BANK FOR INTERNATIONAL SETTLEMENTS

BIS Working Papers

No 861

Dealers' insurance, market structure, and liquidity

by Francesca Carapella and Cyril Monnet

Monetary and Economic Department

May 2020

JEL classification: G11, G23, G28.

Keywords: liquidity, dealers, insurance, central counterparties.

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in them are those of their authors and not necessarily the views of the BIS.

This publication is available on the BIS website (www.bis.org).

© *Bank for International Settlements 2020. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

Dealers' Insurance, Market Structure, And Liquidity*

Francesca Carapella
Federal Reserve Board of Governors
francesca.Carapella@frb.gov

Cyril Monnet
Bank for International Settlements
cyril.monnet@bis.org

May 2020

Abstract

We develop a parsimonious model to study the effect of regulations aimed at reducing counterparty risk on the structure of over-the-counter securities markets. We find that such regulations promote entry of dealers, thus fostering competition and lowering spreads. Greater competition, however, has an indirect negative effect on market making profitability. General equilibrium effects imply that more competition can distort incentives of all dealers to invest in efficient technologies ex ante, and so can cause a social welfare loss. Our results are consistent with empirical findings on the effects of post-crisis regulations and with the opposition of some market participants to those regulations.

Keywords: Liquidity, dealers, insurance, central counterparties

JEL classification: G11, G23, G28

*We thank Ted Temzelides, Alberto Trejos, Pierre-Olivier Weill, Randall Wright, as well as audiences at the second African Search & Matching Workshop, the 2012 Money, Banking, and Liquidity Summer Workshop at the Chicago Fed, the 2012 Banking and Liquidity Wisconsin Conference, and the 2013 meetings of the Society for Economic Dynamics for very helpful comments. The views expressed do not necessarily reflect the ones of the BIS or the Federal Reserve Board.

1 Introduction

Many financial markets operate through dealers, market makers, or similar intermediaries. The main function of dealers is to intermediate trades between buyers and sellers of assets, possibly tailoring trades to specific customers' needs, thus promoting liquidity in over-the-counter (OTC) markets. Following the financial crisis, in 2009 leaders of G20 countries adopted measures that have had or will have an impact on market making activities. One of the most important measures is mandatory clearing of standard derivatives contracts through a central counterparty (CCP). By novation, a CCP becomes the buyer to the seller and the seller to the buyer of every contract it clears.[1] As a result it is commonly thought that CCP clearing reduces counterparty risk, thus rendering markets more stable.[2] The Chairman of the Commodity Futures Trading Commission (CFTC), Gary Gensler, announced the beginning of mandatory clearing for swaps in March 2013 as a "historic change for the market, that would benefit the public and the economy at large", specifying that: "Central clearing lowers the risk of the highly interconnected financial system. It promotes competition in and broadens access to the market by eliminating the need for market participants to individually determine counterparty credit risk, as now clearinghouses stand between buyers and sellers." Currently, although central clearing rates have increased globally, there still is a significant proportion of OTC derivatives that is not cleared centrally. As the regulatory framework is being implemented, and as changes in the infrastructure landscape for trading and settlement take place (e.g., due to Brexit and the introduction of blockchain and distributed ledger technologies), little is known about the effects of these reforms on the structure of the market in which they are implemented. In fact, as the Financial Stability Board (FSB) indicates in its Quarterly Review (Financial Stability Board [2017]), "Further analysis is required to understand the effects of the reforms on spreads, post-trade transparency and liquidity, as the market adjusts to the reforms and implementation continues across member jurisdiction." That's where our paper fits in.

While the existing literature has analyzed how CCP clearing can ease market stress, there have been very few studies analyzing the impact of CCP clearing on the structure of OTC markets

[1]See Monnet [2010] for a primer on CCP clearing. And See BIS [2016]: "novation is defined as the satisfaction and discharge of existing contractual obligations by means of their replacement by new obligations (whose effect, for example, is to replace gross with net payment obligations). The parties to the new obligations may be the same as those to the existing obligations or, in the context of some clearing house arrangements, there may additionally be substitution of parties."

[2]See Dudley [2012]: "The idea is that because some dealers are likely to be long some trades to one dealer and short some trades to other dealers, passing both sets of trades through a single CCP can reduce the aggregate amount of risk in the system".

and their functioning in normal times.[3] In particular, is it obvious that central clearing promotes competition as Chairman Gensler claims? And, if so, is that necessarily associated with a Pareto improvement in the equilibrium allocation?

In this paper, we analyze the effects of introducing measures aimed at reducing counterparty risk on dealers' entry/exit, bid-ask spreads, and welfare of dealers and end-users in normal times.[4] Furthermore, we analyze the impact of reducing counterparty risk on the incentives of dealers to innovate in better market making technologies. One may expect that initiatives aimed at reducing counterparty risk would bring uncontested benefits. However, in line with the theory of the second best, we show that such initiatives may to some extent "back-fire": dealers can have too little incentives to innovate and could take actions yielding to inefficient outcomes.

In order to capture the effect of the reforms on competition, we use a simple set-up with monopolistically competitive dealers who intermediate trades between buyers and sellers. In equilibrium, mandatory central clearing affects competition via entry and exit into market making by dealers. Dealers are heterogeneous, and indexed by their cost of executing transactions. Furthermore, they can invest into superior market making technologies, which lower their expected cost of intermediating trades, thus making them more efficient ex-ante. This technology stands in for more efficient balance sheet management, a larger network of investors, access to lower cost of settlement –for example via blockchain and distributed ledger technologies– etc. Once they decide to invest, dealers post and commit to bid and ask prices. Buyers and sellers sample dealers randomly and decide whether to trade at the posted bid or ask price, or whether they should carry on searching for a dealer next period. The search friction implies that the equilibrium bid-ask spreads will be positive. It also implies that even the least efficient dealers will be active because buyers and sellers may be better off accepting an offer rather than waiting for a better offer. Therefore, our search friction defines the structure of the market, measured by how many and which dealers are operating, and its liquidity, measured by the distribution of bid-ask spreads.

In order to analyze the effects of regulations aimed at lowering counterparty risk[5] on the liquidity and the structure of intermediated markets, we introduce counterparty risk by allowing dealers to be exposed to the risk of having to hold inventories, differently from Duffie et al. [2005]. To make markets, dealers have to accommodate buy-orders with sell-orders. However we assume

[3] Acharya and Bisin [2014] focus on the effect of central clearing on transparency of trades and its consequences for collateral management. Biais et al. [2016] characterize the optimal design of central clearing when risk taking incentives can lead to contagion. Duffie and Zhu [2011], Garratt and Zimmerman [2017], Carapella and Mills [2012] study the role of multilateral netting in reducing exposures and fostering liquidity. One exception is Koeppl [2013] who studies moral hazard in the choice of trading counterparty in away from times of stress.

[4] See Vuillemey [2019] for evidence that central clearing facilitated entry of new dealers when the first CCP was created in France in 1882.

[5] See Financial Stability Board [2017] pg.3, 22.

that buyers (and sellers) can default after placing their orders. If dealers could perfectly forecast how many buyers will default, they would just acquire fewer assets. Otherwise they may find themselves with too many assets in inventory for longer than expected. For simplicity, we make the extreme assumption that dealers cannot sell the asset if the buyer defaults.[6] Facing the risk that a buy order will not be settled, dealers maximize their expected profits by posting bid and ask prices that depend on such inventory risk as well as on their cost of intermediating transactions. Due to these costs, less efficient dealers may find it optimal to stay out of market making activities.

As the model delivers predictions about entry and bid-ask spreads depending on the nature and magnitude of counterparty risk, we are able to address the questions raised by regulators (e.g. the CFTC and the FSB). In particular, we focus on the effects of the mandatory central clearing on (i) the measure of active dealers, buyers and sellers, (ii) the share of the market that each dealer services, and (iii) the equilibrium distribution of bid-ask spreads. Such a comprehensive characterization of the equilibrium also identifies who gains and who loses from such regulations.

We model the central clearing mandate as a reduction in the severity of counterparty risk which affects dealers' inventory risk.[7] Intuitively, dealers should benefit from a reduction or elimination of inventory risk. Everything else constant, a reduction in counterparty risk will result in a reduction of the bid-ask spread. Two distinct mechanisms are responsible for the lower spread. First, facing lower default risk, dealers prefer to charge a lower mark up per transaction and execute a larger volume of transactions. Second, a reduction in counterparty risk improves competition by inducing less efficient dealers to enter the market. As a result, the measure of dealers active in the market increases. More efficient dealers, however, make fewer profits in equilibrium because they lose some market share to the newly entered, less efficient dealers. In fact, the most efficient dealers would prefer some counterparty risk as long as other dealers are also facing some counterparty risk.

We also analyze the impact of a reduction in counterparty risk on dealers' incentives to adopt a better market-making technology that lowers their expected intermediation cost. Protection against risk can induce dealers to opt for a worse market making technology. As discussed, reducing risk allows less efficient dealers to enter the market. This additional competition reduces profits of more efficient dealers, thus lowering their incentives to invest in better market-making technologies. In turn the average dealer is less efficient and this adversely impacts buyers and sellers who face worse terms of trade on average. Under some assumptions on the distribution of dealers' trading cost, their choice of innovation is always inefficiently low. As a consequence, the introduction of a

[6]See Lagos et al. [2011] for a model where dealers hold inventories. Our results go through even if we assumed that dealers could obtain some value for the asset, as long as this value is not too high.

[7]See Vuillemey [2019] for evidence "that central clearing increased the ability of coffee dealers to hedge their inventories".

seemingly beneficial insurance mechanism against counterparty risk reduces welfare of buyers and sellers, unless dealers receive a transfer to compensate them for their investment into more efficient market making technologies.

Finally, in light of these results, this paper makes two additional contributions: first it explains the opposition of some dealers to tighter regulation, such as mandatory central clearing for all standardized derivatives traded OTC.[8] Second, it argues that forcing the adoption of seemingly beneficial regulation can have adverse consequences on welfare by affecting the incentives of some market participants. To be clear, we are not claiming that counterparty risk should not be reduced, rather we are suggesting that mandating central clearing has costs that have been largely ignored in regulatory and academic discussions. So when accounting for the desirability of CCP clearing our paper shows that its benefits for the financial system should compensate for the loss of efficiency due to, for example, a lack of innovation.

1.1 Related literature

The literature on the microstructure of markets is large and has been mostly focused on explaining bid-ask spreads. It is not our intention to cover this literature here, and we refer the interested reader to O'hara [1995]. Among the first to study the inventory problem of market makers are Amihud and Mendelson [1980]. Here, we are not interested in the inventory management problem per-se as much as in how the cost of managing inventories affects liquidity. In particular, we normalize the optimal size of inventory to zero and we analyze how the probability to experience deviations from this optimal inventory level affects liquidity.

Our paper, by focusing on the effect of competition on the adoption of better market-making technologies, is also related to Dennert [1993] and Santos and Scheinkman [2001]. Following the seminal contribution of Kyle [1985], Dennert [1993] analyzes the effect of competition on bid-ask spreads and liquidity, and shows that liquidity traders might prefer to trade with a monopolist market maker. Santos and Scheinkman [2001] study the effects of competing platforms when there is a risk of default. They show that a monopolist intermediary may ask for relatively little guarantee against the risk of default.

The papers that are most related to ours are the equilibrium search models of Spulber [1996] and Rust and Hall [2003], which we extend by introducing inventory risk through the default of

[8]See BIS [2009], pg 52: "The risk management of the CCP [...] should result in a higher overall posting of financial resources relative to potential losses than with OTC markets. However, this, combined with the perceived high costs of operating CCPs, may in part explain the long-standing opposition among key market participants to the introduction of CCPs.". Also, see Vuillemey [2019] for evidence of "early opposition by some trading houses" to the creation of the *Caisse de Liquidation des Affaires en Marchandises* in Le Havre (France) in 1882 (pg. 21).

buyers. Duffie et al. [2005] present an environment where market makers are able to trade their inventory imbalances with each other after each trading rounds. Therefore, market makers never carry any inventory in equilibrium. We depart from Duffie et al. [2005] by assuming that market markets may have to hold inventories and we study the effect of regulations, whose goal is to make market makers closer to the set-up in Duffie et al. [2005], on the structure of the market. In an environment similar to Duffie et al. [2005], Weill [2007] shows that competitive market makers offer the socially optimal amount of liquidity, provided they have access to sufficient capital to hold inventories. Weill [2011] shows that if market makers face a capacity constraint on the number of trades which they can conduct, then delays in reallocating assets among investors emerge, thus creating a time-varying bid ask spread, widening and narrowing as market makers build up and unwind their inventories. In contrast to the last papers, we analyze the incentives of dealers to enter market making activities in the first place. In this respect, our paper is also related to Atkeson et al. [2015], who study the incentives of ex-ante heterogenous banks to enter and exit an OTC market. This allows Atkeson et al. [2015] to identify the banks which behave as end users versus the banks which intermediate transactions, and thus behave as dealers. In contrast, we analyze the impact of current OTC market reforms on dealers' entry and investment decisions, and on the efficiency of the resulting equilibrium allocation.

In the empirical literature, there is strong evidence that inventory costs and spreads are tightly related: Comerton-Forde et al. [2010] find that losses on inventories widen effective spreads for firms trading individual NYSE stocks; Rapp [2016] finds that dealers' inventory financing costs widen bid-ask spreads in the US corporate bonds market. Moreover, our results are consistent with empirical findings on the effects of mandatory central clearing for Credit Default Swaps indexes in the United States. Studying separately the effects of each implementation phase of the Dodd Frank reform, Loon and Zhong [2016] find that the effect of central clearing on a measure of transaction-level spread is significantly different according to the category of market participants affected by the reform. In particular, central clearing is correlated with an increase in spreads for swap dealers and with a decrease in spreads for commodity pools and all other swap market participants.[9] In our model, the final general equilibrium effect of introducing an insurance mechanism against counterparty risk (e.g. central clearing) crucially depends on features of the market participants involved.

Section 2 describes the basic structure of the model. To understand the basic mechanism underlying our main results, we analyze the equilibrium with no counterparty risk (i.e. settlement fails) in Section 3 and the equilibrium with counterparty risk/settlement fails in Section 4. Section

[9] See Loon and Zhong [2016], Table 10 and Appendix A.2.1, pg. 667-9.

5 contains our result about dealers' lack of incentives to invest in a more efficient market making technology ex-ante. Section 6 describes the implications of the model for average bid-ask spreads and relates them to the empirical evidence in Loon and Zhong [2016]. Section 7 presents a numerical solution to the model of Section 4 which relaxes the assumptions necessary for the analytical characterization of the equilibrium in Section 4. Finally, Section 9 concludes.

2 A Model of Dealers and Risk

We base our analysis on a modified version of the search models in Spulber [1996] and Rust and Hall [2003]. The presentation of the model follows closely the one in Rust and Hall [2003].

There is a continuum $[0,1]$ of heterogeneous and risk neutral buyers, sellers, and dealers. A seller of type $v \in [0,1]$ can sell at most one unit of the asset at an opportunity cost v. A buyer of type $v \in [0,1]$ can hold at most one unit of the asset and is willing to pay at most v to hold it. A buyer consumes the asset on the spot. While dealers and sellers are infinitely lived, a fraction $\lambda \in [0,1)$ of buyers exits the market each period. A buyer who exits the market is replaced by a new buyer in the next period, whose type v is drawn from the uniform distribution over $[0,1]$. The timing of exit that we specify later will play an important role in our analysis. In the remainder of the paper we refer to buyers and sellers as traders when we do not need to distinguish between them. Traders cannot trade an asset directly, and all trades must be intermediated by dealers. To do so, traders engage in search for a dealer: if he decides to search, a trader gets a price quote from a dealer, drawn randomly from the distribution of dealers who are active in the market. A dealer of type $k \in [0,1]$ can execute a trade at cost k. Specifically, k denotes the marginal cost of processing a seller's order. We can interpret a dealer's trading cost as the ease of finding a counterparty for a specific trade (e.g. the price at which a trader would be willing to take a certain side in a transaction). As we discuss in detail below, we assume that a dealer must process a seller's order before the buyer's order is settled.[10] The most efficient dealer can process trades at cost $k = 0$. The distribution of dealers is given by the probability distribution function $f(\cdot)$ with cumulative density $F(\cdot)$. In equilibrium, only dealers who can make a profit will operate a trading post and there will be a threshold level of trading cost, $\bar{k} \leq 1$, such that no dealer with a cost greater than $\bar{k}$ operates a post. A dealer of type $k \in \left[0, \bar{k}\right]$ chooses a pair of bid-ask prices $(b(k), a(k))$ that maximizes his expected discounted profit. A dealer is willing to buy the asset at price $b(k)$ from a seller and is willing to sell the asset at the ask price $a(k)$. We consider a stationary equilibrium

[10] This introduces an asymmetry regarding the cost of dealing with a buyer or a seller, which can be justified in real contracts as the cost of handling the asset underlying the contract. Our results would be substantially identical if we introduced a handling cost of the buyer as well, k^b as long as $k^b < k$. Here we set $k^b = 0$.

so that $b(k)$ and $a(k)$ will be constant through time. Traders face distributions $G_a(a)$ and $G_b(b)$ of ask and bid prices. These distributions are equilibrium objects that depend on $F(\cdot)$ and $\bar{k}$.

We assume buyers exit the market after they placed an order with dealers but *before* they have the chance to settle their orders. So dealers face the risk of having to hold inventories. But in and of itself, this type of risk is not interesting: as dealers trade with a continuum of buyers, they will correctly anticipate that a fraction λ of them will not settle their trade. So they can just buy $1-\lambda$ assets. To analyze the effect of counterparty risk on dealers we assume that dealers face idiosyncratic risk: nature does not allocate buyers perfectly across dealers who can then be in two states, $s=1$ and $s=-1$. In state $s=1$, a dealer has a measure $\lambda-\varepsilon$ of his buyers exiting the market, while in state $s=-1$ a measure $\lambda+\varepsilon$ of his buyers exit. This default shock is independent of whether the buyers placed an order at the bid-ask spread posted by the dealer. Also dealers cannot observe the state before it occurs as they only observe the actual measure of buyers exiting the market. The shock ε is i.i.d. across dealers and each state occurs with probability 1/2. In this sense, there is no aggregate uncertainty and the average rate of buyers' exit before settlement is λ. The distribution of types of buyers and sellers is $v \sim U[0,1]$. Because one of the functions of CCPs is the mitigation of counterparty risk, we model central clearing as a reduction in ε.

To be clear, dealers cannot default on buyers, so they must hold enough assets to cover buyers' demand in all states of the world (that is, even if the default rate this period is lower than expected). Notice that there is a missing market: dealers with inventories have nowhere to sell their assets. Also, we do not consider strategic default. Thus we abstract from issues related to asymmetric information and counterparty monitoring.[11] Finally, in contrast to buyers, we assume sellers always settle their orders.[12]

Figure 1 shows the timing, which is as follows: dealer $k \in \left[0, \bar{k}\right]$ chooses bid and ask quotes. Buyers and sellers decide whether they want to search or not. If so, they contact a dealer at random, and they either accept the quoted price or not. If they agree, they place an order to buy/sell a unit of the asset. Then each buyer exits with probability λ. Moreover, if a dealer is in state $s \in \{-1, 1\}$, then a measure $\lambda - s\varepsilon$ of his buyers exit before settlement. Finally, settlement occurs: Each operating dealer receives assets from the sellers who placed an order and delivers one asset to each of the $(1-\lambda+s\varepsilon)$ buyers who settle their orders. Dealers obtain no value from the surplus of assets.[13]

[11] This is akin to focusing on the risk that a counterparty is in default for reasons that are independent of the nature of its trading activities.

[12] This asymmetry between buyers and sellers is not substantial. Analogous results would arise if sellers exited the market before settlement.

[13] We could assume that dealers gets some value $\bar{p}$ for each unit of asset they hold and we normalize $\bar{p}=0$, so that the asset fully depreciates in the hand of the dealers. This low holding-value can be interpreted as high regulatory

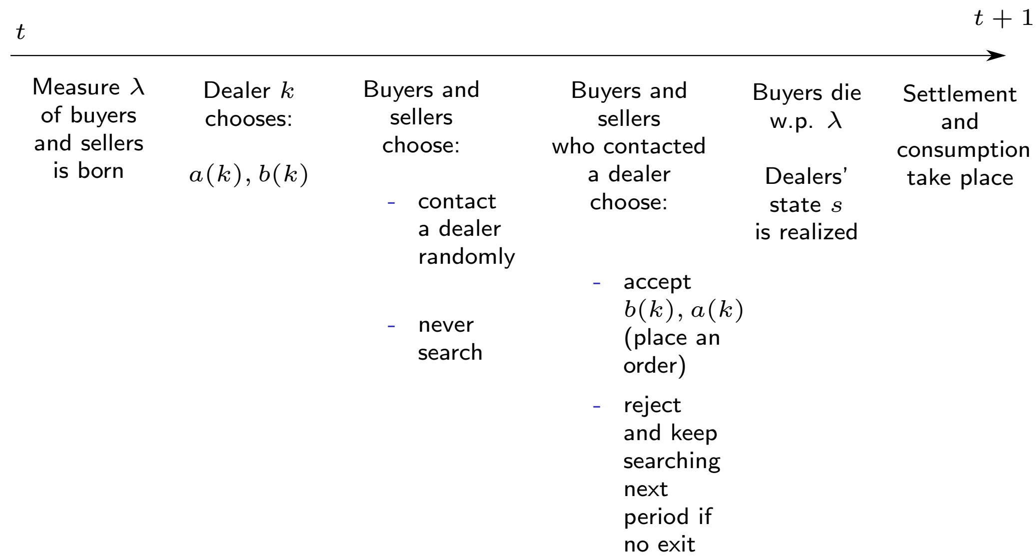

Figure 1: Timing

2.1 Discussion

We model counterparty risk as the risk that a dealer may face defaults from some of the buyers who placed orders at the beginning of the period. We interpret these defaults mainly as being triggered by settlement fails. Settlement risk is relatively common in financial markets: in fact the Treasury Market Practices Group introduced settlement fails charge for Treasury securities in 2009.[14] The goal was to mitigate settlement fails in order to limit market participants' net interest expenses as well as their exposure to the risk of counterparty insolvency. Fails charge incentivizes timely settlement by providing that a buyer of Treasury securities can claim monetary compensation from a seller if the seller fails to deliver on a timely basis.[15] Data from the Depository Trust and Clearing Corporation (DTCC), whose Fixed Income division serves as the clearing house for trades in U.S. government securities, shows that settlement fails are normal time phenomena as the daily average of US Treasury and Agency settlement fails is in a steady order of $40 billions from May 2018 to May 2019.

Counterparty risk in our model, however, is not necessarily only settlement risk, as it can also be triggered by a counterparty filing for bankruptcy. In either case, we think that times of normal

costs of holding some assets (such as higher capital requirements). All our results go through, however, if we allowed dealers to holding inventories, as long as that's sufficiently costly.

[14] See Marshall [2017]. Similarly, settlement risk is defined by the Bank for International Settlements as "the risk that settlement in a transfer system will not take place as expected" – see Glossary of terms used in payments and settlements systems, Committee on Payment and Settlement Systems, March 2003.

[15] See Garbade et al. [2010].

market conditions are times when shocks are primarily idiosyncratic and can be diversified by risk pooling, which is exactly what CCPs do. Since our goal is to characterize the implications of counterparty risk insurance on market activities away from times of stress, when systemic and correlated risks are less likely to arise, we model counterparty risk as idiosyncratic risk. We believe this modeling choice captures some of the regulator's concerns regarding the effect of the reforms on spreads, liquidity and competition.

Besides considering counterparty risk, the main difference between our model and those in Spulber [1996] and Rust and Hall [2003] is that while in those models buyers exit the market after they trade, we allow buyers to have future trading opportunities after they trade in a given period. Thus, in our model, trading decision of buyers and sellers are simpler. Because trading decisions are simpler, we can extend Spulber [1996] and Rust and Hall [2003] to the case where dealers face a generic distribution $F(k)$ of costs for market making, while they assume uniform distributions. This allows us to highlight the role of having a large measure of relatively inefficient dealers. In Section 8 we characterize the (in)sensitivity of our results to introducing the assumption that buyers exit the market once they trade as in Spulber [1996] and Rust and Hall [2003].

A common feature between our environment and those in Spulber [1996] and Rust and Hall [2003] is that each active dealer loses market share when new entrants become active. This is key to our results.

3 No counterparty risk

To gain some intuition, in this section we study the benchmark economy where there is no risk so that $\lambda = \varepsilon = 0$. The decision of buyers is simply to accept the quoted ask price a whenever $v \geq a$ and reject it otherwise. Similarly, the decision of sellers is to accept the quoted bid price b whenever $v \leq b$ and reject it otherwise. Hence, the expected buy orders – or demand – for dealers who post ask-price a is:

$$D(a) = \frac{1}{N}\int_a^1 dv = \frac{1}{N}(1-a) \tag{1}$$

where N denotes the measure of active dealers.[16] Similarly, dealers that post a bid-price b face the following demand

$$S(b) = \frac{1}{N}\int_0^b dv = \frac{1}{N}b \tag{2}$$

[16] Expectation is taken before being matched with buyers.

A dealer of type k maximizes his profit by choosing a and b

$$\Pi(k) = \max_{a,b} \{aD(a) - (b+k)S(b)\}$$

subject to the resource constraint $D(a) \leq S(b)$. The resource constraint is binding, so that $b = 1-a$ and a dealer chooses a to maximize

$$\Pi(k) = \frac{1}{N}(1-a)(2a-1-k)$$

with solution

$$a(k) = \frac{3+k}{4} \quad (3)$$

$$b(k) = \frac{1-k}{4} \quad (4)$$

Notice that, in the models of Spulber [1996] and Rust and Hall [2003], the distribution of bid and ask prices are uniform on $\left[a(0), a(\bar{k})\right]$ and $\left[b(\bar{k}), b(0)\right]$ because the bid and ask prices (3) and (4) are linear and the distribution of dealer cost is uniform. As described in Section 2, we instead allow for a general distribution $F(k)$ of dealers' trading cost, and we denote by $G_a(a)$ and $G_b(b)$ the distributions of ask and bid prices respectively.

In equilibrium, all dealers with intermediation cost k such that $\Pi(k) \geq 0$ will be active. Because dealers' profits are decreasing in k, then all dealers with $k \leq \bar{k}$ will be active, where $\bar{k}$ denotes the marginal active dealer who just breaks even: $\Pi(\bar{k}) = 0$. The measure of active dealers is then $N = F\left(\bar{k}\right)$. It is easy to see that $\bar{k} = 1$, $a(\bar{k}) = 1$, and $b(\bar{k}) = 0$, resulting in the marginal active dealer being indifferent between operating and staying out of the market. In fact, dealer $\bar{k}$ would face a measure zero demand at the price $a(\bar{k}) = 1$. Any dealer $k < \bar{k} = 1$ makes strictly positive profits given by:

$$\Pi(k) = \frac{(1-k)^2}{8F\left(\bar{k}\right)}.$$

Then we can find the extremes of the support of the bid and ask price distributions:

$$\overline{a} = a(\bar{k}) = \frac{3+\bar{k}}{4} = 1 \qquad \underline{a} = a(0) = \frac{3}{4}$$

$$\overline{b} = b(0) = \frac{1}{4} \qquad \underline{b} = b(\bar{k}) = \frac{1-\bar{k}}{4} = 0$$

Clearly, each dealer charges its monopoly price: the bid/ask prices posted by *other* dealers do not influence the decision of traders to accept or reject the price they obtain because, differently from the models in Spulber [1996] and Rust and Hall [2003], traders can search again next period independently of their decision in the current period. Since dealers charge the monopoly price, even relatively inefficient dealers (those with large values of k) can make profits, which implies that they have the incentive to enter the market: hence we should expect that the equilibrium number of active dealers is too high relative to what a planner would choose. We analyze this next.

To characterize the optimal number of dealers, we now define the surplus of dealers, buyers and sellers as a function of $\bar{k}$. The surplus of dealers is:

$$S_d(\bar{k}) \;=\; \int_0^{\bar{k}} \Pi(k)dF(k)$$

In Appendix A.3, we show how to obtain the surplus of buyers and sellers,

$$S_b(\bar{k}) \;=\; \int_{a(0)}^{1} \left[\int_{a(0)}^{a(\bar{k})\vee v} (v-a)dG_a(a)\right] dv = \frac{S_d(\bar{k})}{4} \tag{5}$$

$$S_s(\bar{k}) \;=\; \int_{0}^{b(0)} \left[\int_{b(\bar{k})\wedge v}^{b(0)} (b-v)dG_b\left(b\right)\right] dv = \frac{S_d(\bar{k})}{4} \tag{6}$$

From the above characterization it is easy to see that total surplus is decreasing in $\bar{k}$ whenever the surplus of dealers is. This is the case whenever

$$(1-\bar{k})^2 \leq \int_0^{\bar{k}} \frac{(1-k)^2}{F(\bar{k})} dF(k).$$

which is always satisfied at $\bar{k} = 1$. Therefore neither dealers nor traders benefit from the entry of relatively inefficient dealers. Given that intermediation by dealers is needed to execute transactions between traders, the solution to a planner's problem constrained by intermediation, but unconstrained in the matching process, is to have only the most efficient dealers, those with $k = 0$, intermediate all trades. Notice that this is the case because the most efficient dealer charges the same bid and ask prices independent of the presence of other dealers.[17] In the next section we introduce counterparty risk.

[17] As we discuss in Section 8, this is not true in a models similar to those in Spulber [1996] and Rust and Hall [2003], where even the most efficient dealer may wish to offer narrower spreads when other dealers are operating.

4 Counterparty risk

In this section we introduce counterparty risk for dealers. As discussed in Section 2, we interpret counterparty risk as settlement fails, which, in the context of our model, is defined as the event in which a buyer fails to collect and pay for his buy order. We assume that with probability 1/2 a measure $\lambda + \varepsilon$ of buyers fail to settle and otherwise a measure $\lambda - \varepsilon$ fails to settle. The shock ε captures our notion of counterparty risk.[18] Settlement fails are costly for dealers because they may be left holding inventories. As described in Section 2, we normalize to zero the value of inventories to dealers, by assuming that dealers have to dispose of the asset at the end of each period.[19]

The decision problems of buyers and sellers are the same as in the previous section, so demand and supply are, respectively: $D(a) = (1-a)/F(\bar{k})$ and $S(b) = b/F(\bar{k})$, where $\bar{k}$ denotes the trading cost of the marginal active dealer and is determined endogenously in equilibrium, as in the previous section. Dealers' decision problem is:

$$\Pi(k;\lambda,\varepsilon) = \max_{\{a,b\}} E_s\left\{a\left(1-\lambda+s\varepsilon\right)D\left(a\right)-\left(b+k\right)S\left(b\right)\right\} \tag{7}$$

$$s.t.\ \left(1-\lambda+s\varepsilon\right)D\left(a\right)\leq S\left(b\right) \quad \forall s\in\{-1,1\} \tag{8}$$

The resource constraint (8) binds when $s = 1$ as we do not allow dealers to default on their promises to deliver the asset.[20] Therefore

$$S(b) = \left(1-\lambda+\varepsilon\right)D\left(a\right)\equiv\lambda_{\varepsilon}D(a)=\lambda_{\varepsilon}(1-a)/F(\bar{k}). \tag{9}$$

Notice that dealers expect to have to deliver $(1-\lambda)D(a)$ assets. However, dealers have to purchase more securities than they expect will be necessary, as they have to satisfy their buy orders in all possible states. Hence, counterparty risk implies that dealers over-buy the asset. Substituting out for $D(a)$ and $S(b)$ yields:

$$\Pi(k;\lambda,\varepsilon) = \max_{a}\left\{a\left(1-\lambda\right)-\left[\lambda_{\varepsilon}\left(1-a\right)+k\right]\lambda_{\varepsilon}\right\}\frac{1}{F(\bar{k})}\left(1-a\right) \tag{10}$$

Inspecting (10) reveals that ask prices must be less than one (i.e. the highest valuation buyer) for dealers to make positive profits. Specifically, (10) implies that profits are zero at two roots, $a = 1$

[18] We can extend this to a symmetrically distributed ε around $[-\bar{\varepsilon},\bar{\varepsilon}]$, where $\bar{\varepsilon} < \lambda$ and $E(\varepsilon) = 0$. Then everything below holds with $\varepsilon = \bar{\varepsilon}$.

[19] As mentioned in Section 2, our results go through even if can hold inventories, as long as that is sufficiently costly.

[20] This assumption is without loss of generality for all of our results, to the extent that dealers' default on buy orders is costly and that innovation by dealers can reduce such cost ex ante.

and $a = (k\lambda_\varepsilon + \lambda_\varepsilon^2)/(1-\lambda+\lambda_\varepsilon^2)$, and are positive for some ask price $a \in (0,1)$ whenever the second root is less than one. Since a dealer with higher trading cost k necessarily makes lower profits, the positive profit region shrinks as k increases. Hence, the marginal active dealer is obtained when the positive profit region is empty, that is when the second root equals 1. Formally, the marginal active dealer, $\overline{k}$ such that $\Pi(\bar{k}; \lambda, \varepsilon) = 0$, is

$$\bar{k} = \frac{1-\lambda}{\lambda_\varepsilon} < 1. \tag{11}$$

Equation (11) implies that the measure of active dealers is decreasing in ε because an increase in counterparty risk on their buy-side requires dealers to acquire more assets from their sellers. This is costly, as every trade is executed at cost k. Hence the profits of every dealer decrease, the least efficient dealers exit the market, and the marginal active dealer is relatively more efficient. That is to say $\frac{\partial \pi(k)}{\partial \varepsilon} < 0$ and $\frac{\partial \overline{k}}{\partial \varepsilon} < 0$.

It is then easy to see that $a\left(\overline{k}\right) = 1$. Taking the number of operating dealers as given, Figure 2 shows the profits of dealer k, $\Pi(k, \lambda, \varepsilon)$, as a function of its ask quote as risk increases from $\varepsilon = 0$ to $\varepsilon > 0$. The direct effect of a discrete increase in risk is to reduce dealers' profits. Hence the curve $\Pi(k, \lambda, \varepsilon)$ lies below the curve $\Pi(k, \lambda, 0)$. Dealers respond by increasing their ask price (i.e. $a'(\varepsilon) > 0$). The mechanism driving this result is intuitive: by quoting ask price a, a dealer receives $D(a)$ buy orders but expects only $(1-\lambda)D(a)$ buyers to collect the asset and pay for it. However, a dealer needs to buy sufficiently many securities to cover effective demand in state $s = 1$. Because such demand increases with ε, an increase in ε reduces dealers' profits. To account for the loss in profits, dealers adjust their ask price upwards. As a consequence they face fewer buy orders, resulting in lower effective demand in state $s = 1$.

The first order conditions to dealers' decision problem imply:

$$a(\varepsilon, k) = 1 - \frac{1-\lambda-k\lambda_\varepsilon}{2\left(1-\lambda+\lambda_\varepsilon^2\right)} = \frac{1-\lambda+2\lambda_\varepsilon^2+k\lambda_\varepsilon}{2(1-\lambda+\lambda_\varepsilon^2)} \tag{12}$$

$$b(\varepsilon, k) = \lambda_\varepsilon(1-a(k)) = \lambda_\varepsilon \frac{1-\lambda-k\lambda_\varepsilon}{2\left(1-\lambda+\lambda_\varepsilon^2\right)} \tag{13}$$

It is worth discussing the effect of increasing risk on the bid-ask spread. Since the ask price is increasing with risk, fewer buyers place buy orders and demand is decreasing with risk. Because dealers need to serve fewer buyers one would expect them to reduce their bid price. This direct effect of risk on the bid price is reflected in the term $1 - a(k)$ in (13), and results in a decrease in the bid price. Higher counterparty risk, however, has also an indirect effect on the bid price, captured by the factor λ_ε multiplying $1 - a(k)$ in (13): as ε increases dealers have to buy more

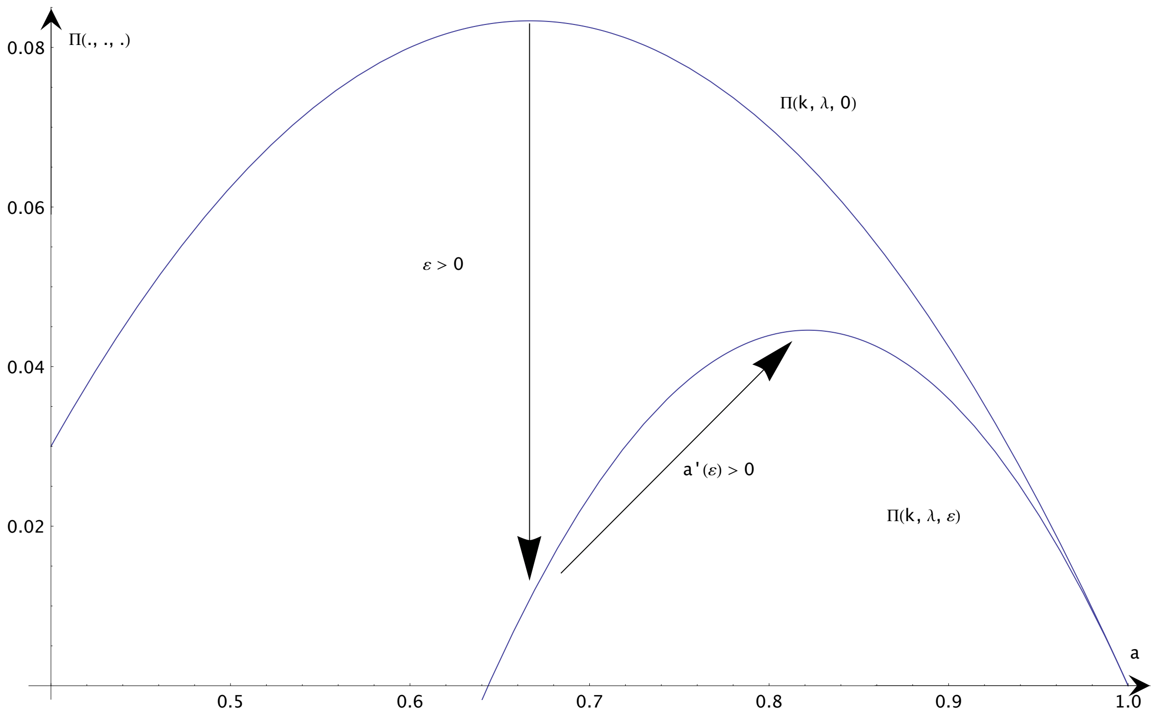

Figure 2: Dealer's profits as a function of ask price a

assets in order to satisfy the resource constraint (8) in state $s = 1$. This indirect effect results in an increase in the bid price. Therefore the overall effect of an increase in ε on the bid price is ambiguous. Still, we show that all dealers quote wider bid-ask spreads in response to an increase in ε.

Lemma 1. *The bid-ask spread $\sigma(\varepsilon, k) \equiv a(\varepsilon, k) - b(\varepsilon, k)$ is increasing in ε for all dealers $k \leq \bar{k}$.*

Proof. See Appendix A.5. □

Using the expressions in (9) for the demand and supply faced by dealer k, and (12) for the ask price, the profit function of dealer k becomes:

$$\Pi(k; \varepsilon) = \frac{\pi(k)}{F(\bar{k})} = \frac{(1 - \lambda - k\lambda_\varepsilon)^2}{4F(\bar{k})(1 - \lambda + \lambda_\varepsilon^2)} \tag{14}$$

where $\pi(k)$ is the per-trade profit of dealer k after being matched. We then obtain the main result of this section.

Lemma 2. *$\pi(k)$ is decreasing in ε for all k. The total surplus of dealers $S_d(\varepsilon) = \int_0^{\bar{k}} \Pi(k; \varepsilon) dF(k)$*

is decreasing in ε if and only if

$$-\frac{\partial \bar{k}}{\partial \varepsilon}\frac{f(\bar{k})}{F(\bar{k})}\int_0^{\bar{k}} \Pi(k)dF(k) + \int_0^{\bar{k}} \frac{\partial \pi(k)}{\partial \varepsilon}\frac{dF(k)}{F\left(\bar{k}\right)} < 0 \tag{15}$$

However, the surplus of the most efficient dealers increases in counterparty risk if and only if

$$\frac{f(\bar{k})}{F(\bar{k})}\bar{k} > \frac{2\lambda_\varepsilon^2}{1-\lambda+\lambda_\varepsilon^2} \tag{16}$$

or, equivalently, $e_{\pi(0),\varepsilon} > e_{f\left(\bar{k}\right),\varepsilon}$, where $e_{\pi(0),\varepsilon}$ denotes the elasticity of the profits of the most efficient dealer with respect to ε, and $e_{f\left(\bar{k}\right),\varepsilon}$ denotes the elasticity of the distribution of dealers' trading costs with respect to ε

Proof. See Appendix A.2.1. □

The result in Lemma 2 relies on a general equilibrium effect via exit of dealers from the market. The first term in (15) reflects the effect of risk on the composition of the market for dealers: this effect is positive because the least efficient dealers exit as risk increases. Hence the market is composed of more efficient dealers whose market share increases, resulting in a larger volume of trades per dealer. We refer to this effect as the intensive margin. However, this is balanced by the overall decline in profits for all active dealers, as shown by the second term in (15). We refer to this effect as the extensive margin. Overall, the total expected surplus is decreasing in ε whenever the effect on profits (the extensive margin, second term) is larger than the effect on exit resulting in more trades per active dealer (the intensive margin). To understand (16) recall that the most efficient dealers process the largest number of trades and make the largest profit per transaction. Thus, they may benefit from an increase in risk if any direct loss to the mark-up they charge on each transactions is compensated by gains from processing a larger number of transactions (i.e. they lose on the intensive margin but earn on the extensive margin). This is true if and only if counterparty risk is sufficiently small relative to the mass of dealers that are active at the margin (i.e. at $\bar{k}$) and exit the market as risk increases further. If instead (16) is violated, the cost of accepting additional buy orders that can fail is too large to be recovered by the profits made on a higher number of transactions. Hence (16) is equivalent to requiring that the elasticity of the profits of the most efficient dealers with respect to risk is larger than the elasticity of the distribution of dealers' trading costs with respect to risk.

Turning to a welfare analysis, we characterize the surplus of traders and dealers. With respect to the previous section, the surplus of buyers now has to take into account that they may not obtain the good if they fail to settle. Therefore, their surplus is scaled down by the probability of experiencing a settlement fail, which, from the perspective of a buyer, is simply λ. In Appendix A.3 we derive the expressions for the surplus of buyers S_b and sellers S_s as

$$S_b(\varepsilon) = \int_0^{\bar{k}} \frac{1}{2}(1-a(k))^2 \frac{f(k)}{F(\bar{k})} dk \tag{17}$$

$$S_s(\varepsilon) = \lambda_\varepsilon^2 S_b(\varepsilon) \tag{18}$$

Differently from our results in the previous section, given by (5) and (6), equation (18) implies that seller's surplus is always smaller than buyers' surplus. This is so even when $\varepsilon = 0$, as apparent from (18), and is due to the assumption that buyers exit the market with probability λ, whereas sellers do not. Consider a dealer k: take as given the posted ask price $a(k)$ and the asset demand, or buy orders, that it generates for the dealer, $D(a(k))$. Dealer k correctly anticipates that only a fraction $1-\lambda$ of the buy orders will settle. Because the dealer needs to purchase fewer assets from sellers, he adapts his pricing strategy and lowers his bid price relative to an economy with $\lambda = 0$. This, in turn, reduces sellers' surplus. [21]

The surplus of buyers may increase or decrease with risk. The ask price increases with risk, which reduces buyers' surplus. However, the least efficient dealers – offering the worst terms of trade – exit the market, as $\partial\overline{k}/\partial\varepsilon < 0$. This tends to benefit buyers. So the overall effect of risk on the surplus of buyers is not clear. The same applies for sellers. Notice, however, that changes in $a(k)$ and $b(k)$ only affect the distribution of the surplus, but not the total surplus from trade. Indeed, total surplus is

$$S(\varepsilon) \equiv S_d(\varepsilon) + S_s(\varepsilon) + S_b(\varepsilon) = \frac{3}{2}\int_0^{\bar{k}} \Pi(k) dF(k) \tag{19}$$

which is decreasing in risk whenever the dealer's total surplus is decreasing in risk, that is when (15) holds. If the mass of dealers at $\bar{k}$ is sufficiently large then an increase in risk results in an increase in total surplus. This is due to the indirect effect of risk on exit of relatively inefficient dealers: as risk increases, a larger fraction of inefficient dealers remain inactive, which raises aggregate surplus. The following Proposition summarizes the results so far.

Proposition 1. *The most efficient dealers benefit from an increase in counterparty risk if (16) holds. The total surplus from trade, given by (19), is increasing in ε whenever the surplus of dealers is, that is when $f(\bar{k})$ is large enough, as defined by (15).*

[21] Because the asymmetry between buyers and sellers is without loss of generality, as we discussed in Section 2, then all of our results would go through even in a model where both buyers and sellers may exit the market. The only difference would be equation (18), which would imply that sellers' surplus is the same buyers' surplus.

Proof. See Appendix A.2.1. □

Before moving to the next section we would like to draw some preliminary conclusions of our analysis for CCP clearing. Our model implies that the introduction of a CCP, by providing insurance against counterparty risk (ε), promotes dealers' entry because more inefficient dealers can make a positive profit. This results in a smaller market share for the most efficient dealers. In the next section, we analyze how these results affect dealers' decision to adopt a better market making technology.

5 Dealers' incentive to innovate

In this section we study the incentives of dealers to innovate on their market making technologies. We model innovation as a costly investment decision of dealers to become more efficient in intermediating transactions between buyers and sellers. Specifically, we assume that dealers are endowed with two distributions of trading cost $F_1(\cdot)$ and $F_2(\cdot)$, and that $F_1(\cdot)$ first order stochastically dominates $F_2(\cdot)$ (i.e. $F_1(k) \leq F_2(k)$ for all k). Since $F_2(k)$ places more weight on lower values of the trading cost k, everything else equal, dealers would prefer to draw from $F_2(\cdot)$. However, we will assume it is costly to draw from the more efficient distribution. Precisely, dealers can choose the probability $\rho \in [0,1]$ of drawing from F_2 at the cost $\gamma(\rho)$. We assume $\gamma(\cdot)$ is increasing, convex, $\lim_{\rho\to 0} \gamma'(\rho) = \gamma(0) = 0$ and $\lim_{\rho\to 1} \gamma'(\rho) = +\infty$. Alternatively, dealers choose the distribution $\rho F_2(\cdot) + (1-\rho)F_1(\cdot)$ from which they will draw their trading cost, by paying the cost $\gamma(\rho)$.

Intuitively, a dealer gains from becoming more efficient because he is more likely to draw a relatively low trading cost k, which results in larger profits from both a wider bid ask spread and from a larger volume of intermediated transactions. Therefore, dealers have an incentive to innovate if the cost $\gamma(\cdot)$ is not too large. Both buyers and sellers benefit from being matched with more efficient dealers, so innovation also has benefits for the economy as a whole.[22]

The introduction of a CCP, by affecting dealers' entry as described in Proposition 1, has the unintended consequence of reducing dealers' incentives to innovate. Intuitively, while the cost of adopting the better technology does not change, the introduction of a CCP reduces the market share of more efficient dealers thus lowering the benefit of becoming more efficient. Buyers and sellers, then, may also be worse off with a CCP because they are less likely to be matched and trade with efficient dealers.

[22] As we explain after Proposition 3, while dealers do not necessarily rip all those benefits, they retain some incentives to invest in the low cost distribution under some assumptions.

5.1 Dealers' choice

One benefit of our modeling choice for innovation is that innovation does not affect the marginal active dealer, whose trading cost is still $\bar{k} = (1-\lambda)/\lambda_\varepsilon$ and independent of ρ. Therefore, given all other dealers choose probability $\bar{\rho} \in [0,1]$, the number of active dealers is $F(\bar{k};\bar{\rho}) \equiv (1-\bar{\rho})F_1(\bar{k}) + \bar{\rho}F_2(\bar{k})$, which is increasing in ρ since $F_2(k) \geq F_1(k)$ for all k. Then the problem of a dealer – given all other dealers choose $\bar{\rho}$ — is to choose $\rho \in [0,1]$ to maximize ex-ante expected profits (before drawing their k and before being matched):

$$(1-\rho)\int_0^{\bar{k}} \frac{\pi(k)}{F(\bar{k};\bar{\rho})} dF_1(k) + \rho \int_0^{\bar{k}} \frac{\pi(k)}{F(\bar{k};\bar{\rho})} dF_2(k) - \gamma(\rho)$$

The first order condition gives

$$\int_0^{\bar{k}} \frac{\pi(k)}{F(\bar{k};\bar{\rho})} dF_2(k) - \int_0^{\bar{k}} \frac{\pi(k)}{F(\bar{k};\bar{\rho})} dF_1(k) = \gamma'(\rho)$$

As $\pi(k)$ is strictly decreasing in k and finite, and $F_1(k) \leq F_2(k)$, the LHS is strictly positive and finite and shows the marginal benefit of investing in the low cost technology: it is the difference in expected profit of drawing more heavily from the more efficient distribution. The RHS is also positive and is the marginal cost of investing more in the low cost technology. Since $\gamma'(0) = 0$ and $\gamma'(1) = +\infty$ there is a unique ρ that solves this equation. The equilibrium is the fixed point $\bar{\rho}$ that solves

$$\int_0^{\bar{k}} \pi(k) dF_2(k) - \int_0^{\bar{k}} \pi(k) dF_1(k) = \gamma'(\bar{\rho}) F(\bar{k};\bar{\rho}) \tag{20}$$

Again, the LHS is positive and constant in $\bar{\rho}$. Since $F_1(k) \leq F_2(k)$ for all k, $F(\bar{k};\bar{\rho})$ is increasing in $\bar{\rho}$. Hence the RHS is increasing in $\bar{\rho}$. So there is a unique $\bar{\rho}$ that solves (20). In this section, we are interested in the effect of a change in risk ε on the dealers' choice to innovate. To do so, it is useful to define the total surplus generated by dealer k as

$$S(k,\varepsilon) = \frac{3}{2} \frac{\pi(k)}{F(\bar{k};\bar{\rho})} \tag{21}$$

so that total surplus is $\int_0^{\bar{k}} S(k,\varepsilon) dF(k;\bar{\rho})$. We have the following result.

Lemma 3. *The equilibrium choice of ρ by dealers is increasing in risk (i.e. $\bar{\rho}'(\varepsilon) \geq 0$) if and only if*

$$H(\bar{\rho}(\varepsilon), \varepsilon) \equiv \int_0^{\bar{k}} \frac{\partial S(k,\varepsilon)}{\partial \varepsilon} \left[f_2(k) - f_1(k)\right] dk \geq 0. \tag{22}$$

Proof. See Appendix A.4.2. □

This result is intuitive: when risk increases dealers innovate more if the loss in expected surplus due to the increased risk is lower when they draw their trading cost from $f_2(\cdot)$ than from $f_1(\cdot)$. A necessary condition for (22) to hold is that at least the most efficient dealers generate more surplus when risk ε increases, so (22) requires (16). In the case with innovation, we can write (16) as

$$\frac{f\left(\bar{k};\bar{\rho}\right)}{F\left(\bar{k};\bar{\rho}\right)}\bar{k} > \frac{2\lambda_{\varepsilon}^{2}}{\left(1-\lambda+\lambda_{\varepsilon}^{2}\right)} \tag{23}$$

and this is a necessary and sufficient condition for $\frac{\partial S(k,\varepsilon)}{\partial\varepsilon}|_{k=0} > 0$. In words, (23) indicates that, following an increase in risk, the measure of least efficient dealers should fall sufficiently so that the matching probability for more efficient dealers increase enough to make up the loss in profit due to more risk.

We can illustrate this result by using the following simple albeit extreme example: Suppose the more efficient distribution F_2 has more mass at $k = 0$ than F_1, while $f_2(k) = f_1(k)$ for all $k \in (0, \bar{k}]$. So f_2 is the same as f_1 but it allocates more mass to the most efficient dealers and less mass to some dealers that are always inactive. So over the range of active dealers $(0, \bar{k}]$, $F_2(\cdot)$ first order stochastically dominates $F_1(\cdot)$. This example is useful as it illustrates the necessary conditions for our results and by continuity our results will also hold in a neighborhood region of that distribution. We will call this example for f_2, the mass-at-zero distribution. Using the mass-at-zero distribution, it is easy to see that the condition for $\bar{\rho}'(\varepsilon) \geq 0$ is exactly (23), since only the surplus of the most efficient dealers matters for the optimal choice of ρ. So dealers choose to innovate whenever the matching probability is sufficiently elastic relative to risk when evaluated at $\bar{\rho}$.

In general, and for completeness, we can derive a sufficient condition such that $\bar{\rho}'(\varepsilon) \geq 0$ as a function of primitives. To ease exposition, it is useful to define several elasticities, first $e_{\pi,\varepsilon} = -\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}\right)/\pi(k,\varepsilon)$ is the ε-risk elasticity of profits of dealer k, second $e_{\pi'(k),\varepsilon} = -\left(\frac{\partial^2\pi(k,\varepsilon)}{\partial\varepsilon\partial k}\right)/\frac{\partial\pi(k,\varepsilon)}{\partial k}$ is the elasticity of the marginal profit with respect to risk, and finally $e_{f(\bar{k}),\varepsilon} = -\frac{f(\bar{k})}{F(\bar{k})}\frac{\partial\bar{k}}{\partial\varepsilon}$ is the ε-risk elasticity of the distribution at $\bar{k}$. Given the formula for $S(k,\varepsilon)$ we have:

$$\frac{\partial S(k,\varepsilon)}{\partial\varepsilon} = S(k,\varepsilon)\left[e_{f(\bar{k}),\varepsilon} - e_{\pi,\varepsilon}\right]$$

and

Proposition 2. *Assume*

$$\frac{f_2(\bar{k})}{F_2(\bar{k})}\bar{k} \quad > \quad \frac{\lambda_\varepsilon^2 - (1-\lambda)}{(1-\lambda+\lambda_\varepsilon^2)} \tag{24}$$

Then $\bar{\rho}'(\varepsilon) \geq 0$ *if*

$$\int_0^{k^0} -\left(F_2\left(k\right) - F_1\left(k\right)\right)\left(e_{f_2\left(\bar{k}\right),\varepsilon} - e_{\pi'(k),\varepsilon}\right)\frac{\partial \pi(k)}{\partial k}dk \quad \geq$$
$$\int_{k^0}^{\bar{k}} \left(F_2\left(k\right) - F_1\left(k\right)\right)\left(e_{f_2\left(\bar{k}\right),\varepsilon} - e_{\pi'(k),\varepsilon}\right)\frac{\partial \pi(k)}{\partial k}dk$$

where $k_0 = \left\{k \in \left(0, \bar{k}\right) : e_{f_2\left(\bar{k}\right),\varepsilon} = e_{\pi'(k),\varepsilon}\right\}$.

Proof. See Appendix A.4.2. □

Proposition 2 highlights a tradeoff arising from the choice of innovation. Once they select their efficiency, dealers can be in two groups. Either they are relatively efficient and $k \leq k_0$ or they are not. Those dealers with $k \leq k_0$ will benefit from added risk because $e_{f_2\left(\bar{k}\right),\varepsilon} \geq e_{\pi'(k),\varepsilon}$ so that their matching probability increases sufficiently to compensate for their loss in profit. To the contrary, less efficient dealers with $k > k_0$ will lose from added risk because they profit decreases too much to be compensated by the higher matching probability, $e_{\pi'(k),\varepsilon} > e_{f_2\left(\bar{k}\right),\varepsilon}$. Then Proposition 2 says that risk will increase innovation if the more efficient distribution places more weight on dealers who benefit from increased risk than on those dealers who don't.

5.1.1 Optimal and constrained-optimal choice of investment

In this section, we compare the equilibrium innovation choice with the innovation level chosen by a planner. We derive sufficient conditions so that a planner would choose to innovate more than dealers would actually do. We first solve the unconstrained planner problem where she can choose the level of risk ε as well as innovation ρ. Then we analyze the planner's problem who can choose the risk ε, but is constrained by the equilibrium choice of innovation $\bar{\rho}$ that we derived above.

The unconstrained planner's problem Suppose the planner can choose the level of risk ε as well as innovation ρ. The planner's choice of ε and ρ would be the one maximizing total surplus net of the cost of selecting k from the better distribution, that is

$$\max_{\varepsilon\in[0,\lambda],\rho\in[0,1]} \int_0^{\bar{k}} S(k,\varepsilon)dF(k;\rho) - \gamma(\rho) \tag{25}$$

where $S(k,\varepsilon)$ is defined in (21). Suppose that the planner's problem is concave in ρ, ε,[23] then the planner's choice of ε^* solves

$$\int_0^{\bar{k}} \frac{\partial S(k,\varepsilon)}{\partial \varepsilon} dF(k;\rho) \quad \leq \quad 0 \tag{26}$$

with strict inequality if $\varepsilon^* = 0$.[24] The planner's choice of investment, given ε, denoted by $\rho^*(\varepsilon)$ solves

$$\int_0^{\bar{k}} \pi(k)\left[f_2(k) - f_1(k)\right]dk - \int_0^{\bar{k}} \pi(k)\frac{\left[F_2\left(\bar{k}\right) - F_1\left(\bar{k}\right)\right]}{F(\bar{k};\rho^*)} dF(k;\rho^*) \quad = \quad \frac{2}{3}\gamma'(\rho^*)F\left(\bar{k},\rho*\right) \tag{27}$$

Define $C(\rho)$ as the loss in total profit due to reduced matching probability,

$$C\left(\rho\right) = \int_0^{\bar{k}} \pi(k)\frac{\left[F_2\left(\bar{k}\right) - F_1\left(\bar{k}\right)\right]}{F(\bar{k};\rho)} dF(k;\rho). \tag{28}$$

The following Proposition characterizes the planner's choice of ρ and compares it with the equilibrium choice of ρ by dealers.

Proposition 3. *There exists a unique $\rho^* \in [0,1]$ satisfying (27), and $\rho^* > \bar{\rho}$ if and only if*

$$\int_0^{\bar{k}} \pi(k)\left[f_2(k) - f_1(k)\right]dk > 3C(\rho^*).$$

A sufficient condition for $\rho^* > \bar{\rho}$ is:

$$3\frac{F_1\left(\bar{k}\right)}{F_2\left(\bar{k}\right)} - 2 \quad > \quad \frac{\int_0^{\bar{k}} \pi(k)dF_1\left(k\right)}{\int_0^{\bar{k}} \pi(k)dF_2\left(k\right)}. \tag{29}$$

Proof. See Appendix A.4. □

A necessary and sufficient condition for $\rho^* > \bar{\rho}$ is that the difference in expected profits between the two distributions F_2 and F_1 exceeds the loss of profits due to reduced matching probability for dealers, buyers, and sellers (hence the multiplication by 3 of the term $C\left(\rho^*\right)$ in Proposition 3).

Notice that the result in Proposition 3 is robust to a different choice of trading protocol where

[23] We show examples where this is the case below.

[24] Also and $\int_0^{\bar{k}} \frac{\partial S(k,\varepsilon)}{\partial \varepsilon} dF(k;\rho) > 0$ if $\varepsilon = \lambda$. To obtain (26) notice that the derivative with respect to $\bar{k}$ cancels out since $S(\bar{k},\varepsilon) = 0$

dealers extract all the surplus from each of the trades they execute. In fact, while buyers and sellers are appropriating some surplus from trade, the inefficient choice of ρ stems from the effect that the entry of less efficient dealers has on profits of already active dealers. As the provision of insurance introduced by central clearing allows them to make positive profits, less efficient dealers effectively steal market share from incumbent and more efficient dealers, eroding the profits of the latter. A decrease in profits then reduces dealers' benefits from innovation. Since the cost of innovation is unaffected by entry, dealers' equilibrium choice of innovation is then lower than in a model with a fixed measure of active dealers.

The constrained planner's problem We now consider the constrained planner's problem, where the planner takes the investment decision of dealers $\bar{\rho}(\varepsilon)$ as given by (20). Then the planner chooses ε to solve:

$$\max_{\varepsilon\in[0,\lambda]} \int_0^{\bar{k}} S(k,\varepsilon;\bar{\rho}(\varepsilon))dF(k;\bar{\rho}(\varepsilon)) - \gamma(\bar{\rho}(\varepsilon)) \tag{30}$$

If the planner's problem is concave in ε the first order condition is necessary and sufficient to characterize a solution. Let $\hat{\varepsilon}$ denote the solution to this maximization problem, that is the constrained optimal level of risk. Using the definitions of $C(\rho)$ in (28) and using $A = \int_0^{\bar{k}} \pi(k)dF_2(k)$, $B = \int_0^{\bar{k}} \pi(k)dF_1(k)$ we show in Appendix (A.4.4) that the first order condition to problem (30) can be rearranged as:

$$H(\bar{\rho}(\varepsilon),\varepsilon) + \frac{\partial\bar{\rho}}{\partial\varepsilon}\frac{[A - B - 3C(\rho)]}{2F(\bar{k},\bar{\rho})} \quad\leq\quad 0 \tag{31}$$

if $\hat{\varepsilon} \in [0,\lambda)$, and with the inequality reversed if $\hat{\varepsilon} = \lambda$, where recall that H is given by (22). Notice that $H(\rho^*,\varepsilon)$ is the unconstrained choice of ε by the planner, defined by (26). Then at an interior solution for ε^*, $H(\rho^*,\varepsilon) = 0$; $H(\rho^*,\varepsilon) < 0$ if $\varepsilon^* = 0$ and $H(\rho^*,\varepsilon) > 0$ if $\varepsilon^* = \lambda$. In Proposition 4 we will use these relations, combined with the results from the previous Lemmas, to conclude that the constrained planner chooses more risk than the unconstrained one, $\hat{\varepsilon} \geq \varepsilon^*$ for all $\lambda \in (0,1)$ under some conditions. To do that, Lemma 4 characterizes useful properties of $H(\rho,\varepsilon)$ under two additional assumptions: 1) a stricter stochastic order between F_1 and F_2, and 2) the elasticity of the distribution function is smaller than the elasticity of efficient dealers' profits with respect to ε.

Lemma 4. *Suppose (24) holds. Also assume that F_1 conditionally stochastically dominates F_2 so* *that* $\frac{F_2(k)}{F_2(\bar{k})} < \frac{F_1(k)}{F_1(\bar{k})}$ *for all* k*. Let* k_0 *be defined as in Proposition 2. Then* $\frac{\partial H(\rho,\varepsilon)}{\partial \rho} \leq 0$ *for all* ρ, ε *if*

$$\int_0^{k_0} \frac{\partial \pi(k)}{\partial k} \left[\left\{ e_{f_2(\bar{k}),\varepsilon} - e_{\pi'(k),\varepsilon} \right\} \frac{\partial}{\partial \rho} \left(\frac{F(k;\rho)}{F(\bar{k},\rho)} \right) - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{F(k;\rho)}{F(\bar{k},\rho)} \frac{\partial}{\partial \rho} \left(\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)} \right) \right] dk \quad \geq$$

$$\int_{k_0}^{\bar{k}} \frac{\partial \pi(k)}{\partial k} \left[- \left\{ e_{f_2(\bar{k}),\varepsilon} - e_{\pi'(k),\varepsilon} \right\} \frac{\partial}{\partial \rho} \left(\frac{F(k;\rho)}{F(\bar{k},\rho)} \right) + \frac{\partial \bar{k}}{\partial \varepsilon} \frac{F(k;\rho)}{F(\bar{k},\rho)} \frac{\partial}{\partial \rho} \left(\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)} \right) \right] dk.$$

Proof. See Appendix A.4.5. □

To understand the intuition for Lemma 4, recall that $H(\rho,\varepsilon)$ denotes the marginal social benefit of an increase in risk for the unconstrained social planner. $H(\rho,\varepsilon)$ is decreasing in ρ if the marginal benefit of keeping inefficient dealers out of the market decreases with ρ. In our model such benefit stems from a trade-off between 1) the intensive and extensive margins for dealers' profits, and 2) the matching probability between a given dealer k and buyers and sellers. The intensive margin is defined as the per dealer profits ($\pi(k)$), and it is larger for dealers indexed by low k as they are relatively more efficient. The extensive margin is defined as the measure of dealers making positive profits, that is equivalent to the dealers operating in the market. The magnitude of these components depends on their interaction with the matching probability, in particular it crucially depends on which values of k the distribution $F(k,\rho)$ places more mass, as the fewer dealers operate in the market the higher the probability for a given dealer k that he will be matched with a customer (buyer and seller).

Consider dealers with a high k, who are relatively inefficient and make little profits. Restricting their access to the market has two effects on aggregate surplus: a) a negative direct one, as after all they would make positive profits if they could access the market, and b) a positive indirect one, as more efficient dealers can now trade more often. The first direct effect dominates when there are many inefficient dealers, that is to say i) when the elasticity of the distribution $F(k,\rho)$ at the marginal active dealer $\bar{k}$, (i.e. $e_{f(\bar{k}),\varepsilon}$) is large, ii) when the profits of dealers who are more efficient than $\bar{k}$ are not very sensitive to an increase in risk (i.e. $e_{\pi',\varepsilon}$ is small). In this case the extensive margin dominates the intensive margin. Higher values of ρ imply that the distribution $F(k,\rho)$ places more mass on dealers with a low k. Hence, as ρ increases the measure of efficient dealers who are active in the market is larger, and, as a consequence, the additional benefit of keeping inefficient dealers out of the market by marginally increasing risk, is lower. Then the second indirect effect dominates and $H(\rho,\varepsilon)$ is decreasing in ρ. The additional sufficient condition in Lemma 4 about F_1 conditionally stochastic dominating F_2 guarantees that the distribution F_2 is sufficiently above F_1 for an increase in ρ to move sufficiently more mass to efficient dealers.

We can now discuss our results about the constrained and unconstrained optimal levels of risk.

Proposition 4. *Suppose i) the assumptions in Proposition 2 are satisfied , ii) $H(\rho, \varepsilon)$ is decreasing in ρ, and iii) $\int_0^{\bar{k}} \pi(k) [f_2(k) - f_1(k)] \, dk > 3C(\rho^*)$ then $\hat{\varepsilon} \geq \varepsilon^*$ for all $\lambda \in (0, 1)$.*

Proof. See Appendix A.4.7. □

In our example with mass-at-zero distribution, condition *i)* in Proposition 4 is always satisfied when condition (16) in Lemma 2 is satisfied, as we show in the proof of Lemma 5 in Appendix A.4.2. Hence, combining this with *ii)*, which we derive in Appendix A.4.6 for the case of mass-at-zero distribution, yields

$$\frac{f(\bar{k})}{F_1(\bar{k})}\bar{k}\left[1 - \frac{f_1(0)}{F_2(\bar{k})}\right] < \frac{2\lambda_\varepsilon^2}{(1 - \lambda + \lambda_\varepsilon^2)} < \frac{f(\bar{k})}{F_2(\bar{k})}\bar{k}$$

Finally, with a mass-at-zero distribution condition *iii)* in Proposition 4 is simply

$$\int_0^{\bar{k}} k \frac{f_2(k)}{F_2(\bar{k})} dk > \frac{2(1 - \lambda)}{3\lambda_\varepsilon}$$

Proposition 4 shows that a planner who is constrained by dealers' innovation decision, $\bar{\rho}$, chooses a higher level of risk than that chosen by an unconstrained planner, who can jointly choose the level of risk and innovation in the market. This result is intuitive in light of our findings in Proposition 3, where we show that the equilibrium choice of innovation by dealers is inefficiently low: $\bar{\rho} < \rho^*$. When constrained by such $\bar{\rho}$, the planner chooses risk so to indirectly affect dealers' choice of innovation and bring it closer to the efficient one, ρ^*. To see this, recall that $\bar{\rho}'(\varepsilon) \geq 0$, as shown in Proposition 2. Then, by increasing risk relative to its unconstrained optimal level ε^*, the planner implements an equilibrium with a higher level of innovation, resulting in more efficient dealers being active in the market. This result can serve as a guideline for policy makers implementing central clearing across a variety of markets.

6 Average Bid-Ask Spread

In this section, we characterize the average bid-ask spread as a function of the risk ε, relegating the details and proofs to Appendix A.5. Let $\sigma(k, \varepsilon)$ denote the bid-ask spread posted by dealer k,

$$\sigma(k, \varepsilon) = (1 + \lambda_\varepsilon)a(k) - \lambda_\varepsilon$$

and $\sigma(\varepsilon)$ the average bid ask spread:

$$\sigma(\varepsilon)=\int_0^{\bar{k}}\frac{\sigma(k,\varepsilon)}{F(\bar{k};\bar{\rho})}dF(\bar{k},\bar{\rho})$$

We can then characterize how $\sigma(\varepsilon)$ changes with risk as follows:

$$\begin{aligned}\sigma'(\varepsilon) &= \int_0^{\bar{k}}\frac{\partial\sigma(k,\varepsilon)}{\partial\varepsilon}\frac{f\left(k;\rho\right)}{F(\bar{k};\rho)}dk-\left[1-\sigma(\varepsilon)\right]\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\frac{\bar{k}}{\lambda_\varepsilon}\\ &\quad -\bar{\rho}'\left(\varepsilon\right)\int_0^{\bar{k}}\sigma(k,\varepsilon)\frac{f\left(k;\rho\right)}{F(\bar{k};\rho)}\left[\frac{(f_2(k)-f_1(k))}{f\left(k;\rho\right)}-\frac{(F_2(\bar{k})-F_1(\bar{k}))}{F(\bar{k};\rho)}\right]dk \qquad (32)\end{aligned}$$

Equation (32) identifies each different effect of risk on the bid ask spread: 1) $\int_0^{\bar{k}}\frac{\partial\sigma(k,\varepsilon)}{\partial\varepsilon}\frac{f(k;\rho)}{F(\bar{k};\rho)}dk$ is the direct effect of risk on the bid ask of dealer k, which is positive as more risk increases the inventory cost of each dealer; 2) $\left[1-\sigma(\varepsilon)\right]\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\frac{\bar{k}}{\lambda_\varepsilon}$ is the indirect effect of risk on the average bid ask spread via entry: risk reduces the measure of active dealers in the market, which channels more transactions through relatively more efficient dealers due to random matching. This effect reduces average bid ask spread because more efficient dealers charge lower spreads and more transactions are taking place with these dealers. This effect is large if a) the average bid-ask spread is narrow to start with (i.e. $\sigma(\varepsilon)$ is relatively small) and b)the distribution function at $\bar{k}$ is very elastic with respect to risk, because an increase in risk effectively shifts transactions toward more efficient dealers; 3) $\bar{\rho}'\left(\varepsilon\right)\int_0^{\bar{k}}\sigma(k,\varepsilon)\frac{f(k;\rho)}{F(\bar{k};\rho)}\left[\frac{(f_2(k)-f_1(k))}{f(k;\rho)}-\frac{(F_2(\bar{k})-F_1(\bar{k}))}{F(\bar{k};\rho)}\right]dk$ is the effect of risk on the equilibrium distribution of dealers through dealers' own choice of ρ: an increase in risk changes dealers' incentives to innovate in higher values of ρ, which, in turn, implies a change in the fraction of efficient dealers that are active. When $\bar{\rho}'\left(\varepsilon\right)\geq 0$ the fraction of efficient dealers that are active increases with risk, and the magnitude of such increase depends on the relative additional mass that the distribution F_2 places on each active dealer $k\leq\bar{k}$ with respect to F_1. Because more efficient dealers charge lower bid ask spreads, then the effect of risk on the equilibrium distribution of dealers results in lower bid ask spreads. We prove these results in the following Proposition.

Proposition 5. *The direct effect of risk on the bid ask spread is always positive for all $\lambda\in(0,1)$ and $\varepsilon\in(0,\lambda)$:*

$$\int_0^{\bar{k}}\frac{\partial\sigma(k,\varepsilon)}{\partial\varepsilon}\frac{f\left(k;\rho\right)}{F(\bar{k};\rho)}dk \quad > \quad 0,$$

the indirect effect of risk on the average bid ask spread via dealers' entry is always negative,

$$-\left[1-\sigma(\varepsilon)\right]\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\frac{\bar{k}}{\lambda_{\varepsilon}} \quad < \quad 0$$

and if F_1 conditionally stochastically dominates (CSD) F_2 and if the assumptions in Proposition 2 are satisfied, then the indirect effect of risk on the average bid ask spread via the equilibrium distribution of dealers is negative:[25]

$$-\bar{\rho}'(\varepsilon)\int_0^{\bar{k}}\sigma(k,\varepsilon)\frac{f(k;\rho)}{F(\bar{k};\rho)}\left[\frac{(f_2(k)-f_1(k))}{f(k;\rho)}-\frac{(F_2(\bar{k})-F_1(\bar{k}))}{F(\bar{k};\rho)}\right]dk \quad < \quad 0$$

Proof. See Appendix A.5. □

In the case of the mass-at-zero distribution, we derive in Appendix A.5 sufficient conditions for the average bid-ask spread to be increasing in risk, that is to say $\sigma'(\varepsilon) > 0$:

$$\begin{aligned}\left[\frac{(1-\lambda)(1-\lambda_{\varepsilon})+2\lambda_{\varepsilon}^2}{2(1-\lambda+\lambda_{\varepsilon}^2)}+\frac{(1+\lambda_{\varepsilon})\lambda_{\varepsilon}}{2(1-\lambda+\lambda_{\varepsilon}^2)}\int_0^{\bar{k}}k\frac{f_2(k)}{F_2(\bar{k})}dk\right]\frac{f_2(\bar{k})}{F_2(\bar{k})} \quad &> \quad 1, \quad \textit{and} \\ F_1(\bar{k};\rho) \quad &> \quad f_2(0)\,.\end{aligned}$$

Proposition 5 is particularly relevant in light of recent empirical findings by Loon and Zhong [2016] who study the effect of the different measures introduced by the Dodd-Frank Act, including central clearing, on transaction-level effective spreads. They analyze the three-implementation phases of mandatory central clearing by the CFTC, with each phase covering a different category of market participants. In phase 1 clearing became mandatory for Swap dealers and private funds active in the swaps market only. In phase 2 clearing became mandatory for commodity pools, private funds, and persons predominantly engaged in activities that are in the business of banking, or in activities that are financial in nature, and in phase 3 for all other Swap market participants. Table 10 in Loon and Zhong [2016] shows that mandatory clearing only for swap dealers and private funds active in the swaps market lead to an increase in the relative effective spread for all CDS indexes in their data set, with the increase ranging from 0.069% for index CDX NA HY, and as much as to 1.15% for iTraxx Europe Crossover. However, their measure of relative effective spread declined in phases 2 and 3.

To interpret these results in light of our model, it is useful to think of swap dealers as catering to a different investor base than the one of, say, commodity pools. Then the different phases of

[25] Conditional stochastic dominance (CSD) is defined in the usual way, as in Maskin and Riley [2000]. Also notice that CSD implies first order stochastic dominance (FOSD).

mandatory clearing apply to different segments of the swap market. For example, phase 1 applies to the segment of the market where swaps dealers are active, while phases 2 and 3 to the segments of the market where other financial market participants operate. Then, taking the characteristics of traders' supply and demand to be the same across these segmented markets, we can apply our findings about the effect of a decrease in counterparty risk on average bid-ask spreads ($\sigma'(\varepsilon)$) to the swap dealers submarket, for the evidence reported for phase 1 of mandatory clearing, and to the other submarkets for phases 2 and 3.

Proposition 5, in fact, implies that in the swap dealers segment, the indirect effects of risk on spreads, working via entry and via the equilibrium distribution of dealers, must dominate the direct effect of risk on spreads working via inventory cost. This happens if a) the swap dealers operating even before the transition to central clearing were relatively efficient, which results in low bid-ask spreads to start with (i.e. $\sigma(\varepsilon)$ is low); b) the distribution function of dealers' transaction cost is relatively elastic at $\overline{k}$; and/or if c) the distribution of dealers after innovation is significantly different from the initial distribution, resulting in innovation by dealers having a sizable effect on their equilibrium distribution. The opposite is true for commodity pools and all other swap dealers, for which the indirect effects are not very strong, given that the relative spread decrease with the introduction of central clearing.

While it is difficult to find tangible evidence that swap dealers are more efficient than other market makers, it is likely that swap dealers have a deeper customer base and larger market access than other market makers. Indeed, the list of all registered swap dealers with the CFTC shows that most of them are large banking corporations, investment banks and hedge funds, which are more likely to have a large and diverse customer base, and to operate in several markets, relative to the classes of financial institutions which were affected by phases 2 and 3 of the central clearing mandate. With access to a large and diverse customer base, swap dealers are able to better manage their inventories by lending or borrowing the assets they need, and able to find counterparty to a trade more easily or quickly.

As further evidence that swap dealers can be more efficient than other market makers, the phased-in introduction of central clearing reflects the relative inefficiency and complexity of the business model of these other market makers. In the Federal Register it is noted that "the Commission believes that certain market participants may require additional time to bring their swaps into compliance with the new regulatory requirement for mandatory clearing of a swap or class of swaps.[26] This is particularly true for market participants that may not be registered with the Commission and those market participants that may have hundreds or thousands of managed ac-

[26]Specifically, see Federal Register 2011, Vol. 76, No. 182, Tuesday, September 20, 2011, p. 58189.

counts, referred to as third-party subaccounts for the purposes of this proposal (...) Moreover, several commenters emphasized the need to have adequate time to educate their clients regarding the new regulatory requirements." The decision of the Commission to phase in swaps dealers first, therefore, suggests that swap dealers were already relatively more efficient with respect to all other market participants. As a consequence, our analysis implies that the indirect effects of competition via entry and of innovation on average spreads for those already efficient dealers dominate the direct effect of risk on their profits. Our model suggests that the opposite is true for all other market participants.

7 Numerical Example

To derive results analytically in the previous section we relied on some relatively strong assumptions about the distribution of dealers' trading cost. In this section we provide a numerical example, showing that the conditions we derived above are sufficient but not necessary. Regarding the cost of innovation, we let $\gamma(\rho) = A\rho^2/2$, which does not satisfy our assumption that $\gamma'(1) = +\infty$. Therefore, the unconstrained planner can choose $\rho = 1$ whenever A is not too large. Then for the initial distribution of trading costs for dealers, we use the following distribution function,

$$F_1(k) = \begin{cases} \frac{1}{2}k & \text{if } k < \frac{1}{2} \\ \frac{3}{2}k - \frac{1}{2} & \text{if } k \in \left[\frac{1}{2}, 1\right] \end{cases}$$

So the density function, f_1, is a piecewise linear function combining the density functions of different uniform distributions: $f_1(k)$ is the pdf of a random variable distributed as $\mathcal{U}[0,2]$ when $x < 1/2$ and it scales up the pdf of a random variable distributed as $\mathcal{U}[0,1]$ otherwise. If dealers innovate, they can draw from a distribution with the following distribution function,

$$F_2(k) = \begin{cases} 7.3k & \text{if } k < 0.05 \\ 7.05 + 0.25k & \text{if } k \in [0.05, 1) \\ 1 & \text{if } k = 1 \end{cases}$$

Notice that this definition of $F_2(k)$ implies that the density function $f_2(k)$ has a mass point at $k = 1$. Figure 3 shows $F_1(k)$ and $F_2(k)$ for $k \leq 1$. Clearly, F_1 does not FOSD F_2 since $F_2(k) \leq F_1(k)$ for high values of k.

Using $\lambda = 0.5$ and $A = 0.05$, we get that the unconstrained planner would choose to select dealers' trading only from the most efficient distribution, $\rho^* = 1$, and she would choose no risk,

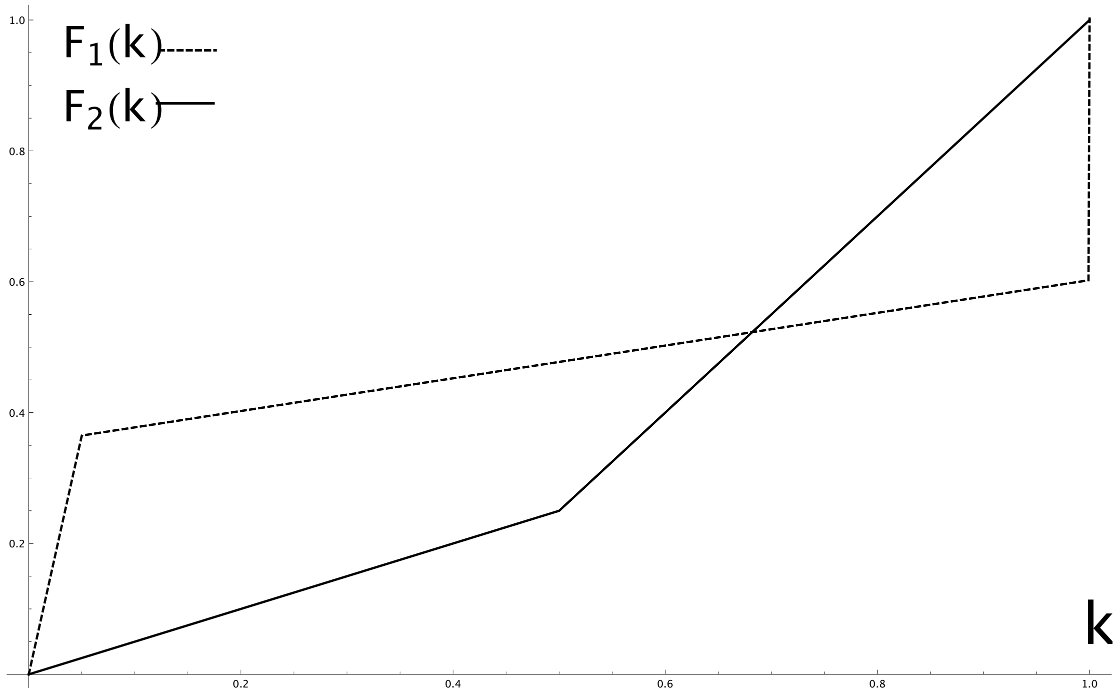

Figure 3: $F_1(k)$ (solid line), and $F_2(k)$ (dashed line).

$\varepsilon^* = 0$. To the contrary, when the planner is constrained by the dealers' innovation choice, she would choose $\hat{\varepsilon} \approx 0.34$, in which case $\bar{\rho}(\hat{\varepsilon}) \approx 0.625$. Figure 4a shows $\bar{\rho}(\varepsilon)$ is increasing with risk if and only if risk is relatively benign. The planner chooses the level of risk ε that maximizes $\bar{\rho}$. Figure 4b shows that the average bid-ask spread $\sigma(\varepsilon)$ is non-monotonic in risk: it is decreasing when there is little risk, that is for values of ε resulting in $\bar{\rho}(\varepsilon)$ steeply increasing with risk. In this region, as the measure of efficient dealers increases with risk, the resulting average bid-ask spread is lower. And the average bid-ask spread is decreasing for larger values of ε, when the increase in the measure of efficient dealers flattens and then stops.

8 Extensions

In this section we describe a version of the benchmark model of Section 4 where the decision of sellers and buyers to accept a bid/ask quote is dynamic. Formally, we assume that traders exit the market after they trade, implying that the option to search and trade in the future is relevant for current reservation prices. While we relegate the description of the model economy and the equilibrium characterization to the online Appendix, here we summarize the main results and discuss the differences with respect to the benchmark model.

The model with the option to search shares similar features with the benchmark model, and

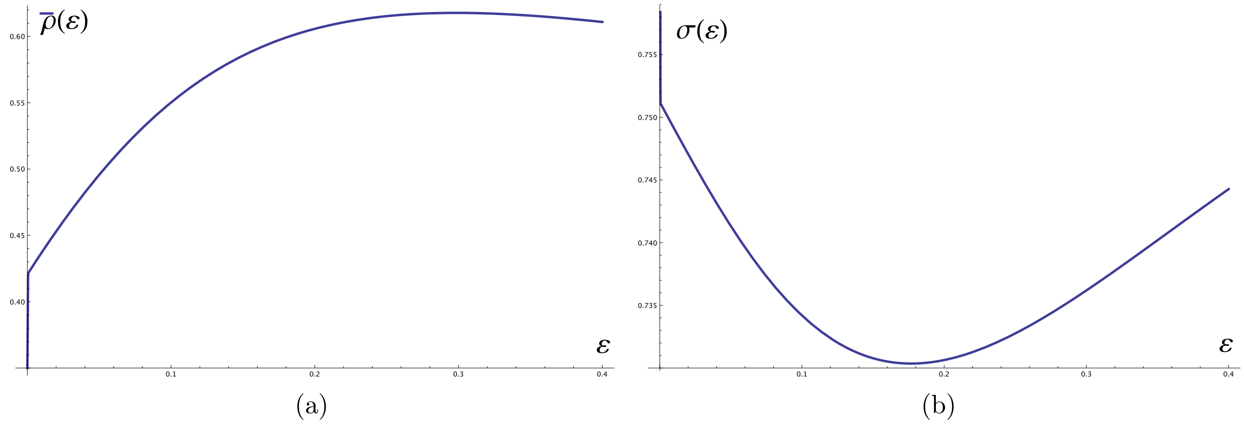

Figure 4: (a) $\bar{\rho}(\varepsilon)$; (b) $\sigma(\varepsilon)$

the qualitative results of the two models are the same. The profits of relatively efficient dealers are always increasing in risk, due to the effect of risk on dealers' entry. Therefore the main results in our benchmark model carry over to the model with a search option. It is interesting to note that the model with a search option lies in between the benchmark model of the previous section and the fully competitive outcome where the only operating dealers are the most efficient ones, as we discuss below and in Proposition 9 in the online Appendix. Propositions 6-9 in the online Appendix characterize the differences between the model with a search option and the benchmark model, which we can summarize as follows:

- The reservation price of buyers (sellers) is lower (higher) in the model with the a search option because the option to search makes traders more wary of accepting current ask and bid prices posted by dealers.

- There are fewer active dealers in the model with a search option. The change in reservation prices implies that marginal active dealer in the benchmark model now makes negative profits because the buyer and seller whom he can serve are no longer willing to pay the ask and bid prices that he can offer.

- The model with a search option has a lower ask price and a higher bid price than the benchmark model. In the model with a search option, 1) buyers' reservation value is lower and sellers' reservation value is higher; 2) dealers are competing with their future selves and with other dealers with whom end customers might match. As a consequence, bid-ask spreads are narrower.

- The order flow (volume of trades) is larger with a search option only for the more efficient dealers. Precisely, there exists $\tilde{k} \in \left(0, \bar{k}\right)$ such that the order flow for a dealer k is larger iff $k \in \left(0, \tilde{k}\right)$. Figure 5 illustrates the last result. We plot the demand for assets in the model with the search option and the benchmark model, $D\left(a\left(k\right)\right)$ and $D_B\left(a\left(k\right)\right)$ respectively, as functions of dealers' trading cost k: for relatively efficient dealers demand is larger in the model with the option to search, while the opposite is true for relatively inefficient dealers. Dealers with trading cost$k \in \left(\bar{k}, \bar{k}_B\right)$ are not active in the model with the search option so we exclude them from the comparison. However, the marginal active dealer in the model with the option to search, $\bar{k}$, makes the zero profits as the marginal active dealer in the benchmark model, $\bar{k}_B$. Therefore, dealer $\bar{k}$ makes positive profits in the benchmark model. Dealer $\tilde{k}$ is the one whose profits are the same in the two economies. Interestingly, while bid-ask spreads are narrower, the order flow can be larger or smaller than in the benchmark model because the reservation values of buyers and sellers also change with the option to search. This is intuitive in light of a comparison between the equilibrium of our model and that in a perfectly competitive market where buyers and sellers trade with dealers. An economy with a perfectly competitive market between end customers and dealers is one where only $k = 0$ dealers are active and the order flow is a point mass at that dealer. Therefore, the model with the option to search lays in between our benchmark economy and a competitive one, as the volume of transaction processed via relatively efficient dealers is larger than in our benchmark but lower than in the competitive case. In fact, when buyers and sellers exit the market after trading, they are more reluctant to accept a given ask or bid quote by a dealer, for a given distribution of ask and bid prices in the market. This implies that dealers are competing with their future selves and with other dealers with whom end customers might match at future dates. This effect makes the model with the option to search closer to the competitive one than our benchmark is.

- The effect of risk on the profits of efficient dealers is larger in the model with a search option than in the benchmark model, as the following Proposition shows. Precisely, we show in the Appendix that for all $\lambda \in (0, 1)$:

$$\frac{\partial \pi(a, b, k)}{\partial \varepsilon} |_{\varepsilon=0,k=0} \quad > \quad \frac{\partial \pi_B(a, b, k)}{\partial \varepsilon} |_{\varepsilon=0,k=0} .$$

This result follows form the effect of competition from future dates on current active dealers in the economy with the search option. In fact, as in the benchmark model, a marginal increase in risk results in limiting entry, but the most efficient dealers gain marginally more than in the benchmark model because competition is fiercer in the economy with the search option. Thus,

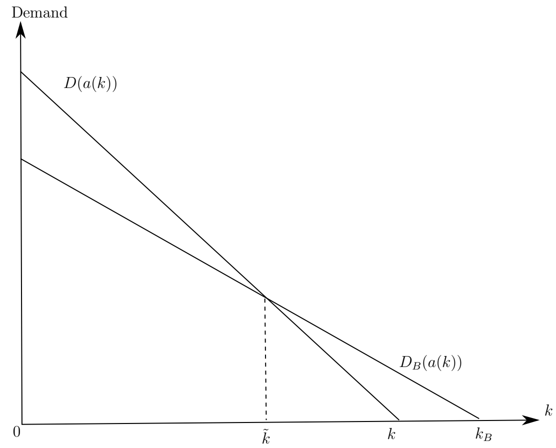

Figure 5: Order flows.

limiting competition by limiting entry has a stronger effect in that economy. Hence, our results from the benchmark model are qualitatively the same, and quantitatively stronger in the economy with the search option.

9 Conclusion

In this paper we study the effect of mandatory central clearing on the structure of OTC markets for financial assets in normal times. Despite the central clearing mandate was triggered in response to the 2008 financial crisis, its effects on the market away from crisis times are not well understood. This is in contrast with the existing literature which analyzes different channels by which central clearing relieves times of stress in financial markets. Our paper fills this gap: we focus on financial markets where transactions are intermediated by dealers, or market makers, and study the effect of introducing CCPs on measures of liquidity, such as bid-ask spreads, entry and exit into market making, and welfare.

We find that the overall impact of central clearing on financial markets is composed of a direct effect and an indirect, general equilibrium, effect. The former is relatively straightforward: by reducing counterparty risk for market makers, central clearing results in an increase in their profits. The latter is novel and key for our results: the general equilibrium channel works through the entry of less efficient dealers into market making activities. By reducing counterparty risk, central clearing allows less efficient dealers to make positive profits and become active. This reduces the market share of more efficient dealers, whose profits are then negatively affected by the introduction of a CCP. When this indirect effect on the profits of the most efficient dealers is sufficiently large, the introduction of a CCP causes a reduction in ex-ante investment by all dealers into becoming more efficient at making markets. In turn, the negative effect on dealers' incentive to innovate can result in aggregate, economy wide, welfare losses.

A natural extension is to consider how the welfare loss would be affected by considering systemic risk. Since dealers do not internalize the effects of their decisions on aggregate outcomes, we would expect that our findings on dealers' lack of incentives to innovate, and the resulting inefficiency, would still hold. However, adding systemic risk would affect the social desirability of CCPs. Provided that CCPs are resilient and appropriately risk managed, central clearing would likely become more desirable from a social point of view.[27] Then, if accounting for the desirability of such insurance, our paper shows that CCPs are desirable whenever the resulting increase in the resiliency of the financial system compensate for the loss of efficiency due to dealers' under-investment in innovation.

Finally, a version of the model we developed may speak to the effects of other post-crisis regulations on market outcomes. For example, some practitioners and academics have argued that the Volcker Rule could significantly reduce turnover and broker-dealers' inventories in corporate bonds markets, together with their willingness to provide liquidity.[28] The empirical evidence in support of these arguments is, however, mixed, perhaps due to markets simultaneously moving towards more electronic trading and to several rounds of quantitative easing. So one could enrich our model to disentangle the impact of regulations from the counteracting effects of better trading technology and QE.[29]

[27] It is worthwhile to notice that the ability of the CCPs to withstand systemic events relies in part on regulators themselves. In fact CCPs can only eliminate idiosyncratic risk, via loss mutualization, but they shift correlated risks from the bilateral counterparties to the CCP itself. For them to be resilient, regulators require ongoing supervision and tail-events testing (see BIS [2012]). For continuity in the provision of clearing services regulators emphasize the need to develop robust recovery and orderly resolution for CCPs (see BIS [2015], pg 71).

[28] See Duffie [2012], OICV-IOSCO [2019], and Fender and Lewrick [2015]

[29] For empirical evidence on the effects of post-crisis regulation on corporate bonds markets see Bao et al. [2018], Di Maggio et al. [2017], Goldstein and Hotchkiss [2019] and Boyarchenko et al. [2018]. For the impact of rising use of electronic and automated trading in fixed income markets see Bech et al. [2016].

References

Viral Acharya and Alberto Bisin. Counterparty risk externality: centralized versus over-the-counter markets. *Journal of Economic Theory*, 149:153–182, 2014. 3

Yakov Amihud and Haim Mendelson. Dealership market: market-making with inventory. *Journal of Financial Economics*, 8(1):31–53, 1980. 1.1

Andrew G Atkeson, Andrea L Eisfeldt, and Pierre-Olivier Weill. Entry and exit in OTC derivatives markets. *Econometrica*, 83(6):2231–2292, 2015. 1.1

Jack Bao, Maureen O'Hara, and Xing Alex Zhou. The Volcker Rule and corporate bond market making in times of stress. *Journal of Financial Economics*, 130(1):95–113, 2018. 29

Morten L Bech, Annamaria Illes, Ulf Lewrick, and Andreas Schrimpf. Hanging up the phone-electronic trading in fixed income markets and its implications. *BIS Quarterly Review March*, 2016. 29

Bruno Biais, Florian Heider, and Marie Hoerova. Risk-sharing or risk-taking? Counterparty risk, incentives, and margins. *The Journal of Finance*, 71(4):1669–1698, 2016. 3

BIS. International banking and financial market developments. Quarterly review, Bank for International Settlements, September 2009. 8

BIS. Principles for financial market infrastructures. Technical report, Committee on Payment and Settlement Systems, Bank for International Settlements, April 2012. 27

BIS. International banking and financial market developments. Technical report, Bank for International Settlements, Quarterly Review, December 2015. 27

BIS. A glossary of terms used in payments and settlement systems. Technical report, Bank for International Settlements, Committee on Payment and Settlement Systems, September 2016. 1

Nina Boyarchenko, Thomas M Eisenbach, Pooja Gupta, Or Shachar, and Peter Van Tassel. Bank-intermediated arbitrage. *Federal Reserve Bank of New York*, Staff Report No. 858, 2018. 29

Francesca Carapella and David C Mills. Information insensitive securities: the benefits of central counterparties. *Federal Reserve Board of Governors, Washington*, 2012. 3

Carole Comerton-Forde, Terrence Hendershott, Charles M Jones, Pamela C Moulton, and Mark S Seasholes. Time variation in liquidity: the role of market-maker inventories and revenues. *The Journal of Finance*, 65(1):295–331, 2010. 1.1

Jürgen Dennert. Price competition between market makers. *The Review of Economic Studies*, 60 (3):735–751, 1993. 1.1

Marco Di Maggio, Amir Kermani, and Zhaogang Song. The value of trading relations in turbulent times. *Journal of Financial Economics*, 124(2):266–284, 2017. 29

William C. Dudley. Reforming the OTC derivatives market. In *Remarks at the Harvard Law School's Symposium on Building the Financial System of the 21st Century, March 22nd, Armonk, New York*. http://www.newyorkfed.org/newsevents/speeches/2012/dud120322.html, 2012. 2

Darrell Duffie. Market making under the proposed volcker rule. *Rock Center for Corporate Governance at Stanford University Working Paper*, (106), 2012. 28

Darrell Duffie and Haoxiang Zhu. Does a central clearing counterparty reduce counterparty risk? *The Review of Asset Pricing Studies*, 1(1):74–95, 2011. 3

Darrell Duffie, Nicolae Gârleanu, and Lasse Heje Pedersen. Over-the-Counter markets. *Econometrica*, 73(6):1815–1847, 2005. 1, 1.1

Ingo Fender and Ulf Lewrick. Shifting tides - market liquidity and market-making in fixed income instruments. Technical report, Bank for International Settlements, BIS Quarterly Review, March 2015. 28

FSB Financial Stability Board. Review of otc derivatives market reforms. effectiveness and broader effects of the reforms. Technical report, Financial Stability Board, June 2017. 1, 5

Kenneth Garbade, Frank M Keane, Lorie Logan, Amanda Stokes Kirby, and Jennifer Wolgemuth. The introduction of the TMPG fails charge for US Treasury securities. *Economic Policy Review*, 16(2), 2010. 15

Rodney Garratt and Peter Zimmerman. Centralized netting in financial networks. *Journal of Banking & Finance*, 2017. 3

Michael A Goldstein and Edith S Hotchkiss. Providing liquidity in an illiquid market: Dealer behavior in US corporate bonds. *Journal of Financial Economics, forthcoming*, 2019. 29

Thorsten V Koeppl. The limits of central counterparty clearing: Collusive moral hazard and market liquidity. Technical report, 2013. 3

Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985. 1.1

Ricardo Lagos, Guillaume Rocheteau, and Pierre-Olivier Weill. Crises and liquidity in over-the-counter markets. *Journal of Economic Theory*, 146(6):2169–2205, 2011. 6

Yee Cheng Loon and Zhaodong (Ken) Zhong. Does Dodd-Frank affect OTC transaction costs and liquidity? Evidence from real-time CDS trade reports. *Journal of Financial Economics*, 119: 645–672, 2016. 1.1, 9, 6

David Marshall. Liquidity, settlement risk, and systemic stability, September, 8 2017. URL https://www.chicagofed.org/publications/speeches/2017/9-08-liquidity-settlement-risks-and-systemic-stability-marshall. A speech delivered on September 8, 2017, before the World Federation of Exchanges Annual Meeting in Bangkok, Thailand. 14

Eric Maskin and John Riley. Asymmetric auctions. *The Review of Economic Studies*, 67(3): 413–438, 2000. 25

Cyril Monnet. Let's make it clear: how central counterparties save(d) the day. *Business Review*, (First Quarter):1–10, 2010. 1

Maureen O'hara. *Market microstructure theory*, volume 108. Blackwell Cambridge, MA, 1995. 1.1

OICV-IOSCO. Liquidity in corporate bond markets under stressed conditions. Technical Report Final Report, The Board of the International Organization of Securities Commissions, June 2019. 28

Andreas C. Rapp. Middlemen matter: corporate bond market liquidity and dealer inventory funding. 2016. URL https://ssrn.com/abstract=2867531. 1.1

John Rust and George Hall. Middlemen versus market makers: A theory of competitive exchange. *The Journal of Political Economy*, 111:2:353–403, 2003. 1.1, 2, 2.1, 3, 17

Tano Santos and Jose A Scheinkman. Competition among exchanges. *The Quarterly Journal of Economics*, 116(3):1027–1061, 2001. 1.1

D. Spulber. Market making by price-setting firms. *Review of Economic Studies*, 63:559–580, 1996. 1.1, 2, 2.1, 3, 17, A.6

Guilaume Vuillemey. The value of central clearing. *Journal of Finance, forthcoming*, 2019. 4, 7, 8

Pierre-Olivier Weill. Leaning against the wind. *The Review of Economic Studies*, 74(4):1329–1354, 2007. 1.1

Pierre-Olivier Weill. Liquidity provision in capacity-constrained markets. *Macroeconomic Dynamics*, 15(S1):119–144, 2011. 1.1

Appendix

A.1 Useful reference equations

From the definition of dealers' profits, equation (14), we have:

$$
\begin{aligned}
\pi(k) &= \frac{(1-\lambda-\lambda_\varepsilon k)^2}{4\left(1-\lambda+\lambda_\varepsilon^2\right)} \\
\frac{\partial \pi(k)}{\partial k} &= -\lambda_\varepsilon \frac{(1-\lambda-\lambda_\varepsilon k)}{2(1-\lambda+\lambda_\varepsilon^2)} < 0 \\
&= -\pi(k)\frac{2\lambda_\varepsilon}{(1-\lambda-\lambda_\varepsilon k)} < 0 \\
&= \frac{\partial \pi(k)}{\partial \varepsilon}\frac{\lambda_\varepsilon\left(1-\lambda+\lambda_\varepsilon^2\right)}{(1-\lambda)(k+\lambda_\varepsilon)} \\
\frac{\partial \pi(k)}{\partial \varepsilon} &= -\frac{(1-\lambda)(k+\lambda_\varepsilon)(1-\lambda-k\lambda_\varepsilon)}{2\left(1-\lambda+\lambda_\varepsilon^2\right)^2} < 0 \\
&= -2\pi(k)\frac{(1-\lambda)(k+\lambda_\varepsilon)}{(1-\lambda-\lambda_\varepsilon k)\left(1-\lambda+\lambda_\varepsilon^2\right)} \\
\frac{\left(1-\lambda+\lambda_\varepsilon^2\right)}{(1-\lambda)(k+\lambda_\varepsilon)}\frac{\partial \pi(k)}{\partial \varepsilon} &= \frac{-2\pi(k)}{(1-\lambda-\lambda_\varepsilon k)} \\
\frac{\partial \pi(k)}{\partial \varepsilon}|_{k=0} &= -\frac{(1-\lambda)^2\lambda_\varepsilon}{2\left(1-\lambda+\lambda_\varepsilon^2\right)^2} < 0 \\
&= -2\pi(k)|_{k=0}\frac{\lambda_\varepsilon}{\left(1-\lambda+\lambda_\varepsilon^2\right)} \\
\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k} &= -\frac{(1-\lambda)\left[1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right]}{2\left(1-\lambda+\lambda_\varepsilon^2\right)^2} \\
&= \frac{\partial \pi(k)}{\partial k}\frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\lambda_\varepsilon\left(1-\lambda-\lambda_\varepsilon k\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} \\
&= \frac{\partial \pi(k)}{\partial \varepsilon}\frac{\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{(k+\lambda_\varepsilon)\left(1-\lambda-\lambda_\varepsilon k\right)} \\
&= -2\pi(k)\frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\left(1-\lambda-\lambda_\varepsilon k\right)^2\left(1-\lambda+\lambda_\varepsilon^2\right)} \\
\left.\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}\right|_{k=0} &= -\frac{(1-\lambda)\left(1-\lambda-\lambda_\varepsilon^2\right)}{2\left(1-\lambda+\lambda_\varepsilon^2\right)^2} \\
\left.\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}\right|_{k=\bar{k}} &= -\frac{(1-\lambda)\left(1-\lambda-2\bar{k}\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{2\left(1-\lambda+\lambda_\varepsilon^2\right)^2} = \frac{(1-\lambda)}{2\left(1-\lambda+\lambda_\varepsilon^2\right)} > 0
\end{aligned}
$$

A.2 Proof of Lemma 2 and Proposition 1

A.2.1 Proof of Lemma 2

Proof. Dealers' expected surplus (conditional on being active) is defined as

$$S_d(\varepsilon) = \int_0^{\bar{k}} \Pi(k) \frac{dF(k)}{F(\bar{k})}$$

Therefore, the dealers' expected surplus is decreasing in ε, if and only if

$$\int_0^{\bar{k}} \frac{\partial \Pi(k)}{\partial \varepsilon} \frac{dF(k)}{F(\bar{k})} - \frac{f(\bar{k})}{F(\bar{k})^2} \frac{\partial \bar{k}}{\partial \varepsilon} \int_0^{\bar{k}} \Pi(k) dF(k) < 0$$

where we use the result that $\Pi(\bar{k}) = 0$. Also, since $\bar{k} = (1-\lambda)/\lambda_\varepsilon$, we obtain $\partial \bar{k}/\partial \varepsilon = -\bar{k}/\lambda_\varepsilon$. Hence $\frac{\partial S_d(\varepsilon)}{\partial \varepsilon} < 0$ if and only if:

$$\begin{aligned}
\frac{f(\bar{k})}{F(\bar{k})} \frac{\bar{k}}{\lambda_\varepsilon} \int_0^{\bar{k}} \Pi(k) dF(k) &< -\int_0^{\bar{k}} \frac{\partial \Pi(k)}{\partial \varepsilon} dF(k) \\
\frac{f(\bar{k})}{F(\bar{k})} \frac{\bar{k}}{\lambda_\varepsilon} \int_0^{\bar{k}} \Pi(k) dF(k) &< -\int_0^{\bar{k}} \left[\frac{1}{F(\bar{k})} \frac{\partial \pi(k)}{\partial \varepsilon} + \Pi(k) \frac{f(\bar{k})}{F(\bar{k})} \frac{\bar{k}}{\lambda_\varepsilon} \right] dF(k) \\
\frac{f(\bar{k})}{F(\bar{k})} \frac{\bar{k}}{\lambda_\varepsilon} \int_0^{\bar{k}} \Pi(k) dF(k) &< -\frac{1}{2} \int_0^{\bar{k}} \frac{\partial \pi(k)}{\partial \varepsilon} \frac{dF(k)}{F(\bar{k})}
\end{aligned}$$

which is the same as (15). To derive (16) recall that, given random matching, the expected profit of dealer k is defined in equation (14). Thus, using $b = \lambda_\varepsilon(1-a)$ and:

$$\begin{aligned}
D(a) &= \frac{1}{F(\bar{k})}(1-a) \\
S(b) &= \frac{1}{F(\bar{k})}\lambda_\varepsilon(1-a)
\end{aligned}$$

we have

$$\Pi(k) = \frac{\pi(k)}{F(\bar{k})} = \frac{(1-\lambda-k\lambda_\varepsilon)^2}{F(\bar{k})4(1-\lambda+\lambda_\varepsilon^2)}$$

where $\pi(k)$ denotes the actual profit per trade of dealer k. From the definition of the surplus of a dealer k we have:

$$\frac{\partial S(k,\varepsilon)}{\partial \varepsilon}|_{k=0} = \frac{\partial \left(\frac{\pi(k)}{F(\bar{k})} \right)}{\partial \varepsilon}|_{k=0}$$

$$
\begin{aligned}
&= \frac{\frac{\partial \pi(k)}{\partial \varepsilon} F\left(\bar{k}\right) - \pi(k) f\left(\bar{k}\right) \frac{\partial \bar{k}}{\partial \varepsilon}}{F\left(\bar{k}\right)^2} \\
&= -\frac{(1-\lambda)\pi(k)}{F\left(\bar{k}\right)} \left[\frac{2(k+\lambda_\varepsilon)}{(1-\lambda-\lambda_\varepsilon k)\left(1-\lambda+\lambda_\varepsilon^2\right)} - \frac{f\left(\bar{k}\right)}{F\left(\bar{k}\right)} \frac{1}{\lambda_\varepsilon^2} \right]
\end{aligned}
$$

and this is strictly positive at $k = 0$ if and only if

$$
\frac{2\lambda_\varepsilon^3}{(1-\lambda)\left(1-\lambda+\lambda_\varepsilon^2\right)} < \frac{f\left(\bar{k}\right)}{F\left(\bar{k}\right)}
$$

which is the same as

$$
\frac{2\lambda_\varepsilon^2}{\left(1-\lambda+\lambda_\varepsilon^2\right)} < \frac{f\left(\bar{k}\right)}{F\left(\bar{k}\right)} \bar{k} \tag{33}
$$

This condition sets a lower bound on the elasticity of the distribution function of k. Such lower bound is increasing in ε. Condition (33) can be rearranged as

$$
\frac{\frac{\partial \pi(k)}{\partial \varepsilon}|_{k=0}}{\pi(k)|_{k=0}} > \frac{f\left(\bar{k}\right)}{F\left(\bar{k}\right)} \frac{\partial \bar{k}}{\partial \varepsilon}
$$

which says that the elasticity of the profits of the most efficient dealer with respect to risk is larger than the elasticity of the distribution function at the marginal active dealer with respect to risk. This proves Lemma 2. □

A.2.2 Proof of Proposition 1

Proof. The result on the surplus of the most efficient dealer follows from Lemma 2. We now concentrate on total surplus. From the analysis above we have

$$
\begin{aligned}
S_d(\varepsilon) &= \int_0^{\bar{k}} \pi(k) \frac{dF(k)}{F(\bar{k})} \\
S_b(\varepsilon) &= \int_0^{\bar{k}} \frac{1}{2}(1-a(k))^2 \frac{f(k)}{F(\bar{k})} dk \\
S_s(\varepsilon) &= \lambda_\varepsilon^2 S_b(\varepsilon)
\end{aligned}
$$

Where, by the law of large numbers, $S_d(\varepsilon)$ denotes aggregate surplus of dealers before they learn their k and before they are matched with buyers and sellers.

Now, recall that the surplus of buyers has to be discounted by $1-\lambda$ because only a fraction

$1-\lambda$ of buyers can obtain the good as the other fraction fails to settle. Therefore, the total surplus of buyers and sellers is

$$(1-\lambda)S_b(\varepsilon)+S_s(\varepsilon)=(1-\lambda+\lambda_\varepsilon^2)\int_0^{\bar{k}}\frac{1}{2}(1-a(k))^2\frac{f(k)}{F(\bar{k})}dk$$

Therefore, total surplus is $S(\varepsilon)\equiv S_d(\varepsilon)+(1-\lambda)S_b(\varepsilon)+S_s(\varepsilon)$, specifically:

$$\begin{aligned}
S(\varepsilon) &= \int_0^{\bar{k}}(1-a(k))\left[(1-\lambda+\lambda_\varepsilon^2)a(k)-(\lambda_\varepsilon+k)\lambda_\varepsilon\right]\frac{dF(k)}{F(\bar{k})}+\\
&\quad +(1-\lambda+\lambda_\varepsilon^2)\int_0^{\bar{k}}(1-a(k))\left[\frac{1}{2}(1+a(k))-a(k)\right]\frac{dF(k)}{F(\bar{k})}\\
&= \int_0^{\bar{k}}(1-a(k))\left[(1-\lambda+\lambda_\varepsilon^2)\frac{1}{2}(1+a(k))-(\lambda_\varepsilon+k)\lambda_\varepsilon\right]\frac{dF(k)}{F(\bar{k})}
\end{aligned}$$

Recall that

$$a(k)=\frac{1}{2}+\frac{\lambda_\varepsilon^2+\lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)}$$

so that

$$(1-\lambda+\lambda_\varepsilon^2)(2a(k)-1)=\lambda_\varepsilon^2+\lambda_\varepsilon k$$

Hence

$$\begin{aligned}
S(\varepsilon) &= \int_0^{\bar{k}}(1-\lambda+\lambda_\varepsilon^2)(1-a(k))\frac{3}{2}(1-a(k))\frac{dF(k)}{F(\bar{k})}\\
&= \frac{3}{2}\int_0^{\bar{k}}(1-\lambda+\lambda_\varepsilon^2)(1-a(k))^2\frac{dF(k)}{F(\bar{k})} && (34)\\
&= \frac{3}{8}\int_0^{\bar{k}}\frac{(1-\lambda-\lambda_\varepsilon k)^2}{(1-\lambda+\lambda_\varepsilon^2)}\frac{dF(k)}{F(\bar{k})}\\
&= \frac{3}{2}\int_0^{\bar{k}}\pi(k)\frac{dF(k)}{F(\bar{k})} && (35)
\end{aligned}$$

From $S(\varepsilon)$ and since $\pi(\bar{k})=0$, we get

$$\frac{2}{3}\frac{\partial S(\varepsilon)}{\partial\varepsilon} = -\frac{\partial\bar{k}}{\partial\varepsilon}\frac{f(\bar{k})}{F(\bar{k})^2}\int_0^{\bar{k}}\pi(k)dF(k)+\int_0^{\bar{k}}\frac{\partial\pi(k)}{\partial\varepsilon}\frac{dF(k)}{F(\bar{k})}$$

where $\frac{\partial\bar{k}}{\partial\varepsilon}>0$ and $\frac{\partial\pi(k)}{\partial\varepsilon}<0$. The first term reflects the effect of risk on the composition of the market for dealers: This effect is positive since the market is composed of more efficient dealers.

However, this is balanced by the overall decline in profits for all active dealers, as shown by the second term. Therefore, the total expected surplus is decreasing whenever the effect on profit is larger than the composition effect. We can use the expressions for $\frac{\partial \bar{k}}{\partial \varepsilon}$ and $\frac{\partial \pi(k)}{\partial \varepsilon}$ to obtain

$$\frac{8}{3}\frac{\partial S(\varepsilon)}{\partial \varepsilon} = \int_0^{\bar{k}} -\frac{(1-\lambda-\lambda_\varepsilon k)(1-\lambda)}{(1-\lambda+\lambda_\varepsilon^2)}\left\{-\frac{(1-\lambda-\lambda_\varepsilon k)}{\lambda_\varepsilon^2}\frac{f(\bar{k})}{F(\bar{k})}+\frac{2(\lambda_\varepsilon+k)}{(1-\lambda+\lambda_\varepsilon^2)}\right\}\frac{dF(k)}{F\left(\bar{k}\right)}$$

notice now that if $f(\bar{k})=0$ then $\frac{\partial S(\varepsilon)}{\partial \varepsilon}<0$ and it is increasing in $f(\bar{k})$. So there is a threshold $\bar{f}$ such that $\frac{\partial S(\varepsilon)}{\partial \varepsilon}\leq 0$ for all $f(\bar{k})\leq \bar{f}$ and $\frac{\partial S(\varepsilon)}{\partial \varepsilon}>0$ otherwise.

This proves Proposition 1. □

A.3 Surpluses

A.3.1 Surplus of buyers

The definition of buyers' surplus is

$$\begin{aligned} S_b(\varepsilon) &= \int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{v}(v-a)dG_a(a)dv+\int_{\bar{a}}^{1}\int_{\underline{a}}^{\bar{a}}(v-a)dG_a(a)dv \\ & \int_{\underline{a}=a(0)}^{a(\bar{k})}\int_{\underline{a}=a(0)}^{v}(v-a)dG_a(a)dv+\int_{a(\bar{k})}^{1}\int_{\underline{a}=a(0)}^{a(\bar{k})}(v-a)dG_a(a)dv \end{aligned}$$

where the first term denotes all buy orders for buyers with valuation below the highest ask price, that is with $v\leq\bar{a}$, and where the second term denotes all buy orders for buyers with valuation above the highest ask prices, that is with $v>\bar{a}$. We now do a change of variable, switching from $a(k)$ to k, and use the equation for the ask price, (12), where

$$a(k)=\frac{1}{2}+\frac{\lambda_\varepsilon^2+\lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)}=x+yk$$

Notice that $a\left(\overline{k}\right)=x+y\overline{k}=1$.

For the change of variable we have $G_a(a)$ denotes the probability that the observed ask price does not exceed a:

$$\begin{aligned} G_a(a) &= \mathbb{P}(\text{observed ask price}\leq a) \\ &= \mathbb{P}\left(x+yk\leq a|k\leq\overline{k}\right) \end{aligned}$$

where the probability that the observed ask price takes into account that a given ask price must be posted by an active dealer, that is to say $k \leq \overline{k}$. Then:

$$\begin{aligned}
G_a(a) &= \mathbb{P}\left(k \leq \frac{a-x}{y} | k \leq \overline{k}\right) \\
&= \frac{\mathbb{P}\left(k \leq \min\left(\frac{a-x}{y}, \overline{k}\right)\right)}{\mathbb{P}\left(k \leq \overline{k}\right)} \\
&= \frac{F\left(\min\left(\frac{a-x}{y}, \overline{k}\right)\right)}{F\left(\overline{k}\right)} = \begin{cases} \frac{F\left(\frac{a-x}{y}\right)}{F\left(\frac{1-x}{y}\right)} & \text{if } \frac{a-x}{y} < \overline{k} \\ 1 & \text{if } \frac{a-x}{y} \geq \overline{k} \end{cases}
\end{aligned}$$

So $g_a(a) = \frac{1}{yF\left(\frac{1-x}{y}\right)} f\left(\frac{a-x}{y}\right) = \frac{1}{yF(\bar{k})} f(k)$. Then $dG_a(a) = g_a(a)\, da$, and $da = ydk$, so when doing the change of variable we have

$$dG_a(a) = \frac{1}{yF(\bar{k})} f(k)\, ydk = \frac{f(k)}{F(\bar{k})} dk$$

Then the two terms in buyers' surplus can be rearranged as follows:

$$\begin{aligned}
\int_{a(0)}^{a(\bar{k})} \int_{a(0)}^{v} (v-a) g_a(a) da dv &= \int_{x}^{x+y\bar{k}} \int_{0}^{\frac{v-x}{y}} (v - a(k)) \frac{1}{y} \frac{f(k)}{F(\bar{k})} (ydk) dv \\
&= \frac{1}{F(\bar{k})} \int_{x}^{x+y\bar{k}} \int_{0}^{\frac{v-x}{y}} [v - x - yk]\, dF(k) dv
\end{aligned}$$

and

$$\int_{a(\bar{k})}^{1} \int_{\underline{a}=a(0)}^{a(\bar{k})} (v-a) dG_a(a) dv = \frac{1}{F(\bar{k})} \int_{x+y\bar{k}}^{1} \int_{0}^{\bar{k}} [v - x - yk]\, dF(k) dv$$

Then the buyers' surplus is $S_b(\varepsilon)$ that satisfies,

$$S_b(\varepsilon) = \int_{x}^{x+y\bar{k}} \int_{0}^{\frac{v-x}{y}} [v - x - yk] \frac{dF(k)}{F(\bar{k})} dv + \int_{x+y\bar{k}}^{1} \int_{0}^{\bar{k}} [v - x - yk] \frac{dF(k)}{F(\bar{k})} dv$$

where x and y are functions of ε, and using $a_k(\varepsilon) = x(\varepsilon) + y(\varepsilon)k$. Substituting out $a\left(\overline{k}\right) = x + y\overline{k} = 1$, yields:

$$S_b(\varepsilon) = \int_{x}^{1} \int_{0}^{\frac{v-x}{y}} [v - x - yk] \frac{dF(k)}{F(\bar{k})} dv = \int_{x}^{1} \int_{0}^{\frac{v-x}{y}} [v - a(k)] \frac{dF(k)}{F(\bar{k})} dv$$

Thus we obtain

$$S_b'(\varepsilon) = \int_x^1 \int_0^{\frac{v-x}{y}} -\frac{\partial a(k)}{\partial \varepsilon} \frac{dF(k)}{F(\bar{k})} dv - \frac{f(\bar{k})}{F(\bar{k})} \bar{k}'(\varepsilon) S_b(\varepsilon)$$

This derivative is indeterminate because while $\frac{\partial a(k)}{\partial \varepsilon} > 0$ plays against the buyer's surplus, the most adverse dealers – offering the worst terms of trade – drop out $\overline{k}'(\varepsilon) < 0$, which plays in favor of buyers. But we can say more on the surplus of buyers, by inverting the integrals,

$$\begin{aligned} S_b(\varepsilon) &= \int_x^1 \int_0^{\frac{v-x}{y}} [v - a(k)] \frac{f(k)}{F(\bar{k})} dk dv \\ &= \int_0^{\bar{k}} \int_{a(k)}^1 [v - a(k)] \, dv \frac{f(k)}{F(\bar{k})} dk \\ &= \int_0^{\bar{k}} \frac{1}{2} (1 - a(k))^2 \frac{f(k)}{F(\bar{k})} dk \end{aligned}$$

A.3.2 Surplus of sellers

The definition of sellers' surplus is

$$\begin{aligned} S_s(\varepsilon) &= \int_{\underline{b}}^{\bar{b}} \int_v^{\bar{b}} [b - v] \, dG_b(b) dv + \int_0^{\underline{b}} \int_{\underline{b}}^{\bar{b}} [b - v] \, dG_b(b) dv \\ &= \int_{b(\bar{k})}^{b(0)} \int_v^{b(0)} [b - v] \, dG_b(b) dv + \int_0^{b(\bar{k})} \int_{b(\bar{k})}^{b(0)} [b - v] \, dG_b(b) dv \end{aligned}$$

where the first term denotes sell orders for sellers with valuation above the lowest bid price, that is with $v \geq \underline{b}$, and where the second term denotes all sell orders for sellers with valuation below the lowset bid prices, that is with $v < \underline{b}$. Using the equation for the bid price, (13), we write

$$b(k) = \tilde{x} - \tilde{y}k$$

where

$$\begin{aligned} \tilde{x} &= \frac{\lambda_\varepsilon (1 - \lambda)}{1 - \lambda + \lambda_\varepsilon^2} \\ \tilde{y} &= \frac{\lambda_\varepsilon^2}{1 - \lambda + \lambda_\varepsilon^2} \end{aligned}$$

and where $b(\overline{k}) = \tilde{x} - \tilde{y}\overline{k} = 0$. Thus

$$\begin{aligned} G_b(b) &= \mathbb{P}(\text{observed bid price} \leq b) \\ &= \mathbb{P}\left(\tilde{x} - \tilde{y}k \leq b | k \leq \overline{k}\right) \end{aligned}$$

where the probability of observing a certain bid price takes into account that it must be posted by an active dealer, that is to say $k \leq \overline{k}$. Then:

$$\begin{aligned} G_b(b) &= \mathbb{P}\left(k \geq \frac{\tilde{x} - b}{\tilde{y}} | k \leq \overline{k}\right) \\ &= \frac{\mathbb{P}\left(\overline{k} \geq k \geq \frac{\tilde{x}-b}{\tilde{y}}\right)}{\mathbb{P}\left(k \leq \overline{k}\right)} \\ &= \frac{\int_{\frac{\tilde{x}-b}{\tilde{y}}}^{\overline{k}} dF(k)}{\mathbb{P}\left(k \leq \overline{k}\right)} \\ &= \frac{F\left(\overline{k}\right) - F\left(\frac{\tilde{x}-b}{\tilde{y}}\right)}{F\left(\overline{k}\right)} \end{aligned}$$

and

$$g_b(b) = \frac{1}{\tilde{y}F\left(\frac{\tilde{x}-b(\bar{k})}{\tilde{y}}\right)} f\left(\frac{\tilde{x} - b}{\tilde{y}}\right) = \frac{1}{\tilde{y}F\left(\bar{k}\right)} f(k)$$

and $db = -\tilde{y}dk$

$$\begin{aligned} S_s(\varepsilon) &= \int_{b(\bar{k})}^{b(0)} \int_v^{b(0)} [b - v]\, g_b(b) db dv + \int_0^{b(\bar{k})} \int_{b(\bar{k})}^{b(0)} [b - v]\, g_b(b) db dv \\ &= \int_{b(\bar{k})}^{b(0)} \int_v^{b(0)} [b - v]\, f\left(\frac{\tilde{x} - b}{\tilde{y}}\right) \frac{1}{\tilde{y}F\left(\bar{k}\right)} db dv + \int_0^{b(\bar{k})} \int_{b(\bar{k})}^{b(0)} [b - v]\, f\left(\frac{\tilde{x} - b}{\tilde{y}}\right) \frac{1}{\tilde{y}F\left(\bar{k}\right)} db dv \\ &= \int_{b(\bar{k})}^{b(0)} \int_{\frac{\tilde{x}-v}{\tilde{y}}}^{0} [\tilde{x} - \tilde{y}k - v]\, (-f(k)) \frac{dk}{F(\bar{k})} dv + \int_0^{b(\bar{k})} \int_{\bar{k}}^{0} [\tilde{x} - \tilde{y}k - v]\, (-f(k)) \frac{dk}{F(\bar{k})} dv \\ &= \int_{\tilde{x}-\tilde{y}\bar{k}}^{\tilde{x}} \int_{\frac{\tilde{x}-v}{\tilde{y}}}^{0} [\tilde{x} - \tilde{y}k - v]\, (-f(k)) \frac{dk}{F(\bar{k})} dv + \int_0^{\tilde{x}-\tilde{y}\bar{k}} \int_{\bar{k}}^{0} [\tilde{x} - \tilde{y}k - v]\, (-f(k)) \frac{dk}{F(\bar{k})} dv \\ &= \int_{\tilde{x}-\tilde{y}\bar{k}}^{\tilde{x}} \int_0^{\frac{\tilde{x}-v}{\tilde{y}}} [(\tilde{x} - \tilde{y}k) - v]\, f(k) \frac{dk}{F(\bar{k})} dv + \int_0^{\tilde{x}-\tilde{y}\bar{k}} \int_0^{\bar{k}} [(\tilde{x} - \tilde{y}k) - v]\, f(k) \frac{dk}{F(\bar{k})} dv \end{aligned}$$

Substituting out $b(\overline{k}) = \tilde{x} - \tilde{y}\overline{k} = 0$ yields

$$\begin{aligned} S_s(\varepsilon) &= \int_0^{\tilde{x}} \int_0^{\frac{\tilde{x}-v}{\tilde{y}}} [(\tilde{x} - \tilde{y}k) - v] f(k) \frac{dk}{F(\bar{k})} dv \\ &= \int_0^{\tilde{x}} \int_0^{\frac{\tilde{x}-v}{\tilde{y}}} [b(k) - v] f(k) \frac{dk}{F(\bar{k})} dv \end{aligned}$$

Now, it will be useful to express this surplus as a function of $a(k)$, so we use

$$\begin{aligned} b(k) &= \lambda_\varepsilon(1 - a(k)) \\ \tilde{x} - \tilde{y}k &= \lambda_\varepsilon(1 - x - yk) \end{aligned}$$

to find

$$\tilde{x} = \lambda_\varepsilon(1 - x)$$

and

$$\tilde{y} = \lambda_\varepsilon y$$

Therefore,

$$\begin{aligned} S_s(\varepsilon) &= \int_0^{\tilde{x}} \int_0^{\frac{\tilde{x}-v}{\tilde{y}}} [b(k) - v] f(k) \frac{dk}{F(\bar{k})} dv \\ &= \int_0^{\lambda_\varepsilon(1-x)} \int_0^{\frac{\lambda_\varepsilon(1-x)-v}{\lambda_\varepsilon y}} [\lambda_\varepsilon(1 - a(k)) - v] f(k) \frac{dk}{F(\bar{k})} dv \end{aligned}$$

changing variables $\tilde{v} = 1 - \frac{v}{\lambda_\varepsilon}$, or

$$\begin{aligned} v &= \lambda_\varepsilon(1 - \tilde{v}) \\ dv &= -\lambda_\varepsilon d\tilde{v} \end{aligned}$$

then $\tilde{v}(0) = 1$ and $\tilde{v}(\lambda_\varepsilon(1 - x)) = x$, so that

$$\begin{aligned} S_s(\varepsilon) &= \int_0^{\lambda_\varepsilon(1-x)} \int_0^{\frac{\lambda_\varepsilon(1-x)-v}{\lambda_\varepsilon y}} [\lambda_\varepsilon(1 - a(k)) - v] f(k) \frac{dk}{F(\bar{k})} dv \\ &= \int_1^{x} \int_0^{\frac{\lambda_\varepsilon(1-x)-\lambda_\varepsilon(1-\tilde{v})}{\lambda_\varepsilon y}} [\lambda_\varepsilon(1 - a(k)) - \lambda_\varepsilon(1 - \tilde{v})] f(k) \frac{dk}{F(\bar{k})} (-\lambda_\varepsilon) d\tilde{v} \\ &= -\lambda_\varepsilon^2 \int_1^{x} \int_0^{\frac{\tilde{v}-x}{y}} [\tilde{v} - a(k)] f(k) \frac{dk}{F(\bar{k})} d\tilde{v} \end{aligned}$$

$$
\begin{aligned}
&= \lambda_\varepsilon^2 \int_x^1 \int_0^{\frac{v-x}{y}} [v - a(k)]\, f(k) \frac{dk}{F(\bar{k})} dv \\
&= \lambda_\varepsilon^2 S_b(\varepsilon)
\end{aligned}
$$

As for buyers, the change in the sellers' surplus with respect to ε is indeterminate.

A.3.3 Surplus of dealers

We define the surplus of dealers as the expected profit of dealers before they select k, that is

$$
\begin{aligned}
S_d(\varepsilon) &= \int_0^{\bar{k}} \frac{\pi(k)}{F(\bar{k})} dF(k) \\
&= \int_0^{\bar{k}} \left\{(1-\lambda)a(k) - [\lambda_\varepsilon(1-a(k)) + k]\, \lambda_\varepsilon\right\} (1 - a(k)) \frac{dF(k)}{F(\bar{k})} \\
&= \int_0^{\bar{k}} (1 - a(k)) \left[(1 - \lambda + \lambda_\varepsilon^2) a(k) - (\lambda_\varepsilon + k)\, \lambda_\varepsilon\right] \frac{dF(k)}{F(\bar{k})}
\end{aligned}
$$

A.4 Proofs of Section 5

A.4.1 Proof of Lemma 3

Proof. To find $\bar{\rho}'(\varepsilon)$, differentiate (20) with respect to ε:

$$
\left[\int_0^{\bar{k}} \frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} dF_2(k) - \int_0^{\bar{k}} \frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} dF_1(k) - \gamma'(\bar{\rho}) f(\bar{k};\bar{\rho}) \frac{\partial \bar{k}}{\partial \varepsilon}\right] = \left[\gamma''(\bar{\rho}) F(\bar{k};\bar{\rho}) + \gamma'(\bar{\rho}) \left(F_2(\bar{k}) - F_1(\bar{k})\right)\right] \bar{\rho}'(\varepsilon)
$$

using (20) to get rid of $\gamma'(\rho)$:

$$
\left[\int_0^{\bar{k}} \frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} dF_2(k) - \int_0^{\bar{k}} \frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} dF_1(k) - \left[\int_0^{\bar{k}} \pi(k) dF_2(k) - \int_0^{\bar{k}} \pi(k) dF_1(k)\right] \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\partial \bar{k}}{\partial \varepsilon}\right] = \left[\gamma''(\bar{\rho}) F(\bar{k};\bar{\rho}) + \gamma'(\bar{\rho}) \left(F_2(\bar{k}) - F_1(\bar{k})\right)\right] \bar{\rho}'(\varepsilon)
$$

$$
\left[\int_0^{\bar{k}} \left(\frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} - \pi(k) \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\partial \bar{k}}{\partial \varepsilon}\right) dF_2(k) - \int_0^{\bar{k}} \left(\frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} - \pi(k) \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\partial \bar{k}}{\partial \varepsilon}\right) dF_1(k)\right] = \left[\gamma''(\bar{\rho}) F(\bar{k};\bar{\rho}) + \gamma'(\bar{\rho}) \left(F_2(\bar{k}) - F_1(\bar{k})\right)\right] \bar{\rho}'(\varepsilon)
$$

$$
\left[\int_0^{\bar{k}} \left(\frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} + \pi(k) \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\bar{k}}{\lambda_\varepsilon}\right) dF_2(k) - \int_0^{\bar{k}} \left(\frac{\partial \pi(k,\varepsilon)}{\partial \varepsilon} + \pi(k) \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\bar{k}}{\lambda_\varepsilon}\right) dF_1(k)\right] =
$$

$$\left[\gamma''(\bar{\rho})F(\bar{k};\bar{\rho})+\gamma'(\bar{\rho})\left(F_2(\bar{k})-F_1(\bar{k})\right)\right]\bar{\rho}'(\varepsilon) \quad (36)$$

$$\int_0^{\bar{k}} \pi(k)\left(\frac{\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}}{\pi(k)}\lambda_\varepsilon+\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k}\right)[f_2(k)-f_1(k)]\,dk \;=\; \lambda_\varepsilon\left[\gamma''(\bar{\rho})F(\bar{k};\bar{\rho})+\gamma'(\bar{\rho})\left(F_2(\bar{k})-F_1(\bar{k})\right)\right]\bar{\rho}'(\varepsilon) \quad (37)$$

So, deriving S and using $\frac{\partial\bar{k}}{\partial\varepsilon}=-\frac{\bar{k}}{\lambda_\varepsilon}$ we have

$$\frac{\partial S(k,\varepsilon)}{\partial\varepsilon}=\frac{3}{2}\frac{\pi(k)}{F(\bar{k};\bar{\rho})\lambda_\varepsilon}\left(\frac{\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}}{\pi(k)}\lambda_\varepsilon+\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k}\right)$$

and can rearrange (37) as

$$F(\bar{k};\bar{\rho})\frac{2}{3}\int_0^{\bar{k}}\frac{\partial S(k,\varepsilon)}{\partial\varepsilon}[f_2(k)-f_1(k)]\,dk=\lambda_\varepsilon\left[\gamma''(\bar{\rho})F(\bar{k};\bar{\rho})+\gamma'(\bar{\rho})\left(F_2(\bar{k})-F_1(\bar{k})\right)\right]\bar{\rho}'(\varepsilon)$$

As the term in square brackets on the right hand side is always positive, it follows that

$$\rho'(\varepsilon)\geq 0 \iff \int_0^{\bar{k}}\frac{\partial S(k,\varepsilon)}{\partial\varepsilon}[f_2(k)-f_1(k)]\,dk\geq 0$$

□

A.4.2 Proof of Proposition 2

Proof. To derive more precise conditions for the result in Lemma 3, consider each term of (36) in the proof of Lemma 3 individually:

$$\int_0^{\bar{k}}\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}+\pi(k)\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dF_j(k)$$

and integrate by parts to obtain

$$\int_0^{\bar{k}}\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}+\pi(k)\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dF_j(k) \;=\; \left[\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}+\pi(k)\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)F_j\right]_0^{\bar{k}}-\int_0^{\bar{k}}F_j\frac{\partial}{\partial k}\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}+\pi(k)\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk \;=$$

$$-\left[\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}|_{k=0}+\pi(k)|_{k=0}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)F_j\left(0\right)\right]-\int_0^{\bar{k}}F_j\left(\frac{\partial^2\pi(k,\varepsilon)}{\partial\varepsilon\partial k}+\frac{\partial\pi(k,\varepsilon)}{\partial k}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk$$

Notice that, when substituted back in the left hand side of (36), we are left with[30]

$$\left[\int_0^{\bar{k}}\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}+\pi(k)\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dF_2(k)-\int_0^{\bar{k}}\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}+\pi(k)\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dF_1(k)\right] \;\geq\; 0$$

which is satisfied if and only if:

$$\left[-\int_0^{\bar{k}}F_2\left(\frac{\partial^2\pi(k,\varepsilon)}{\partial\varepsilon\partial k}+\frac{\partial\pi(k,\varepsilon)}{\partial k}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk-\left(-\int_0^{\bar{k}}F_1\left(\frac{\partial^2\pi(k,\varepsilon)}{\partial\varepsilon\partial k}+\frac{\partial\pi(k,\varepsilon)}{\partial k}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk\right)\right] \;\geq\; 0$$

$$\int_0^{\bar{k}}(F_1-F_2)\left(\frac{\partial^2\pi(k,\varepsilon)}{\partial\varepsilon\partial k}+\frac{\partial\pi(k,\varepsilon)}{\partial k}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk \;\geq\; 0 \tag{38}$$

Notice that F_1 FOSD F_2 implies $(F_1-F_2)\leq 0$. Then, using the equations for the derivative of the profit functions in Appendix A.1, we can rewrite (38) as

$$-\int_0^{\bar{k}}(F_1-F_2)\,2\pi(k)\left(\frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\left(1-\lambda-\lambda_\varepsilon k\right)^2\left(1-\lambda+\lambda_\varepsilon^2\right)}+\frac{\lambda_\varepsilon}{\left(1-\lambda-\lambda_\varepsilon k\right)}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk \;\geq\; 0 \tag{39}$$

Then let

$$g_1\left(k\right) \;=\; \frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\left(1-\lambda-\lambda_\varepsilon k\right)^2\left(1-\lambda+\lambda_\varepsilon^2\right)}=\frac{\lambda_\varepsilon}{\left(1-\lambda-\lambda_\varepsilon k\right)}\frac{\frac{\partial^2\pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial\pi(k)}{\partial k}} \tag{40}$$

$$g_2\left(k\right) \;=\; \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\left(1-\lambda-\lambda_\varepsilon k\right)} \tag{41}$$

so that we can rearrange (39) as

$$-\int_0^{\bar{k}}(F_1-F_2)\,2\pi(k)\left(\frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\left(1-\lambda-\lambda_\varepsilon k\right)^2\left(1-\lambda+\lambda_\varepsilon^2\right)}+\frac{\lambda_\varepsilon}{\left(1-\lambda-\lambda_\varepsilon k\right)}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk \;=\; -\int_0^{\bar{k}}(F_1-F_2)\,2\pi(k)\left(g_1\left(k\right)+g_2\left(k\right)\right)dk \;\geq\; 0$$

[30]In fact the first terms, $-\left[\left(\frac{\partial\pi(k,\varepsilon)}{\partial\varepsilon}|_{k=0}+\pi(k)|_{k=0}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)F_j\left(0\right)\right]$ for $j=1,2$, cancel out because they are constants.

We now characterize some properties of the functions $g_1(k), g_2(k)$, which we will use to characterize the sufficient conditions for $\bar{\rho}'(\varepsilon) \geq 0$. To ease intuition notice that

$$
\begin{aligned}
(g_1(k) + g_2(k)) &= \frac{\lambda_\varepsilon}{(1-\lambda-\lambda_\varepsilon k)}\left[\frac{\frac{\partial^2\pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial\pi(k)}{\partial k}} + \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right] \\
&= \frac{-\frac{\partial\pi(k)}{\partial k}}{2\pi(k)}\left[\frac{\frac{\partial^2\pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial\pi(k)}{\partial k}} + \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right] \\
&= \frac{-\frac{\partial\pi(k)}{\partial k}}{2\pi(k)}\left[-e_{\pi'(k),\varepsilon} + e_{f_2(\bar{k}),\varepsilon}\right]
\end{aligned}
$$

So $g_1(k) + g_2(k) \geq 0$ if and only if $-e_{\pi'(k),\varepsilon} + e_{f_2(\bar{k}),\varepsilon} \geq 0$. Notice also that for $k = 0$ we have

$$
-e_{\pi'(0),\varepsilon} = \frac{\frac{\partial^2\pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial\pi(k)}{\partial k}}|_{k=0} = \frac{[1-\lambda-\lambda_\varepsilon^2]}{\lambda_\varepsilon(1-\lambda+\lambda_\varepsilon^2)}
$$

which is non negative if $1-\lambda-\lambda_\varepsilon^2 \geq 0$. Thus $-e_{\pi'(0),\varepsilon}$ is non negative for $\varepsilon \in (0, \varepsilon_2)$ with $\varepsilon_2 = -(1-\lambda) + \sqrt{(1-\lambda)} \in (0, \lambda)$.

Lemma 5. *If $\varepsilon \in (0, \varepsilon_2)$, with $\varepsilon_2 = -(1-\lambda) + \sqrt{(1-\lambda)}$ then $[g_1(k) + g_2(k)]|_{k=0} > 0$ for all $\lambda \in (0,1)$. If $\varepsilon \in (\varepsilon_2, \lambda)$, then $[g_1(k) + g_2(k)]|_{k=0} > 0$ if and only if*

$$
\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k} > -\frac{(1-\lambda-\lambda_\varepsilon^2)}{(1-\lambda+\lambda_\varepsilon^2)} \tag{42}
$$

Condition (42) is equivalent to $-e_{\pi'(k=0),\varepsilon} + e_{f_2(\bar{k}),\varepsilon} > 0$. A sufficient condition for (42) to be satisfied is

$$
\frac{f_2(\bar{k})}{F_2(\bar{k})}\bar{k} > \frac{\lambda_\varepsilon^2 - (1-\lambda)}{(1-\lambda+\lambda_\varepsilon^2)} \tag{43}
$$

In general, if (33) holds then (42) is always satisfied.

Proof. Condition (42) follows from evaluating $g_1(k) + g_2(k)$ at $k = 0$, and substituting $\bar{k} = \frac{(1-\lambda)}{\lambda_\varepsilon}$. then we have that $[g_1(k) + g_2(k)]|_{k=0} > 0$ if and only if:

$$
\frac{(1-\lambda-\lambda_\varepsilon^2)}{(1-\lambda)(1-\lambda+\lambda_\varepsilon^2)} + \frac{\lambda_\varepsilon}{(1-\lambda)}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon} > 0
$$

which is always satisfied if $\varepsilon \in (0, \varepsilon_2)$, as it implies $(1-\lambda-\lambda_\varepsilon^2) \geq 0$. If $\varepsilon \in (\varepsilon_2, \lambda)$ then this inequality is satisfied if and only if (42) holds. That (33) implies (42) follows from rewriting (33) as

$$\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k} \quad > \quad \frac{2\lambda_\varepsilon^2}{(1-\lambda+\lambda_\varepsilon^2)}$$

Combining then (33) with (42) yields

$$\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k} \quad > \quad \frac{1}{(1-\lambda+\lambda_\varepsilon^2)}\max\left(2\lambda_\varepsilon^2, -\left(1-\lambda-\lambda_\varepsilon^2\right)\right)$$

conditional on $(1-\lambda-\lambda_\varepsilon^2) < 0$. It is then easy to see that $\max\left(2\lambda_\varepsilon^2, -\left(1-\lambda-\lambda_\varepsilon^2\right)\right) = 2\lambda_\varepsilon^2$, as that's true if and only if $\lambda_\varepsilon^2 > -(1-\lambda)$, which is always satisfied. Finally, because $\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}$ is decreasing in ρ, then a sufficient condition for (42) to be satisfied is

$$\frac{f_2(\bar{k})}{F_2(\bar{k})}\bar{k} \quad > \quad \frac{\lambda_\varepsilon^2-(1-\lambda)}{(1-\lambda+\lambda_\varepsilon^2)}$$

□

Lemma 6. *If (42) is satisfied then there exists a unique $k^0 \in \left[0,\bar{k}\right]$ such that $g_1(k)+g_2(k)=0$. This is equivalent to $-e_{\pi'(k=k_0),\varepsilon}+e_{f_2(\bar{k}),\varepsilon}=0$.*

Proof. Follows from i)$(1-\lambda)\lim_{k\to\bar{k}}\left(g_1(k)+g_2(k)\right)=-\infty$, ii) $g_1(0)+g_2(0)>0$ and iii) $g_1(k)+g_2(k)$ strictly monotonic in k.

To show i) evaluate $(1-\lambda)\lim_{k\to\bar{k}}\left(g_1(k)+g_2(k)\right)$:

$$\begin{aligned}
\lim_{k\to\bar{k}}\left(\frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\left(1-\lambda-\lambda_\varepsilon k\right)^2\left(1-\lambda+\lambda_\varepsilon^2\right)}+\frac{\lambda_\varepsilon}{\left(1-\lambda-\lambda_\varepsilon k\right)}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right) &= \\
(1-\lambda)\lim_{k\to\bar{k}}\left(\frac{F(\bar{k};\bar{\rho})\lambda_\varepsilon\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)+f(\bar{k};\bar{\rho})\left(1-\lambda-\lambda_\varepsilon k\right)\left(1-\lambda+\lambda_\varepsilon^2\right)}{\left(1-\lambda-\lambda_\varepsilon k\right)^2\left(1-\lambda+\lambda_\varepsilon^2\right)F(\bar{k};\bar{\rho})\lambda_\varepsilon}\right) &= -\infty
\end{aligned}$$

since

$$\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)|_{\bar{k}}<0,$$

$$\lim_{k\to\bar{k}}\left\{f(\bar{k};\bar{\rho})\left(1-\lambda-\lambda_\varepsilon k\right)\left(1-\lambda+\lambda_\varepsilon^2\right)\right\}=0$$

and

$$\lim_{k\to\bar{k}}\left(1-\lambda-\lambda_\varepsilon k\right)=0$$

.

Result ii) follows from Lemma 5.

To show iii) we evaluate:

$$\begin{aligned}\frac{\partial\left[g_1(k)+g_2(k)\right]}{\partial k} &= \frac{(1-\lambda)}{(1-\lambda+\lambda_\varepsilon^2)}2\lambda_\varepsilon\frac{(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2)-(1-\lambda-\lambda_\varepsilon k)}{(1-\lambda-\lambda_\varepsilon k)^3}+\\ &\quad \frac{-\lambda_\varepsilon}{(1-\lambda-\lambda_\varepsilon k)^2}\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k}<0\end{aligned}$$

where the last inequality follows from

$$\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)-(1-\lambda-\lambda_\varepsilon k) = -k\lambda_\varepsilon-\lambda_\varepsilon^2<0.$$

□

Then we are ready to prove Proposition 2. Under the assumption that (42) is satisfied we can use results from Lemmas (5) and (6). Then $\bar{\rho}'(\varepsilon)\geq 0$ if and only if

$$\begin{aligned}&\int_0^{k^0}(F_2-F_1)\frac{2\pi(k)}{(1-\lambda-\lambda_\varepsilon k)}\left(\frac{(1-\lambda)(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2)}{(1-\lambda-\lambda_\varepsilon k)(1-\lambda+\lambda_\varepsilon^2)}+\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k}\right)dk \geq\\ &\int_{k^0}^{\bar{k}}-(F_2-F_1)\frac{2\pi(k)}{(1-\lambda-\lambda_\varepsilon k)}\left(\frac{(1-\lambda)(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2)}{(1-\lambda-\lambda_\varepsilon k)(1-\lambda+\lambda_\varepsilon^2)}+\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\bar{k}\right)dk\end{aligned}$$

Recall from Appendix A.1 that

$$\begin{aligned}\frac{\partial^2\pi(k)}{\partial\varepsilon\partial k} &= \frac{\partial\pi(k)}{\partial k}\frac{(1-\lambda)(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2)}{\lambda_\varepsilon(1-\lambda-\lambda_\varepsilon k)(1-\lambda+\lambda_\varepsilon^2)}\\ \frac{\partial\pi(k)}{\partial k} &= -\pi(k)\frac{2\lambda_\varepsilon}{(1-\lambda-\lambda_\varepsilon k)}\end{aligned}$$

Then $\bar{\rho}'(\varepsilon)\geq 0$ if and only if (42) is satisfied and

$$\int_0^{k^0}-(F_2-F_1)\frac{\partial\pi(k)}{\partial k}\left(\frac{\frac{\partial^2\pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial\pi(k)}{\partial k}}+\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk \geq \int_{k^0}^{\bar{k}}(F_2-F_1)\frac{\partial\pi(k)}{\partial k}\left(\frac{\frac{\partial^2\pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial\pi(k)}{\partial k}}+\frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})}\frac{\bar{k}}{\lambda_\varepsilon}\right)dk \quad (44)$$

Lemma 7. *Suppose F_1 CSD F_2, i.e.*

$$\frac{f_1(k)}{F_1(k)}\geq\frac{f_2(k)}{F_2(k)}\forall k$$

then

$$\frac{f(k;\rho)}{F(k;\rho)} = \frac{(1-\rho)f_1(k) + \rho f_2(k)}{(1-\rho)F_1(k) + \rho F_2(k)}$$

is decreasing in ρ.

Proof. Differentiating $\frac{f(k;\rho)}{F(k;\rho)}$ with respect to ρ yields:

$$\begin{aligned} & \frac{[f_2(k) - f_1(k)]\,[(1-\rho)F_1(k) + \rho F_2(k)]}{F(\rho)^2} \\ & -\frac{[F_2(k) - F_1(k)]\,[(1-\rho)f_1(k) + \rho f_2(k)]}{F(\rho)^2} \quad = \\ \frac{(1-\rho)F_2(k)F_1(k)\left\{\frac{f_2(k)}{F_2(k)} - \frac{f_1(k)}{F_1(k)}\right\} + \rho F_2(k)F_1(k)\left\{\frac{f_2(k)}{F_2(k)} - \frac{f_1(k)}{F_1(k)}\right\}}{F(\rho)^2} & \quad < \quad 0 \end{aligned}$$

where the last inequality follows from $\frac{f_1(k)}{F_1(k)} \geq \frac{f_2(k)}{F_2(k)} \forall k$. □

Lemma 7 shows that the reverse hazard ratio is decreasing in ρ: $\frac{\partial\left(\frac{f(k,\rho)}{F(k,\rho)}\right)}{\partial \rho} \leq 0$. Then inequality (44) is satisfied if:

$$\begin{aligned} \int_0^{k^0} -(F_2 - F_1)\frac{\partial \pi(k)}{\partial k}\left(\frac{\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}}{\frac{\partial \pi(k)}{\partial k}} + \frac{f_2\left(\bar{k}\right)}{F_2\left(\bar{k}\right)}\frac{\bar{k}}{\lambda_\varepsilon}\right) dk \quad \geq \\ \int_{k^0}^{\bar{k}} (F_2 - F_1)\frac{\partial \pi(k)}{\partial k}\left(\frac{\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}}{\frac{\partial \pi(k)}{\partial k}} + \frac{f_1\left(\bar{k}\right)}{F_1\left(\bar{k}\right)}\frac{\bar{k}}{\lambda_\varepsilon}\right) dk \end{aligned}$$

which is equivalent to

$$\begin{aligned} \int_0^{k^0} (F_1 - F_2)\frac{\partial \pi(k)}{\partial k}\left(-e_{\pi'(k),\varepsilon} + e_{f_2(\bar{k}),\varepsilon}\right) dk \quad \geq \\ \int_{k^0}^{\bar{k}} (F_1 - F_2)\frac{\partial \pi(k)}{\partial k}\left[-\left(-e_{\pi'(k),\varepsilon} + e_{f_2(\bar{k}),\varepsilon}\right)\right] dk \end{aligned}$$

This completes the proof of Proposition 2. □

A.4.3 Proof of Proposition 3

Proof. Let $A = \int_0^{\bar{k}} \pi(k) dF_2(k)$ and $B = \int_0^{\bar{k}} \pi(k) dF_1(k)$. The first order condition (27) can simply be written as

$$\gamma'(\rho^*) F\left(\bar{k}, \rho^*\right) \;=\; A - B + \frac{(A-B)}{2} - \frac{3}{2} C\left(\rho^*\right)$$

Notice that

$$\frac{\partial C(\rho)}{\partial \rho} \;=\; \int_0^{\bar{k}} \pi(k) \left[F_2\left(\bar{k}\right) - F_1\left(\bar{k}\right)\right] \left\{ \frac{\partial \left(\frac{1}{F\left(\bar{k}, \rho\right)}\right) dF(k, \rho) + \frac{1}{F\left(\bar{k}, \rho\right)} \frac{\partial f\left(\bar{k}, \rho\right)}{\partial \rho} dk}{\partial \rho} \right\}$$

which can be simplified to:

$$\frac{\partial C(\rho)}{\partial \rho} \;=\; \frac{\left[F_2\left(\bar{k}\right) - F_1\left(\bar{k}\right)\right]}{F\left(\bar{k}, \rho\right)} \left(A - B - C(\rho)\right)$$

Then, combining (27) with our assumption that $\gamma'(\rho) > 0$, yields $(A - B - C(\rho)) > 0$. Also, because F_1 first order stochastically dominates F_2 then $F_2\left(\bar{k}\right) - F_1\left(\bar{k}\right) \geq 0$. Hence $\frac{\partial C(\rho)}{\partial \rho} \geq 0$. Recall now that $\rho \in [0, 1]$ and that $\gamma(\rho)$ satisfies Inada conditions. Then, for $\rho \to 0$ the right and left hand side of (27) are, respectively: $\lim_{\rho \to 0} \gamma'(\rho) = 0$ and $C(0) = \int_0^{\bar{k}} \pi(k) \frac{\left[F_2\left(\bar{k}\right) - F_1\left(\bar{k}\right)\right]}{F_1\left(\bar{k}\right)} dF_1(k) > 0$. For $\rho \to 1$ the right and left hand side of (27) are, respectively: $\lim_{\rho \to 1} \gamma'(\rho) = +\infty$ and $C(1) = \int_0^{\bar{k}} \pi(k) \frac{\left[F_2\left(\bar{k}\right) - F_1\left(\bar{k}\right)\right]}{F_2\left(\bar{k}\right)} dF_2(k) < +\infty$. Then there exists a unique $\rho \in (0, 1)$ that satisfies (27). The first order condition for the equilibrium choice of ρ by dealers, (20), can then also be written in terms of A, B:

$$\gamma'(\bar{\rho}) F\left(\bar{k}, \bar{\rho}\right) \;=\; A - B$$

Substituting this back into (27) yields:

$$\gamma'(\rho^*) F\left(\bar{k}, \rho^*\right) = \gamma'(\bar{\rho}) F\left(\bar{k}, \bar{\rho}\right) + \frac{(A-B)}{2} - \frac{3}{2} C(\rho^*) \tag{45}$$

Since $\gamma'(\rho) F(\bar{k}, \rho)$ is increasing in ρ, $\rho^* > \bar{\rho}$ whenever

$$\frac{1}{3}(A - B) > C(\rho^*)$$

Moreover, recall that $\frac{\partial C(\rho)}{\partial \rho} \geq 0$. Hence,

$$C\left(\rho^{*}\right) \quad \leq \quad C(1)=\int_{0}^{\bar{k}} \pi(k) \frac{\left[F_{2}\left(\bar{k}\right)-F_{1}\left(\bar{k}\right)\right]}{F_{2}\left(\bar{k}\right)} d F_{2}(k)=\frac{\left[F_{2}\left(\bar{k}\right)-F_{1}\left(\bar{k}\right)\right]}{F_{2}\left(\bar{k}\right)} A$$

So a sufficient condition for $\rho^{*}(\varepsilon)>\bar{\rho}(\varepsilon)$ is $A-B>3 C(1)$ or

$$3 \frac{F_{1}\left(\bar{k}\right)}{F_{2}\left(\bar{k}\right)}-2 \quad > \quad \frac{\int_{0}^{\bar{k}} \pi(k) d F_{1}(k)}{\int_{0}^{\bar{k}} \pi(k) d F_{2}(k)} \tag{46}$$

□

A.4.4 First order condition of the constrained planner

The first order condition of the constrained planner's problem is

$$\begin{aligned}
\frac{\partial \bar{\rho}}{\partial \varepsilon}\left[\int_{0}^{\bar{k}} \frac{3}{2} \frac{\left(F_{2}\left(\bar{k}\right)-F_{1}\left(\bar{k}\right)\right)}{F\left(\bar{k}, \bar{\rho}\right)^{2}} \pi(k) d F(k ; \bar{\rho}(\varepsilon))-\int_{0}^{\bar{k}} \frac{1}{2} \pi(k)\left[f_{2}(k)-f_{1}(k)\right] \frac{d k}{F\left(\bar{k}, \bar{\rho}\right)}\right] \quad \geq \\
\int_{0}^{\bar{k}} \frac{3}{2} \frac{\partial \pi}{\partial \varepsilon} \frac{d F(k ; \bar{\rho}(\varepsilon))}{F\left(\bar{k}, \bar{\rho}\right)}-\int_{0}^{\bar{k}} \frac{3}{2} \frac{\pi(k)}{F\left(\bar{k}, \bar{\rho}\right)^{2}}\left[(1-\bar{\rho}) f_{1}\left(\bar{k}\right)+\bar{\rho} f_{2}\left(\bar{k}\right)\right] \frac{\partial \bar{k}}{\partial \varepsilon} d F(k ; \bar{\rho}(\varepsilon))
\end{aligned} \tag{47}$$

For $\hat{\varepsilon} \in[0, \lambda)$ we can rearrange (47) as

$$\begin{aligned}
\int_{0}^{\bar{k}} \frac{3}{2} \frac{\partial \pi}{\partial \varepsilon} \frac{d F(k ; \bar{\rho}(\varepsilon))}{F\left(\bar{k}, \bar{\rho}\right)}-\int_{0}^{\bar{k}} \frac{3}{2} \frac{\left(F_{2}\left(\bar{k}\right)-F_{1}\left(\bar{k}\right)\right)}{F\left(\bar{k}, \rho\right)^{2}} \pi(k) \frac{\partial \bar{\rho}}{\partial \varepsilon} d F(k ; \bar{\rho}(\varepsilon)) \quad + \\
-\int_{0}^{\bar{k}} \frac{3}{2} \frac{\pi(k)}{F\left(\bar{k}, \bar{\rho}\right)^{2}}\left[(1-\bar{\rho}) f_{1}\left(\bar{k}\right)+\bar{\rho} f_{2}\left(\bar{k}\right)\right] \frac{\partial \bar{k}}{\partial \varepsilon} d F(k ; \bar{\rho}(\varepsilon)) \quad + \\
\int_{0}^{\bar{k}} \frac{3}{2} \pi(k) \frac{\partial \bar{\rho}}{\partial \varepsilon} \frac{d F_{2}(k)}{F\left(\bar{k}, \bar{\rho}\right)}-\int_{0}^{\bar{k}} \frac{3}{2} \pi(k) \frac{\partial \bar{\rho}}{\partial \varepsilon} \frac{d F_{1}(k)}{F\left(\bar{k}, \bar{\rho}\right)}-\gamma^{\prime}(\bar{\rho}) \frac{\partial \bar{\rho}}{\partial \varepsilon} \quad \leq \quad 0
\end{aligned}$$

Similar algebra is applied when $\hat{\varepsilon}=\lambda$. Then group $\frac{\partial \bar{\rho}}{\partial \varepsilon}$ terms

$$\begin{aligned}
\int_{0}^{\bar{k}} \frac{3}{2} \frac{\partial \pi}{\partial \varepsilon} \frac{d F(k ; \bar{\rho}(\varepsilon))}{F\left(\bar{k}, \rho\right)}-\int_{0}^{\bar{k}} \frac{3}{2} \frac{\pi(k)}{F\left(\bar{k}, \bar{\rho}\right)^{2}}\left[(1-\bar{\rho}) f_{1}\left(\bar{k}\right)+\bar{\rho} f_{2}\left(\bar{k}\right)\right] \frac{\partial \bar{k}}{\partial \varepsilon} d F(k ; \bar{\rho}(\varepsilon)) \quad + \\
\frac{\partial \bar{\rho}}{\partial \varepsilon}\left\{-\int_{0}^{\bar{k}} \frac{3}{2} \frac{\left(F_{2}\left(\bar{k}\right)-F_{1}\left(\bar{k}\right)\right)}{F\left(\bar{k}, \rho\right)^{2}} \pi(k) d F(k ; \bar{\rho}(\varepsilon))+\int_{0}^{\bar{k}} \frac{3}{2} \pi(k) \frac{d F_{2}(k)}{F\left(\bar{k}, \bar{\rho}\right)}-\int_{0}^{\bar{k}} \frac{3}{2} \pi(k) \frac{d F_{1}(k)}{F\left(\bar{k}, \bar{\rho}\right)}-\gamma^{\prime}(\bar{\rho})\right\} \quad \leq \quad 0
\end{aligned}$$

which, using $\gamma'(\bar{\rho}) = \int_0^{\bar{k}} \pi(k) \frac{dF_2(k)}{F(\bar{k},\bar{\rho})} - \int_0^{\bar{k}} \pi(k) \frac{dF_1(k)}{F(\bar{k},\bar{\rho})}$, can be rearranged as

$$
\begin{aligned}
& H(k;\bar{\rho}(\varepsilon)) \quad + \\
\frac{\partial \bar{\rho}}{\partial \varepsilon}\left\{-\int_0^{\bar{k}} \frac{3}{2}\frac{\left(F_2(\bar{k})-F_1(\bar{k})\right)}{F(\bar{k},\bar{\rho})^2}\pi(k)\,dF(k;\bar{\rho}(\varepsilon)) + \int_0^{\bar{k}} \frac{1}{2}\pi(k)\frac{dF_2(k)}{F(\bar{k},\bar{\rho})} - \int_0^{\bar{k}} \frac{1}{2}\pi(k)\frac{dF_1(k)}{F(\bar{k},\bar{\rho})}\right\} & \leq \quad 0
\end{aligned}
$$

where we use the definition of $H(\rho)$. This can be further rearranged using the definition of $A, B, 3C(\bar{\rho})$ in Proposition 3, as:

$$
H(\bar{\rho}(\varepsilon)) + \frac{\partial \bar{\rho}}{\partial \varepsilon}\frac{[A-B-3C(\bar{\rho})]}{2F(\bar{k},\bar{\rho})} \quad \leq \quad 0
$$

A.4.5 Proof of Lemma 4

Using the definition of $H(\rho,\varepsilon)$ we have

$$
\begin{aligned}
H(\rho,\varepsilon) &= \int_0^{\bar{k}} \frac{3}{2}\frac{\partial \pi(k)}{\partial \varepsilon}\frac{dF(k;\rho)}{F(\bar{k},\rho)} - \int_0^{\bar{k}} \frac{3}{2}\pi(k)\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)}\frac{\partial \bar{k}}{\partial \varepsilon}\frac{dF(k;\rho)}{F(\bar{k},\rho)} \\
&= \frac{3}{2}\int_0^{\bar{k}} G(k,\rho)\frac{dF(k;\rho)}{F(\bar{k},\rho)}
\end{aligned}
$$

where $G(k,\rho) = \frac{\partial \pi(k)}{\partial \varepsilon} - \pi(k)\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)}\frac{\partial \bar{k}}{\partial \varepsilon}$. Integrating by parts then gives:

$$
\begin{aligned}
\frac{2}{3}H(\rho,\varepsilon) &= \left[\frac{G(k,\rho)F(k;\rho)}{F(\bar{k},\rho)}\right]_0^{\bar{k}} - \int_0^{\bar{k}} \frac{\partial G(\bar{k},\rho)}{\partial k}\frac{F(k;\rho)}{F(\bar{k},\rho)}dk \\
&= -\int_0^{\bar{k}} \left[\left(\frac{\partial \pi(k)}{\partial k}\frac{\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}}{\frac{\partial \pi(k)}{\partial k}}\right)\frac{F(k;\rho)}{F(\bar{k},\rho)} - \frac{\partial \pi(k)}{\partial k}\frac{F(k;\rho)}{F(\bar{k},\rho)}\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)}\frac{\partial \bar{k}}{\partial \varepsilon}\right]dk
\end{aligned}
$$

so that:

$$
\begin{aligned}
\frac{2}{3}\frac{\partial H(\rho,\varepsilon)}{\partial \rho} &= -\int_0^{\bar{k}} \left[\left(\frac{\partial \pi(k)}{\partial k}\frac{\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}}{\frac{\partial \pi(k)}{\partial k}}\right)\frac{\partial}{\partial \rho}\left(\frac{F(k;\rho)}{F(\bar{k},\rho)}\right) - \frac{\partial \pi(k)}{\partial k}\frac{\partial \bar{k}}{\partial \varepsilon}\frac{\partial}{\partial \rho}\left(\frac{F(k;\rho)}{F(\bar{k},\rho)}\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)}\right)\right]dk \\
&= -\int_0^{\bar{k}} \Bigg[\left\{\frac{\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon}\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)}\right\}\frac{\partial \pi(k)}{\partial k}\frac{\partial}{\partial \rho}\left(\frac{F(k;\rho)}{F(\bar{k},\rho)}\right) - \\
&\qquad \left(\frac{\partial \pi(k)}{\partial k}\frac{\partial \bar{k}}{\partial \varepsilon}\right)\frac{F(k;\rho)}{F(\bar{k},\rho)}\frac{\partial}{\partial \rho}\left(\frac{f(\bar{k},\rho)}{F(\bar{k},\rho)}\right)\Bigg]dk
\end{aligned}
$$

then we can consider the functions g_1, g_2 defined in (40) and (41), to define k_0 as in Lemma 6, so that $g_1(k) + g_2(k) = 0$:

$$\frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\left(1-\lambda-\lambda_\varepsilon k\right)^2\left(1-\lambda+\lambda_\varepsilon^2\right)} + \frac{\lambda_\varepsilon}{\left(1-\lambda-\lambda_\varepsilon k\right)} \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\bar{k}}{\lambda_\varepsilon} \quad = \quad 0$$

which can be simplified to:

$$\frac{\lambda_\varepsilon}{\left(1-\lambda-\lambda_\varepsilon k\right)} \left[\frac{\frac{\partial^2 \pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right] \quad = \quad 0$$

Hence notice that k_0 is the value such that $\left[\frac{\frac{\partial^2 \pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right] = \left[-e_{\pi'(k),\varepsilon} + e_{f_2\left(\bar{k}\right),\varepsilon}\right] = 0$ and we assumed that $\left[\frac{\frac{\partial^2 \pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right] > 0$ for $k \in [0, k_0)$ and notice that $\left(\frac{\frac{\partial^2 \pi(k,\varepsilon)}{\partial\varepsilon\partial k}}{\frac{\partial \pi(k,\varepsilon)}{\partial k}} + \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\bar{k}}{\lambda_\varepsilon}\right)$ is decreasing in k:

$$\begin{aligned}\frac{\partial}{\partial k}\left[\frac{(1-\lambda)\left(1-\lambda-2k\lambda_\varepsilon-\lambda_\varepsilon^2\right)}{\lambda_\varepsilon\left(1-\lambda-\lambda_\varepsilon k\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} + \frac{f(\bar{k};\bar{\rho})}{F(\bar{k};\bar{\rho})} \frac{\bar{k}}{\lambda_\varepsilon}\right] \quad &= \\ \frac{(1-\lambda)}{\lambda_\varepsilon\left(1-\lambda+\lambda_\varepsilon^2\right)} \left[\frac{-\lambda_\varepsilon(1-\lambda)-\lambda_\varepsilon^3}{\left(1-\lambda-\lambda_\varepsilon k\right)^2}\right] \quad &< \quad 0\end{aligned}$$

Then for $\frac{\partial H(\rho,\varepsilon)}{\partial \rho} \leq 0$ it is necessary and sufficient that

$$\int_0^{\bar{k}} \left[\left\{\frac{\frac{\partial^2 \pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right\} \frac{\partial \pi(k)}{\partial k} \frac{\partial}{\partial \rho}\left(\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)}\right) - \left(\frac{\partial \pi(k)}{\partial k}\frac{\partial \bar{k}}{\partial \varepsilon}\right) \frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \frac{\partial}{\partial \rho}\left(\frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right)\right] dk \quad \geq \quad 0$$

Because $\frac{\partial \pi(k)}{\partial k} \leq 0$, $\frac{\partial}{\partial \rho}\left(\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)}\right) \leq 0$ for the first term in square brackets, and $\left(\frac{\partial \pi(k)}{\partial k}\frac{\partial \bar{k}}{\partial \varepsilon}\right) \geq 0$, $\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \frac{\partial}{\partial \rho}\left(\frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right) \leq 0$ for the second term in square brackets, then the second term in square brackets is always non negative, while the first is non negative if and only if $k \in [0, k_0]$. Hence a sufficient condition for $\frac{\partial H(\rho,\varepsilon)}{\partial \rho} \leq 0$ is

$$\begin{aligned}&\int_0^{k_0} \left[\left\{\frac{\frac{\partial^2 \pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right\} \frac{\partial \pi(k)}{\partial k} \frac{\partial}{\partial \rho}\left(\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)}\right) - \left(\frac{\partial \pi(k)}{\partial k}\frac{\partial \bar{k}}{\partial \varepsilon}\right) \frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \frac{\partial}{\partial \rho}\left(\frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right)\right] dk \quad \geq \\ &\int_{k_0}^{\bar{k}} \left[-\left\{\frac{\frac{\partial^2 \pi(k)}{\partial\varepsilon\partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right\} \frac{\partial \pi(k)}{\partial k} \frac{\partial}{\partial \rho}\left(\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)}\right) + \left(\frac{\partial \pi(k)}{\partial k}\frac{\partial \bar{k}}{\partial \varepsilon}\right) \frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \frac{\partial}{\partial \rho}\left(\frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)}\right)\right] dk\end{aligned}$$

Or could write it as

$$\int_0^{\bar{k}} \left\{ \frac{\frac{\partial^2 \pi(k)}{\partial \varepsilon \partial k}}{\frac{\partial \pi(k)}{\partial k}} - \frac{\partial \bar{k}}{\partial \varepsilon} \frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)} \right\} \frac{\partial \pi(k)}{\partial k} \frac{\partial}{\partial \rho} \left(\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \right) dk - \int_0^{\bar{k}} \left[\left(\frac{\partial \pi(k)}{\partial k} \frac{\partial \bar{k}}{\partial \varepsilon} \right) \frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \frac{\partial}{\partial \rho} \left(\frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)} \right) \right] dk \;\geq\; 0$$

and then

$$\int_0^{k_0} \frac{\partial \pi(k)}{\partial k} \frac{\partial}{\partial \rho} \left(\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \right) |e_{f_2(\bar{k}),\varepsilon} - e_{\pi'(k),\varepsilon}| dk - \int_0^{\bar{k}} \left[\left(\frac{\partial \pi(k)}{\partial k} \frac{\partial \bar{k}}{\partial \varepsilon} \right) \frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \frac{\partial}{\partial \rho} \left(\frac{f\left(\bar{k},\rho\right)}{F\left(\bar{k},\rho\right)} \right) \right] dk \;\geq\; \int_{k_0}^{\bar{k}} \frac{\partial \pi(k)}{\partial k} \frac{\partial}{\partial \rho} \left(\frac{F(k;\rho)}{F\left(\bar{k},\rho\right)} \right) |e_{f_2(\bar{k}),\varepsilon} - e_{\pi'(k),\varepsilon}| dk$$

Notice that this is a sufficient condition and not a necessary one, because we are using the same threshold k_0 that we used to prove Proposition 2.

A.4.6 Condition for $H'(\rho)$, with mass-at-zero example

We now turn to conditions under which $H'(\rho) < 0$. For our argument, it is enough to show this holds at $\rho^*(\varepsilon^*)$. With the mass-at-zero distribution, notice that $f(k,\rho)$ only depends on ρ when $k = 0$. Therefore $(1-\rho) f_1\left(\bar{k}\right) + \rho f_2\left(\bar{k}\right) = f(\bar{k})$ and

$$H\left(\rho\right) \;=\; \int_0^{\bar{k}} \left[\frac{3}{2} \frac{\partial \pi}{\partial \varepsilon} - \frac{3}{2} \pi\left(k\right) \frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho\right)} \frac{\partial \bar{k}}{\partial \varepsilon} \right] \frac{dF(k;\rho)}{F\left(\bar{k},\rho\right)}$$

Hence, after rearranging, we have:

$$\begin{aligned} H'(\rho) \;=\; & \int_0^{\bar{k}} \left[\frac{3}{2} \pi\left(k\right) \frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho\right)^2} (F_2(\bar{k}) - F_1(\bar{k})) \frac{\partial \bar{k}}{\partial \varepsilon} \right] \frac{dF(k;\rho)}{F\left(\bar{k},\rho\right)} - H(\rho) \frac{(F_2(\bar{k}) - F_1(\bar{k}))}{F(\bar{k},\rho)} \\ & + \left[\frac{3}{2} \frac{\partial \pi(0)}{\partial \varepsilon} - \frac{3}{2} \pi\left(0\right) \frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho\right)} \frac{\partial \bar{k}}{\partial \varepsilon} \right] (f_2(0) - f_1(0)) \frac{1}{F\left(\bar{k},\rho\right)} \end{aligned}$$

Notice that $F_2(\bar{k}) - F_1(\bar{k}) = \int_0^{\bar{k}} f_2(k) - f_1(k) dk = f_2(0) - f_1(0)$. Therefore

$$\begin{aligned} H'(\rho) \;=\; & \int_0^{\bar{k}} \left[\frac{3}{2} \pi\left(k\right) \frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho\right)^2} (f_2(0) - f_1(0)) \frac{\partial \bar{k}}{\partial \varepsilon} \right] \frac{dF(k;\rho)}{F\left(\bar{k},\rho\right)} - H(\rho) \frac{(f_2(0) - f_1(0))}{F(\bar{k},\rho)} \\ & + \left[\frac{3}{2} \frac{\partial \pi(0)}{\partial \varepsilon} - \frac{3}{2} \pi\left(0\right) \frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho\right)} \frac{\partial \bar{k}}{\partial \varepsilon} \right] (f_2(0) - f_1(0)) \frac{1}{F\left(\bar{k},\rho\right)} \end{aligned}$$

and factoring out $(f_2(0) - f_1(0))/F(\bar{k}, \rho) > 0$, we only need to check the sign of

$$h(\rho) \equiv \int_0^{\bar{k}} \left[\frac{3}{2}\pi(k)\frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho\right)^2}\frac{\partial \bar{k}}{\partial \varepsilon}\right] dF(k;\rho) - H(\rho) + \left[\frac{3}{2}\frac{\partial \pi(0)}{\partial \varepsilon} - \frac{3}{2}\pi(0)\frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho\right)}\frac{\partial \bar{k}}{\partial \varepsilon}\right]$$

Evaluated at ρ^*, $H(\rho^*) \geq 0$ (If $H(\rho^*) < 0$ then $\rho^* = 0$ and the result follows). Hence

$$h(\rho^*) \leq \int_0^{\bar{k}} \left[\frac{3}{2}\pi(k)\frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho^*\right)^2}\frac{\partial \bar{k}}{\partial \varepsilon}\right] dF(k;\rho^*) + \frac{3}{2}\frac{\partial \pi(0)}{\partial \varepsilon} - \frac{3}{2}\pi(0)\frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho^*\right)}\frac{\partial \bar{k}}{\partial \varepsilon}$$

which can be rearranged as

$$\frac{2}{3}h(\rho^*) \leq -\pi(0)\frac{f\left(\bar{k}\right)}{F\left(\bar{k},\rho^*\right)^2}\frac{1-\lambda}{\lambda_\varepsilon^2}\left[\rho^* f_2(0) + (1-\rho^*)f_1(0) - F\left(\bar{k},\rho^*\right)\right] - 2\pi(0)\frac{\lambda_\varepsilon}{(1-\lambda+\lambda_\varepsilon^2)}$$

Hence, factoring $\pi(0) > 0$ from the right hand side, a sufficient condition for $h(\rho^*) < 0$ is:

$$-\frac{f\left(\bar{k}\right)}{F(\bar{k},\rho^*)}\frac{1-\lambda}{\lambda_\varepsilon^2}\frac{[\rho^* f_2(0) + (1-\rho^*)f_1(0)]}{\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})} + \frac{1-\lambda}{\lambda_\varepsilon^2}\frac{f\left(\bar{k}\right)}{\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})} - 2\frac{\lambda_\varepsilon}{(1-\lambda+\lambda_\varepsilon^2)} < 0 \quad (48)$$

Since $f(k,\rho)/F(k,\rho)$ is decreasing in ρ, this inequality is satisfied whenever we set the first ρ^* in the first expression to 1 and ρ^* in the second expression to 0. Hence, a sufficient condition for $h(\rho^*) < 0$ is

$$-\frac{f\left(\bar{k}\right)}{F_2(\bar{k})}\frac{1-\lambda}{\lambda_\varepsilon^2}\frac{[\rho^* f_2(0) + (1-\rho^*)f_1(0)]}{\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})} + \frac{f\left(\bar{k}\right)}{F_1(\bar{k})}\frac{1-\lambda}{\lambda_\varepsilon^2} - 2\frac{\lambda_\varepsilon}{(1-\lambda+\lambda_\varepsilon^2)} < 0$$

In the above inequality, consider now the term $\frac{[\rho^* f_2(0)+(1-\rho^*)f_1(0)]}{\rho^* F_2\left(\bar{k}\right)+(1-\rho^*)F_1(\bar{k})}$ and differentiate it with respect to ρ:

$$\begin{aligned}&\frac{\partial}{\partial \rho}\left(\frac{\rho^* f_2(0) + (1-\rho^*)f_1(0)}{\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})}\right)\\ =\ &\frac{[f_2(0) - f_1(0)]\left[\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})\right] - \left[F_2\left(\bar{k}\right) - F_1(\bar{k})\right]\left[\rho^* f_2(0) + (1-\rho^*)f_1(0)\right]}{\left[\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})\right]^2}\end{aligned}$$

which can be rearranged as

$$\frac{\partial}{\partial \rho}\left(\frac{\rho^* f_2(0) + (1-\rho^*)f_1(0)}{\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})}\right)$$

$$= \frac{[f_2(0) - f_1(0)]}{\left[\rho^* F_2\left(\bar{k}\right) + (1-\rho^*)F_1(\bar{k})\right]^2}\left[\rho^*\left[F_2\left(\bar{k}\right) - f_2(0)\right] + (1-\rho^*)\left[F_1(\bar{k}) - f_1(0)\right]\right] > 0$$

So inequality (48) is satisfied whenever

$$-\frac{f\left(\bar{k}\right)}{F_2(\bar{k})}\frac{1-\lambda}{\lambda_\varepsilon^2}\frac{f_1(0)}{F_1(\bar{k})} + \frac{f\left(\bar{k}\right)}{F_1(\bar{k})}\frac{1-\lambda}{\lambda_\varepsilon^2} - 2\frac{\lambda_\varepsilon}{(1-\lambda+\lambda_\varepsilon^2)} \quad < \quad 0$$

which can be rearranged as

$$\frac{f\left(\bar{k}\right)}{F_1(\bar{k})}\bar{k}\left[1 - \frac{f_1(0)}{F_2(\bar{k})}\right] \quad < \quad \frac{2\lambda_\varepsilon^2}{(1-\lambda+\lambda_\varepsilon^2)}$$

A.4.7 Proof of Proposition 4

Proof. Assumption i) implies $\bar{\rho}'(\varepsilon) > 0$. Then, because and $H(\rho, \varepsilon)$ is decreasing in ρ, we obtain

$$H(\rho^*, \varepsilon^*) \quad \leq \quad H(\bar{\rho}(\varepsilon^*), \varepsilon^*) < H(\bar{\rho}(\varepsilon^*), \varepsilon^*) + \frac{\partial\bar{\rho}}{\partial\varepsilon}\frac{[A - B - 3C(\rho)]}{2F\left(\bar{k}, \bar{\rho}\right)} \tag{49}$$

where the last inequality follows from assumptions i) and iii). Then we can argue the following: if $\varepsilon^* \in (0, \lambda)$ then $H(\rho^*, \varepsilon^*) = 0$ and the term on the right hand side of (49) when evaluated at ε^*, namely $H(\bar{\rho}(\varepsilon^*), \varepsilon^*) + \frac{\partial\bar{\rho}}{\partial\varepsilon}\frac{[A-B-3C(\rho)]}{2F\left(\bar{k},\bar{\rho}\right)} > 0$. This is equivalent to the first order condition of the constrained planner when evaluated at ε^*. As a result $\hat{\varepsilon} > \varepsilon^*$ when $\varepsilon^* \in (0, \lambda)$. And similarly for corner solutions we have $\hat{\varepsilon} \geq \varepsilon^*$: If $\varepsilon^* = \lambda$ (this implies $H(\bar{\rho}(\varepsilon^*), \varepsilon^*) > 0$), then $H(\bar{\rho}(\varepsilon^*), \varepsilon^*) + \frac{\partial\bar{\rho}}{\partial\varepsilon}\frac{[A-B-3C(\rho)]}{2F\left(\bar{k},\bar{\rho}\right)} > 0$ if $\varepsilon^* = \lambda$, so $\hat{\varepsilon} = \lambda = \varepsilon^*$. Finally if $\varepsilon^* = 0$ (this implies $H(\bar{\rho}(\varepsilon^*), \varepsilon^*) < 0$)) then $H(\bar{\rho}(\varepsilon^*), \varepsilon^*) + \frac{\partial\bar{\rho}}{\partial\varepsilon}\frac{[A-B-3C(\rho)]}{2F\left(\bar{k},\bar{\rho}\right)}$ can be either non negative or never positive, so $\hat{\varepsilon} \in [0, \lambda] \geq \varepsilon^* = 0$. □

A.5 Average Bid-Ask Spread: proofs of Lemma 1 and Proposition 5

Recall that the ask for a dealer k is

$$a(k) = \frac{1}{2} + \frac{\lambda_\varepsilon^2 + \lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)} = x + yk$$

Notice that $a\left(\bar{k}\right) = x + y\bar{k} = 1$, while the bid for a dealer k is

$$b(k) \quad = \quad \lambda_\varepsilon(1 - a(k))$$

Hence, the bid-ask posted by dealer k is

$$\sigma(k,\varepsilon) = (1+\lambda_\varepsilon)a(k) - \lambda_\varepsilon$$

Now recall also that the distribution of ask price given ρ is $G_a(a;\rho)$, where $G_a(a;\rho) = \frac{F\left(\frac{a-x}{y};\rho\right)}{F\left(\frac{1-x}{y};\rho\right)}$ so that $g_a(a;\rho) = \frac{1}{yF\left(\frac{1-x}{y};\rho\right)} f\left(\frac{a-x}{y};\rho\right) = \frac{1}{yF(\bar{k};\rho)} f(k;\rho)$. Then $dG_a(a;\rho) = g_a(a;\rho)\,da$, and $da = ydk$, so when doing the change of variable we have

$$dG_a(a;\rho) = \frac{1}{yF(\bar{k};\rho)} f(k;\rho)\,ydk = \frac{f(k;\rho)}{F(\bar{k};\rho)}dk$$

So, recall that

$$a(k) = x + yk = \frac{1}{2} + \frac{\lambda_\varepsilon^2 + \lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)}$$

which can be rearranged as

$$\frac{\partial a(\varepsilon,k)}{\partial \varepsilon} = \frac{1}{2}\frac{\lambda_\varepsilon(1-\lambda+\lambda_\varepsilon^2) + (\lambda_\varepsilon+k)(1-\lambda-\lambda_\varepsilon^2)}{(1-\lambda+\lambda_\varepsilon^2)^2}$$

so if $(1-\lambda-\lambda_\varepsilon^2) \geq 0$ then $\frac{\partial a(\varepsilon,k)}{\partial\varepsilon} \geq 0$ for all k. If, however, $(1-\lambda-\lambda_\varepsilon^2) < 0$ (that is to say if $\varepsilon > \varepsilon_2 = -(1-\lambda) + \sqrt{(1-\lambda)}$ then because $\frac{\partial a(\varepsilon,k)}{\partial\varepsilon}$ is decreasing in k in this case, then at $\bar{k}$

$$\begin{aligned}
\frac{\partial a(\varepsilon,k)}{\partial\varepsilon}\Big|_{\bar{k}} &= \frac{1}{2\lambda_\varepsilon}\frac{\lambda_\varepsilon^2(1-\lambda+\lambda_\varepsilon^2) + (1-\lambda+\lambda_\varepsilon^2)(1-\lambda-\lambda_\varepsilon^2)}{(1-\lambda+\lambda_\varepsilon^2)^2} \\
&= \frac{1}{2\lambda_\varepsilon}\frac{(1-\lambda+\lambda_\varepsilon^2)(1-\lambda)}{(1-\lambda+\lambda_\varepsilon^2)^2} > 0
\end{aligned}$$

Hence $\frac{\partial a(\varepsilon,k)}{\partial\varepsilon} \geq 0$ for all k, regardless of whether $(1-\lambda-\lambda_\varepsilon^2) \geq 0$ or $(1-\lambda-\lambda_\varepsilon^2) < 0$.[31]

The average bid-ask spread is

$$\sigma(\varepsilon) = \int_0^{\bar{k}} \sigma(\varepsilon,k)\frac{f(k;\rho)}{F(\bar{k};\rho)}dk$$

---

[31] From

$$\begin{aligned}
\frac{\partial a(\varepsilon,k)}{\partial\varepsilon} &= \frac{1}{2}\frac{\lambda_\varepsilon\left(1-\lambda+\lambda_\varepsilon^2\right) + (\lambda_\varepsilon+k)\left(1-\lambda-\lambda_\varepsilon^2\right)}{(1-\lambda+\lambda_\varepsilon^2)^2} \\
1 - a(\varepsilon,k) &= \frac{1}{2} - \frac{\lambda_\varepsilon^2 + \lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)}
\end{aligned}$$

and the last expression is 0 at $k = \bar{k}$.

then

$$\frac{\partial \sigma(\varepsilon)}{\partial \varepsilon} = \int_0^{\bar{k}} \left\{ \frac{\partial \sigma(\varepsilon,k)}{\partial \varepsilon} \frac{f(k;\rho)}{F(\bar{k};\rho)} + \sigma(\varepsilon,k) \frac{\partial}{\partial \varepsilon} \left(\frac{f(k;\rho)}{F(\bar{k};\rho)} \right) \right\} dk + \\ + \sigma(\varepsilon,\bar{k}) \frac{f(\bar{k};\rho)}{F(\bar{k};\rho)} \frac{\partial \bar{k}}{\partial \varepsilon} \tag{50}$$

The first term under the integral in (50) is always positive:

$$\begin{aligned} \frac{\partial \sigma(\varepsilon,k)}{\partial \varepsilon} &= (1+\lambda_\varepsilon)\frac{\partial a(\varepsilon,k)}{\partial \varepsilon} - (1-a(k)) \\ &= (1+\lambda_\varepsilon)\frac{\partial a(\varepsilon,k)}{\partial \varepsilon} - \frac{1}{2} + \frac{\lambda_\varepsilon^2 + \lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)} \end{aligned} \tag{51}$$

Hence:

$$\frac{\partial \sigma(\varepsilon,k)}{\partial \varepsilon} = (1+\lambda_\varepsilon)\frac{1}{2}\frac{\lambda_\varepsilon(1-\lambda+\lambda_\varepsilon^2) + (\lambda_\varepsilon + k)(1-\lambda-\lambda_\varepsilon^2)}{(1-\lambda+\lambda_\varepsilon^2)^2} - \frac{1}{2} + \frac{\lambda_\varepsilon^2+\lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)}$$

which we can rewrite as

$$\frac{\partial \sigma(\varepsilon,k)}{\partial \varepsilon} = \left\{ \frac{\lambda_\varepsilon k - (1-\lambda)}{2(1-\lambda+\lambda_\varepsilon^2)} + (1+\lambda_\varepsilon)\left[\frac{k(1-\lambda-\lambda_\varepsilon^2) + 2\lambda_\varepsilon(1-\lambda)}{2(1-\lambda+\lambda_\varepsilon^2)^2} \right] \right\}$$

so that $\frac{\partial \sigma(\varepsilon,k)}{\partial \varepsilon} \geq 0$ if and only if

$$(1+\lambda_\varepsilon)\left[k\left(1-\lambda-\lambda_\varepsilon^2\right) + 2\lambda_\varepsilon(1-\lambda)\right] - (1-\lambda-\lambda_\varepsilon k)\left(1-\lambda+\lambda_\varepsilon^2\right) \geq 0$$

which can be simplified to:

$$k\left[\left(1-\lambda-\lambda_\varepsilon^2\right) + 2\lambda_\varepsilon(1-\lambda)\right] + \left[2\lambda_\varepsilon + \lambda_\varepsilon^2 - (1-\lambda)\right](1-\lambda) \geq 0$$

Now notice that $[2\lambda_\varepsilon + \lambda_\varepsilon^2 - (1-\lambda)](1-\lambda) > 0$ as $\lambda_\varepsilon > (1-\lambda)$. Then, for small values of $\varepsilon \in (0,\varepsilon_2)$ with $\varepsilon_2 = -(1-\lambda) + \sqrt{(1-\lambda)}$ this inequality is always satisfied as $\varepsilon \in (0,\varepsilon_2)$ implies $(1-\lambda-\lambda_\varepsilon^2) \geq 0$. For large values of ε, however, $(1-\lambda-\lambda_\varepsilon^2) < 0$. Suppose then that in this case $[(1-\lambda-\lambda_\varepsilon^2) + 2\lambda_\varepsilon(1-\lambda)] < 0$. Then, because $k \leq \bar{k}$, a sufficient condition for $\frac{\partial \sigma(\varepsilon,k)}{\partial \varepsilon} \geq 0$ is

$$\bar{k}\left[\left(1-\lambda-\lambda_\varepsilon^2\right) + 2\lambda_\varepsilon(1-\lambda)\right] + \left[2\lambda_\varepsilon + \lambda_\varepsilon^2 - (1-\lambda)\right](1-\lambda) \geq 0$$

which can be simplified to

$$(1-\lambda)\left[1-\lambda+\lambda_{\varepsilon}^{2}+\lambda_{\varepsilon}(1-\lambda)+\lambda_{\varepsilon}^{3}\right] \quad \geq \quad 0$$

which is always satisfied. Thus we can conclude that $\frac{\partial\sigma(\varepsilon,k)}{\partial\varepsilon} \geq 0$ for all $\lambda \in (0,1)$ and $\varepsilon \in (0,\lambda)$. This proves Lemma 1.

Then the last term in (50) is

$$\begin{aligned}
\sigma\left(\varepsilon,\bar{k}\right)\frac{f\left(\bar{k};\rho\right)}{F\left(\bar{k};\rho\right)}\frac{\partial\bar{k}}{\partial\varepsilon} &= \left[(1+\lambda_{\varepsilon})\frac{(1-\lambda)-\lambda_{\varepsilon}k}{2\left(1-\lambda+\lambda_{\varepsilon}^{2}\right)}-1\right]\frac{f\left(\bar{k};\rho\right)}{F\left(\bar{k};\rho\right)}\frac{\bar{k}}{\lambda_{\varepsilon}} \\
&= -\sigma\left(\varepsilon,\bar{k}\right)\frac{f\left(\bar{k};\rho\right)}{F\left(\bar{k};\rho\right)}\frac{\bar{k}}{\lambda_{\varepsilon}}
\end{aligned}$$

The second term under the integral in (50) is

$$\begin{aligned}
\sigma(\varepsilon,k)\frac{\partial}{\partial\varepsilon}\left(\frac{f(k;\rho)}{F\left(\bar{k};\rho\right)}\right) &= \frac{\sigma(\varepsilon,k)}{F\left(\bar{k};\rho\right)^{2}}\Big\{\left[f_{2}(k)-f_{1}(k)\right]F\left(\bar{k};\rho\right)\frac{\partial\bar{\rho}}{\partial\varepsilon} \\
&\quad -\left[F_{2}\left(\bar{k}\right)-F_{1}\left(\bar{k}\right)\right]f(k;\rho)\frac{\partial\bar{\rho}}{\partial\varepsilon}-f\left(\bar{k};\rho\right)f(k;\rho)\frac{\partial\bar{k}}{\partial\varepsilon}\Big\} \\
&= \sigma(\varepsilon,k)\frac{\partial\bar{\rho}}{\partial\varepsilon}\frac{\left[f_{2}(k)F_{1}\left(\bar{k}\right)-f_{1}(k)F_{2}\left(\bar{k}\right)\right]}{F\left(\bar{k};\rho\right)^{2}} \\
&\quad +\sigma(\varepsilon,k)\frac{\bar{k}}{\lambda_{\varepsilon}}\frac{f\left(\bar{k};\rho\right)f(k;\rho)}{F\left(\bar{k};\rho\right)^{2}}
\end{aligned}$$

so that

$$\begin{aligned}
\int_{0}^{\bar{k}}\sigma(\varepsilon,k)\frac{\partial}{\partial\varepsilon}\left(\frac{f(k;\rho)}{F\left(\bar{k};\rho\right)}\right)dk &= \int_{0}^{\bar{k}}\left\{\sigma(\varepsilon,k)\frac{\partial\bar{\rho}}{\partial\varepsilon}\frac{\left[f_{2}(k)F_{1}\left(\bar{k}\right)-f_{1}(k)F_{2}\left(\bar{k}\right)\right]}{F\left(\bar{k};\rho\right)^{2}}\right\}dk \\
&\quad +\int_{0}^{\bar{k}}\left[\sigma(\varepsilon,k)\frac{\bar{k}}{\lambda_{\varepsilon}}\frac{f\left(\bar{k};\rho\right)f(k;\rho)}{F\left(\bar{k};\rho\right)^{2}}\right]dk
\end{aligned}$$

which can be rearranged as

$$\int_{0}^{\bar{k}}\sigma(\varepsilon,k)\frac{\partial}{\partial\varepsilon}\left(\frac{f(k;\rho)}{F\left(\bar{k};\rho\right)}\right)dk = \frac{F_{1}\left(\bar{k}\right)F_{2}\left(\bar{k}\right)}{F\left(\bar{k};\rho\right)^{2}}\left\{\int_{0}^{\bar{k}}\left[\sigma(\varepsilon,k)\frac{\partial\bar{\rho}}{\partial\varepsilon}\right]\frac{dF_{2}(k)}{F_{2}\left(\bar{k}\right)}-\int_{0}^{\bar{k}}\left[\sigma(\varepsilon,k)\frac{\partial\bar{\rho}}{\partial\varepsilon}\right]\frac{dF_{1}(k)}{F_{1}\left(\bar{k}\right)}\right\}$$

$$+\int_0^{\bar{k}} \left[\sigma(\varepsilon,k)\frac{\bar{k}}{\lambda_\varepsilon}\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\right]\frac{dF(k)}{F(\bar{k};\rho)}$$

Then overall we can rearrange the derivative of the average bid ask spread as

$$\begin{aligned}\frac{\partial\sigma(\varepsilon)}{\partial\varepsilon} &= \int_0^{\bar{k}}\left\{\frac{\partial\sigma(\varepsilon,k)}{\partial\varepsilon}\frac{f(k;\rho)}{F(\bar{k};\rho)}\right\}dk \\ &+\frac{F_1(\bar{k})F_2(\bar{k})}{F(\bar{k};\rho)^2}\left\{\int_0^{\bar{k}}\left[\sigma(\varepsilon,k)\frac{\partial\bar{\rho}}{\partial\varepsilon}\right]\frac{dF_2(k)}{F_2(\bar{k})}-\int_0^{\bar{k}}\left[\sigma(\varepsilon,k)\frac{\partial\bar{\rho}}{\partial\varepsilon}\right]\frac{dF_1(k)}{F_1(\bar{k})}\right\} \\ &+\int_0^{\bar{k}}\left[\sigma(\varepsilon,k)\frac{\bar{k}}{\lambda_\varepsilon}\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\right]\frac{dF(k)}{F(\bar{k};\rho)}-\sigma(\varepsilon,\bar{k})\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\frac{\bar{k}}{\lambda_\varepsilon}\end{aligned}$$

where $\sigma(\varepsilon,\bar{k})=1$ and $\frac{\partial\bar{\rho}}{\partial\varepsilon}$ is independent of k, so that:

$$\begin{aligned}\frac{\partial\sigma(\varepsilon)}{\partial\varepsilon} &= \int_0^{\bar{k}}\left\{\frac{\partial\sigma(\varepsilon,k)}{\partial\varepsilon}\frac{f(k;\rho)}{F(\bar{k};\rho)}\right\}dk+\frac{F_1(\bar{k})F_2(\bar{k})}{F(\bar{k};\rho)^2}\frac{\partial\bar{\rho}}{\partial\varepsilon}\left\{\int_0^{\bar{k}}\sigma(\varepsilon,k)\frac{dF_2(k)}{F_2(\bar{k})}-\int_0^{\bar{k}}\sigma(\varepsilon,k)\frac{dF_1(k)}{F_1(\bar{k})}\right\} \\ &+\int_0^{\bar{k}}\left[\sigma(\varepsilon,k)\frac{\bar{k}}{\lambda_\varepsilon}\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\right]\frac{dF(k)}{F(\bar{k};\rho)}-\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\frac{\bar{k}}{\lambda_\varepsilon}\end{aligned}$$

and finally:

$$\begin{aligned}\frac{\partial\sigma(\varepsilon)}{\partial\varepsilon} &= \int_0^{\bar{k}}\left\{\frac{\partial\sigma(\varepsilon,k)}{\partial\varepsilon}\frac{f(k;\rho)}{F(\bar{k};\rho)}\right\}dk+\frac{F_1(\bar{k})F_2(\bar{k})}{F(\bar{k};\rho)^2}\frac{\partial\bar{\rho}}{\partial\varepsilon}\left\{\int_0^{\bar{k}}\sigma(\varepsilon,k)\frac{dF_2(k)}{F_2(\bar{k})}-\int_0^{\bar{k}}\sigma(\varepsilon,k)\frac{dF_1(k)}{F_1(\bar{k})}\right\} \\ &+\frac{\bar{k}}{\lambda_\varepsilon}\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\left[\sigma(\varepsilon)-1\right]\end{aligned}$$

This is the same as (32). Notice that because $\sigma(\varepsilon,k)$ is increasing in k (since the ask price is) and because we assumed F_1 CSD F_2 then: $\int_0^{\bar{k}}\sigma(\varepsilon,k)\frac{dF_2(k)}{F_2(\bar{k})}\leq\int_0^{\bar{k}}\sigma(\varepsilon,k)\frac{dF_1(k)}{F_1(\bar{k})}$ implying that the second term in the above equation, working through the equilibrium choice of ρ by dealers, is always weakly negative. The last term in the above equation is always negative, as $\sigma(\varepsilon)\in(0,1)$, while the first term is always positive by Lemma 1. These results prove Proposition 5.

Consider now the case of the mass-at-zero distribution, and rearrange $\frac{\partial\sigma(\varepsilon)}{\partial\varepsilon}$ using $\frac{\partial\left(\frac{f(k;\rho)}{F(k;\rho)}\right)}{\partial\rho}=$

$\frac{(f_2(k)-f_1(k))F(\bar{k};\rho)-f(k;\rho)(F_2(\bar{k})-F_1(\bar{k}))}{F(\bar{k};\rho)^2}$, as follows:

$$\begin{aligned}
\sigma'(\varepsilon) &= \underbrace{\int_0^{\bar{k}} \frac{\partial\sigma(k,\varepsilon)}{\partial\varepsilon}\frac{f(k;\rho)}{F(\bar{k};\rho)}dk}_{>0} - [1-\sigma(\varepsilon)]\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\frac{\bar{k}}{\lambda_\varepsilon} \\
&\quad -\rho'(\varepsilon)\int_0^{\bar{k}} \sigma(k,\varepsilon)\frac{f(k;\rho)}{F(\bar{k};\rho)}\left[\frac{(f_2(k)-f_1(k))}{f(k;\rho)} - \frac{(F_2(\bar{k})-F_1(\bar{k}))}{F(\bar{k};\rho)}\right]dk
\end{aligned}$$

So that:

$$\begin{aligned}
\sigma'(\varepsilon) &= \underbrace{\int_0^{\bar{k}} \frac{\partial\sigma(k,\varepsilon)}{\partial\varepsilon}\frac{f(k;\rho)}{F(\bar{k};\rho)}dk}_{>0} - [1-\sigma(\varepsilon)]\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)}\frac{\bar{k}}{\lambda_\varepsilon} \\
&\quad -\bar{\rho}'(\varepsilon)\sigma(0,\varepsilon)\frac{(f_2(0)-f_1(0))}{F(\bar{k};\rho)}\left(1-\frac{f(0;\rho)}{F(\bar{k};\rho)}\right)
\end{aligned}$$

where $\bar{\rho}$ is defined by

$$\pi(0)\left(f_2(0)-f_1(0)\right) = \gamma'(\bar{\rho})\left[\bar{\rho}F_2(\bar{k}) + (1-\bar{\rho}))F_1(\bar{\rho})\right]$$

so $\bar{\rho}'(\varepsilon)$ is

$$\underbrace{\frac{\partial\pi(0)}{\partial\varepsilon}}_{<0}\frac{(f_2(0)-f_1(0))}{\gamma''(\bar{\rho})F(\bar{k};\bar{\rho}) + \gamma'(\bar{\rho})[F_2(\bar{k})-F_1(\bar{k})]} = \bar{\rho}'(\varepsilon) < 0$$

So $\sigma'(\varepsilon) > 0$ whenever

$$\sigma(\varepsilon)\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)} \quad > \quad f(0;\rho)\,1 \quad and \quad < F(\bar{k};\rho)$$

Since $\sigma(k,\varepsilon) \geq 0$ and increasing in k for all k,

$$\begin{aligned}
\sigma(\varepsilon) &= \int_0^{\bar{k}} \left\{1-(1+\lambda_\varepsilon)\frac{(1-\lambda)-\lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)}\right\}\frac{f(k;\rho)}{F(\bar{k};\rho)}dk \\
&\geq \int_0^{\bar{k}} \left\{1-(1+\lambda_\varepsilon)\frac{(1-\lambda)-\lambda_\varepsilon k}{2(1-\lambda+\lambda_\varepsilon^2)}\right\}\frac{f_2(k)}{F_2(\bar{k})}dk \\
&\geq 1-\frac{(1+\lambda_\varepsilon)(1-\lambda)}{2(1-\lambda+\lambda_\varepsilon^2)} + \frac{(1+\lambda_\varepsilon)\lambda_\varepsilon}{2(1-\lambda+\lambda_\varepsilon^2)}\int_0^{\bar{k}} k\frac{f_2(k)}{F_2(\bar{k})}dk \\
&\geq \frac{(1-\lambda)(1-\lambda_\varepsilon)+2\lambda_\varepsilon^2}{2(1-\lambda+\lambda_\varepsilon^2)} + \frac{(1+\lambda_\varepsilon)\lambda_\varepsilon}{2(1-\lambda+\lambda_\varepsilon^2)}\int_0^{\bar{k}} k\frac{f_2(k)}{F_2(\bar{k})}dk
\end{aligned}$$

and since F_1 FOSD F_2 , because CSD implies FOSD, we have

$$\frac{f_1(\bar{k})}{F_1(\bar{k})} \geq \frac{f_2(\bar{k})}{F_2(\bar{k})}$$

so that

$$\frac{f(\bar{k};\rho)}{F(\bar{k};\rho)} \geq \frac{f_2(\bar{k})}{F_2(\bar{k})}$$

Hence sufficient conditions for $\sigma'(\varepsilon) > 0$ are

$$\begin{aligned} \left[\frac{(1-\lambda)(1-\lambda_\varepsilon)+2\lambda_\varepsilon^2}{2(1-\lambda+\lambda_\varepsilon^2)} + \frac{(1+\lambda_\varepsilon)\lambda_\varepsilon}{2(1-\lambda+\lambda_\varepsilon^2)}\int_0^{\bar{k}} k\frac{f_2\left(k\right)}{F_2(\bar{k})}dk\right]\frac{f_2(\bar{k})}{F_2(\bar{k})} &> 1 \\ F_1(\bar{k};\rho) &> f_2\left(0\right) \end{aligned}$$

A.6 Extension: Derivation of reservation prices, minimum ask price and maximum bid price.

In order to solve the dealer's decision problem and characterize the equilibrium prices distribution, we follow Spulber [1996] and consider a relaxed problem where dealers are constrained to meeting buyers' demand in the state of the world where demand is high at settlement, only in present value:

$$\max_{a,b} a(1-\lambda)D(a) - (b+k)S(b) \tag{52}$$

$$\text{s.t. } \lambda_\varepsilon D(a) \leq S(b) \tag{53}$$

We characterize the solution to the relaxed problem where the sequence of constraints $\forall t, s$ are replaced by a single less restrictive constraint, and then show that the solution to the relaxed problem is in the constraint set of the original problem. Any solution to this relaxed problem satisfies the constraint with equality. Otherwise, a dealer would be able to lower b, attract less sellers, but make a higher profit. Hence, using (1) and (2),we must have

$$b = \lambda_\varepsilon(\overline{r_c} - a) + \underline{r_p} \tag{54}$$

We can now use the constraint to substitute for $S(b)$ in the dealers' objective function to rewrite it as:

$$D(a)\{a\left(1-\lambda+\lambda_\varepsilon^2\right) - \left(\lambda_\varepsilon\overline{r_c} + \underline{r_p} + k\right)\lambda_\varepsilon\}$$

Therefore, dealers choose a to solve the first order condition:

$$\frac{\partial D(a)}{\partial a}\left[a\left(1-\lambda+\lambda_{\varepsilon}^{2}\right)-\left(\lambda_{\varepsilon}\overline{r_c}+\underline{r_p}+k\right)\lambda_{\varepsilon}\right]+D(a)\left(1-\lambda+\lambda_{\varepsilon}^{2}\right)=0$$

and using (1) this yields

$$\begin{aligned} a\left(k\right) &= \frac{\overline{r_c}\left(1-\lambda+2\lambda_{\varepsilon}^{2}\right)+(\underline{r_p}+k)\lambda_{\varepsilon}}{2\left(1-\lambda+\lambda_{\varepsilon}^{2}\right)}, &(55)\\ b\left(k\right) &= \frac{\overline{r_c}\left(1-\lambda\right)\lambda_{\varepsilon}+\underline{r_p}\left[2(1-\lambda)+\lambda_{\varepsilon}^{2}\right]-\lambda_{\varepsilon}^{2}k}{2\left(1-\lambda+\lambda_{\varepsilon}^{2}\right)}. &(56)\end{aligned}$$

The linearity of $a(k)$ and $b(k)$, together with k being distributed according to $F(\cdot)$ implies that G_a, G_b will depend on $F(\cdot)$ in a similar way as in the model without option value that we analyzed above. To be able to say more, we will now assume that F is the uniform distribution over $[0,1]$. Then G_a, G_b will also be uniform distributions. In order to fully characterize the equilibrium we need to pin down the extremes of these distributions. We can show that the reservation price of the highest valuation buyer is:

$$\overline{r_c}=\frac{\underline{a}+2\delta}{1+2\delta}$$

The lowest ask-price, $\underline{a}$, is set by the most efficient dealer, so that $\underline{a}=a(0)$ and using (55),

$$\underline{a}=\frac{\overline{r_c}\left(1-\lambda+2\lambda_{\epsilon}^{2}\right)+\underline{r_p}\lambda_{\epsilon}}{2(1-\lambda+\lambda_{\epsilon}^{2})}$$

So

$$\overline{r_c}[1-\lambda+4\delta(1-\lambda+\lambda_{\epsilon}^{2})] \quad = \quad \underline{r_p}\lambda_{\epsilon}+4\delta(1-\lambda+\lambda_{\epsilon}^{2})$$

Similarly, the reservation price of the lowest valuation seller is:

$$\underline{r_p}=\frac{\bar{b}}{1+2\delta}$$

And we know that the most efficient dealer will set the highest bid price, $\bar{b}=b(0)$. So using (**??**)

$$\bar{b}=\frac{\overline{r_c}\left(1-\lambda\right)\lambda_{\varepsilon}+\underline{r_p}\left[2(1-\lambda)+\lambda_{\varepsilon}^{2}\right]}{2\left(1-\lambda+\lambda_{\varepsilon}^{2}\right)} \tag{57}$$

Therefore

$$\underline{r_p} \;=\; \lambda_\varepsilon \frac{\overline{r_c}(1-\lambda)}{\lambda_\varepsilon^2 + 4\delta(1-\lambda+\lambda_\varepsilon^2)}$$

And using this expression for $\underline{r}_p$ back into the equation for $\overline{r_c}$ yields:

$$\overline{r_c} = 1 - \frac{1-\lambda-\lambda_\varepsilon \underline{k}}{(1+4\delta)\,(1-\lambda+\lambda_\varepsilon^2)} \tag{58}$$

which we can replace in the expression for $\underline{r_p}$ to obtain

$$\underline{r_p} = \lambda_\varepsilon \frac{1-\lambda-\lambda_\varepsilon \underline{k}}{(1+4\delta)\,(1-\lambda+\lambda_\varepsilon^2)}$$

Notice that

$$\underline{r_p} = \lambda_\varepsilon\,(1-\overline{r_c})$$

and that $\frac{\partial \overline{r_c}}{\partial \lambda} > 0, \frac{\partial \underline{r_p}}{\partial \lambda} < 0$ as $\frac{\partial \delta}{\partial \lambda} > 0$.

We now solve for a and b as a function of k. Using (55) we obtain

$$a(k) \;=\; \frac{2\lambda_\varepsilon^2 + 4\delta\,(1-\lambda+2\lambda_\varepsilon^2) + (1+4\delta)\,\lambda_\varepsilon k}{2\,(1+4\delta)\,(1-\lambda+\lambda_\varepsilon^2)} \tag{59}$$

$$b(k) \;=\; \lambda_\varepsilon \frac{2\,(1+2\delta)\,(1-\lambda) - \lambda_\varepsilon k\,(1+4\delta)}{2\,(1+4\delta)\,(1-\lambda+\lambda_\varepsilon^2)} \tag{60}$$

We can show

Proposition 6. *The ask price $a(k)$ is increasing in ε for all k.*
If $\varepsilon > \sqrt{1-\lambda}\left(1-\sqrt{1-\lambda}\right)$ then the bid price $b(k)$ is always decreasing in ε for all dealers k. Otherwise, if $\varepsilon \leq \sqrt{1-\lambda}\left(1-\sqrt{1-\lambda}\right)$ then the bid price $b(k)$ is increasing in ε if an only if $k < k_b^*$.
If $\varepsilon^2 < (1-\lambda)(2-\lambda)$ (which is always satisfied if $\lambda < \frac{2}{3}$) then the bid-ask spread $a(k) - b(k)$ is increasing in ε for all k.

The condition $\lambda < 2/3$ is not very strong, as typically λ will be small. As we shall see below, a rough calibration of the model gives us $\lambda = 0.25\%$, well below this threshold. Furthermore we can show

Proposition 7. *The derivative of the ask price with respect to* ε *is a linear function of* k*, i.e.* $\frac{\partial a(k)}{\partial \varepsilon} = a_0(\varepsilon) + a_1(\varepsilon)k$ *where* $a_1(\varepsilon) = \frac{1-\lambda-\lambda_\varepsilon^2}{2(1-\lambda+\lambda_\varepsilon^2)^2}$ *where* $a_1'(\varepsilon) < 0$.

This result implies that the difference between the marginal impact of an increase in idiosyncratic risk on the ask price for two different dealers indexed by different costs k_1, k_2 is a constant function of $k_2 - k_1$ given ε and only depends on the difference in costs of executing trades. Also, for any dealer with a given k, the equilibrium ask price is less sensitive to an increase in idiosyncratic risk for high levels of such risk.

This result also implies that less efficient dealers will increase their ask price by more if and only if $1 - \lambda > \lambda_\varepsilon^2$. We also have different comparative statics results for the supply and demand of the asset with respect to risk for the most and the least efficient dealers. Taking the derivative with respect to λ_ε we obtain

$$\frac{\partial S(k)}{\partial \lambda_\varepsilon} = \frac{\lambda_\varepsilon \left(8(1-\lambda)^2 - k(\frac{1}{\rho(1-\lambda)\delta} + 3)\lambda_\varepsilon \left(3 - 3\lambda + \lambda_\varepsilon^2\right)\right)}{8(1-\lambda)(1-\rho)\left(1-\lambda+\lambda_\varepsilon^2\right)^2}$$

Then

$$\frac{\partial S(0)}{\partial \lambda_\varepsilon} = \frac{\lambda_\varepsilon(1-\lambda)}{(1-\rho)\left(1-\lambda+\lambda_\varepsilon^2\right)^2} > 0$$

and

$$\frac{\partial S(\bar{k})}{\partial \lambda_\varepsilon} = \frac{-\lambda_\epsilon}{2(1-\rho)\left(1-\lambda+\lambda_\epsilon^2\right)} < 0$$

Taking the derivative of $D(k)$ with respect to λ_ε we obtain:

$$\frac{\partial D(k)}{\partial \lambda_\varepsilon} = \frac{2(1-\lambda) - 2\lambda_\epsilon^2 - k(\frac{1}{\rho(1-\lambda)\delta} + 3)\lambda_\epsilon}{4(1-\rho)\left(1-\lambda+\lambda_\epsilon^2\right)^2}$$

Then

$$\frac{\partial D(0)}{\partial \lambda_\varepsilon} = \frac{1-\lambda-\lambda_\varepsilon^2}{2(1-\rho)\left(1-\lambda+\lambda_\varepsilon^2\right)^2}$$

so that $\partial D(0)/\partial \lambda_\varepsilon > 0$ for ε small and $\partial D(0)/\partial \lambda_\varepsilon < 0$ for high ε.

$$\frac{\partial D(\bar{k})}{\partial \lambda_\varepsilon} = \frac{-1}{2(1-\rho)\left(1-\lambda+\lambda_\epsilon^2\right)} < 0$$

Taking the number of dealers M as given and taking the derivative of $D(k)$ with respect to ε we obtain at $k = 0$,

$$\frac{\partial D(0)|_M}{\partial \lambda_\varepsilon} = -\frac{4(1-\lambda)(1-(1-\lambda)\rho)\lambda_\varepsilon}{M(1-\rho)(4-3(1-\lambda)\rho)\left(1-\lambda+\lambda_\varepsilon^2\right)^2} < 0$$

Proposition 8. *The reservation price of buyers (sellers) is lower (higher) in the model with the option to search.*

Proof. Since the reservation price of buyers is implicitly defined by

$$v = r_c(v) + \frac{1}{\delta}\int_{\underline{a}}^{r_c(v)} G_a(a)da$$

and since $G_a(a) \geq 0$ for all a, then $r_c(v) < v$. Similarly, for sellers, it follows from comparing (**??**) with the reservatio price of sellers in the benchmark model, which is simply a seller's valuation, v. Since the reservation price of sellers is defined by

$$v = r_p(v) - \frac{1}{\delta}\int_{r_p(v)}^{\bar{b}} \left[1 - G_b(b)\right] db$$

and since $[1 - G_b(b)] \geq 0$ then for all b, then $r_p(v) > v$. □

This result is intuitive because the option to search makes buyers and sellers more wary of accepting current ask and bid prices posted by dealers.

Proposition 9. *The measure of active dealers is lower in the model with the option to search.*

Proof. From the definition of $\overline{k}$ (defined as $a(\overline{k}) = 1$) we have

$$\overline{k} = \frac{4\delta}{(1+4\delta)}\frac{\lambda_\varepsilon - \varepsilon}{\lambda_\varepsilon} = \frac{4\delta}{(1+4\delta)}\frac{1-\lambda}{\lambda_\varepsilon} = \frac{4\delta}{(1+4\delta)}\overline{k}_B \quad (61)$$

where $\overline{k}_B = \frac{1-\lambda}{\lambda_\varepsilon}$ denotes the equilibrium measure of active dealers in the benchmark model. □

This result is intuitive as the option to search reduces the reservation price of end customers. Therefore the marginal active dealer in the benchmark model of section 4 now makes negative profits because the buyer and seller whom he can serve are no longer willing to pay the ask and bid prices that he can offer. The following lemma describes into further details the relationship between bid and ask prices in the two models.

Proposition 10. *The ask price is lower and the bid price higher in the model with the option to search. Bid ask spread are narrower in the model with the option to search. Let* $\Gamma = \frac{1-\lambda}{2(1+4\delta)(1-\lambda+\lambda_\varepsilon^2)}$, *and let $a(k), b(k), a_B(k), b_B(k)$ denote the ask and bid prices in the model with the option to search and in the benchmark model of section 4 respectively. Then:*

$$\begin{aligned} a(k) &= a_B(k) - \Gamma \\ b(k) &= b_B(k) + \lambda_\varepsilon \Gamma \\ a(k) - b(k) &= a_B(k) - b_B(k) - \Gamma\left(1 + \lambda_\varepsilon\right) \end{aligned}$$

Proof. The equilibrium ask price in the model with the option to search is defined by (**??**), which we can rearrange as

$$\begin{aligned} a(k) &= \frac{-\left(1-\lambda\right) + \left(1+4\delta\right)\left(1-\lambda+2\lambda_\varepsilon^2+\lambda_\varepsilon k\right)}{2\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} \\ &= \frac{-\left(1-\lambda\right)}{2\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} + a_B\left(k\right) \\ &< a_B\left(k\right) \end{aligned} \tag{62}$$

where $a_B\left(k\right)$ denotes the ask price in the benchmark model, defined by (12). Similarly for the bid price,

$$\begin{aligned} b(k) &= \lambda_\varepsilon \frac{\left(2+4\delta\right)\left(1-\lambda\right) - \lambda_\varepsilon k\left(1+4\delta\right)}{2\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} \\ &= \lambda_\varepsilon \left[\frac{1-\lambda+\left(1+4\delta\right)\left(1-\lambda-\lambda_\varepsilon k\right)}{2\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)}\right] \\ &= \lambda_\varepsilon \left[\frac{1-\lambda}{2\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} + b_B\left(k\right)\right] \\ &> b_B\left(k\right) \end{aligned} \tag{63}$$

where $b_B\left(k\right)$ denotes the bid price in the benchmark model, defined by (13). The result on the bid ask spread follows from combining (62) with (63). □

Again these results are intuitive because in the model with the option to search there are two effects going on: 1) buyers' reservation value is lower and sellers' reservation value is higher; 2) dealers are competing with their future selves and with other dealers with whom end customers might match. As a consequence, bid-ask spreads have to be narrower. Whether the order flow (or volume of trades), however, is larger or smaller than in the benchmark model is not a priori clear, as the reservation values of buyers and sellers also change with the option to search. The following lemma characterizes the change in order flow across the two models for the dealers who are active in both models. [32]

Proposition 11. *There exists $\tilde{k} \in \left(0, \bar{k}\right)$ such that the order flow in the model with option to search is larger than in the benchmark model if and only if $k \in \left(0, \tilde{k}\right)$.*

Proof. Consider the demand of assets in the model with the option to search defined and substitute $\overline{r_c} = 1 - \frac{1-\lambda}{(1+4\delta)(1-\lambda+\lambda_\varepsilon^2)}$, using $\underline{k} = 0$, and $N = \bar{k} = \frac{4\delta}{(1+4\delta)}\overline{k}_B$:

$$\begin{aligned} D(a) &= \frac{(\overline{r_c} - a)}{N(1-\beta)} \\ &= \frac{(1 - \frac{1-\lambda}{(1+4\delta)(1-\lambda+\lambda_\varepsilon^2)} - a\left(k\right))}{\frac{4\delta}{(1+4\delta)}\overline{k}_B(1-\beta)} \end{aligned}$$

where, recall that $a\left(k\right) = a_B(k) - \Gamma$, with $\Gamma = \frac{1-\lambda}{2(1+4\delta)(1-\lambda+\lambda_\varepsilon^2)}$. Analogously, consider the demand of assets in the benchmark model

$$D\left(a\right) = \frac{1 - a_B\left(k\right)}{\bar{k}_B}$$

Then the demand of assets in the model with the option to search is larger than that in the benchmark model if and only if

$$\begin{aligned} \frac{1 - \frac{1-\lambda}{(1+4\delta)(1-\lambda+\lambda_\varepsilon^2)} - a_B(k) + \Gamma}{\frac{4\delta}{(1+4\delta)}\overline{k}_B(1-\beta)} &\geq \frac{1 - a_B\left(k\right)}{\bar{k}_B} \\ 1 - a_B(k) - \frac{1-\lambda}{\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} + \Gamma &\geq \frac{4\delta}{\left(1+4\delta\right)}\left(1-\beta\right)\left(1-a_B\left(k\right)\right) \\ \left(1-a_B\left(k\right)\right)\left[1 - \frac{4\delta}{\left(1+4\delta\right)}(1-\beta)\right] &\geq \frac{1-\lambda}{\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} - \Gamma \\ \left(1-a_B\left(k\right)\right)\left[1 - \frac{4\delta}{\left(1+4\delta\right)}(1-\beta)\right] &\geq \Gamma \end{aligned}$$

Now, using (12), we have that $1 - a_B(k) = \frac{1-\lambda-k\lambda_\varepsilon}{2(1-\lambda+\lambda_\varepsilon^2)}$, and substituting out $\Gamma = \frac{1-\lambda}{2(1+4\delta)(1-\lambda+\lambda_\varepsilon^2)}$ yields

$$\begin{aligned} \frac{1-\lambda-k\lambda_\varepsilon}{2\left(1-\lambda+\lambda_\varepsilon^2\right)}\left[1 - \frac{4\delta}{\left(1+4\delta\right)}(1-\beta)\right] &\geq \frac{1-\lambda}{2\left(1+4\delta\right)\left(1-\lambda+\lambda_\varepsilon^2\right)} \\ \left(1-\lambda-k\lambda_\varepsilon\right)\left[1 - \frac{4\delta}{\left(1+4\delta\right)}(1-\beta)\right] &\geq \frac{1-\lambda}{\left(1+4\delta\right)} \end{aligned}$$

[32] Recall that the marginal active dealer in the model with the option to search is indexed by a lower k than in the bechmark model. We restrict the comparison of order flow across dealers $k \in \left[0, \overline{k}\right]$ with $\bar{k}$ defined in (61).

$$
\begin{aligned}
(1-\lambda-k\lambda_{\varepsilon})\left[(1+4\delta)-4\delta(1-\beta)\right] &\geq 1-\lambda \\
(1-\lambda-k\lambda_{\varepsilon})\left[1+4\delta\beta\right] &\geq 1-\lambda
\end{aligned}
$$

Substituting now $\delta=\frac{1}{\beta(1-\lambda)}-1$, yields

$$
\begin{aligned}
(1-\lambda-k\lambda_{\varepsilon})\left[1+4\beta\left(\frac{1}{\beta\,(1-\lambda)}-1\right)\right] &\geq 1-\lambda \\
(1-\lambda-k\lambda_{\varepsilon})\left[1+4\left(\frac{1}{(1-\lambda)}-\beta\right)\right] &\geq 1-\lambda \\
(1-\lambda)\,4\left(\frac{1}{(1-\lambda)}-\beta\right) &\geq k\lambda_{\varepsilon}\left[1+4\left(\frac{1}{(1-\lambda)}-\beta\right)\right]
\end{aligned}
$$

This inequality is satisfied at $k=0$ as $\frac{1}{(1-\lambda)}>\beta$. Notice that the right hand side increases with k, and evaluating the above inequality at $k=\overline{k}$ yields:

$$
\begin{aligned}
(1-\lambda)\,4\left(\frac{1}{(1-\lambda)}-\beta\right) &\geq \left[\frac{4\delta}{(1+4\delta)}\frac{1-\lambda}{\lambda_{\varepsilon}}\right]\lambda_{\varepsilon}\left[1+4\left(\frac{1}{(1-\lambda)}-\beta\right)\right] \\
(1-\lambda)\,4\left(\frac{1}{(1-\lambda)}-\beta\right) &\geq \frac{4\delta}{(1+4\delta)}(1-\lambda)\left[1+4\left(\frac{1}{(1-\lambda)}-\beta\right)\right] \\
\left(\frac{1}{(1-\lambda)}-\beta\right) &\geq \frac{\delta}{(1+4\delta)}+\frac{4\delta}{(1+4\delta)}\left(\frac{1}{(1-\lambda)}-\beta\right) \\
\left(\frac{1}{(1-\lambda)}-\beta\right)\left(1-\frac{4\delta}{(1+4\delta)}\right) &\geq \frac{\delta}{(1+4\delta)} \\
\left(\frac{1}{(1-\lambda)}-\beta\right)\frac{1}{(1+4\delta)} &\geq \delta
\end{aligned}
$$

Recall that $\delta=\frac{1}{\beta(1-\lambda)}-1$, so that the last inequality can be rearranged as

$$
\begin{aligned}
\frac{\beta}{(1+4\delta)} &\geq 1 \\
\beta &\geq 1+4\delta \\
\beta &\geq -3+\frac{4}{\beta\,(1-\lambda)} \\
\beta^2+3\beta-\frac{4}{(1-\lambda)} &\geq 0
\end{aligned}
$$

which is never satisfied as $\beta\in(0,1)$ and $\lambda\in(0,1)$. Hence, demand at the most efficient dealer is larger in the model with the option to search, while demand at the least efficient dealer that is active in both economies (that is $\bar{k}<\bar{k}_B$) is larger in the benchmark model. The intermediate value theorem implies that there exist $\tilde{k} \in \left(0, \bar{k}\right)$ such that $D\left(a\left(k\right)\right) \geq D_B\left(a_B\left(k\right)\right)$ for all $k \in \left[0, \tilde{k}\right]$ and $D\left(a\left(k\right)\right) < D_B\left(a_B\left(k\right)\right)$ for all $k \in (\tilde{k}, \bar{k}]$. Because supply mirrors demand via the resource constraint in dealers' decision problem, then the aggregate order flow in the model with the option to search is larger for all $k \in \left[0, \tilde{k}\right]$ and smaller for all $k \in (\tilde{k}, \bar{k}]$. □

Proposition 12. *The marginal profits of the most efficient dealer evaluated at zero risk are larger in the economy with the options to search: For all* $\lambda \in (0, 1)$*:*

$$\frac{\partial \pi(a, b, k)}{\partial \varepsilon} |_{\varepsilon=0,k=0} \quad > \quad \frac{\partial \pi_B(a, b, k)}{\partial \varepsilon} |_{\varepsilon=0,k=0}$$

Proof. Consider marginal profits of the most efficient dealer and compare them with those of the most efficient dealer in the benchmark model, for a uniform distribution. The results in lemma 2 for a uniform distribution, imply

$$\frac{\partial \pi_B(a, b, k)}{\partial \varepsilon} |_{k=0} \quad = \quad \frac{(1-\lambda)\left(1-\lambda-\lambda_\varepsilon^2\right)}{\left[4\left(1-\lambda+\lambda_\varepsilon^2\right)\right]^2}$$

Then the marginal profits of the most efficient dealer, evaluated at zero risk, $\varepsilon = 0$, are larger in the model with the option to search if and only if

$$\begin{aligned}
\frac{\lambda\left(1-(1-\lambda)\beta\right)}{(2-\lambda)^2(1-\beta)\left[4-3(1-\lambda)\beta\right]} &> \frac{(1-\lambda)\left(1-\lambda-(1-\lambda)^2\right)}{\left[4\left(1-\lambda+(1-\lambda)^2\right)\right]^2} \\
\frac{\lambda\left(1-(1-\lambda)\beta\right)}{(2-\lambda)^2(1-\beta)\left[4-3(1-\lambda)\beta\right]} &> \frac{(1-\lambda)^2\lambda}{\left[4(1-\lambda)(2-\lambda)\right]^2} = \frac{\lambda}{\left[4(2-\lambda)\right]^2} \\
\frac{\left(1-(1-\lambda)\beta\right)}{(1-\beta)\left[4-3(1-\lambda)\beta\right]} &> \frac{1}{4^2} \\
16\left(1-(1-\lambda)\beta\right) &> (1-\beta)\left[4-3(1-\lambda)\beta\right]
\end{aligned}$$

where the last equation follows from $[4-3(1-\lambda)\beta] > 0$ always. Then the last equation can be further rearranged as:

$$\begin{aligned}
16 &> 4(1-\beta) + \left[16-3(1-\beta)\right](1-\lambda)\beta \\
\lambda &> 1 - \frac{16-4(1-\beta)}{\beta\left[16-3(1-\beta)\right]} \\
\lambda &> (1-\beta)\frac{-3\beta-12}{\beta\left[16-3(1-\beta)\right]}
\end{aligned}$$

which is always satisfied. □

Previous volumes in this series

All volumes are available on our website www.bis.org.