A Theory of Participation in OTC and Centralized Markets*

Jérôme Dugast† Semih Üslü‡ Pierre-Olivier Weill§

March 15, 2019

Abstract

Should regulators encourage the migration of trade from over-the-counter (OTC) to centralized markets? To address this question, we consider a model of equilibrium and socially optimal market participation of heterogeneous banks in an OTC market, in a centralized market, or in both markets at the same time. We find that banks have the strongest private incentives to participate in the OTC market if they have the lowest risk-sharing needs and highest ability to take large positions. These banks endogenously assume the role of OTC market dealers. Other banks, with relatively higher risk-sharing needs and lower ability to take large positions, lie at the margin: they are indifferent between the centralized market and the OTC market, where they endogenously assume the role of customers. We show that more customer banks participation in the centralized market can be welfare improving only if investors are mostly heterogeneous in their ability to take large positions in OTC market, and if participation costs induce banks to trade exclusively in one market. Empirical evidence suggest that these necessary conditions for a welfare improvement are met.

*An earlier version of this paper was circulated under the title "Platform Trading with an OTC Market Fringe". We would like to thank, for fruitful comments and suggestions, David Cimon, Valentin Haddad, Gregor Jarosch, Peter Kondor, Carlos Ramirez, and Zhaogang Song, as well as seminar participants at UCLA, University of Luxembourg, Johns Hopkins University, University of Colorado Boulder, the Bank of Canada Workshop on Money, Banking, Payments, and Finance, the 2nd LAEF OTC Markets and Securities Workshop at UCSB, the 2018 ASSA Meetings in Philadelphia, the Spring 2018 Midwest Macroeconomics Meetings at UW-Madison, the 2018 Annual Meeting of the Society for Economic Dynamics in Mexico, the CEPR 2018 European Summer Symposium in Financial Markets in Gerzensee, the European Finance Association 2018 Meetings in Warsaw, and the Northern Finance Association 2018 Meetings in Charlevoix.

†Université Paris-Dauphine, Université PSL, email: jerome.dugast@dauphine.psl.eu

‡Johns Hopkins Carey Business School, e-mail: semihuslu@jhu.edu

§University of California, Los Angeles, NBER, and CEPR, e-mail: poweill@econ.ucla.edu

1 Introduction

Over-the-counter (OTC) markets have a decentralized structure: trade is bilateral, involves bargaining, and generates substantial price dispersion. A common policy concern is that the decentralized structure makes OTC markets excessively fragile, for example because liquidity dries up too quickly during financial turmoils. Another common concern is that the price dispersion in OTC markets benefits dealers but hurts customers, who end up paying high prices for low-quality intermediation services. As a result of these concerns, regulators have made proposals and taken measures to increase investors' participation in centralized markets.[1] But it is not obvious that such policies are welfare improving, since trading behavior and participation decisions are endogenous: if investors had not been content with participating as customers in OTC markets, private parties could have successfully offered them to participate in centralized markets.[2] Therefore, to make a case for these policies, one must answer the following question: can it be socially optimal for investors to participate in a centralized market, when they find it privately optimal to participate as customers in an OTC market?

To address this question, we consider an equilibrium model of banks participation in an OTC market, in a centralized market, or in both markets at the same time. Banks are heterogeneous in two dimensions. First, they have heterogenous risk-sharing needs, represented by differences in their initial endowment of risky assets. Second, they are heterogenous in their ability to take large positions in the OTC market, what we call their OTC market trading capacity. Different trading capacities conveniently represent differences in funding constraints, access to collateral pool, risk-management technology, or trading expertise. Banks incur costs to participate in the OTC market, in the centralized market, or in both markets at the same time.

When making their participation decisions, banks face a trade-off between sharing risk and earning intermediation profits. Specifically, while the centralized market allows them to change their risk exposures more easily, the decentralized OTC market creates price dispersion and so allow them to earn intermediation profits. Accordingly we show that, under natural

[1]For example, regulators have mandated that some swaps trade multilaterally on platforms called "swap execution facilities." In 2009, G20 Leaders agreed that "all standardized OTC derivative contracts should be traded on exchanges or electronic trading platforms." And, as of June 2017, "12 jurisdictions have in force comprehensive assessment standards or criteria for determining when products should be platform traded, and an appropriate authority regularly assesses transactions against these criteria" (Financial Stability Board, 2017).

[2]In reality, when centralized and OTC markets co-exist, trading volume often concentrates in the OTC market. For example, Biais and Green (2006) note that "more than 1000 bond issues are still listed on the Exchange" but "the overwhelming majority of trades are conducted over the counter." Riggs, Onur, Reiffen, and Zhu (2018) document that Swap Execution Facilities allow investors to use different execution mechanisms that differ in their degree of centralization. They find that the most centralized mechanism, a limit-order book, attracts very little volume.

participation cost structures, banks with low risk-sharing needs and high trading capacity participate exclusively in the OTC market. Amongst the exclusive OTC banks, those with relatively lower risk-sharing needs and higher trading capacity behave as dealers: they engage in large offsetting trades and so make large intermediation profits. Those with relatively high risk-sharing needs and low trading capacity behave as customers: they engage in small trades, mostly in the same direction, and so make little intermediation profits. Depending on the structure of participation costs, banks with even higher risk-sharing needs and even lower trading capacity participate either exclusively in the centralized market or in both markets at the same time.

A key observation about equilibrium is that customers are the marginal OTC market participants: relative to dealers, they have a weaker preference for the OTC over the centralized market. However, under natural participation cost structures, customers marginally prefer cheap but low-quality risk sharing in the OTC market over expensive but high-quality risk sharing in the centralized market. The finding that customers are marginal OTC participants is in line with the commonly held view that they benefit relatively less from OTC market trading than dealers do. Of course, this does not necessarily imply that increasing their participation in the centralized market would be welfare improving. Increasing centralized market participation is welfare improving only if customers' private incentives to participate in OTC market are larger than their social incentives, i.e., if the presence of the marginal customers in the OTC market imposes negative externality onto other market participants.

We show that increasing the participation of marginal customer banks in the centralized market can be welfare improving only if two conditions are met. First, banks must differ mostly in terms of their OTC trading capacities. Second, participation costs must induce banks to trade exclusively in the centralized market or in the OTC market. To build intuition, notice that when trade is exclusive, an increase in centralized market participation mechanically leads to a decrease in OTC market participation. This creates two effects going in opposite directions. On the one hand, bilateral trades are destroyed because the marginal bank, who now participates only in the centralized market, no longer matches with infra-marginal banks in the OTC market. On the other hand, bilateral trades are created because infra-marginal banks in the OTC market now match more together. We show that, if banks differ mostly in terms of trading capacities, then the welfare gain from the second effect dominates. Indeed, since the marginal bank has a relatively lower trading capacity than infra-marginal banks, the trades destroyed have *smaller size* than the trades created. In contrast, if banks differ mostly in terms of risk-sharing needs, then it is the welfare loss from the first effect that dominates. In that case, the marginal banks have relatively stronger risk-sharing needs than the infra-marginal banks. As a result, the trades destroyed have *larger value* than the trades created.

An implication of these findings is that, in order to evaluate whether mandating or subsidizing trade in a centralized trading venue is welfare improving, it is crucial to empirically distinguish an economy in which banks differ mostly in terms of OTC trading capacity, from an economy in which banks differ mostly in terms of their risk-sharing needs. Our examples suggest the following empirical distinctions. When investors differ mostly in terms of OTC trading capacity, the per-dealer gross trading volume can be much larger than the per-customer gross volume, and the net trading volume of dealers can be large. In contrast, when banks differ mostly in terms of risk-sharing needs, dealers and customers have comparable gross trading volume, but dealers have lower net trading volume. In reality, OTC markets typically feature a core-periphery structure, most trades are intermediated by a handful of dealers, which implies that the per-dealer gross trading volume is larger than the per-customer gross volume, even after controlling for size.[3] In addition, Siriwardane (2018) finds that, in the CDS market, the net notionals are concentrated in the hands of dealers. In particular, he shows that, dealers are responsible for 55% of all net buying and 60% of all net selling on average between 2010 and 2014, which implies that the net trading volume of dealers in the CDS market is large. Hence, viewed through the lens of the model, this empirical evidence suggests that banks differ mostly in terms of their OTC trading capacity.

In the last part of the paper, we further deepen the scope of our results by extending the model to allow for differential resiliency across market structures. More specifically, we introduce a "financial turmoil" event which occurs with some probability. Upon a turmoil, the OTC market shuts down and OTC market participants are prevented from trading while the centralized market shuts down only with some lower probability. We show that, when banks have rational expectations about shutdown risk, all our results go through.

Literature review

This paper is based on the model of Atkeson, Eisfeldt, and Weill (2015, henceforth AEW), who have developed a tractable framework, using insights from both the search- and network-theoretic literature, to study entry and trading patterns in OTC market. We generalize AEW

[3]See Bech and Atalay (2010), Atkeson, Eisfeldt, and Weill (2013), Di Maggio, Kermani, and Song (2017), Hollifield, Neklyudov, and Spatt (2017), and Li and Schürhoff (2019).

in two main dimensions. First, we allow investors to trade in two markets, one OTC and one centralized, instead of OTC only in AEW. This is clearly essential to analyze our main research question. Second, while AEW only considered heterogeneity in risk-sharing needs, investors in our model are heterogeneous in two dimensions: their risk-sharing needs and their OTC market trading capacities. We show that the trading-volume patterns that arise from the heterogeneity in trading capacities fits better the patterns observed in OTC markets in practice. On the normative side, we show that two dimensions of heterogeneity have opposite policy implications.[4]

A branch of the literature compares the costs and benefits associated with centralized and decentralized trading structures without endogenous participation decision. See, for example, Glode and Opp (2019), Geromichalos and Herrenbrueck (2016), Li and Song (2019), Liu, Vogel, and Zhang (2018), and Vogel (2019). Another branch of the literature has studied the trade-off between exclusive participation in a centralized or a decentralized market. See, for example, Yavaş (1992), Gehrig (1993), Rust and Hall (2003), Miao (2006), Lee and Wang (2018), and Yoon (2018). Our contribution relative to these papers is to generate dealer and customer trading patterns endogenously in the OTC market, and relate these patterns to private and social participation incentives. In addition, we also study non-exclusive participation, i.e., the possibility that investors participate simultaneously in two markets. We show that the positive and normative analysis of non-exclusive participation is conceptually different from that of exclusive participation, and sometimes generates opposite normative results.

Praz (2014) studies a dynamic equilibrium asset pricing model with non-exclusive trade. Namely, investors trade two correlated assets in two markets, the first one in a centralized market, and the second one in a decentralized search market. Our contribution relative to his work is to provide a positive and normative analysis of heterogeneous investors' costly participation decisions in these markets.

Many recent papers have studied the manner in which customer and dealer trading patterns emerge endogenously in models of OTC markets, based on alternative assumptions regarding investors' heterogeneity. An non-exhaustive list of papers includes Babus (2009), Neklyudov (2012), Afonso and Lagos (2015), Hugonnier, Lester, and Weill (2014), Babus and Kondor

[4]Our framework is also technically more sophisticated: we now consider general distributions over risk endowments and trading capacities, instead of discrete distributions in AEW. While the economics of the problem remains essentially the same, this creates important technical difficulties, as all optimization and fixed point arguments must be formulated and solved in infinite dimensional vector spaces. This more general mathematical framework clarifies the economic forces at play, and importantly leads to closed-form characterizations of equilibrium for some important cases of interest, illustrated in Section 4.

(2018), Üslü (2015), Farboodi, Jarosch, and Shimer (2015), Farboodi, Jarosch, and Menzio (2016), Chang and Zhang (2016), Wang (2016), and Bethune, Sultanum, and Trachter (2018). Our paper builds on their insights with a different modeling framework to study equilibrium and socially optimal participation in OTC vs. centralized market. The advantage of our static modeling framework is that it allows for a rigorous, transparent, and simple characterization of the composition externalities induced by participation decisions.

Axelson (2007), Rostek and Yoon (2018), and Babus and Hachem (2019) study how the market structure affects the optimal security design problem of asset issuers. A few papers have explored the manner in which market fragmentation may emerge as an equilibrium outcome due to information and price-setting frictions, and may dominate a centralized exchange. See Kawakami (2017), Malamud and Rostek (2017), Babus and Parlatore (2017), and Cespa and Vives (2018). We do not seek to explain fragmentation per-se, nor study the asset issuance. Instead, we study whether investors make socially optimal participation decision in exogenously fragmented markets.

The rest of the paper is organized as follows. In Section 2, we lay out our model of participation in an OTC and in a centralized market. In Section 3, we define an equilibrium and study its general properties. In Section 4, we consider analytical examples of the general model, under alternative assumptions about investor heterogeneity and participation costs. Then, we derive our main normative results regarding the social gain/loss from increasing the participation of customers in the centralized market.

2 Model

The model presented in this section generalizes Atkeson, Eisfeldt, and Weill (2015, henceforth AEW) to allow for multidimensional banks' heterogeneity at the trading stage, and participation decisions in multiple markets.

2.1 Time, Agents and Assets

There are four dates $t \in \{0, 1, 2, 3\}$, one good consumed at the terminal date, $t = 3$, and one risky asset with normally distributed payoff. There is a measure one of traders with Constant Absolute Risk Aversion (CARA) utility over $t = 3$ consumption. Traders' common CARA coefficient is denoted by η. We assume that traders are organized into a measure one of large coalitions, called "banks." Banks are heterogeneous in two dimensions: their risk-sharing needs and their OTC market trading technology.

Heterogeneous risk-sharing needs are generated by heterogeneity in banks' *endowment* $\omega \in [0, \Omega]$ of the risky asset. Since CARA banks have incentives to trade so as to equalize their holdings, banks have stronger risk-sharing needs if their initial endowment is very large or small relative to the economy-wide average.

Heterogeneity in trading technology is generated by heterogeneity in what we term *OTC market trading capacities*, denoted by $k \in [0, K]$. Banks with larger k can trade larger quantities in the OTC market, in a manner to be specified precisely later.

We let Φ denote the exogenous joint distribution of endowments and trading capacities, (ω, k), over the set $[0, \Omega] \times [0, K]$, equipped with its Borel σ-algebra.

2.2 Participation

At $t = 0$, banks make one of the four market participation decisions: They can choose to trade the risky asset in a decentralized OTC market with bilateral bargaining, $\pi = \text{otc}$, in a centralized market with price taking, $\pi = \text{cent}$, or in both markets at the same time, $\pi = \text{otc+cent}$. They can also stay in autarky, $\pi = \text{aut}$. We let $\Pi \equiv \{\text{otc}, \text{cent}, \text{otc+cent}, \text{aut}\}$ be the set of all possible participation decisions.

The cost of participation π is denoted by $C(\pi) \geq 0$. We normalize $C(\text{aut}) = 0$ and we assume that it is more costly to participate in the centralized than the OTC market, $C(\text{cent}) > C(\text{otc})$. Theoretically, this condition is necessary to obtain equilibria in which banks participate in both markets. Empirically, this condition means that centralized markets impose more stringent regulatory requirements, and involve larger infrastructure costs which are passed on to market participants.[5]

After its participation decision, a bank's type is summarized by the triple $x \equiv (\omega, k, \pi)$. On aggregate, banks' collective participation decisions induce an endogenous measure N over the set X of all possible bank types. We will call the endogenous measure N *the participation path*. The participation path must satisfy a basic consistency condition. Namely, the marginal distribution over endowment and trading capacity, (ω, k) must equal the exogenous distribution

[5] Biais and Green (2006) discuss regulatory and disclosure costs in centralized markets, while Duffie (2013) emphasizes the "fixed costs of setting up and maintaining exchange trading." In a consulting document about the mandate to migrate trade to Swap Execution Facilities, Paulson et al. (2013) write that "operationalizing to trade on SEFs as well as bilaterally could lead to higher transaction costs initially as dealers pass on infrastructure costs and fees for central clearing and execution."

Φ, that is:

$$\Phi(A) = \int \mathbb{I}_{\{(\omega(x),k(x))\in A\}}\, dN(x), \tag{1}$$

for all measurable subsets $A \subseteq [0,\Omega] \times [0,K]$, and where $\omega(x)$ and $k(x)$ denote the projection functions mapping the bank type, x, to its corresponding endowment and trading capacity, ω and k.

2.3 Trading and payoffs

The timing of trade after participation decisions have been made is the following. At $t = 1$, banks who chose $\pi \in \{\text{otc}, \text{otc+cent}\}$ trade in the OTC market. At $t = 2$, banks who chose $\pi \in \{\text{cent}, \text{otc+cent}\}$ trade in the Walrasian market. At $t = 3$, every bank consolidates all its traders' positions, and the risky asset pays off.

Next, we describe the trading process and corresponding payoffs in detail.

OTC market trades. Let $X_{\text{otc}} \equiv \pi^{-1}(\{\text{otc}, \text{otc+cent}\})$ denote the set of banks' types participating in the OTC market, and assume positive participation, that is, $N(X_{\text{otc}}) > 0$. Then, at $t = 1$, all banks with type $x \in X_{\text{otc}}$ send their traders to the decentralized OTC market, where they are paired uniformly to bargain over a bilateral trade. When a trader from a type-x bank is paired with a trader from a type-x' bank the trader of type x buys a quantity $\gamma(x,x')$ of assets from the trader of type x', in exchange for the payment $P_{\text{otc}}(x,x')\gamma(x,x')$. A positive $\gamma(x,x')$ is an outright purchase, and a negative an outright sale. OTC market trades must satisfy two elementary feasibility constraints:

$$\gamma(x,x') + \gamma(x',x) = 0 \text{ for all } (x,x') \in X^2 \tag{2}$$

$$\gamma(x,x') = 0 \text{ if } (x,x') \notin X_{\text{otc}}^2. \tag{3}$$

Condition (2) imposes bilateral feasibility and condition (3) rules out trades between types who do not participate in the OTC market.

We add a crucial constraint which prevents banks from fully sharing their risk by trading only in the OTC market. Namely, we subject traders to a bilateral trading capacity constraint:

$$-M(x',x) \leq \gamma(x,x') \leq M(x,x'). \tag{4}$$

for some continuous and positive-valued function M. In our analytical examples below, M is increasing in the trading capacities of both counterparties, $k(x)$ and $k(x')$. The interpretation is that the bank's risk-management department imposes risk-limit (notional or risk-based) on the position taken by its trading desk, and that banks differ in their ability or willingness to let their traders take large positions.[6]

To prevent perfect risk sharing in the OTC market, we assume that the function M is bounded above by some $\bar{M}$ such that:

$$\bar{M} < \sup_{\omega \in [0,\Omega]} \left| \omega - \int \omega' d\Phi\left(\omega', k'\right) \right|.$$

Taking stock, a collection of OTC market bilateral trades, $\gamma : X \to \mathbb{R}^2$, is *feasible* if it is measurable and if it satisfies (2), (3), and (4).

Centralized market trades. Let $X_{\text{cent}} \equiv \pi^{-1}(\{\text{cent}, \text{otc+cent}\})$ denote the set of banks' type participating in the centralized market and assume positive participation, that is, $N(X_{\text{cent}}) > 0$. Then, at time $t = 2$, banks with types $x \in X_{\text{cent}}$ can trade without frictions in the centralized market: they can purchase unrestricted quantities, $\varphi(x)$, at some fixed price P_{cent}.

We let a collection of centralized-market trades be described by some measurable function $\varphi : X \to \mathbb{R}$. Centralized-market trades are *feasible* if:

$$\int \varphi(x)\, dN(x) = 0 \tag{5}$$

$$\varphi(x) = 0 \text{ if } x \notin X_{\text{cent}}. \tag{6}$$

Condition (5) is the market-clearing condition in the centralized market, and condition (6) rules out trade by banks who do not participate in the centralized market.

[6] At this stage we consider a general specification that can also depend on endowments, $\omega(x)$ and $\omega(x')$, so that our model admits other interpretations. For example, assume that a fraction k of the assets endowed to the (ω, k) bank can be traded in the OTC market, subject to a short-selling constraint. Then, assuming that each bank distributes its tradeable asset endowment uniformly to its traders before they enter the OTC market, the short-selling constraint can be represented by $M(x, x') = k'\omega(x')$. That is the maximum that a trader can purchase is determined by the amount of asset brought by its OTC market counterparty. For another example, assume that the asset being traded is a derivative contract, and that each derivative contract sold must be backed by one unit of collateral. Let $k(x)$ denote the amount of collateral that a type-x bank endows each traders prior to matching in OTC market. Then, the number of contracts that a type-x trader can purchase is bounded by the amount of collateral brought by its counterparty, $M(x, x') = k(x')$.

Consolidation and payoffs. After trading in the OTC and the centralized market, a bank consolidates all its trades. The asset random payoff, denoted by v, realizes. Each trader then receives a consumption equal to the average per-trader payoff of the bank:

$$-C\left[\pi(x)\right] + \omega(x)v + \int \gamma(x,x')\left[v - P_{\text{otc}}(x,x')\right]\,dN(x' \,|\, X_{\text{otc}}) + \varphi(x)\left(v - P_{\text{cent}}\right).$$

The first term is the participation cost. The second term is the payoff of the initial asset endowment. The third term is the net payoff from OTC-market trades. Finally, the fourth term is the net payoff from centralized-market trades. To calculate the certainty equivalent corresponding to this payoff, it is convenient to define the bank's *post-trade* exposure to the risky asset:

$$g(x) \equiv \omega(x) + \int \gamma(x,x')\,dN(x' \,|\, X_{\text{otc}}) + \varphi(x). \tag{7}$$

The first term is the initial endowment, the second term is the exposure gained via OTC-market trades, and the third term is the exposure gained via centralized-market trades. The certainty-equivalent payoff of the bank can be written:

$$-C\left[\pi(x)\right] + U\left[g(x)\right] - \int \gamma(x,x')P_{\text{otc}}(x,x')\,dN(x' \,|\, X_{\text{otc}}) - \varphi(x)P_{\text{cent}}, \tag{8}$$

where $U(g) \equiv \mathbb{E}\left[v\right]g - \frac{\eta}{2}\mathbb{V}\left[v\right]g^2$ is the mean-variance payoff that obtains with CARA utility, absolute risk aversion η, and normally distributed risky asset.[7]

3 Equilibrium

In this section, we define an equilibrium in two steps. First, we define equilibrium trades given participation decisions, summarized by the participation path N. Second, we define equilibrium participation decisions, N, given rational expectations about subsequent equilibrium trades.

[7] In general, if v is not normally distributed, many of our results would go through because $U(g)$ would be a well-behaved concave function. While not crucial, the assumption of quadratic payoffs is used in two places. First, when the support of the distribution of (ω, k) is not discrete, we use the assumption of quadratic payoff to guarantee a continuity property and complete the final steps of the proof of Proposition 2. Second, in Section 4, the assumption of quadratic payoffs guarantees that participation incentives are appropriately symmetric between net buyers and net sellers, which is useful to simplify our parametric examples.

3.1 Equilibrium trades given participation

We assume for simplicity that participation is positive in all markets, $N(X_{\text{otc}}) > 0$ and $N(X_{\text{cent}}) > 0$ (as will be clear, it is straightforward to extend the analysis to the other cases).

Optimal trading in the OTC market. We assume that, in the OTC market, traders view themselves as small relative to their bank's coalition, and do not coordinate their trades with other traders in the same bank coalition. As a result, we assume that a trader's objective is to maximize the value to the bank of its bilateral trade, taking as given all other bilateral and centralized trades in the bank coalition.[8] Formally, the objective of a type-x trader who meets a type-x' trader in the OTC market is to maximize his marginal impact on the certainty-equivalent payoff, (8):

$$\gamma(x, x') \left\{ U_g \left[g(x) \right] - P_{\text{otc}}(x, x') \right\},$$

where $U_g(\cdot)$ is the derivative of $U(\cdot)$, and where an individual trader takes others' decisions as given, as summarized by the post-trade exposure, $g(x)$. Assuming that bilateral trades are the outcome of symmetric Nash bargaining between the two traders, we obtain the following optimality conditions:

$$\gamma(x, x') = \begin{cases} M(x, x') & \text{if } g(x) < g(x') \\ \in \left[-M(x', x), M(x, x') \right] & \text{if } g(x) = g(x') \\ -M(x', x) & \text{if } g(x) > g(x') \end{cases} \tag{9}$$

for all $(x, x') \in X_{\text{otc}}^2$. That is, if the type-$x$ trader expects a lower post-trade exposure than the type-x' trader, then he should purchase some asset. Given that the type-x trader views itself as small relative to its coalition, it is optimal to purchase as much as feasible given the bilateral trading capacity constraint (4). The asset price between x and x' is set to split the bilateral gains from trade in half:

$$P_{\text{otc}}(x, x') = \frac{1}{2} \left\{ U_g \left[g(x) \right] + U_g \left[g(x') \right] \right\}. \tag{10}$$

One sees from (9) that OTC market trades tend to bring banks' post-trade exposures closer together, in that banks with small exposures tend to buy from banks with high exposures.

[8] This approach is used extensively in monetary economics literature as well as by AEW. See Lucas (1990), Andolfatto (1996), Shi (1997), and Shimer (2010), among others.

However, in general, banks do not equalize their post-trade exposures, for two reasons. First, the trading capacity constraint (4) limits the size of OTC market trades. Second, the bilateral trading protocol implies that traders in the same bank will trade in opposite direction depending on who they meet. For example, type-x traders purchase from type-x' traders if $g(x) < g(x')$, but they sell if $g(x) > g(x')$. Trades of the same size going in opposite direction will net out to zero, and so do not contribute to the equalization of post-trade exposures.

Optimal trading in the centralized market. Taking derivative of (8) with respect to φ, we obtain:

$$U_g\left[g(x)\right] = P_{\text{cent}} \;\; \forall x \in X_{\text{cent}}. \tag{11}$$

Clearly, this implies that banks who participate in the centralized market fully equalize their post-trade exposures:

$$g(x) = \int_{X_{\text{cent}}} g(x')\, dN(x' \mid X_{\text{cent}}) \;\; \forall x \in X_{\text{cent}}. \tag{12}$$

Definition of equilibrium given participation. An equilibrium given positive participation in all markets, $N(X_{\text{otc}}) > 0$ and $N(X_{\text{cent}}) > 0$, is a collection $(\gamma, \varphi, g, P_{\text{otc}}, P_{\text{cent}})$ of feasible OTC market bilateral trades, γ, feasible centralized market trades, φ, post-trade exposures, g, OTC market prices, P_{otc}, and Walrasian price, P_{cent}, such that (7), (9), (10), (11) and (12) hold.

Notice that our definition of equilibrium requires that the optimality conditions (9) and (12) hold *everywhere* instead of *almost everywhere*. That is, optimality must hold even for sets of types that have measure zero according to N. This is economically important: it means that we require banks' trading decisions to be optimal both on and off the participation path, which is crucial to evaluate the value of all possible participation decisions and solve for equilibrium patterns of participation.[9]

[9]Suppose, for example, that some banks with endowments and trading capacities (ω, k) in some set A only participate in the centralized market. That is, $N(A \times \{\text{cent}\}) > 0$ but $N(A \times \{\text{otc}, \text{otc+cent}\}) = 0$. To verify whether participating only in the centralized market is indeed optimal, banks $(\omega, k) \in A$ evidently need to compare the value of *all* participation decisions, $\pi \in \{\text{otc}, \text{cent}, \text{otc+cent}\}$. This means that we need to solve for trades, (g, γ, φ), and payoffs for all types $x \in A \times \{\text{otc}, \text{cent}, \text{otc+cent}\}$, even if some of these types have measure zero on the equilibrium participation path, N.

Existence. To establish existence, we show that an equilibrium allocation, (γ, φ, g), solves a planning problem. Namely, we consider the social planning problem:

$$W^{\star}(N) = \sup \int \{U[g(x)] - C[\pi(x)]\} \, dN(x),$$

with respect to feasible OTC market trades, γ, feasible Walrasian market trades, φ, and post-trade exposures g generated by (γ, φ) according to (7). We obtain:

Proposition 1. *There exists an equilibrium given positive participation in all markets. All equilibria solve the planning problem given participation. The equilibrium is essentially unique in the sense that all equilibria share the same post-trade risk exposures, g, OTC prices, P_{otc}, centralized-market price, P_{cent}, and certainty-equivalent payoff, (8). Equilibria may only differ in terms of OTC and centralized-market trades, (γ, φ).*

In the first step of our existence proof, we show that all equilibria solve the planner's problem. This follows by direct comparison of the planner's first-order conditions with the equilibrium optimality condition (9)-(12).

In the second step, we establish that the planner's problem has at least one solution. This follows from standard results on convex optimization in infinite dimensional vector spaces (see, for example, Proposition 1.2, Chapter II in Eckland and Témam, 1987).

Finally, in the third step, we show that an appropriately selected solution of the planner's problem is the basis of an equilibrium. The key difficulty in completing this step is that, since the planner only cares about those types that have positive measure according to the participation path N, the planner's problem has many solutions. Indeed, the planner only needs to determine trading behavior on the participation path, while our definition of equilibrium requires to determine trading behavior both on and off the path.[10] However, the planner's problem uniquely determines aggregate market conditions: the post-trade exposures of all counterparties that can be met with positive probability in the OTC market, $g(x)$, and the price in the centralized market, P_{cent}. This allows us to calculate optimal trading behavior given any off-path participation decision.

[10] Continuing with the example of Footnote 9, consider banks with endowments and trading capacities (ω, k) in some set A who only participate in the centralized market, $N(A \times \{\text{cent}\}) > 0$ and $N(A \times \{\text{otc}, \text{otc+cent}\}) = 0$. As argued above, the equilibrium requires to determine their payoffs and trades on and off the participation path, that is, for all participation decisions $\pi \in \{\text{cent}, \text{otc}, \text{otc+cent}\}$. But since these banks only participate in the centralized market, they are in zero measure in the OTC market. Formally, banks of type $x \in A \times \{\text{otc}, \text{otc+cent}\}$ are in measure zero according to N, which implies that they have zero weight in the planner's objective, and so their socially optimal trades are indeterminate.

3.2 Equilibrium participation

The certainty equivalent of a type-x bank, before participation cost, can be written:

$$U\left[g(x)\right] - \int \gamma(x,x')\, P_{\text{otc}}(x,x')\, dN(x' \,|\, X_{\text{otc}}) - \varphi(x)\, P_{\text{cent}}.$$

The first term is the certainty equivalent utility over post-trade exposure. The second term is the total cost for OTC market trades, and the third term is the total cost of centralized market trades. Using (7), (9), (10), and (11), this formula can be re-written conveniently as follows.

Lemma 1. *Assume that participation is positive in all markets. Then, the certainty equivalent of a bank of type $x \in X$ can be written*

$$U\left[\omega(x)\right] + \text{MPV}(x), \;\; \textit{with}\; \text{MPV}(x) = K(x) + \frac{F(x)}{2},$$

and $\text{MPV}(x)$ *is the marginal private value,* $K(x)$ *the competitive surplus, and* $F(x)$ *the frictional surplus, defined as:*

$$\begin{aligned} K(x) &\equiv U\left[g(x)\right] - U\left[\omega(x)\right] - U_g\left[g(x)\right]\left[g(x) - \omega(x)\right] \\ F(x) &\equiv \mathbb{I}_{\{x \in X_{\text{otc}}\}} \int_{X_{\text{otc}}} \Big\{ \left(U_g\left[g(x)\right] - U_g\left[g(x')\right]\right)^{+} M(x,x') \\ &\qquad\qquad + \left(U_g\left[g(x)\right] - U_g\left[g(x')\right]\right)^{-} M(x',x) \Big\}\, dN(x' \,|\, X_{\text{otc}}). \end{aligned} \tag{13}$$

The marginal private value, or MPV, is the net certainty-equivalent payoff relative to autarky. It can be decomposed into two components.

The first component of the MPV is what we call the competitive surplus. It is the value of changing exposure, assuming that all assets are bought and sold at marginal value, $U_g\left[g(x)\right]$. This marginal value corresponds to the asset price for centralized-market trades.

But the MPV is larger than the competitive surplus because, for OTC-market trades, a bank is able to bargain a price that is more advantageous than its marginal value. Specifically, when a type-x trader expects a lower post-trade exposure than its counterparty, it purchases a quantity $M(x,x')$ below marginal value. Vice versa, when it expects a higher post-trade exposure it sells a quantity $M(x',x)$ above marginal value. The second component of the MPV is the sum of all these OTC bargaining gains for a bank of type-x: it is equal to half of the frictional surplus, $F(x)/2$, due to the symmetry in bargaining powers.

Definition of an equilibrium with positive participation. An equilibrium with positive participation in both markets is a positive measure, N, over the set of banks' types, X, satisfying the following three conditions. First, participation is positive in both markets: $N(X_{\text{otc}}) > 0$ and $N(X_{\text{cent}}) > 0$. Second, the participation path must satisfy (1), that is, it must be consistent with the primitive exogenous distribution of risk endowment and trading capacities, Φ. Third, the participation path must be generated by optimal participation decisions, that is:

$$\int \left(\text{MPV}(x) - C\left[\pi(x)\right] - \max_{\pi' \in \Pi} \left\{ \text{MPV}\left(\omega(x), k(x), \pi'\right) - C(\pi') \right\} \right) dN(x) = 0.$$

It is conceptually more subtle to define an equilibrium in which participation is zero in one or in both markets, $N(X_{\text{otc}}) = 0$ or $N(X_{\text{cent}}) = 0$. Indeed, in that case one needs specify a bank's rational belief regarding its payoff if it chooses to enter a market in which no one else participates.[11] For the remainder of this paper, we will focus on equilibria in which participation is positive in all markets.

3.3 Efficient participation

In this section, we study whether more participation in the centralized market is welfare improving.

A first-order approach. To answer this question, we calculate the change in welfare associated with a small reallocation of banks across markets. Formally, let us start from some arbitrary participation path such that $N(X_{\text{otc}}) > 0$ and $N(X_{\text{cent}}) > 0$. A reallocation of banks across market induces a new participation path of the form:

$$N + \varepsilon\left(n^+ - n^-\right),$$

where ε is a small positive number parameterizing the scale of reallocation, while (n^+, n^-) is a pair of positive finite measures parameterizing the direction of reallocation. Notice that reallocation involve increasing participation in some markets, as parameterized by n^+, and decreasing participation in others, as parameterized by n^-. The direction of reallocation

[11] One possible choice of beliefs is to assume that, if no one else participates in a market, then the payoff of participation is zero. But this creates coordination failures: no participation is always an optimal choice if the market is expected to be empty. Another choice is to assume that some infinitesimal exogenous measure of banks participate in all markets at no cost. Finally, one could also attempt to specify beliefs in the spirit of subgame perfection, as in a competitive search equilibrium. That is, if a bank chooses to enter in an empty market, it expects to attract the banks who have most incentives to enter.

(n^+, n^-) is *admissible* if it satisfies the following two natural conditions. First, the participation path $N + \varepsilon(n^+ - n^-)$ must satisfy (1), so as to remain consistent with the distribution of endowments and trading capacities.[12] Second, the new participation path $N + \varepsilon(n^+ - n^-)$ must remain positive for all ε small enough. Formally, we require that the negative part n^- is absolutely continuous with respect to N, with a bounded Radon-Nikodym derivative.

From Proposition 1, we know that equilibrium social welfare given the participation path $N + \varepsilon(n^+ - n^-)$ solves an optimization problem: it is equal to $W^\star[N + \varepsilon(n^+ - n^-)]$, the maximized value of the social planner's objective given the participation path $N + \varepsilon(n^+ - n^-)$. This observation allows us to use Envelope Theorems to calculate the derivative of social welfare with respect to ε. Precisely, adapting argument from Milgrom and Segal (2002), we obtain:

Proposition 2. *Assume participation is positive in all markets and consider any admissible direction of reallocation, (n^+, n^-). Then, the function $\varepsilon \mapsto W^\star[N + \varepsilon(n^+ - n^-)]$ is right-hand differentiable at $\varepsilon = 0$, with derivative:*

$$\frac{d}{d\varepsilon}\left[W^\star(N + \varepsilon(n^+ - n^-)\right](0^+) = \int \{\mathrm{MSV}(x) - C[\pi(x)]\}\left[dn^+(x) - dn^-(x)\right] \tag{14}$$

$$\textit{with} \quad \mathrm{MSV}(x) \equiv \mathrm{MPV}(x) + \frac{1}{2}\left[F(x) - \mathbb{I}_{\{x \in X_{\text{otc}}\}}\bar{F}\right],$$

where $\mathrm{MPV}(x)$ is the marginal private value, defined in Lemma 1, $F(x)$ is the equilibrium frictional surplus given participation path N, defined in (13), and $\bar{F} = \int F(x')\,dN(x' \mid X_{\text{otc}})$ is the average equilibrium frictional surplus across banks who participate in the OTC market.

The Proposition shows that, when a bank participates exclusively in the centralized market, then the marginal social and private value are equalized $\mathrm{MSV}(x) = \mathrm{MPV}(x)$. It is intuitive that, in this case, private and social incentives are aligned. Indeed, in the centralized market, trading is multilateral with price taking, which promotes efficient outcomes.

However, the Proposition shows that when a bank participates in the OTC market, $x \in X_{\text{otc}}$, there is a wedge between the marginal social value and the marginal private value, $\mathrm{MSV}(x) - \mathrm{MPV}(x) = \frac{1}{2}\left[F(x) - \mathbb{I}_{\{x \in X_{\text{otc}}\}}\bar{F}\right]$. This is because OTC market prices do not incorporate the social value and cost of match creation and destruction induced by OTC market participation.

Participation induces match creation simply because a new participant trades with incumbents. The social value of match creation is equal to the frictional surplus, $F(x)$.[13] However,

[12] Equivalently, (n^+, n^-) must satisfy the conservation equation: $\int \mathbb{I}_{\{(\omega(x),k(x))\in A\}}\,dn^+(x) = \int \mathbb{I}_{\{(\omega(x),k(x))\in A\}}\,dn^-(x)$ for all measurable sets $A \subseteq [0,\Omega]\times[0,K]$ of endowments and trading capacities.

[13] Indeed in the OTC market, a new type-x participant always purchases from incumbents with higher exposure, $g(x') \geq g(x)$. This pushes down the unit social cost of increasing the new participant exposure

when they bargain, banks only appropriate half of the social value of match creation. The other half of the frictional surplus, $F(x)/2$, drives a wedge between the MSV and the MPV.

But match creation has an opportunity cost: when a new participant matches with incumbents, incumbents match less together. This is what we call match destruction. To calculate the quantity and social value of these destroyed matches, notice first that the creation of a match between a new participant and an incumbent requires just one incumbent trader. The destruction of a match between incumbents frees up exactly two traders. Hence, the quantity of match destroyed per match created is equal to one half. Moreover, the matching protocol implies that matches are destroyed at random in the populations of incumbents. Hence, the average social value of a match destroyed is equal to the average frictional surplus. Taken together, these observations imply that the social cost of match destruction is equal to half of the average frictional surplus, $\bar{F}/2$.

Proposition 2 provides a general formula for evaluating the welfare impact of any arbitrary reallocation. This formula can be simplified further in a special case of interest: the reallocation of marginal banks, that is, a bank which is indifferent between participating in the OTC and participating in the centralized market.

The reallocation of a marginal bank under exclusive participation. Suppose that there is exclusive participation that is, banks participate either in the centralized market or in the OTC market, but not in both at the same time. In this context, consider a marginal bank, that is, a bank that is indifferent between exclusive participation in the OTC market, with a type x such that $\pi(x) = \text{otc}$, and exclusive participation in the centralized market, with a type x' such that $\pi(x') = \text{cent}$. Reallocation from the OTC to the centralized market corresponds to a direction (n^+, n^-) such that $dn^+(x') = dn^-(x)$, and zero everywhere else. Then, Proposition 2 implies the following welfare variation:

$$\Delta W = \text{MSV}(x') - C\left[\pi(x')\right] - \left(\text{MSV}(x) - C\left[\pi(x)\right]\right).$$

Using the indifference condition $\text{MPV}(x) - C\left[\pi(x)\right] = \text{MPV}(x') - C\left[\pi(x')\right]$ between the OTC market and the centralized market,

$$\Delta W = \text{MSV}(x') - \text{MPV}(x') - \left[\text{MSV}(x) - \text{MPV}(x)\right].$$

below its marginal value, $U_g\left[g(x)\right]$, by an amount equal to $U_g\left[g(x)\right] - U_g\left[g(x')\right] > 0$. The opposite is true when the new type-x participant sells.

Moreover, as noted before, exclusive participation in the centralized market aligns private with social values, $\mathrm{MPV}(x') = \mathrm{MSV}(x')$. Therefore, according to Proposition 2, the social value of reallocation is

$$\Delta W = -\left[\mathrm{MSV}(x) - \mathrm{MPV}(x)\right] = \frac{1}{2}\left[-F(x) + \bar{F}\right]. \tag{15}$$

To understand this formula, recall that when the marginal bank is reallocated to the centralized market, it no longer matches with infra-marginal OTC banks, with social cost equal to the frictional surplus, $F(x)$. But infra-marginal OTC banks substitute their match with the marginal bank by matches amongst themselves, with social value equal to the average frictional surplus, $\bar{F}$. The formula shows that, if $F(x) < \bar{F}$, then the reallocation of marginal banks from the OTC market to the centralized market is welfare improving.

Inspecting the frictional-surplus formula of Lemma 1, it is clear that $F(x)$ will be smaller than $\bar{F}$, so that $\Delta W > 0$, under the following two broad conditions:

1. the trades of the marginal bank have a sufficiently small <u>size</u>, measured by $M(x, x')$, relative to the average,

2. the trades of the marginal bank must create a small enough <u>surplus</u> per quantity traded, measured by $|U_g\left[g(x)\right] - U_g\left[g(x')\right]|$, relative to the average.

In the next section, we develop analytical examples to better understand circumstances under which these two conditions are likely to hold.

Non-exclusive participation in the centralized market. Now let us make a different assumption: the marginal bank is indifferent between trading exclusively in the OTC market, with a type x such that $\pi(x) = \text{otc}$, and trading non-exclusively in the centralized market, with a type x' such that $\pi(x') = \text{otc+cent}$. Proceeding as above we obtain a different formula for the social value of reallocation:

$$\Delta W = \frac{1}{2}\left[-F(x) + F(x')\right].$$

Relative to the previous formula, in Equation (15), this new formula replaces the average surplus, $\bar{F}$, by the frictional surplus of the bank when it trades in the OTC and the centralized market at the same time. This is because, in that case, there is no match creation and destruction: the bank continues to trade in the OTC market. But its frictional surplus changes, since it has access to the centralized market.

The formula of Proposition 1 now suggest a different condition for a welfare improvement: the reallocation of the marginal banks to non-exclusive participation in the centralized market must increase its surplus per quantity traded in the OTC market.

4 Analytical examples

In the previous section, we derived and discussed conditions for welfare improving reallocation. But these conditions depend on endogenous outcomes: which bank is at the margin, whether that bank is indifferent between exclusive or non-exclusive participation in the centralized market, and its relative post-trade exposure in the OTC market. We now characterize these endogenous outcomes, in the context of tractable parametric examples.

4.1 Exclusive participation with heterogeneous capacities

With exclusive participation, we have seen above that reallocating the marginal bank to the centralized market is welfare improving under two conditions: the marginal bank must trade smaller-than-average quantities, and it must create smaller-than-average surplus per quantity traded. We now construct an analytical example to better understand the first of the two conditions. To do so, we generate heterogeneity in quantity traded by assuming that banks have heterogeneous capacities. We keep the surplus per quantity traded constant by assuming that banks have homogeneous risk-sharing need. Specifically, we assume that

(i) banks are heterogeneous in their trading capacities: the distribution of k across banks is uniform over the interval $[0, 1]$, and independent from the distribution of endowments;

(ii) banks are homogeneous in their risk-sharing needs: half start with endowment $\omega = 0$ and the other half with endowment $\omega = 1$;[14]

(iii) participation costs $C(\pi)$ induce exclusive participation: optimal participation choices are either $\pi = \text{otc}$ or $\pi = \text{cent}$;

It is not completely obvious how to best specify the bilateral trading capacity constraint, $M(k(x), k(x'))$, since we do not provide precise micro-foundations. To guide the specification,

[14] Although endowments are heterogeneous, the risk-sharing needs are effectively the same. Indeed, as will become clear shortly, the trades of $\omega = 0$ and $\omega = 1$ banks are symmetric, and their surpluses are the same.

it is useful to start with an obvious observation: any arbitrary trading capacity constraint depends positively on $\min\{k(x), k(x')\}$ and on $\max\{k(x), k(x')\}$. The positive dependence on the "min" means that matching with a low-capacity counterparty forces smaller trades. The positive dependence on the "max" means that matching with a high-capacity counterparty makes larger trades feasible. Empirically, a positive dependence on the "max" is realistic: it is consistent with the observation that, in OTC market data, the customer-to-customer volume is much smaller than the customer-to-dealer volume. Indeed, the "max" restricts the quantity traded between two customer banks, who both have a small capacity, but it does not restrict the quantity traded when at least one counterparty is a dealer, who has a large capacity.[15] The "min", on the other hand, would restrict the quantity traded in both cases, and so would imply counterfactually small trade between dealers and customers. A related empirical observation in favor of the "max" is that a negative shock affecting dealers' capacity will translate into lower customer-to-dealer trading volume. In line with this implication, Adrian, Boyarchenko, and Shachar (2017) study the post-crisis regulations' differential impact on the US corporate bonds market participants. They find that institutions that face tighter regulatory constraints are less able to intermediate customer trades.

For tractability, in what follows we assume that trading capacity *only* depends on the "max". That is, we assume that $M(k(x), k(x')) = m(\max\{k(x), k(x')\})$ for some increasing function m. But after the change of variable $m(k)$, one immediately sees that one can assume without loss of generality that trading capacity is *equal* to the "max", that is:

$$M(x, x') = \max\{k(x), k(x')\}. \tag{16}$$

Conjectured equilibrium participation patterns. Next, we guess and verify that, under parameter restrictions to be determined, there exists an equilibrium such that participation is independent of ω: banks with endowment 0 and 1 make the same participation decision if they have the same trading capacity. Moreover, there is some $k^\star \in [0, 1]$ such that banks with $k < k^\star$ participate exclusively in the centralized market, while banks with $k \geq k^\star$ participate exclusively in the OTC market. We derive below equilibrium trades and payoff of all participation decisions, and verify that these conjectured participation decisions are indeed optimal.

[15] For a theoretical motivation, consider the recent work of Üslü (2015), who studies a dynamic model where trade quantities are determined bilaterally on the margin without any exogenous restrictions. He shows that, as a result of dealers' *endogenous* willingness to trade in large quantities, trading with a dealer enables a customer to trade in large quantities that would not be possible in a trade with another customer.

Equilibrium trades. We first characterize trading patterns in both the centralized and the OTC market, given any arbitrary threshold $k^\star$.

Lemma 2. *Given our conjectured participation patterns, post-trade exposures are symmetric across the two endowment types, $g(0, k, \pi) = 1 - g(1, k, \pi)$. Conditional on $\pi = \text{cent}$, post-trade exposure are $g(0, k, \text{cent}) = 1/2$. Conditional on $\pi = \text{otc}$, for banks with $\omega = 0$,*

- *Post-trade exposures do not depend on capacity, k:*

$$g(0, k, \text{otc}) = \frac{1}{2}\mathbb{E}\left[\max\{k', k''\} \,|\, k', k'' \geq k^\star\right];$$

- *Bilateral trades depend on capacity k. Small-k banks trade like customers: they tend to buy from all banks. Large-k banks trade like dealers: they tend to buy from $\omega = 1$ banks and sell to $\omega = 0$ banks.*

We already know from (12) that post-trade exposures in the centralized market are equalized. By symmetry, they must be equal to 1/2, which is the average endowment across banks participating in the centralized market. It is natural that the symmetry result extends to the OTC market as well.

What is perhaps more surprising is that banks in the OTC market with identical endowment equalize their post-trade exposures, even though they have different capacities. To gain intuition, let us observe that, in equilibrium, $\omega = 0$ banks buy assets from $\omega = 1$ banks, and more so if they have a large capacity k. This tends to create dispersion in post-trade exposures amongst $\omega = 0$ banks. But then the trades among $\omega = 0$ banks go in the opposite direction, and tend to equalize post-trade exposures. Namely, $\omega = 0$ banks with large post-trade exposures arising from trades with $\omega = 1$ banks optimally sell to other $\omega = 0$ banks with small post-trade exposures arising from trades with $\omega = 1$ banks. Lemma 2 shows that this leads to complete equalization: banks with identical ω have identical post-trade exposures, regardless of their capacity k.

It is then easy to solve for the equalized post-trade exposures of $\omega = 0$ banks who participate in the OTC market. Since the aggregate trade amongst all $\omega = 0$ must net out to zero, the equalized post-trade exposure of $\omega = 0$ banks must be equal to their aggregate trade with $\omega = 1$ banks. Precisely, an $(\omega = 0, k')$ trader matches with an $(\omega = 1, k'')$ trader with probability 1/2, and trades the quantity $\max\{k', k''\}$. Aggregating across all bilateral matches, and then across all banks who participate in the OTC market, we obtain the formula of the Lemma.[16] Note that, given our assumption that $k \leq 1$, risk sharing is imperfect: the post-trade exposure of a $\omega = 0$ bank is less than $1/2$, while the post-trade exposure of a $\omega = 1$ banks is greater than $1/2$.

Although banks with identical endowment reach the same post-trade exposure regardless of their trading capacity, k, their trading behavior depends on k. Indeed, an $\omega = 0$ bank with a small capacity buys less from $\omega = 1$ banks, so the only way it can equalize its post-trade exposures is if it also buys from $\omega = 0$ banks. In that sense, it trades like a customer. Vice versa, an $\omega = 0$ bank with a large capacity buys more from $\omega = 1$ banks, so it must sell to other $\omega = 0$ banks. In that sense, it trades like a dealer. This can be shown formally. The average quantity purchased by an $(\omega = 0, k)$ bank from $\omega = 1$ banks is:

$$\bar{\gamma}(k) = \mathbb{E}\left[\max\{k, k'\} \,|\, k' \geq k^\star\right]$$

To attain the post-trade exposure of Lemma 2, the $\omega = 0$ bank with capacity k must trade with other $\omega = 0$ banks a quantity:

$$\mathbb{E}\left[\bar{\gamma}(k')\right] - \bar{\gamma}(k).$$

This expression is clearly decreasing in k. Since it averages out to zero across $k \geq k^\star$, it must be positive for $k \simeq k^\star$ banks, and strictly negative for $k \simeq 1$ banks, establishing the claim.

The trade-off between OTC and centralized markets. As shown in Section 2, banks who participate in the centralized market equalize their risk exposures to $g(\omega, k, \text{cent}) = 1/2$. Hence the competitive and frictional surpluses conditional on participating exclusively in the centralized market are:

$$K(\omega, k, \text{cent}) = \frac{|U_{gg}|}{8}$$
$$F(\omega, k, \text{cent}) = 0.$$

[16] Notice that this is true both on and off the participation path. That is, if an $\omega = 0$ bank with capacity $k \in [0, k^\star]$ deviates, and participates in the OTC market instead of the centralized market, then its post-trade exposure would still be given by the formula of Lemma 2. Indeed, following the deviation, the bank with $k < k^\star$ only matches with the banks who actually participate in the OTC market, whose capacity is $k' \in [k^\star, 1]$. Our specification of the trading capacity constraint then implies that the trade size is determined by k' and not by k. So the bank has the same opportunities as a $k^\star$ bank, and so attains the same post-trade exposure.

Next, using the formula for post-trade exposures, we find the competitive and frictional surplus conditional on participating exclusively in the OTC market:

$$K(\omega, k, \text{otc}) = \frac{|U_{gg}|}{2} \left[g(0, k, \text{otc})\right]^2$$
$$F(\omega, k, \text{otc}) = \frac{|U_{gg}|}{2} \left[1 - 2g(0, k, \text{otc})\right] \mathbb{E}\left[\max\left\{k, k'\right\} \mid k' \geq k^\star\right],$$

One sees that, since banks equalize their post-trade exposures, and since they have linear marginal utility, the competitive surplus is independent of both ω and k. However, the frictional surplus is an increasing function of k.[17] Although the equalized post-trade exposures imply that the frictional surplus *per unit* traded is the same in all matches between $\omega = 0$ and $\omega = 1$ banks, and equal to $1 - 2g(0, k, \text{otc})$, the quantities traded in the OTC market differ systematically across banks. In particular, banks with larger k trade larger quantities and so generate larger frictional surplus.

Taken together, these calculation lead to a simple but key result in this section:

Lemma 3. *Given our conjectured participation patterns, the difference between the MPV of participating in the centralized market and that of participating in the OTC market,* $\text{MPV}(\omega, k, \text{cent}) - \text{MPV}(\omega, k, \text{otc})$*, is constant in* $k < k^\star$ *and strictly decreasing in* $k \geq k^\star$*.*

Therefore, banks with larger trading capacities have stronger preferences to participate in the OTC market vs. the centralized market. In this example, this stronger preference stems from their ability to trade larger quantities in the OTC market, leading to a larger frictional surplus.

Equilibrium participation. Using the formula for the surpluses shown above together with our assumption that k is uniformly distributed across banks, we obtain after some algebra that, given our guess about participation patterns, the marginal bank is indifferent between the centralized and the OTC market if and only if:

$$\frac{|U_{gg}|}{36} (1 - k^\star)^2 = C(\text{cent}) - C(\text{otc}). \tag{17}$$

This indifference condition implies natural comparative statics. The OTC market is larger, as proxied by a smaller $k^\star$, when risk-sharing considerations matter less to investors (smaller $|U_{gg}|$)

[17] To be precise, the frictional surplus is strictly increasing function of k for $k \in [k^\star, 1]$ and constant for $k \in [0, k^\star)$, due to the max specification of trading capacity constraint, (16).

and when the participation cost is high in the centralized market relative to the OTC market (larger $C(\text{cent}) - C(\text{otc})$).

Next, we need to verify that $k^\star$ solving (17) is the basis of an equilibrium. To do so we must show that (i) banks with $k < k^\star$ prefer exclusive centralized market participation over exclusive OTC participation, and vice versa for banks $k > k^\star$; (ii) autarky and joint participation in the centralized and OTC markets are dominated participation decisions for all banks. We go through these verification steps in the appendix and derive associated necessary and sufficient conditions on costs.

Proposition 3. *There exists an equilibrium in which banks with $k \in [0, k^\star)$ participate exclusively in the Walrasian market, and banks with $k \in [k^\star, 1]$ participate exclusively in the OTC market, if and only if $k^\star$ solve (17), $0 \leq C(\text{otc}) < C(\text{cent}) \leq \min\left\{\frac{|U_{gg}|}{8}, C(\text{otc}) + \frac{|U_{gg}|}{36}\right\}$, and $C(\text{otc+cent})$ is large enough.*

The Proposition imposes natural restrictions on the level of participation costs and on their difference. In particular, the level of $C(\text{otc})$ and $C(\text{cent})$ cannot be too large, otherwise autarky would be preferred to market participation. The difference between $C(\text{cent})$ and $C(\text{otc})$ cannot be too large either, otherwise there could not be positive participation in both markets at the same time: namely, if $C(\text{cent}) \gg C(\text{otc})$ then all banks would find it optimal to participate in the OTC market only. Another restriction is that the cost to participate simultaneously in both markets are large enough, so as to guarantee exclusive participation.

More participation in the centralized market is welfare improving. We now study whether increasing participation in the centralized market is welfare improving. We consider reallocating banks near the margin ($k > k^\star$ but near $k^\star$) from the OTC market to the centralized market. Using the formula derived in Section 3.3, we obtain:

$$\Delta W = \frac{1}{2}\left[-F\left(\omega, k^\star, \text{otc}\right) + \bar{F}\right] > 0.$$

As we have shown above, the frictional surplus is increasing in $k \geq k^\star$, which implies that the frictional surplus of the marginal bank is lower than the average frictional surplus in the OTC market. Therefore, ΔW is positive; i.e., moving the marginal bank from the OTC to the Walrasian market is welfare improving. This means that there is too much participation in the OTC market and too little in the Walrasian market. This stems from the fact that the marginal bank has the lowest OTC trading capacity, and so trade relatively smaller quantities than other banks in the OTC market. Hence, by removing the marginal bank from the OTC market, the social planner destroys small trades and, correspondingly, allows the remaining infra-marginal banks to create larger trades. Since the surplus per quantity traded is the same for the destroyed and for the newly created trades, this results in a welfare improvement.

Put differently, the pricing mechanism gives customer banks participation incentives that are too strong: indeed, when she participate in the OTC market, the marginal bank appropriates half of the frictional surplus, $F(0, k^\star, \text{otc})/2$. Even though, for the marginal bank, the frictional surplus appropriated is small in absolute terms, it is too large in relative terms. That is, it is larger than the net social frictional surplus the bank creates by participating, $F(0, k^\star, \text{otc}) - \bar{F}/2$.

4.2 Exclusive participation with heterogeneous risk-sharing needs

In the analytical example of the previous section, the welfare impact of increasing centralized market participation depends crucially on the endogenous distribution of quantity traded across marginal and infra-marginal banks. We now turn to an example in which it depends on the distribution of surplus per quantity traded. We thus make the following assumptions:

(i) banks are heterogeneous in their endowment: the distribution of ω across banks is uniform over the interval $[0, 1]$;

(ii) banks are homogeneous in their trading capacities: the trading capacity constraint is $M(x, x') = k$ for all x and some $k < \frac{1}{2}$;

(iii) participation costs $C(\pi)$ induce exclusive participation: optimal participation choices are either $\pi = \text{otc}$ or $\pi = \text{cent}$;

Conjectured equilibrium participation patterns. We guess and verify that, under parameter restrictions to be determined, there exists an equilibrium with the following features. First, participation is symmetric around $\omega = 1/2$: banks with endowment ω and $1 - \omega$ make the same participation decision. Second, extreme-ω banks, who have the strongest risk-sharing needs, participate exclusively in the centralized market, which is the most efficient trading venue. Middle-ω banks, on the other hand, participate exclusively in the OTC market.[18] Precisely, there is some $\omega^\star \in [0, 1/2]$ such that banks with $\omega \in [0, \omega^\star) \cup (1 - \omega^\star, 1]$ participate in the centralized market, and banks with $\omega \in [\omega^\star, 1 - \omega^\star]$ participate in the OTC market, where $\omega^\star$ satisfies $\omega^\star + k < 1/2$. As will be clear below, this latter condition ensures that the marginal

[18]These participation patterns are similar to the one obtained by Gehrig (1993), Miao (2006).

bank shares risk imperfectly in the OTC market, and so faces meaningful trade-off between the OTC and the centralized market.

Equilibrium trades. We first characterize trading patterns in the OTC market.

Lemma 4. *Given our conjectured participation patterns, post-trade exposures are symmetric across the two endowment types, $g(\omega, k, \pi) = 1 - g(1-\omega, k, \pi)$. Conditional on $\pi = \text{cent}$, post-trade exposures are equal to $g(\omega, k, \text{cent}) = 1/2$. Conditional on $\pi = \text{otc}$:*

- *Post-trade exposures are strictly increasing in ω;*
- *Bilateral trades differ depending on ω. Extreme-ω banks trade like customers: their trade with other banks tend to go in the same direction. Middle-ω banks trade like dealers: they tend to buy from banks with higher endowment, and sell to banks with lower endowment.*

Figure 1 illustrates the post-trade exposures conditional on $\pi = \text{cent}$ and $\pi = \text{otc}$. One can obtain a closed-form solution for the post-trade exposures in the OTC market. Indeed, since $g(\omega, k, \text{otc})$, is strictly increasing in ω, the optimality condition (9) implies that an ω trader always sells k units to $\omega' < \omega$ traders, and always purchases k units from $\omega' > \omega$ traders:

$$g(\omega, k, \text{otc}) = \omega - kN(\{\omega' < \omega\} \mid X_{\text{otc}}) + kN(\{\omega' > \omega\} \mid X_{\text{otc}}), \tag{18}$$

where $N(\cdot \mid X_{\text{otc}})$ is the conditional distribution of types in the OTC market. Given our assumed participation decisions and given uniform distribution,

$$N(\{\omega' < \omega\} \mid X_{\text{otc}}) = \begin{cases} 0 & \text{if } \omega \in [0, \omega^\star) \\ \dfrac{\omega - \omega^\star}{1 - 2\omega^\star} & \text{if } \omega \in [\omega^\star, \frac{1}{2}], \end{cases}$$

and, by symmetry, $N(\{\omega' < \omega\} \mid X_{\text{otc}}) = 1 - N(\{1 - \omega' < \omega\} \mid X_{\text{otc}})$ for $\omega \geq \frac{1}{2}$. Plugging the expression for the conditional distribution in (18), and keeping in mind that $\omega^\star + k < \frac{1}{2}$, one easily sees that $g(\omega, k, \text{otc})$ is strictly increasing, so our guess is verified.[19]

The post-trade exposures and associated trading patterns are the same as in AEW. Banks with extreme ω trade like "customers," in the sense that most of their trades go in the same

[19] As before, the formula (18) applies on and off the participation path, for all banks: the one who actually decide to participate in the OTC market (in the support of $N(\cdot \mid X_{\text{otc}})$) and the one who do not (outside the support). Figure 1 shows $g(\omega, k, \text{otc})$, with solid lines for the banks in the support and with dashed lines for the banks outside the support.

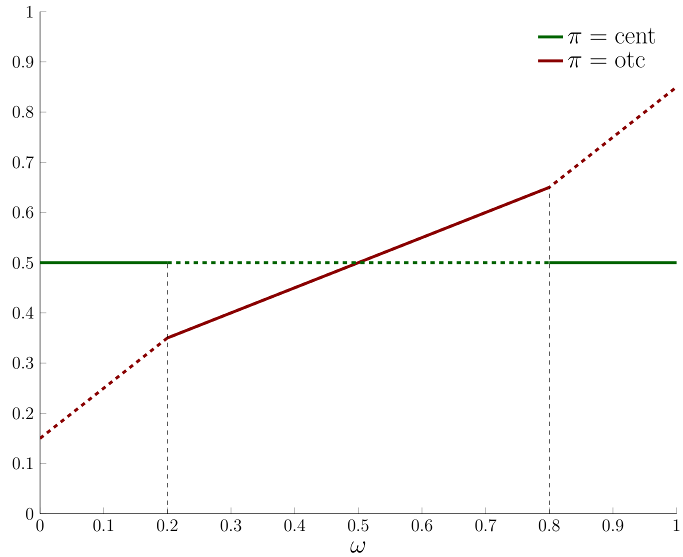

Figure 1: Post-trade exposures as a function of endowment in the example with heterogeneous endowment and exclusive participation. Plain lines show the post-trade exposure on the participation path, that is, conditional on a bank's actual participation decision. Dotted lines show the post-trade exposure off the participation path, conditional on making the alternative exclusive market participation decision.

direction. Specifically, low-ω banks mostly purchase assets, and high-ω banks mostly sell assets. Middle-ω banks, on the other hand, trade like "intermediaries." They trade in all directions, buying from high-ω banks and selling to low-ω banks.

The trade-off between OTC and centralized markets. Banks who participate in the centralized market equalize their risk exposures, that is, $g(\omega, k, \text{cent}) = \frac{1}{2}$. Therefore, their competitive and frictional surpluses are:

$$K(\omega, k, \text{cent}) = \frac{|U_{gg}|}{2}\left(\frac{1}{2} - \omega\right)^2$$
$$F(\omega, k, \text{cent}) = 0.$$

One sees that the competitive surplus is strictly U shaped in ω. This is simply because extreme-ω banks have the strongest incentive to trade and share risk.

In the OTC market, we obtain the following competitive and frictional surpluses:

$$K(\omega, k, \text{otc}) = \frac{|U_{gg}|}{2}\left[g(\omega, k, \text{otc}) - \omega\right]^2$$
$$F(\omega, k, \text{otc}) = |U_{gg}|k \int \left|g(\omega', k, \text{otc}) - g(\omega, k, \text{otc})\right| dN(\omega' \mid X_{\text{otc}}).$$

Just as for the centralized market the surpluses are U-shaped and symmetric around $\frac{1}{2}$. The property holds for the competitive surplus because extreme-ω banks have the strongest incentive to trade and share risk. It also holds for the frictional surplus because extreme-ω are able to bargain larger discounts (premia) relative to their marginal value when they buy (sell).

The above calculations show that extreme-ω banks have the strongest absolute incentives to trade – in other word, they strongly prefer participating in some market rather than staying in autarky. Next, we determine the optimal participation decision by studying the relative incentives to participate in the centralized vs. the OTC market.

Lemma 5. *Given our conjectured participation patterns, the difference between the* MPV *of participating in the centralized market and that of participating in the OTC market* $\text{MPV}(\omega, k, \text{cent}) - \text{MPV}(\omega, k, \text{otc})$ *is strictly U shaped and symmetric around* $\omega = 1/2$.

Therefore, although extreme-ω banks have higher MPV of participating in the OTC market, their MPV of participating in the centralized market is even higher. This is a key step in verifying that our guessed participation patterns are indeed optimal.

Equilibrium participation. Given our conjectured participation patterns, an $\omega^\star$-bank must be indifferent between participating in the centralized market and the OTC market, that is $\text{MPV}(\omega^\star, k, \text{cent}) - \text{MPV}(\omega^\star, k, \text{otc}) = C(\text{cent}) - C(\text{otc})$. After some algebra, we obtain that this condition writes:

$$\frac{|U_{gg}|}{2}\left(\frac{1}{2} - \omega^\star\right)\left(\frac{1}{2} - (\omega^\star + k)\right) = C(\text{cent}) - C(\text{otc}). \tag{19}$$

The left-side is a decreasing function of $\omega^\star \in \left[0, \frac{1}{2} - k\right]$. Hence, as long as $C(\text{cent}) - C(\text{otc})$ is not too large relative to $|U_{gg}|$, there is a unique solution such that $\omega^\star + k < \frac{1}{2}$.

The indifference conditions implies three natural comparative statics. Participation in the OTC market increases (the threshold $\omega^\star$ decreases) when the OTC market trading technology is relatively more efficient, as proxied by an increase in k, when the difference in participation costs $C(\text{cent}) - C(\text{otc})$ decreases, and when risk-sharing is less important to banks, as proxied by a decrease in $|U_{gg}|$.

We then verify that our conjectured participation patterns are optimal and we obtain:

Proposition 4. *There exists an equilibrium in which banks with* $\omega \in [0, \omega^\star) \cup (1 - \omega^\star, 1]$ *participate exclusively in the centralized market, and banks with* $\omega \in [\omega^\star, 1 - \omega^\star]$ *participate exclusively in the OTC market, for some* $\omega^\star + k < 1/2$ *if and only if* $\omega^\star$ *is the unique solution* *of the indifference condition* (19) *belonging to* $\left[0, \frac{1}{2} - k\right]$, $3\,C(\text{otc}) + \frac{8}{|U_{gg}|k^2} C(\text{otc})^2 < C(\text{cent}) < C(\text{otc}) + \frac{|U_{gg}|}{8}(1-2k)$, *and* $C(\text{otc+cent})$ *is large enough.*

More participation in the centralized market is welfare reducing. As before, we reallocate banks near the $\omega^\star$ margin from the OTC to the centralized market. Using equation (15), we obtain that the change in social welfare is:

$$\Delta W = \frac{1}{2}\left[-F\left(\omega^\star, k, \text{otc}\right) + \bar{F}\right] < 0.$$

Indeed, the frictional surplus is U-shaped: this implies that, amongst all banks who participate in the OTC market, an $\omega^\star$-bank has the largest frictional surplus. In particular, this frictional surplus is larger than $\bar{F}$, the average surplus of banks who participate in the OTC market.

The result arises because the marginal bank has the strongest unfulfilled risk-sharing needs. In equilibrium, the distance between its post-trade exposure and the post-trade exposures of other banks participating in the OTC market is largest. Correspondingly, it creates a larger frictional surplus per quantity traded than any other bank participating in the OTC market. Hence, by reallocating a marginal bank from the OTC market to the centralized market, the social planner destroys matches in which the surplus per quantity traded is large, and creates matches in which the surplus per quantity traded is small. Since we are assuming here that banks have the same capacity, the trade size is the same in all matches. Hence, the reallocation of the marginal bank reduces welfare.

We conclude that, with heterogeneity in endowments, the centralized market is inefficiently large. Encouraging further centralized market participation is welfare reducing.

4.3 Non-exclusive participation

In the Appendix, we provide an analytical example with non-exclusive participation. However, based on the intuition developed above, one can readily guess the answer to our key question: with non-exclusive participation, reallocating banks to the centralized market turns out to be welfare reducing.

To gain intuition, suppose there exists an equilibrium in which banks either participate exclusively in the OTC market, or participate at the same time in the OTC and in the centralized markets. That is, equilibrium participation decisions are either $\pi = \text{otc}$, or $\pi = \text{otc+cent}$. Assume as before that participation decisions are symmetric. In that case, banks in the centralized market have post-trade exposure $g(\omega, k, \text{otc+cent}) = 1/2$. In the OTC market, on the other hand, post-trade exposures are dispersed with a median of $1/2$. Finally, the social value of reallocating a bank to the centralized market, that is, from $\pi = \text{otc}$ to $\pi = \text{otc+cent}$ is

$$\Delta W = \frac{1}{2}\left[-F(\omega, k, \text{otc}) + F(\omega, k, \text{otc+cent})\right].$$

The key observation is that, when banks participate more in the centralized market, their post trade exposures become $g(\omega, k, \text{otc+cent}) = 1/2$. But then their average surplus per quantity traded in the OTC market is smaller. Indeed, the frictional surplus of a bank with post-trade exposure g is proportional to the absolute distance between g and the post-trade exposures of other banks. As is well known, such an average distance is minimized if g is the median post-trade exposure, that is, if $g = 1/2$.[20] The quantity traded stays the same, since a bank's trading capacity constraint does not change. Taken together, this implies that the frictional surplus of the bank falls. Hence, with non-exclusive participation, reallocating banks to the centralized market is welfare reducing.

4.4 Empirical implications of different heterogeneities

Our analytical examples so far suggest that, when participation is exclusive, increasing participation in centralized market is welfare improving when banks differ mostly in their ability to take large positions in OTC market (trading capacity), and welfare reducing when banks differ mostly in their risk-sharing needs (endowment). Therefore, it is crucial to empirically distinguish an economy in which banks differ mostly in terms of their trading capacities, from an economy in which banks differ mostly in terms of their risk-sharing needs. To do so, we study banks' net and gross OTC trading volume, defined as:

$$NV(x) \equiv \int \gamma(x, x')\, dN(x' \mid X_{\text{otc}}) = g(x) - \omega(x),$$
$$GV(x) \equiv \int |\gamma(x, x')|\; N(x' \mid X_{\text{otc}}).$$

The net volume only depends on post-trade exposures, $g(x)$, which are uniquely pinned down in an equilibrium conditional on participation. The gross volume depends on bilateral exposures, $\gamma(x, x')$, which are not uniquely pinned down when traders expect the same post-trade expo-

[20]When trade size are heterogeneous, the median must be calculated after a change of measure that puts more weights on banks with larger trading capacities. But as long as the distribution of endowment is symmetric, the median does not change.

sures. This never happens in our example with heterogeneous endowment, but it happens in our example with heterogeneous capacities. In that case, we need to pick a particular equilibrium selection for bilateral trade, as explained below.

Heterogeneity in trading capacities. Since all $\omega = 0$-banks who participate in the OTC market have the same exposure, the bilateral trades between them are not uniquely determined. We make the natural assumption that, when two $\omega = 0$-traders meet, they "swap" the exposures their banks acquired from $\omega = 1$-banks, and vice versa when two $\omega = 1$-traders meet. Precisely, let

$$\bar{\gamma}(k) \equiv \mathbb{E}\left[\max\{k, k'\} \,|\, k' \geq k^{\star}\right],$$

denote the net trade of a $(\omega = 0, k)$ bank with all $\omega = 1$ banks. We assume that, when an $(\omega = 0, k)$-trader meets an $(\omega = 0, k')$-trader, their bilateral trade is $\bar{\gamma}(k') - \bar{\gamma}(k)$. It is easy to check that these bilateral exposures satisfy the trading capacity constraint. Moreover, when aggregated across all possible $\omega = 0$ counterparties, these swaps mechanically equalize exposure of all $\omega = 0$ banks who participate in the OTC market. Hence, these swaps implement the equilibrium post-trade exposure. Given our selection for bilateral trade, the net and gross volume are:

$$\begin{aligned} NV(0, k, \text{otc}) &= \frac{1}{2}\mathbb{E}\left[\bar{\gamma}(k') \,|\, k' \geq k^{\star}\right] \\ GV(0, k, \text{otc}) &= \frac{1}{2}\mathbb{E}\left[|\bar{\gamma}(k') - \bar{\gamma}(k)| \;|\, k' \geq k^{\star}\right] + \frac{1}{2}\bar{\gamma}(k), \end{aligned}$$

One sees that the net volume is independent of capacity. The gross volume, on the other hand, is greater than the net volume, and is easily seen to be increasing in k.[21]

[21] Alternatively, we could focus on bilateral trades that minimize the gross volume as in AEW; i.e., the total gross volume an $\omega = 0$-bank creates by trading with other $\omega = 0$-banks is equal to the minimum amount of trades that is necessary to reach the post-trade exposure of the atom of the $\omega = 0$-banks. In that case, the gross volume would be $\frac{1}{2}\mathbb{E}\left[|g - \bar{\gamma}(k)| \;|\, k' \geq k^{\star}\right] + \frac{1}{2}\bar{\gamma}(k)$, where $g = \frac{1}{2}\mathbb{E}\left[\bar{\gamma}(k') \,|\, k' \geq k^{\star}\right]$. One easily sees that this minimum gross volume is weakly increasing in k. In particular, it is constant for $k < \bar{k}$ and increasing for $k \geq \bar{k}$, where $\bar{\gamma}(\bar{k}) = g$.

Heterogeneity in endowment. With heterogeneous endowments, the net and gross volume for a bank with endowment ω are:

$$NV(\omega, k, \text{otc}) = k\left(1 - 2N\left[\{\omega' < \omega\} \mid X_{\text{otc}}\right]\right)$$
$$GV(\omega, k, \text{otc}) = k.$$

The net volume is largest for banks with extreme endowments, and smallest for banks with intermediate endowments. The gross volume, on the other hand, is the same for all banks. This is because all banks have the same trading capacity, k, and because there are strict gains from trade in all bilateral matches. Therefore, the same quantity is traded in all bilateral matches for all banks, leading to constant gross volume.

Stylized facts. Empirical evidence about net volume is reported by Siriwardane (2018) in the context of CDS markets: he finds that dealers have large net volume. This observation is better in line with the heterogeneous capacity model, in which the net volume of endogenous dealer can be large. Indeed, dealers and customers have identical risk-sharing needs, leading them to take identical post-trade exposures, and so have identical net volume. In the heterogeneous endowment model, by contrast, dealers tend to have low net volume: indeed intermediate-endowment banks provide intermediation services precisely because they do not need to use their capacity to change their net exposures.

Empirical evidence further suggests that dealers concentrate a very large fraction of gross volume as well (Bech and Atalay, 2010; Di Maggio, Kermani, and Song, 2017; Hollifield, Neklyudov, and Spatt, 2017; Li and Schürhoff, 2019). This observation holds after controlling for natural measure of bank size (see for example Atkeson, Eisfeldt, and Weill, 2013, for the CDS market), which is relevant for our model in which all banks have the same number of traders and hence the same size. This goes in favor of the heterogeneous capacity model, in which the gross volume of dealers is larger. In the heterogeneous endowment model, by contrast, all agents have the same gross volume.

4.5 Market resiliency differential

Regulators' concern about OTC markets also originate from the commonly held view that these markets are less stable than exchanges during financial turmoils. In this section, we show that our model can be extended by introducing differential probabilities of markets' shutdown upon a crisis. The main take-away is that as long as investors have rational expectations about shutdown risk, our previous results go through.

Assume that between the initial participation decision stage and the following trading stages, a crisis occurs with probability $\theta < 1$. Assume further that, upon a crisis the OTC market shuts down with probability 1 while the centralized market shuts down only with probability $\delta < 1$. Once a market shuts down, participants are prevented from trading in this specific market. By assuming that the centralized market does not necessarily shut in a crisis period while the OTC market does, we introduce the idea that not only centralized markets induce better risk sharing but also are more resilient.

In this setup, the formulas for the marginal private and social values naturally derive from our previous analysis.[22] Indeed, conditional on a specific market being shut, the marginal private value of a given participant is zero, and so is his marginal social value. Then, the marginal private values of entering the OTC and centralized market for a bank (ω, k) are, respectively,

$$\begin{aligned}\mathrm{MPV}(\omega, k, \mathrm{otc}) &= (1-\theta)\left[K(\omega, k, \mathrm{otc}) + \frac{1}{2}F(\omega, k, \mathrm{otc})\right]\\ \mathrm{MPV}(\omega, k, \mathrm{cent}) &= (1-\delta\theta)K(\omega, k, \mathrm{cent}).\end{aligned}$$

Similarly, the marginal social values of entering the OTC and centralized market of a bank (ω, k) are, respectively,

$$\begin{aligned}\mathrm{MSV}(\omega, k, \mathrm{otc}) &= (1-\theta)\left[K(\omega, k, \mathrm{otc}) + F(\omega, k, \mathrm{otc}) - \frac{\bar{F}}{2}\right]\\ \mathrm{MSV}(\omega, k, \mathrm{cent}) &= (1-\delta\theta)K(\omega, k, \mathrm{cent}).\end{aligned}$$

As before, in an equilibrium with exclusive participation in which both markets operate, one can find a marginal bank $(\omega^\star, k^\star)$ indifferent between participating in either market, that is, $\mathrm{MPV}(\omega^\star, k^\star, \mathrm{otc}) - C(\mathrm{otc}) = \mathrm{MPV}(\omega^\star, k^\star, \mathrm{cent}) - C(\mathrm{cent})$. If we again consider reallocating this specific bank near the margin from the OTC market to the centralized market, we obtain the following welfare variation:

$$\Delta W = \frac{1-\theta}{2}\left[-F\left(\omega^\star, k^\star, \mathrm{otc}\right) + \bar{F}\right].$$

[22] Because traders have CARA utility, our previous analysis applies up to a suitable first-order approximation. See Appendix C for details.

As in Sections 4.1 and 4.2, welfare improves in the case of heterogeneous capacities, and deteriorates in the case of heterogeneous risk-sharing needs.

5 Conclusion

This paper uses a venue choice model to study theoretically whether the migration of trades from OTC to centralized markets is socially desirable. Our model provides us with two necessary conditions for this migration to be welfare improving: First, banks must differ from each other mostly in terms of their trading ability rather than their trading need. Second, participation costs must induce exclusive participation decisions. By comparing trading-volume patterns that arise in our model and are observed in practice, we argue that these necessary conditions for a welfare improvement are met.

A Proofs

A.1 Proof of Proposition 1

A.1.1 The planning problem

Existence of a solution. Consider the space of square integrable measurable trades (γ, φ). Social welfare can be written

$$\begin{aligned} W(\gamma, \varphi \,|\, N) =& \mathbb{E}\left[v\right] \int \left[\omega(x) + \varphi(x) + \int \gamma(x,x') dN(x' \,|\, \mathrm{otc})\right] dN(x) \\ & - \frac{\eta}{2} \mathbb{V}\left[v\right] \int \left[\omega(x) + \varphi(x) + \int \gamma(x,x')\, dN(x' \,|\, \mathrm{otc})\right]^2 dN(x). \end{aligned}$$

Given that the measure N is finite, $N(X) < \infty$, repeat applications of the Cauchy Schwartz inequality show that the social welfare function is continuous in (γ, φ).[23]

Because $(\gamma, \varphi) \mapsto W(\gamma, \varphi \,|\, N)$ is continuous, it is lower semi-continuous. Clearly, the function is also concave and the constraint set is bounded. Existence of a solution then follows from an application of Proposition 1.2, Chapter II in Eckland and Témam (1987).

Almost everywhere uniqueness of $g(x)$. Because the objective is strictly concave, all solutions must share the same $g(x)$, almost everywhere according to N.

Post-trade exposures are constant almost everywhere over $X_{\mathbf{cent}}$. Otherwise, given strict concavity, one could achieve a strictly higher welfare by pooling the exposures of all Walrasian market participants. Given that $M(x, x')$ is bounded on the support of N, and given that pre-trade exposures are bounded, the Walrasian trades required to pool exposures are also bounded, hence feasible.

First-order conditions. Take any optimal γ and let

$$\begin{aligned} \hat{\gamma}(x,x') &= \gamma(x,x') + \varepsilon \left[M(x,x') - \gamma(x,x')\right] \mathbb{I}_{\{g(x)<g(x')\}} - \varepsilon \left[M(x',x) + \gamma(x,x')\right] \mathbb{I}_{\{g(x)>g(x')\}} \\ &\equiv \gamma(x,x') + \varepsilon \Delta(x,x'). \end{aligned}$$

[23]Consider for example the function $\gamma \mapsto \int\int \gamma(x,x') dN(x)\, dN(x')$. We have that

$$\begin{aligned} &\left| \int\int \left(\gamma(x,x') - \hat{\gamma}(x,x')\right)\, dN(x \,|\, \mathrm{otc})\, dN(x') \right| \leq \int\int \left|\gamma(x,x') - \hat{\gamma}(x,x')\right|\, dN(x \,|\, \mathrm{otc})\, dN(x') \\ &\leq \int\int \left|\gamma(x,x') - \hat{\gamma}(x,x')\right|^2\, dN(x \,|\, \mathrm{otc})\, dN(x') = \|\gamma - \hat{\gamma}\|_{L^2} N(X), \end{aligned}$$

by the Cauchy Schwartz inequality. Similar arguments can be applied to all the other terms.

One easily sees that $\hat{\gamma}$ is feasible for the planning problem, as long as $\varepsilon \in [0,1]$. Hence, for small ε, we obtain that up to second-order terms:

$$\begin{aligned}
&\frac{W(\hat{\gamma},\varphi) - W(\gamma,\varphi)}{N(X_{\text{otc}})} = \varepsilon \int\int U'\left[g(x)\right]\Delta(x,x')\,dN(x\,|\,\text{otc})\,dN(x'\,|\,\text{otc}) \\
=&\frac{\varepsilon}{2}\int\int U'\left[g(x)\right]\Delta(x,x')\,dN(x\,|\,\text{otc})\,dN(x'\,|\,\text{otc}) \\
&+\frac{\varepsilon}{2}\int\int U'\left[g(x')\right]\Delta(x',x)\,dN(x'\,|\,\text{otc})\,dN(x\,|\,\text{otc}) \\
=&\frac{\varepsilon}{2}\int\int \left\{U'\left[g(x)\right] - U'\left[g(x')\right]\right\}\Delta(x,x')\,dN(x\,|\,\text{otc})\,dN(x'\,|\,\text{otc}) \\
=&\frac{\varepsilon}{2}\int\int \left\{U'\left[g(x)\right] - U'\left[g(x')\right]\right\}\left[M(x,x') - \gamma(x,x')\right]\mathbb{I}_{\{g(x)<g(x')\}}\,dN(x\,|\,\text{otc})\,dN(x'\,|\,\text{otc}) \\
&-\frac{\varepsilon}{2}\int\int \left\{U'\left[g(x)\right] - U'\left[g(x')\right]\right\}\left[M(x',x) + \gamma(x,x')\right]\mathbb{I}_{\{g(x)>g(x')\}}\,dN(x\,|\,\text{otc})\,dN(x'\,|\,\text{otc}).
\end{aligned}$$

If $\gamma(x,x')$ is optimal, this must be negative. Since both integrands are positive, they must be zero almost everywhere, or:

$$N(\,\cdot\,|\,\text{otc})\text{ a.e. over }X_{\text{otc}}^2,\ \gamma(x,x') = \begin{cases} M(x,x') & \text{if } g(x) < g(x') \\ \in [-M(x',x), M(x,x') & \text{if } g(x) = g(x') \\ -M(x',x) & \text{if } g(x) > g(x') \end{cases}$$

A.1.2 A preliminary result

In this section we study what can be viewed as the "partial equilibrium" determination of post-trade exposures $g(x)$. That is, we consider the problem of determining the post-trade exposure of an individual bank, taking as given arbitrary post-trade exposures in the OTC market. This preliminary result is important because it allows us to determine the post-trade exposures of any (ω,k), even those who choose not to participate in the OTC market.

Formally, we consider some arbitrary $x \in X$, and we fix some arbitrary function for the post-trade exposures, $h(x')$, of other banks in the OTC market. We then seek a solution to the problem:

$$g = \omega(x) + \int_{X_{\text{otc}}} \gamma(x,x')\,dN(x'\,|\,\text{otc}), \tag{20}$$

where, for all $x' \in X_{\text{otc}}$:

$$\gamma(x,x') = \begin{cases} M(x,x') & \text{if } g < h(x') \\ \in [-M(x',x), M(x,x')] & \text{if } g = h(x') \\ -M(x',x) & \text{if } g > h(x') \end{cases} \tag{21}$$

To that end we define the following two functions. First:

$$\begin{aligned}\overline{V}(g) &= \omega(x) + \int M(x,x')\mathbb{I}_{\{g \le h(x')\}}\, dN(x' \,|\, \text{otc}) - \int M(x',x)\mathbb{I}_{\{g > h(x')\}}\, dN(x' \,|\, \text{otc}) \\ &= \omega(x) + \int \left[M(x,x') + M(x',x)\right]\mathbb{I}_{\{g \le h(x')\}}\, dN(x' \,|\, \text{otc}) - \int M(x',x)\, dN(x' \,|\, \text{otc}).\end{aligned}$$

The function $\overline{V}(g)$ represents the maximum post-trade exposure of the bank of type x, if all its traders take position anticipating that the post-trade will be g. One easily sees that $\overline{V}(g)$ is decreasing and left-continuous. Second, we let:

$$\begin{aligned}\underline{V}(g) &= \omega(x) + \int M(x,x')\mathbb{I}_{\{g < h(x')\}}\, dN(x' \,|\, \text{otc}) - \int M(x',x)\mathbb{I}_{\{g \ge h(x')\}}\, dN(x' \,|\, \text{otc}) \\ &= \omega(x) + \int M(x',x)\, dN(x' \,|\, \text{otc}) - \int \left[M(x,x') + M(x',x)\right]\mathbb{I}_{\{g \ge h(x')\}}\, dN(x' \,|\, \text{otc}).\end{aligned}$$

The function $\underline{V}(g)$ represents the minimum post-trade exposure of a bank active in the OTC market only, if all its traders take position anticipating that the post-trade will be g. One easily sees that $\underline{V}(g)$ is decreasing and right-continuous. One also sees easily that:

$$\overline{V}(g^+) = \underline{V}(g) \text{ and } \overline{V}(g) = \underline{V}(g^-)$$

where the notation g^+ and g^- is for right- and left-limit. Finally, given that pre-trade exposures, $\omega(x)$, are bounded, and given that $M(x,x')$ is bounded over the support of N, there exists $a < b$ such that $\underline{V}(g) \in [a,b]$ and $\overline{V}(g) \in [a,b]$ for all g.

We then have:

Lemma 6. *A post-trade exposure g solves (20)-(21) if and only if $g \in [\underline{V}(g), \overline{V}(g)]$.*

Proof. For the "only if" part, take a solution of (20)-(21) and use the optimality conditions (21) to show that it belongs to $[\underline{V}(g), \overline{V}(g)]$. For the "if" part, take some $g(x) \in [\underline{V}(g), \overline{V}(g)]$, let $\gamma(x,x') = M(x,x')$ if $g(x) < h(x')$, let $\gamma(x,x') = -M(x',x)$ if $g(x) > h(x')$, and let $\gamma(x,x') = \alpha M(x,x') - (1-\alpha)M(x',x)$ if $g = h(x')$, where α is chosen so that

$$\begin{aligned}g = \omega(x) &+ \int M(x,x')\mathbb{I}_{\{g < h(x')\}}\, dN(x' \,|\, \text{otc}) - \int M(x',x)\mathbb{I}_{\{g > h(x')\}}\, dN(x' \,|\, \text{otc}) \\ &+ \int \left[\alpha M(x,x') - (1-\alpha)M(x',x)\right]\mathbb{I}_{\{g = h(x')\}}\, dN(x' \,|\, \text{otc}).\end{aligned}$$

Given that $g \in [\underline{V}(g), \overline{V}(g)]$, it follows that $\alpha \in [0,1]$, hence $\gamma(x,x') \in [-M(x',x), M(x,x')]$ if $g = h(x')$. □

Next we show that:

Lemma 7. *The fixed point problem $g \in [\underline{V}(g), \overline{V}(g)]$ has a unique solution.*

Proof. To show that a solution exists, we apply Kakutani's fixed point theorem to the correspondence

$$g \rightrightarrows V(g) \equiv [\underline{V}(g), \overline{V}(g)].$$

It is clear that $V(g \,|\, h)$ takes values that are convex sets included in $[a, b]$. To see that $V(g)$ has a closed graph consider any converging sequence $(g_n, v_n) \to (g, v)$ such that $v_n \in V(g_n \,|\, h)$ for all n. Then we can extract a subsequence of g_n such that either $g_n \leq g$ for all n or $g_n \geq g$ for all n. Suppose that we are in the former case (the latter case is symmetric). Then, since $\underline{V}(g)$ is decreasing, it follows that $v_n \geq \underline{V}(g_n) \geq \underline{V}(g)$. Going to the limit, we obtain $v \geq \underline{V}(g)$. Since $v_n \leq \overline{V}(g_n)$ and since $\overline{V}(g)$ is left-continuous, it follows that $v \leq \overline{V}(g)$. Therefore, $v \in V(g)$.

For uniqueness consider any $g \in V(g)$. Then $g \geq \underline{V}(g) = \overline{V}(g^+)$. But since $\overline{V}(g)$ is decreasing, it follows that $g' > \overline{V}(g')$ for all $g' > g$. Hence, $g' \notin V(g')$. A similar argument applies to $g' < g$. $\square$

Finally, we obtain

Lemma 8. *The solution of the fixed point problem* (20)-(21) *remains the same:*

- *If the post-trade exposure function h is changed but remains the same $N(\,\cdot\,|\,\mathrm{otc})$-almost everywhere;*
- *If the optimality conditions are required to hold $N(\,\cdot\,|\,\mathrm{otc})$-almost everywhere.*

This follows directly because these change do not impact the functions $\underline{V}(g)$ and $\overline{V}(g)$, hence the fixed point remains the same.

A.1.3 Equilibrium existence

We construct an equilibrium from a solution of the planning problem. The main difficulty in doing so is that the planner's problem determines trades only N-almost everywhere, that is, only for investors' types who are actually present in the market. An equilibrium, by contrast, requires that trade be well defined for all types.

Consider, then a solution (γ, φ) to the planning problem, and the associated post-trade exposures g. These may not constitute an equilibrium because the equilibrium conditions only hold almost everywhere. To obtain an equilibrium based on (γ, φ, g), we modify these functions on a set on measure zero so that the equilibrium conditions do hold everywhere. The key part of the modification is to determine the post-trade exposure $g(x)$ for measure zero subsets of X_{otc}, i.e. for the (ω, k) who choose not to participate in the OTC market. For this we rely on the result of the previous subsection: given the post-trade exposures of types who do participate, the post-trade exposure of any other $x \in X_{\mathrm{otc}}$ is uniquely determined.

Step 1: modify trades of $x \in X_{\mathbf{cent}}$. From the planner's problem we know that there exists some constant g_{cent} such that $g(x) = g_{\mathrm{cent}}$, N-almost everywhere for $x \in X_{\mathrm{cent}}$. For any $x \in X_{\mathrm{cent}}$ such that

$g(x) \neq g_{\text{cent}}$ we pick $\gamma(x,x') = 0$ if $(x,x') \notin X^2_{\text{otc}}$, and otherwise we pick $\gamma(x,x')$ that satisfies (21) for all $x' \in X_{\text{otc}}$, given $g(x) = g_{\text{cent}}$. Finally, we let $\varphi(x) = g_{\text{cent}} - \omega(x) - \int \gamma(x,x')\, dN(x' \,|\, \text{otc})$. Because these changes are made for a measure zero set of x, they do not impact the post-trade exposures of any other x', nor do they impact the market-clearing condition in the Walrasian market.

Step 2: modify trades of $(x,x') \in X^2_{\text{otc}}$. We define the following sets:

$$
\begin{aligned}
\Phi &= \left\{(x,x') \in X^2_{\text{otc}} \,:\, (21) \text{ holds for } (x,x')\right\} \\
\Phi(x) &= \left\{x' \in X_{\text{otc}} \,:\, (21) \text{ holds for } (x,x')\right\} \\
\Psi &= \left\{x \in X_{\text{otc}} \,:\, N(\Phi(x) \,|\, \text{otc}) = 1\right\}.
\end{aligned}
$$

One easily shows that $N(\Psi \,|\, \text{otc}) = 1$. Then, we define:

$$
A = \Phi \cap (\Psi \times \Psi)
$$

The set A has measure one because it is the intersection of two sets of measure one. It contains pairs (x,x') with the following properties. First, they together satisfy the optimality condition (21). Second, they each satisfy the optimality condition (21) with almost every other $\hat{x}$. Next we define:

$$
\begin{aligned}
B &= \left\{x \in X_{\text{otc}} \,:\, (x,x') \in A \text{ for some } x'\right\} \\
C &= X_{\text{otc}} \backslash B.
\end{aligned}
$$

The set B has also measure one because $A \subseteq B \times B$. Notice that any $x \in B$ is such that $N(\Phi(x) \,|\, \text{otc}) = 1$.

Our modification goes as follows:

- For all $(x,x') \in A$, the optimality condition (21) holds and so we keep $\gamma(x,x')$ the same.
- For all $(x,x') \in B^2$ but not in A, we modify $\gamma(x,x')$ so that it satisfies (21). Notice that since $N(\Phi(x) \,|\, \text{otc}) = N(\Phi(x') \,|\, \text{otc}) = 1$, these modifications concern a measure zero sets of counterparties for both x and x', and so they do not change the post-trade exposures $g(x)$ or $g(x')$.
- For all (x,x') such that $x \in C$ and $x' \in B$, we pick $\gamma(x,x')$ and $g(x)$ that solves the fixed point problem of Section A.1.2. For any $x' \in B$, this changes the bilateral trades for a measure zero set of counterparties and so does not change $g(x')$.
- For $(x,x') \in C^2$, then we change the bilateral trades so that they satisfy optimality condition (21). For either x or x', this changes the bilateral trades for a measure zero of counterparties, and so does not change $g(x)$ or $g(x')$.

A.2 Proof of Proposition 2

A.2.1 Notations

We start from a finite positive measure N over the set X of type and we add to it a "small" amount ε of a finite signed measure n such that

$$N(X_{\text{otc}}) + \varepsilon\, n(X_{\text{otc}}) > 0 \tag{22}$$

$$N(X_{\text{cent}}) + \varepsilon\, n(X_{\text{cent}}) > 0, \tag{23}$$

for all $\varepsilon > 0$.[24] This restriction ensures that the measure of agents in each market remains strictly positive (the case of weak inequalities can be handled similarly). The planner chooses a measurable function $\gamma : X^2 \to \mathbb{R}$ for bilateral exposures in the constraint set Γ such that:

$$\begin{aligned}
&\gamma(x, x') = 0 \text{ if } x \notin X_{\text{otc}} \\
&\gamma(x, x') + \gamma(x', x) = 0 \\
&- M(x', x) \leq \gamma(x, x') \leq M(x, x'),
\end{aligned}$$

where the function M is uniformly bounded in $(x, x') \in X^2$.

Given any γ and any ε, we define the post-otc exposure by:

$$h(x, \gamma, \varepsilon) = \omega(x) + \int \gamma(x, x')\, \frac{dN(x') + \varepsilon dn(x')}{N(X_{\text{otc}}) + \varepsilon n(X_{\text{otc}})}.$$

Then, the centralized market trade can be defined to "pool exposures," something we already know is optimal:

$$\varphi(x, \gamma, \varepsilon) = \int_{x' \in X_{\text{cent}}} \left[h(x', \gamma, \varepsilon) - h(x, \gamma, \varepsilon)\right] \frac{dN(x') + \varepsilon dn(x')}{N(X_{\text{cent}}) + \varepsilon n(X_{\text{cent}})}$$

if $x \in X_{\text{cent}}$, and $\varphi(x, \gamma, \varepsilon) = 0$ otherwise. Finally, we define the post-trade exposure by:

$$g(x, \gamma, \varepsilon) = h(x, \gamma, \varepsilon) + \varphi(x, \gamma, \varepsilon).$$

Social welfare before participation costs is:

$$W(\gamma, \varepsilon \,|\, N) = \int U\left[g(x, \gamma, \varepsilon)\right] \left(dN(x) + \varepsilon\, dn(x)\right).$$

[24] That is, n is the difference between two finite positive measures, i.e., $n = n^+ - n^-$. This means that are potentially both adding and subtracting to N, which happens for example if we move some agent from one market to some other.

Suppressing N, the planner's problem can be defined in notations that closely follow Milgrom and Segal (2002):

$$\begin{aligned} W^\star(\varepsilon) &= \sup_{\gamma\in\Gamma} W(\gamma,\varepsilon) \\ \Gamma^\star(\varepsilon) &= \{\gamma\in\Gamma \,:\, W(\gamma,\varepsilon) = W^\star(\varepsilon)\}. \end{aligned}$$

We know from our earlier results that the planner's problem has at least a solution, i.e. that $\Gamma^\star(\varepsilon)$ is not empty.

Next, we adapt results in Milgrom and Segal (2002) to show that the right-hand derivative maximizes marginal social value.

A.2.2 Right-hand differentiability

To show that social welfare is right-hand differentiable, we check that the assumptions of Theorem 1 and 3 in Milgrom and Segal (2002) are satisfied in our setting. The main technical difficulty lies in establishing that the right-derivative can be calculated by taking the maximum of the partial derivative over all maximizers. This is a result that Milgrom and Segal (2002) provide in their Corollary 4 for "continuous functions on compact choice sets". Since our choice set is only weakly compact, we must check that the required continuity properties hold in the weak topology.

A useful preliminary result is that:

Lemma 9. *The functions h, φ, g, $\partial h/\partial\varepsilon$, $\partial\varphi/\varepsilon$, $\partial g/\partial\varepsilon$, $\partial^2 h/\partial\varepsilon^2$, $\partial^2\varphi/\varepsilon^2$, $\partial^2 g/\partial\varepsilon^2$, are all uniformly bounded in $(x,\gamma,\varepsilon)\in X\times\Gamma\times[0,\bar\varepsilon]$, for some $\bar\varepsilon$ small enough.*

Proof. For h, g and φ, this follows directly because γ and ω are uniformly bounded. The first derivatives of h, φ, g, can be calculated explicitly as:

$$\begin{aligned} \frac{\partial h}{\partial\varepsilon} =& \int \gamma(x,x')\,\frac{dn(x')N(X_{\text{otc}}) - dN(x')n(X_{\text{otc}})}{\left[N(X_{\text{otc}})+\varepsilon\, n(X_{\text{otc}})\right]^2} \\ \frac{\partial\varphi}{\partial\varepsilon} =& \mathbb{I}_{\{x\in X_{\text{cent}}\}}\Bigg\{\int\left[\frac{\partial h}{\partial\varepsilon}(x',\gamma,\varepsilon)-\frac{\partial h}{\partial\varepsilon}(x,\gamma,\varepsilon)\right]\frac{dN(x')+\varepsilon dn(x')}{N(X_{\text{cent}})+\varepsilon n(X_{\text{cent}})} \\ &+\int\left[h(x',\gamma,\varepsilon)-h(x,\gamma,\varepsilon)\right]\frac{dn(x')N(X_{\text{cent}})-dN(x')n(X_{\text{cent}})}{\left[N(X_{\text{cent}})+\varepsilon\, n(X_{\text{cent}})\right]^2}\Bigg\}, \end{aligned} \tag{24}$$

and $\partial g/\partial\varepsilon = \partial h/\partial\varepsilon + \partial\varphi/\partial\varepsilon$. The result then follows because γ is uniformly bounded, because N and n are finite signed measures, and from our maintained assumption (22) and (23). Indeed, if $N(X_{\text{otc}})>0$ and $N(X_{\text{cent}})>0$, then (22) and (23) ensure that the numerators, $N(X_{\text{otc}})+\varepsilon n(X_{\text{otc}})$ and $N(X_{\text{cent}})+\varepsilon n(X_{\text{cent}})$ are bounded away from zero for small enough ε. If $N(X_{\text{otc}})=0$, then $\partial h/\partial\varepsilon = 0$, and so is evidently uniformly bounded. If $N(X_{\text{cent}})=0$, then the second term in $\partial\varphi/\partial\varepsilon$ is zero, so uniform boundedness obtains as well. Similar arguments imply uniform boundedness of the second derivatives

$$\begin{aligned}
\frac{\partial^2 h}{\partial \varepsilon^2} =& - 2n(X_{\text{otc}}) \int \gamma(x,x') \, \frac{dn(x')N(X_{\text{otc}}) - dN(x')n(X_{\text{otc}})}{[N(X_{\text{otc}}) + \varepsilon \, n(X_{\text{otc}})]^3} \\
\frac{\partial^2 \varphi}{\partial \varepsilon^2} =& \mathbb{I}_{\{x \in X_{\text{cent}}} \bigg\{ \int \left[\frac{\partial^2 h}{\partial \varepsilon^2}(x',\gamma,\varepsilon) - \frac{\partial h}{\partial \varepsilon}(x,\gamma,\varepsilon) \right] \frac{dN(x') + \varepsilon dn(x')}{N(X_{\text{cent}}) + \varepsilon n(X_{\text{cent}})} \\
& 2 \int \left[\frac{\partial h}{\partial \varepsilon}(x',\gamma,\varepsilon) - \frac{\partial h}{\partial \varepsilon}(x,\gamma,\varepsilon) \right] \frac{dn(x')N(X_{\text{cent}}) - dN(x')n(X_{\text{cent}})}{[N(X_{\text{cent}}) + \varepsilon \, n(X_{\text{cent}})]^2} \bigg\} \\
& - 2n(X_{\text{cent}}) \int \left[h(x',\gamma,\varepsilon) - h(x,\gamma,\varepsilon) \right] \frac{dn(x')N(X_{\text{cent}}) - dN(x')n(X_{\text{cent}})}{[N(X_{\text{cent}}) + \varepsilon \, n(X_{\text{cent}})]^3} \bigg\},
\end{aligned}$$

and $\partial^2 g / \partial \varepsilon^2 = \partial^2 h / \partial \varepsilon^2 + \partial^2 \varphi / \partial \varepsilon^2$. □

Uniform boundedness allows us to apply Leibniz' rule to obtain the first and second partial derivative of W with respect to ε:

$$\begin{aligned}
\frac{\partial W}{\partial \varepsilon} =& \int U \left[g(x,\gamma,\varepsilon) \right] \, dn(x) + \int \frac{dU}{dg} \left[g(x,\gamma,\varepsilon) \right] \frac{\partial g}{\partial \varepsilon}(x,\gamma,\varepsilon) \, \left[dN(x) + \varepsilon dn(x) \right] \\
\frac{\partial^2 W}{\partial \varepsilon^2} =& 2 \int \frac{dU}{dg} \left[g(x,\gamma,\varepsilon) \right] \frac{\partial g}{\partial \varepsilon}(x,\gamma,\varepsilon) \, dn(x) + \int \frac{d^2 U}{dg^2} \left[g(x,\gamma,\varepsilon) \right] \left[\frac{\partial g}{\partial \varepsilon}(x,\gamma,\varepsilon) \right]^2 \, \left[dN(x) + \varepsilon dn(x) \right] \\
& + \int \frac{dU}{dg}(x,\gamma,\varepsilon) \frac{\partial^2 g}{\partial \varepsilon^2}(x,\gamma,\varepsilon) \left[dN(x) + \varepsilon dn(x) \right].
\end{aligned}$$

The uniform boundedness properties of Lemma 9 then implies that both $\partial W / \partial \varepsilon$ and $\partial^2 W / \partial \varepsilon^2$ are uniformly bounded in (γ, ε). This further implies that both W and $\partial W / \partial \varepsilon$ are Lipchitz continuous functions of ε, with Lipchitz coefficients that do not depend on γ. Therefore, the equi-continuity and equi-differentiability properties required in Theorem 1 and 3 in Milgrom and Segal (2002) hold. It then follows from these two Theorems that:

Lemma 10. *Given any selection $\gamma^\star(\varepsilon)$ of the maximum correspondence:*

$$\frac{dW^\star}{d\varepsilon}(0^+) = \lim_{\varepsilon \to 0^+} \frac{\partial W}{\partial \varepsilon}(\gamma^\star(\varepsilon), \varepsilon) \geq \max_{\gamma^\star \in \Gamma^\star(0)} \frac{\partial W}{\partial \varepsilon}(\gamma^\star, 0).$$

A.2.3 The right-hand derivative maximizes marginal social value

Next, we show the equality by adapting the argument of Corollary 4 in Milgrom and Segal. To that end consider a sequence $\varepsilon_m \to 0^+$ and some associated sequence of bilateral exposures $\gamma_m^\star \in \Gamma^\star(\varepsilon_m)$. Let $h_m^\star(x) \equiv h(x, \gamma_m^\star, \varepsilon_m)$, $\varphi_m^\star(x) \equiv \varphi(x, \gamma_m^\star, \varepsilon_m)$, $g_m^\star(x) \equiv h_m^\star(x) + \varphi_m^\star(x)$, and $\partial g_m^\star / \partial \varepsilon(x) \equiv \partial g / \partial \varepsilon(x, \gamma_m^\star, \varepsilon_m)$. Similarly, let $h^\star(x) \equiv h(x, \gamma^\star, 0)$, $\varphi^\star(x) \equiv \varphi(x, \gamma^\star, 0)$, $g^\star(x) \equiv g(x, \gamma^\star, 0)$ and $\partial g^\star / \partial \varepsilon = \partial g / \partial \varepsilon(x, \gamma^\star, 0)$.

Weak convergence. Given that that bilateral exposures are uniformly bounded, the Riez Weak Compactness Theorem (Royden and Fitzpatrick, 2010, Section 19.4) allows us to successively extract weakly convergent subsequences, so that we can assume without loss of generality that γ_m converges weakly to some $\gamma^\star$ in $L^2(N \times N)$, $L^2(N \times n)$, $L^2(n \times N)$ and $L^2(n \times n)$, and that the sequences of real numbers $\int U\left[g_m^\star(x)\right] dN(x)$, $\int U\left[g_m^\star(x)\right] dn(x)$ and $\int dU/dg[g_m^\star(x)]\partial g_m^\star/\partial\varepsilon(x)\, dN(x)$ all converge. It then follows from direct calculations using the explicit formula for partial derivatives shown in the proof of Lemma 9 that $h_m^\star(x)$, $\varphi_m^\star(x)$, $g_m^\star(x)$, $\partial g_m^\star/\partial\varepsilon(x)$ converge to $h^\star(x)$, $\varphi^\star(x)$, $g^\star(x)$, $\partial g^\star/\partial\varepsilon(x)$ weakly in $L^2(N)$ and $L^2(n)$.

Strong convergence and asymptotic optimality of post-trade exposures. Given that $g \mapsto \int U\left[g(x)\right] dN(x)$ is strongly continuous and convex, it is weakly upper semi-continuous (see Corollary 2.2 in Eckland and Témam, 1987), which implies that:

$$\int U\left[g^\star(x)\right] dN(x) \geq \lim_{m\to\infty} \int U\left[g_m^\star(x)\right] dN(x). \tag{25}$$

Given any $\gamma \in \Gamma$, the optimality of $\gamma_m^\star$ given the distribution $N + \varepsilon_m n$ implies that:

$$\int U\left[g_m^\star(x)\right] \left[dN(x) + \varepsilon_m dn(x)\right] \geq \int U\left[g(x,\gamma,\varepsilon_m)\right] \left[dN(x) + \varepsilon_m dn(x)\right]. \tag{26}$$

It can be easily checked that, holding γ fixed, $g(x,\gamma,\varepsilon_m) \to g(x,\gamma,0)$ strongly in $L^2(N)$. Given that $g \mapsto \int U\left[g(x)\right] dN(x)$ is strongly continuous, we can go to the limit in the inequality (26) and, combining with (25), we obtain:

$$\int U\left[g^\star(x)\right] dN(x) \geq \lim_{m\to\infty} \int U\left[g_m^\star(x)\right] dN(x) \geq \int U\left[g(x,\gamma,0)\right] dN(x).$$

It follows that $\gamma^\star$ is an optimum for $\varepsilon = 0$, i.e. $\gamma^\star \in \Gamma^\star(0)$. Taking the supremum over $\gamma \in \Gamma$ implies that $\lim_{m\to\infty} U\left[g_m^\star(x)\right] dN(x) = \int U\left[g^\star(x)\right] dN(x)$. Since $U\left[g\right]$ is quadratic and $g_m^\star \to g^\star$ weakly in $L^2(N)$, it follows that $\int \left[g_m^\star(x)\right]^2 dN(x) \to \int \left[g^\star(x)\right]^2 dN(x)$. Therefore $g_m^\star \to g^\star$ weakly in $L^2(N)$, and the $L^2(N)$ norm of g_m converges to that of $g^\star$. It thus follows that $g_m^\star \to g^\star$ strongly in $L^2(N)$.

The derivative maximizes marginal social value. With these results in mind, consider

$$\frac{\partial W}{\partial \varepsilon}(\gamma_m^\star, \varepsilon_m) = \int U\left[g_m^\star(x)\right] dn(x) + \int \frac{dU}{dg}\left[g_m^\star(x)\right] \frac{\partial g_m^\star}{\partial \varepsilon} \left[dN(x) + \varepsilon_m dn(x)\right]. \tag{27}$$

Using the weak upper semi continuity of $g \mapsto \int U\left[g(x)\right] dn(x)$ as above, we obtain that

$$\int U\left[g^\star(x)\right] dn(x) \geq \lim_{m\to\infty} U\left[g_m^\star(x)\right] dn(x). \tag{28}$$

Now recall that $dU/dg[g(x)]$ is linear, that $g_m^\star \to g^\star$ strongly in $L^2(N)$, that $\partial g_m^\star/\partial\varepsilon$ is uniformly bounded and converges weakly in $L^2(N)$ toward $\partial g^\star/\partial\varepsilon$. It thus follows that $dU/dg\left[g_m^\star\right] \partial g_m^\star/\partial\varepsilon$ converges weakly in $L^2(N)$ towards $dU/dg\left[g^\star\right] \partial g^\star/\partial\varepsilon$. Together with (28), this allows us to go to the limit as in (27) and obtain:

$$\frac{\partial W}{\partial \varepsilon}(\gamma^\star, 0) = \int U\left[g^\star(x)\right] dn(x) + \int \frac{dU}{dg}\left[g^\star(x)\right] \frac{\partial g^\star}{\partial \varepsilon} dN(x) \geq \lim_{m\to\infty} \frac{\partial W}{\partial \varepsilon}(\gamma_m^\star, \varepsilon_m).$$

Combining with Lemma 10 we obtain

Lemma 11. *The right-hand derivative maximizes marginal social value:*

$$\frac{dW^\star}{d\varepsilon}(0^+) = \max_{\gamma^\star \in \Gamma^\star(0)} \frac{\partial W}{\partial \varepsilon}(\gamma^\star, 0).$$

The formula shows that the partial derivatives is obtained by "maximizing marginal social value over all maximizers." To understand the economic significance of this maximization, consider some measure N such that investors of certain endowment and capacity type, $(\omega(x), k(x)) = (\omega, k)$, only participate in the Walrasian market. As a result, the planner does not care about the trades of (ω, k) in the OTC market. Formally, the bilateral trades of (ω, k, otc) are indeterminate: any feasible bilateral trades can be part of the planner's solution. Now suppose that the planner moves a "small" measure of (ω, k) in the OTC market. Then, the socially optimal bilateral trades of (ω, k, otc) are no longer indeterminate. The envelope formula above tells that, for a small measure of (ω, k) in the OTC market, the socially optimal bilateral trade are found by maximizing the marginal social value over all maximizers.

A.2.4 An expression of the partial derivative

Lemma 12. *For any* $\gamma \in \Gamma$*:*

$$\begin{aligned}\frac{\partial W}{\partial \varepsilon}(\gamma, 0) = &\int U\left[g(x,\gamma,0)\right] dn(x) \\ &- \int \frac{dU}{dg}\left[\int_{X_{\text{cent}}} g(x',\gamma,0) d\nu(x')\right] \varphi(x,\gamma,0)\, dn(x) \\ &- \int \frac{dU}{dg}\left[g(x',\gamma,0)\right] \gamma(x,x') d\mu(x')\, dn(x) \\ &- \frac{n(X_{\text{otc}})}{2} \int\int \left\{\frac{dU}{dg}\left[g(x',\gamma,0)\right] - \frac{dU}{dg}\left[g(x'',\gamma,0)\right]\right\} \gamma(x',x'') d\mu(x') d\mu(x''),\end{aligned}$$

where

$$d\nu = \frac{dN}{N(X_{\text{cent}})} \text{ if } N(X_{\text{cent}}) > 0 \text{ and } d\nu = \frac{dn}{n(X_{\text{cent}})} \text{ otherwise}$$
$$d\mu = \frac{dN}{N(X_{\text{otc}})} \text{ if } N(X_{\text{otc}}) > 0 \text{ and } d\mu = \frac{dn}{n(X_{\text{otc}})} \text{ otherwise.}$$

Proof. We first calculate:

$$\begin{aligned}\frac{\partial W}{\partial \varepsilon}(x,\gamma,0) &= \int U\left[g(x)\right]\,dn(x) + \int \frac{dU}{dg}\left[g(x)\right]\frac{\partial g}{\partial \varepsilon}(x)\,dN(x)\\ &= \int U\left[g(x)\right]\,dn(x) + \int \frac{dU}{dg}\left[g(x)\right]\frac{\partial h}{\partial \varepsilon}(x)\,dN(x) + \int \frac{dU}{dg}\left[g(x)\right]\frac{\partial \varphi}{\partial \varepsilon}(x)\,dN(x)\end{aligned}$$

where, to simplify notations, we let $g(x) \equiv g(x,\gamma,0)$ and $\partial g/\partial\varepsilon(x) \equiv \partial g/\partial\varepsilon\,(x,\gamma,0)$, and where we used that $g(x) = h(x) + \varphi(x)$. To help with the calculations, define:

$$\begin{aligned}A &\equiv \int U\left[g(x)\right]\,dn(x) \qquad (29)\\ B &\equiv \int \frac{dU}{dg}\left[g(x)\right]\frac{\partial h}{\partial \varepsilon}(x)\,dN(x)\\ C &\equiv \int \frac{dU}{dg}\left[g(x)\right]\frac{\partial \varphi}{\partial \varepsilon}(x)\,dN(x).\end{aligned}$$

An expression for B. We first work on:

$$B = \int \frac{dU}{dg}\left[g(x)\right]\frac{\partial h}{\partial \varepsilon}(x)\,dN(x).$$

Using formula (24) we see that, if $N(X_{\text{otc}}) = 0$, then the second term is zero. If $N(X_{\text{otc}}) > 0$, then the formula (24) implies that

$$\begin{aligned}&\int\int \frac{dU}{dg}\left[g(x)\right]\frac{\partial h}{\partial \varepsilon}(x)\,dN(x)\\ &= \int\int \frac{dU}{dg}\left[g(x)\right]\gamma(x,x')\frac{dn(x')N(X_{\text{otc}}) - dN(x')n(X_{\text{otc}})}{N(X_{\text{otc}})^2}\,dN(x)\\ &= \int\int \frac{dU}{dg}\left[g(x')\right]\gamma(x',x)\frac{dN(x')}{N(X_{\text{otc}})}\,dn(x)\\ &\qquad - n(X_{\text{otc}})\int\int \frac{dU}{dg}\left[g(x)\right]\gamma(x,x')\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2}\\ &= -\int\int \frac{dU}{dg}\left[g(x')\right]\gamma(x,x')\frac{dN(x')}{N(X_{\text{otc}})}\,dn(x)\\ &\qquad - n(X_{\text{otc}})\int\int \frac{dU}{dg}\left[g(x)\right]\gamma(x,x')\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2} \qquad (30)\end{aligned}$$

where the second-to-last equality follows by exchanging the name of variables, namely replacing x by x' in the first term, and the last equality follows by bilateral feasibility, i.e. $\gamma(x',x) = -\gamma(x,x')$. The second term can be simplified as follows:

$$\begin{aligned}
& - n(X_{\text{otc}}) \int\int \frac{dU}{dg}\,[g(x)]\,\gamma(x,x')\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2} \\
=& - \frac{n(X_{\text{otc}})}{2} \int\int \frac{dU}{dg}\,[g(x)]\,\gamma(x,x')\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2} \\
& - \frac{n(X_{\text{otc}})}{2} \int\int \frac{dU}{dg}\,\left[g(x')\right]\,\gamma(x',x)\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2} \\
=& - \frac{n(X_{\text{otc}})}{2} \int\int \frac{dU}{dg}\,[g(x)]\,\gamma(x,x')\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2} + \\
& \frac{n(X_{\text{otc}})}{2} \int\int \frac{dU}{dg}\,\left[g(x')\right]\,\gamma(x,x')\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2} \\
=& - \frac{n(X_{\text{otc}})}{2} \left\{ \frac{dU}{dg}\,[g(x)] - \frac{dU}{dg}\,\left[g(x')\right] \right\} \gamma(x,x')\frac{dN(x)\,dN(x')}{N(X_{\text{otc}})^2},
\end{aligned}$$

where: the first equality follows by breaking the integral into two identical halves and exchanging the name of variables in the second term, replacing x by x' in the second half; the second equality follows by bilateral feasibility $\gamma(x',x) = -\gamma(x,x')$; and the third equality by collecting terms. Taken together we obtain that, if $N(X_{\text{otc}}) > 0$:

$$\begin{aligned}
B =& - \int\int \frac{dU}{dg}\,\left[g(x')\right]\,\gamma(x,x')d\mu(x)\,dn(x) \\
& - \frac{n(X_{\text{otc}})}{2} \int\int \left\{ \frac{dU}{dg}\,\left[g(x')\right] - \frac{dU}{dg}\,\left[g(x'')\right] \right\} \gamma(x',x'')d\mu(x')d\mu(x''),
\end{aligned} \tag{31}$$

where $d\mu(x) = dN(x)/N(X_{\text{otc}})$. If $N(X_{\text{otc}}) = 0$ then $B = 0$. One easily check, using equation (30), that the value $B = 0$ is obtained by setting $d\mu(x) = dn(x)/n(X_{\text{otc}})$ in (31).

An expression for C. Next, we turn to

$$C = \int \frac{dU}{dg}\,[g(x)]\,\frac{\partial\varphi}{\partial\varepsilon}(x)\,dN(x)$$

If $N(X_{\text{cent}}) = 0$, then clearly $C = 0$, since $\varphi(x) = 0$ if $x \notin X_{\text{cent}}$. If $N(X_{\text{cent}}) > 0$, then $g(x)$ constant over $x \in X_{\text{cent}}$ and evidently equal to

$$\bar{g} = \int_{X_{\text{cent}}} g(x')\frac{dN(x')}{N(X_{\text{cent}})},$$

for all $x \in X_{\text{cent}}$. Hence

$$
\begin{aligned}
C =& \frac{dU}{dg}(\bar{g}) \int \frac{\partial \varphi}{\partial \varepsilon}(x)\, dN(x) \\
=& \frac{dU}{dg}(\bar{g}) \int\int_{X_{\text{cent}}^2} \left[\frac{\partial h}{\partial \varepsilon}(x') - \frac{\partial h}{\partial \varepsilon}(x)\right] \frac{dN(x')\, dN(x)}{N(X_{\text{cent}})} \\
&+ \frac{dU}{dg}(\bar{g}) \int\int_{X_{\text{cent}}^2} \left[h(x') - h(x)\right] \frac{dn(x')N(X_{\text{cent}}) - dN(x')n(X_{\text{cent}})}{N(X_{\text{cent}})^2}\, dN(x).
\end{aligned}
$$

Note that, on a symmetric domain, integrals of the form $\int\int [f(x') - f(x)]\, dN(x')dN(x)$ are equal to zero. Hence we are only left with:

$$
\begin{aligned}
C =& \frac{dU}{dg}(\bar{g}) \int\int_{(x,x')\in X_{\text{cent}}^2} \left[h(x') - h(x)\right] dn(x') \frac{dN(x)}{N(X_{\text{cent}})} \\
=& -\frac{dU}{dg}(\bar{g}) \int\int_{(x,x')\in X_{\text{cent}}^2} \left[h(x') - h(x)\right] \frac{dN(x')}{N(X_{\text{cent}})}\, dn(x) \\
=& -\frac{dU}{dg}(\bar{g}) \int \varphi(x)\, dn(x) \qquad (32)
\end{aligned}
$$

where the second equality follows from exchanging the name of variables, replacing x by x', and the third equality follows by definition of $\varphi(x)$. Note that, if $N(X_{\text{cent}}) = 0$ this expression is equal to zero given that $g(x)$ is a constant n-almost everywhere for $x \in X_{\text{cent}}$, and given the market clearing condition. Hence, the above equality holds in all case, $N(X_{\text{cent}}) > 0$ and $N(X_{\text{cent}}) = 0$.

Collecting terms. Adding up A, B and C given in equation (29), (31) and (32), we arrive at the formula of the lemma.

□

The expression for the partial derivative is intuitive. On the first line, we have the utility gain of the new entrant, $U\left[g(x, \gamma, 0)\right]$. On the other lines, we have the costs that entry imposes on incumbents. The second line is the resources cost of Walrasian trades, $\varphi(x)$, evaluated at the marginal utility of incumbents who participate in the Walrasian market. The third line shows similarly the resource costs of OTC trades, $\gamma(x, x')$, evaluated at the marginal utility of incumbents who participate in the OTC market. Finally, the fourth line shows the displacement cost: indeed, the creation of new entrant-incumbent matches displaces matches amongst incumbents. The matching protocol implies that these incumbent-incumbent matches are effectively displaced at random. Therefore, the displacement cost is simply equal to the average surplus destroyed.

Notice that, in the formula, the measure used to calculate average depend on whether there is, under N, positive participation in a market. For instance, if $N(X_{\text{otc}}) > 0$, then the average is calculated based on the conditional distribution of incumbent in the OTC market. If $N(X_{\text{otc}}) = 0$, then the average is calculated based on the conditional distribution of entrant.

To proceed we assume that $N(X_{\text{otc}}) > 0$ and $N(X_{\text{cent}}) > 0$.[25]

A.2.5 Equilibrium exposures maximize the partial derivative

We first show that equilibrium bilateral exposures maximize $\frac{\partial W}{\partial \varepsilon}(\gamma, 0)$ with respect to $\gamma \in \Gamma^\star(0)$. Recall first that:

$$\begin{aligned}\frac{\partial W}{\partial \varepsilon}(\gamma,0) = &\int U\left[g(x)\right]\, dn(x) \\ &- \int \frac{dU}{dg}\left[\int_{X_{\text{cent}}} g(x')\frac{dN(x')}{N(X_{\text{cent}})}\right]\varphi(x,\gamma,0)\, dn(x) \\ &- \int \frac{dU}{dg}\left[g(x')\right]\gamma(x,x')\frac{dN(x')}{N(X_{\text{otc}})}\, dn(x) \\ &- \frac{n(X_{\text{otc}})}{2}\int\int \left\{\frac{dU}{dg}\left[g(x')\right] - \frac{dU}{dg}\left[g(x'')\right]\right\}\gamma(x',x'')\frac{dN(x')dN(x'')}{N(X_{\text{otc}})^2},\end{aligned} \tag{33}$$

where, as before, we let $g(x) = g(x,\gamma,0)$ to simplify notations. Now consider any socially optimal $\gamma \in \Gamma^\star(0)$. Because the first-order conditions hold almost everywhere according to $N(\,\cdot\,|\,X_{\text{otc}}) \times N(\,\cdot\,|\,X_{\text{otc}})$, it follows that the integrand of the last term is equal to

$$\left\{\frac{dU}{dg}\left[g(x')\right] - \frac{dU}{dg}\left[g(x'')\right]\right\}^+ M(x',x'') + \left\{\frac{dU}{dg}\left[g(x')\right] - \frac{dU}{dg}\left[g(x'')\right]\right\}^- M(x'',x'),$$

almost everywhere according to $N(\,\cdot\,|\,X_{\text{otc}}) \times N(\,\cdot\,|\,X_{\text{otc}})$. Therefore, the last term is constant and equal to

$$-n(X_{\text{otc}})\frac{\bar{F}}{2},$$

for any socially optimal $\gamma \in \Gamma^\star(0)$. We also recall that socially optimal post-trade exposures, $g(x)$ are uniquely determined almost everywhere according to N, which implies that:

$$\bar{g} \equiv \int_{X_{\text{cent}}} g(x')\frac{dN(x')}{N(X_{\text{cent}})}$$

is also a constant.

[25] The proposition restricts attention to the case of strictly positive participation in both market. But formula (14) can be also extended to the case of $N(X_{\text{otc}}) = 0$ or $N(X_{\text{cent}}) = 0$, after redefining the equilibrium in an appropriate way. Consider for example participation patterns such that $N(X_{\text{otc}}) = 0$ and $N(X_{\text{cent}}) > 0$, with a perturbation such that $n(X_{\text{otc}}) > 0$. Then, one needs to define equilibrium post-trade exposures when there is a large group of investor, of size $N(X_{\text{cent}} \setminus X_{\text{otc}})$, participating in the Walrasian market, and a "infinitesimal" group of investors, with type distribution $n(x)/n(X_{\text{otc}})$, participating in the OTC and possibly simultaneously in the Walrasian market.

Now let γ denote an equilibrium bilateral exposures, and let $\hat{\gamma}$ denote any socially optimal bilateral exposures. We calculate:

$$
\begin{aligned}
& U\left[g(x)\right] - \frac{dU}{dg}\left[\bar{g}\right]\varphi(x) - \int \frac{dU}{dg}\left[g(x')\right]\gamma(x,x')\frac{dN(x')}{N(X_{\text{otc}})} \\
& - U\left[\hat{g}(x)\right] + \frac{dU}{dg}\left[\bar{g}\right]\hat{\varphi}(x) + \int \frac{dU}{dg}\left[\hat{g}(x')\right]\hat{\gamma}(x,x')\frac{dN(x')}{N(X_{\text{otc}})} \\
\geq & \frac{dU}{dg}\left[g(x)\right]\left[g(x) - \hat{g}(x)\right] - \frac{dU}{dg}\left[\bar{g}\right]\varphi(x) + \frac{dU}{dg}\left[\bar{g}\right]\hat{\varphi}(x) \\
& - \int \frac{dU}{dg}\left[g(x')\right]\left\{\gamma(x,x') - \hat{\gamma}(x,x')\right\}\frac{dN(x')}{N(X_{\text{otc}})} \\
\geq & \left\{\frac{dU}{dg}\left[g(x)\right] - \frac{dU}{dg}\left[\bar{g}\right]\right\}\left[\varphi(x) - \hat{\varphi}(x)\right] \\
& + \int \left\{\frac{dU}{dg}\left[g(x)\right] - \frac{dU}{dg}\left[g(x')\right]\right\}\left\{\gamma(x,x') - \hat{\gamma}(x,x')\right\}\frac{dN(x')}{N(X_{\text{otc}})},
\end{aligned}
$$

where: the first inequality follows by concavity, and because $g(x') = \hat{g}(x')$ almost everywhere according to N; the second inequality follows using $g(x) = h(x) + \varphi(x)$ as well as the explicit expression of $h(x)$ in terms of $\gamma(x,x')$.

Now the first term on the right-side of the last inequality is zero. Indeed if $x \notin X_{\text{cent}}$, then $\hat{\varphi}(x) = 0$. If $x \in X_{\text{cent}}$ then, by construction, $g(x) = \bar{g}$. The second term on the right-side of the last inequality is positive because in equilibrium, the first-order conditions hold everywhere. Hence we have shown that the integrand of (33) is greatest when evaluated at equilibrium bilateral exposures, and the result follows.

A.2.6 The partial derivative in terms of marginal social value.

Our calculations so far show that the partial derivative evaluated at equilibrium bilateral exposures is equal to:

$$
\begin{aligned}
\frac{\partial W}{\partial \varepsilon}(\gamma, 0) = & \int U\left[g\right]\,dn(x) - \int \frac{dU}{dg}\left[\bar{g}\right]\varphi(x)\,dn(x) \\
& - \int \frac{dU}{dg}\left[g(x')\right]\gamma(x,x')\frac{dN(x)}{N(X_{\text{otc}})}\,dn(x) - n(X_{\text{otc}})\frac{\bar{F}}{2} \\
= & \int U\left[g\right]\,dn(x) - \int \frac{dU}{dg}\left[g(x)\right]\varphi(x)\,dn(x) \\
& - \int \frac{dU}{dg}\left[g(x')\right]\gamma(x,x')\frac{dN(x)}{N(X_{\text{otc}})}\,dn(x) - \int \mathbb{I}_{\{x\in X_{\text{otc}}\}}\frac{\bar{F}}{2}\,dn(x),
\end{aligned}
$$

where the second equality follows because either $\varphi(x) = 0$, or $\varphi(x) \neq 0$ and $g(x) = \bar{g}$, which implies that $\frac{dU}{dg}[\bar{g}]\,\varphi(x) = \frac{dU}{dg}[g(x)]\,\varphi(x)$. Next we add and subtract:

$$\begin{aligned}&\int U[\omega(x)]\;dn(x) + \int \frac{dU}{dg}[g(x)]\,\{h(x) - \omega(x)\}\;dn(x)\\ =&\int U[\omega(x)]\;dn(x) + \int\int \frac{dU}{dg}[g(x)]\,\frac{dN(x')}{N(X_{\text{otc}})}\,dn(x)\end{aligned}$$

Collecting terms we obtain:

$$\begin{aligned}\frac{\partial W}{\partial \varepsilon}(\gamma, 0) = &\int U[\omega(x)]\;dn(x)\\ &+\int \left\{U[g(x)] - U[\omega(x)] - \frac{dU}{dg}[g(x)]\,\{g(x) - \omega(x)\}\right\}\,dn(x)\\ &+\int\int \left\{\frac{dU}{dg}[g(x)] - \frac{dU}{dg}\left[g(x')\right]\right\}\gamma(x, x')\frac{dN(x')}{N(X_{\text{otc}})}\,dn(x)\\ &-\int \mathbb{I}_{\{x \in X_{\text{otc}}\}}\frac{\bar{F}}{2}\,dn(x),\end{aligned}$$

and the result follows by observing that equilibrium exposure satisfy the first order conditions everywhere and by re-defining function $\varepsilon \mapsto W^\star\left[N + \varepsilon\left(n^+ - n^-\right)\right]$ to capture the participation costs as well:

$$W^\star\left[N + \varepsilon\left(n^+ - n^-\right)\right](\varepsilon) = \sup_{\gamma \in \Gamma}\int \{U[g(x, \gamma, \varepsilon)] - C[\pi(x)]\}\,\left[dN(x) + \varepsilon\left(dn^+(x) - dn^-(x)\right)\right].$$

A.3 Proof of Lemma 2

The lemma starts with the intuitive result that banks in the Walrasian market have the same post-trade exposure. Given symmetry in participation patterns, this common post-trade exposure has to be $1/2$.

The atom property. Let us guess and verify that post-trade exposures conditional on participating in the OTC market have the "atom property": all banks with the same endowment ω who participate in the OTC markets have the same post-trade exposures, regardless of their capacity. Formally, for each $\omega \in \{0, 1\}$, $g(\omega, k, \text{otc})$ is independent of $k \in [0, 1]$.[26] In addition, we conjecture that risk-sharing is imperfect, that is $\omega = 0$-banks do not equalize their post-trade exposures with $\omega = 1$-banks, $g(0, k, \text{otc}) < 1/2 < g(1, k, \text{otc})$.

If the atom property holds, and if there is imperfect risk sharing, the optimality condition (9) implies that, when an $\omega = 0$ trader is paired with an $\omega = 1$ trader, the $\omega = 0$ trader buys $\max\{k, k'\}$

[26] Notice that this property is assumed to hold for all k. That is, if an ω-bank makes the possibly suboptimal choice to participate in the OTC market, it will equalize its exposure with that other ω banks in the OTC market.

units from $\omega = 1$ trader. When two $\omega = 0$-traders meet, they anticipate the same post-trade exposures and so are indifferent between any quantity in $[-\max\{k, k'\}, \max\{k, k'\}]$. The post-trade exposure of an $\omega = 0$ bank conditional on participating in the OTC market can thus be written:

$$g(0, k, \text{otc}) = \frac{1}{2} \int \gamma(k, k') \, dN(k' \,|\, X_{\text{otc}}) + \frac{1}{2} \int \max\left\{k, k'\right\} dN(k' \mid X_{\text{otc}}), \tag{34}$$

where, with a slight abuse of notation, "$\gamma(k, k')$" stands for "$\gamma\left[(0, k, \text{otc}), (0, k', \text{otc})\right]$." The first term is the net trade with $\omega = 0$ banks. The second term is the net trade with $\omega = 1$ banks. Now if we aggregate (34) over $k \in [k^\star, 1]$, the bilateral feasibility constraint (2) implies that the aggregate net trade between $\omega = 0$ banks is equal to zero. Therefore, if the atom property holds, we must have:

$$g(0, k, \text{otc}) = \int g(0, k', \text{otc}) \, dN(k' \,|\, X_{\text{otc}}) = \frac{1}{2} \mathbb{E}\left[\max\{k', k''\} \,|\, (k', k'') \in X_{\text{otc}}^2\right], \tag{35}$$

where, with a slight abuse of notation, "$k' \in X_{\text{otc}}$" stands for "$(0, k', \text{otc}) \in X_{\text{otc}}$". In words, the post-trade exposure of $\omega = 0$ bank is equal to $\frac{1}{2}$, which is the probability of a meeting between an $\omega = 0$ and an $\omega = 1$ traders, multiplied by $\mathbb{E}\left[\max\{k', k''\} \,|\, (k', k'') \in X_{\text{otc}}^2\right]$, which is the average trade size between an $\omega = 0$ and an $\omega = 1$ trader.

To verify our guess, we need to find bilateral trades between $\omega = 0$-traders, $\gamma(k, k')$, that have two properties. First, they must satisfy the bilateral trading capacity constraint (4) for all (k, k'). Second, the post-trade exposures resulting from these bilateral trades, (35), must be equalized. A natural candidate for these bilateral trades is:

$$\gamma(k, k') = \mathbb{E}\left[\max\{k', k''\} \,|\, k'' \in X_{\text{otc}}\right] - \mathbb{E}\left[\max\{k, k''\} \,|\, k'' \in X_{\text{otc}}\right].$$

That is, when two $\omega = 0$-traders meet, they "swap" the exposures their banks acquired from $\omega = 1$ banks. When aggregated across all possible $\omega = 0$ counterparties, these swaps mechanically equalize exposure of all $\omega = 0$ banks who participate in the OTC market. To see that these swaps also satisfy the bilateral trading capacity constraint, note that

$$\max\{k', k''\} - \max\{k, k''\} = \max\{k' - k'', 0\} - \max\{k - k'', 0\} \in [-k, k'],$$

and so satisfies the bilateral trading capacity constraint.

Similarly, one can show that the atom property holds for $\omega = 1$-banks as well with $g(1, k, \text{otc}) = 1 - g(0, k, \text{otc})$. We have thus verified our conjecture that the post-trade exposures of OTC market participants are independent of k. That small-k (resp. large-k) banks trade like customers (resp. dealers) is proved in the main text.

A.4 Proof of Proposition 3

First, for equation (17) to have a solution, we need that the left-hand side is larger than the right-hand side at $k^\star = 0$, and smaller at $k^\star = 1$, which can be written $0 \leq C(\text{cent}) - C(\text{otc}) \leq \frac{U_{gg}}{36}$.

Now given a solution $k^\star$ to the (17), Lemma 3 implies that banks with $k \leq k^\star$ prefer the Walrasian market to the OTC market, while banks with $k \geq k^\star$ prefer the OTC market to the Walrasian market. To ensure that they prefer participating in some market to stay in autarky, we need that:

$$\max\{\text{MPV}(\omega, k, \text{cent}) - C(\text{cent}), \text{MPV}(\omega, k, \text{otc}) - C(\text{otc})\} \geq 0$$

for all $k \in [0, 1]$. Lemma 3 implies that this function is increasing in k, hence it is positive if and only if it is positive at $k = 0$. But when $k = 0$, the bank prefers the Walrasian market to the OTC market. We thus conclude that banks prefer to participate in some market than to stay in autarky if and only if $C(\text{cent}) \leq \frac{|U_{gg}|}{8}$. Finally, we can rule out participation in two market by setting $C(\text{otc+cent})$ large enough, and we are done.

A.5 Proof of Lemma 5

Direct calculations show that:

$$\frac{d}{d\omega} K(\omega, k, \text{otc}) = |U_{gg}| \, (g - \omega) \left(\frac{dg}{d\omega} - 1 \right)$$
$$\frac{d}{d\omega} F(\omega, k, \text{otc}) = |U_{gg}| k \left[2N(\{\omega' < \omega\} \,|\, X_{\text{otc}}) - 1 \right] \frac{dg}{d\omega},$$

where g and $dg/d\omega$ denote, respectively, the post-trade exposure and its right-derivative for an ω-bank who participates in the OTC market. Using that $g - \omega = k \left[1 - 2N(\{\omega' < \omega\} \,|\, X_{\text{otc}})\right]$, we obtain:

$$\frac{d}{d\omega} \text{MPV}(\omega, k, \text{otc}) = -|U_{gg}| k \left[1 - 2N(\{\omega' < \omega\} \,|\, X_{\text{otc}}) \right] \left[\frac{1}{2} \frac{dg}{d\omega} - 1 \right].$$

Now use that $\frac{d}{d\omega} \text{MPV}(\omega, k, \text{cent}) = -|U_{gg}| \left(\frac{1}{2} - \omega \right)$, and obtain:

$$\frac{d}{d\omega} \left[\text{MPV}(\omega, k, \text{cent}) - \text{MPV}(\omega, k, \text{otc}) \right]$$
$$= -|U_{gg}| \left\{ \frac{1}{2} - \omega + k \left[1 - 2N(\{\omega' < \omega\} \,|\, X_{\text{otc}}) \right] \left[\frac{1}{2} \frac{dg}{d\omega} - 1 \right] \right\}$$

For $\omega < \omega^\star$, then $dg/d\omega = 1$, $N(\{\omega' < \omega\} \,|\, X_{\text{otc}}) = 0$, so that:

$$\frac{d}{d\omega} \left[\text{MPV}(\omega, k, \text{cent}) - \text{MPV}(\omega, k, \text{otc}) \right] = -|U_{gg}| \left\{ \frac{1}{2} - \omega - \frac{k}{2} \right\} < 0,$$

given that $\omega < \omega^\star$ and $\omega^\star + k < \frac{1}{2}$. Next, for $\omega \in \left[\omega^\star, \frac{1}{2}\right]$, $N(\{\omega' < \omega\} \,|\, X_{\text{otc}}) = \frac{\omega - \omega^\star}{1 - 2\omega^\star}$ and $\frac{dg}{d\omega} = 1 - \frac{2k}{1-2\omega^\star}$. Substituting and rearranging, we obtain after a few lines of algebra:

$$\frac{d}{d\omega}\left[\text{MPV}(\omega, k, \text{cent}) - \text{MPV}(\omega, k, \text{otc})\right] = -|U_{gg}| \frac{1-2\omega}{1-2\omega^\star}\left(\frac{1}{2} - \omega^\star - k\right)\left(\frac{1}{2} + \frac{k}{1-2\omega^\star}\right) \leq 0,$$

with equality if $\omega = \frac{1}{2}$.

Using the symmetry, i.e.,

$$\begin{aligned} MPV(1-\omega, k, \text{cent}) &= MPV(\omega, k, \text{cent}) \\ MPV(1-\omega, k, \text{otc}) &= MPV(\omega, k, \text{otc}), \end{aligned}$$

one easily sees that, for $\omega \in \left[\frac{1}{2}, 1\right]$

$$\frac{d}{d\omega}\left[\text{MPV}(\omega, k, \text{cent}) - \text{MPV}(\omega, k, \text{otc})\right] \geq 0,$$

with equality if $\omega = \frac{1}{2}$.

A.6 Proof of Proposition 4

Two conditions. The left side of equation (19) is zero at $\omega = \frac{1}{2} - k$ and it is strictly decreasing for $\omega \in \left[0, \frac{1}{2} - k\right]$. Hence, a necessary and sufficient condition for (19) to have a solution in $\left[0, \frac{1}{2} - k\right]$ is that the left-side evaluated at $\omega = 0$ is greater than the right side:

$$\frac{|U_{gg}|}{8}(1 - 2k) > C(\text{cent}) - C(\text{otc}). \tag{36}$$

Once $\omega^\star$ is found, Lemma 5 ensures that banks in $\omega \in [0, \omega^\star]$ prefer the Walrasian market to the OTC market, and banks in $\omega \in \left[\omega^\star, \frac{1}{2}\right]$ prefer the OTC market to the Walrasian market (incentives are symmetric for $\omega \geq \frac{1}{2}$.

We next need to make sure that banks prefer their choices of market to autarky.

$$\max\left\{\text{MPV}(\omega, k, \text{cent}) - C(\text{cent}), \text{MPV}(\omega, k, \text{otc}) - C(\text{otc})\right\} > 0.$$

Clearly, this function is decreasing since both MPV's are decreasing. Therefore, the condition holds if and only if it holds for $\omega = \frac{1}{2}$.

$$\begin{aligned} \text{MPV}\left(\frac{1}{2}, k, \text{otc}\right) - C(\text{otc}) > 0 \Leftrightarrow\ & \frac{1}{2}F\left(\frac{1}{2}, k, \text{otc}\right) - C(\text{otc}) > 0 \\ \Leftrightarrow\ & \frac{|U_{gg}|k}{4}\left(\frac{1}{2} - \omega^\star - k\right) - C(\text{otc}) > 0. \end{aligned} \tag{37}$$

The first equivalence follows because, for an $\omega = \frac{1}{2}$-bank, the competitive surplus is zero. The second equivalence because the frictional surplus is proportional to the average distance between the post-trade exposure of the $\omega = \frac{1}{2}$ bank, and the post-trade exposures of other banks. Given uniform distribution and symmetry, this average distance is equal to half the distance between the post trade exposure of the $\omega = \frac{1}{2}$ bank, $g\left(\frac{1}{2}, k, \text{otc}\right) = \frac{1}{2}$, and the post-trade exposure of the $\omega^\star$-bank, $g(\omega^\star, k, \text{otc}) = \omega^\star + k$.

Conditions (36) and (37) in the C(otc), C(cent) plane. Condition (36) can be re-written:

$$C(\text{cent}) < C(\text{otc}) + \frac{|U_{gg}|}{8}(1 - 2k),$$

i.e., $C(\text{cent})$ must be below a line with slope one and intercept, $\frac{|U_{gg}|}{8}(1-2k)$. Condition (37) requires some work because it depends on $\omega^\star$, which is itself a function of $C(\text{otc})$ and $C(\text{cent})$. To obtain a workable representation, note first that (37) can be equivalently written as

$$\omega^\star \leq \frac{1}{2} - k - \frac{4C(\text{otc})}{|U_{gg}|k},$$

where $\omega^\star$ solves (19). Since the left-hand side of (19) is strictly decreasing in $\omega < \frac{1}{2} - k$, this is equivalent to requiring that, when evaluated at $\frac{1}{2} - k - \frac{4C(\text{otc})}{|U_{gg}|k}$ the left-hand side of (19) is less than the right-hand side:

$$\frac{|U_{gg}|}{2}\left(k + \frac{4\,C(\text{otc})}{|U_{gg}|k}\right)\frac{4\,C(\text{otc})}{|U_{gg}|k} < C(\text{cent}) - C(\text{otc}) \Leftrightarrow 3\,C(\text{otc}) + \frac{8\,C(\text{otc})^2}{|U_{gg}|k^2} < C(\text{cent}),$$

and we are done.

B Non-exclusive participation with heterogeneous capacities

We now consider our special case without exclusivity. We assume that there are two endowment types, $\omega \in \{0, 1\}$, and that there is a continuous uniform distribution over trading capacities $k \in [0, 1]$.

Proposition 5. *There exists $k^\star$ such that banks with capacities $k < k^\star$ participate in the both markets and banks with capacities $k \geq k^\star$ participate only in the OTC market. The post-trade exposures of banks who participate in both markets are equalized $g(\omega, k, \text{otc+cent}) = 1/2$. The post-trade exposures of $\pi =$ otc-banks are*

$$g(0, k, \text{otc}) = \frac{1}{3}\left(1 + k^\star + (k^\star)^2\right),$$
$$g(1, k, \text{otc}) = \frac{1}{3}\left(2 - k^\star - (k^\star)^2\right).$$

where $k^\star$ solves

$$\frac{|U_{gg}|}{2}\left(\frac{1}{2}-g(0,k,\text{otc})\right)\left(\frac{1}{2}+g(0,k,\text{otc})-(k^\star)^2\right)=C(\text{cent}).$$

$C(\text{cent}) \in \left(0, \frac{5|U_{gg}|}{72}\right)$ *is a necessary condition for this equilibrium to exist (i.e., to make $k^\star$ interior). If, on top of this condition, investors who enter both markets do not have incentive to leave the OTC market, i.e.,*

$$C(\text{otc}) < \frac{|U_{gg}|}{4}\left(\frac{1}{2}-g(0,k,\text{otc})\right)\left(1-(k^\star)^2\right),$$

then this equilibrium exists. In this equilibrium, the frictional surpluses created by a $\pi = \text{otc}$-bank and by a $\pi = \text{otc+cent}$-bank are

$$\begin{aligned} F(0,k,\text{otc}) &= |U_{gg}|\left(\frac{1}{2}-g(0,k,\text{otc})\right)k^\star\mathbb{E}\left[\max\left\{k,k'\right\}|k'<k^\star\right] \\ &\quad +|U_{gg}|\left(\frac{1}{2}-g(0,k,\text{otc})\right)(1-k^\star)\,\mathbb{E}\left[\max\left\{k,k'\right\}|k'>k^\star\right] \end{aligned}$$

and

$$F(0,k,\text{otc+cent}) = |U_{gg}|\left(\frac{1}{2}-g(0,k,\text{otc})\right)(1-k^\star)\,\mathbb{E}\left[\max\left\{k,k'\right\}|k'>k^\star\right],$$

respectively.

Proof. Suppose banks with capacities $k < k^\star$ participate in both markets and banks with capacities $k \geq k^\star$ participate only in the OTC market. As a general result, we know post-trade exposures are uniquely determined. By keeping this in mind, we guess and verify that the post-trade exposures of pure OTC banks are independent of k. We start with the banks with endowment of 0:

$$g(0,k,\text{otc}) = \int_0^{k^\star} k dk' + \frac{1}{2}\int_{k^\star}^{1} \max\left\{k,k'\right\}dk' + \frac{1}{2}\int_{k^\star}^{1}\gamma\left[(0,k),(0,k')\right]dk'.$$

Aggregating over $k \geq k^\star$ and using the bilateral feasibility constraint, the post-trade exposure independent of k must be

$$g(0,k,\text{otc}) = k^\star\mathbb{E}\left[k'|k'>k^\star\right] + \frac{1}{2}(1-k^\star)\,\mathbb{E}\left[\max\left\{k,k'\right\}|k,k'>k^\star\right].$$

Here in this case post-trade exposures turn out very simple: The expression reveals that the post-trade exposure of low endowment banks captures the average trade size with the banks with different post-trade exposures. The first term equals the average trade size with the banks who choose to participate in the centralized market, multiplied by the total mass of those banks; and the second term equals the average trade size with the high endowment banks who participate only in the OTC market, multiplied by the total mass of those banks. Using the fact that k is distributed uniformly on $[0,1]$, the expression in the proposition for $g(0,k,\text{otc})$ obtains. By symmetry, $g(1,k,\text{otc}) = 1 - g(0,k,\text{otc})$.

For brevity, let $g = g(0,k,\text{otc})$. Then, Lemma 1 implies the frictional surplus formulas stated in the proposition and that

$$\begin{aligned}
\text{MPV}(0,k,\text{otc}) &= \frac{|U_{gg}|}{2}g^2 + \frac{|U_{gg}|}{2}\left(\frac{1}{2}-g\right)k^\star \mathbb{E}\left[\max\left\{k,k'\right\}|k'<k^\star\right] \\
&+ \frac{|U_{gg}|}{2}\left(\frac{1}{2}-g\right)(1-k^\star)\,\mathbb{E}\left[\max\left\{k,k'\right\}|k'>k^\star\right] \\
\text{MPV}(0,k,\text{cent}) &= \frac{|U_{gg}|}{8} \\
\text{MPV}(0,k,\text{otc+cent}) &= \frac{|U_{gg}|}{8} + \frac{|U_{gg}|}{2}\left(\frac{1}{2}-g\right)(1-k^\star)\,\mathbb{E}\left[\max\left\{k,k'\right\}|k'>k^\star\right].
\end{aligned}$$

Therefore,

$$\text{MPV}(0,k,\text{otc+cent}) - \text{MPV}(0,k,\text{cent}) = \frac{|U_{gg}|}{2}\left(\frac{1}{2}-g\right)(1-k^\star)\,\mathbb{E}\left[\max\left\{k,k'\right\}|k'>k^\star\right],$$

$$\text{MPV}(0,k,\text{otc+cent}) - \text{MPV}(0,k,\text{otc}) = \frac{|U_{gg}|}{2}\left(\frac{1}{2}-g\right)\left(\frac{1}{2}+g-k^\star\right)\mathbb{E}\left[\max\left\{k,k'\right\}|k'<k^\star\right].$$

The equilibrium conditions are

$$\text{MPV}(0,k^\star,\text{otc+cent}) - C(\text{cent}) - C(\text{otc}) = \text{MPV}(0,k^\star,\text{otc}) - C(\text{otc}) > 0$$

and

$$\text{MPV}(0,k^\star,\text{cent}) - C(\text{cent}) < \text{MPV}(0,k^\star,\text{otc+cent}) - C(\text{cent}) - C(\text{otc}),$$

which imply the conditions in the proposition. □

Now we conduct the same thought experiment as in the exclusive case. The effect on social welfare of changing the marginal OTC bank's entry decision to $\pi' = \text{otc+cent}$ is

$$\Delta W = \frac{1}{2}\left[F(0,k^\star,\text{otc+cent}) - F(0,k^\star,\text{otc})\right].$$

The proposition implies

$$\Delta W = -|U_{gg}|\left(\frac{1}{2} - g(0,k,\text{otc})\right)(k^\star)^2,$$

which is negative; i.e., letting the marginal bank enter the centralized market besides the OTC market decreases the social welfare, which implies that there is too much participation in the centralized market. In the exclusivity case, the interpretation of the opposite normative result was based on the match creation-destruction interpretation. In this case, the process of match creation and destruction will not matter because the both scenarios of our thought experiment, otc and otc+cent, feature trading in the OTC market, and hence, the set of OTC matches are exactly the same. However, the composition of match surpluses is different.

In particular, the total frictional surplus a pure OTC agent creates by trading with other OTC participants is larger than the total frictional surplus an OTC+centralized agent creates. This is because a positive-surplus trade between two pure OTC agents has a higher surplus than a trade between one pure OTC and one OTC+centralized agent (i.e., the size of the surplus in the former case is exactly twice the size of the surplus in the latter case, in our symmetric equilibrium with linear marginal benefit). By noting that the half of matches between pure OTC agents are positive-surplus trades, it is easy to see that letting the marginal bank enter the centralized market besides the OTC market will not change the total surplus it creates with pure OTC agents. However, the value of its earlier matches with OTC+centralized agents will be destroyed as it is also an OTC+centralized agent now. Thus, this constitutes the marginal social loss caused by letting it enter the centralized market besides the OTC market.

In the MPV formula, the marginal bank internalizes only the half of the social loss. As it compares this private loss against a "constant" competitive gain, it ends up having too high an incentive to enter the centralized market.

C Details of the exercise with market resiliency differential

Suppose participation costs induce exclusive equilibrium participation. The expected utility of participating in the OTC market and in the centralized market are

$$-\left(1-\theta\right)e^{-\eta\left\{-C(\mathrm{otc})+U[g(x)]-\int\gamma(x,x')P_{\mathrm{otc}}(x,x')\,dN(x'\,|\,X_{\mathrm{otc}})\right\}}-\theta e^{-\eta\{-C(\mathrm{otc})+U[\omega(x)]\}} \tag{38}$$

and

$$-\left(1-\delta\theta\right)e^{-\eta\{-C(\mathrm{cent})+U[g(x')]-\varphi(x')P_{\mathrm{cent}}\}}-\delta\theta e^{-\eta\{-C(\mathrm{cent})+U[\omega(x')]\}}, \tag{39}$$

respectively, where $U(g)\equiv\mathbb{E}\left[v\right]g-\frac{\eta}{2}\mathbb{V}\left[v\right]g^{2}$. Equivalently, (38) and (39) can be specified as

$$\frac{1-\left(1-\theta\right)e^{-\eta\left\{-C(\mathrm{otc})+U[g(x)]-\int\gamma(x,x')P_{\mathrm{otc}}(x,x')\,dN(x'\,|\,X_{\mathrm{otc}})\right\}}-\theta e^{-\eta\{-C(\mathrm{otc})+U[\omega(x)]\}}}{\eta} \tag{40}$$

and

$$\frac{1-(1-\delta\theta)\,e^{-\eta\{-C(\text{cent})+U[g(x')]-\varphi(x')P_{\text{cent}}\}}-\delta\theta e^{-\eta\{-C(\text{cent})+U[\omega(x')]\}}}{\eta}, \tag{41}$$

respectively.

Fix parameters $\overline{\eta}$ and $\overline{\mathbb{V}}[v]$ and let $\mathbb{V}[v]=\overline{\mathbb{V}}[v]\,\overline{\eta}/\eta$. In the limit as $\eta\to 0$, (40) and (41) become

$$-C\,(\text{otc})+\bar{U}\,[\omega(x)]+(1-\theta)\left\{\bar{U}\,[g(x)]-\bar{U}\,[\omega(x)]-\int\gamma(x,x')P_{\text{otc}}(x,x')\,dN(x'\,|\,X_{\text{otc}})\right\}$$

and

$$-C\,(\text{cent})+\bar{U}\left[\omega(x')\right]+(1-\delta\theta)\left\{\bar{U}\left[g(x')\right]-\bar{U}\left[\omega(x')\right]-\varphi(x')P_{\text{cent}}\right\},$$

respectively, where $\bar{U}(g)\equiv\mathbb{E}\,[v]\,g-\frac{\overline{\eta}}{2}\overline{\mathbb{V}}\,[v]\,g^2$. Treating these as the certainty-equivalents, the formulas for the marginal private and social values naturally derive from our earlier analysis.

References

Tobias Adrian, Nina Boyarchenko, and Or Shachar. Dealer balance sheets and bond liquidity provision. *Journal of Monetary Economics*, 89:92–109, 2017.

Gara Afonso and Ricardo Lagos. Trade dynamics in the market for federal funds. *Econometrica*, 83:263–313, 2015.

David Andolfatto. Business cycles and labor-market search. *American Economic Review*, 86 (1):112–132, 1996.

Andrew G. Atkeson, Andrea L. Eisfeldt, and Pierre-Olivier Weill. The market for otc credit derivatives. Working paper, UCLA, 2013.

Andrew G. Atkeson, Andrea L. Eisfeldt, and Pierre-Olivier Weill. Entry and exit in over-the-counter markets. *Econometrica*, 83:2231–2292, 2015.

Ulf Axelson. Security design with investor private information. *Journal of Finance*, 62(6): 2587–2632, 2007.

Ana Babus. Endogenous intermediation in over-the-counter markets. Working paper, Federal Reserve Bank of Chicago, 2009.

Ana Babus and Kinda Cheryl Hachem. Markets for financial innovation. Working paper, 2019.

Ana Babus and Péter Kondor. Trading and information diffusion in over-the-counter markets. *Econometrica*, 86(5):1727–1769, 2018.

Ana Babus and Cecilia Parlatore. Strategic fragmented markets. Working paper, 2017.

Morten L. Bech and Enghin Atalay. The topology of federal funds market. *Physica A*, 389(22): 5223–5246, 2010.

Zachary Bethune, Bruno Sultanum, and Nicholas Trachter. An information-based theory of financial intermediation. Working paper, 2018.

Bruno Biais and Richard Green. The microstructure of the bond market in the 20th century. Working paper, TSE, HEC, and CMU, 2006.

Giovanni Cespa and Xavier Vives. Exchange competition, entry, and welfare. Working paper, 2018.

Briana Chang and Shengxin Zhang. Endogenous market making and network formation. Working paper, 2016.

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.

Darrell Duffie. Futurization of swaps. Technical report, Bloomberg Government, 2013.

Ivan Eckland and Roger Témam. *Convex Analysis and Variational Problems*. Society for Industrial Mathematics, 1987.

Maryam Farboodi, Gregor Jarosch, and Robert Shimer. The emergence of market structure. Working paper, 2015.

Maryam Farboodi, Gregor Jarosch, and Guido Menzio. Intermediation as rent extraction. Working paper, 2016.

Financial Stability Board. Otc derivatives market reform, twelfth progress report on implementation. Technical report, 2017.

Thomas Gehrig. Intermediation in search markets. *Journal of Economics and Management Strategy*, 2:97–120, 1993.

Athanasios Geromichalos and Lucas Herrenbrueck. Monetary policy, asset prices, and liquidity in over-the-counter markets. *Journal of Money, Credit, and Banking*, 48(1):35–79, 2016.

Vincent Glode and Christian C. Opp. Over-the-counter vs. limit-order markets: The role of traders' expertise. *Review of Financial Studies*, Forthcoming, 2019.

Burton Hollifield, Artem Neklyudov, and Chester Spatt. Bid-ask spreads, trading networks, and the pricing of securitizations. *The Review of Financial Studies*, 30(9):3048–3085, 2017.

Julien Hugonnier, Benjamin Lester, and Pierre-Olivier Weill. Heterogeneity in decentralized asset market. Working paper, 2014.

Kei Kawakami. Welfare consequences of information aggregation and optimal market size. *American Economic Journal: Microeconomics*, 9(4):303–323, 2017.

Tomy Lee and Chaojun Wang. Why trade over the counter? when investors want price discrimination. Working paper, 2018.

Dan Li and Norman Schürhoff. Dealer networks. *Journal of Finance*, 74(1):91–144, 2019.

Wei Li and Zhaogang Song. Dealers as information intermediaries in over-the-counter markets. Working paper, 2019.

Ying Liu, Sebastian Vogel, and Yuan Zhang. Electronic trading in otc markets vs. centralized exchange. Working paper, 2018.

Robert E. Jr. Lucas. Liquidity and interest rates. *Journal of Economic Theory*, 50:237–264, 1990.

Semyon Malamud and Marzena Rostek. Decentralized exchange. *American Economic Review*, 107(11):3320–3362, 2017.

Jianjun Miao. A search model of centralized and decentralized trade. *Review of Economic Dynamics*, 9(1):68–92, 2006.

Paul Milgrom and Ilya Segal. Envelope theorem for arbitrary choice sets. *Econometrica*, 2: 583–601, 2002.

Artem Neklyudov. Bid-ask spreads and the over-the-counter interdealer markets: Core and peripheral dealers. Working paper, CMU, 2012.

Richard Paulson, Margaret Paulsen, William Penner, Scott Breman, and Eric Thompson. Derivatives: Sefs – opening bell sounds. Technical report, Price Waterhouse Coopers Regulatory Briefs, 2013.

Remy Praz. *Essays in Asset Pricing with Search Frictions*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2014.

Lynn Riggs, Esen Onur, David Reiffen, and Haoxiang Zhu. Swap trading after dodd-frank: Evidence from index cds. Working paper CFTC and MIT, 2018.

Marzena Rostek and Ji Hee Yoon. Decentralized markets and derivatives. Working paper, 2018.

Halsey L. Royden and Patrick M. Fitzpatrick. *Real Analysis*. Pearson, 4th edition, 2010.

John Rust and George Hall. Middlemen versus market makers: A theory of competitive exchange. *Journal of Political Economy*, 111(2):pp. 353–403, 2003.

Shouyang Shi. A divisible search model of fiat money. *Econometrica*, 65:75–102, 1997.

Robert Shimer. *Labor Markets and Business Cycles (CREI Lectures in Macroeconomics)*. Princeton University Press, 2010.

Emil Siriwardane. Limited investment capital and credit spreads. *Journal of Finance*, Forthcoming, 2018.

Semih Üslü. Pricing and liquidity in decentralized asset markets. Working paper, University of California at Los Angeles, 2015.

Sebastian Vogel. When to introduce electronic trading platforms in over-the-counter markets? Working paper, 2019.

Chaojun Wang. Core-periphery trading networks. Working paper, 2016.

Abdullah Yavaş. Marketmakers versus matchmakers. *Journal of Financial Intermediation*, 2: 33–58, 1992.

Ji Hee Yoon. Endogenous market structure: Over-the-counter versus exchange trading. Working paper, 2018.