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

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

This version: September 2019

Abstract

Post-crisis bank regulations raised the market-making costs of bank-affiliated dealers. We show that this can, somewhat surprisingly, improve the overall welfare of investors 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 intensify competitive pressure from non-bank dealers and incentivize bank dealers to invest in technology that shifts their business toward matchmaking. Thus, post-crisis bank regulations have the (unintended) benefit of replacing the costly balance sheet of banks 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

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

*Gideon Saar is from the Johnson Graduate School of Management, Cornell University (gs25@cornell.edu). Jian Sun is from MIT's Sloan School of Management (jiansun@mit.edu). Ron Yang is from Harvard Business School (rnyang@g.harvard.edu). Haoxiang Zhu is from MIT's Sloan School of Management and the NBER (zhuh@mit.edu). We thank Philip Bond, Christopher Hrdlicka, Stacey Jacobsen, Jonah Platt, Dimitri Vayanos, Kumar Venkataraman, Yao Zeng, Zhuo Zhong, and Hao Zhou as well as seminar participants at the Chinese University of Hong Kong, the Chinese University of Hong Kong Shenzhen, the City University of Hong Kong, Manchester Business School, MIT, the PBC School of Finance, University at Buffalo, the University of British Columbia, the University of Illinois, the University of Washington, Warwick Business School, Wharton, and the 4th Sydney Market Microstructure Meeting for helpful comments.

1 Introduction

The aftermath of the financial crisis saw several regulatory initiatives to curtail banks' risk-taking desire, hindering proprietary trading and increasing the cost of market making. In part, these reflect a wide-spread 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 they were meant to improve market-maker resiliency, the post-crisis changes to 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. Some market observers have portrayed these regulatory initiatives as introducing a trade-off between resiliency (or lessening contagion) in times of stress and liquidity during normal times.

Such a perspective, however, overlooks the microstructure aspects of liquidity. In particular, the over-the-counter corporate bond market features two parallel trading mechanisms corresponding to the dual capacity of broker-dealers. In the first mechanism, market making, a bank intermediary functions as a dealer that provides immediacy to customers by taking bonds onto his balance sheet. In the second mechanism, matchmaking, a bank intermediary functions as a broker that searches for couterparties to affect customer trades.[1] While regulations aimed at boosting resiliency have increased the cost of taking a bond onto a bank's balance sheet, they did not increase the costs associated with the search process. In fact, technological innovations over the past several years have bolstered the efficiency and reduced the costs of matching, stimulating investment in matchmaking technology by bank dealers and attracting customers to the search process. Our focus in this paper is therefore not on the trade-off between resiliency in times of stress and liquidity during normal times. Rather, we investigate whether the regulatory push to increase resiliency could make investors better off during normal times by changing the market microstructure they experience when trading over-the-counter securities such as corporate bonds.

A growing body of empirical literature investigates the impact of post-crisis regulation on liquidity. The US corporate bond market is the most commonly studied in this context given its importance and dealer-centric nature. On balance, this literature finds improvement or at least no deterioration in the average transaction costs of corporate bond trades

[1] The matchmaking mechanism therefore encompasses both agency and riskless principal trading. We elaborate on this issue in Section 3.1.

(Mizrach (2015), Adrian, Fleming, Shachar, and Vogt (2017), Anderson and Stulz (2017), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), and Trebbi and Xiao (2017)). Still, some papers document an increase in the cost of immediacy or transacting in times of stress (Bao, O'Hara, and Zhou (2018), Choi and Huh (2017), and Dick-Nielsen and Rossi (2018)). These studies also find a reduction in the amount of capital bank dealers commit to market making and a shift in their activity from market making towards matchmaking (Bao, O'Hara, and Zhou (2018), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), and Choi and Huh (2017)). In contrast, non-bank dealers appear to increase capital commitments and principal trading.

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 the average transaction costs could suggest that a shift from market making to matchmaking is beneficial to customers. The empirical evidence pertaining to the change in average transaction costs, however, may be accompanied by two less measurable but potentially very 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 due to the higher costs of immediacy (and their unwillingness to wait longer for execution), their welfare loss cannot be ascertained by observing executed trades. To assess these two potential costs, which are hard to measure empirically, one needs a model in which the trade-off between the cost of delay (matchmaking) and the cost of immediacy (market making) is explicitly considered, and customers take into account the various costs associated with trading in their decision whether to trade. This is the model we set out to investigate.

Our model features infinitesimal customers (buyers and sellers) who arrive at a constant rate and wish to trade bonds in an over-the-counter market. There are two representative broker-dealers standing for two groups of dealers: bank-affiliated dealers and non-bank-affiliated dealers (henceforth bank dealers and non-bank dealers, respectively). Our use of a representative bank dealer and a representative non-bank dealer is meant to recognize the market power dealers in the corporate bond market possess, while at the same time enabling us to explicitly consider the changing nature of competition between the two broker-dealer groups as a result of bank regulations. Bank dealers offer their customers two mechanisms for trading bonds. The first, market making, enables customers to trade immediately at a spread that reflects the bank dealers' cost of market making. The second, matchmaking, involves the bank dealers helping customers search for trading counterparties and earning a fee to facilitate such trades.

Banks dealers optimally set the market-making spread and the matchmaking fee while they also decide how much to invest in the search technology that determines the matching rate. Non-bank dealers offer market-making services to customers at a spread that they set to reflect their cost of market making and the competitive environment.[2] Customers are price takers and heterogeneous with regard to patience (or the value they attach to immediacy), which we model with private values. They optimize when deciding whether to trade as well as how to trade: immediately with the bank dealers, immediately with the non-bank dealers, or using the matchmaking search process facilitated by the bank dealers. Searching for a counterparty takes time, though, and the cost of using the matchmaking mechanism consists of both the cost of waiting for a match and the trading fee.

Our analysis focuses on investigating what happens to customer welfare, dealer behavior, and market outcomes when the regulatory costs imposed on bank market making rise. We show that if the market-making costs bank dealers incur are much lower than those that non-bank dealers incur, a small increase in bank regulatory costs raises the cost of immediacy. While some customers shift to the matchmaking mechanism, bank dealers do not have sufficiently strong incentives to significantly increase their investment in the matchmaking technology, and hence these customers incur high waiting costs and the combination of these effects could worsen overall customer welfare.

The economics of intermediation discussed thus far involve only customers and bank dealers when the cost of market making is low. An increase in bank dealers' regulatory costs beyond a certain point, however, renders market making by non-bank dealers viable, which injects competition for market-making services into the picture. Initially, the profitability of bank dealers' market-making operations is squeezed by the competitive threat even as they maintain their dominance. They find it optimal to increase their investment in the matchmaking technology to serve customers who otherwise would be deterred by the high cost of market making, and some of these customers choose to engage in a time-consuming search process to transact at a lower cost. Bank dealers may also lower their matchmaking fee, attracting customers who previously did not trade at all. The overall welfare of the population of customers therefore goes up as some customers are made better off even as others are worse off, and average transaction costs in the market can decrease despite the increase in the cost of immediacy.

[2] Non-bank dealers are typically smaller and are not subject to bank regulations that increase balance sheet costs (Bao, O'Hara, and Zhou (2018), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018)).

The story does not end here, however. As bank regulatory costs rise even further, customers who require immediacy switch to non-bank dealers that offer a more attractive market-making spread. Bank dealers in turn focus on extracting maximum rents from the matchmaking mechanism. Increasing regulatory costs beyond this point only makes non-bank dealers more profitable as it reduces the competitive threat posed by bank dealers. This clearly makes those seeking immediacy worse off, while engendering very little in the way of benefits to customers who use the matchmaking mechanism. As such, overall customer welfare can go down and average transaction costs in the market may increase. Beyond a certain regulatory cost threshold, non-bank dealers set the market-making spread as unconstrained monopolists, and further increases in bank regulatory costs have no impact on customer welfare or the cost of trading.

Our model demonstrates the complex and interesting economic interactions that evolve as bank regulatory costs increase, pointing to subtleties not fully recognized in the extant literature. Our contribution lies in characterizing how changes in bank regulatory costs, both in absolute terms and relative to the market-making costs of non-bank dealers, affect the nature of equilibrium in the market, with each equilibrium giving rise to a set of implications for customer welfare, dealer behavior, and market outcomes. These complex interactions also highlight why empirical work thus far has struggled to articulate the overall impact of increased bank regulatory costs on customer welfare in the corporate bond market.

Can our theory contribute to evaluating how the post-crisis regulatory initiatives impacted customer welfare in the corporate bond market? We address this question towards the end of the paper by relating the wealth of empirical implications generated by the model to findings reported in the empirical papers. As we increase bank regulatory costs in the model, our predictions for observable market outcomes change as we move from one equilibrium to another. We believe that the patterns uncovered by the empirical literature map onto an equilibrium region in our model where an increase in bank regulatory costs improves overall customer welfare.[3] While bank dealers are clearly worse off in the post-crisis regulatory environment, our model vividly demonstrates that there is no one-to-one correspondence between the profitability of bank dealers on the one hand and average transaction costs or overall customer welfare on the other hand. This wedge reflects bank dealers' endogenous switch from using their balance sheets to using search technology in liquidity provision. Hence, regulation that negatively impacts dealer profitability and capital commitment need

[3]It is important to stress that the empirical patterns documented in these papers reflect the overall effects of post-crisis regulations. Hence, our results do not imply that a particular rule (or any specific feature of a particular rule) is beneficial for customer welfare.

not translate into lower overall customer welfare.

The regulation of banks' proprietary trading is still a work in progress. In May 2018, the five agencies responsible for the administration of the Volcker Rule proposed changes to the rule. While most of the changes appear to be focused on simplifying reporting as opposed to significantly changing the operations of bank trading desks, some have the potential to affect market-making incentives. Therefore, we believe that it is of paramount importance to understand how bank regulations that change the cost of market making impact the welfare of customers and the functioning of the corporate bond market.

The rest of the paper proceeds as follows. In Section 2, we provide a brief exposition of the relevant post-crisis bank regulatory changes, discuss the findings of empirical papers that examine how liquidity has changed in the corporate bond market, and relate our paper to several recent theoretical models. In Section 3, we discuss the motivation behind the modeling approach we pursue and provide an exposition of the theoretical model. In Section 4, we formally solve for the different cases of equilibria that prevail as bank regulatory costs increase. In Section 5, we discuss the robustness of our implications to alternative specifications of the model. In Section 6, we use the insights of our model to address the question whether post-crisis bank regulations harmed investors in the corporate bond market, and Section 7 presents our conclusions.

2 Background and Literature

2.1 Post-Crisis Bank Regulations and the Corporate Bond Market

The Basel regulatory framework and the Dodd-Frank Act are the cornerstones of post-crisis bank regulations that heavily impacted market-making activities in the United States, including in the corporate bond market. Interviews with market participants suggest that the revision of the Basel II market-risk framework ("Basel 2.5"), finalized in June 2012 in the United States, increased inventory costs for corporate bonds by boosting regulatory capital charges through the incremental risk capital (IRC) charge and the trading book stressed VaR requirement (BIS Committee on the Global Financial System (2014)). Furthermore, the Basel III framework, finalized in July 2013 in the United States, not only raised the risk-based capital requirements on banks, but also raised the non-risk-based capital requirement through the supplementary leverage ratio (SLR). For example, global systemically important banks (G-SIBs) are required to maintain an SLR of 5% or higher at the bank holding company level and an SLR of 6% or higher at the depository subsidiary (see Davis Polk (2014) for more details). Greenwood, Hanson, Stein, and Sunderam (2017) and Duffie (2018) find that the leverage ratio requirement is the most tightly binding constraint for most US G-SIBs, according to data derived from the Federal Reserve's stress tests in 2017. Besides capital requirements, another pillar of Basel III is higher liquidity standards, including the liquidity coverage ratio (LCR) and the net stable funding ratio (NSFR), which require banks to hold enough high-quality liquid assets or use sufficiently stable financing sources to guard against adverse conditions in the funding markets.

The most relevant part of the Dodd-Frank Act for trading in the corporate bond market is the Volcker Rule. While the Act was signed into law in July 2010, implementation of this rule was delayed until April 2014, with full compliance required by July 2015. The Volcker Rule aims at discouraging banks with access to FDIC insurance or the Federal Reserve's discount window from engaging in proprietary trading of risky securities. While it contains an exception for market making, the rule mandates the reporting of various measures (e.g., inventory turnover, standard deviation of daily trading profits, customer-facing trade ratio) to aid regulators in distinguishing between market making and other forms of proprietary trading. The scrutiny that these measures receive from regulatory agencies changes the incentives of bank dealers and thereby increases market-making costs.[4]

The suite of post-crisis bank regulations now in place discourages risk taking to strengthen banks' resilience to various risks (e.g., market, counterparty, and funding). As such, these rules have an impact on all trading activities undertaken by banks, including market making. Some market participants therefore warned that these rules would increase transaction costs in the corporate bond market, and these claims motivated subsequent empirical research examining post-crisis changes in corporate bond liquidity.

2.2 Post-Crisis Changes in Corporate Bond Liquidity: Empirical Evidence

Most empirical papers that examine corporate bond market liquidity following the financial crisis find no deterioration, indeed even improvement, in liquidity subsequent to the regulatory interventions. For example, Mizrach (2015) finds that bid-ask spreads and the price impact of trades dropped after the crisis to below pre-crisis levels. Similarly, Adrian,

[4]The regulatory costs we model in this paper probably best represent the explicit costs imposed by the Basel framework rather than the implicit costs associated with the Volcker Rule. Still, most empirical findings concerning the impact of post-crisis regulations on the corporate bond market reflect the combined effects brought about by all these rules and therefore we model a single regulatory cost of market making rather than investigating specific features of these rules.

Fleming, Shachar, and Vogt (2017) find that spreads and price impact declined. Furthermore, they document that trading volume and issuance are at record levels, although trade size and turnover have declined. Anderson and Stulz (2017) look at a variety of measures, including price-based measures (price impact and bid-ask spreads) and trading measures. They find overall that price-based measures are marginally better following the onset of regulation (2013-2014) relative to before the crisis. Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018) also document that average customer trade execution costs have not increased after these regulations were imposed. Trebbi and Xiao (2017) look for a structural break in various liquidity measures but find no evidence for deteriorating liquidity during the period that corresponds to the Dodd-Frank and Basel III regulatory interventions.[5]

Several papers find worsening in a particular dimension of trade execution: the cost of immediacy. Bao, O'Hara, and Zhou (2018) use downgrades of bonds to junk status as stress events to examine the cost of immediacy, and find that it increased following the implementation of the Volcker Rule. Similarly, Dick-Nielsen and Rossi (2018) use exclusions from the Barclays Capital investment-grade bond index as events that create demand for immediacy, and document an increase in the cost of immediacy after the financial crisis. Choi and Huh (2017) show that trading costs for unmatched (i.e., market-making) trades increased substantially in the post-regulation period relative to pre-crisis levels, and this increase is driven by bank dealers.

Our model is able to reconcile these two lines of empirical evidence that seem at odds with one another, namely that average transaction costs in corporate bonds have declined while the cost of immediacy has gone up. The driving force behind these results in our theory is the bank dealers' endogenous shift from market making to matchmaking. The empirical papers we have cited indeed provide ample evidence that matchmaking has increased following the crisis and the implementation of post-crisis regulations (e.g., Bao, O'Hara, and Zhou (2018), Choi and Huh (2017), Schultz (2017)).[6] As a result of the shift to matchmaking, the execution of large trades now tends to require more time (BIS Committee on the Global

[5]Two papers find more nuanced effects. Allahrakha, Cetina, Munyan, and Watugala (2019) find higher markups for a subset of the 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.

[6]Ederington, Guan, and Yadav (2014), Randall (2015), and Anand, Jotikasthira, and Venkataraman (2018) provide additional evidence about matchmaking and provision of liquidity by customers in the corporate bond market.

Financial System (2014, 2016)).

It is important to stress that there are differences in the manner in which bank dealers and non-bank dealers responded to the regulatory changes. Bao, O'Hara, and Zhou (2018) find that dealers affected by the Volcker Rule increase their matchmaking activity while committing less capital to market making.[7] At the same time, competition in market making from smaller non-bank dealers appears to intensify: they increase capital commitments and the amount of principal trading (see also Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018)). Non-bank-dealer matchmaking, on the other hand, has decreased after the implementation of the Volcker Rule. We return to the findings of the empirical literature to motivate our modeling approach and guide our discussion of how bank regulations following the financial crisis affected customer welfare.

2.3 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 different approach to studying the two mechanisms. An, Song, and Zhang (2017) study intermediation chains by modeling the 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 not only the pricing of their services but also how much to invest in the matchmaking technology.

[7] Goldstein and Hotchkiss (2017) investigate the trades that dealers seek to offset versus those that they hold in inventory overnight.

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.[8] Transparency plays no role in our model because we assume homogeneous common-value information. Our emphasis, instead, is on competition from non-bank dealers and the role it plays in determining customer welfare and the extent of matchmaking.

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, we show that investment in technology can make matchmaking a less expensive form of intermediation than market making, and the market power of dealers provides a role for bank regulations in enhancing competition and improving customer welfare.

3 Model

3.1 Motivating Our Modeling Framework

Before going into the technical presentation of the model, we motivate our choices regarding its key modeling features. As our discussion of the theoretical literature in the previous section suggests, the particular features of the bond trading environment one chooses to include depend on the objective of the analysis.

Our goal is to investigate how structural changes in the corporate bond market, brought about by bank regulations that increase the cost of market making, impact customer welfare and market liquidity. The empirical finding of a shift from market making to matchmaking for bank dealers tells us that including these two trading mechanisms and allowing both bank dealers and customers to make optimal choices are necessary components of a model that seeks to understand the evolution of the bond market in the post-crisis era. Furthermore, the last several years have witnessed increased broker-dealer investment in fixed-income electronic trading, giving rise to new trading venues and protocols (BIS Markets Committee

[8] 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.

(2016)).[9] Enhancing matchmaking in the corporate bond market depends critically on bank dealers' prioritizing investment in these technologies, and therefore our model requires bank dealers to optimally choose their level of investment.

It is important to acknowledge the difficulty involved in empirically measuring the extent of matchmaking. The empirical papers we cite in Section 2.2 use various forms of the TRACE database. 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. If a trade moves through a proprietary trading account, it is considered a principal trade by the Financial Industry Regulatory Authority (FINRA). Because TRACE does not currently support riskless principal reporting as a separate category, such trades will be reported as principal despite the fact that their economics are identical to those of agency trading. 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.[10] Given that agency and riskless principal have similar balance-sheet implications, we lump them together under our matchmaking mechanism. In other words, matchmaking consists of all dealer-facilitated trading that does not involve taking a trade onto the bank dealer's balance sheet.

An important aspect of liquidity provision in the corporate bond market that we choose to model explicitly is competition between bank and non-bank dealers. The significant role that non-bank dealers can play by stepping into the void left by banks was emphasized even before post-crisis regulations were implemented (e.g., Duffie (2012)). The empirical findings cited above, that non-bank dealers have increased their capital commitments and market-making activity, must be allowed to arise endogenously in our model, and therefore we include non-bank dealers that offer market-making services and compete with bank dealers.

A related choice we need to make is whether to allow non-bank dealers to operate a matchmaking mechanism. The model in this paper does not include this feature so that we can obtain closed-form solutions for clarity of intuition. Still, there are several reasons to believe that omitting the matchmaking mechanism for non-bank dealers does not detract from the external validity of the model's implications. First, Bao, O'Hara, and Zhou (2018)

[9]In particular, electronic trading that is typically based on permutations of the request-for-quote (RFQ) protocol grew in the dealer-client segment.

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

and Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018) find that most non-bank dealers are small. Many of these non-bank dealers are regional in focus, and it is reasonable to assume that most of them do not possess vast networks of customers that would allow them to operate an efficient matchmaking mechanism. Bank dealers, on the other hand, have long dominated the OTC markets and have rich information about a large group of customers. Second, Bao, O'Hara, and Zhou (2018) find that non-bank dealers decrease their matchmaking activity following the implementation of the Volcker Rule (by about 40% of the pre-crisis level), which could suggest they are being crowded out by the bank dealers. The main dimension along which they inject competition into the corporate bond market following the financial crisis appears to be market making (by boosting both capital commitments and trading). As such, market making is the dimension we choose to model explicitly for non-bank dealers.

Insofar as our focus is on bank regulation that applies to all bank dealers and, by the same token, does not apply to any non-bank dealers, we choose to model each type using a representative dealer that stands for the entire group of dealers. In so doing, we abstract from the inter-dealer market and competition within each group. While other papers (e.g., An, Song, and Zhang (2017)) focus on the inter-dealer market, we believe that the basic economics of intermediation chains did not materially change with the implementation of post-crisis bank regulations. Furthermore, while we stress competition from non-bank dealers as a factor that is important for understanding the impact of bank regulations, we strongly believe that dealers in the corporate bond market have market power. 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. The usual explanation for this empirical regularity is dealer market power: large customers have greater bargaining power and hence obtain better prices or lower fees than small customers can. In the model, the spread charged by a dealer does not depend on the identity of the customer. While clearly a simplification, we believe that specifying several types of customers with varying degrees of bargaining power would not materially change the main implications of our model. Hence, we choose to simplify the exposition by having only one type of customer and giving the full market power to the dealer.[11] In other words, each representative dealer optimally sets prices to maximize profits subject to competition from the other representative dealer. The market power of the bank dealer is important for the implications we derive concerning

[11] We note that some bank dealers may find it optimal to stop offering market making services as bank regulatory costs go up, increasing the market power of the remaining bank dealers.

customer welfare. In Section 5.2 we discuss this issue and solve a variation of the model with a benevolent bank dealer who makes zero profit.

Two elements of the model that are critical to evaluating the impact of post-crisis bank regulations on the corporate bond market involve the customers. First, we assume that customers are heterogeneous with respect to their degree of patience, or the private value they attach to trading immediately. The customers in our model can be thought of as institutional investors whose trades are relatively large. These trades, which account for most of the volume in the corporate bond market, require the active intermediation of OTC broker-dealers either in a principal (market making) or an agent (matchmaking) capacity.[12] Market making allows customers to trade immediately at a cost, while searching via the matchmaking mechanism takes time. Given the ample empirical evidence that matchmaking is a viable option for some customers in the corporate bond market, it has to be the case that customers are heterogeneous with respect to patience or the desire for immediacy. Our model provides for such heterogeneity.

Second, we let customers optimize over whether to trade and how to trade. Optimizing customers are an important feature of the model because we seek to contribute insights that are difficult to observe empirically. As we stress in the introduction, the empirical finding of lower average transaction costs can be accompanied by other outcomes that are not easily measured but that negatively impact customer welfare (e.g., delayed execution of trades or the loss of welfare of customers who choose not to trade because the cost of immediacy rises). In the model, customers optimally choose between (i) not trading, (ii) using the matchmaking mechanism and incurring both a delay cost and a trading fee, and (iii) choosing to trade immediately with the dealer that posts the best spread (either the bank dealer or the non-bank dealer). This optimization is at the core of the model's ability to speak to the manner in which bank regulations affect customer welfare through their impact on the microstructure of the corporate bond market.

Lastly, we model bank regulatory costs as per-share balance sheet costs imposed on all trades accommodated by the bank dealer in his capacity as a market maker. This is a parsimonious way to differentiate market-making activity from trade facilitation via the matchmaking mechanism that is not subject to such regulatory costs. For symmetry, the non-bank dealer's market-making costs are also modeled as per-share costs. Given this

[12]In contrast, retail-size orders are often traded on MarketAxess, which is a centralized system that currently executes almost 20% of the volume in the corporate bond market (investment grade and high yield combined). MarketAxess operates an all-to-all trading platform that gives access to both dealers and customers.

specification of balance sheet costs and the risk-neutrality of dealers in the model, the level of inventory does not play a role in pricing.[13] As such, there is no loss of generality in assuming equal arrival rates of customers who wish to buy and customers who wish to sell, which simplifies the exposition of the model.

3.2 Model Setup

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 $r > 0$.[14]

Customers and dealers. Infinitesimal buyers arrive in the market at the 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 the private benefit (or "value") for trading immediately is a random variable $x \in [0, \infty)$, with cumulative distribution function G. Heterogeneity in this private value is the manner in which we model differences across buyers in their degree of patience. 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. 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 private benefit for trading is not observable by others, and the customer exits the market upon trading.

There are two representative yet distinct strategic intermediaries, called dealers, who help customers trade this asset. One of the dealers is a bank affiliate, subject to bank regulations, whereas the other security dealer is unaffiliated with any bank and hence is not subject to bank regulations. In practice, all intermediaries face some types of regulatory constraints, but the specific post-crisis regulations we discuss in Section 2.1 (Basel 2.5, Basel III, and the Volcker Rule) apply only to banks and their broker-dealer affiliates.

Trading protocols: market making and matchmaking. There are two trading mechanisms (or protocols), which we call "market making" and "matchmaking." 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. To fill an order immediately, the bank

[13] 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.

[14] In our model, the discount rate r captures two effects: the rate at which customers and dealers discount future profits as well as the rate at which trading opportunities decay over time.

(non-bank) dealer charges customers a per-unit spread of $S_B > 0$ ($S_{NB} > 0$), which is publicly observable. The bank dealer's balance sheet cost is assumed to be c_B per unit of the asset regardless of whether he is accommodating a buy or a sell order (that is, the cost is incurred on the gross amount traded). The corresponding balance sheet cost for the non-bank dealer is c_{NB}.

The bank dealer's cost is conceptually comprised of three components:

$$c_B = \mathit{OperatingCosts} - \mathit{ImplicitSubsidy} + \mathit{PostCrisisRegulatoryCosts}.$$

The first component reflects the direct operating costs involved in running the market-making business. We assume that the bank dealer operates optimally to minimize this cost. This component is similar to the cost of the non-bank dealer, c_{NB}, and can be influenced by the same economic forces. The second component (*ImplicitSubsidy*) has been discussed extensively in regulatory circles in the aftermath of the financial crisis. 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)). This implicit too-big-to-fail subsidy lowers the capital costs of the bank dealer's trading book relative to that of the non-bank dealer 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. Fixing the first two components, the imposition of post-crisis regulations unequivocally increases the bank dealer's cost, but it is unclear whether the sum of the three components is lower than, equal to, or greater than the cost of the non-bank dealer. We therefore investigate how post-crisis bank regulations impact dealer actions, investor choices, and market outcomes by looking at the effects of increasing c_B over the entire relevant range (below and above c_{NB}).[15]

Bank dealers can invest in developing a matchmaking mechanism. Under the matchmaking protocol, a dealer searches for a counterparty for the customer's order, but the dealer does not use his balance sheet to accommodate the order. Using the matchmaking mechanism, a customer is matched with a counterparty at an exponentially distributed time τ with intensity $H \in [0, \infty)$.[16] While the dealer conducts the search, the customer incurs a

[15]Throughout the paper, when we discuss an increase in c_B we mean an increase in the third component, holding the other two components constant.

[16]Specifically, τ has a cumulative distribution function $P(\tau < t) = 1 - e^{-Ht}$. The matching intensity in

delay cost by discounting the private benefit of trading. Given the exponential distribution of matching time τ, we define the benefit customers obtain from matchmaking with intensity H as

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

Higher $\mathcal{H}$ implies a shorter waiting time (with a lower cost of delay) for the searching customers, and hence we refer to $\mathcal{H}$ as the *speed* of matchmaking.

To achieve matchmaking speed $\mathcal{H}$, the bank dealer invests $K(\mathcal{H})$ at time 0, where $K(0) = 0$, $K'(0) = 0$, $K'(\mathcal{H}) > 0$ for all $\mathcal{H} \in (0,1]$, and $K''(\mathcal{H}) > 0$ for all $\mathcal{H} \in [0,1]$.[17] This investment represents various aspects involved in operating a matchmaking mechanism such as the cost of installing trading technology and deploying salespeople. Certain investments in trading technology have the potential to benefit both matchmaking and market making, and we assume that these are made optimally by the bank dealer to minimize the operating-cost component of c_B. $K(\mathcal{H})$ is therefore an incremental investment that enhances only the matchmaking mechanism over and above other investments that benefit the bank dealer's entire operation.[18]

To simplify the exposition, we model the bank dealer's investment in technology as a one-time cost at time 0. This enables us to maintain general functional forms for both the distribution of investor types and the cost of investment while obtaining closed-form solutions and explicit comparative statics. The main implications of our model do not depend on specifying a one-time investment in technology. In Section 5.3, we discuss a variation of the model in which the bank dealer incurs a positive marginal cost for every trader who chooses the matchmaking service. We solve this variation of the model with a specific distribution for G and a particular functional form for K, obtaining implications similar to those of our main model.

our model does not depend on the mass of customers who choose matchmaking. This departure from the Duffie, Gârleanu, and Pedersen (2005) framework is made for analytical simplicity.

[17]The assumptions regarding $K(\mathcal{H})$ are meant to ensure that the equilibrium investment in matchmaking technology is positive and finite.

[18]The search technology we describe increases the matching speed for the bank dealer's customers. The report on electronic trading in fixed income markets issued by the Markets Committee of the Bank for International Settlements notes that growth in electronic trading has occurred primarily in the dealer-client segment. Many of the systems in this segment are single-dealer platforms. Investment in such technology by one bank dealer does not increase the matching speed of another dealer. In practice, the search for a counterparty for a customer's trade may take several forms. For example, a bank dealer may tap liquidity on MarketAxess, which is an all-to-all trading platform primarily for smaller trades. We view MarketAxess simply as one of the tools that bank dealers can use to search, although it often not the tool of choice for large institutional trades that are our primary focus, and we do not model each such tool separately in this paper.

When a match is made, the bank dealer receives a publicly observable fee of f from both the buyer and the seller. As Choi and Huh (2017) note, a dealer searching for a counterparty may need to offer better terms of trade to the counterparty in order to execute the trade. Hence, the two sides of the trade need not be symmetric in the amount they pay for participating in the transaction, and one side may even buy (sell) at a lower (higher) price than the fundamental value. We view $2f$ as the net compensation earned by a dealer that executes both legs of such an agency (or riskless principal) cross.

Equilibrium. An equilibrium consists of:

1. The bank dealer's choices of market-making spread S_B, matching speed $\mathcal{H}$, and matching 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.[19]

3.3 Customer and Dealer Objectives

The customer's optimization problem is simple: she needs to choose between trading immediately with the bank dealer, trading immediately with the non-bank dealer, searching for a counterparty using the matchmaking service, or not trading. The bank and non-bank dealers' market-making services are identical from the customer perspective. Therefore, a customer who opts to trade immediately will choose the market-making service that charges the lower spread, 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

[19]Note that in our definition of equilibrium, all price decisions of the two dealers, $\{S_B, f, S_{NB}\}$, are made at time 0. That is, we assume that they can commit to the prices determined at time 0 and do not dynamically adjust them. In a stationary environment like ours, an equilibrium obtained under committed prices is also an equilibrium if we allow dealers to change their prices $\{S_B, f, S_{NB}\}$ dynamically, under off-equilibrium beliefs that deviations are "punished" sufficiently severely. We do not, however, claim that the equilibrium obtained under committed prices is the unique equilibrium if dealers can adjust prices over time.

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. Obviously, 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 for $\mathcal{H} < 1$ 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 $x \in [f, b]$, or choose market making with the dealer offering the lower spread if $x > b$.[20]

Given the two thresholds f and b, the overall welfare of customers aggregated across the three ranges of x is:

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

The bank dealer's profit is comprised of three components: the discounted matchmaking profit (that depends on the fee and the matching speed), the discounted market-making profit (if the bank offers the lower spread), and the up-front cost $K(\mathcal{H})$, such that

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

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

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

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

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

[20] In Section 4 we show that any equilibrium in which the bank dealer invests in matchmaking ($\mathcal{H} > 0$) must satisfy $f < S$ (i.e., the matchmaking fee is lower than the market-making spread).

4 Equilibrium

Before delving into the full equilibrium analysis, we provide an intuitive description of the equilibrium structure and how it evolves as bank regulatory costs increase. Readers wishing to see the formal statements first are invited to jump to Section 4.2 and then come back to Section 4.1.

4.1 Properties of the Equilibrium: An Intuitive Exposition

A key qualitative insight of our model is that increasing the regulatory balance sheet cost of the bank dealer encourages investment in a matchmaking technology that the bank dealer, seeking to preserve market-making profits, otherwise shuns. Figure 1 illustrates this idea by plotting the customers' transaction costs (Panel A), choice of trading mechanism (Panel B), and overall welfare (Panel C) as functions of the bank dealer's balance sheet cost, c_B, holding all other model parameters constant. For this example, we use an exponential distribution for x (the gains from trade) with a mean of 10 basis points and a quadratic function for the cost of the matchmaking technology $K(\mathcal{H}) = k\mathcal{H}^2$, where $k = 0.1$.[21]

We observe that the equilibrium changes its properties as bank regulatory costs increase, and we distinguish between four regions (separated by the vertical dashed lines). In region 1, the bank dealer's cost is low enough that 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 region, the bank dealer has an unconstrained monopoly on the provision of immediacy. As the bank dealer's balance sheet cost increases, the spread rises as well. Certain customers stop paying the spread and opt instead to use the matchmaking mechanism (the blue area shrinks and the green area expands in Panel B). Due to this endogenous shift, the rate of increase in customers' average transaction costs is lower than the rate of increase in the bank dealer's spread (dotted line in Panel A) and overall customer welfare decreases (dotted line in Panel C). Region 1 corresponds to the common intuition that a higher bank dealer balance sheet cost has negative consequences for market liquidity and customer welfare.

As the bank dealer's balance sheet cost continues to rise and approaches (but is still lower than) the non-bank dealer's balance sheet cost, the bank dealer would have liked to set $S_B > c_{NB}$. However, the non-bank dealer now exerts competitive pressure on the bank dealer and therefore the market-making spread, S_B, needs to equal the non-bank dealer's balance sheet cost, c_{NB} (Panel A). This is region 2 of the equilibrium. Because the bank

[21] Other parameters we use for this numerical example are $\mu = 1$, $r = 0.01$, and $c_{NB} = 20$ bps.

dealer is now unable to increase his spread without losing the market-making business, he seeks to extract higher profits from the matchmaking business by making it more attractive to customers. Therefore, the bank dealer increases his investment in matching speed and lowers the matching fee. Certain customers who were previously excluded from this market start to trade, which can be observed as the shrinking “no trade” (red) area in Panel B. As a result, customers’ average transaction costs are now lower (Panel A) and their overall welfare is higher (Panel C). This region is the most interesting case in our model, where an increase in the bank dealer’s balance sheet cost encourages the bank to invest in matchmaking technology and shift his business model towards one that involves matching buyers and sellers as an agent.

We emphasize that customer welfare increases in region 2 when the balance sheet cost rises because the bank dealer possesses market power and extracts rents from customers. The bank dealer’s optimal response to the increase in the regulatory costs of market making is to attract customers to the matchmaking mechanism, resulting in a decline in the rents he collects and a concomitant increase in overall customer welfare. The results would be quite different if the bank dealer were benevolent. In Section 5.2, 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.

Region 3 and region 4 of the equilibrium are obtained when the bank dealer’s balance sheet cost, c_B, exceeds the non-bank dealer’s balance sheet cost, c_{NB}. In these regions, the non-bank dealer takes over market making and the bank dealer operates only the matchmaking mechanism. In region 3, the non-bank dealer’s spread, S_{NB} (red line in Panel A), is equal to the bank dealer’s balance sheet cost, c_B. As c_B rises, the non-bank dealer increases his market-making spread to extract more rents, creating an opportunity for the bank dealer to increase his matchmaking fee as well. As a result, average transaction costs increase (dotted line in Panel A) and overall customer welfare declines (dotted line in Panel C). In region 4, the bank dealer no longer exerts any competitive pressure on the non-bank dealer. Therefore, the non-bank dealer charges his unconstrained monopoly spread, and further increases in the bank dealer’s balance sheet cost no longer affect the equilibrium outcomes.

4.2 Equilibrium: Formal Statements

We begin by making the following assumption about the distribution of the private values customers attach to trading immediately:

Assumption 1. *For all x, the distribution G of customers' values satisfies*

$$\Phi(x) \geq 0, \quad \Phi'(x) \leq 0, \quad G''(x) \leq 0, \tag{8}$$

where

$$\Phi(x) \equiv G'(x) + \frac{(1 - G(x))\, G''(x)}{G'(x)}. \tag{9}$$

The first of these three conditions, $\Phi(x) \geq 0$, implies that the hazard rate $G'(x)/(1 - G(x))$ is weakly increasing in x.[22] The hazard rate plays a role when we consider the bank dealer's incentive to marginally increase the threshold b between matchmaking and market making (through changes in S_B, $\mathcal{H}$, and f) and it has a natural economic interpretation: the bank dealer collects market-making revenues from a population of customers equal to $1 - G(b)$ but loses the market-making revenues from the marginal customer who switches to matchmaking at rate $G'(b)$. The technical conditions $\Phi'(x) \leq 0$ and $G''(x) \leq 0$ are useful for proving the existence of an equilibrium. Assumption 1 is not overly restrictive and can be shown to hold for many distributions. For example, with the uniform distribution, $G(x) = x/B$ for $B > 0$, and this satisfies the conditions in Assumption 1 because $\Phi(x) = 1/B$, $\Phi'(x) = 0$, and $G''(x) = 0$. The exponential distribution also satisfies this assumption with $G(x) = 1 - e^{-x/\beta}$, and therefore $\Phi(x) \equiv 0$ and $G''(x) = -e^{-x/\beta}/\beta^2 < 0$.

As we discuss in Section 4.1, the equilibrium structure is comprised of several distinct cases depending on the degree of competition between the bank and non-bank dealers (or the relative magnitudes of c_B and c_{NB}). Case 1 occurs when the bank dealer's unconstrained monopoly spread is lower than the market-making cost of the non-bank dealer (and hence there is no competitive pressure from the non-bank dealer). In Case 2, the bank dealer still provides market-making services but his spread is constrained to equal the non-bank dealer's cost of market making. Case 3 is where the c_B exceeds c_{NB}, and as a result market-making services are supplied by the non-bank dealer. We consider these cases one by one. As we will see, most of our economic insights are borne out in Case 1 and Case 2. For this reason, the proofs for Cases 1 and 2 are provided in the internal Appendix, whereas the proofs for Case 3 are provided in the Internet Appendix.

[22] We have $\frac{d}{dx}\frac{1-G(x)}{G'(x)} = \frac{-G'(x)^2-(1-G(x))G''(x)}{G'(x)^2} = -\frac{\Phi(x)}{G'(x)}$.

4.2.1 Case 1: Unconstrained Bank Dealer ($S_1^\star < c_{NB}$)

We define the bank dealer's unconstrained optimization problem as Problem 1:

$$\max_{S \geq f \geq 0, \mathcal{H} \in [0,1]} \Pi_1 \equiv \frac{2\mu}{r} \left[\mathcal{H} f \left(G\left(b\right) - G\left(f\right) \right) + \left(S - c_B\right) \left(1 - G\left(b\right)\right) \right] - K\left(\mathcal{H}\right). \tag{10}$$

The following proposition establishes the existence and uniqueness of the solution to Problem 1.

Proposition 1. *Suppose $K''(\mathcal{H}) \geq \beta_1$ for all $\mathcal{H} \in (0,1)$, where β_1 is a constant.*[23] *Then, Problem 1 has a unique interior solution $(S_1^\star, \mathcal{H}_1^\star, f_1^\star)$, where $\mathcal{H}_1^\star \in (0,1)$ and $S_1^\star > f_1^\star > 0$. This solution satisfies the following first-order conditions:*

$$S_1^\star - c_B - \mathcal{H}_1^\star f_1^\star = \frac{1 - G\left(b_1^\star\right)}{G'\left(b_1^\star\right)} (1 - \mathcal{H}_1^\star), \tag{11}$$

$$f_1^\star = \frac{1 - G\left(f_1^\star\right)}{G'\left(f_1^\star\right)}, \tag{12}$$

$$\frac{r}{2\mu} K'\left(\mathcal{H}_1^\star\right) = f_1^\star \left(1 - G\left(f_1^\star\right)\right) - b_1^\star \left(1 - G\left(b_1^\star\right)\right), \tag{13}$$

where $b_1^\star = \frac{S_1^\star - \mathcal{H}_1^\star f_1^\star}{1 - \mathcal{H}_1^\star}$ is the private value of the marginal customer who is indifferent between matchmaking and market making.

The three equations that characterize the equilibrium are the (rearranged) first-order conditions with respect to S, f, and $\mathcal{H}$. The condition $K''(\mathcal{H}) \geq \beta_1$ plays two roles. First, it guarantees an interior solution for matchmaking (i.e., $\mathcal{H} < 1$), which we believe is the more empirically plausible case. Second, it is a sufficient condition for the equilibrium to be unique. Note that our equilibrium definition from Section 3.2 reduces in Case 1 to merely solving the bank dealer's optimization problem (given the simple discrete choice customers face). As such, uniqueness means that the bank dealer's objective function is single-peaked.

To determine how bank regulatory costs impact dealer behavior, market outcomes, and customer welfare over the entire range of c_B, we need to investigate how these attributes change within each of the cases. Proposition 2 provides the comparative statics for Case 1.

[23] $\beta_1 \in \left(\frac{8\mu\bar{c}}{r}, \infty\right)$ is the solution of

$$\frac{r}{2\mu}\beta = \frac{1}{\left(1 - \sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta}}\right)^2} c_B \max_b \left\{ bG'\left(b\right) \right\}.$$

Specifically, we look at how an increase in c_B impacts the bank dealer's optimal choices (spread, fee, and speed), the customers' choice of trading outcome and venue (i.e., matchmaking, market making, or refraining from trading altogether), the average transaction costs of customers, bank dealer profits, and overall customer welfare.[24]

Proposition 2. *In the equilibrium characterized by Proposition 1, if the bank dealer's balance sheet cost c_B increases and holding all else equal, then:*

1. *The bank dealer's spread ($S_1^\star$) increases, the matching speed ($\mathcal{H}_1^\star$) increases, and the matchmaking fee ($f_1^\star$) is unchanged.*

2. *The fraction of customers using the market-making service, $1 - G(b_1^\star)$, decreases; the fraction of customers choosing the matchmaking mechanism, $G(b_1^\star) - G(f_1^\star)$, increases; and the fraction of customers who refrain from trading, $G(f_1^\star)$, is unchanged.*

3. *Customers' average transaction costs, $\frac{1}{1-G(f_1^\star)}\left[\left(G(b_1^\star) - G(f_1^\star)\right) f_1^\star + \left(1 - G(b_1^\star)\right) S_1^\star\right]$, increase if and only if the following condition holds at the equilibrium $(S_1^\star, \mathcal{H}_1^\star, f_1^\star)$:*

$$\frac{c_B G'(b_1^\star)}{\frac{r}{2\mu} K''(\mathcal{H}_1^\star)(1 - \mathcal{H}_1^\star)^2} \leq \frac{1}{b_1^\star - f_1^\star} - \frac{G'(b_1^\star)}{1 - G(b_1^\star)}. \tag{14}$$

   Moreover, if $K''(\mathcal{H})(1-\mathcal{H})^2$ is non-increasing in $\mathcal{H}$, then customers' average transaction costs is a hump-shaped function of c_B, and the decreasing region may not exist (depending on the parameters of the economy).

4. *The Bank dealer's profit, π_B, decreases.*

5. *Overall customer welfare, π_c (defined in (5)), decreases if and only if the following condition holds at the equilibrium $(S_1^\star, \mathcal{H}_1^\star, f_1^\star)$:*

$$\frac{r}{2\mu} K''(\mathcal{H}_1^\star)(1 - \mathcal{H}_1^\star)^2 \geq c_B \frac{G'(b_1^\star)}{1 - G(b_1^\star)} \int_{f_1^\star}^{b_1^\star} (1 - G(x))\, dx. \tag{15}$$

   Moreover, if $K''(\mathcal{H})(1-\mathcal{H})^2$ is non-increasing in $\mathcal{H}$, then π_c is a U-shaped function of c_B, and the increasing region may not exist (depending on the parameters of the economy).

[24]The non-bank dealer has zero market share and makes zero profit in Case 1.

The bank dealer sets the matchmaking fee and speed in this case to maximize his unconstrained monopolist profit. Given a particular distribution of customers' private value (or patience), a lower fee increases the number of customers who choose to trade while decreasing the compensation from each trade. This tradeoff results in a unique fee that maximizes the bank dealer's expected profit from matchmaking that depends only on the distribution of private values customers attach to trading immediately (and hence does not change as bank regulatory costs increase). The bank passes any increase in regulatory costs to market-making customers by increasing the spread, leaving the bank dealer's profit per trade in the market-making mechanism unchanged. This means that increasing c_B shifts some customers from market making to matchmaking (as the spread increases), but the population of customers who refrain from trading (which depends only on the magnitude of the matchmaking fee) does not change.

Average transaction costs can be computed as the weighted average of the market-making spread and the matchmaking fee weighed by the respective populations of customers who choose each of these mechanisms. When regulatory costs increase, the spread increases but the fraction of customers choosing the market making mechanism decreases, and hence the direction of the change in average transaction costs can go either way. For c_B approaching zero, the right-hand-side of expression (14) approaches infinity while the left-hand-side is close to zero, ensuring that average transaction costs increase. As c_B goes up, either the condition in (14) holds until the equilibrium reaches Case 2 or the condition is violated for all c_B greater than a certain threshold, resulting in an inverse U-shape.

The last two comparative statics involve bank dealer profits and overall customer welfare. Bank dealer profits are unequivocally lower as regulatory costs increase (or, put differently, as the bank dealer's implicit "too-big-to-fail" subsidy is reduced). The rationale is clear: market making is more profitable for the bank dealer than matchmaking, and customers are shifting from the former to the latter. It is clear that some customers who wish to trade immediately are worse off when c_B increases. Whether the overall population of customers is benefitting or not depends on the extent of bank dealer investment in the matchmaking technology. Both $\mathcal{H}_1^\star$ and $b_1^\star$ increase when c_B goes up in Case 1, and the right-hand side (left-hand side) of the condition in (15) is increasing (non-increasing) in c_B. When balance sheet costs are close to zero, this condition is satisfied. As regulatory costs increase, either the condition is satisfied until the equilibrium reaches Case 2 (and welfare decreases over the entire range of Case 1) or the condition is violated at some threshold and is violated for all c_B greater than the threshold (resulting in welfare starting to increase in the interior of

Case 1).

Case 1 corresponds to region 1 in the example depicted in Figure 1. For this particular choice of parameters, average transaction costs increase and overall customer welfare decreases over the entire region rather than having an interior maximum and minimum, respectively. The comparative statics with respect to transaction costs, the fractions of customers choosing each trading mechanism, and customer welfare are clearly visible in the three panels of the figure. The lack of competition from the non-bank dealer in Case 1 allows the bank dealer to push the increase in regulatory costs to customers who trade via the market-making mechanism, making them worse off. In addition, the bank dealer does not have sufficient incentives in Case 1 to lower the matching fee to attract more cost-conscious customers to the market. These outcomes change dramatically when competition starts constraining the bank dealer's profit maximizing strategies.

4.2.2 Case 2: Constrained Bank Dealer ($S_2^\star = c_{NB}$)

Case 2 arises when the bank dealer's balance sheet cost is high enough that his unconstrained optimal spread (from Case 1) would be higher than the non-bank dealer's balance sheet cost c_{NB}. Given that for Problem 1 we would obtain $\frac{\partial \pi_B}{\partial S} > 0$ at $S = c_{NB} < S_1^\star$, the constraint binds. As a result, the bank dealer sets $S_2^\star = c_{NB}$.

We define the bank dealer's constrained optimization problem as Problem 2:

$$\max_{S \geq f \geq 0, \mathcal{H} \in [0,1]} \Pi_2 = \frac{2\mu}{r}\left[\mathcal{H} f\left(G\left(b\right) - G\left(f\right)\right) + \left(S - c_B\right)\left(1 - G\left(b\right)\right)\right] - K\left(\mathcal{H}\right), \tag{16}$$

such that $S = c_{NB}$.

The following proposition establishes the existence of equilibrium in Case 2:

Proposition 3. *If $K''(\mathcal{H}) \geq \beta_1$ for all $\mathcal{H} \in (0,1)$, then the solution for Problem 2 exists. Denote the solution as $(S_2^\star, \mathcal{H}_2^\star, f_2^\star)$. It satisfies $\mathcal{H}_2^\star \in (0,1)$, $S_2^\star = c_{NB}$, and the following first-order conditions:*

$$G\left(b_2^\star\right) - G\left(f_2^\star\right) = -\left(c_{NB} - c_B - \mathcal{H}_2^\star f_2^\star\right) G'\left(b_2^\star\right) \frac{1}{1 - \mathcal{H}_2^\star} + f_2^\star G'\left(f_2^\star\right), \tag{17}$$

$$\frac{r}{2\mu} K'\left(\mathcal{H}_2^\star\right) = f_2^\star\left(G\left(b_2^\star\right) - G\left(f_2^\star\right)\right) - \left(c_{NB} - c_B - \mathcal{H}_2^\star f_2^\star\right) G'\left(b_2^\star\right) \frac{c_{NB} - f_2^\star}{\left(1 - \mathcal{H}_2^\star\right)^2}, \tag{18}$$

where $b_2^\star = \frac{c_{NB} - f_2^\star \mathcal{H}_2^\star}{1 - \mathcal{H}_2^\star}$ is the private value of the marginal customer who is indifferent between matchmaking and market making.

While we do not require additional conditions (beyond $K''(\mathcal{H}) \geq \beta_1$ from Case 1) to establish the existence of the equilibrium in Case 2, the constrained nature of the optimization problem prevents us from demonstrating uniqueness. Still, the lack of uniqueness does not preclude characterizing the comparative statics, and the following proposition provides very sharp predictions regarding the impact of an increase in c_B on the economic environment in Case 2.

Proposition 4. *In the equilibrium characterized by Proposition 3, if the bank dealer's balance sheet cost c_B increases and holding all else equal, then:*

1. *The bank dealer's spread ($S_2^\star$) does not change, the matching speed ($\mathcal{H}_2^\star$) increases, and the matchmaking fee ($f_2^\star$) decreases.*

2. *The fraction of customers using the market-making service, $1 - G(b_2^\star)$, decreases; the fraction of customers choosing the matchmaking mechanism, $G(b_2^\star) - G(f_2^\star)$, increases; and the fraction of customers who refrain from trading, $G(f_2^\star)$, decreases.*

3. *Customers' average transaction costs, $\frac{1}{1-G(f_2^\star)}\left[\left(G(b_2^\star) - G(f_2^\star)\right) f_2^\star + \left(1 - G(b_2^\star)\right) S_2^\star\right]$, decrease.*

4. *The bank dealer's profit, π_B, decreases.*

5. *Overall customer welfare, π_c (defined in (5)), increases.*

The comparative statics in Case 2 are all unambiguously signed. As competition from the non-bank dealer constrains the bank dealer's ability to pass the rising regulatory costs onto market-making customers, the bank dealer seeks to extract higher profits from the matchmaking business. For that purpose, he increases overall trading volume (which is equivalent in our setting to the fraction of customer types who trade, $G(f_2^\star)$) by lowering the matching fee to attract customers who refrain from trading in Case 1. At the same time, he increases investment to achieve a higher matching speed and thereby to shorten the time it takes to collect the fee from each searching customer.

Shifting more customers to the less expensive matchmaking service lowers average transaction costs in the market, but even more notable is that overall customer welfare increases. The welfare loss of customers with a high need for immediacy who pay a higher spread is being offset by the welfare gains of two groups of customers: the first group contributed zero welfare in Case 1 (because these customers did not trade) but contributes a positive amount to welfare in Case 2 (by trading via the matchmaking mechanism), and the second group includes some of the switching customers who are better off in Case 2 with a lower fee and a higher matching speed than in Case 1 where they paid the spread. Case 2 corresponds to region 2 in the example depicted in Figure 1, and the patterns described by the comparative statics are very prominent in the figure. Here customers derive the greatest benefit from increased bank regulatory costs because the industrial organization angle (increased competition from the non-bank dealer) interacts with the market microstructure angle (improvements in the matchmaking mechanism that attract more customers to using it), leading to reduced bank dealer rents and increased overall customer welfare. In Section 6, we return to discuss these comparative statics and how they map onto the empirical evidence.

4.2.3 Case 3: $c_B \geq c_{NB}$

When $c_B > c_{NB}$, the bank dealer does not provide market making services because the non-bank dealer can always charge $S_{NB} = c_B - \epsilon$ and attract all customers who are willing to pay the spread to trade immediately.[25] Given the non-bank dealer's choice of spread S, the optimization problem of the bank dealer is

$$\max_{\mathcal{H}\in[0,1],f\geq 0} \Pi_{B,3} = \begin{cases} \frac{2\mu}{r}\mathcal{H}f\left(G\left(b\right)-G\left(f\right)\right)-K\left(\mathcal{H}\right) & \text{if } S \geq f \\ 0 & \text{if } S < f \end{cases}. \tag{19}$$

Given the bank dealer's choice of matchmaking speed and fee, $(\mathcal{H}, f)$, the optimization problem of the non-bank dealer is

$$\max_{c_B\geq S_{NB}} \Pi_{NB,3} = \begin{cases} \frac{2\mu}{r}\left[\left(S_{NB}-c_{NB}\right)\left(1-G\left(b\right)\right)\right] & \text{if } S_{NB} \geq f \\ \frac{2\mu}{r}\left[\left(S_{NB}-c_{NB}\right)\left(1-G\left(S_{NB}\right)\right)\right] & \text{if } c_{NB} < S_{NB} < f \end{cases}. \tag{20}$$

The strategy space for the two dealers is $\Lambda_0 = \{(S,\mathcal{H},f)\,|S\in[0,c_B]\,,\mathcal{H}\in[0,1]\,,f\geq 0\}$, where S is set by the non-bank dealer and $\mathcal{H}$ and f are set by the bank dealer. The following lemma shows useful restrictions on the optimal strategies of the bank and non-bank dealers.

Lemma 1. *In equilibrium, the optimal fee (f) of the bank dealer and the optimal spread (S) of the non-bank dealer satisfy* $f < \min\{\tilde{f}, S\}$*, where* $\tilde{f}$ *is the unique solution to* $\frac{1-G(\tilde{f})}{G'(\tilde{f})} - \tilde{f} = 0$.

[25] 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. This boundary point can be included in either Case 2 or Case 3, and we could also assume that customers who would like to use the market-making mechanism randomize between the bank and non-bank dealers.

Given Lemma 1, the new strategy space can be written as

$$\Lambda_1 = \left\{ (S, \mathcal{H}, f) \, | S \in [f, c_B] \, , \mathcal{H} \in [0,1] \, , f \in \left(0, \min \left\{ \tilde{f}, S \right\} \right) \right\}. \tag{21}$$

The bank dealer's optimization problem can be rewritten as

$$\max_{\mathcal{H} \in [0,1]; f \in \left(0, \min\left\{\tilde{f}, S\right\}\right)} \Pi_{B,3} = \frac{2\mu}{r} \mathcal{H} f \left(G(b) - G(f) \right) - K(\mathcal{H}) \tag{22}$$

and the optimization problem of the non-bank dealer can be rewritten as

$$\max_{c_B \geq S_{NB} \geq f} \Pi_{NB,3} = \frac{2\mu}{r} \left[(S_{NB} - c_{NB}) \left(1 - G(b)\right) \right]. \tag{23}$$

The equilibrium definition from Section 3.2 can be formally restated in Case 3 as follows:

Equilibrium in Case 3. A triple $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$ is called an equilibrium if

1. $S_3^\star \in \arg\max_{c_B \geq S \geq f_3^\star} \Pi_{NB,3}(S, \mathcal{H}_3^\star, f_3^\star)$;
2. $(\mathcal{H}_3^\star, f_3^\star) \in \arg\max_{\mathcal{H} \in [0,1], S_3^\star \geq f \geq 0} \Pi_{B,3}\left(S_3^\star, \mathcal{H}, f\right)$ and
3. Arriving customers maximize their expected profits by choosing between the non-bank dealer's market-making service, the bank dealer's matchmaking service, or refraining from trading altogether.

The bank dealer's strategy space in the equilibrium definition is specified as $S_3^\star \geq f \geq 0$ because it is implied by $\min\left\{\tilde{f}, S_3^\star\right\} \geq f \geq 0$. We call an equilibrium with $S_3^\star = c_B$ a *constrained equilibrium* and an equilibrium with $S_3^\star < c_B$ an *unconstrained equilibrium*. The following proposition characterizes the equilibrium.

Proposition 5. *Suppose that $K''(\mathcal{H}) > \beta_2$ for all $\mathcal{H} \in (0,1)$, where β_2 is a constant.*[26] *Then, for $c_B \geq c_{NB}$, the following results are true.*

1. *(Existence) There always exists at least one equilibrium $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$.*

[26] $\beta_2 \in \left(\frac{8\mu\bar{c}}{r}, \infty\right)$ is the solution to

$$\frac{r}{2\mu}\beta = \frac{1}{\left(1 - \sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta}}\right)^3} \frac{5}{2} c_B \max_b \left\{ bG'(b) \right\}. \tag{24}$$

2. *(Structure) There exist* $\bar{c}_1$ *and* $\bar{c}_2$ *satisfying* $c_{NB} < \bar{c}_1 < \bar{c}_2$ *such that*

   (a) *If* $c_B \in [c_{NB}, \bar{c}_1)$*, there exists a unique equilibrium, and the equilibrium is a constrained equilibrium.*

   (b) *If* $c_B \in [\bar{c}_2, \infty)$*, all equilibria are unconstrained.*

3. *(Optimality)* $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$ *satisfy the conditions below.*

   (a) $(\mathcal{H}_3^\star, f_3^\star)$ *satisfy the following first-order conditions for all possible equilibria:*

$$G\left(b_3^\star\right) - G\left(f_3^\star\right) = f_3^\star G'\left(f_3^\star\right) + f_3^\star \frac{\mathcal{H}_3^\star}{1 - \mathcal{H}_3^\star} G'\left(b_3^\star\right), \tag{25}$$

$$\frac{r}{2\mu} K'\left(\mathcal{H}_3^\star\right) = f_3^\star \left(G\left(b_3^\star\right) - G\left(f_3^\star\right)\right) + \frac{S_3^\star - f_3^\star}{\left(1 - \mathcal{H}_3^\star\right)^2} f_3^\star \mathcal{H}_3^\star G'\left(b_3^\star\right), \tag{26}$$

   where $b_3^\star = \frac{S_3^\star - \mathcal{H}_3^\star f_3^\star}{1 - \mathcal{H}_3^\star}$.

   (b) *In a constrained equilibrium,* $S_3^\star = c_B$*; in an unconstrained equilibrium,* $S_3^\star$ *satisfies the following first-order condition:*

$$1 - G\left(b_3^\star\right) - \left(S_3^\star - c_{NB}\right) \frac{G'\left(b_3^\star\right)}{1 - \mathcal{H}_3^\star} = 0. \tag{27}$$

Case 3 is more complex because the equilibrium definition does not reduce to the bank dealer's optimization problem (as in the two other cases). Rather, when c_B exceeds c_{NB} the bank and non-bank dealers need to take each other's best responses into account when choosing their strategies. We can show uniqueness of the constrained equilibrium as long as the bank regulatory costs are not too high. If $c_B \in [\bar{c}_1, \bar{c}_2)$, however, a constrained equilibrium may exist but multiple equilibria are possible. In the unconstrained equilibrium, the bank regulatory costs are greater than the non-bank dealer's optimal spread, and therefore nothing changes in the equilibrium if c_B increases. In other words, since the non-bank dealer does not increase his spread, the bank dealer does not have an incentive to change his choices of matchmaking fee or speed. As such, customers do not alter their behavior and overall welfare is unchanged. While the unconstrained equilibrium appears somewhat extreme, the constrained equilibrium is more nuanced and we now turn to characterizing its comparative statics.

Proposition 6. *In the case characterized by Proposition 5, when c_B increases, if $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$ is an unconstrained equilibrium, then nothing will be changed in the equilibrium; if $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$ is a constrained equilibrium, then*

1. *The non-bank dealer's optimal spread, $S_3^\star$, increases and the bank dealer's optimal fee, $f_3^\star$, increases. The matching speed, $\mathcal{H}_3^\star$, increases if and only if the following condition holds at the equilibrium $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$:*

$$\frac{\mathcal{H}}{1-\mathcal{H}} S G'^2(b) > f\left[G'(b) + \mathcal{H}(b-f) G''(b)\right]\left[-fG''(f) - 2G'(f)\right]. \tag{28}$$

   If, in addition, $K''(\mathcal{H}) > \beta_3$ for all $\mathcal{H} \in (0,1)$, where β_3 is a constant,[27] *then the bank dealer's optimal speed $\mathcal{H}_3^\star$ increases in c_B.*

2. *The fraction of customers paying the spread, $1 - G(b_3^\star)$, decreases. The fraction of customers who refrain from trading, $G(f_3^\star)$, increases. The fraction of customers choosing matchmaking, $G(b_3^\star) - G(f_3^\star)$, increases if and only if the following condition holds at the equilibrium $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$:*

$$\frac{\left(1 - \Gamma \frac{c_B}{1-\mathcal{H}}\right) G'(b) - \frac{\mathcal{H} f}{1-\mathcal{H}} G''(b)}{(1 - \Gamma(b-f))(2G'(f) + fG''(f)) + \frac{\mathcal{H}}{1-\mathcal{H}} G'(b)} < \frac{G'(b)}{G'(f)}, \tag{29}$$

   where $\Gamma = \frac{fG'(b)}{\frac{r}{2\mu} K''(\mathcal{H})(1-\mathcal{H})^2}$. In particular, $(G(b_3^\star) - G(f_3^\star))$ increases if c_{NB} is sufficiently small and c_B is greater than but sufficiently close to c_{NB}.

3. *Customers' average transaction costs $\frac{1}{1-G(f_3^\star)}\left[(G(b_3^\star) - G(f_3^\star)) f_3^\star + (1 - G(b_3^\star)) S_3^\star\right]$ increase if c_{NB} is sufficiently small and c_B is greater than but sufficiently close to c_{NB}.*

4. *The bank dealer's profit π_B increases. The non-bank dealer's profit, π_{NB}, increases if c_{NB} is sufficiently small and c_B is greater than but sufficiently close to c_{NB}. On the other hand, if c_B is sufficiently larger than c_{NB} (yet the equilibrium is still constrained) and K'' is sufficiently large, the non-bank dealer's profit, π_{NB}, decreases.*

[27] $\beta_3 \in \left(\frac{8\mu\bar{c}}{r}, \infty\right)$ is the solution to

$$\sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta}} = \frac{1}{1 + \bar{c}\left(-\frac{G''(0)}{G'(0)}\right)}$$

5. *Overall customer welfare, π_c (defined in (5)), decreases if K'' is sufficiently large. π_c also decreases if c_{NB} is sufficiently small and c_B is greater than but sufficiently close to c_{NB}.*

It is clear that raising regulatory costs in this case works to weaken competition rather than strengthening it. As such, the non-bank dealer becomes less constrained as c_B rises and can increase his spread. The bank dealer focuses on maximizing his profits from matchmaking and can do so by raising his fee. The higher fee means that more customers choose to refrain from trading, which decreases volume and negatively affects welfare. The impact on average transaction costs and overall customer welfare in the constrained equilibrium hinges on the investment in matching speed and the fraction of customers who choose to utilize the matchmaking service, both of which can rise or fall depending on the parameters of the economy. Overall customer welfare declines if the rate at which the cost of speed goes up is sufficiently high. Also, if the market-making cost of the non-bank dealer is sufficiently small and the bank dealer's cost is greater than but not too far from that of the non-bank dealer, then the investment in speed cannot offset the negative impact of the increases in the spread and fee, resulting in lower overall customer welfare. Both dealers benefit (at least when c_{NB} is sufficiently small and c_B is greater than but sufficiently close to c_{NB}) when each dealer specializes in a different business (the non-bank dealer in market making and the bank dealer in matchmaking) and is able to use his market power to extract higher rents.[28]

Case 3 corresponds to region 3 and region 4 in the example depicted in Figure 1. Specifically, region 3 shows the patterns associated with the constrained equilibrium. For the parameters in this specific example, we observe that average transaction costs rise only slightly, but overall customer welfare decreases by a large amount due to the increase in the fraction of customers who choose to refrain from trading. Region 4 in this example is consistent with the unconstrained equilibrium, where further increases in c_B have no impact on any of the outcomes.

5 Robustness

In this section, we elaborate on three aspects of our model. In Section 5.1 we consider a potential alternative metric for welfare that includes, in addition to customer welfare,

[28]The non-bank dealer's profits could potentially fall if the bank dealer's strategy attracts enough customers to the matchmaking mechanism, but this can happen only for very high values of c_B that are close to the boundary of the unconstrained equilibrium.

dealer profits and the implicit too-big-to-fail subsidy the bank dealer receives. We examine what happens to this more expansive welfare metric when the regulatory costs of bank dealers increase. In Section 5.2 we study a variation of our model in which the bank dealer is benevolent: he maximizes customer welfare subject to a zero-profit condition. As we mention in Section 4.1, solving this variation of the model demonstrates that the bank dealer's market power is a driving force behind the increase in customer welfare in region 2 (or Case 2) in our model. In Section 5.3 we investigate a variation of our model in which the cost of operating the matchmaking business increases with the number of customers utilizing it. This demonstrates the robustness of our main finding—the increase in customer welfare in region 2—to the assumption in our model that investment in matchmaking occurs at time zero and the bank dealer incurs zero marginal cost for additional customer utilization. Proofs of the results reported in this section are provided in the Internet Appendix.

5.1 Further Discussion of Welfare

Throughout the paper, we focus on customer welfare as the metric with which to evaluate the ultimate benefit or cost of post-crisis regulations. This quantity, defined in (5), reflects the gains from trade of customers in the model (who represent the investor population in our economy). For us, this is a natural benchmark for regulators to consider.

Our customer welfare measure, however, lacks two components that one may care about. First, the too-big-to-fail subsidy the bank dealer receives—reflected in the difference in balance sheet costs between the bank and non-bank dealers—ultimately comes out of the pocket of investors (or the public in general), so we may want to include this adjustment. Second, both the bank and non-bank dealers are agents in the economy, and one may want to include them in the definition of welfare.

Therefore, we consider an alternative welfare measure as the sum of customer welfare and dealer profits, while adjusting for the implicit too-big-to-fail subsidy of the bank dealer:

$$W = \pi_c + \pi_B + \pi_{NB} - \underbrace{\frac{2\mu}{r}\left(1 - G\left(b\right)\right)\left(c_{NB} - c_B\right)\mathbb{I}_{S=S_B}}_{\text{Too-big-to-fail subsidy of the bank dealer, if } c_B < c_{NB}}. \tag{30}$$

The first three terms of this expression are defined in Section 3.3, but the last term requires additional discussion. Our assumption is that, without giving an implicit public subsidy to the bank dealer, the balance sheet costs of the bank dealer would be similar to those of the non-bank dealer—that is, the same market-making activity would require the same cost of capital regardless of who provides it.[29] Hence, the difference between the two balance sheet costs, multiplied by the number of customers who trade using the market-making services of the bank dealer, represents the welfare loss to investors associated with granting this subsidy to the bank dealer. As post-crisis regulations increase the market-making costs of the bank dealer, the likelihood that the dealer would need to be rescued decreases, and with it the implicit subsidy. As c_B approaches c_{NB}, the subsidy given to the bank dealer's balance sheet cost disappears. The indicator function at the end of the term is there to ensure that the subsidy is considered only when the bank dealer makes markets.

Because the subsidy to the bank dealer is relevant if and only if the bank dealer makes markets, we can make the intuition more transparent by expanding (30) in Case 1 and Case 2 of the equilibrium, where $c_B < c_{NB}$ and $\pi_{NB} = 0$:

$$W = \frac{2\mu}{r}\left[\int_{x=f}^{b} \mathcal{H} x dG(x) + \int_{x=b}^{\infty} (x - c_{NB}) dG(x)\right] - K(\mathcal{H}), \text{ if } c_B < c_{NB}. \tag{31}$$

This expression essentially measures investors' gains from trades, after adjusting for the cost $K(\mathcal{H})$ of the technology investment and the non-bank dealer's balance sheet cost c_{NB}, which is taken to be the market-determined, risk-based balance sheet cost of market-making activity. While c_B does not show up explicitly in (31), an increase in c_B changes the equilibrium values of f, b, and $\mathcal{H}$ as described in sections 4.2.1 and 4.2.2. Overall, W reflects the allocative efficiency of the market.

The following proposition shows what happens to W as the bank dealer's balance sheet costs increase in Case 1 and Case 2 of the equilibrium.

Proposition 7. *In the equilibria characterized by Proposition 1 and Proposition 3, if the bank dealer's balance sheet cost c_B increases and holding all else equal, then the alternative welfare measure, W, increases.*

Proposition 7 shows that the implicit too-big-to-fail subsidy has substantial influence on how the alternative welfare measure behaves. We know from Proposition 2 that in Case 1, the bank dealer's profit π_B decreases in c_B while customer welfare π_c can decrease over the entire region or start increasing in the interior for some parameter values (the non-bank dealer does not trade in Case 1). Hence, the sum $\pi_B + \pi_c$ can either decrease or have an ambiguous sign for some parameter values. In Case 2, we know from Proposition 4 that

[29] The cost of capital in this context means the cost of capital of the activity, not of the entire institution that may have many subsidiaries. In the data, the cost of capital for each business line might be difficult to observe.

customer welfare π_c increases while the bank dealer's profit π_B decreases when c_B rises. Again, the manner in which $\pi_c + \pi_B$ is changing as we increase c_B could have an ambiguous sign. The unambiguous result expressed in Proposition 7, that W increases in both cases, is therefore a testament to the importance of accounting for the implicit too-big-to-fail subsidy, $\frac{2\mu}{r}(1 - G(b))(c_{NB} - c_B)\mathbb{I}_{S=S_B}$.

While the results using the alternative welfare measure are unambiguous, there are good reasons to focus our analysis on customer welfare, π_c, rather than on W. First, we set out to model trading during normal times, not resiliency of dealers in times of stress or the valuation of the implicit public subsidy. Our goal is to show the mechanism through which post-crisis bank regulations can impact customer welfare by changing the market structure for trading. Without explicitly modeling dealer bankruptcy or the formation of the public subsidy, using the difference between the bank and non-bank dealers' market-making costs to represent the implicit subsidy is clearly a simplification. In practice, part of the difference in market-making costs may arise for other reasons, such as operational efficiency, and we have no way in this model to evaluate the actual costs imposed on the public by the too-big-to-fail guarantee. Second, the bank dealer's profit stems from his market power. While regulators clearly see improving investor welfare during normal times as one of their goals, it is hard to imagine a situation in which regulators justify their actions by appealing to the need to maximize banks' market power rents. As such, we think that overall customer welfare remains the most natural metric with which to judge the impact of post-crisis regulations during normal times, and in Section 6 we use it specifically to evaluate the regulations' impact on investors in the corporate bond market.

5.2 Benevolent Bank Dealer

Why does customer welfare increase in region 2 (or Case 2) of the model? Competition from the non-bank dealer prevents the bank dealer from increasing the spread as regulatory costs rise, decreasing the rents it collects from the market-making business and incentivising a shift into the matchmaking business. The competition element is important because the bank dealer wields market power in his interaction with customers, and therefore it prices the market-making spread to extract economic rents. As we discuss in Section 3.1, it has long been established that dealers in the over-the-counter corporate bond market have market power. This market power is critical to delivering our results. To illustrate this point, in this section we discuss a variation of our model in which the bank dealer is benevolent. By benevolent we mean that the bank dealer maximizes overall customer welfare subject to the constraint that he is making zero profit on providing liquidity. We focus on the parameter range $c_B < c_{NB}$ (i.e., Case 1 and Case 2 in our model), where the bank dealer is the sole liquidity provider, because our goal is to scrutinize the result that customer welfare increases. The bank dealer's problem in this variation of the model is therefore

$$\max_{\substack{\mathcal{H} \in [0,1] \\ c_{NB} \geq S \geq f \geq 0}} \quad \pi_c(S, \mathcal{H}, f) = \frac{2\mu}{r} \left[\int_f^b (x-f)\, \mathcal{H} dG(x) + \int_b^\infty (x-S)\, dG(x) \right] \tag{32}$$

such that

$$\pi_B(S, \mathcal{H}, f, c_B) = \frac{2\mu}{r} \left[\mathcal{H} f (G(b) - G(f)) + (S - c_B)(1 - G(b)) \right] - K(\mathcal{H}) = 0. \tag{33}$$

We define the maximum customer welfare as $V(c_B) = \pi_c(S^\star(c_B), \mathcal{H}^\star(c_B), f^\star(c_B))$, where we use $(S^\star(c_B), \mathcal{H}^\star(c_B), f^\star(c_B))$ to denote the optimal solution to the above problem when the balance sheet cost is c_B.

Proposition 8. *If $K''(\mathcal{H}) > \beta_1$ for all $\mathcal{H} \in (0,1)$, then $V(c_B)$ is decreasing in c_B when $c_B < c_{NB}$.*[30]

The setting with a benevolent bank dealer who maximizes customer welfare and breaks even on liquidity provision delivers a stark result: customer welfare will always decrease when regulatory costs rise. The welfare result we obtain in our model is therefore attributable to the inefficiency introduced by the market power of over-the-counter bank dealers. Regulatory intervention in this case facilitates competition (from non-bank dealers) that improves efficiency and increases customer welfare.

5.3 An Alternative Formulation of the Bank Dealer's Cost of Matchmaking

In this section we consider an alternative formulation of the bank dealer's matchmaking cost, such that $K(\mathcal{H})$ is incurred by the bank dealer when he provides matchmaking services to each of the customers ($\int_{x=f}^b K(\mathcal{H}) dG(x)$). In other words, the bank dealer incurs a positive marginal cost for every customer who chooses matchmaking, and hence the term $[\mathcal{H}f - K(\mathcal{H})](G(b) - G(f))$ appears in the bank dealer's optimization problem.

[30] β_1 is defined in Proposition 1 and is needed here to rule out the unrealistic case of $\mathcal{H} = 1$.

Investigating this alternative formulation is worthwhile for two reasons. First, some readers may question the external validity of specifying a one-time cost for matchmaking speed.[31] In particular, we want to investigate whether our result regarding the increase in customer welfare in Case 2 is driven by the assumption that the cost of investing in matchmaking occurs at time 0 and does not increase with the number of customers who choose to utilize the matchmaking service. Second, the formulation in which the bank dealer incurs a cost for each customer has intuitive appeal in the context of a search model. It represents the effort the dealer invests in finding a counterparty, which can involve electronic systems as well as non-electronic means. Such efforts were costly in the days of telephone searches, and are still costly today even though investment in technology over the past several years has substantially reduced the cost of searching. Hence, this cost-of-effort formulation does not depend on any assumptions about the nature of the electronic systems that dealers may adopt.

We are interested in investigating the case in which the bank dealer provides both market-making and matchmaking services ($c_B < c_{NB}$), where our main model delivers the result that customer welfare increases when regulatory costs rise. The bank dealer's problem using the alternative formulation can be written as

$$\max_{c_{NB} \geq S \geq f \geq 0, \mathcal{H} \in [0,1]} \pi_B = \frac{2\mu}{r} \left[\left(G(b) - G(f) \right) \left[\mathcal{H} f - K(\mathcal{H}) \right] + \left(S - c_B \right) \left(1 - G(b) \right) \right] \tag{34}$$

and we characterize the equilibrium in the following proposition:

Proposition 9. *If $K(0) = 0$, $K'(0) = 0$, and $K(1) > c_{NB}$, then the equilibrium $(S_a, \mathcal{H}_a, f_a)$ exists.*

If the equilibrium is unconstrained, i.e., $S_a < c_{NB}$, then $(S_a, \mathcal{H}_a, f_a)$ satisfies

$$S_a - c_B - \mathcal{H}_a f_a + K(\mathcal{H}_a) = \frac{1 - G(b_a)}{G'(b_a)} (1 - \mathcal{H}_a), \tag{35}$$

$$f_a - \frac{K(\mathcal{H}_a)}{\mathcal{H}_a} = \frac{1 - G(f_a)}{G'(f_a)}, \tag{36}$$

$$\left(f_a - K'(\mathcal{H}_a) \right) \left(G(b_a) - G(f_a) \right) = \frac{S_a - c_B}{1 - \mathcal{H}_a} \left(1 - G(b_a) \right). \tag{37}$$

[31] We should note that specifying a one-time cost is equivalent to the bank dealer's paying a flow cost $K(\mathcal{H})dt$ at every instant dt. In the steady state, the bank dealer will choose a constant flow cost $\mathcal{H}$ for all t. The present value of this flow cost will therefore equal a constant times $K(\mathcal{H})$ and is identical to the one-time cost, as in our main model.

If the equilibrium is constrained, i.e., $S_a = c_{NB}$, then $(\mathcal{H}_a, f_a)$ satisfies

$$\mathcal{H}_a \left(G\left(b_a\right) - G\left(f_a\right)\right) - \left(\mathcal{H}_a f_a - K\left(\mathcal{H}_a\right)\right) G'\left(f_a\right) + \frac{\mathcal{H}_a}{1-\mathcal{H}_a} G'\left(b_a\right)\left(S_a - c_B - \mathcal{H}_a f_a + K\left(\mathcal{H}_a\right)\right) = 0, \tag{38}$$

$$\left(f_a - K'\left(\mathcal{H}_a\right)\right)\left(G\left(b_a\right) - G\left(f_a\right)\right) = \frac{S_a - c_B}{\left(1-\mathcal{H}_a\right)^2} G'\left(b_a\right)\left(S_a - c_B - \mathcal{H}_a f_a + K\left(\mathcal{H}_a\right)\right). \tag{39}$$

Unfortunately, we are unable to obtain interpretable comparative statics without imposing a specific distribution for G and a functional form for K. In the following analysis, we consider the comparative statics when G follows a uniform distribution and K is a quadratic function.[32]

Assumption 2. *$G(x)$ is uniformly distributed in the interval $[0, B]$, where $B \geq c_{NB}$. $K(\mathcal{H}) = \frac{1}{2} k \mathcal{H}^2$, where $k \geq 4 c_{NB}$.*

The assumption that $B \geq c_{NB}$ means that there exist customers who can afford the market-making service even if $S = c_{NB}$.

With Assumption 2, the optimization problem becomes

$$\begin{aligned}
\max_{\substack{c_{NB} \geq S \geq f \geq 0 \\ \mathcal{H} \in [0,1]}} \quad & \frac{2\mu}{r} \left[\left(G(b) - G(f)\right)\left[\mathcal{H} f - K(\mathcal{H})\right] + \left(S - c_B\right)\left(1 - G(b)\right)\right] \\
&= \frac{2\mu}{r} \left[\frac{b-f}{B}\left[\mathcal{H} f - K(\mathcal{H})\right] + \left(S - c_B\right)\left(1 - \frac{b}{B}\right)\right] \\
&= \frac{2\mu}{r} \frac{1}{B} \left[\frac{S-f}{1-\mathcal{H}}\left(\mathcal{H} f - K(\mathcal{H})\right) + \left(S - c_B\right)\left(B - \frac{S - \mathcal{H} f}{1-\mathcal{H}}\right)\right]
\end{aligned} \tag{40}$$

As we mention above, we focus on examining the robustness of the main contribution of our model: establishing that overall customer welfare can increase when the bank dealer's regulatory cost rises. This happens when the bank dealer is constrained (Case 2), and the following proposition provides the equivalent comparative statics for the constrained bank dealer case when we use the alternative formulation of bank dealer matchmaking cost.

Proposition 10. *Under Assumption 2, the unconstrained bank dealer equilibrium is obtained in the range $c_B \in (0, 2c_{NB} - B)$, and the constrained bank dealer equilibrium is obtained in the range $c_B \in (2c_{NB} - B, c_{NB})$. In the constrained equilibrium $(S_a, \mathcal{H}_a, f_a)$, if the bank*

[32] The uniform distribution satisfies Assumption 1 that we discuss in Section 4.2.

dealer's balance sheet cost c_B increases and all else are held equal, then the bank dealer's spread does not change ($S_a = c_{NB}$), the matching speed ($\mathcal{H}_a$) increases, the matchmaking fee (f_a) decreases, and overall customer welfare (π_c) increases.

We observe that the comparative statics in the constrained bank dealer case using the alternative formulation of matchmaking costs are similar to those we obtain in the main model. As the regulatory cost rises, the bank dealer has to keep the spread equal to the balance sheet cost of the non-bank dealer and hence the profitability of his market-making business declines. Therefore, he expends greater effort (i.e., invests in higher matchmaking speed) and lowers the matchmaking fee, resulting in higher overall customer welfare, exactly as in our main model.

Why then in our main model do we utilize the one-time cost of investing in the matchmaking technology rather than the per-trade cost of effort? We believe there is added value in solving the model in closed form and obtaining interpretable expressions for the comparative statics with a general distribution for customer types and a general functional form for the cost of matchmaking. Alas, we are unable to obtain interpretable expressions for the comparative statics when using the alternative formulation of the bank dealer's matchmaking costs without using specific G and K. Hence, we use the general functions with the one-time investment formulation as our main model, and provide the cost-of-effort formulation with a specific distribution and a specific cost function to show the robustness of our result pertaining to customer welfare.

6 Did Post-Crisis Bank Regulations Harm Investors in the Corporate Bond Market?

We can use the model's insights to evaluate whether bank regulations following the financial crisis lowered the overall welfare of investors. Given the wide availability of empirical evidence regarding the US corporate bond market, we focus our discussion on this market, but the model's insights would apply similarly to other markets affected by post-crisis bank regulations. As we stress in the introduction, empirical estimates of transaction-based measures of liquidity cannot capture investors' loss of welfare when finding a trading counterparty takes more time or when they forgo trading to avoid paying the high cost of immediacy. In contrast, our theoretical model takes these potential welfare losses into account alongside the out-of-pocket cost of trading. Our analysis shows that the implications of bank regulations for overall customer welfare depend on the manner in which competition between the bank and non-bank dealers shapes the equilibrium. Hence, to evaluate the impact of post-crisis regulations on welfare we need to tie the equilibrium regions in our model to the state of affairs in the corporate bond market around the enactment of the bank regulations. In this section we attempt to do so by mapping the implications of our model for observable market outcomes onto the findings reported in the empirical literature.

The first set of results involves price-based measures: average transaction costs and the cost of immediacy. The general finding in the empirical literature is that average transaction costs in the form of bid-ask spreads or price impacts 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 (2017)) following the enactment of post-crisis regulations. On the other hand, 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)).

Referring back to the example depicted in Figure 1, the blue line in Panel A traces the cost of immediacy (trading via the market-making mechanism) as we increase bank regulatory costs (c_B) on the x-axis. We observe that the market-making spread increases in region 1, remains constant in region 2 (where it is constrained by the market-making cost of the non-bank dealer), increases further in region 3 (where competitive pressure from the bank dealer is easing), and is constant again in region 4 as the non-bank dealer charges his unconstrained optimal spread. The matchmaking fee represented by the red line is constant in region 1, declines in region 2 (where competition from the non-bank dealer makes it optimal for the bank dealer to substantially increase his investment in matchmaking technology), increases somewhat in region 3 (where the non-bank dealer's market-making spread rises, enabling the bank dealer to extract more matchmaking rents) and remains constant at the unconstrained maximum in region 4. Average transaction costs, which take into account both the price charged to customers and the amount of trading in each mechanism, are shown by the black dotted line. They increase in region 1, decline in region 2, increase in region 3, and remain unchanged in region 4.

The empirical findings of an increase in the cost of immediacy together with a decline in average transaction costs around the enactment of post-crisis bank regulations must mean we have moved from the right portion of region 1 to region 2 (or from the left portion of region 2 to the left portion of region 3). Therefore, the key to observing both an increase in the cost of immediacy and a decline in average transaction costs is that the market-making costs of bank dealers move through region 2, where the gains from lower matchmaking fees materialize.[33] Propositions 2, 4, and 6 show that this is a robust implication of our model that is not tied to a particular example.

Panel B plots the fraction of traders choosing the market-making mechanism (light blue), trading via the matchmaking mechanism (light green), or refraining from trade altogether (red). A transition from region 1 to region 2 (or the left portion of region 3) implies a large increase in matchmaking volume accompanied by a decline in market making. These patterns also match the empirical findings: an increase in matchmaking volume driven by bank dealers (Bao, O'Hara, and Zhou (2018), Choi and Huh (2017), Schultz (2017)) and a significant decline in capital commitment to market making by bank dealers that is not sufficiently offset by an increase in market making by non-bank dealers (Bao, O'Hara, and Zhou (2018), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018)).[34]

Proposition 4 predicts an increase in the overall volume of trading in region 2, where customers who refrain from trading in region 1 are drawn to the market by lower matchmaking fees and higher matching speed. The empirical evidence regarding volume is somewhat nuanced. Overall trading volume in bonds has significantly increased, as our model predicts, 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)). Some market observers note that the 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, resulted in decreased turnover in each bond issue (BIS Committee on the Global Financial System (2014)). In other words, the higher observed overall volume could be driven by lower average transaction costs that attract more customers to trade bonds (as in our model), but the record number of new issues means that each bond issue would be traded less often.

Summarizing our discussion, the empirical findings relating to the implementation of bank regulations in the aftermath of the financial crisis all point to the particular role that

[33]The figure also reflects the intuitive result that the fee charged to customers in the matchmaking mechanism is lower than the cost of trading in the market-making mechanism, which is consistent with the empirical evidence reported in Choi and Huh (2017) and Schultz (2017).

[34]In the model, there is a clear distinction between region 2, where the bank dealer captures the entire market-making business, and region 3, where the non-bank dealer accommodates all customers who demand immediacy. We would expect a more gradual transition in reality. The empirical evidence of an increase (decline) in market-making activity by non-bank (bank) dealers following the enactment of post-crisis regulations suggests that we are probably rather close to the border between region 2 and region 3 (where the cost differential between the bank and non-bank dealers is very small).

region 2 plays in our model. Perhaps we were in region 1 in the pre-crisis period and ended up in region 2 (or the left portion of region 3) after bank regulations were imposed. We might also have started in the left portion of region 2 and ended up further to the right in region 2 (or the left portion of region 3). Either way, the empirical evidence suggests that the regulations caused a move across region 2. With this in mind, we can now consider our model's predictions for overall customer welfare.

Panel C plots overall customer welfare (the dotted black line) as well as the welfare of customers who utilize market making (red line) and matchmaking (blue line). We observe for this particular example that welfare declines in region 1, increases in region 2, and declines again in region 3. Propositions 2, 4, and 6 show that welfare can actually start increasing already in the right portion of region 1 and continue to increase in the left portion of region 3 (before starting to decline) for some combination of parameters. Irrespective of parameter choices, however, overall customer welfare unequivocally increases as the regulatory costs of bank dealers rise throughout region 2.

The clear mapping of the empirical findings to an increase in regulatory costs across region 2 in the model coupled with our welfare implications indicate that investors in the corporate bond market are likely better off under the current bank regulatory regime compared with the one that prevailed before the financial crisis. We cannot say whether a particular feature of the Basel framework is responsible for this result and our results do not imply that the Volcker Rule improved liquidity. Rather, considering the empirical findings on the impact of the entire suite of post-crisis regulations, our model would suggest that the increase in market-making costs benefits corporate bond investors even during normal times. We believe that this insight is an important contribution of our work to the debate over the merits of bank regulations that were imposed in the aftermath of the financial crisis.

7 Concluding Remarks

Our paper highlights the complex and multifaceted consequences that post-crisis bank regulations have for market liquidity and investor welfare. While the explicit goal of those regulations is to enhance financial stability in times of stress, our work shows that these regulations can improve market efficiency even during normal times by prompting a change in the nature of liquidity provision. Pre-crisis, bank dealers underinvest in matchmaking technology because they wish to keep their profitable market-making business, even though such investment would benefit investors. Post-crisis regulations eliminate obstacles to competition in the most profitable business (market making) and incentivize bank dealers to invest in the less profitable but more efficient market structure (matchmaking). The industrial organization angle combines with the market microstructure angle to deliver this positive outcome.

The key insight we offer in this paper is that an increase in regulatory costs imposed on banks can make customers in the corporate bond market better off. Specifically, our focus is on balance sheet costs that were imposed on banks following the financial crisis to deter proprietary trading and as a result increased the costs of market making. A notable aspect of our work is that we do not focus on whether these costs reduce the likelihood or severity of a crisis, which are presumably the stated goals of the regulation. Rather, we show that the increase in regulatory costs can lower average transaction costs during normal times and, more importantly, increase the overall welfare of customers. The driving forces behind our results are four elements that characterize the corporate bond market: the coexistence of two distinct trading mechanisms (market making and matchmaking), the market power enjoyed by bank dealers, potential market-making competition from non-bank dealers, and the investment in technology that is reshaping the trading environment. Although our paper is motivated by and specifically speaks to observations in the corporate bond market, the model we present can be applied to other over-the-counter markets that feature these four elements.

How can customers be made better off by regulations that deter market making? The rationale is that while market making is a very profitable form of intermediation for bank dealers, it can be an expensive form of intermediation for customers. When regulation increases the cost of market making for bank dealers and competition prevents them from transferring the cost increase to their customers, bank dealers are incentivized to develop their matchmaking business. They do so by lowering fees and investing in technology that makes matchmaking more efficient, thereby rendering it more attractive to customers. The increase in welfare of customers who obtain better execution using the matchmaking mechanism can be greater than the decline in the welfare of customers who must pay more for market making when regulation increases its cost. This aggregate increase in welfare is not a Pareto improvement—some customers are worse off—but the overall population of customers benefits.

Regulators always worry about the unintended consequences of their regulatory interventions. While increasing the costs of market making was meant to enhance bank dealers' resiliency in times of market stress, we believe that one (perhaps) "unintended" benefit of these bank regulations was to push bank dealers to invest in matchmaking. This can improve overall customer welfare during normal times and therefore materially changes the supposed tradeoff between resiliency in times of stress and day-to-day liquidity. The market microstructure angle in this case is crucial to delivering this result. It is also important to emphasize that the improvement in customer welfare in our model likely understates the extent of the true effect. Our model intentionally shuts off one of the routes through which customers are made better off when they switch to matchmaking. Specifically, bank dealers in our model directly specify the matching rate by their investment in technology. While this simplification allows us to present analytic solutions, waiting costs in a more general search model would also fall as more customers switch to trading through the matchmaking mechanism, further enhancing overall customer welfare.

Raising regulatory costs beyond a certain level would inevitably worsen welfare. A natural question is whether we can help regulators identify the optimal level of such costs. While one could calibrate our model with specifics pertaining to observable quantities to arrive at a prescription for the optimal level of regulatory costs, we caution against taking an admittedly simplified theoretical model and requiring it to provide such a prescription. Nonetheless, our model does point to an approach that could help bank regulators. In particular, we show that beneficial investment in matchmaking technology is boosted by competition in market making as the cost differential between bank and non-bank dealers shrinks. Bank regulators can therefore consider the market-making costs of non-bank dealers to help them judge the level they should impose on bank dealers to ensure sufficient competition.

Five regulatory agencies—the Fed, the FDIC, the Office of the Comptroller of the Currency, the SEC, and the CFTC—have recently adopted the revised Volcker Rule.[35] The revision was intended, among other things, to "provide banking entities with clarity about what activities are prohibited." While our model is not about the Volcker Rule per se, we stress that simply lowering the regulatory costs of bank dealers need not help liquidity in the corporate bond market nor would it necessarily improve customer welfare. Our analysis suggests that, in addition to examining the market-making costs of non-bank dealers as a benchmark, regulators should scrutinize their proposed revisions to ensure that they do not induce bank dealers to reduce their investment in matchmaking technology. In fact, regulators could specifically incentivize further investments in matchmaking to help ensure that customers continue to benefit even as bank regulatory costs decline. It is this investment, and the partial shift from market making to matchmaking, that is key in our model

[35] See https://www.sec.gov/rules/final/2019/bhca-7.pdf.

to generating higher overall customer welfare.

All theoretical models employ simplifications in the process of creating the appropriate structure and deriving the results. Our model is no different. Aiming for robustness and clarity while obtaining closed-form solutions necessitates abstracting from some features of the economic environment. One example is that we do not model the matchmaking business of non-bank dealers. Several other such features—intermediation chains, dealer inventory, the nature of dealer funding, and the moral hazard problem that arises when a customer employs an agent for searching—have recently been discussed in other papers (An, Song, and Zhang (2017), An and Zheng (2017), Li and Li (2017), and Cimon and Garriott (2019)). Partially for that reason we forgo incorporating them in our model (e.g., we abstract from the inventory consideration by using the same arrival rates for buyers and sellers). Instead, we focus on dimensions we deem critical to understanding how customer welfare in the corporate bonds market changes when regulation increases the costs of market making by bank dealers: the dual trading mechanisms, dealer market power, and the nature of competition between bank and non-bank dealers. The resulting model provides insights that we hope capture the more salient trade-offs. Nonetheless, integrating additional economic features and specifying a more general search model in which the size of the pool of searching customers impacts the match rate could be valuable extensions of the work we pursue in this paper.

Crucial to our analysis are customers who optimize over several key choices—from the basic decision of whether to trade to the tradeoff between search time and the cost of execution in the two trading mechanisms—but we recognize that we could not capture all aspects of customer behavior. For example, increased market-making costs could incentivize customers to shift their trading to newly issued securities to minimize transaction costs. Furthermore, the low interest-rate environment following the financial crisis encouraged new bond issues and made such substitutions easier to achieve. Such coping strategies may limit portfolio flexibility and adversely affect welfare. Similarly, customers may choose to reduce their average trade size in response to the higher cost of market making. The implications of a reduction in trade size for customer welfare are ambiguous: larger trades have traditionally exhibited lower transaction costs in the corporate bonds market but the increase in electronic trading could make order-splitting optimal (as has been the case in equity markets). We believe that these aspects of customer behavior (choosing newer issues or a smaller trade size) play a lesser role in determining customer welfare than the pronounced shift from market making to matchmaking, but nonetheless they suggest a complexity to the manner in which customers respond to changes in their economic environment.

The evolving regulatory frameworks and the breathtaking pace at which technology impacts securities markets continue to dominate the agenda of regulators, practitioners, and academics. We hope that our work—studying how regulation impacts incentives to invest in technology and as a consequence alters the market structure for trading securities—would serve to both highlight important tradeoffs and spur additional work on the changing nature of our securities markets.

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.

Aubin, J., and H. Frankowska. 2009. Set-Valued Analysis. *Springer*.

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.

Anand, A., C. Jotikasthira, and K. Venkataraman. 2018. Do Buy-Side Institutions Supply Liquidity in Bond Markets? Evidence from Mutual Funds. doi:10.2139/ssrn.3003189. Working Paper.

Anderson, M., and R. Stulz. 2017. Is Post-Crisis Bond Liquidity Lower? doi:10.3386/w23317. Working Paper.

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. Working Paper.

———. 2016. Fixed income market liquidity. Working Paper.

BIS Markets Committee. 2016. Electronic trading in fixed income markets. 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. doi:10.17016/FEDS.2017.116. Working Paper.

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

Davis Polk. 2014. Supplementary Leverage Ratio. Memo.

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

Duffie, D. 2012. Market Making Under the Proposed Volcker Rule. doi:10.2139/ssrn.1990472. Working Paper.

Duffie, D. 2018. Post-Crisis Bank Regulations and Financial Market Liquidity. Baffi Lecture.

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

Ederington, L. H., W. Guan, and P. K. Yadav. 2014. Dealer Spreads in the Corporate Bond Market: Agent vs. Market-Making Roles. doi:10.2139/ssrn.2378000. Working Paper.

Goldstein, M. A., and E. S. Hotchkiss. 2017. Providing Liquidity in an Illiquid Market: Dealer Behavior in U.S. Corporate Bonds. doi:10.2139/ssrn.2977635. Working Paper.

Greenwood, R., S. G. Hanson, J. C. Stein, and A. Sunderam. 2017. Strengthening and Streamlining Bank Capital Regulation. *Brookings Papers on Economic Activity*.

Li, J., and W. Li. 2017. Agency Trading and Principal Trading. doi:10.2139/ssrn.2760457. Working Paper.

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

Osborne, M. J., and A. Rubinstein. 1994. A Course in Game Theory. *The MIT Press*.

Randall, O. 2015. How Do Inventory Costs Affect Dealer Behavior in the US Corporate Bond Market? doi:10.2139/ssrn.2590331. Working Paper.

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. doi:10.2139/ssrn.2966919. Working Paper.

Trebbi, F., and K. Xiao. 2017. Regulation and Market Liquidity. doi:https://doi.org/10.1287/mnsc.2017.2876. *Management Science*.

Appendix

A Proofs of Propositions 1–4

A.1 Proof of Proposition 1

First, we want to show that $\mathcal{H}_1^\star > 0$. If the bank dealer chooses $\mathcal{H}_1^\star = 0$, then it has zero matchmaking capacity and its profit is invariant to $f_1^\star$. Suppose, in this case, the bank dealer chooses $(S_{NM}^\star, \mathcal{H}^\star = 0, f^\star)$, where "NM" means no matching. Now let's think about $\frac{\partial \Pi_1\left(S_{NM}^\star, \mathcal{H}, f^\star\right)}{\partial \mathcal{H}}|_{\mathcal{H}=0}$. If this derivative is positive for some reasonable $f^\star$, then it is profitable for the bank dealer to choose a positive $\mathcal{H}$. It is easy to show that

$$\frac{\partial \Pi_1\left(S_{NM}^\star, \mathcal{H}, f^\star\right)}{\partial \mathcal{H}}|_{\mathcal{H}=0} \propto f^\star\left(G\left(b^\star\right) - G\left(f^\star\right)\right) - \left(S_{NM}^\star - c_B\right) G'\left(b^\star\right)\left(S_{NM}^\star - f^\star\right), \tag{41}$$

where $b^\star$ is the cutoff value b evaluated at $(S_{NM}^\star, \mathcal{H}^\star, f^\star)$. In this case $b^\star = S_{NM}^\star$ because we set $\mathcal{H}^\star = 0$. Then with $K'(0) = 0$, we have

$$\begin{aligned}\frac{\partial \Pi_1\left(S_{NM}^\star, \mathcal{H}, f^\star\right)}{\partial \mathcal{H}}|_{\mathcal{H}=0} &\propto f^\star\left(G\left(S_{NM}^\star\right) - G\left(f^\star\right)\right) - \left(S_{NM}^\star - c_B\right) G'\left(S_{NM}^\star\right)\left(S_{NM}^\star - f^\star\right) \\ &\propto \left(S_{NM}^\star - f^\star\right)\left(f^\star \frac{G\left(S_{NM}^\star\right) - G\left(f^\star\right)}{S_{NM}^\star - f^\star} - \left(S_{NM}^\star - c_B\right) G'\left(S_{NM}^\star\right)\right).\end{aligned}$$

We only need to find one $f^\star$ such that the above equation is positive. Now let's set $f^\star \uparrow S_{NM}^\star$, in which case $\left(S_{NM}^\star - f^\star\right) > 0$, $\frac{G\left(S_{NM}^\star\right) - G(f^\star)}{S_{NM}^\star - f^\star} \to G'\left(S_{NM}^\star\right)$, and $f^\star > S_{NM}^\star - c_B$. Thus, in this case, $\frac{\partial \Pi_1\left(S_{NM}^\star, \mathcal{H}, f^\star\right)}{\partial \mathcal{H}}|_{\mathcal{H}=0} > 0$. So it is never optimal to choose $\mathcal{H} = 0$.

With the assumption $K'' > \beta_1$, we can show that it is not optimal to choose $\mathcal{H}_1^\star = 1$ (if $\mathcal{H}_1^\star = 1$, the bank dealer's profit would be negative). So the optimal solution must satisfy $\mathcal{H}_1^\star \in (0, 1)$, i.e., $\mathcal{H}_1^\star$ is an interior solution.

To characterize the equilibrium, let's make the following change of variables

$$\begin{bmatrix} S \\ f \\ \mathcal{H} \end{bmatrix} \to \begin{bmatrix} b \\ f \\ \mathcal{H} \end{bmatrix} \tag{42}$$

where $b = \frac{S - f\mathcal{H}}{1 - \mathcal{H}}$. It is easy to show that $S = (1 - \mathcal{H})\, b + f\mathcal{H}$. The optimization problem now becomes

$$\begin{aligned}
&\max_{S\geq f\geq 0,\mathcal{H}\in[0,1]} \frac{2\mu}{r}\left[\mathcal{H}f\left(G\left(b\right)-G\left(f\right)\right)+\left(S-c_B\right)\left(1-G\left(b\right)\right)\right]-K\left(\mathcal{H}\right)\\
\iff &\max_{b\geq f\geq 0,\mathcal{H}\in[0,1]} \frac{2\mu}{r}\left[\mathcal{H}f\left(G\left(b\right)-G\left(f\right)\right)+\left(\left(1-\mathcal{H}\right)b+f\mathcal{H}-c_B\right)\left(1-G\left(b\right)\right)\right]-K\left(\mathcal{H}\right).
\end{aligned} \tag{43}$$

It can be shown that if $b\leq f$, we must have $S\leq f$, and in this case the value from the above formula must be less than or equal to the case in which the bank dealer offers only the market-making service, and we have already proved that it is not optimal for the bank dealer to abandon matchmaking completely. So without loss of generality, we just need to consider the following relaxed problem:

$$\max_{b\geq 0,f\geq 0,\mathcal{H}\in[0,1]} \frac{2\mu}{r}\left[\mathcal{H}f\left(G\left(b\right)-G\left(f\right)\right)+\left(\left(1-\mathcal{H}\right)b+f\mathcal{H}-c_B\right)\left(1-G\left(b\right)\right)\right]-K\left(\mathcal{H}\right). \tag{44}$$

Rewriting the problem, we obtain

$$\begin{aligned}
&\max_{b\geq 0,f\geq 0,\mathcal{H}\in[0,1]} \frac{2\mu}{r}\left[\mathcal{H}f\left(G\left(b\right)-G\left(f\right)\right)+\left(\left(1-\mathcal{H}\right)b+f\mathcal{H}-c_B\right)\left(1-G\left(b\right)\right)\right]-K\left(\mathcal{H}\right)\\
\iff &\max_{b\geq 0}\max_{f\geq 0,\mathcal{H}\in[0,1]} \frac{2\mu}{r}\left[\mathcal{H}f\left(G\left(b\right)-G\left(f\right)\right)+\left(\left(1-\mathcal{H}\right)b+f\mathcal{H}-c_B\right)\left(1-G\left(b\right)\right)\right]-K\left(\mathcal{H}\right)\\
\iff &\max_{b\geq 0}\tilde{\Pi}_1\left(b\right),
\end{aligned} \tag{45}$$

where

$$\begin{aligned}
\tilde{\Pi}_1\left(b\right) &= \max_{f\geq 0,\mathcal{H}\in[0,1]} \frac{2\mu}{r}\left[\mathcal{H}f\left(G\left(b\right)-G\left(f\right)\right)+\left(\left(1-\mathcal{H}\right)b+f\mathcal{H}-c_B\right)\left(1-G\left(b\right)\right)\right]-K\left(\mathcal{H}\right)\\
&= \max_{f\geq 0,\mathcal{H}\in[0,1]} \tilde{\Pi}\left(f,\mathcal{H};b\right).
\end{aligned} \tag{46}$$

Let's consider the first-order derivative of the above equation:

$$\frac{\partial\tilde{\Pi}\left(f,\mathcal{H};b\right)}{\partial f} = \frac{2\mu}{r}\mathcal{H}\left[\left(G\left(b\right)-G\left(f\right)\right)-fG'\left(f\right)+1-G\left(b\right)\right] \tag{47}$$

$$= \frac{2\mu}{r}\mathcal{H}\left[1-G\left(f\right)-fG'\left(f\right)\right] \tag{48}$$

$$= \frac{2\mu}{r}\mathcal{H}G'\left(f\right)\left[\frac{1-G\left(f\right)}{G'\left(f\right)}-f\right]. \tag{49}$$

Given the assumption $\Phi\geq 0$, we know $\frac{1-G(f)}{G'(f)}$ is weakly decreasing in f, $\frac{\partial\tilde{\Pi}(f,\mathcal{H};b)}{\partial f}|_{f=0}>0$, and $\frac{\partial\tilde{\Pi}(f,\mathcal{H};b)}{\partial f}|_{f=\infty}<0$. So, there exists a unique $f_1^\star>0$ such that $\frac{\partial\tilde{\Pi}(f,\mathcal{H};b)}{\partial f}>0$ if $f<f_1^\star$

and $\frac{\partial\tilde{\Pi}(f,\mathcal{H};b)}{\partial f}<0$ if $f>f_1^\star$. And $f_1^\star$ is independent of $(b,\mathcal{H})$. Thus, $f_1^\star$ is the global optimal value. Then

$$\tilde{\Pi}_1(b)=\max_{\mathcal{H}\in[0,1]}\frac{2\mu}{r}\left[\mathcal{H}f_1^\star\left(G(b)-G(f_1^\star)\right)+\left((1-\mathcal{H})b+f_1^\star\mathcal{H}-c_B\right)(1-G(b))\right]-K(\mathcal{H}). \tag{50}$$

And the first-order derivative of $\tilde{\Pi}(f_1^\star,\mathcal{H};b)$ with respect to $\mathcal{H}$ is

$$\frac{\partial\tilde{\Pi}(f_1^\star,\mathcal{H};b)}{\partial\mathcal{H}}=\frac{2\mu}{r}\left[f_1^\star\left(G(b)-G(f_1^\star)\right)+\left(-b+f_1^\star\right)(1-G(b))\right]-K'(\mathcal{H}). \tag{51}$$

We know that K is strictly convex, so there must exist a unique $\mathcal{H}^\star(f_1^\star,b)$, such that $\frac{\partial\tilde{\Pi}(f,\mathcal{H};b)}{\partial\mathcal{H}}|_{f=f_1^\star}>0$ if and only if $\mathcal{H}<\mathcal{H}^\star(f_1^\star,b)$, where $\mathcal{H}^\star(f_1^\star,b)$ is solved by

$$K'(\mathcal{H}^\star)=\max\left\{\frac{2\mu}{r}\left[f^\star\left(G(b)-G(f_1^\star)\right)+\left(-b+f_1^\star\right)(1-G(b))\right],0\right\} \tag{52}$$

$$=\max\left\{\frac{2\mu}{r}\left[f_1^\star\left(1-G(f_1^\star)\right)-b\,(1-G(b))\right],0\right\}. \tag{53}$$

Then

$$\tilde{\Pi}_1(b)=\frac{2\mu}{r}\left[\mathcal{H}^\star f_1^\star\left(G(b)-G(f_1^\star)\right)+\left((1-\mathcal{H}^\star)b+f_1^\star\mathcal{H}^\star-c_B\right)(1-G(b))\right]-K(\mathcal{H}^\star). \tag{54}$$

By the envelope theorem,

$$\frac{d\tilde{\Pi}_1(b)}{db}=\frac{2\mu}{r}G'(b)(1-\mathcal{H}^\star)\left[\frac{c_B}{1-\mathcal{H}^\star}-\left(b-\frac{1-G(b)}{G'(b)}\right)\right]. \tag{55}$$

If $\left(b-\frac{1-G(b)}{G'(b)}\right)\leq 0$, we would have $\frac{d\tilde{\Pi}_1(b)}{db}>0$. In particular, at $b=0$, $b-\frac{1-G(b)}{G'(b)}=-\frac{1}{G'(0)}<0$ and $\frac{d\tilde{\Pi}_1(b)}{db}|_{b=0}>0$. In the region $\left(b-\frac{1-G(b)}{G'(b)}\right)>0$, we can show that $\frac{c_B}{1-\mathcal{H}^\star}$ is non-decreasing in b. To see this, note that the expression $\frac{\partial\tilde{\Pi}\left(f_1^\star,\mathcal{H};b\right)}{\partial\mathcal{H}}$ is equal to zero at the optimal solution. We can then take the total derivative again to obtain

$$\frac{r}{2\mu}K''(\mathcal{H}^\star)\,d\mathcal{H}+\left[\frac{1-G(b)}{G'(b)}-b\right]G'(b)\,db=0\Longleftrightarrow\frac{d\mathcal{H}^\star}{db}>0. \tag{56}$$

Then $c_B/(1-\mathcal{H})$ is also increasing in $\mathcal{H}$.

Lemma 2 below shows that if $K''(\cdot)\geq\beta_1$, then

$$\frac{d\left[\frac{c_B}{1-\mathcal{H}^\star(b)}\right]}{db}<\frac{d\left[b-\frac{1-G(b)}{G'(b)}\right]}{db}\quad\text{if }\left(b-\frac{1-G(b)}{G'(b)}\right)>0, \tag{57}$$

and if $\lim_{b\to\infty} \frac{d\tilde{\Pi}_1(b)}{db} < 0$, then there must exist a unique $b^\star$ such that $\frac{d\tilde{\Pi}_1(b)}{db} > 0$ iff $b < b^\star$, because

$$\frac{d\tilde{\Pi}_1(b)}{db} > 0 \Longleftrightarrow \frac{c_B}{1-\mathcal{H}^\star} - \left(b - \frac{1-G(b)}{G'(b)}\right) > 0.$$

Define $y(b) = \frac{c_B}{1-\mathcal{H}^\star} - \left(b - \frac{1-G(b)}{G'(b)}\right)$. First we know that $b - \frac{1-G(b)}{G'(b)}$ is increasing in b, so there exists a threshold b_1 such that $\left\{b | b - \frac{1-G(b)}{G'(b)} \le 0\right\} = [0, b_1]$. (Actually $b_1 = f_1^\star$, as shown before.) Then $y(b) > 0$ when $b \in [0, b_1]$. In region $b \in (b_1, \infty)$, we know $b - \frac{1-G(b)}{G'(b)} > 0$. With the assumption $K'' > \beta_1$, Lemma 2 below shows that $y'(b) < 0$ if $b \in (b_1, \infty)$. From the assumption $K'' > \beta_1$, we know there exists an $\tilde{\mathcal{H}} < 1$ such that the optimal $\mathcal{H}^\star$ satisfies $\mathcal{H}^\star < \tilde{\mathcal{H}} < 1$. Then as $b \to \infty$, $\frac{c_B}{1-\mathcal{H}^\star}$ must be bounded. Moreover, we know both b and $-\frac{1-G(b)}{G'(b)}$ are increasing in b, and $-\frac{1-G(0)}{G'(0)}$ is bounded, so $b - \frac{1-G(b)}{G'(b)} \to \infty$ as $b \to \infty$. Then we must have $y(\infty) = -\infty$. Together with the conditions $y(b_1) > 0$ and $y'(b) < 0$ for all $b \in (b_1, \infty)$, there must exist a $b_1^\star$ such that $y(b) > 0$ if and only if $b < b_1^\star$, which is equivalent to $\frac{d\tilde{\Pi}_1(b)}{db} > 0$ if and only if $b < b_1^\star$.

Therefore, the global optimal solution is $(b_1^\star, f_1^\star, \mathcal{H}^\star(f_1^\star, b_1^\star))$ and, from the above derivation, we know that the global optimal solution can be solved by the first-order approach. Because of the one-to-one mapping from $(S, f, \mathcal{H})$ to $(b, f, \mathcal{H})$, we also know that the global optimal solution to the original problem can also be solved by the first-order approach and the solution is unique.

Lemma 2. *Let β_1 be the unique solution of*

$$\frac{r}{2\mu}\beta = \frac{1}{\left(1 - \sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta}}\right)^2} c_B \max_b \{bG'(b)\}. \tag{58}$$

If $K''(\mathcal{H}) \ge \beta_1$ for all $\mathcal{H}$, then

$$\frac{d\left[\frac{c_B}{1-\mathcal{H}^\star(b)}\right]}{db} < \frac{d\left[b - \frac{1-G(b)}{G'(b)}\right]}{db} \tag{59}$$

for all b that satisfies $\left(b - \frac{1-G(b)}{G'(b)}\right) > 0$.

Proof. The LHS is

$$\begin{aligned} \frac{d\left[\frac{c_B}{1-\mathcal{H}^\star(b)}\right]}{db} &= \frac{c_B}{(1-\mathcal{H}^\star)^2}\frac{d\mathcal{H}^\star}{db} \\ &= \frac{c_B}{(1-\mathcal{H}^\star)^2}\frac{bG'(b) - (1-G(b))}{\frac{r}{2\mu}K''(\mathcal{H}^\star)}. \end{aligned}$$

The RHS is

$$\frac{d\left[b-\frac{1-G(b)}{G'(b)}\right]}{db}=2+\frac{\left(1-G\left(b\right)\right)G''\left(b\right)}{\left(G'\left(b\right)\right)^2}.$$

Then

$$\frac{c_B}{\left(1-\mathcal{H}^\star\right)^2}\frac{bG'\left(b\right)-\left(1-G\left(b\right)\right)}{\frac{r}{2\mu}K''\left(\mathcal{H}\right)}<2+\frac{\left(1-G\left(b\right)\right)G''\left(b\right)}{\left(G'\left(b\right)\right)^2}$$

$$\iff\quad \frac{r}{2\mu}K''\left(\mathcal{H}^\star\right)\left(1-\mathcal{H}^\star\right)^2\geq c_B\frac{bG'\left(b\right)-\left(1-G\left(b\right)\right)}{2+\frac{\left(1-G(b)\right)G''(b)}{\left(G'(b)\right)^2}}<c_B bG'\left(b\right). \tag{60}$$

Thus, a sufficient condition for the desired inequality is

$$\frac{r}{2\mu}K''\left(\mathcal{H}^\star\right)\left(1-\mathcal{H}^\star\right)^2\geq c_B\max_b bG'\left(b\right). \tag{61}$$

Now we want to show that (61) is implied by $K''\geq\beta_1$. Given $K(0)=0$ and $K^{'}(0)=0$, for any β such that $K^{''}>\beta$, we must have $K(\mathcal{H})>\frac{1}{2}\beta\mathcal{H}^2$. The total profit from the bank dealer's liquidity provision in market making and matchmaking is less than $\frac{2\mu}{r}2\bar{c}$ (a customer pays the bank dealer no more than the non-bank dealer's cost $c_{NB}<\bar{c}$). Then in any equilibrium, we must have $\frac{1}{2}\beta\mathcal{H}^2<\frac{2\mu}{r}2\bar{c}$ (the bank dealer's profit is non-negative), which is $\mathcal{H}<\sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta}}$. This implies that, if $K^{''}>\beta$, then $\mathcal{H}^\star<\sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta}}$. Now assume that $K''>\beta_1$. Then the LHS of (61) becomes $\frac{r}{2\mu}K''\left(\mathcal{H}^\star\right)\left(1-\mathcal{H}^\star\right)^2>\frac{r}{2\mu}\beta_1\left(1-\sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta_1}}\right)^2=c_B\max_b bG'(b)$, that is, (61) is satisfied.

□

A.2 Proof of Proposition 2

1. **Spread, fee, and speed.** Because $f_1^\star-\frac{1-G\left(f_1^\star\right)}{G'\left(f_1^\star\right)}=0$, $f_1^\star$ is a constant.

   Consider the first-order condition $\frac{d\tilde{\Pi}_1(b)}{db}=0$, which is

$$\frac{c_B}{1-\mathcal{H}^\star}-\left(b_1^\star-\frac{1-G\left(b_1^\star\right)}{G'\left(b_1^\star\right)}\right)=0. \tag{62}$$

   Taking the total derivative of the above equality, we have

$$dc_B\frac{1}{1-\mathcal{H}^\star}+c_B\frac{d\frac{1}{1-\mathcal{H}^\star}}{db^\star}db^\star-\frac{d\left(b_1^\star-\frac{1-G\left(b_1^\star\right)}{G'\left(b_1^\star\right)}\right)}{db^\star}db^\star=0. \tag{63}$$

From Lemma 2 we know

$$c_B \frac{d\frac{1}{1-\mathcal{H}^\star}}{db^\star} < \frac{d\left(b_1^\star - \frac{1-G\left(b_1^\star\right)}{G'\left(b_1^\star\right)}\right)}{db^\star}. \tag{64}$$

Rearranging the expressions in the total derivative, we have $\frac{db^\star}{dc_B} > 0$.

We already show that $\frac{d\mathcal{H}^\star}{db} > 0$, so $\mathcal{H}_1^\star$ also increases.

Moreover, because the function $x - \frac{1-G(x)}{G'(x)}$ is increasing in x, we know that $b_1^\star > f_1^\star$.

We know $S_1^\star = (1-\mathcal{H}_1^\star)\, b_1^\star + f_1^\star \mathcal{H}_1^\star$, in which case

$$\begin{aligned}\frac{dS_1^\star}{db_1^\star} &= (1-\mathcal{H}_1^\star) - b_1^\star \frac{d\mathcal{H}_1^\star}{db_1^\star} + f_1^\star \frac{d\mathcal{H}_1^\star}{db_1^\star} \\ &= (1-\mathcal{H}_1^\star) + (f_1^\star - b_1^\star)\, \frac{b_1^\star G'\left(b_1^\star\right) - \left(1-G\left(b_1^\star\right)\right)}{\frac{r}{2\mu} K''\left(\mathcal{H}_1^\star\right)}.\end{aligned}$$

Given the assumption $K'' \geq \beta_1$, we have

$$\frac{r}{2\mu} K''\left(\mathcal{H}_1^\star\right)\left(1-\mathcal{H}_1^\star\right)^2 > c_B \cdot \sup_b bG'\left(b\right) > c_B \cdot \left[b_1^\star G'\left(b_1^\star\right) - \left(1-G\left(b_1^\star\right)\right)\right], \tag{65}$$

Because $f_1^\star - b_1^\star < 0$, we have

$$\frac{dS_1^\star}{db_1^\star} > (1-\mathcal{H}_1^\star) + (f_1^\star - b_1^\star)(1-\mathcal{H}_1^\star)^2 \frac{1}{c_B}. \tag{66}$$

A sufficient condition for $\frac{dS_1^\star}{db_1^\star} > 0$ is $\frac{c_B}{1-\mathcal{H}_1^\star} > b_1^\star - f_1^\star$. This condition holds because

$$\frac{c_B}{1-\mathcal{H}_1^\star} = \left(b_1^\star - \frac{1-G\left(b_1^\star\right)}{G'\left(b_1^\star\right)}\right) \geq b_1^\star - \frac{1-G\left(f_1^\star\right)}{G'\left(f_1^\star\right)} = b_1^\star - f_1^\star, \tag{67}$$

where we have used Assumption 1, that $\frac{1-G(x)}{G'(x)}$ is weakly decreasing in x.

2. **Customers' choices.** By the change of variables, $b = \frac{S-\mathcal{H}f}{1-\mathcal{H}} = f + \frac{S-f}{1-\mathcal{H}}$. As shown in the previous part, as c_B increases, $S_1^\star$ increases, $\mathcal{H}_1^\star$ increases, and $f_1^\star$ stays the same. So $b_1^\star$ also increases and the fraction of customers paying the market making spread, $1-G\left(b_1^\star\right)$, decreases.

   The fraction of customers doing matching, $G(b_1^\star) - G(f_1^\star)$, increases because $b_1^\star$ increases.

   The fraction of customers not trading, $G(f_1^\star)$, remains the same because $f_1^\star$ is invariant to c_B.

3. **Customers' average transaction costs.**

As c_B increases, the change in customers' average transaction costs is

$$\frac{1}{1-G\left(f_1^{\star}\right)}\left[\left(1-G\left(b_1^{\star}\right)\right)dS_1^{\star}-\left(S_1-f\right)G'\left(b_1^{\star}\right)db_1^{\star}\right], \tag{68}$$

so the average transaction costs increases if and only if

$$\left(1-G\left(b_1^{\star}\right)\right)dS_1^{\star}\geq\left(S_1^{\star}-f_1^{\star}\right)G'\left(b_1^{\star}\right)db_1. \tag{69}$$

From $S=\mathcal{H}f+(1-\mathcal{H})\,b$, we have

$$dS_1^{\star}=\left(f_1^{\star}-b_1^{\star}\right)d\mathcal{H}_1^{\star}+\left(1-\mathcal{H}_1^{\star}\right)db_1^{\star}, \tag{70}$$

in which case

$$\begin{aligned}
&\left(1-G\left(b_1^{\star}\right)\right)dS_1^{\star}\geq\left(S_1^{\star}-f_1^{\star}\right)G'\left(b_1^{\star}\right)db_1 \qquad (71)\\
\Longleftrightarrow&\left(1-G\left(b_1^{\star}\right)\right)\left(db_1^{\star}-\frac{b_1^{\star}-f_1^{\star}}{1-\mathcal{H}_1^{\star}}d\mathcal{H}_1^{\star}\right)\geq\left(b_1^{\star}-f_1^{\star}\right)G'\left(b_1^{\star}\right)db_1\\
\Longleftrightarrow&\frac{d\mathcal{H}_1^{\star}}{db_1^{\star}}\leq\left(1-\mathcal{H}_1^{\star}\right)\left(\frac{1}{b_1^{\star}-f_1^{\star}}-\frac{G'\left(b_1^{\star}\right)}{1-G\left(b_1^{\star}\right)}\right).
\end{aligned}$$

From the proof in Proposition 1, we know that

$$\frac{d\mathcal{H}_1^{\star}}{db_1^{\star}}=\frac{b_1^{\star}G'\left(b_1^{\star}\right)-\left(1-G\left(b_1^{\star}\right)\right)}{\frac{r}{2\mu}K''(\mathcal{H}_1^{\star})}, \tag{72}$$

so

$$\begin{aligned}
&\frac{d\mathcal{H}_1^{\star}}{db_1^{\star}}\leq\left(1-\mathcal{H}_1^{\star}\right)\left(\frac{1}{b_1^{\star}-f_1^{\star}}-\frac{G'\left(b_1^{\star}\right)}{1-G\left(b_1^{\star}\right)}\right) \qquad (73)\\
\Longleftrightarrow&\frac{b_1^{\star}G'\left(b_1^{\star}\right)-\left(1-G\left(b_1^{\star}\right)\right)}{\frac{r}{2\mu}K''(\mathcal{H}_1^{\star})}\leq\left(1-\mathcal{H}_1^{\star}\right)\left(\frac{1}{b_1^{\star}-f_1^{\star}}-\frac{G'\left(b_1^{\star}\right)}{1-G\left(b_1^{\star}\right)}\right).
\end{aligned}$$

In equilibrium we have $\frac{c_B}{1-\mathcal{H}_1^{\star}}=\frac{b_1^{\star}G'\left(b_1^{\star}\right)-\left(1-G\left(b_1^{\star}\right)\right)}{G'\left(b_1^{\star}\right)}$, and substituting this condition into the above result, we obtain

$$\frac{c_B G'\left(b_1^{\star}\right)}{\frac{r}{2\mu}K''\left(\mathcal{H}_1^{\star}\right)\left(1-\mathcal{H}_1^{\star}\right)^2}\leq\frac{1}{b_1^{\star}-f_1^{\star}}-\frac{G'\left(b_1^{\star}\right)}{1-G\left(b_1^{\star}\right)}. \tag{74}$$

We proved (from the above result) that, as c_B increases, average transaction costs increase if and only if (74) holds. Rewriting this condition gives us

$$\frac{c_B}{\frac{r}{2\mu}K''\left(\mathcal{H}_1^{\star}\right)\left(1-\mathcal{H}_1^{\star}\right)^2}\leq\frac{1}{G'\left(b_1^{\star}\right)\left(b_1^{\star}-f_1^{\star}\right)}-\frac{1}{1-G\left(b_1^{\star}\right)}. \tag{75}$$

As c_B increases, both $\mathcal{H}_1^\star$ and $b_1^\star$ increase, so the LHS is non-decreasing based on the assumption stated in the proposition. Now we want to show that the RHS is a decreasing function of $b_1^\star$. To show this, define

$$y\left(b_1^\star\right) = \frac{1}{G'\left(b_1^\star\right)\left(b_1^\star - f_1^\star\right)} - \frac{1}{1 - G\left(b_1^\star\right)}. \tag{76}$$

Then

$$\frac{dy}{db_1^\star} = \frac{-1}{\left(G'\left(b_1^\star\right)\left(b_1^\star - f_1^\star\right)\right)^2}\left(G''\left(b_1^\star\right)\left(b_1^\star - f_1^\star\right) + G'\left(b_1^\star\right)\right) - \frac{G'\left(b_1^\star\right)}{\left(1 - G\left(b_1^\star\right)\right)^2}. \tag{77}$$

From our Assumption 1, we know

$$-\frac{G''\left(b_1^\star\right)}{G'\left(b_1^\star\right)} \leq \frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)}. \tag{78}$$

Then we have

$$\begin{aligned} \frac{dy}{db_1^\star} &\leq \frac{1}{1 - G\left(b_1^\star\right)} \frac{1}{b_1^\star - f_1^\star} - \frac{1}{G'\left(b_1^\star\right)\left(b_1^\star - f_1^\star\right)^2} - \frac{G'\left(b_1^\star\right)}{\left(1 - G\left(b_1^\star\right)\right)^2} \\ &= \frac{1}{G'\left(b_1^\star\right)}\left[\frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)} \frac{1}{b_1^\star - f_1^\star} - \frac{1}{\left(b_1^\star - f_1^\star\right)^2} - \left(\frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)}\right)^2\right]. \end{aligned} \tag{79}$$

We know

$$\begin{aligned} &\frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)} \frac{1}{b_1^\star - f_1^\star} - \frac{1}{\left(b_1^\star - f_1^\star\right)^2} - \left(\frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)}\right)^2 \\ &\leq 2\frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)} \frac{1}{b_1^\star - f_1^\star} - \frac{1}{\left(b_1^\star - f_1^\star\right)^2} - \left(\frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)}\right)^2 \\ &= -\left(\frac{1}{b_1^\star - f_1^\star} - \frac{G'\left(b_1^\star\right)}{1 - G\left(b_1^\star\right)}\right)^2 < 0, \end{aligned} \tag{80}$$

so we must have $\frac{dy}{db_1^\star} < 0$. Then the RHS of (75) is decreasing in c_B. As $c_B \to 0$, the LHS of (75) is close to zero, while the RHS converges to infinity (because in this case $b \to f$). So if c_B is close to zero, customers' average transaction costs must be increasing in c_B. As c_B increases, either the condition (75) holds until the equilibrium reaches Case 2, or the condition is violated at some value of c_B and will be violated for all c_B larger than that value. So customers' average transaction costs is a hump-shaped function of c_B, and the decreasing region may not exist.

4. **Bank dealer's profit.** By the envelope theorem, $\frac{d\pi_B}{dc_B} = \frac{\partial \pi_B}{\partial c_B} = \frac{2\mu}{r}\left(-1\right)\left(1 - G\left(b_1^\star\right)\right) < 0$.

5. **Customer welfare.**

As c_B increases, we have

$$d\pi_c = \int_{f_1^\star}^{b_1^\star} (x - f_1^\star)\, d\mathcal{H}_1^\star dG(x) - \int_{b_1^\star}^{\infty} dS_1^\star dG(x)\,. \tag{81}$$

By $S_1^\star = (1 - \mathcal{H}_1^\star)\, b_1^\star + \mathcal{H}_1^\star f_1^\star$, we have

$$d\pi_c = d\mathcal{H}_1^\star \left(\int_{f_1^\star}^{b_1^\star} x dG(x) \right) - f_1^\star (1 - G(f_1^\star))\, d\mathcal{H}_1^\star - (-d\mathcal{H}_1^\star b_1^\star + (1 - \mathcal{H}_1^\star)\, db_1^\star)(1 - G(b_1^\star))\,. \tag{82}$$

Using

$$\int_{f_1^\star}^{b_1^\star} x dG(x) = -b_1^\star (1 - G(b_1^\star)) + f_1^\star (1 - G(f_1^\star)) + \int_{f_1^\star}^{b_1^\star} (1 - G(x))\, dx, \tag{83}$$

we have

$$d\pi_c = d\mathcal{H}_1^\star \int_{f_1^\star}^{b_1^\star} (1 - G(x))\, dx - (1 - \mathcal{H}_1^\star)(1 - G(b_1^\star))\, db_1^\star. \tag{84}$$

By $\frac{d\mathcal{H}_1^\star}{db_1^\star} = \frac{b_1^\star G'(b_1^\star) - (1 - G(b_1^\star))}{\frac{r}{2\mu} K''(\mathcal{H}_1^\star)}$ and $\frac{c_B}{1 - \mathcal{H}_1^\star} = \frac{b_1^\star G'(b_1^\star) - (1 - G(b_1^\star))}{G'(b_1^\star)}$, we have

$$\begin{aligned} & d\pi_c \leq 0 \\ \Longleftrightarrow & \frac{d\mathcal{H}_1^\star}{db_1^\star} \leq \frac{(1 - \mathcal{H}_1^\star)(1 - G(b_1^\star))}{\int_{f_1^\star}^{b_1^\star} (1 - G(x))\, dx} \\ \Longleftrightarrow & \frac{r}{2\mu} K''(\mathcal{H}_1^\star)(1 - \mathcal{H}_1^\star)^2 \geq c_B \frac{G'(b_1^\star)}{1 - G(b_1^\star)} \int_{f_1^\star}^{b_1^\star} (1 - G(x))\, dx. \end{aligned} \tag{85}$$

From the above proof, we know that π_c decreases if and only if

$$\frac{r}{2\mu} K''(\mathcal{H}_1^\star)(1 - \mathcal{H}_1^\star)^2 \geq c_B \frac{G'(b_1^\star)}{1 - G(b_1^\star)} \int_{f_1^\star}^{b_1^\star} (1 - G(x))\, dx. \tag{86}$$

In the unconstrained bank dealer case, as c_B increases, both $\mathcal{H}_1^\star$ and $b_1^\star$ increase, so the RHS of condition (86) is increasing in c_B and, by the assumption stated in the proposition, the LHS is non-increasing in c_B. When c_B is close to zero, condition (86) must be satisfied. As c_B increases from zero, there are two possibilities: first, the condition is satisfied until the equilibrium becomes the constrained bank dealer case; second, the condition is violated at some value and it will be violated for all c_B higher than that value.

A.3 Proof of Proposition 3

We prove the result in the following steps.

1. The solution must exist because the domain is a closed set and the objective function is continuous.

2. We must have $\mathcal{H}_2^\star \in (0,1)$, i.e., $\mathcal{H}_2^\star$ is an interior solution. The proof is the same as in Proposition 1.

3. We must have $S_2^\star = c_{NB}$. If $S_2^\star < c_{NB}$, then $(S_2^\star, \mathcal{H}_2^\star, f_2^\star)$ must be a local maximum in the unconstrained problem. We have already shown, however, that the objective function in the unconstrained problem is single-peaked with a unique local maximum (under the assumption $K'' > \beta_1$), so in this constrained problem, we must have $S_2^\star = c_{NB}$.

4. We must have $f_2^\star \in (0, c_{NB})$. If $f_2^\star = 0$, we must have $\mathcal{H}_2^\star = 0$, which is not optimal, as proved before. It is also impossible to have $f_2^\star = S_2^\star = c_{NB}$ because, in this case, no customer would choose matching and the optimal speed becomes $\mathcal{H}_2^\star = 0$, which again is not optimal.

5. $(\mathcal{H}_2^\star, f_2^\star)$ satisfies the first-order conditions. From the previous steps, we know that $S_2^\star = c_{NB}$ and $(\mathcal{H}_2^\star, f_2^\star)$ is an interior solution, and because the objective function is continuously differentiable, the optimal solution must satisfy the first-order conditions.

A.4 Proof of Proposition 4

1. **Spread, fee, and speed.** $S_2^\star = c_{NB}$ is invariant to c_B.

   $(\mathcal{H}_2^\star, f_2^\star)$ satisfies the first-order conditions for $\Pi_2(\mathcal{H}, f; c_B)$. By the implicit function theorem,

$$\frac{\partial f_2^\star}{\partial c_B} \propto \left(\frac{\partial^2 \Pi_2}{\partial f \partial \mathcal{H}} \frac{\partial^2 \Pi_2}{\partial \mathcal{H} \partial c_B} - \frac{\partial^2 \Pi_2}{\partial \mathcal{H}^2} \frac{\partial^2 \Pi_2}{\partial f \partial c_B} \right) \Bigg|_{(\mathcal{H}_2^\star, f_2^\star)}. \tag{87}$$

   For simplicity of notation, we drop the subscripts "2" from the remainder of the proof wherever it is not ambiguous. At the optimal solution, we have

$$\begin{aligned} \frac{r}{2\mu} \frac{\partial^2 \Pi_2}{\partial f \partial \mathcal{H}} = {} & \frac{\mathcal{H}}{1-\mathcal{H}} (S-f)(S-c_B-f\mathcal{H}) \frac{G''(b)}{(1-\mathcal{H})^2} - [fG'(f) - G(b) + G(f)] \\ & + \frac{G'(b)}{(1-\mathcal{H})^2}(S-c_B-f\mathcal{H}) + \frac{G'(b)}{(1-\mathcal{H})^2}\mathcal{H}(S-f-f(1-\mathcal{H})). \end{aligned} \tag{88}$$

   From the first-order conditions, we have $(S - c_B - f\mathcal{H}) = \frac{1-\mathcal{H}}{G'(b)} [fG'(f) - G(b) + G(f)]$, in which case the above expression becomes

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_2}{\partial f \partial \mathcal{H}} = \frac{fG'(f) - G(b) + G(f)}{(1-\mathcal{H})^2} \mathcal{H} \left[(S-f) \frac{G''(b)}{G'(b)} + 1 - \mathcal{H} \right] + \mathcal{H} \frac{G'(b)}{(1-\mathcal{H})^2} (S-f-f(1-\mathcal{H})). \tag{89}$$

Moreover, we can solve the following results at the optimum $(S_2^\star, \mathcal{H}_2^\star, f_2^\star)$:

$$\frac{r}{2\mu}\frac{\partial^2\Pi_2}{\partial\mathcal{H}\partial c_B} = \frac{S-f}{(1-\mathcal{H})^2}G'(b) > 0, \tag{90}$$

$$\begin{aligned}\frac{r}{2\mu}\frac{\partial^2\Pi_2}{\partial\mathcal{H}^2} =& \frac{S-f}{(1-\mathcal{H})^4}\left[-(S-f)(S-c_B-f\mathcal{H})G''(b) - 2(1-\mathcal{H})(S-c_B-f)G'(b)\right] \\ &- \frac{r}{2\mu}K''(\mathcal{H}),\end{aligned}$$

$$\frac{r}{2\mu}\frac{\partial^2\Pi_2}{\partial f\partial c_B} = -\frac{\mathcal{H}}{1-\mathcal{H}}G'(b) < 0. \tag{91}$$

The necessary and sufficient condition for $\frac{\partial f_2^\star}{\partial c_B} < 0$ is

$$\frac{\partial^2\Pi_2}{\partial f\partial\mathcal{H}}\frac{\partial^2\Pi_2}{\partial\mathcal{H}\partial c_B} < \frac{\partial^2\Pi_2}{\partial\mathcal{H}^2}\frac{\partial^2\Pi_2}{\partial f\partial c_B} \Longleftrightarrow \frac{\partial^2\Pi_2}{\partial f\partial\mathcal{H}}\frac{b-f}{\mathcal{H}} < -\frac{\partial^2\Pi_2}{\partial\mathcal{H}^2}. \tag{92}$$

Using $b-f = \frac{S-f}{1-\mathcal{H}}$, we have

$$\frac{r}{2\mu}\frac{\partial^2\Pi_2}{\partial f\partial\mathcal{H}} = (fG'(f) - G(b) + G(f))\frac{\mathcal{H}}{1-\mathcal{H}}\left[(b-f)\frac{G''(b)}{G'(b)} + 1\right] + G'(b)\frac{\mathcal{H}}{1-\mathcal{H}}(b-2f). \tag{93}$$

By the first-order condition (17), we know at the optimum,

$$\begin{aligned}fG'(f) - G(b) + G(f) &= \frac{S-c_B-\mathcal{H}f}{1-\mathcal{H}}G'(b) \\ &= \left(b - \frac{c_B}{1-\mathcal{H}}\right)G'(b).\end{aligned}$$

Substituting this into (93), we obtain

$$\frac{r}{2\mu}\frac{\partial^2\Pi_2}{\partial f\partial\mathcal{H}} = G'(b)\frac{\mathcal{H}}{1-\mathcal{H}}\left[\left(b - \frac{c_B}{1-\mathcal{H}}\right)\left[(b-f)\frac{G''(b)}{G'(b)} + 1\right] + (b-2f)\right]. \tag{94}$$

Similarly, we obtain

$$\begin{aligned}\frac{r}{2\mu}\frac{\partial^2 \Pi_2}{\partial \mathcal{H}^2} =& \frac{S-f}{(1-\mathcal{H})^4}\left[-(S-f)(S-c_B-f\mathcal{H})G''(b) - 2(1-\mathcal{H})(S-c_B-f)G'(b)\right] \\ &- \frac{r}{2\mu}K''(\mathcal{H}) \\ =& \frac{b-f}{(1-\mathcal{H})^2}\left[-(b-f)(S-c_B-f\mathcal{H})G''(b) - 2(S-c_B-f)G'(b)\right] \\ &- \frac{r}{2\mu}K''(\mathcal{H}). \end{aligned} \tag{95}$$

Using the following relations

$$S - c_B - f\mathcal{H} = (1-\mathcal{H})\left(b - \frac{c_B}{1-\mathcal{H}}\right), \tag{96}$$

$$S - c_B - f = (1-\mathcal{H})\left(b - f - \frac{c_B}{1-\mathcal{H}}\right), \tag{97}$$

we obtain

$$\frac{r}{2\mu}\frac{\partial^2 \Pi_2}{\partial \mathcal{H}^2} = \frac{b-f}{1-\mathcal{H}}\left[-(b-f)\left(b-\frac{c_B}{1-\mathcal{H}}\right)G''(b) - 2\left(b-f-\frac{c_B}{1-\mathcal{H}}\right)G'(b)\right] - \frac{r}{2\mu}K''(\mathcal{H}). \tag{98}$$

Substituting (94) and (98) into (92), the condition becomes

$$\begin{aligned}&\frac{r}{2\mu}K''(\mathcal{H}) - \frac{b-f}{1-\mathcal{H}}\left[-(b-f)\left(b-\frac{c_B}{1-\mathcal{H}}\right)G''(b) - 2\left(b-f-\frac{c_B}{1-\mathcal{H}}\right)G'(b)\right] \\ &\geq \frac{b-f}{\mathcal{H}}G'(b)\frac{\mathcal{H}}{1-\mathcal{H}}\left[\left(b-\frac{c_B}{1-\mathcal{H}}\right)\left[(b-f)\frac{G''(b)}{G'(b)}+1\right] + (b-2f)\right]. \end{aligned} \tag{99}$$

Simplifying the above condition, we obtain

$$\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2 \geq c_B(b-f)G'(b). \tag{100}$$

If $K''(\mathcal{H}) \geq \beta_1$, we can verify that the above holds, and $f_2^\star$ is decreasing in c_B.
We now move on to $\mathcal{H}_2^\star$. By the implicit function theorem, we have

$$\frac{\partial \mathcal{H}_2^\star}{\partial c_B} \propto \frac{\partial^2 \Pi_2}{\partial f \partial \mathcal{H}}\frac{\partial^2 \Pi_2}{\partial f \partial c_B} - \frac{\partial^2 \Pi_2}{\partial f^2}\frac{\partial^2 \Pi_2}{\partial \mathcal{H}\partial c_B}. \tag{101}$$

At the optimum, we have

$$\frac{\partial^2 \Pi_2}{\partial f^2} = \frac{\mathcal{H}^2 \left[-\left(S - c_B - f\mathcal{H}\right) G''\left(b\right) - 2\left(1-\mathcal{H}\right) G'\left(b\right)\right] + \mathcal{H}\left(1-\mathcal{H}\right)^2 \left(-fG''\left(f\right) - 2G'\left(f\right)\right)}{\left(1-\mathcal{H}\right)^2} \tag{102}$$

Substituting equations (89), (90), (91), and (102) into equation (101), we obtain

$$\begin{aligned} &\frac{\partial^2 \Pi_2}{\partial f \partial \mathcal{H}} \frac{\partial^2 \Pi_2}{\partial f \partial c_B} - \frac{\partial^2 \Pi_2}{\partial f^2} \frac{\partial^2 \Pi_2}{\partial \mathcal{H} \partial c_B} \propto \\ &\quad \left\{ - \left[\left(S - c_B - f\mathcal{H}\right) - c_B \frac{\mathcal{H}}{1-\mathcal{H}} \right] G'\left(b\right) + \left(1-\mathcal{H}\right)\left(fG'\left(f\right) - G\left(b\right) + G\left(f\right)\right) \right\} \\ &\quad + \left\{ \left(S - f\right)\left(fG''\left(f\right) + 2G'\left(f\right)\right) \right\}. \end{aligned} \tag{103}$$

Simplifying the terms in the first pair of braces using the following first-order condition,

$$G\left(b_2^\star\right) - G\left(f_2^\star\right) = -\left(c_{NB} - c_B - \mathcal{H}_2^\star f_2^\star\right) G'\left(b_2^\star\right) \frac{1}{1-\mathcal{H}_2^\star} + f_2^\star G'\left(f_2^\star\right), \tag{104}$$

we can obtain

$$\frac{\partial^2 \Pi_2}{\partial f \partial \mathcal{H}} \frac{\partial^2 \Pi_2}{\partial f \partial c_B} - \frac{\partial^2 \Pi_2}{\partial f^2} \frac{\partial^2 \Pi_2}{\partial \mathcal{H} \partial c_B} \propto c_B \frac{\mathcal{H}}{1-\mathcal{H}} G'\left(b\right) + \left\{ \left(S - f\right)\left(fG''\left(f\right) + 2G'\left(f\right)\right) \right\}. \tag{105}$$

Because $fG''\left(f\right) + 2G'\left(f\right) > fG''\left(f\right) + G'\left(f\right) > 0$ at the optimum,[36] we must have

$$\frac{\partial^2 \Pi_2}{\partial f \partial \mathcal{H}} \frac{\partial^2 \Pi_2}{\partial f \partial c_B} - \frac{\partial^2 \Pi_2}{\partial f^2} \frac{\partial^2 \Pi_2}{\partial \mathcal{H} \partial c_B} > 0. \tag{106}$$

Thus $\frac{\partial \mathcal{H}_2^\star}{\partial c_B} > 0$.

2. **Customers' choices.** We know $b = \frac{c_{NB} - f\mathcal{H}}{1-\mathcal{H}} = c_{NB} + \left(c_{NB} - f\right)\frac{\mathcal{H}}{1-\mathcal{H}}$. So as f decreases and $\mathcal{H}$ increases in c_B, b increases, and the fraction of customers paying the spread, $\left(1 - G\left(b_2^\star\right)\right)$, decreases.

   Because $b_2^\star$ increases and $f_2^\star$ decreases in c_B, the fraction of customers using the match-making service, $G(b_2^\star) - G(f_2^\star)$, increases.

   Because $f_2^\star$ decreases in c_B, the fraction of customers choosing not to trade decreases.

3. **Customers' average transaction costs.** Define $\xi = \frac{1-G\left(b_2^\star\right)}{1-G\left(f_2^\star\right)}$, which makes the average transaction costs $\left(1-\xi\right) f_2^\star + \xi S_2^\star$. If c_B increases, the above comparative statics imply that $f_2^\star$ decreases and $\left(1 - G\left(b_2^\star\right)\right)$ decreases and thus ξ decreases. Then

[36] This can be proved by $\Phi \geq 0$ and $f < \frac{1-G(f)}{G'(f)}$ at the optimum.

the change in average transaction costs is

$$(1-\xi)\, df_2^\star + f_2^\star\,(-d\xi) + S_2^\star d\xi = (1-\xi)\, df_2^\star + d\xi\,(S_2^\star - f_2^\star)\,. \tag{107}$$

We know that $df_2^\star < 0$, $(S_2^\star - f_2^\star) > 0$ and $d\xi < 0$, so average transaction costs decrease.

4. **Bank dealer's profit.** The bank dealer's profit is $\pi_B = \Pi_2\,(S_2^\star, \mathcal{H}_2^\star, f_2^\star)$. By the envelope theorem,

$$\frac{d\pi_B}{dc_B} = -\frac{2\mu}{r}\,(1 - G\,(b_2^\star)) < 0. \tag{108}$$

5. **Customer welfare.** Customer welfare is

$$\pi_c = \frac{2\mu}{r}\left[\int_{f_2^\star}^{b_2^\star} (b - f_2^\star)\,\mathcal{H}_2^\star dG\,(b) + \int_{b_2^\star}^{\infty} (b - c_{NB})\, dG\,(b)\right], \tag{109}$$

so

$$\frac{r}{2\mu}\frac{d\pi_C}{dc_B} = (b_2^\star - f_2^\star)\,\mathcal{H}_2^\star G'\,(b_2^\star)\,\frac{\partial b_2^\star}{\partial c_B} + \int_{f_2^\star}^{b_2^\star} \frac{\partial\left[(b - f_2^\star)\,\mathcal{H}_2^\star\right]}{\partial c_B} dG\,(b) - \frac{\partial b_2^\star}{\partial c_B}\,(b_2^\star - c_{NB})\,G'\,(b_2^\star) \tag{110}$$

$$= \int_{f_2^\star}^{b_2^\star} \frac{\partial\left[(b - f_2^\star)\,\mathcal{H}_2^\star\right]}{\partial c_B} dG\,(b) > 0, \tag{111}$$

where in the second equality we have used $b_2^\star = \frac{c_{NB} - \mathcal{H}_2^\star f_2^\star}{1 - \mathcal{H}_2^\star}$.

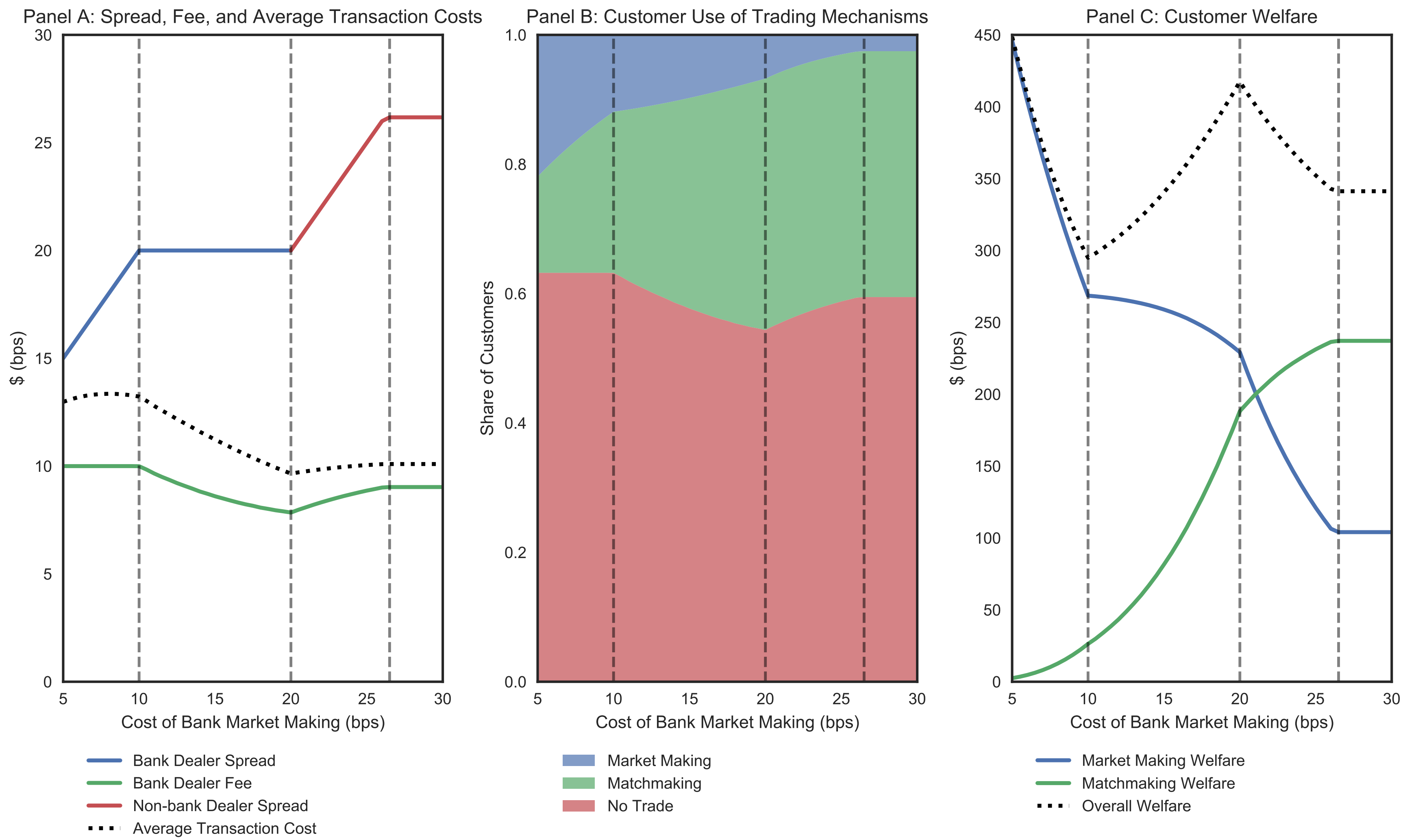

Figure 1: **Example** *Panel A presents the market-making spread, the matchmaking fee, and average transaction costs. Panel B plots the fractions of traders choosing market making, matchmaking, or refraining from trading. Panel C depicts overall customer welfare and, separately, its market-making and matchmaking components. The dashed vertical lines in each panel indicate the transitions between regions 1, 2, 3, and 4. The example uses an exponential distribution for customer private values,* $G(b)$*, and a quadratic function with parameter* k *for the cost of speed,* $K(\mathcal{H})$*. The parameters used in the example are* $\mu = 1$, $r = 0.01$, $c_{NB} = 20bps$, *and* $k = 0.1$.

Internet Appendix for

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

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

B Proofs of Case 3 of the Main Model

B.1 Proof of Lemma 1

As proven in previous propositions, we can show that for any S, it is never optimal for the bank dealer to choose $\mathcal{H} = 0$ or $f > S$, i.e., the bank dealer can always make a weakly positive profit by choosing $f \leq S$. (The bank dealer can make a strictly positive profit if $S > 0$ and the bank dealer sets some $f < S$.)

So without loss of generality, the optimization problem for the bank dealer given the non-bank dealer's spread S is

$$\max_{\mathcal{H}\in[0,1],S\geq f\geq 0} \Pi_{B,3} = \frac{2\mu}{r}\mathcal{H}f\left(G\left(b\right) - G\left(f\right)\right) - K\left(\mathcal{H}\right). \tag{112}$$

Given S,

$$\begin{aligned}\frac{\partial \Pi_{B,3}\left(S,\mathcal{H},f\right)}{\partial f} &\propto \mathcal{H}\left[G\left(b\right) - G\left(f\right) - f\left(G'\left(b\right)\frac{\mathcal{H}}{1-\mathcal{H}} + G'\left(f\right)\right)\right] \\ &\propto \mathcal{H}\left[G\left(b\right) - fG'\left(b\right)\frac{\mathcal{H}}{1-\mathcal{H}} - 1 + G'\left(f\right)\left[\frac{1-G\left(f\right)}{G'\left(f\right)} - f\right]\right].\end{aligned}$$

By Assumption 1, $\left(\frac{1-G(f)}{G'(f)} - f\right)$ is decreasing in f. Since $\tilde{f}$ solves $\frac{1-G(\tilde{f})}{G'(\tilde{f})} - \tilde{f} = 0$, we know that whenever $\mathcal{H} > 0$, $\frac{\partial \Pi_{B,3}(S,\mathcal{H},f)}{\partial f} < 0$ if $f \geq \tilde{f}$. And we have shown before that the optimal $\mathcal{H} > 0$. This implies that the optimal $f < \tilde{f}$.

Now we show that the non-bank dealer never wants to choose $S \leq f$, conditional on $f < \tilde{f}$. If the non-bank dealer chooses $S < f$, then

$$\frac{\partial \Pi_{NB,3}}{\partial S} \propto \left[\frac{1-G\left(S\right)}{G'\left(S\right)} - S + c_{NB}\right]. \tag{113}$$

By the definition of $\tilde{f}$, we must have $\frac{1-G(S)}{G'(S)} - S + c_{NB} > 0$ for any $S \leq f < \tilde{f}$. So the optimal S must be greater than f.

B.2 Proof of Proposition 5

We prove the result in the following steps:

1. Both the bank dealer and the non-bank dealer's best response functions are single-valued.

2. Use Kakutani's fixed point theorem to show equilibrium existence.

3. Discuss equilibria if c_B is very small or very large.

Based on the assumption $K'' > \beta_2$, we know there exists an upper bound $\bar{\mathcal{H}} < 1$ that the bank dealer will never choose any $\mathcal{H} > \bar{\mathcal{H}}$ (see the derivations for $\beta_1^\star$ for details). So we just focus on the strategy space

$$\Omega = \left\{ (S, \mathcal{H}, f) \, | c_B \geq S \geq f \geq 0, \mathcal{H} \in \left[0, \bar{\mathcal{H}}\right] \right\}. \tag{114}$$

Step 1-1: Non-bank dealer's best response function

Let's first consider the non-bank dealer's best response function $BR_{NS}(\mathcal{H}, f)$.

$$BR_{NB}(\mathcal{H}, f) = \arg \max_{c_B \geq S \geq f} \frac{2\mu}{r} \left[(S_{NB} - c_{NB}) (1 - G(b)) \right]. \tag{115}$$

It is obvious that the objective function is continuous in all arguments, and the correspondence $(\mathcal{H}, f) \to [f, c_B]$ is a compact-valued correspondence and is continuous at any $(\mathcal{H}, f)$. Then the best response function (correspondence) BR_{NB} is non-empty, compact-valued, and upper hemicontinuous (by the maximum theorem). Besides, we can show that BR_{NB} is single-valued. To see this, we have

$$\begin{aligned}
\frac{r}{2\mu} \frac{\partial \Pi_{NB,3}(S, \mathcal{H}, f)}{\partial S} &= 1 - G(b) - (S - c_{NB}) G'(b) \frac{1}{1 - \mathcal{H}} \\
&= G'(b) \left[\frac{1 - G(b)}{G'(b)} - \frac{S - f\mathcal{H}}{1 - \mathcal{H}} + \frac{c_{NB} - f\mathcal{H}}{1 - \mathcal{H}} \right] \\
&= G'(b) \left[\frac{1 - G(b)}{G'(b)} - b + \frac{c_{NB} - f\mathcal{H}}{1 - \mathcal{H}} \right].
\end{aligned} \tag{116}$$

We know that b is increasing in S, and $\frac{1-G(b)}{G'(b)} - b$ is decreasing in b. So given $(\mathcal{H}, f)$, $\left(\frac{1-G(b)}{G'(b)} - b \right)$ is decreasing in S. Then, there exists a unique best response value $S_{NB}(\mathcal{H}, f) \in [f, \infty)$ for each $(\mathcal{H}, f)$.

Step 1-2: Bank dealer's best response function

Let's look at the bank dealer's best response function $BR_S(S)$:

$$BR_B(S) = \arg \max_{\mathcal{H}\in[0,\bar{\mathcal{H}}],S\geq f\geq 0} \frac{2\mu}{r}\mathcal{H}f\left(G(b)-G(f)\right)-K(\mathcal{H}). \tag{117}$$

It is obvious that the objective function is continuous in all arguments, and the correspondence $S \to \left\{(\mathcal{H},f)\,|\mathcal{H}\in\left[0,\bar{\mathcal{H}}\right], S\geq f\geq 0\right\}$ is a compact-valued correspondence and is continuous at any S. Then, the best response function (correspondence) BR_S is non-empty, compact-valued and upper hemicontinuous. Next, we will show that $BR_B(S)$ is a single-valued function when $K''(\mathcal{H})$ is large enough. To see this, we will find a new and concave strategy space for the bank dealer, and show that the bank dealer's objective function is always strictly concave in this strategy space. Let's first look at

$$\begin{aligned}\frac{\partial \Pi_{B,3}(S,\mathcal{H},f)}{\partial f} &\propto \mathcal{H}\left[G(b)-G(f)-f\left(G'(b)\frac{\mathcal{H}}{1-\mathcal{H}}+G'(f)\right)\right]\\ &\propto \mathcal{H}\left[G(b)-fG'(b)\frac{\mathcal{H}}{1-\mathcal{H}}-1+G'(f)\left[\frac{1-G(f)}{G'(f)}-f\right]\right].\end{aligned} \tag{118}$$

First we know $\left(\frac{1-G(f)}{G'(f)}-f\right)$ is decreasing in f. Recall $\tilde{f}$ is the unique solution to $\frac{1-G(\tilde{f})}{G'(\tilde{f})}-\tilde{f}=0$. If $\left(\frac{1-G(f)}{G'(f)}-f\right)\leq 0$, i.e., if $f\geq\tilde{f}$, we must have $\frac{\partial\Pi_{B,3}(S,\mathcal{H},f)}{\partial f}<0$ as the RHS of the above expression is negative. In the interval $f\in\left[0,\tilde{f}\right]$, given $(S,\mathcal{H})$, $G(b)$ is decreasing in f, $\left(-fG'(b)\frac{\mathcal{H}}{1-\mathcal{H}}\right)$ is decreasing in f, $G'(f)$ is decreasing in f, and $\left(\frac{1-G(f)}{G'(f)}-f\right)$ is positive and decreasing in f. With these results, we can show that given $\mathcal{H}\neq 0$, if $f\in\left[0,\tilde{f}\right]$, $\frac{\partial\Pi_{B,3}(S,\mathcal{H},f)}{\partial f}$ is decreasing in f, and if $f\in\left[\tilde{f},S\right]$, $\frac{\partial\Pi_{B,3}(S,\mathcal{H},f)}{\partial f}$ is always negative. By iterated elimination of strictly dominated strategies, the bank dealer will never choose $f\geq\tilde{f}$, so we can focus on the new strategy space for the bank dealer:

$$\Omega_0(S)=\left\{(\mathcal{H},f)\,|\mathcal{H}\in\left(0,\bar{\mathcal{H}}\right); f\in\left(0,\min\left\{\tilde{f},S\right\}\right)\right\}, \tag{119}$$

and it is obvious that $\Omega_0(S)$ is a convex set for any S.

Now let's show that the bank dealer's objective function

$$\Pi_{B,3}=\frac{2\mu}{r}\mathcal{H}f\left(G(b)-G(f)\right)-K(\mathcal{H}) \tag{120}$$

is a strictly concave function in $\Omega_0(S)$. To show this, let's consider its Hessian matrix

$$H_B = \begin{bmatrix} \frac{\partial^2 \Pi_{B,3}}{\partial f^2} & \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} \\ \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} & \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} \end{bmatrix}. \tag{121}$$

The sufficient and necessary condition for $\Pi_{B,3}$ to be strictly concave is that H_B is negative definite, which is equivalent to the following conditions

$$\frac{\partial^2 \Pi_{B,3}}{\partial f^2} < 0, \tag{122}$$

and

$$\frac{\partial^2 \Pi_{B,3}}{\partial f^2} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} - \left(\frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} \right)^2 > 0. \tag{123}$$

To check these conditions, first we have

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial f^2} = \mathcal{H} \left[\frac{-2\mathcal{H}}{1-\mathcal{H}} G'(b) + \frac{f\mathcal{H}^2}{(1-\mathcal{H})^2} G''(b) - 2G'(f) - fG''(f) \right]. \tag{124}$$

In $\Omega_0(S)$, we must have $fG''(f) + G'(f) > 0$ (derived from $f < \frac{1-G(f)}{G'(f)}$ and $\Phi \geq 0$), so $\frac{\partial^2 \Pi_{B,3}}{\partial f^2} < 0$ for all $(\mathcal{H}, f) \in \Omega_0(S)$.

To check condition (123), first we derive the following results :

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} = \frac{(b-f)f}{(1-\mathcal{H})^2} \left[2G'(b) + \mathcal{H}(b-f) G''(b) \right] - \frac{r}{2\mu} K''(\mathcal{H}), \tag{125}$$

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial f \partial \mathcal{H}} = \frac{\mathcal{H}}{(1-\mathcal{H})^2} \left[-\mathcal{H} f (b-f) G''(b) + (S-2f) G'(b) \right]. \tag{126}$$

Let's define

$$\begin{aligned} LHS_1 = & \left[\frac{-2\mathcal{H}}{1-\mathcal{H}} G'(b) + \frac{f\mathcal{H}^2}{(1-\mathcal{H})^2} G''(b) - 2G'(f) - fG''(f) \right] \\ & \cdot \left[(b-f) f \left[2G'(b) + \mathcal{H}(b-f) G''(b) \right] - \frac{r}{2\mu} K''(\mathcal{H}) (1-\mathcal{H})^2 \right], \end{aligned} \tag{127}$$

$$RHS_1 = \frac{\mathcal{H}}{(1-\mathcal{H})^2} \left[-\mathcal{H} f (b-f) G''(b) + (S-2f) G'(b) \right]^2. \tag{128}$$

Then it is easy to verify

$$\frac{\partial^2 \Pi_{B,3}}{\partial f^2} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} - \left(\frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} \right)^2 > 0 \Longleftrightarrow LHS_1 > RHS_1. \tag{129}$$

Expanding LHS_1,

$$LHS_1 = LHS_{1-1} + LHS_{1-2} + LHS_{1-3}, \tag{130}$$

where

$$LHS_{1-1} = \{-2G'(f) - fG''(f)\}\left\{(b-f)f[2G'(b) + \mathcal{H}(b-f)G''(b)] - \frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2\right\}, \tag{131}$$

$$LHS_{1-2} = -\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2\left[\frac{-2\mathcal{H}}{1-\mathcal{H}}G'(b) + \frac{f\mathcal{H}^2}{(1-\mathcal{H})^2}G''(b)\right], \tag{132}$$

$$LHS_{1-3} = (b-f)f[2G'(b) + \mathcal{H}(b-f)G''(b)]\left[\frac{-2\mathcal{H}}{1-\mathcal{H}}G'(b) + \frac{f\mathcal{H}^2}{(1-\mathcal{H})^2}G''(b)\right]. \tag{133}$$

Expanding LHS_{1-3} and RHS_1, we get

$$\begin{aligned} LHS_{1-3} &= (b-f)^2 f^2\mathcal{H}\left(\frac{\mathcal{H}}{1-\mathcal{H}}\right)^2 G''^2(b) - (b-f)f\frac{4\mathcal{H}}{1-\mathcal{H}}G'^2(b) \\ &\quad + (b-f)f\left[2f\left(\frac{\mathcal{H}}{1-\mathcal{H}}\right)^2 - \frac{2\mathcal{H}^2}{1-\mathcal{H}}(b-f)\right]G'(b)G''(b), \end{aligned} \tag{134}$$

$$\begin{aligned} RHS_1 &= (b-f)^2 f^2\mathcal{H}\left(\frac{\mathcal{H}}{1-\mathcal{H}}\right)^2 G''^2(b) - \frac{2\mathcal{H}^2 f(b-f)(S-2F)}{(1-\mathcal{H})^2}G'(b)G''(b) \\ &\quad + \frac{\mathcal{H}}{(1-\mathcal{H})^2}(S-2f)^2 G'^2(b), \end{aligned} \tag{135}$$

$$RHS_1 - LHS_{1-3} = G'^2(b)\frac{\mathcal{H}}{(1-\mathcal{H})^2}S^2. \tag{136}$$

So

$$LHS_1 = LHS_{1-1} + LHS_{1-2} + LHS_{1-3} > RHS_1 \Longleftrightarrow LHS_{1-1} + LHS_{1-2} > G'^2(b)\frac{\mathcal{H}}{(1-\mathcal{H})^2}S^2. \tag{137}$$

Expanding $LHS_{1-1} + LHS_{1-2}$, we get

$$LHS_{1-1} + LHS_{1-2} > G'^2(b)\frac{\mathcal{H}}{(1-\mathcal{H})^2}S^2 \Longleftrightarrow \frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2 \cdot LHS_2 > RHS_2 \tag{138}$$

where

$$LHS_2 = 2G'(f) + fG''(f) + \frac{2\mathcal{H}}{1-\mathcal{H}}G'(b) - \frac{f\mathcal{H}^2}{(1-\mathcal{H})^2}G''(b), \tag{139}$$

$$RHS_2 = G'^2(b)\frac{\mathcal{H}}{(1-\mathcal{H})^2}S^2 + (b-f)f\left[2G'(b) + \mathcal{H}(b-f)G''(b)\right]\left[2G'(f) + fG''(f)\right]. \tag{140}$$

We know $LHS_2 > 0$ in $\Omega_0(S)$, so the condition (138) is equivalent to

$$\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2 > \frac{RHS_2}{LHS_2}. \tag{141}$$

Consider the sufficient condition for the above inequality

$$\begin{aligned}\frac{RHS_2}{LHS_2} &< \frac{G'^2(b)\frac{\mathcal{H}}{(1-\mathcal{H})^2}S^2}{\frac{2\mathcal{H}}{1-\mathcal{H}}G'(b)} + (b-f)f\left[2G'(b) + \mathcal{H}(b-f)G''(b)\right] \\ &< \frac{1}{2}\frac{S^2G'(b)}{1-\mathcal{H}} + 2fbG'(b).\end{aligned} \tag{142}$$

By $S < c_B$ and $S < b$, we know

$$\frac{1}{2}\frac{S^2G'(b)}{1-\mathcal{H}} + 2fbG'(b) < \frac{5}{2}\frac{c_BbG'(b)}{1-\mathcal{H}}. \tag{143}$$

From assumption $K'' > \beta_2$ and the definition of β_2, we know

$$\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^3 \geq \frac{5}{2}c_B\max_b\{bG'(b)\} \tag{144}$$

holds in the whole space $\Omega_0(S)$, so in this space

$$\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2 > \frac{5}{2}\frac{c_BbG'(b)}{1-\mathcal{H}} > \frac{RHS_2}{LHS_2}, \tag{145}$$

which means

$$\frac{\partial^2\Pi_{B,3}}{\partial f^2}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}^2} - \left(\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial f}\right)^2 > 0. \tag{146}$$

So the Hessian matrix in $\Omega_0(S)$ is negative definite. Thus, the bank dealer's objective function is strictly concave in $\Omega_0(S)$ for all S. Then, the bank dealer's best response function $BR_B(S)$ must be single-valued.

Step 2: Applying Kakutani's fixed point theorem

Now let's consider a new function defined on $(\mathcal{H}, f) \in [0, \bar{\mathcal{H}}] \times \left[0, \min\left\{\tilde{f}, S\right\}\right]$

$$BR\left((S, \mathcal{H}, f)\right) = \left(BR_{NB}\left(\mathcal{H}, f\right), BR_B\left(S\right)\right). \tag{147}$$

- Both BR_{NB} and BR_B are upper hemicontinuous; thus, BR is also upper hemicontinuous.
- By the Closed Graph Theorem,[37] because BR is hemicontinuous, the domain is closed, then the graph for BR (denoted as $Gr\left(BR\right)$) is closed.
- Because both BR_{NB} and BR_B are single valued, $BR\left((S, \mathcal{H}, f)\right)$ is single valued for any $(S, \mathcal{H}, f)$. Thus $BR\left((S, \mathcal{H}, f)\right)$ is convex set for any $(S, \mathcal{H}, f)$.
- Then by Kakutani's fixed point theorem,[38] because the domain is a non-empty, compact, and convex subset of Euclidean space, BR has closed graph and $BR\left((S, \mathcal{H}, f)\right)$ is non-empty and convex for all $(S, \mathcal{H}, f)$, then BR must have a fixed point. And by the definition of equilibrium, this fixed point is the equilibrium.

Step 3: Effect of c_B on the equilibrium structure

Next, we will show that if $c_B > c_{NB}$ but close enough, the equilibrium must be a constrained equilibrium with $S_3^\star = c_B$. And from the above we know that the bank dealer's best response function is single-valued, so the equilibrium is unique. To see this, suppose the bank dealer chooses $(\mathcal{H}, f)$, then the non-bank dealer's first order derivative is

$$\frac{\partial \Pi_{NB,3}\left(S, \mathcal{H}, f\right)}{\partial S} = 1 - G\left(b\right) - \left(S - c_{NB}\right) G'\left(b\right) \frac{1}{1 - \mathcal{H}}. \tag{148}$$

Then if the non-bank dealer chooses $S = c_B$ and $\frac{\partial \Pi_{NB,3}(S,\mathcal{H},f)}{\partial S} > 0$, then it is still optimal for the non-bank dealer to choose $S = c_B$. We have

$$\begin{aligned} & 1 - G\left(b\right) - \left(c_B - c_{NB}\right) G'\left(b\right) \frac{1}{1 - \mathcal{H}} > 0 \\ \iff \quad & c_B - c_{NB} < \left(1 - \mathcal{H}\right) \frac{1 - G\left(b\right)}{G'\left(b\right)}. \end{aligned} \tag{149}$$

The RHS of above inequality is decreasing in $\mathcal{H}$ and decreasing in S. And based on our assumption $K''\left(\mathcal{H}\right) > \beta_2$, we know in any equilibrium the bank dealer will always choose $\mathcal{H} < \bar{\mathcal{H}} = \sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta_2}}$, so if we define $\bar{c}_1$ as the unique solution of

$$\bar{c}_1 - c_{NB} = \left(1 - \sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta_2}}\right) \frac{1 - G\left(b\right)}{G'\left(b\right)}, \tag{150}$$

[37] See Proposition 1.4.8 of Aubin and Frankowska (2009).

[38] See Osborne and Rubinstein (1994).

where $b = \frac{\bar{c}_1}{1-\sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta_2}}}$, then if $c_B \in [c_{NB}, \bar{c}_1]$, there exists a unique constrained equilibrium.

If c_B is very large, we want to show that $\frac{\partial \Pi_{NB,3}(S,\mathcal{H},f)}{\partial S}|_{S=c_B} < 0$ is always true for any plausible $(\mathcal{H}, f)$. Then any equilibrium must be an interior equilibrium. To see this, note

$$1 - G(b) - (c_B - c_{NB}) G'(b) \frac{1}{1-\mathcal{H}} < 0 \tag{151}$$

$$\iff \quad c_B - c_{NB} > (1-\mathcal{H}) \frac{1-G(b)}{G'(b)}. \tag{152}$$

We know $\frac{1-G(b)}{G'(b)}$ is decreasing in b, so if $c_B - c_{NB} > \frac{1}{G'(0)}$, then all equilibria must be interior. Let $\bar{c}_2 = c_{NB} + \frac{1}{G'(0)}$.

B.3 Proof of Proposition 6

First it is obvious that for an unconstrained equilibrium, the change of c_B has no impact on the equilibrium. Now we prove results for the constrained equilibrium.

1. **Spread, fee, speed.** Because $S_3^\star = c_B$, an increase in c_B obviously increases the spread.

   By the implicit function theorem, we have

$$\frac{\partial f_3^\star}{\partial c_B} \propto \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} - \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} \frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B} \Bigg|_{\left(S_3^\star, \mathcal{H}_3^\star, f_3^\star\right)}. \tag{153}$$

   At $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$, by direct calculation, we get

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B} = \frac{\mathcal{H}}{(1-\mathcal{H})^2} \left[(1-\mathcal{H}) G'(b) - \mathcal{H} f G''(b)\right], \tag{154}$$

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} = \frac{\mathcal{H}}{(1-\mathcal{H})^2} \left[-f\mathcal{H}(b-f) G''(b) + (S-2f) G'(b)\right], \tag{155}$$

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} = \frac{(b-f) f}{(1-\mathcal{H})^2} \left\{2G'(b) + \mathcal{H}(b-f) G''(b)\right\} - \frac{r}{2\mu} K''(\mathcal{H}), \tag{156}$$

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} = \frac{f G'(b) + \mathcal{H} f (b-f) G''(b)}{(1-\mathcal{H})^2}. \tag{157}$$

   At $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$, it is obvious that $\frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B} > 0$, so

$$\left(\frac{r}{2\mu}\right)^2 \left(\frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} - \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} \frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B}\right) > 0 \iff \frac{\frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f}}{\frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B}} > \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2}. \tag{158}$$

Here, we have

$$\frac{r}{2\mu}\frac{\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial c_B}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial f}}{\frac{\partial^2\Pi_{B,3}}{\partial f\partial c_B}} = \frac{\frac{r}{2\mu}}{(1-\mathcal{H})^2}\frac{\left[-f\mathcal{H}(b-f)G''(b)+(S-2f)G'(b)\right]\left[fG'(b)+\mathcal{H}f(b-f)G''(b)\right]}{\mathcal{H}G'(b)-\frac{f\mathcal{H}^2}{1-\mathcal{H}}G''(b)}. \tag{159}$$

It is easy to verify the following results (using $S-f=(1-\mathcal{H})(b-f)$):

$$fG'(b)+\mathcal{H}f(b-f)G''(b) = SG'(b)-(b-f)\left[(1-\mathcal{H})G'(b)-\mathcal{H}fG''(b)\right], \tag{160}$$

$$-f\mathcal{H}(b-f)G''(b)+(S-2f)G'(b) = -fG'(b)+(b-f)\left[(1-\mathcal{H})G'(b)-\mathcal{H}fG''(b)\right]. \tag{161}$$

Then

$$\begin{aligned}&\frac{r}{2\mu}\frac{\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial c_B}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial f}}{\frac{\partial^2\Pi_{B,3}}{\partial f\partial c_B}}\\ &= \frac{\frac{r}{2\mu}}{(1-\mathcal{H})^2}\left\{\frac{-SfG'^2(b)}{(1-\mathcal{H})G'(b)-\mathcal{H}fG''(b)}+(S+f)(b-f)G'(b)-(b-f)^2\left[(1-\mathcal{H})G'(b)-\mathcal{H}fG''(b)\right]^2\right\}. \end{aligned}\tag{162}$$

Using the above expression and the expression for $\frac{r}{2\mu}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}^2}$, we have

$$\frac{\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial c_B}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial f}}{\frac{\partial^2\Pi_{B,3}}{\partial f\partial c_B}} > \frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}^2} \tag{163}$$

$$\Longleftrightarrow \frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2 \geq (b-f)f\left\{2G'(b)+\mathcal{H}(b-f)G''(b)\right\} \tag{164}$$

$$+\frac{SfG'^2(b)}{(1-\mathcal{H})G'(b)-\mathcal{H}fG''(b)}-(S+f)(b-f)G'(b)+(b-f)^2\left[(1-\mathcal{H})G'(b)-\mathcal{H}fG''(b)\right]^2$$

$$\Longleftrightarrow \frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2 \geq \frac{SfG'(b)}{(1-\mathcal{H})-\mathcal{H}f\frac{G''(b)}{G'(b)}}$$

$$\Longleftarrow \frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^3 \geq SfG'(b). \tag{165}$$

By the assumption $K''(\mathcal{H})>\beta_2$ and the expression of β_2, we know at the equilibrium $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$, we must have

$$\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^3 \geq \frac{5}{2}c_B bG'(b). \tag{166}$$

This is a sufficient condition for

$$\frac{r}{2\mu} K''(\mathcal{H})(1-\mathcal{H})^3 \geq S f G'(b) . \tag{167}$$

Thus we must have

$$\frac{\frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f}}{\frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B}} > \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H}^2} \tag{168}$$

in this case, which implies that

$$\frac{d f_3^\star}{d c_B} > 0. \tag{169}$$

By the implicit function theorem, we have

$$\frac{\partial \mathcal{H}_3^\star}{\partial c_B} \propto \left. \frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} - \frac{\partial^2 \Pi_{B,3}}{\partial f^2} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} \right|_{\left(S_3^\star, \mathcal{H}_3^\star, f_3^\star\right)} . \tag{170}$$

By direct calculation, we have

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B} = \frac{\mathcal{H}}{(1-\mathcal{H})^2} \left[(1-\mathcal{H}) G'(b) - \mathcal{H} f G''(b)\right], \tag{171}$$

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} = \frac{\mathcal{H}}{(1-\mathcal{H})^2} \left[-f \mathcal{H} (b-f) G''(b) + (S-2f) G'(b)\right], \tag{172}$$

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} = \frac{f G'(b) + \mathcal{H} f (b-f) G''(b)}{(1-\mathcal{H})^2}, \tag{173}$$

$$\frac{r}{2\mu} \frac{\partial^2 \Pi_{B,3}}{\partial f^2} = \mathcal{H} \left[-f G''(f) - 2G'(f) + \left(\frac{\mathcal{H}}{1-\mathcal{H}}\right)^2 f G''(b) - \frac{2\mathcal{H}}{1-\mathcal{H}} G'(b)\right]. \tag{174}$$

Then

$$\begin{aligned} & \frac{\partial^2 \Pi_{B,3}}{\partial f \partial c_B} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial f} \\ & = \left(\frac{\mathcal{H}}{(1-\mathcal{H})^2}\right)^2 \left\{ \frac{\mathcal{H}}{1-\mathcal{H}} (S-2f) G'^2(b) - \left(\frac{f\mathcal{H}^2}{1-\mathcal{H}} (b-f) + \frac{f\mathcal{H}^2}{(1-\mathcal{H})^2} (S-2f)\right) G'(b) G''(b) + \frac{f^2 \mathcal{H}^3}{(1-\mathcal{H})^2} (b-f) G''(b) \right\}. \end{aligned} \tag{175}$$

Let's write $\frac{\partial^2 \Pi_{B,3}}{\partial f^2} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B}$ as

$$\frac{\partial^2 \Pi_{B,3}}{\partial f^2} \frac{\partial^2 \Pi_{B,3}}{\partial \mathcal{H} \partial c_B} = RHS_1 + RHS_2, \tag{176}$$

where

$$RHS_1 = \frac{\mathcal{H}f}{(1-\mathcal{H})^2}\left[G'(b) + \mathcal{H}(b-f)G''(b)\right]\left[\left(\frac{\mathcal{H}}{1-\mathcal{H}}\right)^2 fG''(b) - \frac{2\mathcal{H}}{1-\mathcal{H}}G'(b)\right] \tag{177}$$

$$= \frac{\mathcal{H}f}{(1-\mathcal{H})^2}\left\{-\frac{2\mathcal{H}}{1-\mathcal{H}}G'^2(b) + \left[\mathcal{H}(b-f)\left(\frac{-2\mathcal{H}}{1-\mathcal{H}}\right) + \left(\frac{\mathcal{H}}{1-\mathcal{H}}\right)^2 f\right]G'(b)G''(b) + \mathcal{H}(b-f)\left(\frac{\mathcal{H}}{1-\mathcal{H}}\right)^2 fG''^2(b)\right\},$$

$$RHS_2 = \frac{\mathcal{H}f}{(1-\mathcal{H})^2}\left[G'(b) + \mathcal{H}(b-f)G''(b)\right]\left[-fG''(f) - 2G'(f)\right]. \tag{178}$$

Then

$$\frac{\partial^2\Pi_{B,3}}{\partial f\partial c_B}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial f} > \frac{\partial^2\Pi_{B,3}}{\partial f^2}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial c_B} \iff \frac{\partial^2\Pi_{B,3}}{\partial f\partial c_B}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial f} - RHS_1 > RHS_2. \tag{179}$$

We can directly check

$$\frac{\partial^2\Pi_{B,3}}{\partial f\partial c_B}\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial f} - RHS_1 = \frac{\mathcal{H}}{(1-\mathcal{H})^2}\frac{\mathcal{H}}{1-\mathcal{H}}SG'^2(b), \tag{180}$$

then $\mathcal{H}_3^\star$ increases if and only if at the optimum $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$

$$\frac{\mathcal{H}}{(1-\mathcal{H})^2}\frac{\mathcal{H}}{1-\mathcal{H}}SG'^2(b) > RHS_2 = \frac{\mathcal{H}f}{(1-\mathcal{H})^2}\left[G'(b) + \mathcal{H}(b-f)G''(b)\right]\left[-fG''(f) - 2G'(f)\right].$$

At the optimum, from equation (25), we know that $fG'(f) + G(f) < 1$, with the assumption on Φ we can show $-fG''(f) - 2G'(f) < 0$. So a sufficient condition for the above inequality is

$$G'(b) + \mathcal{H}(b-f)G''(b) > 0 \iff G'(b) + \frac{\mathcal{H}}{1-\mathcal{H}}(S-f)G''(b) > 0$$

$$\iff \mathcal{H} < \frac{1}{1 + (S-f)\left(-\frac{G''(b)}{G'(b)}\right)}, \tag{181}$$

which is equivalent to $\frac{\partial^2\Pi_{B,3}}{\partial\mathcal{H}\partial c_B} > 0$. With the assumption $K'' > \beta_3$, we know the equilibrium speed $\mathcal{H}_3^\star$ must satisfy

$$\mathcal{H}_3^\star < \sqrt{\frac{2\mu}{r}}\sqrt{\frac{4\bar{c}}{\beta}} < \frac{1}{1 + \bar{c}\left(-\frac{G''(0)}{G'(0)}\right)}. \tag{182}$$

From the assumption $\Phi' < 0$, we can show that $-\frac{G''(x)}{G'(x)}$ is a decreasing function of x, so the above inequality is a sufficient condition for (181). Then with the assumption $K'' > \beta_3$, $\mathcal{H}_3^\star$ increases in c_B.

2. **Customers' choices.** From equilibrium conditions (25) and (26), we have

$$b\left(G\left(b\right)-G\left(f\right)\right)-\left(b-f\right)fG'\left(f\right)=\frac{r}{2\mu}K'\left(\mathcal{H}\right). \tag{183}$$

Taking the total derivative of the above condition, we have

$$\left[G\left(b\right)+bG'\left(b\right)-G\left(f\right)-fG'\left(f\right)\right]\frac{db}{dc_B}=\left(b-f\right)\left[2G'\left(f\right)+fG''\left(f\right)\right]\frac{df}{dc_B}+\frac{r}{2\mu}K''\left(\mathcal{H}\right)\frac{d\mathcal{H}}{dc_B}. \tag{184}$$

Using equation (25), the above result becomes

$$\frac{c_B}{1-\mathcal{H}}G'\left(b\right)\frac{db}{dc_B}=\left(b-f\right)\left[2G'\left(f\right)+fG''\left(f\right)\right]\frac{df}{dc_B}+\frac{r}{2\mu}K''\left(\mathcal{H}\right)\frac{d\mathcal{H}}{dc_B}. \tag{185}$$

Then taking the total derivative of equation (25), we have

$$\left[G'\left(b\right)-f\frac{\mathcal{H}}{1-\mathcal{H}}G''\left(b\right)\right]\frac{db}{dc_B}=\left[\frac{\mathcal{H}}{1-\mathcal{H}}G'\left(b\right)+2G'\left(f\right)+fG''\left(f\right)\right]\frac{df}{dc_B}+\frac{fG'\left(b\right)}{\left(1-\mathcal{H}\right)^2}\frac{d\mathcal{H}}{dc_B}. \tag{186}$$

Finally, taking the total derivative of $b=\frac{c_B-\mathcal{H}f}{1-\mathcal{H}}$, we have

$$\frac{db}{dc_B}=\frac{1}{1-\mathcal{H}}+\frac{-\mathcal{H}}{1-\mathcal{H}}\frac{df}{dc_B}+\frac{b-f}{1-\mathcal{H}}\frac{d\mathcal{H}}{dc_B}. \tag{187}$$

From (185) and (186), we can get

$$\frac{\frac{df_3^\star}{dc_B}}{\frac{db_3^\star}{dc_B}}=\frac{\frac{fG'(b)}{\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2}\frac{c_B}{(1-\mathcal{H})}G'\left(b\right)-G'\left(b\right)+\frac{\mathcal{H}f}{1-\mathcal{H}}G''\left(b\right)}{\left(b-f\right)\frac{fG'(b)}{\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2}\left(2G'\left(f\right)+fG''\left(f\right)\right)-\frac{\mathcal{H}}{1-\mathcal{H}}G'\left(b\right)-\left(2G'\left(f\right)+fG''\left(f\right)\right)}. \tag{188}$$

From the assumption $K''\left(\mathcal{H}\right)>\beta_2$, we know at the equilibrium, we must have

$$\frac{r}{2\mu}K''\left(\mathcal{H}\right)\left(1-\mathcal{H}\right)^3>\frac{5}{2}c_BbG'\left(b\right). \tag{189}$$

Then we get (consider $c_B\geq S\geq f$)

$$\frac{fG'\left(b\right)}{\frac{r}{2\mu}K''\left(\mathcal{H}\right)\left(1-\mathcal{H}\right)^2}\frac{c_B}{\left(1-\mathcal{H}\right)}<1, \tag{190}$$

and

$$\left(b-f\right)\frac{fG'\left(b\right)}{\frac{r}{2\mu}K''\left(\mathcal{H}\right)\left(1-\mathcal{H}\right)^2}=\frac{\left(S-f\right)fG'\left(b\right)}{\frac{r}{2\mu}K''\left(\mathcal{H}\right)\left(1-\mathcal{H}\right)^3}<1. \tag{191}$$

So both the numerator and denominator of the RHS of (188) are negative. Then

$$\frac{\frac{df_3^\star}{dc_B}}{\frac{db_3^\star}{dc_B}} > 0. \tag{192}$$

Because we already show that $\frac{df_3^\star}{dc_B} > 0$, we must have $\frac{db_3^\star}{dc_B} > 0$. That is, the fraction of customers who refrain from trading goes down.

At the equilibrium $(S_3^\star, \mathcal{H}_3^\star, f_3^\star)$, $(G(b) - G(f))$ increases if and only if

$$\frac{df}{db} < \frac{G'(b)}{G'(f)}. \tag{193}$$

Define $\Gamma = \frac{fG'(b)}{\frac{r}{2\mu}K''(\mathcal{H})(1-\mathcal{H})^2}$ and by (185)$\times\Gamma-$(186), we can solve

$$\frac{df}{db} = \frac{\left(1 - \Gamma\frac{c_B}{1-\mathcal{H}}\right) G'(b) - \frac{\mathcal{H}f}{1-\mathcal{H}}G''(b)}{(1 - \Gamma(b-f))(2G'(f) + fG''(f)) + \frac{\mathcal{H}}{1-\mathcal{H}}G'(b)}. \tag{194}$$

So $(G(b) - G(f))$ increases if and only if

$$\frac{\left(1 - \Gamma\frac{c_B}{1-\mathcal{H}}\right) G'(b) - \frac{\mathcal{H}f}{1-\mathcal{H}}G''(b)}{(1 - \Gamma(b-f))(2G'(f) + fG''(f)) + \frac{\mathcal{H}}{1-\mathcal{H}}G'(b)} < \frac{G'(b)}{G'(f)}. \tag{195}$$

When $c_{NB} \to 0$ and $c_B \downarrow c_{NB}$, there is no matchmaking because in equilibrium $S_3^\star = c_B \to 0$. As c_B increases, $S_3^\star$ increases and some customers start to use the matchmaking mechanism. Because $G(b_3^\star) - G(f_3^\star)$ is a continuous function of c_B, in the region that c_{NB} is small enough and c_B is mildly above c_{NB}, the share of matchmaking must increase.

Because $f_3^\star$ increases in c_B, the fraction of customers who refrain from trading increases.

3. **Customers' average transaction cost** can be rewritten as

$$f + \frac{1 - G(b)}{1 - G(f)}(S - f). \tag{196}$$

As c_B increases, we know $f_3^\star$ increases, so at the equilibrium, we must have

$$\begin{aligned} d\left(f+\frac{1-G(b)}{1-G(f)}(S-f)\right) &> df+\frac{d\left[(1-G(b))(S-f)\right]}{1-G(f)} \\ &= df+\frac{1}{1-G(f)}\left\{d\left[(1-G(b))S\right]-d\left[(1-G(b))f\right]\right\} \\ &= df+\frac{1}{1-G(f)}\left\{d\left[(1-G(b))S\right]+G'(b)f db-(1-G(b))df\right\} \\ &> \frac{d\left[(1-G(b))S\right]}{1-G(f)}+\frac{G(b)-G(f)}{1-G(f)}df. \end{aligned} \tag{197}$$

So a sufficient condition for $d\left(f+\frac{1-G(b)}{1-G(f)}(S-f)\right)>0$ is

$$d\left[(1-G(b))S\right]>0 \tag{198}$$

which is equivalent to

$$(1-G(b))dS-G'(b)Sdb>0 \Longleftrightarrow S\frac{db}{dS}<\frac{1-G(b)}{G'(b)}. \tag{199}$$

If c_{NB} is small enough and $c_B \downarrow c_{NB}$, we have a constrained equilibrium, so the above condition becomes

$$c_B\frac{db}{dc_B}<\frac{1-G(b)}{G'(b)}. \tag{200}$$

Here, $\frac{db}{dc_B}$ can be solved by (185), (186), and (187). So if c_B-c_{NB} is small enough, the above inequality must hold (because the LHS $\to 0$ while the RHS is positive). In this case, average transaction costs must increase.

4. **The profits of dealers (bank dealer and non-bank dealer).** By the envelope theorem, we have

$$\frac{r}{2\mu}\frac{d\pi_B}{dc_B}=\mathcal{H}_3^\star f_3^\star G'(b_3^\star)\frac{1}{1-\mathcal{H}_3^\star}>0. \tag{201}$$

Now we turn to the non-bank dealer's profit. First, we have

$$\frac{r}{2\mu}\frac{d\pi_{NB}}{dc_B}=1-G(b_3^\star)-G'(b_3^\star)(c_B-c_{NB})\frac{db_3^\star}{dc_B}. \tag{202}$$

Then if $\frac{db_3^\star}{dc_B}$ is bounded and $c_B \downarrow c_{NB}$, we must have $\frac{d\pi_{NB}}{dc_B}>0$. Let's show that $\frac{db_3^\star}{dc_B}$ is bounded. By (185), (186) and (187), as $c_B \downarrow c_{NB}$, $\frac{db}{dc_B}$ can be solved and must be bounded.[39] Then, we must have

$$\lim_{c_B \downarrow c_{NB}} \frac{d\pi_{NB}}{dc_B} > 0. \tag{203}$$

For the case in which c_B is close to the unconstrained level, let's use $\frac{db_3^\star}{dc_B} = \frac{1}{1-\mathcal{H}_3^\star} + \frac{-\mathcal{H}_3^\star}{1-\mathcal{H}_3^\star}\frac{df_3^\star}{dc_B} + \frac{c_B - f_3^\star}{\left(1-\mathcal{H}_3^\star\right)^2}\frac{d\mathcal{H}_3^\star}{dc_B}$. If c_B is large enough, $1 - G\left(b_3^\star\right) - G'\left(b_3^\star\right)\left(c_B - c_{NB}\right)\frac{1}{1-\mathcal{H}_3^\star} \to 0$ (this is the non-bank dealer's first-order condition). In this case, if $\frac{-\mathcal{H}_3^\star}{1-\mathcal{H}_3^\star}\frac{df_3^\star}{dc_B} + \frac{c_B - f_3^\star}{\left(1-\mathcal{H}_3^\star\right)^2}\frac{d\mathcal{H}_3^\star}{dc_B} > 0$, we have $\frac{d\pi_{NB}}{dc_B} < 0$. We claim that this is true when K'' is large enough. We now show it.

Suppose that K'' is sufficiently high (the equilibrium speed $\mathcal{H}$ is close to zero) and c_B is sufficiently high so that it is close to the unconstrained equilibrium level. From Proposition 5, define $c_B^u = \inf_{c_x}\{\text{there exists at least one unconstrained equilibrium for } c_B = c_x\}$. First c_B^u is well defined and is greater than c_{NB}. Second, at the unconstrained equilibrium $S_3^\star = c_B^u$, the non-bank dealer's first-order condition satisfies

$$1 - G\left(b_3^\star\right) - G'\left(b_3^\star\right)\left(c_B^u - c_{NB}\right)\frac{1}{1-\mathcal{H}_3^\star} = 0. \tag{204}$$

Besides, for any $c_{NB} < c_B < c_B^u$, there exists only one constrained equilibrium with $S_3^\star = c_B$. If K'' is sufficiently high (say $K'' \to \infty$), then in any equilibrium $\mathcal{H}_3^\star \to 0$. In this case $b_3^\star \to S_3^\star$ and the equilibrium $S_3^\star$ can be solved from the non-bank dealer's first-order condition. $f_3^\star$ can be solved by (taking $\mathcal{H}_3^\star \to 0$)

$$G\left(S_3^\star\right) - G\left(f_3^\star\right) = f_3^\star G'\left(f_3^\star\right) > 0. \tag{205}$$

From (26) we know

$$\frac{r}{2\mu}K'\left(\mathcal{H}_3^\star\right) \to f_3^\star\left(G\left(S_3^\star\right) - G\left(f_3^\star\right)\right)$$

is finite. Then, we must have $\frac{d\mathcal{H}_3^\star}{dc_B} \to 0$ because $K''\left(\mathcal{H}_3^\star\right)\frac{d\mathcal{H}_3^\star}{dc_B} = \frac{d\left(K'\left(\mathcal{H}_3^\star\right)\right)}{dc_B}$. Then (186) becomes

$$G'\left(b\right)\frac{db}{dc_B} = \left[2G'\left(f\right) + fG''\left(f\right)\right]\frac{df}{dc_B}. \tag{206}$$

Substituting this into (185), we have (taking $\mathcal{H}_3^\star \to 0$)

$$fG'\left(b\right)\frac{db}{dc_B} = \frac{r}{2\mu}K''\left(\mathcal{H}\right)\frac{d\mathcal{H}}{dc_B}. \tag{207}$$

[39]the exception is when the coefficient matrix for the system is singular when $c_B = c_{NB}$, but this is a degenerate case.

Consider (26), we have

$$\frac{r}{2\mu}\frac{K'(\mathcal{H})}{\mathcal{H}} = f\frac{G(b)-G(f)}{b-f}\frac{b-f}{\mathcal{H}} \to \frac{r}{2\mu}K''(\mathcal{H}). \tag{208}$$

Then $\mathcal{H}\frac{df}{dc_B} < (b-f)\frac{d\mathcal{H}}{dc_B}$ because

$$\begin{aligned}
&\frac{df}{dc_B} < \frac{b-f}{\mathcal{H}}\frac{d\mathcal{H}}{dc_B} \\
\Longleftrightarrow & f\frac{G(b)-G(f)}{b-f}\frac{df}{dc_B} < \frac{r}{2\mu}K''(\mathcal{H})\frac{d\mathcal{H}}{dc_B} \\
\Longleftrightarrow & \frac{G(b)-G(f)}{b-f} < G'(f) + (G'(f) + fG''(f)).
\end{aligned} \tag{209}$$

If $K''(\mathcal{H})$ is sufficiently large, we must have

$$\mathcal{H}\frac{df}{dc_B} < (b-f)\frac{d\mathcal{H}}{dc_B} \tag{210}$$

in any equilibrium. Now consider

$$\begin{aligned}
\frac{r}{2\mu}\frac{d\pi_{NB}}{dc_B} &= 1 - G(b) - G'(b)(c_B - c_{NB})\frac{db}{dc_B} \\
&= 1 - G(b) - G'(b)(c_B - c_{NB})\left[\frac{1}{1-\mathcal{H}} + \frac{-\mathcal{H}}{1-\mathcal{H}}\frac{df}{dc_B} + \frac{b-f}{1-\mathcal{H}}\frac{d\mathcal{H}}{dc_B}\right].
\end{aligned} \tag{211}$$

If K'' to be sufficiently high, $\left(\frac{-\mathcal{H}}{1-\mathcal{H}}\frac{df}{dc_B} + \frac{b-f}{1-\mathcal{H}}\frac{d\mathcal{H}}{dc_B}\right)$ is always positive and bounded away from zero. This is because $\frac{df}{dc_B}$ is bounded away from zero and $\mathcal{H}$ increases in c_B if K'' is sufficiently high (from the comparative statics of $\mathcal{H}_3^\star$).

Then, we move c_B to the unconstrained level c_B^u (given K''), in this case

$$1 - G(b) - G'(b)(c_B - c_{NB})\frac{1}{1-\mathcal{H}} \to 0, \tag{212}$$

and

$$\frac{r}{2\mu}\frac{d\pi_{NB}}{dc_B} \to -G'(b)(c_B - c_{NB})\left(\frac{-\mathcal{H}}{1-\mathcal{H}}\frac{df}{dc_B} + \frac{b-f}{1-\mathcal{H}}\frac{d\mathcal{H}}{dc_B}\right) < 0. \tag{213}$$

5. **Customer welfare.**

As $c_{NB} \to 0$ and $c_B \downarrow c_{NB}$, the only equilibrium is a constrained equilibrium with $S_3^\star = c_B$, and $b_3^\star \to 0$ so all customers choose market making with zero cost. Then, customers welfare reaches the maximum in this case. Now, if c_B increases, the spread will increase because $S_3^\star = c_B$. In the meantime $f_3^\star$ also increases, so the customer welfare decreases. Customer welfare is a continuous function of both c_B and c_{NB}, so it decreases if c_{NB} is small enough and $c_B - c_{NB}$ is sufficiently small (and positive).

Using $\mathcal{H}(b-f) = b - S$, we can get

$$\begin{aligned}\frac{d\pi_c}{dc_B} &= \int_f^b \mathcal{H}\left(-\frac{df}{dc_B}\right) G'(x)\, dx + \int_f^b (x-f)\frac{d\mathcal{H}}{dc_B} G'(x)\, dx - (1 - G(b)) \\ &< \int_f^b (x-f)\frac{d\mathcal{H}}{dc_B} G'(x)\, dx - (1-G(b)).\end{aligned}$$

From the analysis in the proof for the comparative statics of π_{NB}, we know that when K'' is sufficiently high, $\frac{d\mathcal{H}}{dc_B} \to 0$, and both b and f converge to constant numbers. Then in this case, we must have

$$\frac{d\pi_c}{dc_B} < \int_f^b (x-f)\frac{d\mathcal{H}}{dc_B} G'(x)\, dx - (1-G(b)) \to -(1-G(b)) < 0.$$

This is intuitive: when c_B increases, both the matchmaking fee f and market making spread S increase, and both make consumers worse off. The only possible benefit that consumers can get is from the increase of matching speed $\mathcal{H}$, but if the cost of the matching technology is very high, the increase in matching speed will be very small, and the overall customer welfare will decrease.

C Proofs of Propositions in the Robustness Section

C.1 Proof of Proposition 7

Case 1 of the equilibrium: In this case, we know based on previous results that at the equilibrium $(S_1^\star, \mathcal{H}_1^\star, f_1^\star)$,

$$\frac{r}{2\mu} d\pi_c = d\mathcal{H}\int_f^b (1-G(x))\, dx - (1-\mathcal{H})(1-G(b))\, db, \quad (214)$$

$$\frac{r}{2\mu} d\left(\pi_B + (c_B - c_{NB})(1-G(b))\right) = -(c_B - c_{NB}) G'(b)\, db, \quad (215)$$

$$\pi_{NB} \equiv 0. \quad (216)$$

Then $dW > 0$ is equivalent to

$$\frac{d\mathcal{H}}{db}\int_f^b (1-G(x))\, dx - (1-\mathcal{H})(1-G(b)) - (c_B - c_{NB}) G'(b) > 0. \quad (217)$$

We know that at the equilibrium,

$$\frac{c_B}{1-\mathcal{H}} = \frac{bG'(b) - (1-G(b))}{G'(b)}, \tag{218}$$

so

$$(1-\mathcal{H})(1-G(b)) + c_B G'(b) = (1-\mathcal{H})\, bG'(b). \tag{219}$$

Then

$$\begin{aligned} &\frac{d\mathcal{H}}{db}\int_f^b (1-G(x))\,dx - (1-\mathcal{H})(1-G(b)) - (c_B - c_{NB})\, G'(b) > 0 \\ \Longleftrightarrow &\frac{d\mathcal{H}}{db}\int_f^b (1-G(x))\,dx > (1-\mathcal{H})\, bG'(b) - c_{NB}G'(b). \end{aligned} \tag{220}$$

From previous results we know that at the equilibrium

$$\frac{d\mathcal{H}}{db} > 0. \tag{221}$$

So the LHS of (220) is positive. The RHS of (220) becomes

$$\begin{aligned} (1-\mathcal{H})\, bG'(b) - c_{NB}G'(b) &= G'(b)\left[(1-\mathcal{H})\, b - c_{NB}\right] && (222) \\ &= G'(b)\left[S - \mathcal{H}f - c_{NB}\right]. && (223) \end{aligned}$$

In Case 1, we have $S - \mathcal{H}f - c_{NB} < 0$, so (220) always holds. Thus, $\frac{dW}{dc_B} > 0$.
Case 2 of the equilibrium: Based on our previous results, at the equilibrium $(S_2^\star, \mathcal{H}_2^\star, f_2^\star)$ we have

$$\begin{aligned} \frac{d\pi_B}{dc_B} &= -\frac{2\mu}{r}(1-G(b)), && (224) \\ d\pi_c &= \frac{2\mu}{r}\int_f^b \frac{\partial\left[(x-f)\,\mathcal{H}\right]}{\partial c_B}\, dG(x), && (225) \\ \pi_{NB} &\equiv 0. && (226) \end{aligned}$$

So

$$\begin{aligned} \frac{dW}{dc_B} &= -\frac{2\mu}{r}(1-G(b)) + \frac{2\mu}{r}\int_f^b \frac{d\left[(x-f)\,\mathcal{H}\right]}{dc_B}\, dG(x) + \frac{2\mu}{r}\frac{d\left((c_B - c_{NB})(1-G(b))\right)}{dc_B} && (227) \\ &= \frac{2\mu}{r}\left[-(c_B - c_{NB})\, G'(b)\frac{db}{dc_B} + \int_f^b \frac{d\left[(x-f)\,\mathcal{H}\right]}{dc_B}\, dG(x)\right]. && (228) \end{aligned}$$

In Case 2, we have $c_B - c_{NB} < 0$, $\frac{db}{dc_B} > 0$, $\frac{df}{dc_B} < 0$, and $\frac{d\mathcal{H}}{dc_B} > 0$, so $\frac{dW}{dc_B} > 0$.

C.2 Proof of Proposition 8

If we can show $V(c_1) > V(c_2)$ for any $0 < c_1 < c_2 < c_{NB}$, then this result is proved. Define

$$\begin{pmatrix} S_2 \\ \mathcal{H}_2 \\ f_2 \end{pmatrix} = \begin{pmatrix} S^\star(c_2) \\ \mathcal{H}^\star(c_2) \\ f^\star(c_2) \end{pmatrix}, V_2 = V(c_2); \tag{229}$$

and

$$\begin{pmatrix} S_1 \\ \mathcal{H}_1 \\ f_1 \end{pmatrix} = \begin{pmatrix} S^\star(c_1) \\ \mathcal{H}^\star(c_1) \\ f^\star(c_1) \end{pmatrix}, V_1 = V(c_1). \tag{230}$$

Start with $c_B = c_2$. Then the equilibrium is $(S_2, \mathcal{H}_2, f_2)$, and overall customer welfare is $V_2 = V(c_2)$. Now let's decrease c_B from c_2 to c_1. If the bank dealer still chose the strategy $(S_2, \mathcal{H}_2, f_2)$, the zero profit condition would be violated and the bank dealer's profit would be positive because

$$\frac{2\mu}{r}\left[\mathcal{H}_2 f_2 (G(b_2) - G(f_2)) + (S_2 - c_1)(1 - G(b_2))\right] - K(\mathcal{H}_2) > \frac{2\mu}{r}\left[\mathcal{H}_2 f_2 (G(b_2) - G(f_2)) + (S_2 - c_2)(1 - G(b_2))\right] - K(\mathcal{H}_2) = 0. \tag{231}$$

Consider the following adjustment by the bank dealer: keeping S_2 and f_2 unchanged, but increasing $\mathcal{H}$ such that the bank dealer finds an $\tilde{\mathcal{H}} \in (\mathcal{H}_2, 1)$ to satisfy the zero profit condition:

$$\pi_B\left(S_2, \tilde{\mathcal{H}}, f_2, c_1\right) = \frac{2\mu}{r}\left[\tilde{\mathcal{H}} f_2 \left(G\left(\tilde{b}\right) - G(f_2)\right) + (S_2 - c_1)\left(1 - G\left(\tilde{b}\right)\right)\right] - K\left(\tilde{\mathcal{H}}\right) = 0. \tag{232}$$

This can always be done because

1. If the bank dealer chooses $\mathcal{H} = 1$, his profit will be negative (from the assumption about β_1):
$$\pi_B(S_2, \mathcal{H} = 1, f_2, c_1) < 0. \tag{233}$$

2. We have shown above that
$$\pi_B(S_2, \mathcal{H}_2, f_2, c_1) > 0. \tag{234}$$

Therefore, by the mean value theorem, there must exist $\tilde{\mathcal{H}} \in (\mathcal{H}_2, 1)$ such that $\pi_B\left(S_2, \tilde{\mathcal{H}}, f_2, c_1\right) = 0$.

Having shown that the zero profit condition will be satisfied if the bank dealer chooses $\left(S_2, \tilde{\mathcal{H}}, f_2\right)$ if $c_B = c_1$, we show that this choice delivers higher customer welfare than $(S_2, \mathcal{H}_2, f_2)$ (which is the equilibrium if $c_B = c_2$). To see this, we fix $S = S_2$ and $f = f_2$, and consider $\frac{\partial \pi_c}{\partial \mathcal{H}}$:

$$\frac{\partial \pi_c}{\partial \mathcal{H}}|_{(S=S_2,\mathcal{H},f=f_2)} = \frac{2\mu}{r}\left[(b - f_2)\mathcal{H} G'(b)\frac{\partial b}{\partial \mathcal{H}} - (b - S_2) G'(b)\frac{\partial b}{\partial \mathcal{H}} + \int_{f_2}^{b} (x - f_2)\, dG(x)\right], \tag{235}$$

where $b = \frac{S_2 - \mathcal{H} f_2}{1-\mathcal{H}}$. It is straightforward to show that $\frac{\partial b}{\partial \mathcal{H}} = \frac{S_2 - f_2}{(1-\mathcal{H})^2}$ and $(b - f_2)\mathcal{H} - (b - S_2) = 0$. Then,

$$\frac{\partial \pi_c}{\partial \mathcal{H}}|_{(S=S_2, \mathcal{H}, f=f_2)} = \frac{2\mu}{r}\left[((b - f_2)\mathcal{H} - (b - S_2)) G'(b) \frac{S_2 - f_2}{(1-\mathcal{H})^2} + \int_{f_2}^{b} (x - f_2)\, dG(x)\right] \tag{236}$$

$$= \frac{2\mu}{r}\left[0 + \int_{f_2}^{b} (x - f_2)\, dG(x)\right] > 0. \tag{237}$$

So, given $S = S_2$ and $f = f_2$, π_c is increasing in $\mathcal{H}$. Since we know $\tilde{\mathcal{H}} > \mathcal{H}_2$, we must have

$$\pi_c\left(S_2, \tilde{\mathcal{H}}, f_2\right) > \pi_c(S_2, \mathcal{H}_2, f_2) = V_2 \tag{238}$$

i.e.,

$$\frac{2\mu}{r}\left[\int_{f_2}^{\tilde{b}} (x - f_2)\tilde{\mathcal{H}} dG(x) + \int_{\tilde{b}}^{\infty} (x - S_2)\, dG(x)\right] > \frac{2\mu}{r}\left[\int_{f_2}^{b_2} (x - f_2)\mathcal{H}_2 dG(x) + \int_{b_2}^{\infty} (x - S_2)\, dG(x)\right]. \tag{239}$$

Hence, when $c_B = c_1$ and if the bank dealer chooses $S = S_2$, $\mathcal{H} = \tilde{\mathcal{H}}$, and $f = f_2$, the bank dealer's zero profit condition will be satisfied and customer welfare will be higher than that when the bank dealer chooses $S = S_2$, $\mathcal{H} = \mathcal{H}_2$ and $f = f_2$ (which is the equilibrium customer welfare in the case when $c = c_2$).

By definition, $(S_1, \mathcal{H}_1, f_1)$ should generate weakly better customer welfare than any other strategy $(S, \mathcal{H}, f)$ that satisfies the zero profit condition when $c_B = c_1$, i.e.,

$$V_1 = \pi_c(S_1, \mathcal{H}_1, f_1) \geq \pi_c(S, \mathcal{H}, f) \quad \forall (S, \mathcal{H}, f) \in \{(S, \mathcal{H}, f) \,|\pi_B(S, \mathcal{H}, f, c_1) = 0\}. \tag{240}$$

Since $\left(S_2, \tilde{\mathcal{H}}, f_2\right)$ satisfies the zero profit condition when $c_B = c_1$, we must have

$$V_1 = \pi_c(S_1, \mathcal{H}_1, f_1) \geq \pi_c\left(S_2, \tilde{\mathcal{H}}, f_2\right). \tag{241}$$

Combining (238) and (241), we have

$$V_1 = \pi_c(S_1, \mathcal{H}_1, f_1) \geq \pi_c\left(S_2, \tilde{\mathcal{H}}, f_2\right) > \pi_c(S_2, \mathcal{H}_2, f_2) = V_2. \tag{242}$$

The above analysis is true for any $0 < c_1 < c_2 < c_{NB}$, so $V(c_B)$ must be decreasing in c_B.

C.3 Proof of Proposition 9

The objective function is a continuous function of $(S, f, \mathcal{H})$, and the domain is a compact set, so the maximum must exist. To show the FOCs are satisfied, we just need to show that the optimum is not a corner solution, i.e., the optimal solution cannot be $f_a = 0$, $\mathcal{H}_a = 0$, or $\mathcal{H}_a = 1$.

First, $\mathcal{H}_a = 1$ is not an optimum. Suppose $\mathcal{H} = 1$, then $b = \infty$ and the bank dealer's profit becomes

$$\begin{aligned}&\frac{2\mu}{r}\left[\left(G(b)-G(f)\right)\left[\mathcal{H}f-K\left(\mathcal{H}\right)\right]+\left(S-c_B\right)\left(1-G\left(b\right)\right)\right]\\=&\frac{2\mu}{r}\left[\left(1-G(f)\right)\left[f-K\left(1\right)\right]+\left(S-c_B\right)\left(1-1\right)\right]\end{aligned}$$

Since $K(1) > c_{NB} > f$, the bank dealer's profit will be negative thus this cannot be an optimum.

Second, $\mathcal{H}_a = 0$ is not an optimum. We prove this by contradiction. Suppose $\mathcal{H} = 0$, then the bank dealer's profit becomes

$$\frac{2\mu}{r}\left[\left(S-c_B\right)\left(1-G\left(S\right)\right)\right] \tag{243}$$

which is invariant to f. Then, the optimal solution (S_a, f_a) is given by

$$S_a = \arg\max_{S\leq c_{NB}}\frac{2\mu}{r}\left[\left(S-c_B\right)\left(1-G\left(S\right)\right)\right], \tag{244}$$

$$f_a \in \left[0, S_a\right]. \tag{245}$$

Here f_a can be any value in $[0, S_a]$. Consider $\frac{\partial\pi_B}{\partial\mathcal{H}}|_{(S_a,\mathcal{H}_a=0,f_a)}$. If this derivative is positive for some $f_a \in [0, S_a]$, then it is profitable for the bank dealer to choose a positive $\mathcal{H}$. We can show that

$$\frac{\partial\pi_B}{\partial\mathcal{H}}|_{(S_a,\mathcal{H}_a=0,f_a)} = \frac{2\pi}{r}\left[\left(G\left(b_a\right)-G\left(f_a\right)\right)f_a-\left(S_a-c_B\right)G'\left(b_a\right)\left(S_a-f_a\right)\right], \tag{246}$$

where $b_a = \frac{S_a-\mathcal{H}_a f_a}{1-\mathcal{H}_a} = S_a$. So,

$$\frac{\partial\pi_B}{\partial\mathcal{H}}|_{(S_a,\mathcal{H}_a=0,f_a)} = \frac{2\pi}{r}\left(S_a-f_a\right)\left[f_a\frac{G\left(S_a\right)-G\left(f_a\right)}{S_a-f_a}-\left(S_a-c_B\right)G'\left(S_a\right)\right]. \tag{247}$$

We only need to find one f_a such that the above derivative is positive. Consider $f_a \to S_a$. Then, $S_a - f_a > 0$, $\frac{G(S_a)-G(f_a)}{S_a-f_a} \to G'(S_a)$, and $f_a > S_a - c_B$. The derivative of the bank dealer's profit with respect to the matching speed is positive, and hence it is not optimal to set $\mathcal{H}_a = 0$.

Third, $f_a = 0$ is not an optimum. If $f_a = 0$, it is obvious that the bank dealer should not invest in matchmaking, i.e., $\mathcal{H}_a = 0$. But we have already shown that this cannot be an optimum.

C.4 Proof of Proposition 10

Based on Assumption 2, the optimization problem becomes

$$\begin{aligned}
\max_{\substack{c_{NB} \geq S \geq f \geq 0 \\ \mathcal{H} \in [0,1]}} & \frac{2\mu}{r} \left[\left(G(b) - G(f) \right) \left[\mathcal{H} f - K\left(\mathcal{H}\right) \right] + \left(S - c_B \right) \left(1 - G\left(b\right) \right) \right] \\
&= \frac{2\mu}{r} \left[\frac{b-f}{B} \left[\mathcal{H} f - K\left(\mathcal{H}\right) \right] + \left(S - c_B \right) \left(1 - \frac{b}{B} \right) \right] \\
&= \frac{2\mu}{r} \frac{1}{B} \left[\frac{S-f}{1-\mathcal{H}} \left(\mathcal{H} f - K\left(\mathcal{H}\right) \right) + \left(S - c_B \right) \left(B - \frac{S - \mathcal{H} f}{1-\mathcal{H}} \right) \right]. \quad (248)
\end{aligned}$$

Suppose the constraint $c_{NB} \geq S$ is not binding. Given any $\mathcal{H}$ and f, the objective function is a quadratic function of S and the FOC is

$$\begin{aligned}
& \frac{1}{1-\mathcal{H}} \left(\mathcal{H} f - K\left(\mathcal{H}\right) \right) + B - \frac{S - \mathcal{H} f}{1-\mathcal{H}} + \left(S - c_B \right) \left(\frac{-1}{1-\mathcal{H}} \right) = 0 \\
\Rightarrow \quad & \mathcal{H} f - K\left(\mathcal{H}\right) + B\left(1-\mathcal{H}\right) - S + \mathcal{H} f - S + c_B = 0 \\
\Rightarrow \quad & 2S = \mathcal{H} f - K\left(\mathcal{H}\right) + B\left(1-\mathcal{H}\right) + \mathcal{H} f + c_B \\
\Rightarrow \quad & S = \mathcal{H} f + \frac{B\left(1-\mathcal{H}\right) + c_B - K\left(\mathcal{H}\right)}{2}. \quad (249)
\end{aligned}$$

Similarly, the objective function is also a quadratic function of f and the FOC is

$$\begin{aligned}
& -\frac{1}{1-\mathcal{H}} \left(\mathcal{H} f - K\left(\mathcal{H}\right) \right) + \frac{S-f}{1-\mathcal{H}} \mathcal{H} + \left(S - c_B \right) \frac{\mathcal{H}}{1-\mathcal{H}} = 0 \\
\Rightarrow \quad & -f + \frac{K\left(\mathcal{H}\right)}{\mathcal{H}} + S - f + S - c_B = 0 \\
\Rightarrow \quad & f = S - \frac{1}{2} \left(c_B - \frac{K\left(\mathcal{H}\right)}{\mathcal{H}} \right). \quad (250)
\end{aligned}$$

Given any $\mathcal{H}$ (and with the constraint $S \geq f \geq 0$), solving the above two FOCs gives us the best response functions

$$S = \frac{B + c_B}{2}, \quad (251)$$

$$f = \min \left\{ \frac{B + \frac{K(\mathcal{H})}{\mathcal{H}}}{2}, S \right\} = \frac{B}{2} + \frac{1}{2} \min \left\{ \frac{K\left(\mathcal{H}\right)}{\mathcal{H}}, c_B \right\}. \quad (252)$$

To find the optimum, we need to substitute (251) and (252) into the objective function, and find the optimal $\mathcal{H}$. We can then find the optimal f using (252). Note that $\frac{K(\mathcal{H})}{\mathcal{H}}$ is increasing in $\mathcal{H}$ (since K is convex). When $\frac{K(\mathcal{H})}{\mathcal{H}} > c_B$, $f = S$ and the objective function will be independent of $\mathcal{H}$. So without loss of generality, we just need to consider the region

$\frac{K(\mathcal{H})}{\mathcal{H}} \leq c_B$. Then, the best response function of f is

$$f = \frac{B}{2} + \frac{1}{2}\frac{K(\mathcal{H})}{\mathcal{H}}. \tag{253}$$

We substitute (251) and (253) into the objective function to obtain

$$\begin{aligned}
&\frac{S-f}{1-\mathcal{H}}\left(\mathcal{H}f - K(\mathcal{H})\right) + (S - c_B)\left(B - \frac{S-\mathcal{H}f}{1-\mathcal{H}}\right) &(254)\\
=&\frac{1}{2(1-\mathcal{H})}\left(c_B - \frac{K(\mathcal{H})}{\mathcal{H}}\right)\left(\frac{\mathcal{H}B}{2} - \frac{1}{2}K(\mathcal{H})\right) + \frac{B-c_B}{2}\left(\frac{B}{2} - \frac{c_B - K(\mathcal{H})}{2(1-\mathcal{H})}\right). &(255)
\end{aligned}$$

The optimal $\mathcal{H}$ can be solved from the above objective function with the constraint $\frac{K(\mathcal{H})}{\mathcal{H}} \leq c_B$.

The optimal spread $S = \frac{B+c_B}{2}$ is an increasing function of c_B. Hence, when $\frac{B+c_B}{2} \leq c_{NB} \Longleftrightarrow c_B \leq 2c_{NB} - B$, the equilibrium is unconstrained, and when $c_B \in (2c_{NB} - B, c_{NB})$, the equilibrium is constrained.

We focus on the constrained equilibrium because our goal is to examine whether the increase in customer welfare in Case 2 of our main model is robust to the alternative formulation of the bank dealer's cost of matchmaking. We now derive the comparative statics for the bank dealer's matchmaking fee, matching speed, and most importantly overall customer welfare. In the constrained equilibrium, $S_a = c_{NB}$ and the objective function is

$$\Pi_B = \frac{2\mu}{r}\frac{1}{B}\left[\frac{c_{NB}-f}{1-\mathcal{H}}\left(\mathcal{H}f - K(\mathcal{H})\right) + (c_{NB} - c_B)\left(B - \frac{c_{NB}-\mathcal{H}f}{1-\mathcal{H}}\right)\right]. \tag{256}$$

Given any $\mathcal{H} \in (0,1)$, Π_B is a quadratic function of f, and the FOC with respect to f is

$$\begin{aligned}
&\frac{-1}{1-\mathcal{H}}\left(\mathcal{H}f - K(\mathcal{H})\right) + \frac{c_{NB}-f}{1-\mathcal{H}}\mathcal{H} + (c_{NB} - c_B)\frac{\mathcal{H}}{1-\mathcal{H}} = 0\\
\Rightarrow &f = c_{NB} - \frac{1}{2}\left(c_B - \frac{K(\mathcal{H})}{\mathcal{H}}\right). &(257)
\end{aligned}$$

As we argued before, we just need to consider the case when $c_B \geq \frac{K(\mathcal{H})}{\mathcal{H}}$ and verify that the solution (257) satisfies the constraint $f \leq S = c_{NB}$.

Then, the objective function becomes

$$\begin{aligned}
\frac{rB}{2\mu}\Pi_B &= \frac{\mathcal{H}}{2(1-\mathcal{H})}\left[\left(c_B - \frac{K(\mathcal{H})}{\mathcal{H}}\right)\left(c_{NB} - \frac{1}{2}\left(c_B + \frac{K(\mathcal{H})}{\mathcal{H}}\right)\right)\right] + (c_{NB} - c_B)(B - c_{NB}) - \frac{\mathcal{H}}{2(1-\mathcal{H})}\left(c_B - \frac{K(\mathcal{H})}{\mathcal{H}}\right)(c_{NB} - c_B)\\
&= \frac{\mathcal{H}}{2(1-\mathcal{H})}\left[-\frac{1}{2}\left(c_B - \frac{K(\mathcal{H})}{\mathcal{H}}\right)\left(c_B + \frac{K(\mathcal{H})}{\mathcal{H}}\right) + c_B\left(c_B - \frac{K(\mathcal{H})}{\mathcal{H}}\right)\right] + (c_{NB} - c_B)(B - c_{NB})\\
&= \frac{\mathcal{H}}{4(1-\mathcal{H})}\left(c_B - \frac{K(\mathcal{H})}{\mathcal{H}}\right)^2 + (c_{NB} - c_B)(B - c_{NB}) \qquad (258)
\end{aligned}$$

For $K(\mathcal{H}) = \frac{1}{2}k\mathcal{H}^2$,

$$\frac{rB}{2\mu}\Pi_B = \frac{\mathcal{H}}{4(1-\mathcal{H})}\left(c_B - \frac{1}{2}k\mathcal{H}\right)^2 + (c_{NB} - c_B)(B - c_{NB}). \tag{259}$$

The derivative with respect to $\mathcal{H}$ is

$$\frac{rB}{2\mu}\frac{d\Pi_B}{d\mathcal{H}} = \frac{1}{4}\left[\frac{1}{(1-\mathcal{H})^2}\left(c_B - \frac{1}{2}k\mathcal{H}\right)^2 - \frac{\mathcal{H}}{1-\mathcal{H}}k\left(c_B - \frac{1}{2}k\mathcal{H}\right)\right] \tag{260}$$

$$= \frac{c_B - \frac{1}{2}k\mathcal{H}}{4(1-\mathcal{H})^2}\left[c_B - \frac{1}{2}k\mathcal{H} - k\mathcal{H}(1-\mathcal{H})\right]. \tag{261}$$

We have already shown that one only needs to consider the value of $\mathcal{H}$ when $\frac{K(\mathcal{H})}{\mathcal{H}} = \frac{1}{2}k\mathcal{H} \leq c_B$. Therefore, if $k \geq 4c_{NB}$, we have $\frac{1}{2}k\mathcal{H} \leq c_B \Rightarrow \mathcal{H} \leq \frac{1}{2}$, then $\frac{c_B - \frac{1}{2}k\mathcal{H}}{4(1-\mathcal{H})^2} > 0$ and $c_B - \frac{1}{2}k\mathcal{H} - k\mathcal{H}(1-\mathcal{H})$ is decreasing in $\mathcal{H}$. The unique optimal $\mathcal{H}_a$ is solved by

$$c_B - \frac{1}{2}k\mathcal{H}_a - k\mathcal{H}_a(1-\mathcal{H}_a) = 0. \tag{262}$$

It is clear that when c_B increases, the optimal $\mathcal{H}_a$ must increase. Besides,

$$c_B - \frac{1}{2}k\mathcal{H}_a = k\mathcal{H}_a(1-\mathcal{H}_a) \tag{263}$$

also increases. Then, the optimal fee

$$f_a = c_{NB} - \frac{1}{2}\left(c_B - \frac{1}{2}k\mathcal{H}_a\right) \tag{264}$$

must decrease. Similar to Case 2 of our main model, the change in overall customer welfare when the regulatory cost increases is

$$\frac{r}{2\mu}\frac{d\pi_c}{dc_B} = \int_{f_a}^{b_a} \frac{d\left[(x - f_a)\mathcal{H}_a\right]}{dc_B} dG(x)$$

$$= \int_{f_a}^{b_a}\left[-\frac{df_a}{dc_B}\mathcal{H}_a + (x - f_a)\frac{d\mathcal{H}_a}{dc_B}\right]dG(x), \tag{265}$$

where $b_a = \frac{c_{NB} - \mathcal{H}_a f_a}{1-\mathcal{H}_a}$. We already proved $\frac{d\mathcal{H}_a}{dc_B} > 0$ and $\frac{df_a}{dc_B} < 0$, and therefore we must have

$$-\frac{df_a}{dc_B}\mathcal{H}_a + (x - f_a)\frac{d\mathcal{H}_a}{dc_B} > 0 \tag{266}$$

for all $x \in (f_a, b_a)$. Then,

$$\int_f^b \left[-\frac{df_a}{dc_B} \mathcal{H}_a + (x - f_a) \frac{d\mathcal{H}_a}{dc_B} \right] dG(x) \tag{267}$$

must be positive, and thus $\frac{d\pi_c}{dc_B} > 0$.