NBER WORKING PAPER SERIES

A THEORY OF PARTICIPATION IN OTC AND CENTRALIZED MARKETS

Jérôme Dugast
Semih Üslü
Pierre-Olivier Weill

Working Paper 25887
http://www.nber.org/papers/w25887

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2019

We would like to thank, for fruitful comments and suggestions, Ana Babus, David Cimon, Darrell Duffie, Selman Erol, Valentin Haddad, Gregor Jarosch, Peter Kondor, Jordi Mondria, Carlos Ramirez, Marzena Rostek, and Zhaogang Song, as well as seminar participants at UCLA, University of Luxembourg, Johns Hopkins University, University of Colorado Boulder, the Federal Reserve Bank of St. Louis, University of Toronto, 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, the Northern Finance Association 2018 Meetings in Charlevoix, the 20th Members Meeting of the Finance Theory Group at CMU, and the Spring 2019 Southwest Search and Matching Workshop. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Jérôme Dugast, Semih Üslü, and Pierre-Olivier Weill. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Theory of Participation in OTC and Centralized Markets
Jérôme Dugast, Semih Üslü, and Pierre-Olivier Weill
NBER Working Paper No. 25887
May 2019
JEL No. G0,G1

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 bank participation in the centralized market can be welfare improving only if banks are mostly heterogeneous in their ability to take large positions in the OTC market, and if participation costs induce banks to trade exclusively in one market. Empirical evidence suggests that these conditions for a welfare improvement are met.

Jérôme Dugast
Université Paris-Dauphine
Place du Maréchal de Lattre de Tassigny
75016 Paris
France
jerome.dugast@dauphine.psl.eu

Semih Üslü
The Johns Hopkins Carey Business School
100 International Drive
Baltimore, MD 21202
semihuslu@jhu.edu

Pierre-Olivier Weill
Department of Economics
University of California, Los Angeles
Bunche Hall 8283
Los Angeles, CA 90095
and NBER
poweill@econ.ucla.edu

Proofs are available at http://www.nber.org/data-appendix/w25887

1 Introduction

Over-the-counter (OTC) markets have a decentralized structure: trade is bilateral, opaque, 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, which they consider to be transparent and resilient trading venues.[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 heterogeneous risk-sharing needs, represented by differences in their initial endowment of risky assets. Second, they are heterogeneous 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

[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.

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 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. Nevertheless, these customers prefer to incur low participation cost for low-quality risk-sharing in the OTC market, instead of high participation cost for 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 the 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 among the infra-marginal banks' traders who are freed up by the marginal bank's exit. 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, where 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 turmoil 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.

[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).

We generalize AEW in two main dimensions. First, while AEW only considered the margin of participation between autarky and the OTC market, we add a new margin: between the OTC and the centralized market. This is clearly essential to analyze our main research question. Second, we allow investors to differ in two dimensions: their risk-sharing needs and their OTC market trading capacities. In contrast, AEW only considered differences in risk-sharing needs. We argue 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, Colliard, Foucault, and Hoffmann (2018), 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. If a marginal bank departs to trade in both markets, then it neither destroys nor creates any trade in the OTC market. However, its social contribution to the infra-marginal banks in the OTC market is minimized due to the fact that it exhausts all its risk-sharing need by accessing the centralized market. Thus, encouraging further non-exclusive participation in the centralized market is always welfare reducing although further exclusive participation in the centralized market can be welfare improving depending on the dominating heterogeneity.

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

[4]The mathematical framework is also more general: 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.

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 (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.

Biais and Mariotti (2005), 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 two exogenously given trading venues.

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 divisible 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." While many non-bank institutions such as hedge funds and insurance companies trade in asset markets in reality, we employ the label bank, for brevity, to refer to all of them. We assume that banks differ in two dimensions: their risk-sharing needs and their OTC market trading technology.

Differences in 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,[5] banks have stronger risk-sharing needs if their initial endowment is very large or small relative to the economy-wide average.

Differences in trading technology are 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, $C(\text{cent}) > C(\text{otc})$. This condition is necessary to obtain equilibria in which banks participate in both markets. It is realistic: indeed, relative to OTC markets, a more centralized market imposes more stringent

[5]Indeed, this is the allocation of assets in any Pareto optimum and, by the First Welfare Theorem, in any competitive equilibrium.

regulatory requirements, involves reengineering of participants' infrastructure to prepare for electronic trading, and requires costly membership to a central clearing party (CCP).[6]

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 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 centralized 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

[6] Securities and Exchange Commission (2011) discusses technological, regulatory, and disclosure costs of trading in centralized Swap Execution Facilities (SEF). In a survey about the incremental costs resulting from the mandate to migrate trade to SEFs, ISDA Research Staff (2011) reports that Buy-Side users "expect to spend an average of \$2.1 million in technology, \$1.3 million amending client/counterparty documentation and \$200 thousand (annually) in additional regulatory reporting." Finally, Duffie, Li, and Lubke (2010) state that "beyond demonstrating its financial strength and providing margin, each CCP member must also contribute capital to a pooled CCP guarantee fund. The guarantee fund is an additional layer of defense, after initial margin, to cover losses stemming from the failure of a member to perform on a cleared derivative."

positive $\gamma(x,x')$ is an outright purchase, and a negative an outright sale. OTC market trades must satisfy two elementary feasibility constraints:

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

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

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.[7]

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

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

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. In practice, there are at least three ways in which regulation made trading more centralized than in a traditional bilateral OTC market. First, trading can be

[7]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 $M(x,x') = k(x')\omega(x')$. 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 trader prior to matching in the 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, $M(x,x') = k(x')$.

multilateral instead of bilateral, for example with the introduction of Swap Execution Facilities. Second, to reduce rent extraction by dealers, information about past-trades can be disseminated broadly, for example after the introduction of the Trade Reporting and Compliance Engine (TRACE) in the corporate bond market. Third, counterparty risk and collateral management can be concentrated in Central Clearing Parties (CCP). To capture these three dimensions of centralization in a simple way, we assume that, in the centralized market, banks trade multilaterally, with price taking, and without capacity constraint.

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}.[8]

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.

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

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

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

[8] The assumption of unrestricted trades in the centralized market facilitates some of the welfare analysis —in particular, it makes it easier to apply the envelope arguments of Proposition 2. In Appendix C, we provide an example in which trades in the centralized market are subject to the trading capacity constraint too, and show that our main results regarding equilibrium participation patterns and welfare go through.

risky asset:

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

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:

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

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.[9]

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.

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.[10] Formally, the objective of a type-x trader who

[9]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.

[10]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. It is also the continuum-population analogue of the Nash-in-Nash solution (i.e., the Nash equilibrium in Nash bargains) used in bilateral oligopoly settings

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:

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

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, he finds it 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. 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:

with interdependent payoffs. See Horn and Wolinsky (1988), Stole and Zwiebel (1996), Crawford and Yurukoglu (2012), and Collard-Wexler, Gowrisankaran, and Lee (2019), among others.

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

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 centralized 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.[11]

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

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

with respect to feasible OTC market trades, γ, feasible centralized 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*

[11]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.

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.[12] 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.

3.2 Equilibrium participation

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

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

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.

[12] Continuing with the example of Footnote 11, 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.

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*

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

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

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

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, $U_g\left[g(x)\right]$. If the bank trades in the centralized market, $\pi(x) \in \{\text{cent}, \text{cent+otc}\}$, then the marginal value is equal to the centralized market price.

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.[13]

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,

[13]In the formula of the frictional surplus, $F(x)$, the positive and negative part of any $z \in \mathbb{R}$ are defined as $z^+ \equiv \max\{z, 0\}$ and $z^- \equiv -\min\{z, 0\}$.

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.[14] 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 (n^+, n^-) is *admissible* if it satisfies the following two natural conditions. First, the participation path $N + \varepsilon\left(n^+ - n^-\right)$ must satisfy (1), so as to remain consistent with the distribution of endowments and trading capacities.[15] Second, the new participation path $N + \varepsilon\left(n^+ - n^-\right)$

[14] 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.

[15] 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.

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 arguments 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:*

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

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

where $\text{MPV}(x)$ 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 values are equalized $\text{MSV}(x) = \text{MPV}(x)$. 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, $\text{MSV}(x) - \text{MPV}(x) = \frac{1}{2}\left[F(x) - \bar{F}\right]$. As in the classical welfare analysis of matching models (see, e.g. Hosios, 1990), the wedge arises 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, $F(x)$.[16] However, 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.

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

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, $\bar{F}/2$.

Now imagine that a policy maker seeks to encourage participation in centralized market by way of tax, subsidy, or regulation. One should expect that such a policy will mostly impact marginal banks, that is, banks which are indifferent between participating in the OTC and participating in the centralized market. This observation motivates us to study the formula of Proposition 2 for the special case of marginal bank reallocation.

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,

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

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

reallocation is

$$\Delta W = -\left[\mathrm{MSV}(x) - \mathrm{MPV}(x)\right] = \frac{1}{2}\left[-F(x) + \bar{F}\right]. \tag{14}$$

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 size, measured by $M(x, x')$, relative to the average,
2. the trades of the marginal bank must create a small enough surplus per quantity traded, measured by $|U_g[g(x)] - U_g[g(x')]|$, 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.[17]

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 (14), 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

[17] Our example in Appendix C shows, in the context of a parametric example, that this formula continues to hold when banks are also subject to some trading capacity constraint in the centralized market. Indeed, this formula follows from a general property of the price-setting mechanism in the centralized market: the centralized market price is equal to the Lagrange multiplier on the resource constraint in the corresponding planner's problem, which implies that the marginal private and social value are equalized.

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 suggests 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 are incomplete because they depend on endogenous outcomes: whether the bank at the margin tends to make large or small trades, whether it has large or small surplus per quantity traded, which depends, in turn, on its endogenous post-trade exposure and whether it is indifferent between exclusive or non-exclusive participation in the centralized 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 $\omega = 1$;[18]

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

[18] 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 not 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, it is useful to start with a simple 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. We argue that, 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. To understand why, let us anticipate on results below and admit that small k banks endogenously trade like customers, while large k banks endogenously trade like dealers. Then, 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.[19] The "min", on the other hand, would restrict the quantity traded in both cases, and so would imply counterfactually small trades 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 U.S. 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 $M(k(x), k(x')) = m\left(\max\{k(x), k(x')\}\right) = \max\{m(k(x)), m(k(x'))\}$ for some increasing and continuous 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, in fact, *equal* to the "max", that is:

$$M(x, x') = \max\left\{k(x), k(x')\right\}. \tag{15}$$

Conjectured equilibrium participation patterns. Next, we guess and verify that, under parameter restrictions to be determined, there exists an equilibrium such that participation only depends on k. That is, banks with the same k make the same participation decision

[19] 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.

regardless of their endowment $\omega \in \{0,1\}$. 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.

Equilibrium trades. We first characterize trading patterns in both the centralized and the OTC market, given any arbitrary participation 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 exposures 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:*

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

- *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.[20] 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:

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

In terms of this notation, the post trade exposure of Lemma 2 can be written $\frac{1}{2}\mathbb{E}\left[\bar{\gamma}(k') \,|\, k' \geq k^\star\right]$. To attain this post-trade exposure, the $\omega = 0$-bank with capacity k must trade with other $\omega = 0$-banks a quantity:

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

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 $g(\omega, k, \text{cent}) = 1/2$.

[20] 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 since $k < k'$, 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.

Hence the competitive and frictional surpluses conditional on participating exclusively in the centralized market are:

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

where U_{gg} is the second derivative of U, which is constant since the certainty equivalent is quadratic. Next, using the formula for post-trade exposures, we find the competitive and frictional surplus conditional on participating exclusively in the OTC market:

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

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.[21] 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,

[21] 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, (15).

given our guess about participation patterns, the marginal bank is indifferent between the centralized and the OTC market if and only if:

$$\frac{|U_{gg}|}{36}(1-k^\star)^2 = C(\text{cent}) - C(\text{otc}). \tag{16}$$

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 $|U_{gg}|$)[22] 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 (16) 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 centralized market, and banks with $k \in [k^\star, 1]$ participate exclusively in the OTC market, if and only if $k^\star$ solve (16), $0 \leq C(\text{otc}) < C(\text{cent}) \leq \min\left\{\frac{|U_{gg}|}{8}, C(\text{otc}) + \frac{|U_{gg}|}{36}\right\}$, 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

[22] A natural interpretation of a smaller $|U_{gg}|$ is that the security is less risky: hence, our model predicts that volume is more likely to concentrate in OTC markets for fixed-income securities than for equity. Another interpretation is that institutional investors are becoming more important players. Indeed, since institutional investors can net out redemptions internally, they have, on a per client basis, less need to trade with others. Our model thus predicts that, when institutional investors become more important players, trade volume concentrates more in OTC markets. This may help explain the historical observations of Biais and Green (2006). See also the earlier work of Babus and Hachem (2019) for similar interpretation of comparative statics.

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 centralized market is welfare improving. This means that there is too much participation in the OTC market and too little in the centralized 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 OTC 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.[23] 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 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:

$$g(\omega, k, \text{otc}) = \omega - kN(\{\omega' < \omega\} \mid X_{\text{otc}}) + kN(\{\omega' > \omega\} \mid X_{\text{otc}}), \tag{17}$$

[23] These participation patterns are similar to the one obtained by Gehrig (1993) and Miao (2006), who consider models in which investors differ in their private valuation, which is conceptually analogous to our assumption that banks differ in their initial endowment. In these papers, however, all investors participate as customers in the OTC market.

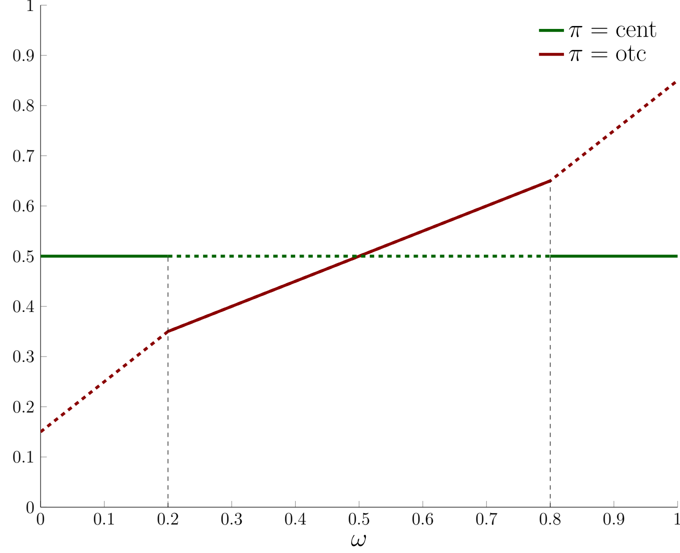

Figure 1: Post-trade exposures as a function of endowment in the example with heterogeneous endowment and exclusive participation. Plain lines show the post-trade exposure on the participation path, that is, conditional on a bank's actual participation decision. Dotted lines show the post-trade exposure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

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

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

and, by symmetry, $N(\{\omega' < \omega\} \,|\, X_{\text{otc}}) = 1 - N(\{1 - \omega' < \omega\} \,|\, X_{\text{otc}})$ for $\omega \geq \frac{1}{2}$. Plugging the expression for the conditional distribution in (17), 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.[24]

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 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.

[24] As before, the formula (17) 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 $N(\,\cdot \,|\, X_{\text{otc}})$) and the one who do not (outside the support). Figure 1 shows $g(\omega, k, \text{otc})$, with plain lines for the banks in the support and with dotted lines for the banks outside the support.

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:

$$\frac{|U_{gg}|}{2}\left(\frac{1}{2}-\omega^{\star}\right)\left(\frac{1}{2}-(\omega^{\star}+k)\right)=C(\text{cent})-C(\text{otc}). \tag{18}$$

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* (18) *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 (14), 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 $g(\omega, k, \text{otc+cent}) = 1/2$. However, at the same time, their average surplus per quantity traded in the OTC market becomes 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 $g = 1/2$.[25] 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.

[25]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.

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:

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

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 exposures. 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:

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

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.

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

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

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.[26] 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,

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

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

[26] When banks make their participation decisions at date $t = 0$ based on the mathematical expectation of their CARA utility over terminal wealth, our previous analysis applies up to a suitable first-order approximation when the probability of turmoil is small. In Appendix D, we follow Van Nieuwerburgh and Veldkamp (2010) and assume that banks make their participation decision based on the expectation of a convex transformation of their CARA utility over terminal wealth, which introduces a preference for early resolution of uncertainty. In this case, the specific "two-stage" preferences that we employ bring back the quadratic certainty-equivalents and our previous analysis regarding participation and welfare applies exactly.

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

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, $\text{MPV}(\omega^\star, k^\star, \text{otc}) - C(\text{otc}) = \text{MPV}(\omega^\star, k^\star, \text{cent}) - C(\text{cent})$. 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:

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

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.

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.

Bruno Biais and Thomas Mariotti. Strategic liquidity supply and security design. *Review of Economic Studies*, 72:615–649, 2005.

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.

Allan Collard-Wexler, Gautam Gowrisankaran, and Robin S. Lee. "nash-in-nash" bargaining: A microfoundation for applied work. *Journal of Political Economy*, 127(1):163–195, 2019.

Jean-Edouard Colliard, Thierry Foucault, and Peter Hoffmann. Inventory management, dealers' connections, and prices in otc markets. Working paper, 2018.

Gregory S. Crawford and Ali Yurukoglu. The welfare effects of bundling in multichannel television markets. *American Economic Review*, 102(2):643–685, 2012.

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, Ada Li, and Theo Lubke. Policy perspectives on otc derivatives market infrastructure. Federal Reserve Bank of New York Staff Report no. 424, March 2010.

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.

Henrick Horn and Asher Wolinsky. Bilateral monopolies and incentives for merger. *RAND Journal of Economics*, 19(3):408–419, 1988.

Arthur J. Hosios. On the efficiency of matching and related models of search and unemployment. *The Review of Economic Studies*, 57:279–298, 1990.

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

ISDA Research Staff. Costs and benefits of mandatory electronic execution requirements for interest rate products. Discussion Papers Series 2, International Swaps and Derivatives Association, Inc., 2011.

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.

John Nachbar. Fixed point theorems. Technical report, Washington University in St. Louis, 2017.

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

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.

Securities and Exchange Commission. Registration and regulation of security-based swap execution facilities. Exchange Act Release 34-63825, Federal Register, 2011.

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.

Lars A. Stole and Jeffrey Zwiebel. Intra-firm bargaining under nonbinding contracts. *The Review of Economic Studies*, 63(3):375–410, 1996.

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

Stijn Van Nieuwerburgh and Laura Veldkamp. Information acquisition and under-diversification. *Review of Economic Studies*, 77:779–805, 2010.

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.