Chapter Twelve

OTC vs. Centralized Markets

The existence of OTC markets may seem puzzling. After all, technologies like centralized limit order books—that closely resemble idealized Walrasian auctions—have been available for decades and would seemingly produce more efficient economic outcomes. The existence of OTC markets becomes even more puzzling when such markets coexist with more centralized exchanges; see, e.g., the discussion of bond trading in Biais and Green [2019]. In such instances, despite the availability of a centralized exchange, we observe some participants *choosing* to trade OTC instead. These observations have given rise to two open questions: Why did OTC trading arrangements emerge, and why have they persisted to this day?

These theoretical questions have important practical implications as well. For instance, inspired by the belief that OTC markets can produce undesirable outcomes, regulators have introduced measures to make such markets more centralized. As a concrete example, the Dodd Frank Wall Street Reform Act of 2010 mandated that swaps be traded on centralized, multilateral platforms called Swap Execution Facilities (SEFs). Similarly, the European Market Infrastructure Regulation (EMIR) regulation adopted by the EU in 2012 aims at enhancing transparency and limiting credit risk by requiring most derivatives trades to be centrally cleared and reported. What are the consequences, intended or otherwise, of migrating trading activity from OTC markets to more centralized venues?

To satisfactorily answer these questions one needs to model the demand of investors for trading in OTC vs. centralized markets. In doing so, one could focus on a variety of distinctions between OTC and centralized markets. These include differences in pre-trade price transparency and competition for order flow; different methods of price determination and settlement in bilateral vs. multilateral trading arrangements; or differences in the way that information is revealed to other market participants after a trade has occurred. In this chapter, we develop a very simple search-theoretic model of venue choice that focuses on differences in the time it takes to locate a counterparty in OTC markets, relative to centralized exchanges.

In this model, we assume that investors can either search for a dealer to trade with in a semi-centralized OTC market or pay an exogenous cost to trade instantly in a centralized market. To endogenize the contact rate in the OTC market, we assume that dealers choose whether to enter the market. Within this model, we characterize the equilibrium and determine the set of investors who find it optimal to trade in OTC vs. centralized markets. We find that concentration of volume in OTC markets can, in some cases, be a self-fulfilling phenomenon. Finally, we conduct a welfare analysis to study whether OTC markets are too large. We find that, relative to the social optimum, investors participate too little in the OTC market and that dealers enter either too little or too much, depending on their bargaining power.

12.1 THE MARKET SETTING

To introduce a meaningful trade-off between OTC and centralized markets we modify the semi-centralized market model of Chapter 4 in two ways. First, we endogenize the contact rate by assuming that there is free entry of dealers in the OTC market. Second, we assume that investors can pay a fixed cost to participate directly in the competitive inter-dealer market. We borrow the assumption of a fixed participation cost from the work of Spulber [1996], Hall and Rust [2003] and Miao [2005]. This device is perhaps the simplest way to create a trade-off between OTC and centralized markets but it also captures an important aspect of real-world markets. For example, Allen and Wittwer [2023] argue that fixed costs of trading are an important determinant of the decision

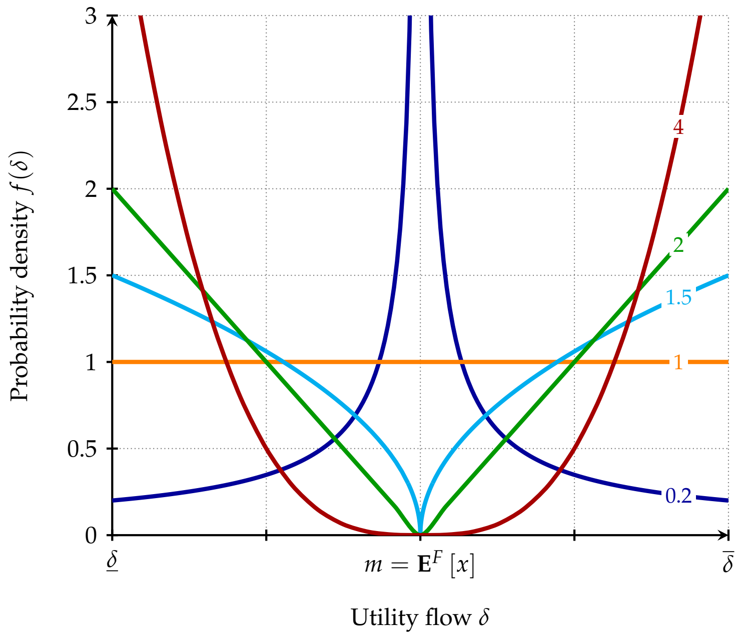

FIGURE 12.1: Symmetric power distributions

This figure illustrates the family of symmetric densities obtained from (12.1) by using the function $g_b(x)$ of (12.2) with $b \in \{0.2, 1, 1.5, 2, 4\}$.

to participate the OTC or the centralized market for Canadian government bonds. In the Notes and References of this chapter, we review papers that have explored other determinants of this tradeoffs.

To simplify the analysis we assume throughout this chapter that the asset supply is $s = \frac{1}{2}$ and that the distribution of investors' utility flows admits a density that is symmetric around the mean

$$m \equiv \mathbf{E}^F[x] = \int_{\underline{\delta}}^{\overline{\delta}} x f(x) dx = \frac{1}{2}\left(\underline{\delta} + \overline{\delta}\right)$$

in the sense that

$$f(\delta) = g\left(|m - \delta|\right) \tag{12.1}$$

for some nonnegative function such that $g(x) = 0$ for $x \geq \underline{\delta} - m$ and which integrates to $\frac{1}{2}$ on $[0, \underline{\delta} - m]$. As illustrated in Figure 12.1, a flexible class of such distributions can be obtained from (12.1) by using the family of power functions defined by

$$g_b(x) \equiv \mathbf{1}_{\{m+x \leq \bar{\delta}\}} \frac{b}{|\mathcal{D}|^b} |2x|^{b-1} \tag{12.2}$$

for some $b > 0$. We will use this particular functional form in our illustrations of the properties of the equilibrium later in the chapter.

The asset is initially allocated randomly across investors. As a result, the density of utility flows among investors with $q \in \{0, 1\}$ assets is

$$\phi_{q0}(\delta) = (qs + (1-q)(1-s))\, f(s).$$

At time $t = 0$, each dealer decides once-and-for-all whether to enter the OTC market by paying a cost $c > 0$. We assume that customers meet dealers more frequently in the OTC market when more dealers enter. Specifically, investors' contact rate with dealers is given by

$$\lambda \equiv \Lambda(\mu_d)$$

where μ_d is the measure of dealers in the OTC market and $\Lambda(\mu)$ is a continuous, strictly increasing, strictly concave, and continuously differentiable function such that $\Lambda(0) = 0$. The economic meaning of concavity is that the meeting technology has decreasing returns to scale: when more dealers enter, the flow rate of meetings *per dealer*, $\Lambda(\mu_d)/\mu_d$, is decreasing.

Investors can trade in two ways. First, they can at any time pay a fixed cost $\kappa > 0$ to trade instantly in the inter-dealer market. Second, they can trade indirectly through dealers as in the semi-centralized market model of Chapter 4: they contact a randomly selected dealer with Poisson intensity λ defined as above and bargain over the terms of trade. As in previous chapters, the bargaining power of the dealer is denoted by $\theta \in [0, 1]$.

We solve for an equilibrium in two steps. First, in Section 12.2 we solve for the equilibrium price and allocation taking as given the search intensity in the OTC market induced by dealer entry. Second, in Section 12.3, we endogenize the search intensity via the dealers' free entry condition. In Section 12.4 we compare the equilibrium outcome to the social optimum.

12.2 EQUILIBRIUM GIVEN DEALER ENTRY

Our symmetry assumption allows us to guess and verify that there exists an equilibrium in which the inter-dealer price P immediately reaches its steady state level at time $t = 0$. Since, in semi-centralized markets, investors decisions may depend on time only through the inter-dealer price, this in turn implies that their value functions are time invariant.

12.2.1 Reservation values

Let $\mathcal{P}_q \subseteq \mathcal{D}$ denote the set of utility flows such that investors who hold q unit of the asset find it optimal to participate in the centralized market. For non owners, $\mathcal{P}_0$ is the set of utility flows such that:

$$V(0,\delta) = V(1,\delta) - (P+\kappa) \tag{12.3}$$

and

$$rV(0,\delta) \geq \gamma \mathbf{E}^F \left[V(0,x) - V(0,\delta)\right] + \lambda_\theta \left(R(\delta) - P\right)^+ \tag{12.4}$$

where $\lambda_\theta \equiv \lambda(1-\theta)$ denotes the bargaining-adjusted rate at which an investor contacts a a dealer, $V(q,\delta)$ denotes the value function of an investor with utility flow $\delta \in \mathcal{D}$ and asset holdings $q \in \{0,1\}$, and

$$R(\delta) \equiv V(1,\delta) - V(0,\delta)$$

denotes her reservation value. Condition (12.3) requires that the value function $V(0,\delta)$ of a non-owner who finds it optimal to participate in the centralized market be equal to the cost of acquiring the asset at price $P+\kappa$ and continuing with the value function $V(1,\delta)$. On the other hand, condition (12.4) requires that the value function of a non owner with utility flow in $\mathcal{P}_0$ be at least as large as the value associated with the decision of searching in the OTC market instead of participating in the centralized market.

Consider a non owner with utility flow $\delta \in \mathcal{P}_0$. Equation (12.3) and the strict positivity of the participation cost imply that

$$V(0,\delta) + (P-\kappa) = V(1,\delta) - 2\kappa < V(1,\delta).$$

Therefore, $\delta \in \mathcal{P}_0 \Rightarrow \delta \notin \mathcal{P}_1$. In words, if a non-owner finds it optimal to participate in the centralized market, then an owner with the same utility flow cannot find it optimal to do the same. This intuitive result means that it cannot be optimal to buy and sell at the same time, as this would incur the strictly positive *round trip* cost 2κ. In addition, it implies that the value function of an owner with $\delta \in \mathcal{P}_0$ equals the value of searching:

$$rV(1,\delta) = \delta + \gamma \mathbf{E}^F\left[V(1,x) - V(1,0)\right] + \lambda_\theta \left(P - R(\delta)\right)^+. \tag{12.5}$$

Combining (12.3), (12.4), and (12.5) shows that that for $\delta \in \mathcal{P}_0$ the reservation value function satisfies

$$P + \kappa = R(\delta) \geq P - \kappa$$

and

$$rR(\delta) \leq \delta + \gamma \mathbf{E}^F\left[R(x) - R(\gamma)\right] + \lambda_\theta \left(P - R(\delta)\right).$$

A similar analysis for utility flows in the sets $\mathcal{P}_1$ and $\mathcal{D}\backslash(\mathcal{P}_0 \cup \mathcal{P}_1)$ allows us to establish the following result.

Lemma 12.1. *The reservation value function of investors is the unique solution to the variational inequality*

$$\begin{aligned} rR(\delta) = \min\Big\{ & r\left(P+\kappa\right), \max\big\{ r\left(P-\kappa\right), \\ & \delta + \gamma \mathbf{E}^F\left[R(x) - R(\delta)\right] + \lambda_\theta (P - R(\delta))\big\}\Big\}. \end{aligned} \tag{12.6}$$

In particular, the reservation value function is bounded from below by $P-\kappa$ and from above by $P+\kappa$, nondecreasing, and absolutely continuous.

Proof. The result follows from contraction mapping arguments similar to those of Chapter 3. See Exercise 12.6 for details. □

Lemma 12.1 shows that the reservation value function solves the usual equation from Chapter 4, adjusted for the option of trading in the centralized market at a fixed cost. Importantly, the reservation value of an investor can never be smaller than $P-\kappa$ or larger than $P+\kappa$ because if it was the investor would then find it optimal to pay the fixed cost and instantly trade in the centralized market at the interdealer market price.

As in previous chapters, in equilibrium, there must be some marginal utility flow $\delta^\star \in \mathcal{D}$ such that $P = R(\delta^\star)$ for otherwise the market would not clear because all investors would be on the same side of both markets. This in turn implies that some investors must find it optimal not to participate in the centralized market and that $\mathcal{D}\backslash\left(\mathcal{P}_0 \cup \mathcal{P}_1\right)$ is an interval $(\delta_s, \delta_b) \ni \delta^\star$. For all utility flows in that set, searching in the OTC market is optimal for both owners and non-owners, meaning that

$$(r+\gamma)R(\delta) = \delta + \gamma \mathbf{E}^F\left[R(x)\right] + \lambda_\theta\left(P - R(\delta)\right).$$

In particular, differentiating both sides with respect to δ shows that on this interval $R'(\delta) = 1/(r+\gamma+\lambda_\theta)$ and combining this result with the fact that the reservation value of a marginal investor is equal to the interdealer price we deduce that

$$R(\delta) = P + \frac{\delta - \delta^\star}{r+\gamma+\lambda_\theta}, \qquad \delta \notin \mathcal{P}_0 \cup \mathcal{P}_1. \tag{12.7}$$

Substituting this expression into the value matching conditions $R(\delta_s) = P - \kappa$ and $R(\delta_b) = P + \kappa$ gives a system of two equations that explicitly characterizes the participation decisions of investors.

Lemma 12.2. *Let*

$$\delta_s \equiv \max\left\{\underline{\delta}, \delta^\star - \kappa\left(r+\gamma+\lambda_\theta\right)\right\}, \tag{12.8a}$$

$$\delta_b \equiv \min\left\{\overline{\delta}, \delta^\star + \kappa\left(r+\gamma+\lambda_\theta\right)\right\}. \tag{12.8b}$$

Then the sets of investors who find it optimal to participate in the centralized market are given by $\mathcal{P}_1 = (\underline{\delta}, \delta_s]$ *and* $\mathcal{P}_0 = [\delta_b, \overline{\delta})$.

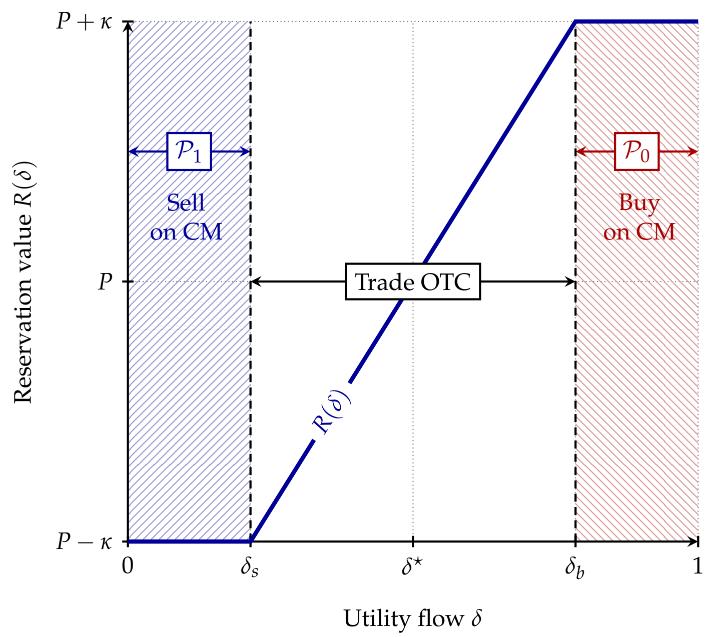

FIGURE 12.2: The reservation value function

This figure illustrates the shape of the reservation value function and the optimal participation decisions of investors in the centralized (CM) and OTC markets. The parameters of the model used to generate the figure are $\mathcal{D} = (0,1)$, $r = 0.05$, $\theta = 0.5$, $\gamma = 1$, $\lambda = 80$, $\kappa = 0.008$, and $P = 10$.

Investors find it optimal to participate in the centralized market if and only if their trade surplus

$$\frac{|\delta - \delta^\star|}{r + \gamma + \lambda_\theta} \tag{12.9}$$

exceeds the fixed participation cost κ. Hence, an owner participates if her utility flow is sufficiently small while a non owner participates if her utility flow is sufficiently large, see Figure 12.2 for an illustration. Furthermore, the explicit expression of the thresholds (δ_s, δ_b) shows that the set of participants in the centralized market naturally shrinks as the participation cost κ increases and as the interest rate r, the switching rate γ, or the bargaining-adjusted contact rate λ_θ increase since any such change results in a lower trade surplus. In particular, if

$$\kappa\left(r+\gamma+\lambda_{\theta}\right) \geq \max\left\{\overline{\delta}-\delta^{\star}, \delta^{\star}-\underline{\delta}\right\}$$

then no investor participates in the centralized because the cost is too large and/or the marginal surplus is too small.

12.2.2 Symmetric equilibrium

Given our assumptions that the asset supply $s = \frac{1}{2}$ and that the probability density of utility flows is symmetric around its mean, it is natural to conjecture that in equilibrium the marginal utility flow that determines the interdealer price is equal to the mean utility flow: $\delta^{\star} = m$. To justify this conjecture we need to verify that markets clear at all times.

INITIAL MARKET CLEARING.

At time $t = 0$, owners with $\delta < \delta_s$ and non owners with $\delta > \delta_b$ pay the participation cost to instantly trade in the centralized market. Keeping in mind that the asset is initially allocated randomly, the market clearing condition can thus be stated as

$$sF(\delta_s) = (1-s)\left(1-F(\delta_b)\right) \tag{12.10}$$

and one can easily verify that this condition holds because $s = 1-s$ and the thresholds (δ_s, δ_b) in (12.8) are symmetric around $m = \delta^{\star}$.

MARKET CLEARING AT $t > 0$.

After the initial date, owners with utility flows in $[\delta_s, \delta^{\star}]$ search to sell in the OTC market and all owners immediately pay κ to sell in the centralized market if they switch to a utility flow $\delta < \delta_s$. Likewise, non owners with utility flows in $[\delta^{\star}, \delta_b]$ search to buy in the OTC market and all non owners pay κ to instantly buy in the centralized market , if they switch to a utility flow $\delta > \delta_b$. Taken together, these observations show that after the initial time the market clearing condition is given by

$$\gamma s F(\delta_s) + \lambda \int_{\delta_s}^{\delta^\star} \phi_{1t}(x)dx = \gamma\,(1-s)\,(1-F(\delta_b)) + \lambda \int_{\delta^\star}^{\delta_b} \phi_{0t}(x)dx.$$

On each side of this condition, the first term collects the flow of investors who trade in the centralized market whereas the second term captures the flow of investors who trade with a dealer in the OTC market. To see that this condition holds when $\delta^\star = m$ observe that the first terms on each side are equal due to the initial market clearing condition (12.10) and that, in a symmetric equilibrium, the densities of utility flows must be such that

$$\phi_{1t}(\delta) = \phi_{0t}(2\delta^\star - \delta), \qquad \delta \in \mathcal{D},$$

to ensure the symmetry of the asset allocation.

We close this section by deriving the equilibrium densities of utility flows among owners and non owners which we will need to study both entry by dealers and the social planner's problem.

Proposition 12.3. *In equilibrium*

$$\phi_{1t}(\delta) = f(\delta) - \phi_{0t}(\delta) = f(\delta) \cdot \begin{cases} 0, & \delta \le \delta_s, \\ \frac{\gamma+\lambda e^{-(\gamma+\lambda)t}}{2(\gamma+\lambda)}, & \delta \in (\delta_s, \delta^\star], \\ 1 - \frac{\gamma+\lambda e^{-(\gamma+\lambda)t}}{2(\gamma+\lambda)}, & \delta \in [\delta^\star, \delta_b), \\ 1, & \delta \ge \delta_b, \end{cases} \tag{12.11}$$

for any value $\lambda \equiv \Lambda(\mu_d)$ of the contact rate with dealers.

Proof. The first equality in (12.11) follows from the fact that the densities must sum up to $f(\delta)$. Since owners with $\delta \le \delta_s$ and non owners with $\delta \ge \delta_b$ instantly trade in the centralized market it clear that $\phi_{1t}(\delta) = f(\delta) - \phi_{0t}(\delta)$ must be equal zero for all $\delta \le \delta_s$ and to $f(\delta)$ for all $\delta \ge \delta_b$.

Over the interval $(\delta_s, \delta^\star]$ of owners who find it optimal to search in the OTC market we have that the density evolves according to

$$\dot{\phi}_{1t}(\delta) = \gamma s f(\delta) - \gamma \phi_{1t}(\delta) - \lambda \phi_{1t}(\delta)$$

where the first two terms on the right account for the inflow and outflow due to type switching, while the third term accounts for the outflow due to trade with dealers. Solving this ODE subject to $\phi_{10}(\delta) = sf(\delta)$ gives the second line in (12.11). By symmetry we then deduce that

$$\phi_{0t}(\delta) = \phi_{1t}(2\delta^\star - \delta) = \frac{\gamma + \lambda e^{-(\gamma+\lambda)t}}{2(\gamma+\lambda)} f(\delta)$$

on the interval $[\delta^\star, \delta_b)$ of non owners who find it optimal to search in the OTC market and the third line in (12.11) now follows from the fact that the densities sum up to the population density $f(\delta)$. □

12.3 FREE ENTRY OF DEALERS

Next, we study the equilibrium arrival rate of meetings with dealers in the OTC market, assuming that dealers can enter at a fixed cost c.

To formulate the free entry condition, recall that when a dealer meets an investor with utility flow $\delta \in \mathcal{D}\backslash(\mathcal{P}_0 \cup \mathcal{P}_1)$, she appropriates a fraction θ of the surplus and thus makes a profit equal to

$$\theta|R(\delta) - P| = \frac{\theta\,|\delta - \delta^\star|}{r + \gamma + (1-\theta)\Lambda(\mu_d)}$$

where the equality follows from (12.7) and the definition of the contact rate. At time $t \geq 0$, the flow rate of such trades *per dealer* is

$$\Gamma_t\,(\delta, \mu_d) \equiv \Lambda(\mu_d)\,\left(\mathbf{1}_{\{\delta<\delta^\star\}}\phi_{1t}(\delta) + \mathbf{1}_{\{\delta>\delta^\star\}}\phi_{0t}(\delta)\right)\Big/\,\mu_d$$

where the numerator represents the aggregate flow of trades and the denominator is the measure of dealers. Therefore, aggregating across utility flows and over time shows that the *free entry condition* is given by:

$$c \geq \int_0^\infty e^{-rt}\left(\int_{\delta_s(\mu_d)}^{\delta_b(\mu_d)} \frac{\theta\,|x-\delta^\star|\,\Gamma_t\,(x,\mu_d)}{r+\gamma+(1-\theta)\,\Lambda(\mu_d)}dx\right)dt$$

with equality if $\mu_d > 0$. Substituting the densities of Proposition 12.3 into this expression, integrating, and rearranging, we obtain that the free entry condition can be equivalently stated as:

$$\frac{C(\lambda)}{\lambda} \geq \theta S\left(\lambda(1-\theta)\right) M(\lambda) \text{ with equality if } \lambda > 0, \tag{12.12}$$

where the function

$$C(x) \equiv c\Lambda^{-1}(x)$$

captures the total entry costs incurred by the dealer sector when the contact rate with investors is equal to λ, and

$$S(x) \equiv \int_0^{\kappa(r+\gamma+x)} \frac{yg(y)\,dy}{r\left(r+\gamma+x\right)},$$
$$M(x) \equiv \frac{r+\gamma}{r+\gamma+x}.$$

The left-hand side of (12.12) is the average cost of search intensity incurred by the dealer sector. The right-hand side is the profit per-unit of search intensity, which is the product of two terms. The first term, $\theta S(\lambda(1-\theta))$, is the present discounted value of dealers' profits when they trade with investors, assuming a constantly random distribution of assets among those investors who find it optimal to search in the OTC market (with utilty flows in (δ_s, δ_b)). The second term, $M(\lambda) \leq 1$, adjusts this present value down to properly account for the fact that assets do not, in fact, remain allocated at random: over time, the OTC market allocates more and more assets to investors with $\delta \geq \delta^\star$, so dealers meet less and less investors in (δ_s, δ_b) with whom they have gain from trade.

The free entry condition (12.12) reveals that there are several forces creating *strategic substitutability* in the entry decisions of dealers. First, when λ increases, the average cost of search intensity, $C(\lambda)/\lambda$, increases as well, making it more costly to enter. Second, $M(\lambda)$ decreases: the fact that the asset is better allocated when λ is larger implies that there are fewer investors who gain from trading with dealers. Third, since investors can meet dealers more frequently going forward, the trade surplus in (12.9) decreases for all utility flows which in trun implies a reduction in the profits that dealers make in every OTC trade.

However, the free entry condition also reveals the existence of a force for *strategic complementarity* stemming from the participation decisions of investors. Namely, when λ increases, investors participate more in the OTC market and less in the centralized market because their trade surplus (12.9) decreases. This extensive margin effect may lead to increased dealer profits, especially if the increase in the contact rate λ induces a large mass of investors to migrate from the centralized to the OTC market.

Proposition 12.4. *There exists an equilibrium. Moreover,*

1. *If the map $\lambda \mapsto \theta S(\lambda(1-\theta))M(\lambda)$ is increasing in some interval of λ then there are cost functions such that multiple equilibria exist;*
2. *If the map $x \mapsto xg(x)$ is strictly decreasing, then the equilibrium is unique and the equilibrium measure of dealers is increasing in θ and κ.*

Proof. To prove the existence claim, note first that if

$$C'(0) = \lim_{\lambda \to 0} \frac{C(\lambda)}{\lambda} \geq \theta S(0)M(0)$$

then there exists an equilibrium in which no dealer enters. Otherwise, using the fact $g(y) = 0$ for all sufficiently large $y > 0$ we deduce that

$$\lim_{\lambda \to \infty} \theta S(\lambda(1-\theta))M(\lambda) = 0 \tag{12.13}$$

and existence follows from an application of the intermediate value theorem because the functions C, S, and M are clearly continuous.

To establish the existence of multiple equilibria, suppose that the condition of the statement holds so that there exist $\lambda_1 < \lambda_2$ such that

$$\theta S(\lambda_1(1-\theta))M(\lambda_1) < \theta S(\lambda_2(1-\theta))M(\lambda_2)$$

Combining this property with (12.13) and the intermediate value theorem shows that there exists $\lambda_3 > \lambda_2$ such that

$$\mathfrak{c} \equiv \theta S(\lambda_1(1-\theta))M(\lambda_1) = \theta S(\lambda_3(1-\theta))M(\lambda_3)$$

and it follows that both λ_1 and λ_3 are equilibria for $C(\lambda) \equiv \mathfrak{c}\lambda$. Finally, a direct calculation using (A.20) shows that

$$S'(x) = \int_0^{\kappa(r+\gamma+x)} \frac{\kappa\,(r+\gamma+x)\,g(\kappa(r+\gamma+x)) - yg(y)}{r+\gamma+x} dy.$$

If the map $y \mapsto yg(y)$ is strictly decreasing, then this derivative is strictly negative. Since the function $M(\lambda)$ is also strictly decreasing this shows that the right handside of (12.12) is strictly decreasing and uniqueness follows. The remaining comparative statics follow from straightforward algebra. □

The first result of the proposition establishes the possibility of multiple equilibria, as illustrated in Figure 12.3. In this numerical example, there are three equilibria. In the first equilibrium $\lambda = 0$ so that no dealer enters the OTC market and the free entry condition holds as a strict inequality. In the intermediate equilibrium, some dealers enter the OTC market and investors are active in both the OTC and centralized markets. Finally, in the equilibrium with the highest value of λ even more dealers enter the OTC market and, as a result, no investors participate in the centralized market.

The existence of multiple equilibria sheds light on certain historical patterns in the data. For example, Biais and Green [2019] document that bond trading migrated from a centralized exchange to an OTC market in the mid 1940s and conjecture that this change may have been due to liquidity spillovers. In particular, they write on page 251: "as institutions and dealers became more prevalent in bond trading, they tipped the balance in favor of the OTC markets." Our analysis offers a more formal, yet nuanced, description of this transition. To see why, note that an increase in the prevalence of dealers is not, in and of itself, sufficient to tip the balance towards OTC trading. Indeed, as more dealers enter, the pool of profitable trades shrinks and transaction costs decline, thus discouraging further entry. Hence, for the balance to tip towards OTC trading, it is necessary that the entry of dealers attracts a large mass of high surplus investors that one could interpret as the large institutions that Biais and Green argue were becoming more prevalent during that time period.

The second result in Proposition 12.4 provides a condition for equilibrium uniqueness. Intuitively, this condition requires that the total surplus utility flow of marginal OTC market participants, $xg(x)$, decreases as participation increases. Under this condition, the strategic complementarity induced by the participation margin is sufficiently weak that multiplicity does not arise. Finally, the third result in the proposition establishes that more dealers find it optimal to enter the OTC market when they have more bargaining power and when participation costs in the centralized market are large.

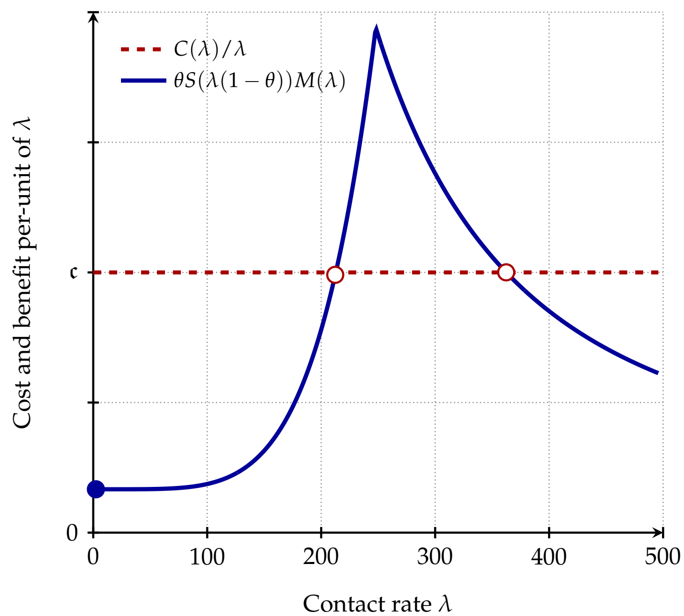

FIGURE 12.3: Multiplicity of entry equilibria

This figure illustrates the existence of multiple equilibria in a model with a linear cost function $C(\lambda) = \mathfrak{c}\lambda$, the same parameters as in figure 12.2, anda probability density of utility flows obtained $g_b(x)$ in (12.2) by setting $b = 7$. The $\circ$ indicate an equilibrium with entry while the $\bullet$ indicates the equilibrium where no dealer enters the OTC market. The kink in the solid curve corresponds to the lowest contact rate such that all investors find it optimal to participate in the OTC market.

12.4 WELFARE ANALYSIS

Given the above description of the equilibrium, we now turn to the normative task of characterizing the socially optimal market structure. We address this problem in two steps. First, we solve for the socially optimal participation and trade decisions at all times taking as given any measure of dealers μ_d or, equivalently, any contact rate λ. Second, we solve for the socially optimal level of dealers entry at the initial time $t = 0$.

12.4.1 Socially optimal participation and trade

The planner chooses feasible trades for all investors but is bound to the same same trading technology. As a result, trades in the OTC market must occur at rate λ while trades in the centralized market can be arranged instantly but incur a fixed cost κ. Using optimal control as in Weill [2007], one can show that the solution of the planner's problem has the same structure as the above equilibrium. Specifically, it is socially optimal for investors to trade in the centralized market if their utility flow satisfies $|\delta - \delta^\star| > \beta$ and to trade OTC otherwise, where $\delta^\star = m$ and β is a participation threshold to be determined.

As a first step, we note that for a given contact rate λ the social welfare function can be written as:

$$\begin{aligned} W(\beta|\lambda) = & \int_0^\infty e^{-rt} \left(\int_{\delta^\star}^{\bar{\delta}} \delta f(\delta) d\delta \right) dt \\ & + \int_0^\infty e^{-rt} \left(\int_{\delta^\star - \beta}^{\delta^\star} \delta \phi_{1t}(\delta) d\delta - \int_{\delta^\star}^{\delta^\star + \beta} \delta \phi_{0t}(\delta) d\delta \right) dt \\ & - \kappa \left(1 + \gamma \int_0^\infty e^{-rt}\, dt \right) \big(sF(\delta^\star - \beta) + (1-s)(1 - F(\delta^\star + \beta)) \big). \end{aligned}$$

where the densities of utility flows $\phi_{1t}(\delta)$ and $\phi_{0t}(\delta)$ among owners and non owners are given as Proposition 12.3 but with the investor thresholds (δ_s, δ_b) replaced by the planer's thresholds.

Taken together, the terms on the first two lines represent the present value of utility flows enjoyed by investors, in deviation from the frictionless allocation which prescribes that the asset be held by investors with $\delta \geq \delta^\star$. Namely, the first line is the present value of utility flows enjoyed by investors in the frictionless allocation. The two terms on the second line account for the fact that the allocation is not frictionless: it adds the utility flows enjoyed by investors with $\delta < \delta^\star$ who actually hold assets but should not, and subtracts the utility flows that are not actually enjoyed because some investors with $\delta \geq \delta^\star$ do not hold assets. The terms on the last line give the present value of the trading costs incurred by investors, at time zero and whenever they switch to utility flows crossing the participation thresholds.

Combining our symmetry assumptions with the explicit expressions of the densities in (12.11), one can show that

$$W\left(\beta|\lambda\right) = W^{\star} - \frac{r+\gamma}{r}\int_0^{\beta}\frac{xg(x)}{r+\gamma+\lambda}dx - \frac{r+\gamma}{r/\kappa}\left(\frac{1}{2} - G(\beta)\right) \tag{12.14}$$

where the constant $W^{\star}$ is the welfare generated by the frictionless allocation and $G(\beta)$ is the cumulative distribution derived from $g(x)$. The second term is the cost of misallocation, while the third term is the cost of participation in the centralized market. Taking derivatives with respect to β yields

$$\frac{\partial W}{\partial \beta}\left(\beta|\lambda\right) = \frac{r+\gamma}{r}g(\beta)\left(\kappa - \beta\left(r+\gamma+\lambda\right)\right).$$

The second term in the bracket is negative because increasing OTC market participation increases misallocation whereas the first term is positive because investors incur less fixed costs as participation in the OTC market increases. Using this first order condition, one can clearly see that the socially optimal threshold is explicitly given by

$$\beta^{\star}(\lambda) \equiv \kappa\left(r+\gamma+\lambda\right) > \kappa\left(r+\gamma+\lambda(1-\theta)\right).$$

This inequality shows that, in equilibrium, investors participate too little in the OTC market whenever $\theta > 0$, i.e., whenever dealers have some bargaining power. The reason is that, when formulating their participation decisions, investors calculate their reservation values based on the bargaining-adjusted contact rate $\lambda\left(1-\theta\right)$ instead of the actual contact rate λ

This finding can be interpreted using the celebrated results of Hosios [1990] who provided a condition on bargaining shares for efficient participation. To intuitively explain this condition, consider a search market between two types of agents, A and B, in respective measures μ_A and μ_B. Assume that the contact rate between types A and B is given by a matching function $\mathcal{M}(\mu_A, \mu_B)$ that is homogenous of degree one. In such a setting, Hosios [1990] showed that, under fairly general conditions, equilibrium participation by agents of type A is efficient if their bargaining power θ_A is equal to the elasticity of the matching function with respect to μ_A:

$$\theta_A \equiv \frac{\mu_A}{\mathcal{M}} \cdot \frac{\partial \mathcal{M}}{\partial \mu_A}.$$

The intuition that is commonly given in the literature is that efficient entry requires that agents of type A capture a share of the surplus equal to their contribution to the matching process, as measured by the marginal increase in the number of matches generated by additional entry of type A agents.

In our model, the condition of Hosios [1990] can be applied as follows. Since dealers have already made their entry decision at the initial time, the only participation decision lies with investors. Therefore, the matching function between dealers and any group of investor of size μ_A is $\lambda\mu_A$ and the elasticity of this function with respect to μ_A is one. Therefore, the reasoning of Hosios [1990] dictates that investors should have all the bargaining power which in turn implies that the participation of investors in the OTC market is inefficient unless the bargaining power of dealers is $\theta = 0$.

12.4.2 Socially optimal entry of dealers

Next, we determine the socially optimal mass of dealers in the OTC market by maximizing the social welfare function (12.14) net of entry costs when investors participate optimally, i.e. by solving

$$\max_{\lambda \geq 0} \left\{ W\left(\beta^\star(\lambda)|\lambda\right) - C(\lambda) \right\}.$$

An application of the envelope theorem shows that at the optimum of this problem we must have

$$C'(\lambda) \geq \frac{\partial W}{\partial \lambda}\left(\beta^\star(\lambda)|\lambda\right) = S(\lambda)M(\lambda) \text{ with equality if } \lambda > 0. \tag{12.15}$$

This necessary condition is similar to the free entry condition (12.12), but with three differences. First, one sees on the left-hand side that the planner uses the marginal cost of search intensity, $C'(\lambda)$, instead of the average cost in the free-entry equilibrium. Equivalently, given that

$$C(\lambda) = c\Lambda^{-1}(\lambda) \tag{12.16}$$

the planner cares about the marginal increase, $\Lambda'(\mu_d)$, in the contact rate with investors whereas an individual dealer's profits are determined by the average meeting rate, $\Lambda(\mu_d)/\mu_d$. Second, the right-hand side is not multiplied by the bargaining power of dealers because the planner considers the full benefit of entry, not just the share of profit that accrues to dealers. Third, participation in the centralized market is chosen in a socially optimal way, so that the argument of the function $S(x)$ on the right handside is λ instead of $\lambda\,(1-\theta)$.

These differences imply that, in general, it is not possible to achieve, at the same time, socially optimal entry of dealers in the OTC market and socially optimal participation of investors in the centralized market. Given socially optimal entry by dealers, attaining socially optimal participation of investors requires setting the dealers' bargaining power to zero. However, if $\theta = 0$, then dealers make no profits ex-post and hence do not enter ex-ante. Likewise, taking the ratio of (12.12) to (12.15) at equality, one obtains that, given socially optimal participation of investors in the centralized market, attaining socially optimal entry of dealers in the OTC market requires

$$\theta = \frac{C(\lambda)}{\lambda C'(\lambda)} = \frac{\mu_d \Lambda'(\mu_d)}{\Lambda(\mu_d)},$$

where the second equality follows from (12.16). This is again the condition of Hosios [1990], but applied to dealers instead of investors. It states that bargaining power of dealers must equal the elasticity of the meeting technology, $\mu_d \Lambda'(\mu_d)/\Lambda(\mu_d)$. This occurs because the entry decisions of dealers are based on the average number of match they create, whereas the planner cares about the marginal increase in the number of matches or, equivalently, about the marginal cost of search intensity.

If the map $x \mapsto xg(x)$ is strictly decreasing, then the free-entry equilibrium is unique and the planner's first-order condition (12.15) is sufficient, which facilitates their comparison.

Proposition 12.5. *Assume that the map $x \mapsto xg(x)$ is strictly decreasing and denote by $(\lambda_P, \lambda_E(\theta))$ the contact rates in, respectively, the planner's problem and the free entry equilibrium when the dealers' bargaining power equals θ. If $\lambda_P > 0$ then:*

1. *There exists $\theta_P > 0$ such that $\lambda_E(\theta) < \lambda_P$ if and only if $\theta < \theta_P$.*
2. *$\lambda_E(\theta)(1-\theta) < \lambda_P$ for all $\theta \in [0,1]$.*

Proof. To establish the first claim recall from Proposition 12.4 that under the stated condition, $S(x)$ is strictly decreasing and $\lambda_E = \lambda_E(\theta)$ is increasing in θ. As the bargaining power of dealers decreases we clearly have

$$\lim_{\theta \to 0} (\lambda_P - \lambda_E(\theta)) = \lambda_P > 0.$$

On the other hand, the convexity of the cost function, the strict decrease of $S(x)$, and the assumption of the statement jointly imply that

$$\lim_{\lambda \to 0} \frac{C(\lambda)}{\lambda} \leq \frac{C(\lambda_P)}{\lambda_P} \leq C'(\lambda_P) = S(\lambda_P)M(\lambda_P) < S(0)M(0).$$

This shows that $\lambda_E(1) > 0$ so that (12.12) holds as an equality when $\theta = 1$. using this property together with the convexity of the cost function then gives

$$C'(\lambda_E(1)) \geq \frac{C(\lambda_E(1))}{\lambda_E(1)} = S(0)M(\lambda_E(1)) > S(\lambda_E(1))M(\lambda_E(1)).$$

This shows that the increasing function $C'(x) - S(x)M(x)$ is strictly positive at $\lambda_E(1)$ and since it is zero at $\lambda_P > 0$ we must have that $\lambda_E(1) > \lambda_P$. Taking stock, we have shown that

$$\lambda_E(0) < \lambda_P < \lambda_E(1)$$

and the desired result now from the intermediate value theorem and the strict monotonicity of the map $\theta \mapsto \lambda_E(\theta)$.

To establish the second claim assume towards a contradiction there exists some $\theta \in [0,1]$ such that $\lambda_E(\theta)(1-\theta) = \lambda_P$ so that the planner's first order condition and the free entry condition can be written as

$$C'(\lambda_P) = S(\lambda_P)M(\lambda_P)$$

and

$$\frac{C(\lambda_P/(1-\theta))}{\lambda_P/(1-\theta)} = \theta S(\lambda_P)M(\lambda_P/(1-\theta)).$$

ratio between the two and rearranging gives

$$\frac{M(\lambda_P/(1-\theta))}{M(\lambda_P)} C'(\lambda_P) \frac{\theta\lambda_P}{1-\theta} = C(\lambda_P/(1-\theta))$$
$$\geq C(\lambda_P) + C'(\lambda_P)\frac{\theta\lambda_P}{1-\theta},$$

which contradicts the fact that the function $M(x)$ is strictly decreasing and thus establishes the desired result. □

The first result in the proposition shows that, to induce efficient entry of dealers in the OTC market, dealers must have strictly positive bargaining power. As discussed above, this generates inefficiently low participation of investors in the centralized market and the second result shows that, irrespective of the dealers' bargaining power, it is not possible to induce socially optimal participation in the centralized market, even at the cost of creating excessive entry of dealers in the OTC market.

12.5 EXERCISES

Exercise 12.6. Consider the equation for reservation values in Lemma 12.1.

1. In the text we verified that this equation has to hold on the set $\mathcal{P}_0$. Verify that it also holds on the sets $\mathcal{P}_1$ and $\mathcal{D}\backslash(\mathcal{P}_0 \cup \mathcal{P}_1)$.
2. Consider the equation obtained by adding $(\gamma + \lambda_\theta)R(\delta)$ to both sides of equation (12.6) and dividing by $r + \gamma + \lambda_\theta$. Show that this modified equation defines a contraction on the set bounded functions on $\mathcal{D}$
3. Show that the reservation value equation admits a unique bounded solution and that is solution is both continuous and nondecreasing.

12.6 NOTES AND REFERENCES

The model in this chapter is original. Some technical comments about our assumptions are in order. To start, the symmetry assumptions regarding the asset supply, the utility flow distribution, and the initial distribution of assets help with tractability: they keep the allocation symmetric and ensure that the equilibrium inter-dealer price immediately adjusts to its steady-state level at time $t = 0$. Similar symmetry assumptions are made by Farboodi, Jarosch, and Shimer [2022] and Liu [2020b,c]. Another important simplification is that entry is a static decision made at the initial time, as otherwise maximizing social welfare would require solving for dealers' optimal *dynamic* entry decisions, which is a much more difficult optimization problem.

While reminiscent of Diamond [1982], the multiplicity result of Proposition 12.4 arises for different reasons. Indeed, in Diamond [1982], multiplicity is driven by increasing returns in the matching technology, which implies that dealers match with investors more frequently when more dealers enter. OTC market models that explore liquidity spillovers induced by similar matching technologies with increasing returns include Vayanos and Wang [2007], Weill [2008], Vayanos and Weill [2008], Sambalaibat [2018], Li and Song [2023], and Coppola, Krishnamurthy, and Xu [2023]. However, this source of strategic complementarity is not present in the model studied here because the matching function $\Lambda(\mu_d)$ is concave: as more dealers enter, they match with investors *less* frequently since $\Lambda(\mu_d)/\mu_d$ is decreasing in μ_d. Instead, the multiplicity in our model is driven by the fact that both sides of the market make an extensive margin decision, so that dealers' entry can encourage investors' participation in OTC markets [see Rocheteau and Wright, 2013, for a related result].

To highlight the distinction between exchanges and OTC trading platforms, the model of this chapter allows investors to choose between a centralized and a semi-centralized trading venue. Several papers in the existing literature allow investors two choose between different venues with the *same* market structure in order to study competition across trading platforms. For example, Pagnotta and Philippon [2018] assume that investors commit ex ante to trade forever in one of two semi-centralized venues, and venues compete by investing in speed and setting fees. In Back et al. [2022], investors are allowed to choose between two semi-centralized venues as their utility type change over time.

Another key, simplifying assumption in this chapter is that investors must pay some exogenous fixed cost in order to participate in the centralized market. The recent literature has developed models to explain why, in reality, it can indeed be more costly to trade in centralized than in OTC markets, based on detailed analysis of price setting mechanisms and informational frictions.

In particular, Rostek and Weretka [2015] argue that centralized market are often thin because they involve a finite (often small) number of participants. In thin markets, investors no longer take prices as given, as in frictionless markets. Instead, they rationally anticipate that their trades have price impact, raising the price when they want to buy and lowering it when they want to sell. Investors who anticipate price impact trade less aggressively than absent frictions, which creates misallocation. In that context, Malamud and Rostek [2017], Babus and Parlatore [2022], Chen and Duffie [2021], Rostek and Yoon [2021], and Wittwer [2021] have shown that breaking down the centralized market into fragmented venues can sometimes improve welfare.

To gain some intuition on the welfare impact of fragmentation, let us briefly elaborate on the paper by Chen and Duffie [2021]. In their work, fragmentation has two opposing effects. On the one hand, smaller marketplaces imply greater price impact, which leads to investors to trade even less aggressively. On the other hand, since investors are able to split their orders across markets, the larger price impact now affects a smaller quantity of infra-marginal units. Chen and Duffie [2021] show that it is possible to choose the level of fragmentation that implements the efficient allocation of asset.

In the above papers, centralized markets are imperfect because of market thinness and price impact. Another crucial consideration when evaluating the merits of centralized vs. OTC markets is the presence and interaction of market structure with information frictions. In particular, since OTC markets traditionally exhibit less pre- and post-trade price transparency, information is potentially transmitted (or leaked) more slowly in OTC markets relative to centralized exchanges. Following Kyle [1985], there is a large literature in finance studying how concerns over information leakage affects investors' trading behavior; see, for example, Burdett and O'Hara [1987], Easley and O'hara [1987], and Seppi [1990].

Recent attempts to study how market structure affects information leakage and trading activity include Hendershott and Madhavan [2015], Kawakami [2010, 2017], and Baldauf and Mollner [Forthcoming]. For example, Kawakami [2010, 2017] considers a setting with ex-ante identical investors who receive an idiosyncratic endowment of a risky asset, and some independent private information about the fundamental value of that asset. He shows that gains from trade are a hump-shaped function of N, the number of investors trading in the market. When N is very small, adverse selection destroys gain from trade, as in Bhattacharya and Spiegel [1991]. When N is large, the Hirshleifer [1971] effect causes the price to become almost perfectly revealing of the true payoff. As a result, there is almost no risk left to share, which also destroys ex-ante gains from trade. Fragmentation emerges endogenously: competitive trading platforms offer investors to trade in small exclusive markets.

The idea that trading frictions can be welfare-improving in the presence of asymmetric information emerges in several other recent papers, including Lester, Shourideh, Venkateswaran, and Zetlin-Jones [2019] and Glode and Opp [2019]. For example, Glode and Opp [2019] show that, when the asymmetry of information between customers is severe, moderately informed agents can be natural intermediaries. They also endogenize market structure in a network formation game where investors choose ex-ante to form chains of bilateral meetings along which trading occurs ex post.

Others have highlighted additional dimensions along which centralized markets can be imperfect (and OTC markets can create value). Vogel [2019] takes the view that, in markets such as the corporate bond market, existing centralized platforms allow investors to quickly query the entire market but only attract a random number of quotes from dealers. The OTC market, by contrast, allows to receive quotes from all dealers, but only sequentially and by incurring non trivial search costs. In Dugast, Üslü, and Weill [2022], trading large quantities requires some capacity, either in the centralized or in the OTC market. These capacities represent investors' access to funding, collateral, or trading expertise. In equilibrium, large capacity investors endogenously play the role of dealers in the OTC market by providing intermediation services to small capacity investors. Intermediate capacity investors sort themselves into the multilateral centralized market. Altermatt and Gottardi [2023] also study sorting but generate a form of trading capacity endogenously by imposing collateral requirements. Farboodi, Jarosch, and Shimer [2022] study a search model in which investors make an ex-ante choice of search intensity. They find that the equilibrium features mixed strategies: investors choose heterogenous intensities. Interestingly, under some conditions on the cost of intensity, some investors endogenously choose an infinite intensity and so resemble the dealers in models of semi-centralized OTC markets.