_{NB})I}{\mathcal{H}_B - \mathcal{H}_{NB} + 3\mathcal{H}_B(1-\mathcal{H}_{NB})} + \frac{I}{\mathcal{H}_{NB}}\right), \tag{14}$$

$$f^*_B = \frac{(4c_{NB} - 2c_B)(\mathcal{H}_B - \mathcal{H}_{NB}) + (3 - \mathcal{H}_B - 2\mathcal{H}_{NB})I}{\mathcal{H}_B - \mathcal{H}_{NB} + 3\mathcal{H}_B(1-\mathcal{H}_{NB})}, \tag{15}$$

$$S^* = c_{NB}. \tag{16}$$

Before we discuss how customer welfare and market outcomes change when bank regulatory

[24] The condition that the matchmaking fees are greater than $\frac{I}{\mathcal{H}_{NB}}$ is needed to ensure that the search technology is not too costly to prevent matchmaking, while the condition $x_2 < A$ is needed to ensure that the support of G, $[0, A]$, is large enough to create sufficient "space" to fit all customer choices.

costs increase, we should comment on two aspects of these equilibria. First, since $\mathcal{H}_{NB} < \mathcal{H}_B$, the non-bank dealer and the bank dealer are providing two distinct matchmaking services: they have vertically differentiated products. As such, they charge different prices and cater to distinct segments of the population of customers, and both of them earn positive profits even under Bertrand price competition. Second, the competition introduced into the provision of matchmaking services by the non-bank dealer is binding and impacts the bank dealer's pricing strategy in all equilibria covered by Proposition 3.[25] A key question in this extension of the model is whether competition in matchmaking services prevents the bank dealer from reducing his matchmaking fee, thereby hindering the improvement in customer welfare. In the following proposition we show that this is not the case: overall customer welfare increases when the bank dealer is constrained by competition in both services.

Proposition 4. *As c_B increases from c_1 to c_3 in the equilibria described in Proposition 3:*

	Unconstrained Mkt Making	*Constrained Mkt Making*
Spread	↑	*flat*
Matchmaking fee f_B^*	*flat*	↓
Matchmaking fee f_{NB}^*	*flat*	↓
Avg transaction costs	↓ *iff* $c_B > \frac{(3+\mathcal{H}_B-\mathcal{H}_{NB})I+2(1-\mathcal{H}_B)(\mathcal{H}_B-\mathcal{H}_{NB})A}{4\mathcal{H}_B-\mathcal{H}_{NB}}$	↓ *iff* $c_B > \hat{c}$ *(expression in proof)*
Volume	*flat*	↑
Matchmaking (total mkt shr)	↑	↑
Bank dealer	↑	↑
Non-bank dealer	*flat*	↓
Market making (mkt shr)	↓	↓
Overall customer welfare	↓	↑

[25] The labels "constrained" and "unconstrained" that we attach to these equilibrium regions apply only with respect to competition in the market-making service. The matchmaking strategy of the bank dealer in this extension of the model is always constrained

The impact of an increase in c_B in the constrained equilibrium works in this extension as it does in the main model. Figure 2 extends the numerical example in Figure 1 to the case with non-bank dealer matchmaking.[26] An increase in regulatory costs incentivizes the bank dealer to shift business activities towards matchmaking by reducing his fee, as we observe in Panel A of the figure. In fact, he reduces his fee much more aggressively than does the non-bank dealer: the derivative of the matchmaking fee of the bank dealer with respect to c_B is negative and exactly twice the magnitude of the derivative of the non-bank dealer's matchmaking fee, and this difference in aggressiveness is clearly evident in the figure. The reduction in fees in this region of the equilibrium attracts new customers to the market (hence shrinking the "no trade" purple area in Panel B of the figure) and incentivizes customers to switch from market making to matchmaking. These effects are similar to those we obtain in the main model.

We use the same parameter values in Figure 2 as in Figure 1, so it is clear from comparing the figures that customers are unequivocally better off when the non-bank dealer competes in matchmaking. Even if the non-bank dealer's matchmaking service is slower than that operated by the bank dealer ($\mathcal{H}_{NB} < \mathcal{H}_B$), competition from the non-bank dealer causes the bank dealer to set a lower matchmaking fee than in the main model.[27] Because the bank dealer is also more aggressive than the non-bank dealer in reducing the matchmaking fee when regulatory costs increase, the economic mechanism behind our result that overall customer welfare can increase with the imposition of higher regulatory costs appears very robust to introducing this additional dimension of competition.

It is interesting to note that average transaction costs in the constrained equilibrium in Proposition 4 do not decrease unambiguously (as in our main model) but rather only if c_B is above a certain threshold. This is because an increase in c_B causes customers to shift from the cheaper of the two matchmaking services (operated by the non-bank dealer) to the more expensive service (operated by the bank dealer). These customers shift because they are better off, but the end result is that average transaction costs can increase unless the bank dealer's fee is sufficiently low (when the regulatory costs are above the threshold). The divergence in implications between overall customer

[26] We need to add only one parameter, $\mathcal{H}_{NB} = 0.10$.

[27] We provide a proof of this result in the Internet Appendix.

welfare and average transaction costs in this extension further reinforces the earlier point that one cannot simply look at average transaction costs to judge whether investors are better off.

The extension also adds another comparative static: the market share of the bank dealer's matchmaking service increases while the market share of the non-bank dealer's matchmaking service declines (shown visually as an expanding green area and a shrinking red area in Panel B). This is consistent with the empirical findings reported in Bao, O'Hara, and Zhou (2018) and Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018) that bank dealers increased and non-bank dealers decreased their matchmaking activity following the post-crisis regulatory reform. Most importantly, the changes in pricing and customers' optimal shifting between liquidity-provision services unequivocally improve overall customer welfare as c_B rises in this region. Hence, our key result from the main model remains intact when the non-bank dealer is allowed to operate a matchmaking service because in both cases the result is driven by the heterogenous effects of bank regulatory costs on the market-making operations of bank and non-bank dealers.

4.2 Cournot Competition

Our main model utilizes Bertrand competition to highlight the competition on prices in the corporate bond market. Using Bertrand competition means that only the dealer with the lower balance sheet cost operates a market-making service, because he or she can always undercut the other dealer. In this section we examine the robustness of our welfare result to using an alternative equilibrium concept: Cournot quantity competition. As a side benefit, Cournot competition allows both bank and non-bank dealers to simultaneously operate market-making services in equilibrium, generating an additional comparative static about their respective market shares. The only change we make from the main model is the form of the equilibrium; all other elements remain as in Section 3.

The dealers' choice variables are the quantities of market making by the bank and non-bank dealers (q_B and q_{NB}, respectively) and the quantity of bank dealer matchmaking (q_M). Note that because both dealers offer identical market-making services, they are perfect substitutes. Hence, there will be a single equilibrium market-making spread, which we denote by S as in our main model.

As is standard in the literature, customers' demand curves depend on these prices. Conditional on S and the matchmaking fee f, total customer demand for trading is $1 - G(f)$, and customer demand for trading immediately via the market-making service is $1 - G(b)$, where $b = \frac{S-\mathcal{H}f}{1-\mathcal{H}}$ as in the main model. This implies that

$$q_B + q_{NB} + q_M = 1 - G(f) = 1 - \frac{f}{A}, \tag{17}$$

$$q_B + q_{NB} = 1 - G(b) = 1 - \frac{1}{A}\frac{S - \mathcal{H}f}{1 - \mathcal{H}}. \tag{18}$$

The right-hand sides of these two equations represent the customers' downward-sloping demand curves, and equilibrium prices can therefore be expressed in terms of quantities as follows:

$$f = A(1 - q_B - q_{NB} - q_M), \tag{19}$$

$$S = A(1 - q_B - q_{NB} - \mathcal{H}q_M). \tag{20}$$

With this alternative form of competition in market making, the bank dealer and the non-bank dealer choose the quantities that maximize their profits. Specifically, the bank dealer's optimization problem is

$$\begin{aligned}\max_{q_B, q_M} \Pi_B &= \frac{2\mu}{r}[(\mathcal{H}f - I)q_M + (S - c_B)q_B] \\ &= \frac{2\mu}{r}\left[(\mathcal{H}A(1 - q_B - q_{NB} - q_M) - I)q_M + (A(1 - q_B - q_{NB} - \mathcal{H}q_M) - c_B)q_B\right].\end{aligned} \tag{21}$$

while the non-bank dealer solves the following problem:

$$\begin{aligned}\max_{q_{NB}} \Pi_{NB} &= \frac{2\mu}{r}(S - c_{NB})q_{NB} \\ &= \frac{2\mu}{r}\left[(A(1 - q_B - q_{NB} - \mathcal{H}q_M) - c_{NB})q_{NB}\right].\end{aligned} \tag{22}$$

We focus on equilibria in which there is a positive amount of matchmaking and both dealers operate market-making services (an interior solution). The first-order conditions of both problems are linear and admit a unique solution. We need only to check that the three equilibrium quantities are positive and add up to no more than one. It can be shown that the following conditions must hold for an interior solution:

$$q_B > 0 \Longleftrightarrow c_B \leq \frac{2c_{NB}(1-\mathcal{H}) + 3I + 2(1-\mathcal{H})A}{4-\mathcal{H}},$$

$$q_{NB} > 0 \Longleftrightarrow c_B > 2c_{NB} - A,$$

$$q_M > 0 \Longleftrightarrow c_B > \frac{I}{\mathcal{H}}.$$

The first condition ensures that the regulatory cost of the bank dealer is not so high as to render his market-making service inferior to the matchmaking service for all customer types. The second condition ensures that the regulatory cost is not so low as to prevent the non-bank dealer from competing effectively in market-making services. The last condition, which is the same condition we impose in the main model, implies that a nonzero amount of matchmaking service is offered in equilibrium. The following proposition establishes the existence of the equilibrium in this version of the model.

Proposition 5. *If*

$$\max\left(\frac{I}{\mathcal{H}}, 2c_{NB} - A\right) < c_B < \frac{2(1-\mathcal{H})(c_{NB}+A) + 3I}{4-\mathcal{H}}, \tag{23}$$

there exists a unique equilibrium in which both dealers provide market-making services and the bank dealer operates the matchmaking service. The equilibrium prices are

$$S^* = \frac{1}{3}(c_B + c_{NB} + A), \quad f^* = \frac{1}{3}(c_B + c_{NB} + A) - \frac{c_B\mathcal{H} - I}{2\mathcal{H}}. \tag{24}$$

The manner in which customer welfare and market outcomes are affected when bank regulatory costs rise is summarized in the following proposition.

Proposition 6. *As c_B increases in the equilibria described in Proposition 5:*

Spread	$\uparrow$
Matchmaking fee	$\downarrow$
Average transaction costs	$\downarrow$ *iff* $c_B > \hat{c}$ *(expression in proof)*
Volume	$\uparrow$
Matchmaking (market share)	$\uparrow$
Market making (total market share)	$\downarrow$
Bank dealer	$\downarrow$
Non-bank dealer	$\uparrow$
Overall customer welfare	$\uparrow$ *iff* $c_B > \frac{9I+4(1-\mathcal{H})(2A-c_{NB})}{4+5\mathcal{H}}$

This proposition shows how the "internal" competition (between the bank dealer's matchmaking and market-making services) and the "external" competition (in market-making services between the bank and non-bank dealers) interact in the Cournot extension to generate a blend of the implications we observed in the unconstrained and constrained bank-dealer equilibria of our main model. The spread increases and the matchmaking fee decreases as regulatory costs rise, causing matchmaking to go up, market making to go down, and total volume to increase. We observe a new empirical implication in this version of the model: market making by the bank dealer falls at the same time that market making by the non-bank dealer increases. This is quite intuitive given that the increase in cost applies only to the bank dealer, and it matches well with the stylized facts that were documented empirically (e.g., Bao, O'Hara, and Zhou (2018) and Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018)).

Most importantly, overall customer welfare can increase with the rise in regulatory costs and the result has the same flavor as in our main model. Specifically, when the bank dealer's cost is very low in the main model, increasing it initially reduces welfare (the unconstrained bank dealer equilibrium), while beyond a certain threshold customer welfare starts to increase as we move to the constrained bank dealer equilibrium. The Cournot equilibrium has a single equilibrium region with respect to the pricing of liquidity-provision services, but the essence of the result remains exactly the same: for low c_B welfare declines as we move to the right, but once we pass the threshold expression in Proposition 6, customers are made better off when bank regulatory costs rise. To summarize, the insights of our model regarding how an increase in bank regulatory costs impacts customer welfare are the same irrespective of whether we use competition in prices (Bertrand) or quantities (Cournot), which enhances our confidence in the generality of these results.

4.3 Multiple Bank and Non-Bank Dealers

The market failure that lies at the core of our result according to which increasing bank regulatory costs can make customers better off is the market power that the bank dealer wields. In fact, we show in the Internet Appendix that if the bank dealer were benevolent and set the spread and matchmaking fee to maximize customer welfare, increasing c_B always makes customers worse off. A natural question is what happens when multiple bank dealers and multiple non-bank dealers are available to provide liquidity. Would having more than one representative dealer eliminate our main result or would we observe that the result weakens gradually as more and more dealers provide liquidity and each dealer's market power diminishes? If the latter is the case, then our result is highly relevant to the corporate bond market in which 10 to 12 bank dealers have a 70% market share (Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018)). The Cournot model of quantity competition in the previous section provides a convenient framework within which to examine this question.

We assume for this extension that there are J bank dealers (indexed by j) with the same balance sheet cost λc_B ($\lambda > 0$), matchmaking cost I, and speed $\mathcal{H}$. Similarly, there are K non-bank dealers (indexed by k) with the same balance sheet cost βc_{NB} ($\beta > 0$). Each bank dealer j chooses matchmaking quantity q_M^j and market-making quantity q_B^j. Each non-bank dealer k chooses market-making quantity q_{NB}^k. Let

$$Q_M = \sum_j^J q_M^j, \quad Q_B = \sum_j^J q_B^j, \quad Q_{NB} = \sum_k^K q_{NB}^k. \tag{25}$$

Following the exact same logic as in Section 4.2, the equilibrium market-making spread and matchmaking fee satisfy

$$S = A(1 - Q_{NB} - Q_B - \mathcal{H}Q_M), \tag{26}$$

$$f = A(1 - Q_{NB} - Q_B - Q_M). \tag{27}$$

When maximizing profits, each dealer recognizes the impact of his quantity choices on the price. The first-order conditions are linear in quantities and have unique solutions. Our goal is to extend the equilibrium in Proposition 5 to examine how the number of competing bank and non-bank dealers impacts our results, and therefore we focus on the symmetric equilibrium in which the quantities offered by all dealers are positive.

Proposition 7. *There exists a unique equilibrium in which all bank and non-bank dealers provide market-making services and all bank dealers provide matchmaking services. The equilibrium prices are*

$$S^* = \frac{\Lambda c_B + B c_{NB} + A}{1 + J + K}, \tag{28}$$

$$f^* = \frac{-\Lambda c_B \mathcal{H} K + IJ(1 + J + K) + B c_{NB} \mathcal{H}(1 + J) + \mathcal{H}A(1 + J)}{\mathcal{H}(1 + J)(1 + J + K)}, \tag{29}$$

where $\Lambda = J\lambda$ *and* $B = K\beta$.

It turns out that the basic properties of the equilibrium do not change when there are multiple bank and non-bank dealers instead of a representative dealer from each group.

Proposition 8. *As* c_B *increases in the equilibrium described in Proposition 7, the comparative statics for prices and market share are similar to those in Proposition 6 with the only difference being that the cutoffs for customer welfare and average transaction costs depend on the number of bank and non-bank dealers. In particular, overall customer welfare increases iff*

$$c_B > \frac{I \cdot J\,(1 + J + K)^2 + (1 - \mathcal{H})\,(1 + J)^2\,((J + K)\,A - B \cdot c_{NB})}{\Lambda \left((1 + J)^2 + \mathcal{H}K\,(2 + 2J + K)\right)}. \tag{30}$$

The market power of bank dealers in our model distorts the relative pricing of the market-making and matchmaking services, thereby creating a role to regulatory intervention that can make customers better off. As such, we would expect that, as the number of bank dealers increases and hence their market power diminishes, there will be less room for welfare improvement. Panel A of Figure 3 plots an example to illustrate how the number of bank dealers impacts our main result under the assumption that $\lambda = \beta = 1$ (as in Section 4.2), which means that $\Lambda = J$ and $B = K$. Specifically, we examine how the cutoff in equation (30), above which an increase in bank regulatory costs improves welfare, changes with the number of dealers. The y-axis in the figure is this cutoff, while the x-axis is the number of bank dealers. Each line in the plot represents an economy with a different number of non-bank dealers. Indeed, we observe that, as the number of bank dealers increases, the cutoff increases gradually. Most importantly from our perspective, the range of c_B over which our main result holds does not vanish when we abandon the representative dealer case that we use to simplify the exposition of our main model; rather, it shrinks commensurately with the decline in the friction that generates it.

Panel B of the figure illustrates how the cutoff changes as the number of non-bank dealers rises, and we observe that the cutoff declines. Why is that the case? Non-bank dealers in our model provide competition in the provision of market-making services, thereby depressing bank dealers' profitability and incentivizing them to shift more business to the matchmaking service. As such, a higher number of non-bank dealers enhances the effectiveness of the increase in regulatory costs in bringing down the profitability of the bank dealers' market-making business. To gain a better understanding of the role they play in pricing, we examine how the number of non-bank dealers affects the rate at which market prices (the spread and the matchmaking fee) respond to an increase in bank regulatory costs. We can show the following unambiguous results:

$$\frac{\partial^2 S^\star}{\partial c_B \partial K} < 0, \quad \frac{\partial^2 f^\star}{\partial c_B \partial K} < 0 \tag{31}$$

When the number of non-bank dealers (K) is higher, there are more providers of immediacy competing in the market-making business that are not subject to the increase in c_B. This causes $S^\star$ to increase at a lower rate when c_B goes up. As a result, the bank dealers' market-making profitability is squeezed even more, causing them to reduce the matchmaking fee at a higher rate to attract more customers to the matchmaking service. Thus, a smaller increase in bank regulatory costs is required to reach the range at which overall customer welfare improves, which is why in Panel B of Figure 4 we observe the cutoff declines when the number of non-bank dealers rises.

Dealer Heterogeneity. A key driver behind our main result is the heterogenous impact of bank regulatory costs on the balance sheets of bank and non-bank dealers. One could argue that bank regulatory costs could potentially have heterogeneous impacts even among bank dealers (e.g., differing impacts on systemically important financial institutions and others). An interesting question is whether such heterogeneity could weaken the competitive pressure that non-bank dealers exert and how would this, in turn, impact our main result regarding customer welfare.

The extension in this section can be used to shed some light on this question. We modify our specification of the balance sheet costs of dealers slightly by assuming that bank dealer j's balance sheet cost is $\lambda_j c_B$ and non-bank dealer k's balance sheet cost is $\beta_k c_{NB}$, where $\{\lambda_j\}$ and $\{\beta_k\}$ are positive constants. Thus, changes in c_B affect bank dealers heterogeneously. It turns out that the equilibrium in Proposition 7 and the comparative statics in Proposition 8 also exist in this amended setup, and the only difference from the expressions in the proposition is that $\Lambda = \sum_j \lambda_j$ and $B = \sum_k \beta_k$. In other words, when all dealers provide market-making services and all bank dealers operate matchmaking services, the dispersion in λ_j or β_k does not matter for our main result.

Still, the equilibrium we investigate in Proposition 7 is one in which all bank dealers operate both services. One could imagine a situation in which the increase in bank regulatory costs is so large that it renders the high-cost bank dealers unable to offer market-making services. Would it still be the case that customer welfare can improve when bank regulatory costs rise in such an economy? We believe this is indeed the case, and to demonstrate this result we solve in the Internet Appendix a model with one non-bank dealer and two heterogeneous bank dealers: a high-cost bank dealer with $\lambda_1 = 1 + \delta$ and a low-cost bank dealer with $\lambda_2 = 1 - \delta$.

We show that, as c_B increases, there are three cutoffs that are relevant for customer welfare. Cutoff 1 is similar to that in Proposition 8 with $J = 2$; when c_B is above Cutoff 1, overall customer welfare increases as c_B rises. There is, however, another cutoff, Cutoff 2, such that if

c_B increases above it we move to a new equilibrium in which the high-cost bank dealer drops out of providing market-making services (though he still operates the matchmaking business). In this new equilibrium, customer welfare initially declines as c_B rises, but as c_B continues to rise we reach Cutoff 3, above which customer welfare improves again with further increases in c_B. Importantly, we show that equilibrium outcomes are continuous at these cutoff points even when the high-cost dealer drops out of market making. While Cutoff 2 can be above or below Cutoff 1 (depending on the parameters), it is always the case that, when c_B is sufficiently high (above Cutoff 3), increasing bank regulatory costs improves overall customer welfare.

All the extensions discussed in Section 4, therefore, point to the same conclusion. Namely, our main result and the economic rationale behind it are very robust to the exact specification of the model.

5 Our Theory through the Lens of the Empirical Literature

In this section, we discuss the implications of our model for observable market outcomes and how they map onto the findings of the empirical literature about changes in the corporate bond market following the implementation of post-crisis financial regulations. At the outset it is important to note that empirical work on the corporate bond market in the United States has largely used various forms of the TRACE database. While our model emphasizes the distinction between market making and matchmaking, TRACE does not allow a clean identification of the trading mechanism that facilitated the execution of any given trade. In particular, an in-house cross—a dealer buying from a customer and immediately selling to another customer—is reported in TRACE as two transactions. Whether TRACE reports these two transactions as agency or principal depends on the internal accounts of the dealer involved, not the economics of the transaction. The use of agency or proprietary accounts appears to be idiosyncratic to specific dealers and can be influenced by a dealer's preference over reporting the price inclusive of the mark-up/mark-down or a separate commission.[28] Hence, empirical studies must use various algorithms to (imperfectly) infer the

[28] A FINRA rule implemented in May 2018 further changed incentives by requiring dealers to report mark-ups or mark-downs from the prevailing market price for all trades involving retail customers that are offset within a day.

trading mechanism that executed each trade.

Notwithstanding these empirical limitations, examining the trading mechanisms used by customers is of principal interest. Our model predicts a robust shift from market making to matchmaking as bank regulatory costs rise. Indeed, empirical papers provide evidence of an increase in matchmaking (e.g., Bao, O'Hara, and Zhou (2018) Choi and Huh (2017), Schultz (2017)) and a reduction in capital commitment to market making (Bao, O'Hara, and Zhou (2018), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018)) following the crisis and the implementation of post-crisis regulations. The extensions we investigate in Section 4 generate further predictions regarding how the market shares of bank and non-bank dealers in the same trading mechanism would change. In particular, when in Section 4.2 we have both dealers provide market-making services, an increase in bank regulatory costs causes the bank dealer to cut back on his market-making operations and the non-bank dealer to increase her activity (with total market making still declining). This is exactly the picture documented empirically with regard to capital commitment to market making by Bao, O'Hara, and Zhou (2018) and Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018). When in Section 4.1 we have both dealers provide matchmaking services, the bank dealer increases his market share of matchmaking activity while the non-bank dealer decreases hers as bank regulatory costs rise. Indeed, Bao, O'Hara, and Zhou (2018) report that bank dealers increased the share of their volume through the matchmaking mechanism while the opposite was observed for the non-bank dealers. Hence, our model's predictions regarding customer choices with respect to the trading modality match the empirical evidence well.

Our model generates important empirical implications concerning the pricing of liquidity services: the spread (or the cost of immediate execution), the matchmaking fee, and average transaction costs. There are robust findings indicating that the cost of immediacy, or the cost of trading via the market-making mechanism, has risen (Bao, O'Hara, and Zhou (2018), Dick-Nielsen and Rossi (2018), Choi and Huh (2017)). This is hardly surprising given that post-crisis regulations increased the bank dealers' cost of balance-sheet financing. At the same time, there are somewhat surprising empirical findings that average transaction costs have declined (Mizrach (2015), Adrian, Fleming, Shachar, and Vogt (2017), Anderson and Stulz (2017)) or have not changed (Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), Trebbi and Xiao (2019)) following the enactment of post-crisis regulations.[29]

In our model, these findings appear consistent with an increase in c_B that starts in the first region of Figure 1 (the unconstrained bank dealer equilibrium) and ends in the second region (the constrained bank dealer equilibrium). The spread is certainly rising in the first region, but average transaction costs fall most prominently in the second region.[30] What makes the second region special, in addition to the increase in volume, is the decrease in the matchmaking fee. While the decline in average transaction costs constitutes indirect evidence that the matchmaking fee could have decreased, under some conditions this result can reflect the change in the fractions of traders utilizing these two services rather than a change in pricing. The aforementioned empirical studies do not provide more direct evidence regarding changes in execution costs for agency and riskless-principal trades (the empirical equivalents of our matchmaking fee) following the implementation of post-crisis regulations. Hence, determining whether the corporate bond market has shifted to the constrained bank dealer equilibrium may need to await additional empirical findings pertaining to the matchmaking fee.

Our model also provides empirical implications for volume, which can rise, fall, or remain the same depending on the parameter region for c_B. Here, the empirical evidence appears somewhat nuanced. Overall dollar trading volume in bonds has significantly increased, while turnover in each particular bond issue appears to have declined (increased) in more (less) active bonds (BIS Committee on the Global Financial System (2014), Mizrach (2015), Adrian, Fleming, Shachar, and Vogt (2017)).[31] Because turnover is computed as dollar trading volume divided by the value of the bond issue, though, patterns in issuance affect this measure. Some market observers note that the

[29]Two papers find more nuanced effects. Allahrakha, Cetina, Munyan, and Watugala (2019) find higher markups for a subset of trades when looking at Volcker Rule exemptions (e.g., trades in newly issued bonds for which a bank dealer is part of the bond's underwriting group) to infer cost differentials. Chernenko and Sunderam (2018) develop an indirect measure of aggregate corporate bond market liquidity by relating mutual funds' cash holdings to the volatility of their fund flows. They find that, while the liquidity of investment-grade bonds in the post-crisis period essentially recovered to the pre-crisis level, liquidity for speculative grade bonds has not.

[30]If both the matchmaking cost and the bank dealer's balance sheet cost are low enough, the decline in average transaction costs could possibly start already in the right portion of the first region if c_B increases beyond $(1-\mathcal{H})\, f^{\star}+I$ before the equilibrium changes to the constrained bank dealer equilibrium.

[31]Anderson and Stulz (2017) find that turnover increased for investment-grade bonds when looking at the entire universe of bonds in Mergent's FISD database, but the result is reversed when focusing on bond trades in the TRACE database.

low interest-rate environment greatly boosted the attractiveness of bond financing, and business enterprises responded by issuing a record number of bonds. According to this explanation, abnormal issuance, not a decline in the desire to trade, reduced turnover in some bond issues (BIS Committee on the Global Financial System (2014)). A better understanding of the abnormal issuance effect and therefore whether volume in fact increased awaits further empirical work. Researchers may also want to examine how the breadth of investor participation in the corporate bond market, which is equivalent to volume in our model, has changed after post-crisis regulations were imposed to further our understanding of which parameter regions of the model better reflect the experience of the market.

6 Conclusions

Our paper highlights the complex and multifaceted consequences that post-crisis bank regulations have for market liquidity and investor welfare. We have focused on demonstrating how an increase in regulatory costs can improve customer welfare by changing the market structure for trading securities. The regulatory reform that increased the cost of balance-sheet financing of bank dealers has been promoted as a means of creating more resilient intermediaries that "should be better able to absorb risks under stressed market conditions and reduce the risk of market disruptions" (BIS Committee on the Global Financial System (2016)). The potential for investor welfare improvements from this reform presumably comes from preventing very bad outcomes during those stressed periods as well as from reducing the too-big-to-fail subsidy to bank dealers that is paid for in one way or another by investors.

Our model points to another source of welfare gain that arises even during normal times: the increase in bank dealer regulatory costs prompts a change in the nature of liquidity provision. In particular, post-crisis regulations eliminate obstacles to competition in the most profitable business (market making) and incentivize bank dealers to reprice their services to steer customers to an alternative (matchmaking) that better serves the needs of many customers. Three elements of the corporate bond market drive our results: the coexistence of two distinct trading mechanisms (market making and matchmaking), the market power enjoyed by bank dealers, and potential market-making competition from non-bank dealers. Although our paper is motivated by and specifically addresses observations in the corporate bond market, our model can be applied to other over-the-counter markets that feature these three elements. The industrial organization angle combines with the market microstructure angle to deliver this positive outcome, which highlights the role of regulations in influencing the structure of securities markets by fostering competition.

While the key insight we offer in this paper is that imposing higher regulatory costs on banks can make customers in the corporate bond market better off, there is a limit to what such a model can show. Specifically, our model does not allow us unequivocally to ascertain whether customer welfare in the corporate bond market indeed increased as a result of the regulatory reform that followed the financial crisis. Our discussion in the previous section suggests that with additional empirical analysis one could potentially determine whether we have moved from the first region (the unconstrained bank dealer equilibrium) to the second region (the constrained bank dealer equilibrium) in which increases in balance sheet costs improve welfare. Even if additional empirical evidence establishes more firmly that the market has shifted into the constrained bank dealer equilibrium, however, at least conceptually the change in overall customer welfare could be ambiguous because it aggregates both the welfare decline in the first region and the welfare improvement in the second region as bank regulatory costs increased.

Still, it is important to bear in mind the origin of the difference in balance sheet costs between bank and non-bank dealers, which we take as exogenous in our model. If part of this difference reflects a too-big-to-fail subsidy, investors are presumably paying for such a subsidy elsewhere in the economy. The reduction in welfare in the first region of our model would then be overstated because it is offset to some degree by the increase in welfare as the subsidy is reduced. On the other hand, the improvement in welfare in the second region is very robust because it is generated by intensified competition between bank and non-bank dealers that incentivizes dealers to shift business activities to the less expensive matchmaking service. Such an improvement in welfare adds to any welfare gain achieved by reducing the too-big-to-fail subsidy, generating a win-win situation for investors.

The welfare gains could be even more pronounced when combined with another important development that took place in the past decade: technological advances that reduced the cost of matchmaking and therefore rendered the matchmaking mechanism more attractive (see, for example, BIS Markets Committee (2016)). While one might assume that a reduction in the cost of search (or effort) required to effect a transaction in the matchmaking mechanism would always improve overall customer welfare, we establish in the Internet Appendix that such an unambiguous result can be shown only in the constrained equilibria. Hence, the post-crisis regulatory reform could potentially foster an environment in which benefits from technological advancement accrue not just to the dealers but also to their customers.

The evolving regulatory frameworks and the breathtaking pace at which technology impacts securities markets continue to dominate the agendas of regulators, practitioners, and academics. We hope that our work serves to both highlight important tradeoffs and spur additional work on the changing nature of our securities markets.

Appendices

A Proofs

A.1 Proof of Proposition 1

Consider first the case when $c_B \in \left(\frac{I}{\mathcal{H}}, c_{NB}\right]$. It is clear that in this case only the bank dealer operates a market-making service, and the non-bank dealer is passive. The bank dealer's problem is

$$\max_{\frac{I}{\mathcal{H}}<f<S<c_{NB}} \pi_B = \frac{2\mu}{r}\left[(\mathcal{H}f - I)\frac{\frac{S-\mathcal{H}f}{1-\mathcal{H}} - f}{A} + (S - c_B)\left(1 - \frac{\frac{S-\mathcal{H}f}{1-\mathcal{H}}}{A}\right)\right].$$

We verify later that the assumption

$$A > c_{NB} + \frac{\mathcal{H}}{2(1-\mathcal{H})}\left(c_{NB} - \frac{I}{\mathcal{H}}\right)$$

guarantees that in equilibrium we have $b = \frac{S_1^* - \mathcal{H}f_1^*}{1-\mathcal{H}} < A$. The first-order derivatives are

$$\begin{aligned}\frac{\partial \pi_B}{\partial f} &= \frac{2\mu}{r}\frac{2\mathcal{H}(S-f) + I - \mathcal{H}c_B}{(1-\mathcal{H})A} \\ &= \frac{2\mu}{r}\frac{2\mathcal{H}}{(1-\mathcal{H})A}\left(S + \frac{I - \mathcal{H}c_B}{2\mathcal{H}} - f\right),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \pi_B}{\partial S} &= \frac{2\mu}{r}\left(\frac{-2S + 2\mathcal{H} + c_B - I}{(1-\mathcal{H})A} + 1\right) \\ &= \frac{2\mu}{r}\frac{2}{(1-\mathcal{H})A}\left(\mathcal{H}f + \frac{1}{2}(c_B - I + (1-\mathcal{H})A) - S\right).\end{aligned}$$

The solutions to the FOCs are

$$f_0^* = \frac{I}{2\mathcal{H}} + \frac{A}{2}$$

and

$$S_0^* = \frac{c_B}{2} + \frac{A}{2}.$$

Since we consider the region $c_B > \frac{I}{\mathcal{H}}$, we must have $S_0^* > f_0^*$. This is an unconstrained equilibrium only when

$$S_0^* < c_{NB} \Longleftrightarrow c_B < \underline{c} = 2c_{NB} - A.$$

Because the objective function is a quadratic function of (S, f), the above FOCs are also sufficient conditions in this optimization problem. So when $c < \underline{c}$, there is an unconstrained bank dealer equilibrium in which only the bank dealer operates both market-making and matchmaking services, and the prices are

$$S^* = S_0^* = \frac{c_B}{2} + \frac{A}{2}$$

and

$$f^* = f_0^* = \frac{I}{2\mathcal{H}} + \frac{A}{2}.$$

When $S_0^* > c_{NB}$, the solution must be

$$S^* = c_{NB},$$

otherwise if $S^* < c_{NB}$, the solution is interior and violates the condition $S_0^* > c_{NB}$, and if $S^* > c_{NB}$, the non-bank dealer will operate a market-making service and attract all market-making customers. This is the constrained bank dealer equilibrium. Then the solution to

$$\frac{\partial \pi_B}{\partial f} = 0$$

is

$$f^* = \frac{-c_B\mathcal{H} + I + 2\mathcal{H}c_{NB}}{2\mathcal{H}}.$$

Obviously the condition $\frac{I}{\mathcal{H}} < f^* < c_{NB}$ always holds in this equilibrium. The condition

$$A > c_{NB} + \frac{\mathcal{H}}{2\left(1-\mathcal{H}\right)}\left(c_{NB} - \frac{I}{\mathcal{H}}\right)$$

guarantees that $b < A$ always holds in this case.

Consider the case when $c_B > c_{NB}$. In this equilibrium, it is clear that only the non-bank dealer operates a market-making service. The bank dealer's objective function is

$$\pi_B = \frac{2\mu}{r}\left(\mathcal{H}f - I\right)\frac{\frac{S-\mathcal{H}f}{1-\mathcal{H}} - f}{A}.$$

The non-bank dealer's objective function is

$$\pi_{NB} = \frac{2\mu}{r}\left(S - c_B\right)\left(1 - \frac{\frac{S-\mathcal{H}f}{1-\mathcal{H}}}{A}\right).$$

The FOCs are

$$\frac{\partial \pi_B}{\partial f} = \frac{2\mu}{rA} \cdot \frac{I + \mathcal{H}S - 2\mathcal{H}f}{1-\mathcal{H}}$$

and

$$\frac{\partial \pi_{NB}}{\partial S} = \frac{2\pi}{rA} \cdot \frac{A\left(1-\mathcal{H}\right) + c_{NB} + \mathcal{H}f - 2S}{1-\mathcal{H}}.$$

Then, $\frac{\partial \pi_B}{\partial f} = 0$ and $\frac{\partial \pi_{NB}}{\partial S} = 0$ imply that

$$S_1^* = \frac{2c_{NB} + I + 2A\left(1-\mathcal{H}\right)}{4-\mathcal{H}}$$

and

$$f_1^* = \frac{c_{NB}\mathcal{H} + 2I + \mathcal{H}\left(1-\mathcal{H}\right)A}{\left(4-\mathcal{H}\right)\mathcal{H}}.$$

The assumption

$$A > c_{NB} + \frac{\mathcal{H}}{2(1-\mathcal{H})}\left(c_{NB} - \frac{I}{\mathcal{H}}\right)$$

guarantees that

$$S_1^* > c_{NB}$$

and

$$\frac{S_1^* - \mathcal{H} f_1^*}{1-\mathcal{H}} < A.$$

Then, when $c_B \leq \bar{c} = \frac{2A(1-\mathcal{H})+2c_{NB}+I}{4-\mathcal{H}}$, the solution must be

$$S^* = c_B.$$

This is because the non-bank dealer's objective function is a quadratic function of S, and the non-bank dealer will never choose $S > c_B$. This gives us the constrained non-bank dealer equilibrium. In this case, the bank dealer's best response is

$$f^* = \frac{1}{2}\left(\frac{I}{\mathcal{H}} + c_B\right).$$

When $c_B > \bar{c}$, the solution is interior, and thus in the unconstrained non-bank dealer equilibrium

$$S^* = \bar{c} = \frac{2A(1-\mathcal{H}) + 2c_{NB} + I}{4-\mathcal{H}}$$

and

$$f^* = f_1^* = \frac{\mathcal{H} c_{NB} + \mathcal{H}(1-\mathcal{H})A + 2I}{(4-\mathcal{H})\mathcal{H}}.$$

A.2 Proof of Proposition 2

Unconstrained Bank Dealer Equilibrium

In the unconstrained bank dealer equilibrium, the spread $S^* = \frac{1}{2}(c_B + A)$ is increasing in c_B, and $f^* = \frac{1}{2}\left(A + \frac{I}{\mathcal{H}}\right)$ is independent of c_B. The cutoff

$$\begin{aligned} b^* &= \frac{S^* - \mathcal{H} f^*}{1-\mathcal{H}} \\ &= \frac{A}{2} + \frac{1}{2}\frac{c_B - I}{1-\mathcal{H}} \end{aligned}$$

is also increasing in c_B. Hence, total volume $(1-f^*)$ is independent of c_B, matchmaking volume $b^* - f^*$ is increasing in c_B, and market-making volume $(1-b^*)$ is decreasing in c_B.

Average transaction costs are

$$ATC = \frac{(b^* - f^*) f^* + (A - b^*) S^*}{A - f^*}.$$

We can show that

$$\frac{dATC}{dc_B} = \frac{-2c_B\mathcal{H} + \mathcal{H}(1-\mathcal{H})A + (1+\mathcal{H})I}{2(1-\mathcal{H})(\mathcal{H}A - I)}.$$

Since $\frac{dATC}{dc_B}$ is a decreasing linear function of c_B, ATC must be a hump-shaped function of c_B.

Overall customer welfare is

$$\pi_c = \frac{2\pi}{r}\int_{f^*}^{b^*} \mathcal{H}(x - f^\star)\frac{dx}{A} + \frac{2\pi}{r}\int_{b^*}^{A}(x - S^*)\frac{dx}{A},$$

and

$$\frac{d\pi_c}{dc_B} = \frac{2\pi}{r}\int_{b^*}^{A} -\frac{dS^*}{dc_B}\frac{dx}{A} < 0.$$

Constrained Bank Dealer Equilibrium

In the constrained bank dealer equilibrium, the spread $S^* = c_{NB}$ is constant, and $f^* = \frac{-c_B\mathcal{H}+I+2\mathcal{H}c_{NB}}{2\mathcal{H}}$ is decreasing in c_B. The cutoff

$$\begin{aligned} b^* &= \frac{S^* - \mathcal{H}f^*}{1-\mathcal{H}} \\ &= c_{NB} + \frac{\mathcal{H}c_B - I}{2(1-\mathcal{H})} \end{aligned}$$

is increasing in c_B. So total volume $(1 - f^*)$ is increasing c_B, matchmaking volume $b^* - f^*$ is increasing in c_B, and market-making volume $(1 - b^*)$ is decreasing in c_B.

Average transaction costs are

$$ATC = \frac{(b^* - f^*)f^* + (A - b^*)S^*}{A - f^*}.$$

We can show that

$$\frac{dATC}{dc_B} = -\frac{(c_B\mathcal{H} - I)(4(A - c_{NB})\mathcal{H} + c_B\mathcal{H} - I)}{2(1-\mathcal{H})(-2A\mathcal{H} - c_B\mathcal{H} + 2c_{NB}\mathcal{H} + I)^2} < 0.$$

Overall customer welfare is

$$\pi_c = \frac{2\pi}{r}\int_{f^*}^{b^*} \mathcal{H}(x - f^\star)\frac{dx}{A} + \frac{2\pi}{r}\int_{b^*}^{A}(x - S^*)\frac{dx}{A},$$

and

$$\frac{d\pi_c}{dc_B} = \frac{2\pi}{r}\int_{b^*}^{A} -\frac{df^*}{dc_B}\frac{dx}{A} > 0.$$

Constrained Non-Bank Dealer Equilibrium

In the constrained non-bank dealer equilibrium, the spread $S^* = c_B$ is increasing in c_B and $f^* = \frac{1}{2}\left(\frac{I}{\mathcal{H}} + c_B\right)$ is increasing in c_B. The cutoff

$$\begin{aligned} b^* &= \frac{S^* - \mathcal{H} f^*}{1 - \mathcal{H}} \\ &= \frac{c_B (2 - \mathcal{H}) - I}{2(1 - \mathcal{H})} \end{aligned}$$

is also increasing in c_B. So total volume $(1 - f^*)$ is decreasing in c_B, matchmaking volume

$$b^* - f^* = \frac{\mathcal{H} c_B - I}{2\mathcal{H} - 2\mathcal{H}^2}$$

is increasing in c_B, and market-making volume $(1 - b^*)$ is decreasing in c_B.

Average transaction costs are

$$ATC = \frac{(b^* - f^*) f^* + (A - b^*) S^*}{A - f^*}.$$

We can show that

$$\frac{dATC}{dc_B} = \frac{\begin{array}{c} 8A^2(-1+\mathcal{H})\mathcal{H}^2 + c_B^2\mathcal{H}^2(-3+2\mathcal{H}) + 2c_B\mathcal{H}(-3+2\mathcal{H})I \\ +(1+2\mathcal{H})I^2 + 4A\mathcal{H}(c_B(3-2\mathcal{H})\mathcal{H} + I - 2\mathcal{H}I) \end{array}}{2(-1+\mathcal{H})(-2A\mathcal{H} + c_B\mathcal{H} + I)^2}.$$

The sign of the numerator of $\frac{dATC}{dc_B}$ is ambiguous.

Overall customer welfare is

$$\pi_c = \frac{2\pi}{r}\int_{f^*}^{b^*} \mathcal{H}(x - f^\star)\frac{dx}{A} + \frac{2\pi}{r}\int_{b^*}^{A}(x - S^*)\frac{dx}{A},$$

and

$$\frac{d\pi_c}{dc_B} = \frac{2\pi}{r}\int_{f^*}^{b^*} -\mathcal{H}\frac{df^*}{dc_B}\frac{dx}{A} + \frac{2\pi}{r}\int_{b^*}^{A} -\frac{dS^*}{dc_B}\frac{dx}{A} < 0.$$

Unconstrained Non-Bank Dealer Equilibrium

All equilibrium variables in this case are independent of the increase in c_B.

A.3 Proof of Lemma 1

Since $\mathcal{H}_B > \mathcal{H}_N$, we must have $f_B > f_N$ in equilibrium because otherwise no customer will choose the non-bank dealer matchmaking service. The cutoff type who is indifferent between non-bank dealer matchmaking and bank dealer matchmaking satisfies

$$\mathcal{H}_B(x - f_B) = \mathcal{H}_{NB}(x - f_{NB}) \Longleftrightarrow x = x_1 = \frac{\mathcal{H}_B f_B - \mathcal{H}_N f_N}{\mathcal{H}_B - \mathcal{H}_N}.$$

The cutoff type who is indifferent between bank dealer matchmaking and bank dealer market making satisfies

$$\mathcal{H}_B\left(x - f_B\right) = \left(x - S\right) \Longleftrightarrow x = x_2 = \frac{S - \mathcal{H}_B f_B}{1 - \mathcal{H}_B}.$$

A.4 Proof of Proposition 3

For notational simplicity, in this proof let us introduce

$$\Delta = \mathcal{H}_B - \mathcal{H}_{NB} > 0.$$

We focus on the equilibrium in which the non-bank dealer operates a matchmaking service and the bank dealer operates both matchmaking and market-making services. Suppose the equilibrium fee for the non-bank dealer's matchmaking service is f_{NB}^*, the equilibrium fee for the bank dealer's matchmaking service is f_B^*, and the spread for the bank dealer's market-making service is S^*. In this equilibrium, we must have

$$\frac{I}{\mathcal{H}_N} < f_{NB}^* < f_B^* < S^* \leq c_{NB}. \tag{32}$$

Let

$$x_1^* = \frac{\mathcal{H}_B f_B^* - \mathcal{H}_{NB} f_{NB}^*}{\Delta}$$

and

$$x_2^* = \frac{S^* - \mathcal{H}_B f_B^*}{1 - \mathcal{H}_B}.$$

Then, customers with private value $x \in [f_{NB}^*, x_1^*]$ choose non-bank dealer matchmaking, customers with private value $x \in (x_1^*, x_2^*)$ choose bank dealer matchmaking, and customers with private value $x \in [x_2^*, A]$ choose bank dealer market making. Another condition that must be satisfied in this equilibrium is

$$\frac{I}{\mathcal{H}_{NB}} < f_{NB}^* < x_1^* < x_2^* < A. \tag{33}$$

When both the bank dealer's and the non-bank dealer's choice variables $(f_{NB}, f_B, S \leq c_{NB})$ change locally around the equilibrium (f_{NB}^*, f_B^*, S^*), the non-bank dealer's objective function is

$$\begin{aligned}\Pi_{NB}\left(f; f_B\right) &= \frac{2\mu}{r}\left(\mathcal{H}_{NB} f - I\right)\frac{\frac{\mathcal{H}_B f_B - \mathcal{H}_{NB} f}{\Delta} - f}{A}\\ &= \frac{2\mu}{r}\frac{\mathcal{H}_B}{\Delta \cdot A}\left(\mathcal{H}_{NB} f - I\right)\left(f_B - f\right)\end{aligned}$$

and the bank dealer's objective function is

$$\Pi_B\left(f, S; f_{NB}\right) = \frac{2\mu}{r}\left[\left(\mathcal{H}_B f - I\right)\frac{\frac{S - \mathcal{H}_B f}{1 - \mathcal{H}_B} - \frac{\mathcal{H}_B f - \mathcal{H}_{NB} f_{NB}}{\Delta}}{A} + \left(S - c_B\right)\left(1 - \frac{\frac{S - \mathcal{H}_B f}{1 - \mathcal{H}_B}}{A}\right)\right].$$

For the non-bank dealer, the best response is

$$f_{NB}^{br}(f_B) = \frac{1}{2}\left(f_B + \frac{I}{\mathcal{H}_{NB}}\right). \tag{34}$$

Since f_N^* is an interior solution in the non-bank dealer's optimization problem (see condition (32)), we must have

$$f_{NB}^* = \frac{1}{2}\left(f_B^* + \frac{I}{\mathcal{H}_{NB}}\right). \tag{35}$$

For the bank dealer, the first-order derivatives are

$$\frac{\partial \Pi_B(f,S)}{\partial f} = \frac{2\mu}{rA}\frac{\mathcal{H}_B\left(-c_B\Delta - 2f\mathcal{H}_B(1-\mathcal{H}_{NB}) + f_N\mathcal{H}_{NB}(1-\mathcal{H}_B) + (1-\mathcal{H}_{NB})I + 2S\Delta\right)}{(1-\mathcal{H}_B)\Delta}$$

and

$$\frac{\partial \Pi_B(f,S)}{\partial S} = \frac{2\mu}{rA}\frac{c_B + 2f\mathcal{H}_B - I - 2S + A(1-\mathcal{H}_B)}{(1-\mathcal{H}_B)}.$$

Since f_B^* is an interior solution in the bank dealer's optimization problem (see condition (32)), we must have

$$\frac{\partial \Pi_B(f,S)}{\partial f}|_{(f_{NB}^*, f_B^*, S^*)} = 0$$

hence

$$f_B^* = \frac{-c_B\Delta + f_{NB}^*(1-\mathcal{H}_B)\mathcal{H}_{NB} + I(1-\mathcal{H}_{NB}) + 2S^*\Delta}{2\mathcal{H}_B(1-\mathcal{H}_{NB})}. \tag{36}$$

Both (35) and (36) are sufficient conditions for the equilibrium (f_N^*, f_B^*, S^*), and they jointly imply that

$$f_{NB}^* = \frac{1}{2}\left(\frac{-2c_B\Delta + (3-\mathcal{H}_B - 2\mathcal{H}_{NB})I + 4S^*\Delta}{\Delta + 3\mathcal{H}_B(1-\mathcal{H}_{NB})} + \frac{I}{\mathcal{H}_{NB}}\right) \tag{37}$$

and

$$f_B^* = \frac{-2c_B\Delta + (3-\mathcal{H}_B - 2\mathcal{H}_{NB})I + 4S^*\Delta}{\Delta + 3\mathcal{H}_B(1-\mathcal{H}_{NB})}. \tag{38}$$

Substituting (37) and (38) into $\frac{\partial \Pi_B(f,S)}{\partial S}$, we get

$$\frac{\partial \Pi_B(f,S)}{\partial S}|_{(f_{NB}^*, f_B^*, S^*)} = \frac{2\mu(8\mathcal{H}_B - 2\mathcal{H}_{NB})}{r(\Delta + 3\mathcal{H}_B(1-\mathcal{H}_{NB}))}\left[\frac{c_B}{2} + \frac{I(2\mathcal{H}_B + \mathcal{H}_{NB}) + A(\Delta + 3\mathcal{H}_B(1-\mathcal{H}_{NB}))}{8\mathcal{H}_B - 2\mathcal{H}_{NB}} - S^*\right].$$

Let

$$S^u = \frac{c_B}{2} + \frac{I(2\mathcal{H}_B + \mathcal{H}_{NB}) + A(\Delta + 3\mathcal{H}_B(1-\mathcal{H}_{NB}))}{8\mathcal{H}_B - 2\mathcal{H}_{NB}},$$

$$f_B^u = \frac{3I + 2A\Delta}{4\mathcal{H}_B - \mathcal{H}_{NB}}$$

and

$$f_{NB}^u = \frac{I(2\mathcal{H}_B + \mathcal{H}_{NB}) + A\mathcal{H}_{NB}\Delta}{4\mathcal{H}_{NB}\Delta}$$

be the prices in the unconstrained market-making equilibrium.

So, when

$$\frac{I}{\mathcal{H}_{NB}} < f_{NB}^u < f_B^u < S^u < c_{NB}$$

and

$$\frac{I}{\mathcal{H}_{NB}} < f_{NB}^u < \frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{f_B^u - f_{NB}^u} < \frac{S^u - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B} < A,$$

we have

$$S^* = S^u, \; f_B^* = f_B^u, \; f_{NB}^* = f_{NB}^u.$$

This is the unconstrained market-making equilibrium. Otherwise we have

$$S^* = c_{NB}.$$

The matchmaking fees for the bank and non-bank dealers in the constrained equilibrium are

$$f_B^* = f_B^c = \frac{-2c_B\Delta + (3 - \mathcal{H}_B - 2\mathcal{H}_{NB})\, I + 4c_{NB}\Delta}{\Delta + 3\mathcal{H}_B\,(1 - \mathcal{H}_{NB})}$$

and

$$f_{NB}^* = f_{NB}^c = \frac{1}{2}\left(\frac{-2c_B\Delta + (3 - \mathcal{H}_B - 2\mathcal{H}_{NB})\, I + 4c_{NB}\Delta}{\Delta + 3\mathcal{H}_B\,(1 - \mathcal{H}_{NB})} + \frac{I}{\mathcal{H}_{NB}}\right).$$

In summary, the necessary conditions for (f_{NB}^u, f_B^u, S^u) to be an unconstrained market-making equilibrium are

$$\frac{I}{\mathcal{H}_{NB}} < f_{NB}^u < f_B^u < S^u < c_{NB}$$

and

$$\frac{I}{\mathcal{H}_{NB}} < f_{NB}^u < \frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{S^u - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B} < A.$$

From (34), we can simplify the above conditions based on the following observations:

- $f_{NB}^u < f_B^u \Longrightarrow \frac{I}{\mathcal{H}_{NB}} < f_{NB}^u < f_B^u$;
- $\frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{S^u - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B} \Longrightarrow S^u > f_B^u + \frac{\mathcal{H}_{NB}(1-\mathcal{H}_B)}{\Delta}\left(f_B^u - f_{NB}^u\right)$.

Then,

- $\left\{f_{NB}^u < f_B^u \& \frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{S^u - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B}\right\} \Longrightarrow S^u > f_B^u$;
- $f_{NB}^u < f_B^u \Longrightarrow f_{NB}^u < \frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}}$.

Then,

$$\left\{\begin{array}{c} \frac{I}{\mathcal{H}_{NB}} < f_B^u \\ S^u < c_{NB} \\ \frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{S^u - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B} < A \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{c} \frac{I}{\mathcal{H}_{NB}} < f_{NB}^u < f_B^u < S^u < c_{NB} \\ \frac{I}{\mathcal{H}_{NB}} < f_{NB}^u < \frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{S^u - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B} < A \end{array}\right\}$$

Condition $f_{NB}^u < f_B^u$ always holds because our assumptions imply that $\mathcal{H}_{NB}A - 2I > 0$, and thus

$$f_B^u - f_{NB}^u = \frac{\Delta\,(\mathcal{H}_{NB}A - 2I)}{\mathcal{H}_{NB}\,(4\mathcal{H}_B - \mathcal{H}_{NB})} > 0.$$

Condition $\frac{S^u-\mathcal{H}_B f_B^u}{1-\mathcal{H}_B} < A$ is also implied by our assumption on the lower bound of A.

Also,

$$\left\{ \begin{array}{c} S^u \leq c_{NB} \\ \frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{S^u - \mathcal{H}_B f_B^u}{1-\mathcal{H}_B} \end{array} \right\} \Longleftrightarrow c_1 < c_B \leq c_2,$$

where

$$c_1 = \frac{(2+2\mathcal{H}_B - \mathcal{H}_{NB}) I - (1-\mathcal{H}_B)\mathcal{H}_{NB} A}{4\mathcal{H}_B - \mathcal{H}_{NB}}$$

and

$$c_2 = 2c_{NB} - \frac{I(2\mathcal{H}_B + \mathcal{H}_{NB}) + A(\Delta + 3\mathcal{H}_B(1-\mathcal{H}_{NB}))}{4\mathcal{H}_B - \mathcal{H}_{NB}}.$$

These are also sufficient conditions for (f_{NB}^u, f_B^u, S^u) to be an equilibrium. To see this, we just need to verify that f_{NB}^u is the global optimum in the non-bank dealer's optimization problem and that (f_B^u, S^u) is the global optimum in the bank dealer's optimization problem.

It is easy to verify the first condition because, from the FOC, f_{NB}^* is the non-bank dealer's best response when she chooses from $f_{NB} \in \left(\frac{I}{\mathcal{H}_{NB}}, f_B^u\right)$. If the non-bank dealer chooses $f_{NB} \leq \frac{I}{\mathcal{H}_{NB}}$, the profit will be non-positive, which is suboptimal. If the non-bank dealer chooses $f_{NB} \geq f_B^u$, the profit will be zero, which is also suboptimal. Hence, f_{NB}^u is the global optimum in the non-bank dealer's optimization problem.

We discussed above the local optimality of (f_B^u, S^u), i.e., for the bank dealer, (f_B^u, S^u) is the best response if the bank dealer chooses from

$$\left\{ (f_B, S) \,|\, f_{NB}^u < f_B \leq S \leq c_{NB} \& f_{NB}^u < \frac{\mathcal{H}_B f_B - \mathcal{H}_{NB} f_{NB}}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{S - \mathcal{H}_B f_B}{1-\mathcal{H}_B} < A \right\}.$$

To show global optimality, first, it is suboptimal for the bank dealer to choose $f_B < f_{NB}^u$ and provide all the matchmaking service. To see this, note that $f_B \leq f_{NB}^u \Longleftrightarrow x_1 = \frac{\mathcal{H}_B f_B - \mathcal{H}_{NB} f_{NB}^u}{\Delta} \leq f_B$, and then the bank dealer's objective becomes

$$\begin{aligned} &\frac{2\mu}{r}\left[(\mathcal{H}_B f - I)\frac{\frac{S-\mathcal{H}_B f}{1-\mathcal{H}_B} - f_B}{A} + (S - c_B)\left(1 - \frac{\frac{S-\mathcal{H}_B f_B}{1-\mathcal{H}_B}}{A}\right)\right] \\ \leq &\frac{2\mu}{r}\left[(\mathcal{H}_B f - I)\frac{\frac{S-\mathcal{H}_B f}{1-\mathcal{H}_B} - \frac{\mathcal{H}_B f_B - \mathcal{H}_{NB} f_{NB}^u}{\Delta}}{A} + (S - c_B)\left(1 - \frac{\frac{S-\mathcal{H}_B f_B}{1-\mathcal{H}_B}}{A}\right)\right], \end{aligned}$$

which implies that choosing $f_B < f_{NB}^u$ is suboptimal.

It is also suboptimal to choose $\frac{S-\mathcal{H}_B f}{1-\mathcal{H}_B} > A$ and shut down the market-making service. If the bank dealer chooses $\frac{S-\mathcal{H}_B f}{1-\mathcal{H}_B} > A$, the choice of S becomes irrelevant and WLOG we can choose S such that $\frac{S-\mathcal{H}_B f}{1-\mathcal{H}_B} = A$, and we have already shown that this is suboptimal.

It is suboptimal for the bank dealer to choose f_B such that $x_1 \geq x_2$ and there is no bank dealer matchmaking. To see this, let f_x be the solution to

$$\frac{\mathcal{H}_B f_x - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} = \frac{S^u - \mathcal{H}_B f_x}{1-\mathcal{H}_B}.$$

When the bank dealer chooses $f_B \geq f_x$, the choice of f_B becomes irrelevant, and we know that

$\Pi_B(f_x, S^u) < \Pi_B(f_B^u, S^u)$, so it is suboptimal to choose $f \geq f_x$.

It is suboptimal for the bank dealer to choose S such that $x_1 \geq x_2$ and there is no bank matchmaking. To see this, let S_x be the solution to

$$\frac{\mathcal{H}_B f_B^u - \mathcal{H}_{NB} f_{NB}^u}{\mathcal{H}_B - \mathcal{H}_{NB}} = \frac{S_x - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B},$$

when the bank chooses $S \leq S_x$, $x_1 \geq x_2$ and there is no bank dealer matchmaking. In this case, we must have

$$\frac{S - \mathcal{H}_B f_B^u}{1 - \mathcal{H}_B} \leq \frac{S - \mathcal{H}_{NB} f_{NB}^u}{1 - \mathcal{H}_{NB}};$$

otherwise, customers whose private values are in $\left(\frac{S-\mathcal{H}_{NB} f_{NB}^u}{1-\mathcal{H}_{NB}}, \frac{S-\mathcal{H}_B f_B^u}{1-\mathcal{H}_B}\right)$ would prefer bank dealer market making to non-bank matchmaking while preferring bank dealer matchmaking to bank dealer market making. We must then have $x_1 < x_2$. The bank dealer's profit is then

$$\frac{2\mu}{r}(S - c_B)\frac{A - \frac{S-\mathcal{H}_{NB} f_{NB}^u}{1-\mathcal{H}_{NB}}}{A} \leq \frac{2\mu}{r}(S - c_B)\frac{A - \frac{S-\mathcal{H}_B f_B^u}{1-\mathcal{H}_B}}{A},$$

and

$$\begin{aligned}
\left.\frac{\partial\left[\frac{2\mu}{r}(S - c_B)\frac{A - \frac{S-\mathcal{H}_B f_B^u}{1-\mathcal{H}_B}}{A}\right]}{\partial S}\right|_{f_{NB}^u, f_B^u, S} &= \frac{2\mu}{rA}\frac{c_B + 2f_B^u\mathcal{H}_B - I - 2S + A(1-\mathcal{H}_B)}{(1-\mathcal{H}_B)} \\
&> \frac{2\mu}{rA}\frac{c_B + 2f_B^u\mathcal{H}_B - I - 2S_x + A(1-\mathcal{H}_B)}{(1-\mathcal{H}_B)} \\
&> 0
\end{aligned}$$

The last inequality is from

$$\left.\frac{\partial \Pi_B(f, S)}{\partial S}\right|_{f_{NB}^u, f_B^u, S_x} = \frac{2\mu}{rA}\frac{c_B + 2f_B^u\mathcal{H}_B - I - 2S_x + A(1-\mathcal{H}_B)}{(1-\mathcal{H}_B)} > 0.$$

Therefore we know that for any $S < S_x$,

$$\frac{2\mu}{r}(S - c_B)\frac{A - \frac{S-\mathcal{H}_{NB} f_{NB}^u}{1-\mathcal{H}_{NB}}}{A} < \Pi_B(f_B^u, S_x) < \Pi_B(f_B^u, S^u).$$

For the constrained market-making equilibrium, the necessary conditions for $(f_{NB}^c, f_B^c, S^c = c_{NB})$ to be the equilibrium are

$$\frac{I}{\mathcal{H}_{NB}} < f_{NB}^c < f_B^c < c_{NB} \Longleftrightarrow \frac{I}{\mathcal{H}_{NB}} < f_B^c < c_{NB}$$

$$c_{NB} < S^u$$

and

$$\frac{I}{\mathcal{H}_{NB}} < f_{NB}^c < \frac{\mathcal{H}_B f_B^c - \mathcal{H}_{NB} f_{NB}^c}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{c_{NB} - \mathcal{H}_B f_B^c}{1 - \mathcal{H}_B} < A.$$

By a similar argument, we can also show that these are sufficient conditions for (f_{NB}^c, f_B^c, S^c) to be the constrained market-making equilibrium. We can simplify the above conditions based on the following observations:

- $\frac{I}{\mathcal{H}_{NB}} < f_B^c \Longrightarrow \frac{I}{\mathcal{H}_{NB}} < f_{NB}^c < f_B^c \Longrightarrow f_{NB}^c < \frac{\mathcal{H}_B f_B^c - \mathcal{H}_{NB} f_{NB}^c}{\mathcal{H}_B - \mathcal{H}_{NB}}$;
- $\frac{\mathcal{H}_B f_B^c - \mathcal{H}_{NB} f_{NB}^c}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{c_{NB} - \mathcal{H}_B f_B^c}{1 - \mathcal{H}_B} \Longrightarrow c_{NB} > f_B^c + \frac{\mathcal{H}_{NB}(1-\mathcal{H}_B)}{2\Delta}\left(f_B^c - \frac{I}{\mathcal{H}_{NB}}\right) > f_B^c$.

Since $f_B^c < f_B^u$, we have $f_B^c + \frac{\mathcal{H}_{NB}(1-\mathcal{H}_B)}{2\Delta}\left(f_B^c - \frac{I}{\mathcal{H}_{NB}}\right) < f_B^u + \frac{\mathcal{H}_{NB}(1-\mathcal{H}_B)}{2\Delta}\left(f_B^u - \frac{I}{\mathcal{H}_{NB}}\right) < c_{NB}$, where the last inequality is implied by our assumption on the lower bound of A. So

$$\left\{\begin{array}{c} \frac{I}{\mathcal{H}_{NB}} < f_B^c < c_{NB} \\ c_{NB} < S^u \\ \frac{I}{\mathcal{H}_{NB}} < f_{NB}^c < \frac{\mathcal{H}_B f_B^c - \mathcal{H}_{NB} f_{NB}^c}{\mathcal{H}_B - \mathcal{H}_{NB}} < \frac{c_{NB} - \mathcal{H}_B f_B^c}{1-\mathcal{H}_B} < A \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{c} \frac{I}{\mathcal{H}_{NB}} < f_B^c \\ c_{NB} < S^u \end{array}\right\} \Longleftrightarrow c_2 < c_B < c_3,$$

where

$$c_3 = I - \frac{2I}{\mathcal{H}_{NB}} + 2c_{NB}.$$

Then, when $c_B \in (c_2, c_3)$, the equilibrium is the constrained market-making equilibrium.

A.5 Proof of Proposition 4

Unconstrained Market-Making Equilibrium

We consider how the equilibrium changes when c_B increases in this region. In the unconstrained equilibrium, we have

$$f_{NB}^* = \frac{I\left(2\mathcal{H}_B + \mathcal{H}_{NB}\right) + A\mathcal{H}_{NB}\Delta}{4\mathcal{H}_{NB}\Delta}$$

$$f_B^* = \frac{3I + 2A\Delta}{4\mathcal{H}_B - \mathcal{H}_{NB}}$$

$$S^\star = \frac{c_B}{2} + \frac{I\left(2\mathcal{H}_B + \mathcal{H}_{NB}\right) + A\left(\Delta + 3\mathcal{H}_B\left(1 - \mathcal{H}_{NB}\right)\right)}{8\mathcal{H}_B - 2\mathcal{H}_{NB}}.$$

When c_B increases, both f_{NB}^* and f_B^* are unchanged, and S^* increases. Overall customer welfare decreases because for each customer all prices weakly increase. These results also imply that $x_1^* = \frac{\mathcal{H}_B f_B^* - \mathcal{H}_{NB} f_{NB}^*}{\Delta}$ is unchanged and $x_2^* = \frac{S^* - \mathcal{H}_B f_B^*}{1-\mathcal{H}_B}$ increases.

Matchmaking volume $(x_2^* - f_{NB}^*)$ increases because x_2^* increases and f_{NB}^* is unchanged. Non-bank dealer matchmaking volume $(x_1^* - f_{NB}^*)$ is unchanged because both x_1^* and f_{NB}^* are unchanged. Bank dealer matchmaking volume $(x_2^* - x_1^*)$ increases because x_2^* increases and x_1^* is unchanged. Market-making volume $(A - x_2^*)$ decreases because x_2^* increases.

Average transaction costs are

$$ATC = \frac{\left(x_1^* - f_{NB}^*\right) f_{NB}^* + \left(x_2^* - x_1^*\right) f_B^* + \left(A - x_2^*\right) S^*}{A - f_{NB}^*} \propto \left(x_2^* - x_1^*\right) f_B^\star + \left(A - x_2^*\right) S^*.$$

It can be shown that

$$\frac{d\left[\left(x_2^*-x_1^*\right)f_B^{\star}+\left(A-x_2^*\right)S^*\right]}{dc_B}=-\frac{c_B\left(3\mathcal{H}_B+\Delta\right)-\left(3+\Delta\right)I-2\left(1-\mathcal{H}_B\right)\Delta A}{2\left(1-\mathcal{H}_B\right)\left(3\mathcal{H}_B+\Delta\right)}.$$

Hence,

$$\frac{dATC}{dc_B}\propto-\left(c_B\left(3\mathcal{H}_B+\Delta\right)-\left(3+\Delta\right)I-2\left(1-\mathcal{H}_B\right)\Delta A\right).$$

and

$$\frac{dATC}{dc_B}>(<)0\Longleftrightarrow c_B<(>)\frac{\left(3+\Delta\right)I+2\left(1-\mathcal{H}_B\right)\Delta A}{3\mathcal{H}_B+\Delta}.$$

Constrained Market-Making Equilibrium

We consider how the equilibrium changes when c_B increases in this region. In the constrained equilibrium, we have

$$f_{NB}^*=\frac{1}{2}\left(\frac{-2c_B\Delta+\left(3-\mathcal{H}_B-2\mathcal{H}_{NB}\right)I+4c_{NB}\Delta}{\Delta+3\mathcal{H}_B\left(1-\mathcal{H}_{NB}\right)}+\frac{I}{\mathcal{H}_{NB}}\right),$$

$$f_B^*=\frac{-2c_B\Delta+\left(3-\mathcal{H}_B-2\mathcal{H}_{NB}\right)I+4c_{NB}\Delta}{\Delta+3\mathcal{H}_B\left(1-\mathcal{H}_{NB}\right)}$$

and

$$S^*=c_{NB}.$$

When c_B increases, both f_{NB}^* and f_B^* decrease, and S^* is unchanged. Overall customer welfare increases because all prices weakly decrease for all customers. The cutoff

$$x_1^*=\frac{\mathcal{H}_Bf_B-\mathcal{H}_{NB}f_{NB}}{\Delta}=\frac{\left(\mathcal{H}_B-\frac{\mathcal{H}_{NB}}{2}\right)f_B-\frac{1}{2}\mathcal{H}_{NB}I}{\Delta}$$

decreases and

$$x_2^*=\frac{S^*-\mathcal{H}_Bf_B^*}{1-\mathcal{H}_B}$$

increases because f_B^* decreases.

Therefore, total volume $(A-f_N^*)$ increases, non-bank dealer matchmaking volume

$$x_1^*-f_{NB}^*=\frac{\mathcal{H}_B}{\Delta}\frac{1}{2}\left(f_B^*-\frac{I}{\mathcal{H}_{NB}}\right)$$

decreases, and market-making volume

$$A-x_2^*=A-\frac{c_{NB}-\mathcal{H}_Bf_B^*}{1-\mathcal{H}_B}$$

decreases. Total matchmaking volume

$$x_2^*-f_{NB}^*=\left(A-f_{NB}^*\right)-\left(A-x_2^*\right)$$

increases and so does the bank dealer matchmaking volume

$$x_2^* - x_1^* = (A - f_{NB}^*) - (x_1^* - f_{NB}^*) - (A - x_2^*).$$

The change in average transaction costs

$$ATC = \frac{(x_1^* - f_{NB}^*) f_{NB}^* + (x_2^* - x_1^*) f_B^* + (A - x_2^*) S^*}{A - f_{NB}^*}$$

is non-monotone and depends on model parameters. Note that in equilibrium we have

$$f_{NB}^* = \frac{1}{2}\left(f_B^* + \frac{I}{\mathcal{H}_{NB}}\right) \Longleftrightarrow f_B^* = 2f_{NB}^* - \frac{I}{\mathcal{H}_{NB}},$$

thus

$$\begin{aligned} ATC &= \frac{(x_1^* - f_{NB}^*) f_{NB}^* + (x_2^* - x_1^*) f_B^* + (A - x_2^*) S^*}{A - f_{NB}^*} \\ &= \frac{\frac{\mathcal{H}_B}{\Delta}\left(f_{NB}^* - \frac{I}{\mathcal{H}_{NB}}\right) f_{NB}^* + \left(\frac{c_{NB} + \frac{\mathcal{H}_B}{\mathcal{H}_{NB}} I - 2\mathcal{H}_B f_{NB}^*}{1 - \mathcal{H}_B} - \frac{(2\mathcal{H}_B - \mathcal{H}_{NB}) f_{NB}^* - \frac{\mathcal{H}_B}{\mathcal{H}_{NB}} I}{\Delta}\right)\left(2f_{NB}^* - \frac{\mathcal{H}_B}{\mathcal{H}_{NB}} I\right) + \left(A - \frac{c_{NB} + \frac{\mathcal{H}_B}{\mathcal{H}_{NB}} I - 2\mathcal{H}_B f_{NB}^*}{1 - \mathcal{H}_B}\right) c_{NB}}{A - f_{NB}^*} \end{aligned}$$

The denominator of ATC is a linear function of f_{NB}^* and the numerator is a quadratic function of f_{NB}^*. Let

$$y = A - f_{NB}^*.$$

It can be shown that

$$ATC = a_2 y + a_1 + \frac{a_0}{y},$$

where

$$\begin{aligned} a_2 &= -\frac{\mathcal{H}_B^2 + \mathcal{H}_B(3 - 2\mathcal{H}_{NB}) - 2\mathcal{H}_{NB}}{4(1 - \mathcal{H}_B)^2 \Delta^2 \mathcal{H}_{NB}^2} \\ &= -\frac{\mathcal{H}_B \Delta + \mathcal{H}_B(1 - \mathcal{H}_{NB}) + 2\Delta}{4(1 - \mathcal{H}_B)^2 \Delta^2 \mathcal{H}_{NB}^2}, \end{aligned}$$

$$a_1 = -\frac{\begin{array}{c} 2c_{NB}(1 + \mathcal{H}_B)\Delta\mathcal{H}_{NB} + 3\mathcal{H}_B(1 - \mathcal{H}_{NB}) I - 2\mathcal{H}_B\mathcal{H}_{NB}(3 - 2\mathcal{H}_{NB}) A \\ + \mathcal{H}_B^2(I - 2\mathcal{H}_{NB}A) + \mathcal{H}_{NB}(4\mathcal{H}_{NB}A - I) \end{array}}{(1 - \mathcal{H}_B)\Delta\mathcal{H}_{NB}},$$

and

$$a_0 = -4\left(\begin{array}{c} A^2\mathcal{H}_{NB}^2\left(\mathcal{H}_B^2 + \mathcal{H}_B(3 - 2\mathcal{H}_{NB}) - 2\mathcal{H}_{NB}\right) \\ + A\mathcal{H}_{NB}\left(-c_{NB}(3 + \mathcal{H}_B)\Delta\mathcal{H}_{NB} + I\left(-\mathcal{H}_B^2 - 3\mathcal{H}_B(1 - \mathcal{H}_{NB}) + \mathcal{H}_{NB}\right)\right) \\ + \left(c_{NB}^2\Delta\mathcal{H}_{NB}^2 + c_{NB}(1 + \mathcal{H}_B)\Delta\mathcal{H}_{NB} I + \mathcal{H}_B(1 - \mathcal{H}_{NB}) I^2\right) \end{array}\right).$$

It is clear that $a_2 < 0$. We also show that $a_0 < 0$. Note that a_0 is a quadratic function of $A\mathcal{H}_{NB}$.

To show that $a_0 < 0$, we just need to verify that its discriminant is negative:

$$\begin{aligned} D = & \left(-c_{NB}\left(3+\mathcal{H}_B\right)\Delta\mathcal{H}_{NB} + I\left(-\mathcal{H}_B^2 - 3\mathcal{H}_B\left(1-\mathcal{H}_{NB}\right)+\mathcal{H}_{NB}\right)\right)^2 \\ & -4\left(\mathcal{H}_B^2 + \mathcal{H}_B\left(3-2\mathcal{H}_{NB}\right) - 2\mathcal{H}_{NB}\right)\left(c_{NB}^2\Delta\mathcal{H}_{NB}^2 + c_{NB}\left(1+\mathcal{H}_B\right)\Delta\mathcal{H}_{NB}I + \mathcal{H}_B\left(1-\mathcal{H}_{NB}\right)I^2\right) \\ = & -\left(1-\mathcal{H}_B\right)\left(\Delta\mathcal{H}_B\left(2-\mathcal{H}_{NB}+\mathcal{H}_B\right)+\left(\mathcal{H}_B^2-\mathcal{H}_{NB}^2\right)\right)\left(-c_{NB}\mathcal{H}_{NB}+I\right)^2 \\ < & \ 0. \end{aligned}$$

Then

$$\frac{dATC}{dc_B} \propto \frac{dATC}{dy} = a_2 - \frac{a_0}{y^2} = \frac{-a_2}{y^2}\left(\frac{a_0}{a_2} - y^2\right).$$

$$\begin{aligned} & y = A - f_{NB}^{\star} = A - \frac{1}{2}\left(\frac{-2c_B\Delta + \left(3-\mathcal{H}_B-2\mathcal{H}_{NB}\right)I + 4c_{NB}\Delta}{\Delta + 3\mathcal{H}_B\left(1-\mathcal{H}_{NB}\right)} + \frac{I}{\mathcal{H}_{NB}}\right) < \sqrt{\frac{a_0}{a_2}} \\ \Longleftrightarrow & c_B < c_4 \end{aligned}$$

where

$$c_4 = \left(\sqrt{\frac{a_0}{a_2}} - A + \frac{I}{2\mathcal{H}_{NB}}\right)\left(\frac{\Delta + 3\mathcal{H}_B\left(1-\mathcal{H}_{NB}\right)}{\Delta}\right) - 2c_{NB} - \frac{\left(3-\mathcal{H}_B-2\mathcal{H}_{NB}\right)I}{2\Delta}.$$

So, when $c < (>) c_4$, $\frac{dATC}{dc_B} > (<) 0$.

A.6 Proof of Proposition 5

The non-bank dealer's FOC is

$$\frac{r}{2\mu}\frac{\partial \Pi_{NB}}{\partial q_{NB}} = -c_{NB} + \left(1 - q_B - \mathcal{H}q_M - 2q_{NB}\right)A = 0.$$

The bank dealer's FOCs are

$$\frac{r}{2\mu}\frac{\partial \Pi_B}{\partial q_B} = -c_B + A\left(1 - 2q_B - q_{NB} - 2\mathcal{H}q_M\right) = 0$$

and

$$\frac{r}{2\mu}\frac{\partial \Pi_B}{\partial q_M} = -I + \mathcal{H}\left(1 - 2q_B - 2q_M - q_{NB}\right)A = 0.$$

There is a unique solution (q_B^u, q_{NB}^u, q_M^u) for the above three FOCs:

$$\begin{aligned} q_M^u &= \frac{c_B\mathcal{H} - I}{2\left(1-\mathcal{H}\right)\mathcal{H}A} \\ q_{NB}^u &= \frac{c_B - 2c_{NB} + A}{3A} \\ q_B^u &= \frac{c_{NB} - 2c_B + A}{3A} - \frac{\mathcal{H}c_B - I}{2\left(1-\mathcal{H}\right)A}. \end{aligned}$$

Lemma 2. *(q_B^u, q_{NB}^u, q_M^u) is an equilibrium if and only if*

$$q_B^u, q_{NB}^u, q_M^u > 0.$$

Proof. (q_B^u, q_{NB}^u, q_M^u) is an equilibrium if $q_B^u, q_{NB}^u, q_M^u > 0$ and $q_B^u + q_{NB}^u + q_M^u < 1$. From $\frac{r}{2\mu}\frac{\partial \Pi_B}{\partial q_M} = -I + \mathcal{H}(1 - 2q_B - 2q_M - q_{NB})A = 0$, we know that $q_B^u, q_{NB}^u, q_M^u > 0$ implies $q_B^u + q_{NB}^u + q_M^u < 1$. So the condition $q_B^* + q_{NB}^* + q_M^* < 1$ is redundant. □

If in equilibrium the bank dealer is active in matchmaking and market making, and the non-bank dealer is active in market making, the equilibrium $(S^\star, f^\star)$ are

$$\begin{aligned} f^* &= A\left(1 - q_M^u - q_{NB}^u - q_B^u\right) \\ &= \frac{1}{3}\left(c_B + c_{NB} + A\right) - \frac{c_B\mathcal{H} - I}{2\mathcal{H}} \end{aligned}$$

and

$$\begin{aligned} S^* &= A\left(1 - \mathcal{H}q_M^u - q_{NB}^u - q_B^u\right) \\ &= \frac{1}{3}\left(c_B + c_{NB} + A\right). \end{aligned}$$

It can be shown that the following conditions hold:

$$q_B^u > 0 \iff c_B \leq \frac{2c_{NB}(1-\mathcal{H}) + 3I + 2(1-\mathcal{H})A}{4-\mathcal{H}},$$

$$q_{NB}^u > 0 \iff c_B > 2c_{NB} - A,$$

$$q_M^u > 0 \iff c_B > \frac{I}{\mathcal{H}}.$$

So

$$\left\{\begin{array}{c} q_M^* = q_M^u > 0 \\ q_{NB}^* = q_{NB}^u > 0 \\ q_B^* = q_B^u > 0 \end{array}\right\} \iff \max\left\{\frac{I}{\mathcal{H}}, 2c_{NB} - A\right\} < c_B < \frac{2(1-\mathcal{H})(c_{NB}+A) + 3I}{4-\mathcal{H}}.$$

A.7 Proof of Proposition 6

When c_B increases, f^* decreases and S^* increases. Note that both q_M^* and q_{NB}^* increase, and q_B^* decreases. Total market-making volume is

$$q_B^* + q_{NB}^* = \frac{-2c_{NB}(1-\mathcal{H}) - c_B(2+\mathcal{H}) + 3I + 4(1-\mathcal{H})A}{6(1-\mathcal{H})A},$$

which decreases when c_B increases. The cutoff is

$$\begin{aligned} x^* &= \frac{S^* - \mathcal{H}f^*}{1-\mathcal{H}} \\ &= \frac{2(1-\mathcal{H})c_{NB} + c_B(2+\mathcal{H}) - 3I + 2(1-\mathcal{H})A}{6(1-\mathcal{H})}. \end{aligned}$$

Overall customer welfare is

$$\begin{aligned}
\pi_c &= \frac{2\mu}{r}\left[\int_{f^*}^{x^*}\mathcal{H}(x-f^*)\frac{1}{A}dx+\int_{x^*}^{A}(x-S^*)\frac{1}{A}dx\right]\\
&= \frac{\mu}{r}\frac{1}{A}\left[\mathcal{H}(x^*-f^*)^2+(A-S^*)^2-(x^*-S^*)^2\right]\\
&= \frac{\mu}{r}\frac{1}{A}\left[\frac{(\mathcal{H}c_B-I)^2}{4\mathcal{H}(1-\mathcal{H})^2}+\frac{1}{9}(2A-c_B-c_{NB})^2-\frac{(\mathcal{H}c_B-I)^2}{4(1-\mathcal{H})^2}\right]\\
&= \frac{\mu}{r}\frac{1}{A}\left[\frac{(\mathcal{H}c_B-I)^2}{4\mathcal{H}(1-\mathcal{H})}+\frac{1}{9}(2A-c_B-c_{NB})^2\right]
\end{aligned}$$

This is a quadratic function of c_B. The first-order derivative

$$\frac{\partial CS}{\partial c_B}=\frac{\mu}{r}\frac{1}{A}\left[\frac{(\mathcal{H}c_B-I)}{2(1-\mathcal{H})}-\frac{2}{9}(2A-c_B-c_{NB})\right]$$

is positive if and only if

$$c_B>\frac{9I+4(1-\mathcal{H})(2A-c_{NB})}{4+5\mathcal{H}}.$$

Average transaction costs are

$$\begin{aligned}
ATC &= \frac{q_M^*f^*+(q_B^*+q_{NB}^*)S^*}{q_B^*+q_M^*+q_{NB}^*}\\
&= -\frac{\begin{array}{c}3(c_B\mathcal{H}-I)(c_B\mathcal{H}-2c_{NB}\mathcal{H}-3I-2\mathcal{H}A)\\ +2\mathcal{H}^2(c_B+c_{NB}+A)(2c_{NB}(1-\mathcal{H})+c_B(2+\mathcal{H})-3I-4(1-\mathcal{H})A)\end{array}}{6\mathcal{H}(1-\mathcal{H})(c_B\mathcal{H}-2c_{NB}\mathcal{H}-3I+4\mathcal{H}A)}.
\end{aligned}$$

Let

$$z=c_B\mathcal{H}-2c_{NB}\mathcal{H}-3I+4\mathcal{H}A,$$

we have

$$ATC=-\frac{a_zz^2+b_zz+c_z}{6\mathcal{H}(1-\mathcal{H})z},$$

where

$$a_z=7+2\mathcal{H}$$

$$b_z=6((c_{NB}\mathcal{H}+I)(5+\mathcal{H})-\mathcal{H}(11+\mathcal{H})A)$$

and

$$c_z=36(c_{NB}\mathcal{H}+I-2\mathcal{H}A)^2.$$

Then,

$$\frac{dATC}{dc_B}\propto\frac{dATC}{dz}=\frac{a_z}{6\mathcal{H}(1-\mathcal{H})z^2}\left(\frac{c_z}{a_z}-z^2\right).$$

It is clear that both a_z and c_z are positive. We have

$$\frac{dATC}{dc_B} > (<) 0 \Longleftrightarrow \frac{c_z}{a_z} > (<) z^2 \Longleftrightarrow z < (>) \sqrt{\frac{c_z}{a_z}} = 6\frac{2\mathcal{H}A - c_{NB}\mathcal{H} - I}{\sqrt{7+2\mathcal{H}}}$$

and

$$\begin{aligned} & c_B\mathcal{H} - 2c_{NB}\mathcal{H} - 3I + 4\mathcal{H}A < (>) 6\frac{2\mathcal{H}A - c_{NB}\mathcal{H} - I}{\sqrt{7+2\mathcal{H}}} \\ \Longleftrightarrow & c_B < (>) \hat{c} = 2\left(\frac{3}{\sqrt{7+2\mathcal{H}}} - 1\right)(A - c_{NB}) + 2\left(\frac{3}{\sqrt{7+2\mathcal{H}}} - 1\right) A + 3\left(1 - \frac{2}{\sqrt{7+2\mathcal{H}}}\right)\frac{I}{\mathcal{H}}. \end{aligned}$$

A.8 Proof of Proposition 7

In equilibrium, customers with private value $x > A\left(1 - Q_{NB} - Q_B\right)$ choose market making, and customers with private value $x \in \left[A\left(1 - Q_M - Q_{NB} - Q_B\right), A\left(1 - Q_{NB} - Q_B\right)\right]$ choose matchmaking. The equilibrium (f, S) must satisfy

$$f = A\left(1 - Q_M - Q_{NB} - Q_B\right),$$

and

$$\mathcal{H}\left(A\left(1 - Q_{NB} - Q_B\right) - f\right) = A\left(1 - Q_{NB} - Q_B\right) - S \Longleftrightarrow S = A\left(1 - Q_{NB} - Q_B - \mathcal{H}Q_M\right).$$

Then, non-bank dealer k's profit is

$$\begin{aligned} \Pi_{NB}^k &= \frac{2\mu}{r}\left(S \cdot q_{NB}^k - \beta c_{NB} q_{NB}^k\right) \\ &= \frac{2\mu}{r}\left(A\left(1 - Q_{NB} - Q_B - \mathcal{H}Q_M\right) \cdot q_{NB}^k - \beta c_{NB} q_{NB}^k\right), \end{aligned}$$

and bank dealer j's profit is

$$\begin{aligned} \Pi_B^j &= \frac{2\mu}{r}\left(\left(\mathcal{H}f - I\right) q_M^j + S \cdot q_B^j - \lambda c_B q_B^j\right) \\ &= \frac{2\mu}{r}\left(\left(\mathcal{H}A\left(1 - Q_{NB} - Q_B - Q_M\right) - I\right) q_M^j + A\left(1 - Q_{NB} - Q_B - \mathcal{H}Q_M\right) \cdot q_B^j - \lambda c_B q_B^j\right). \end{aligned}$$

Non-bank k's FOC is

$$\frac{r}{2\mu}\frac{\partial \Pi_{NB}^k}{\partial q_{NB}^k} = -\beta c_{NB} + \left(-q_{NB}^k + 1 - Q_{NB} - Q_B - \mathcal{H}Q_M\right) A = 0. \quad (39)$$

Bank dealer j's FOCs are

$$\frac{r}{2\mu}\frac{\partial \Pi_B^j}{\partial q_B^j} = -\lambda c_B - \mathcal{H}A q_M^j - A q_B^j + A\left(1 - Q_{NB} - Q_B - \mathcal{H}Q_M\right) = 0 \quad (40)$$

and

$$\frac{r}{2\mu}\frac{\partial \Pi_B^j}{\partial q_M^j} = -\mathcal{H}Aq_M^j + \left(\mathcal{H}A\left(1 - Q_{NB} - Q_B - Q_M\right) - I\right) - \mathcal{H}Aq_B^j = 0. \tag{41}$$

FOCs (39), (40), and (41) imply that

$$-K\beta c_{NB} - Q_{NB}A + M\left(1 - Q_{NB} - Q_B - \mathcal{H}Q_M\right)A = 0,$$

$$-J\lambda c_B - \mathcal{H}AQ_M - AQ_B + AJ\left(1 - Q_{NB} - Q_B - \mathcal{H}Q_M\right) = 0,$$

and

$$-\mathcal{H}AQ_M + J\left(\mathcal{H}A\left(1 - Q_{NB} - Q_B - Q_M\right) - I\right) - \mathcal{H}AQ_B = 0.$$

The unique solution to the linear system is

$$Q_{NB}^u = \frac{-K\beta \cdot c_{NB}\left(1+J\right) + K\left(J\lambda \cdot c_B + A\right)}{\left(1+K+J\right)A}$$

$$Q_B^u = \frac{-J\lambda \cdot c_B\left(1 + J + K\left(1 + J - \mathcal{H}J\right)\right) + J\left(K\beta \cdot c_{NB}\left(1-\mathcal{H}\right)\left(1+J\right) + I\left(1+J+K\right) + \left(1-\mathcal{H}\right)\left(1+J\right)A\right)}{\left(1-\mathcal{H}\right)\left(1+J\right)\left(1+K+J\right)A}$$

$$Q_M^u = \frac{J\lambda \cdot c_B\mathcal{H} + J \cdot I}{\left(1-\mathcal{H}\right)\mathcal{H}\left(1+J\right)A}.$$

The prices in equilibrium are

$$\begin{aligned} S^u &= A\left(1 - Q_{NB}^u - Q_B^u - \mathcal{H}Q_M^u\right) \\ &= \frac{J\lambda \cdot c_B + K\beta \cdot c_{NB} + A}{1+J+K} \end{aligned}$$

$$\begin{aligned} f^u &= A\left(1 - Q_M - Q_{NB} - Q_B\right) \\ &= S^u - \frac{J\lambda \cdot c_B\mathcal{H} - I \cdot J}{\left(1+J\right)\mathcal{H}} \\ &= \frac{-J\lambda \cdot c_B\mathcal{H}K + I \cdot J\left(1+J+K\right) + K\beta \cdot c_{NB}\mathcal{H}\left(1+J\right) + \mathcal{H}A\left(1+J\right)}{\mathcal{H}\left(1+J\right)\left(1+J+K\right)}. \end{aligned}$$

The equilibrium strategy of bank dealer j is

$$q_B^{j,u} = \frac{-\lambda c_B - \left(\mathcal{H}f^u - I\right) + S^u}{\left(1-\mathcal{H}\right)A} \tag{42}$$

$$q_M^{j,u} = \frac{\lambda c_B\mathcal{H} + \mathcal{H}f^u - I - \mathcal{H}S^u}{\mathcal{H}\left(1-\mathcal{H}\right)A}, \tag{43}$$

and the strategy of non-bank dealer k is

$$q_{NB}^{k,u} = \frac{-\beta c_{NB} + S^u}{A}. \tag{44}$$

We need to provide parameter restrictions such that the solution solved above is indeed an equilibrium.

Lemma 3. $\left\{q_M^{j,u}, q_B^{j,u}, q_{NB}^{k,u}\right\}_{j,k}$ *is an equilibrium if and only if*

$$q_M^{j,u}, q_B^{j,u}, q_{NB}^{k,u} > 0$$

for all j, k.

Proof. (q_B^u, q_{NB}^u, q_M^u) is an equilibrium if $q_M^{j,u}, q_B^{j,u}, q_{NB}^{k,u} > 0$ for all j, k, and $Q_B^u + Q_{NB}^u + Q_M^u < 1$. From (41), we know that

$$\begin{aligned} & -\mathcal{H}Aq_M^j + (\mathcal{H}A(1 - Q_{NB} - Q_B - Q_M) - I) - \mathcal{H}Aq_B^j = 0 \\ \Longleftrightarrow & \mathcal{H}A(1 - Q_{NB} - Q_B - Q_M) = \mathcal{H}Aq_M^j + I + \mathcal{H}Aq_B^j. \end{aligned}$$

When $q_M^{j,u}, q_B^{j,u}, q_{NB}^{k,u} > 0$ for all j, k, this implies $Q_B^u + Q_{NB}^u + Q_M^u < 1$. So the condition $Q_B^u + Q_{NB}^u + Q_M^u < 1$ is redundant. □

From (42), (43), and (44), the condition in Lemma 3 is equivalent to

$$S^u \geq c_{NB} \cdot \beta \tag{45}$$

and

$$S^u - c_B \cdot \lambda \geq \mathcal{H}f^u - I \geq \mathcal{H}(S^u - c_B \cdot \lambda). \tag{46}$$

So, when conditions (45) and (46) are satisfied, the equilibrium is

$$\begin{aligned} S^* &= \frac{J\lambda \cdot c_B + K\beta \cdot c_{NB} + A}{1 + J + K} \\ &= \frac{\Lambda \cdot c_B + B \cdot c_{NB} + A}{1 + J + K} \end{aligned}$$

$$f^* = \frac{-\Lambda \cdot c_B \mathcal{H} K + I \cdot J(1 + J + K) + B \cdot c_{NB}\mathcal{H}(1 + J) + \mathcal{H}A(1 + J)}{\mathcal{H}(1 + J)(1 + J + K)}.$$

A.9 Proof of Proposition 8

When c_B increases, $f^\star$ decreases and $S^\star$ increases. Overall customer welfare is

$$\begin{aligned} \pi_c &= \frac{2\mu}{r}\left[\int_{f^*}^{x^*} \mathcal{H}(x - f^*)\frac{1}{A}dx + \int_{x^\star}^{A}(x - S^*)\frac{1}{A}dx\right] \\ &= \frac{\mu}{r}\frac{1}{A}\left[\mathcal{H}(x^* - f^*)^2 + (A - S^*)^2 - (x^* - S^*)^2\right] \\ &= \frac{\mu}{r}\frac{1}{A}\left[\frac{\begin{array}{c}(1 + J + K)^2(\Lambda c_B \mathcal{H} - I \cdot J)^2 - \mathcal{H}(1 + J + K)^2(\Lambda \cdot c_B \mathcal{H} - J \cdot I)^2 \\ + (1 - \mathcal{H})^2 \mathcal{H}(1 + J)^2(\Lambda c_B + B c_{NB} - (J + K)A)^2\end{array}}{(1 - \mathcal{H})^2 \mathcal{H}(1 + J)^2(1 + J + K)^2}\right] \end{aligned}$$

The first-order derivative with respect to c_B

$$\frac{\partial \pi_c}{\partial c_B} = \frac{\mu}{r}\frac{2\Lambda}{A}\left[\frac{\Lambda\left((1+J)^2 + \mathcal{H}M(2+K+2J)\right)}{(1-\mathcal{H})(1+J)^2(1+J+K)^2}c_B - \frac{I\cdot J}{(1-\mathcal{H})(1+J)^2} - \frac{(J+K)A - B\cdot c_{NB}}{(1+K+J)^2}\right].$$

is positive if and only if

$$c_B > \frac{I\cdot J(1+J+K)^2 + (1-\mathcal{H})(1+J)^2((J+K)A - B\cdot c_{NB})}{\Lambda\left((1+J)^2 + \mathcal{H}K(2+K+2J)\right)}.$$

The cutoff

$$\begin{aligned} x^* &= \frac{S^* - \mathcal{H}f^*}{1-\mathcal{H}} \\ &= \frac{-I\cdot J(1+J+K) + Bc_{NB}(1-\mathcal{H})(1+J) + \Lambda c_B(1+\mathcal{H}K+J) + (1-\mathcal{H})(1+J)A}{(1-\mathcal{H})(1+J)(1+J+K)} \end{aligned}$$

increases in c_B.

For volume, both Q_M^* and Q_{NB}^* increase, and Q_B^* decreases. Market-making volume is

$$Q_B^* + Q_{NB}^* = \frac{-\Lambda(1+\mathcal{H}K+J)}{(1+J)(1-\mathcal{H})(1+J+K)A}c_B + \text{constant}(\perp c_B)$$

which decreases when c_B increases.

Average transaction costs are

$$\begin{aligned} ATC &= \frac{Q_M^* f^* + (Q_B^* + Q_{NB}^*)S^*}{Q_B^\star + Q_M^\star + Q_{NB}^\star} \\ &= \frac{\mathcal{H}(1+J)\left(\frac{(-\Lambda c_B\mathcal{H}+I\cdot J)(-\Lambda c_B\mathcal{H}K + I\cdot J(1+J+K) + Bc_{NB}\mathcal{H}(1+J) + \mathcal{H}A(1+J))}{-(1-\mathcal{H})\mathcal{H}^2(1+J)^2} + \frac{(\Lambda c_B + Bc_{NB} + A)\left(-Bc_{NB}(1+J) - \frac{\Lambda c_{NB}(1+J+K(1+J-\mathcal{H}J)) + J\left(-\left(Bc_{NB}+A\right)(1-\mathcal{H})(1+J) - I(1+J+K)\right)}{(1-\mathcal{H})(1+J)}\right)}{(1+J+K)}\right)}{\Lambda c_B\mathcal{H}K - IJ(1+J+K) - Bc_{NB}\mathcal{H}(1+J) + A\mathcal{H}(J+K)(1+J)}. \end{aligned}$$

Let

$$z = \Lambda c_B\mathcal{H}K - IJ(1+J+K) - Bc_{NB}\mathcal{H}(1+J) + A\mathcal{H}(J+K)(1+J).$$

Then,

$$ATC = \frac{a_z z^2 + b_z z + c_z}{z},$$

where

$$a_z = -\frac{K^2 + (1+\mathcal{H})K(1+J) + (1+J)^2}{(1-\mathcal{H})\mathcal{H}K^2(1+J)(1+K+J)}$$

$$b_z = -\frac{(Bc_{NB}\mathcal{H} + I\cdot J)(2+K+\mathcal{H}K+2J) - \mathcal{H}A\left(2K^2 + 2J(1+J) + K(2+(3+\mathcal{H})J)\right)}{(1-\mathcal{H})\mathcal{H}M^2}$$

and

$$c_z = -\frac{(1+J)(1+J+K)\left(Bc_{NB}\mathcal{H} + I\cdot K - \mathcal{H}(K+J)A\right)^2}{(1-\mathcal{H})\mathcal{H}K^2}.$$

Then,

$$\frac{dATC}{dc_B} \propto \frac{dATC}{dz} = \frac{-a_z}{z^2}\left(\frac{c_z}{a_z} - z^2\right).$$

Since both a_z and c_z are negative, we have

$$\frac{dATC}{dc_B} > (<) 0 \iff \frac{c_z}{a_z} > (<) z^2 \iff z < (>) \sqrt{\frac{c_z}{a_z}} = \frac{(1+J)(1+J+K)\left(J(\mathcal{H}A - I) + \mathcal{H}(KA - Bc_{NB})\right)}{\sqrt{K^2 + (1+\mathcal{H})K(1+J) + (1+J)^2}}$$

and

$$\Lambda c_B \mathcal{H} K - IJ(1+J+K) - Bc_{NB}\mathcal{H}(1+J) + A\mathcal{H}(J+K)(1+J) < (>) \frac{(1+J)(1+J+K)\left(J(\mathcal{H}A - I) + \mathcal{H}(KA - Bc_{NB})\right)}{\sqrt{K^2 + (1+\mathcal{H})K(1+J) + (1+J)^2}}$$

$$\iff c_B < (>) \frac{IJ(1+J+K) + Bc_{NB}\mathcal{H}(1+J) - A\mathcal{H}(J+K)(1+J)}{\Lambda \mathcal{H} K} + \frac{(1+J)(1+J+K)\left(J(\mathcal{H}A - I) + \mathcal{H}(KA - Bc_{NB})\right)}{\Lambda \mathcal{H} K \sqrt{K^2 + (1+\mathcal{H})K(1+J) + (1+J)^2}}.$$

References

Adrian, T., M. Fleming, O. Shachar, and E. Vogt. 2017. Market Liquidity After the Financial Crisis. *Annual Review of Financial Economics* 9:43–83.

Allahrakha, M., J. Centina, B. Munyan, and S. Watugala. 2019. The Effects of the Volcker Rule on Corporate Bond Trading: Evidence from the Underwriting Exemption. Working Paper.

An, Y., Y. Song, and X. Zhang. 2017. The Intermediary Rat Race. Working Paper.

An, Y., and Z. Zheng. 2017. Conflicted Immediacy Provision. Working Paper.

Anderson, M., and R. Stulz. 2017. Is Post-Crisis Bond Liquidity Lower? Working Paper.

Bagnoli, M., and T. Bergstrom. 2005. Log-concave probability and its applications. *Economic Theory* 26:445–469.

Bao, J., M. O'Hara, and X. (Alex) Zhou. 2018. The Volcker Rule and corporate bond market making in times of stress. *Journal of Financial Economics* 130:95–113.

Bessembinder, H., S. Jacobsen, W. Maxwell, and K. Venkataraman. 2018. Capital Commitment and Illiquidity in Corporate Bonds. *Journal of Finance* 73:1615–61.

BIS Committee on the Global Financial System. 2014. Market-making and proprietary trading: industry trends, drivers and policy implications. Report.

———. 2016. Fixed income market liquidity. Report.

BIS Markets Committee. 2016. Electronic trading in fixed income markets. Report.

Bulow, J., and P. Klemperer. 2002. Prices and the Winner's Curse. *The RAND Journal of Economics* 33:1–21.

Chang, B., and S. Zhang. 2020. OTC Market Structure under Reforms. Working Paper.

Chernenko, S., and A. Sunderam. 2018. Corporate Bond Liquidity: A Revealed Preference Approach. Working Paper, Harvard Business School.

Choi, J., and Y. Huh. 2017. Customer Liquidity Provision: Implications for Corporate Bond Transaction Costs. Working Paper.

Cimon, D. A., and C. Garriott. 2019. Banking Regulation and Market Making. Working Paper.

Dick-Nielsen, J., and M. Rossi. 2018. The Cost of Immediacy for Corporate Bonds. *The Review of Financial Studies* 32:1–41.

Duffie, D., N. Gârleanu, and L. H. Pedersen. 2005. Over-the-Counter Markets. *Econometrica* 73:1815–47.

Fudenberg, D., and J. Tirole. 1991. *Game Theory*. MIT Press.

Johnson, J.P., and D.P. Myatt. 2013. Multiproduct Quality Competition: Fighting Brands and Product Line Pruning. *The American Economic Review* 93:748–774.

Kargar, M., B. Lester, D. Lindsay, S. Liu, P. Weill, and D. Zuniga. 2021. Corporate Bond Liquidity during the COVID-19 Crisis. *Review of Financial Studies* 34:5352–5401.

Li, J., and W. Li. 2017. Agency Trading and Principal Trading. Working Paper.

Katz. M. 1984. Firm-Specific Differentiation and Competition Among Multiproduct Firms. *Journal of Business* 57:149–166.

Mussa, M., and S. Rosen. 1978. Monopoly and Product Quality. *Journal of Economic Theory* 18:301–317.

Mizrach, B. 2015. Analysis of Corporate Bond Liquidity. Working Paper.

Nocke, V., and N. Schutz. 2018. Multiproduct-Firm Oligopoly: An Aggregative Games Approach. *Econometrica* 86:523–557.

Schultz, P. 2001. Corporate bond trading costs: A peek behind the curtain. *Journal of Finance* 56:677–98.

———. 2017. Inventory Management by Corporate Bond Dealers. Working Paper.

Trebbi, F., and K. Xiao. 2019. Regulation and Market Liquidity. *Management Science* 65:1949–1968.

Weyl, G., and M. Fabinger. 2013. Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition. *Journal of Political Economy* 121:528–583.

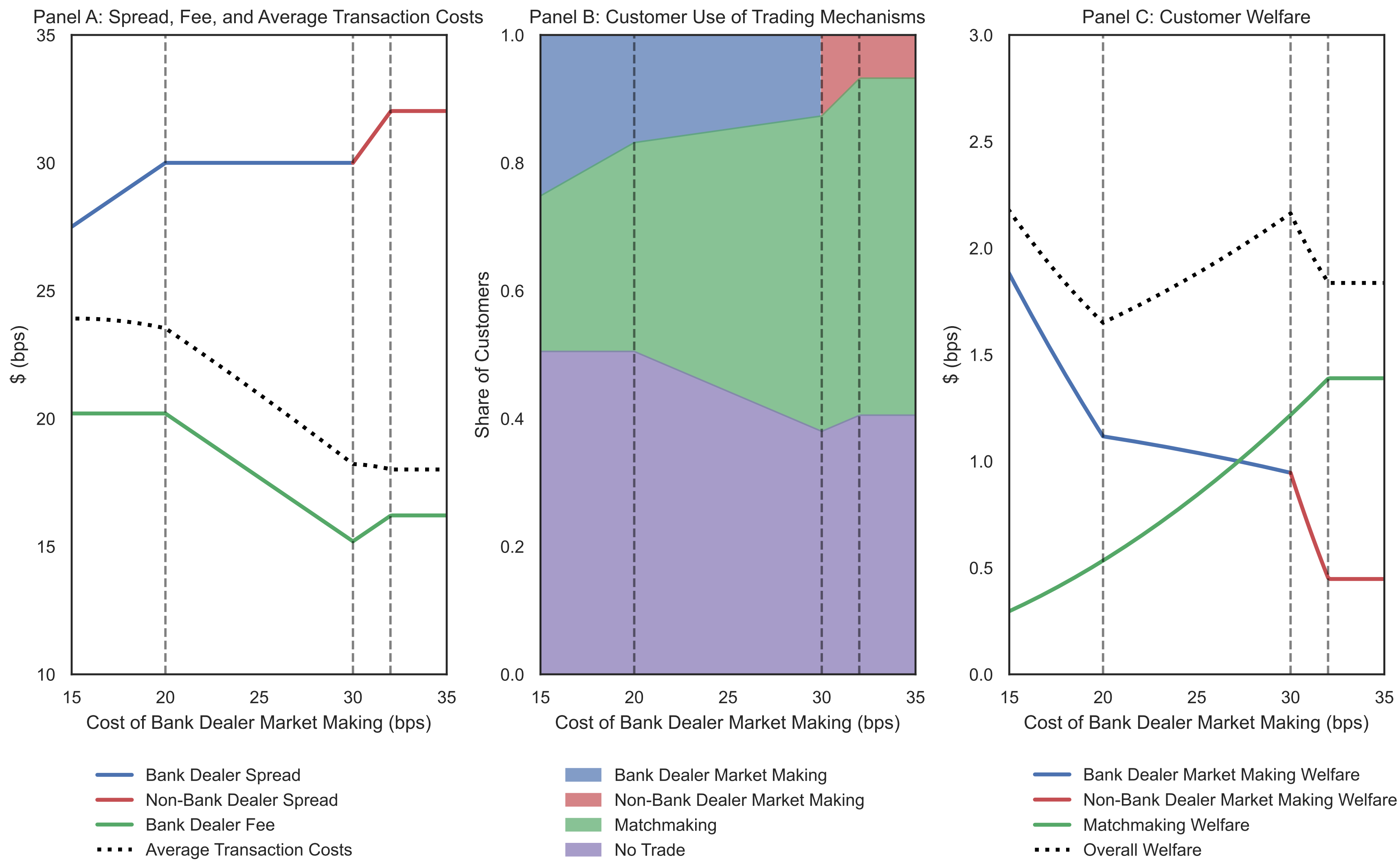

Figure 1: **Main Model.** *Panel A presents the market-making spread, the matchmaking fee, and average transaction costs. Panel B plots the fractions of traders choosing market making, matchmaking, or refraining from trade. Panel C depicts overall customer welfare and, separately, its market-making and matchmaking components. The dashed vertical lines in each panel indicate the transitions from one equilibrium region to another. The parameters used in the numerical example are* $c_{NB} = 30$ *basis points,* $A = 40$ *basis points,* $I = 0.10$ *basis points, and* $\mathcal{H} = 0.25$.

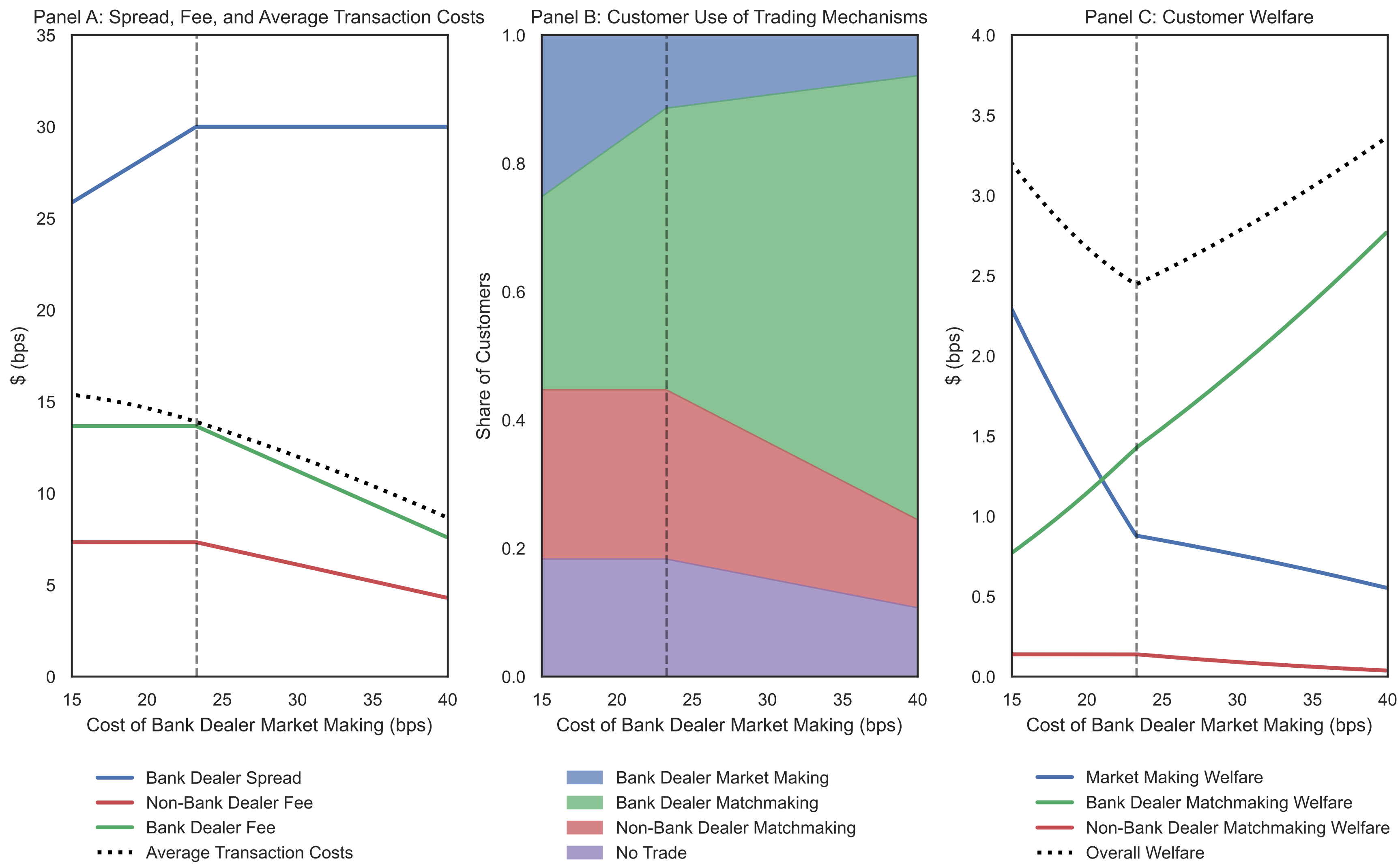

Figure 2: **Non-Bank Dealer Matchmaking Service.** *Panel A presents the market-making spread, the matchmaking fees, and average transaction costs. Panel B plots the fractions of traders choosing market making, bank dealer matchmaking, non-bank dealer matchmaking, or refraining from trade. Panel C depicts overall customer welfare and, separately, its market making and matchmaking components. The dashed vertical line in each panel indicates the transition from one equilibrium region to another. The parameters used in the numerical example are* $c_{NB} = 30$ *basis points,* $A = 40$ *basis points,* $I = 0.10$ *basis points,* $\mathcal{H}_B = 0.25$*, and* $\mathcal{H}_{NB} = 0.10$.

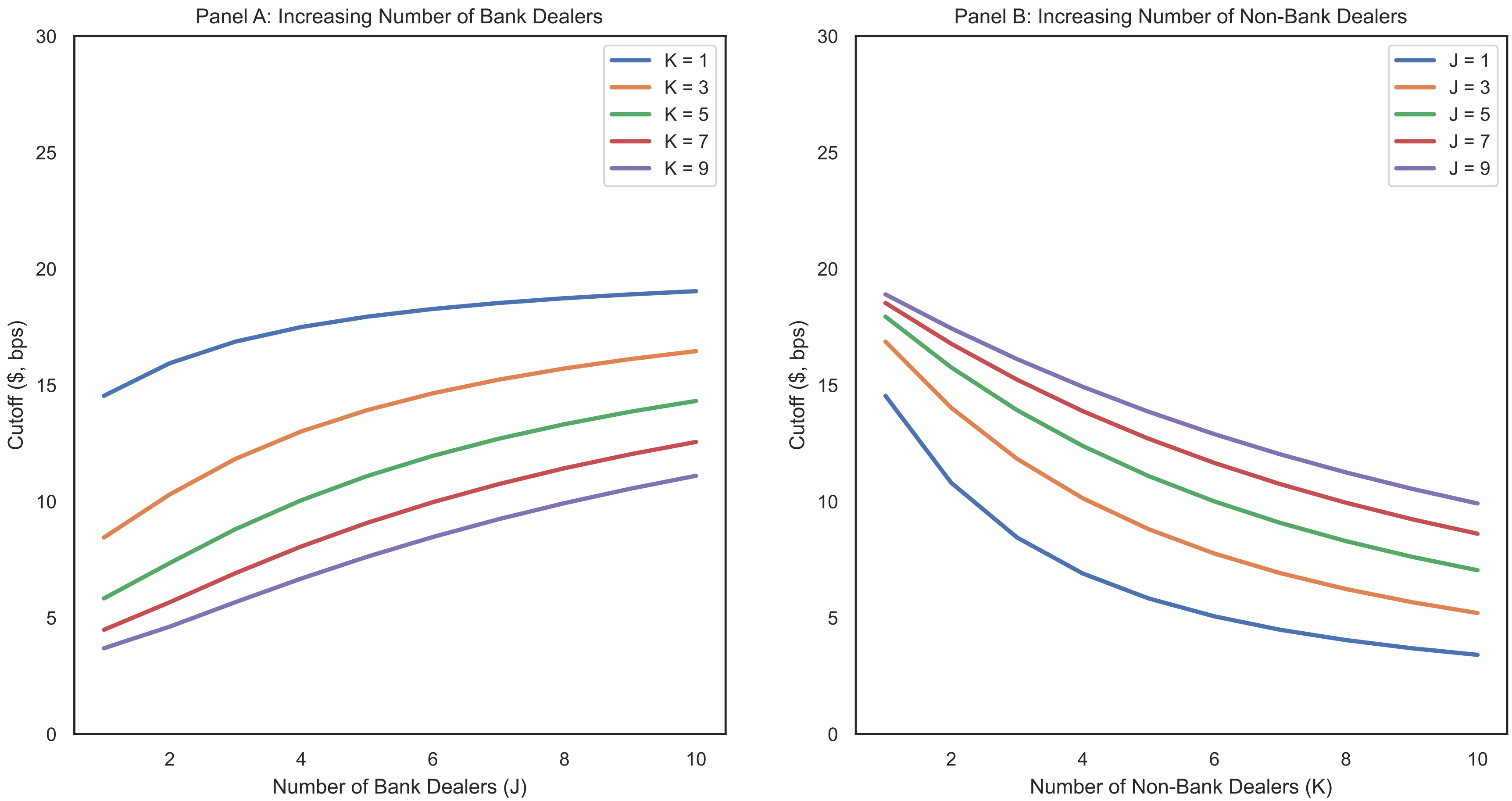

Figure 3: **Multiple Bank and Non-Bank Dealers.** *This figure shows how the cutoff above which an increase in bank regulatory costs improves welfare changes with the number of dealers. In Panel A, the y-axis is this cutoff, while the x-axis is the number of bank dealers. Each line in the plot represents an economy with a different number of non-bank dealers. In Panel B, the y-axis is the cutoff, while the x-axis is the number of non-bank dealers. Each line in the plot represents an economy with a different number of bank dealers. The parameters used in the numerical example are* $c_{NB} = 35$ *basis points,* $A = 40$ *basis points,* $I = 0.50$ *basis points,* $\mathcal{H} = 0.50$*, and* $\lambda = \beta = 1$.