Chapter Four

Semi-centralized markets

In this chapter we study equilibrium prices, trading volume, and allocations under a market structure that we refer to as a *semi-centralized* OTC market. In a nutshell, we assume that the market is populated by two classes of agents: investors and dealers. Dealers can trade among themselves in a frictionless market but investors can only trade through dealers. As in Chapter 3, investors have to sequentially search for a dealer with whom to trade. After contacting a dealer, we assume that the ensuing bid and ask prices are determined by Nash bargaining within each meeting beween and investor and a dealer. Figure 4.1 illustrates the basic market structure.

The semi-centralized framework is attractive for at least two reasons. First, it captures some of the key features of real-world OTC markets in a simple way. For example, in many OTC markets, such as the U.S. Corporate Bond market discussed in Chapter 2, investors trade almost exclusively through dealers, so that almost all of the trading volume is made up of investor-to-dealer trades and dealer-to-dealer trades. Moreover, the interdealer market is typically more competitive than the investor-dealer market. This was also demonstrated in Chapter 2, when we documented that interdealer trades of corporate bonds exhibit significantly less price dispersion than investor-dealer trades.

The second attractive feature of the semi-centralized market structure is that it turns out to be extremely tractable. As a result, this model offers a simple and transparent account of the effects of search and bargaining frictions on prices and allocations. Furthermore, as we establish in later chapters, this workhorse

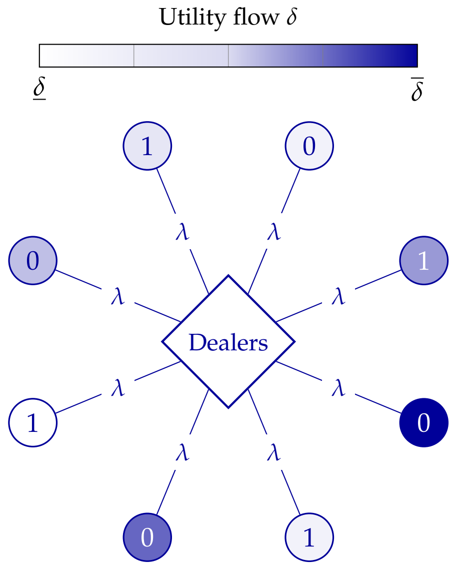

FIGURE 4.1: Semi-centralized OTC market

This figure illustrates the structure of a semi-centralized over-the-counter market in which heterogenous investors holding 0 or 1 unit of the asset meet dealers randomly over time and bargain bilaterally on the terms of trade.

model can be extended and enriched in a variety of ways without sacrificing its tractability.

The Chapter is organized as follows. Section 4.1 first introduces the model. Section 4.2 then derives the frictionless, competitive equilibrium, which is the benchmark against which we evaluate the impact of search and bargaining frictions. Section 4.3 studies asset demand, taking the interdealer price as given. Section 4.4 then determine the interdealer price in equilibrium and compares it with its frictionless counterpart. Section 4.5 derives the empirical implications for various observable outcomes, including liquidity yield spread and trading costs.

4.1 THE MODEL

We consider a continuous-time, infinite-horizon model with a measure one of investors and a measure one of dealers. Investors are risk-neutral, discount the future at rate $r > 0$, and derive utility over two goods: an asset in positive supply $s \in (0,1)$ and a numéraire good. Agents can produce (consume) the numéraire good at constant a marginal cost (utility) normalized to one.

Each investor can hold either 0 or 1 unit of the asset and enjoys a utility flow $\delta_t dt$ from holding one unit of the asset at time $t \geq 0$. The utility flow of each investor is independent from that of any other investor and follows a Markov jump process with arrival rate γ and reset distribution $F(\delta)$ on some open interval $\mathcal{D} = (\underline{\delta}, \overline{\delta})$. Hence, relative to the model of Chapter 3, we make the simplifying assumption that the successive utility flows of an individual investor are independent and identically distributed. Finally, we assume that the distribution of utility flows across the population of investors is initially equal to $F(\delta)$ and thus remains constant since $F(\delta)$ is the stationary distribution of the utility flow process, see Corollary A.41 in Appendix A.

We start from the premise that investors have no way of contacting each other and must trade through dealers. Specifically, we assume that an investor contacts a randomly selected dealer at the jump times of a Poisson process with intensity λ. The contact times are independent of the investor's utility flow process and independent across the population of investors. Conditional on contact, we assume that an investor and a dealer determine the terms of a possible trade by solving a simple Nash bargaining game in which the dealer has bargaining power $\theta \in (0,1)$.

Dealers cannot hold any inventory but are able to trade among themselves instantly, at all times, in a frictionless Walrasian market. Hence, the market structure is *semi-centralized* in the sense that search and bargaining frictions only affect transactions between investors and dealers.

4.2 THE FRICTIONLESS BENCHMARK

As a first step towards the solution to the semi-centralized market, it is helpful to derive prices and allocations in the frictionless benchmark where investors have continuous access to a Walrasian market. Since the distribution of utility flows across investors is constant, one can show that the equilibrium price P in this frictionless market must be constant (a corollary of Exercise 4.6 below). As a result, the optimization problem of an individual investor is given by

$$V(q,\delta) = \sup_{\vartheta \in \mathcal{Q}_q} \mathbf{E}_\delta \left[\int_0^\infty e^{-rt} \left(\vartheta_t \delta_t dt - P d\vartheta_t \right) \right]$$

where $\mathcal{Q}_q$ denotes the set of $\{0,1\}$-valued processes of bounded variations with initial value $\vartheta_0 = q$. Integrating by parts (which is licit here as ϑ_t has bounded variations and $e^{-rt}P$ is bounded and continuous) shows that this problem can be equivalently stated as

$$V(q,\delta) = Pq + \sup_{\vartheta \in \mathcal{Q}_0} \mathbf{E}_\delta \left[\int_0^\infty e^{-rt} q_t \left(\delta_t - rP \right) dt \right].$$

This intuitive formulation shows that the optimal strategy is determined as if, in every small time interval $(t, t+h]$ the investor was choosing whether to purchase a unit of the asset at the beginning of the interval before reselling it at the end. Indeed, if the investor chooses to purchase the asset then she enjoys utility flow $\delta_t h$ but incurs the holding cost

$$\left(P - \frac{P}{1+rh} \right) \simeq rPh,$$

and comparing these flows shows that, taking the price as given, the investor's optimal asset holdings satisfy

$$q_t^\star \begin{cases} = 0 & \text{if } \delta_t < \delta^\star \\ \in \{0,1\} & \text{if } \delta_t = \delta^\star \\ = 1 & \text{if } \delta_t > \delta^\star, \end{cases}$$

where the constant $\delta^\star \equiv rP$ is referred to as the *marginal type*, i.e. the utility flow of an investor who is indifferent between trading or not.

Aggregating this optimality condition across investors and using the fact that the cross-sectional distribution of utility flows is equal to $F(\delta)$ at all times shows that market clearing imposes two conditions. First, the marginal type must lie in the convex hull of the support of $F(\delta)$, for otherwise all investors would be on same side of the market at all times. Second, the marginal type must be a solution to the inclusion

$$s \in [1 - F(\delta^\star), 1 - F(\delta^\star -)]\,, \tag{4.1}$$

where $F(\delta^\star -) \equiv \lim_{\delta \uparrow \delta^\star} F(\delta)$ is the left-limit of the CDF at $\delta^\star$. The left-hand side is the supply while the interval on the right-hand side is the evaluation of the aggregate demand correspondence given P and the induced marginal utility flow $\delta^\star = rP$. In particular, the lower bound of the interval is obtained when only investors with utility flows $\delta > \delta^\star$ purchase the asset while the upper bound is obtained when all investors with utility flows $\delta \geq \delta^\star$, including those at $\delta^\star$ who are indifferent, purchase the asset.

The inclusion (4.1) shows that the marginal utility flow is a quantile of the distribution of utility flows at level $1 - s$. Such a quantile is uniquely defined except in the non generic case where the distribution is equal to $1 - s$ over an open set. Given a solution $\delta^\star$, the corresponding price

$$P = \frac{\delta^\star}{r} = \int_0^\infty e^{-rt} \delta^\star dt \tag{4.2}$$

capitalizes the value of the utility flows derived by a hypothetical investor who remains at the marginal utility type forever.

Next, we solve for the equilibrium asset allocation. Let $\delta^\star$ be a solution to (4.1) and denote by

$$\pi^\star = \mathbf{1}_{\{\Delta F(\delta^\star)=0\}} + \mathbf{1}_{\{\Delta F(\delta^\star)\neq 0\}} \frac{F(\delta^\star) - (1-s)}{\Delta F(\delta^\star)},$$

where $\Delta F(\delta) \equiv F(\delta) - F(\delta -)$, the fraction of investors with utility flow equal to $\delta^\star$ who must hold the asset to clear the market. With this notation, we have that the distribution of utility flows among owners is

$$\begin{aligned}\Psi_1(\delta) &= \mathbf{1}_{\{\delta \geq \delta^\star\}} \left(F(\delta) - F(\delta^\star) + \pi^\star \Delta F(\delta^\star)\right) \\ &= \mathbf{1}_{\{\delta \geq \delta^\star\}} \left(F(\delta) - (1-s)\right) = \max\left\{0, F(\delta) - (1-s)\right\}. \end{aligned} \tag{4.3}$$

Importantly, this expression does not depend on the choice of a marginal type, so that the distribution is uniquely determined even when the marginal type is not. Correspondingly, we can solve for the distribution of utility flows amongst investors who do not hold the asset, $\Psi_0(\delta)$. Indeed, since the measures of holders and non-holder must add up to the population distribution we have

$$\Psi_0(\delta) + \Psi_1(\delta) = F(\delta)$$

and it immediately follows that

$$\Psi_0(\delta) = \min\{F(\delta), 1 - s\}. \tag{4.4}$$

To summarize, a steady state equilibrium of the frictionless benchmark consists in a marginal utility flow $\delta^\star$ satisfying (4.1), a price P satisfying (4.2), and distributions $\Psi_0(\delta)$ and $\Psi_0(\delta)$ satisfying (4.3) and (4.4), respectively.

To close this section, let us briefly discuss trading volume in the frictionless benchmark. Since market participants can buy and sell simultaneously, trading volume is indeterminate in the Walrasian market. Nonetheless, we can pin down the minimum volume that must be traded to support the equilibrium. Indeed, at all times the flow of sell orders must be at least

$$\gamma s F(\delta^\star -),$$

which is the flow of owners who find it strictly optimal to sell because their utility flow has just been reset to some $\delta < \delta^\star$. By the same token, the flow of buy orders must be at least equal to the flow

$$\gamma\,(1-s)\,(1 - F(\delta^\star))\,.$$

of non-owners who find it strictly optimal to buy because their utility flow has just been reset to some $\delta > \delta^\star$, It follows that the minimum volume in the frictionless market is given by

$$v^\star \equiv \gamma \max\left\{sF(\delta^\star -), (1-s)F(\delta^\star)\right\}.$$

Exercise 4.7 below offers a detailed, step-by-step derivation of this lower bound on the equilibrium trading volume.

Example 4.1. To illustrate the frictionless equilibrium consider the two-point distribution from Example 3.1, where

$$F(\delta) = f_L \mathbf{1}_{\{\delta_L \le \delta\}} + f_H \mathbf{1}_{\{\delta_H \le \delta\}}$$

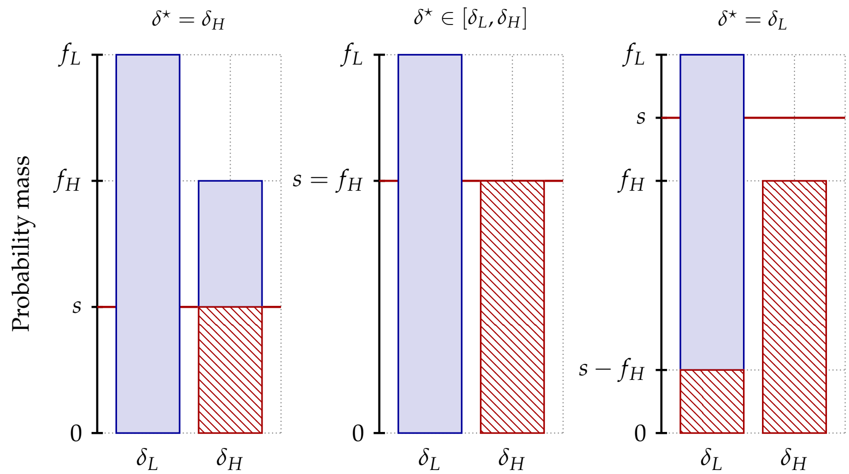

FIGURE 4.2: Frictionless allocation with two types

This figure illustrates the frictionless equilibrium distribution of utility flows among owners (hatched) and non owners (solid) in each of the paremeter configurations of Example 4.1. To produce the figure we fix $f_H = 0.4$ and vary $s \in \{0.2, 0.4, 0.5\}$.

for some $\delta_L < \delta_H$ and $f_L, f_H > 0$ such that $f_L + f_H = 1$. In this specification, three cases may occur depending on the parameter configuration. If the mass of investors with high utility flow exceeds the asset supply, as in the left panel of Figure 4.2, then the marginal type is uniquely defined by $\delta^\star = \delta_H$. In this case, the minimal flow of trades is $\gamma s f_L$ and the equilibrium prescribes that the asset should be held by a randomly selected fraction s/f_H of investors with utility flow δ_H. Alternatively, if the asset supply exceeds the mass of investors with utility flow δ_H, as in the right panel of the figure, then the marginal type is uniquely defined by $\delta^\star = \delta_L$. In this case, the minimal flow of trades is $\gamma(1-s)f_H$ and the equilibrium prescribes that the asset should be held by all investors with utility flow δ_H and a randomly selected fraction $(s - f_H)/f_L$ of investors with utility flow δ_L.

Finally, in the non-generic case in which the asset supply coincides with the mass of investors with utility type δ_H, we have that any $\delta^\star \in [\delta_L, \delta_H]$ defines a marginal type. In that case, the equilibrium allocation naturally requires that the entire supply be in the hands of investors at δ_H and the minimal flow of trades is $\gamma s(1-s)$ regardless of the marginal utlity flow.

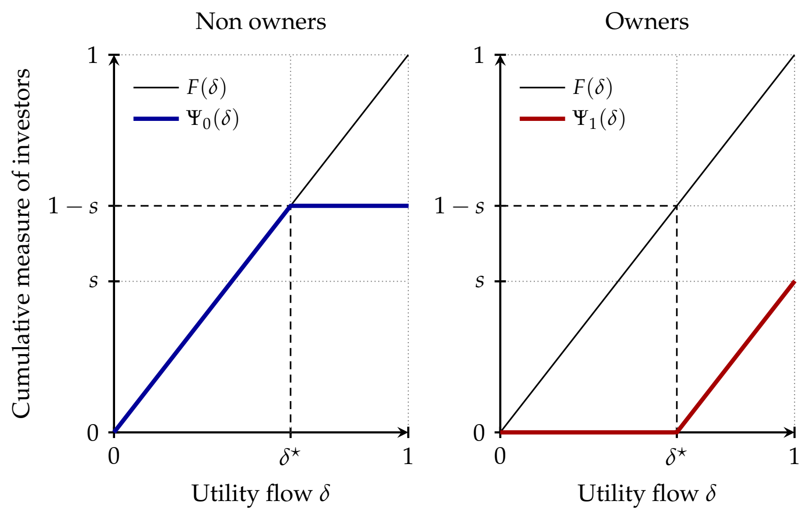

FIGURE 4.3: Frictionless equilibrium with uniform types

This figure illustrates the frictionless equilibrium distribution of utility flows among owners (right panel) and non owners (left panel) in a continuous model with supply $s = 0.4$ and a reset distribution $F(\delta) = \delta$ on $\mathcal{D} = (0, 1)$ so that $\delta^\star = 0.6$.

Example 4.2. Consider the class of models where the distribution of utility flow resets is continuous and strictly increasing. In any such model, the marginal type is uniquely determined by the inverse relation

$$F(\delta^\star) = 1 - s \Leftrightarrow \delta^\star = F^{-1}(1 - s).$$

As a result, the flow of trades per unit of time must exceed $\gamma s(1 - s)$ and the equilibrium allocation prescribes that the asset should only be held by investors with utility flow larger than the marginal type. See Figure 4.3 for an illustration with $s = 0.4$ and a uniform distribution on the unit interval.

4.3 ASSET DEMAND

We now return to the semi-centralized model with search frictions. As a first step towards the characterization of equilibrium, we study the optimization problem of an individual investor who takes as given the interdealer price. We first derive the solution to the bargaining problem between the investor and a dealer, taking as given the relevant outside options. Then, we use the solution to the bargaining problem as building block to derive the HJB equation associated with the investor's dynamic optimization problem. We provide closed form solution to the resulting equation, which in turn delivers an explicit characterization of the optimal trading strategy as a function of the interdealer price and the investor's utility flow.

4.3.1 Bargaining between an investor and a dealer

Consider a meeting between a dealer and an onwer with utility flow $\delta \in \mathcal{D}$. Since investors cannot hold more than one unit, the only possible trade is that the investor sells her asset to the dealer at some bid price $B(\delta)$, which will be determined as the outcome of a Nash bargaining game. If the sale occurs, then the investor becomes a non-owner and receives the continuation utility

$$B(\delta) + V(0, \delta),$$

where the function $V(q, \delta)$ denotes the maximal attainable utility of an investor in state (q, δ). Otherwise, the investor remains an owner and receives the continuation utility $V(1, \delta)$. Hence, the investor's net utility from trading with the dealer is given by

$$B(\delta) + V(0, \delta) - V(1, \delta) = B(\delta) - R(\delta)$$

where, as in Chapter 3, $R(\delta)$ denotes the reservation value of the investor as a function of her utility flow.

On the other side of the transaction, the dealer's net utility from the trade is $P - B(\delta)$. It thus follows that the total surplus from the trade is simply given by the difference between the interdealer price P and the reservation value of the investor, i.e.,

$$(P - B(\delta)) + (B(\delta) - R(\delta)) = P - R(\delta).$$

If the surplus is negative, then there are no gains from trade and the investor simply walks away. If there are gains from trade, the investor and the dealer split the surplus according to the solution of the generalized Nash bargaining program:

$$\max_{B\in[R(\delta),P]} (P-B)^{\theta}\,(B-R(\delta))^{1-\theta}.$$

Since the dealer's bargaining power, θ, lies in $(0,1)$, the objective function of this program is concave in B on the feasible interval $[R(\delta),P]$. Therefore, the first order condition

$$0=\frac{d}{dB}\left((P-B)^{\theta}\,(B-R(\delta))^{1-\theta}\right)$$

is necessary and sufficient for an interior maximum, and solving this equation shows that the bid price is given by a convex combination of the interdealer price and the investor's reservation value:

$$B(\delta)=\theta R(\delta)+(1-\theta)\,P. \tag{4.5}$$

Intuitively, if $\theta\approx 1$ then the dealer has most of the bargaining power and the bid price is close to the reservation value of the investor, which is the price that would be set by a fully discriminating monopsonist. Alternatively, if $\theta\approx 0$ then the dealer has almost no bargaining power and the bid price approaches the competitive interdealer price.

Consider next a meeting between a dealer and a generic non owner with utility flow $\delta\in\mathcal{D}$. In such a meeting, the only possible trade is that the investor buys from the dealer. As in the analysis above, this trade occurs if and only if the surplus $R(\delta)-P$ is nonnegative, in which case the ask price is

$$A(\delta)\equiv\underset{A\in[P,R(\delta)]}{\operatorname{argmax}}\,(A-P)^{\theta}\,(R(\delta)-A)^{1-\theta}=\theta R(\delta)+(1-\theta)\,P. \tag{4.6}$$

Note that the bid and the ask are given by the same formula so that the price at which a given investor can trade with dealers depends on his utility flow but not on the direction of the trade. As we will see below, this feature leads to a significant simplification of the HJB equation, which is ultimately responsible for the tractability of the semi-centralized market model.

We close this section by observing that the results of the Nash bargaining protocol are consistent with the setting of Chapter 3, where we assumed that an investor with a higher utility flow tends to trade at higher prices. Indeed, we will show below that the reservation value function is increasing, which implies that the cumulative price distributions faced by the investor in the semi-centralized market,

$$G_q(x|\delta) = \mathbf{1}_{\{x \geq A(\delta)\}} = \mathbf{1}_{\{x \geq B(\delta)\}} = \mathbf{1}_{\{x \geq \theta R(\delta)+(1-\theta)P\}}, \tag{4.7}$$

are increasing with respect to δ in the first order stochastic dominance sense. Note, however, that here the price distributions are partially endogenous, as they depend on the (yet undetermined) reservation value function.

4.3.2 Reservation values

Substituting $\lambda_q = \lambda$ and (4.7) into (3.17) shows that the HJB equation for the investor's value function is given by

$$\begin{aligned} rV(q,\delta) &= q\delta + \gamma \int_{\underline{\delta}}^{\overline{\delta}} (V(q,x) - V(q,\delta))\, dF(x) \\ &\quad + \lambda q \max\{B(\delta) - R(\delta), 0\} + \lambda(1-q)\max\{R(\delta) - A(\delta, 0)\}. \end{aligned}$$

Subtracting this equation with $q = 0$ from its counterpart with $q = 1$, and substituting the bid and ask prices given in (4.5) and (4.6), then shows that the reservation value function satisfies

$$\begin{aligned} rR(\delta) &= \delta + \gamma \int_{\underline{\delta}}^{\overline{\delta}} (R(x) - R(\delta))\, dF(x) \\ &\quad + \lambda(1-\theta)\max\{P - R(\delta), 0\} - \lambda(1-\theta)\max\{R(\delta) - P, 0\}. \end{aligned} \tag{4.8}$$

Since this functional equation is a special case of (3.10), we can easily derive the existence, uniqueness, and basic features of its solution from the results of Proposition 3.6 and Corollary 3.8.

Proposition 4.3. *The reservation value equation* (4.8) *admits a unique solution in* $\mathcal{B}$ *that is both continuous and strictly increasing in utility flows.*

Comparing equation (4.8) to it sequential search counterpart (3.18) reveals that trading through dealers with bargaining power θ is payoff-equivalent to trading in the interdealer market, but with a reduced intensity

$$\lambda_\theta \equiv \lambda\,(1-\theta)\,.$$

To understand why, recall that Nash bargaining is payoff-equivalent to a static game in which the dealer makes a take-it-or-leave it offer with probability θ and the investor makes a take-it-or-leave-it offer with probability $1-\theta$. For the investor, receiving a take-it-or-leave-it offer is payoff-equivalent to not trading, while making the offer is payoff-equivalent to trading in the interdealer market. As a result, what ultimately matters for reservation values is not the physical contact rate, λ, but the adjusted contact rate, λ_θ, that reflects the combined likelihood of meeting a dealer and being able to make an offer. Hence, even if investors are able to contact dealers quickly, search frictions can have a large impact on valuations if dealers have sufficient bargaining power.

Before moving on to the construction of an equilibrium, we perform some manipulations of equation (4.8) that deliver several alternative representations, as well as an explicit solution for the reservation value function. While these derivations may appear a bit tedious, they turn out to be quite useful down the road, as they shed light on the trading incentives of investors and simplify the characterization of the equilibrium. As a first step, since

$$\max\{P-R(\delta),0\}-\max\{R(\delta)-P,0\}=P-R(\delta),$$

note that we can rewrite (4.8) as

$$rR(\delta)=\delta+\lambda_\theta P+\gamma\int_{\underline{\delta}}^{\overline{\delta}}(R(x)-R(\delta))\,dF(x)-\lambda_\theta R(\delta). \tag{4.9}$$

Combining Proposition A.37 with the basic properties of exponential random variables reported in Lemma A.17 reveals that the unique bounded solution to this functional equation can be represented as

$$\begin{aligned}R(\delta)&=\mathbf{E}_\delta\left[\int_0^{\tau_\theta}e^{-rt}\delta_t dt+e^{-r\tau_\theta}P\right]\\&=\mathbf{E}_\delta\left[\int_0^{\infty}e^{-(r+\lambda_\theta)t}\,(\delta_t+\lambda_\theta P)\,dt\right]\end{aligned} \tag{4.10}$$

where the expectations are conditional on the initial utility flow $\delta_0 = \delta$ and τ_θ is an exponentially distributed variable with intensity equal to the bargaining adjusted contact rate λ_θ.

Equation (4.10) shows that the reservation value can be computed as the present value of the utility flows derived by holding the asset until making contact with the interdealer market and then selling it at the interdealer price. A key insight of this representation is again that the relevant contact rate for the calculation of investors' reservation values reflects *both* the search and the bargaining frictions in the market.

To derive a second useful representation of the reservation value function, recall the elementary identity

$$e^{-r\tau_\theta}P = P + \left(e^{-r\tau_\theta} - 1\right)P = P - \int_0^{\tau_\theta} e^{-rt}\, rP\, dt.$$

Substituting this identity into (4.10) reveals that

$$R(\delta) = P + \mathbf{E}_\delta\left[\int_0^{\tau_\theta} e^{-rt}\left(\delta_t - rP\right)dt\right]. \tag{4.11}$$

This representation highlights that the only reason reservation values differ from the interdealer price is the presence of search and bargaining frictions: the second term is non-zero precisely because the adjusted contact time is strictly positive, so that investors cannot trade at the interdealer price P instantaneously. In particular, as the meeting rate $\lambda \to \infty$, the contact time $\tau_\theta \to 0$ and hence the reservation value $R(\delta) \to P$. Intuitively, if the investor is constantly in contact with dealers, then her reservation value is the interdealer price at all times regardless of her utility flow. Alternatively, as $\lambda \to 0$ or $\theta \to 1$, the adjusted contact time $\tau_\theta \to \infty$ and so

$$\lim_{\lambda_\theta \to 0} R(\delta) = \mathbf{E}_\delta\left[\int_0^{\infty} e^{-rt}\delta_t dt\right].$$

At this limit, investors are effectively in autarky: Either they never meet a dealer or, when they do, the dealer extracts all the surplus. Therefore, they are indifferent between trading and not and their reservation values are simply equal to the autarky values.

Another insight from equation (4.11) that we will use later to solve for the equilibrium price in the interdealer market is that

$$R(\delta) = P \Leftrightarrow P = D(\delta), \tag{4.12}$$

where

$$rD(\delta) = \frac{\mathbf{E}_\delta\left[\int_0^{\tau_\theta} e^{-rt}\delta_t dt\right]}{\mathbf{E}_\delta\left[\int_0^{\tau_\theta} e^{-rt} dt\right]}$$

is the unique solution to

$$\mathbf{E}_\delta \int_0^{\tau_\theta} e^{-rt}\left(\delta_t - x\right) dt = 0.$$

In words, the quantity $rD(\delta)$ represents the *flow* certainty equivalent of the utility that an investor with utility type δ derives by holding the asset until the next bargaining-adjusted contact time. Since the latter is independent from the investor's utility type, we have that

$$r\mathbf{E}_\delta\left[\int_0^{\tau_\theta} e^{-rt} dt\right] = 1 - \mathbf{E}\left[e^{-r\tau_\theta}\right] = \frac{r}{r+\lambda_\theta}$$

where the second equality follows from Lemma A.17. Substituting back into equation (4.10) reveals that the reservation value function can be expressed as a convex combination of the interdealer price and the present value $D(\delta)$ of receiving the certainty equivalent flow in perpetuity, i.e.,

$$R(\delta) = \frac{\lambda_\theta}{r+\lambda_\theta}P + \frac{r}{r+\lambda_\theta}D(\delta). \tag{4.13}$$

To derive an explicit solution for $D(\delta)$ we now proceed as follows (see Proposition A.37 for a similar argument in a more general context). Integrating against the reset distribution $dF(\delta)$ on both sides of (4.9) shows that

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

where

$$\mathbf{E}^F\left[g(\delta)\right] \equiv \int_{\underline{\delta}}^{\overline{\delta}} g(x)dF(x) \tag{4.14}$$

denotes the mean of $g(\delta)$ under the assumption that δ is distributed according to $F(x)$. This identity implies that the average reservation value across the population of investors is explicitly given by

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

regardless of the arrival rate γ of preference shocks. Substituting back into (4.9) and comparing with (4.13), we finally obtain that

$$R(\delta) = \frac{\lambda_\theta}{r + \lambda_\theta}P + \frac{1}{r + \lambda_\theta}\left(\frac{r + \lambda_\theta}{r + \gamma + \lambda_\theta}\delta + \frac{\gamma}{r + \gamma + \lambda_\theta}\mathbf{E}^F\left[x\right]\right). \tag{4.15}$$

This implies in particular that:

$$rD(\delta) \equiv \frac{r + \lambda_\theta}{r + \gamma + \lambda_\theta}\delta + \frac{\gamma}{r + \gamma + \lambda_\theta}\mathbf{E}^F\left[x\right].$$

Together with (4.13), this explicit solution confirms the continuity and strict monotonicity predicted by Proposition 4.3 but also permits a more detailed analysis. In particular, it shows that reservation values are linear in utility flow with a slope

$$R'(\delta) = \frac{1}{r + \gamma + \lambda_\theta}. \tag{4.16}$$

Note that $R'(\delta)$ is independent from the reset distribution $F(\delta)$ and decreasing monotonically in r, γ, and λ_θ, so that an investor's utility flow matters less for her reservation value as she becomes more impatient, as she changes types more often, and as she contact dealers more frequently. The linearity of the reservation value function and its independence from the reset distribution reflect the fact that all investors in the market are equally likely to contact a dealer. As such, these properties are specific to the semi-centralized market setting. In particular, we will see in Chapters 5 and 8 that, when investors can choose their contact rate with dealers or meet each other directly in a frictional market, the marginal reservation value $R'(\delta)$ varies to reflect the different trading opportunities of each investor in the market.

4.4 EQUILIBRIUM

In this section we use the optimal trading behavior of investors to construct the equilibrium allocation of assets across the population of investors and the equilibrium price in the interdealer market. As a first step, we start by deriving a convenient formulation of the market clearing condition.

4.4.1 Market clearing

The optimal trading behavior of investors follows directly from our analysis of the reservation value function in the previous section: Conditional on meeting a dealer, an asset owner with utility flow δ weakly prefers to sell if and only if his reservation value $R(\delta) \leq P$, while a non owner with utility flow δ weakly prefers to buy if his reservation value $R(\delta) \geq P$.

An immediate consequence of this trading strategy at the aggregate level is that *in equilibrium* it must hold that $R(\delta') \leq P \leq R(\delta'')$ for some δ', δ'' in the support of $F(\delta)$, as otherwise all investors would be on the same side of the market at all times. Since the reservation value function is continuous and strictly increasing, by Proposition 4.3, this observation implies that in any equilibrium there exists a unique utility flow $\delta^\star \in \mathcal{D}$ such that $R(\delta^\star) = P$ and $R(\delta) \leq P$ if and only if $\delta \leq \delta^\star$. The traditional approach to pin down this *marginal* type would be to derive the allocation induced by a given $\delta^\star$ and then to vary this threshold until the allocation is stationary and consistent with market clearing. However, and as noted by Lagos and Rocheteau [2009], the specific features of the semi-centralized market setting allow us to circumvent this approach by directly computing the gross demand and supply induced by a given marginal type, as follows.

Since all investors meet dealers at the same rate, it follows from the law of large numbers that the joint distribution of utility flows and asset holdings among the group of investors who meet dealers at a given instant is the same as in the whole population. In particular, the marginal distribution of utility flows in that group is $F(\delta)$, with a proportion s of owners. Hence, over a small time interval of length h, the gross asset supply in the interdealer market is λsh. On the other hand, given a threshold utility flow $\delta^\star$ we have that the post-contact asset holdings of an investor with utility type $\delta \in \mathcal{D}$ must satisfy

$$q \begin{cases} = 0, & \text{if } \delta < \delta^\star \\ \in \{0, 1\} & \text{if } \delta = \delta^\star \\ = 1, & \text{if } \delta > \delta^\star. \end{cases}$$

This implies that, over a small time interval of length h, the gross demand for assets in the interdealer market lies in

$$\lambda h \left[1 - F(\delta^\star), 1 - F(\delta^\star -)\right].$$

Equating gross supply and demand, we conclude that the marginal type is a solution to the inclusion

$$\delta^\star \in \mathcal{S} \text{ such that } s \in \left[1 - F(\delta^\star), 1 - F(\delta^\star -)\right], \tag{4.17}$$

where $\mathcal{S}$ denotes the convex hull of the support of $F(\delta)$. This simple expression of the market clearing condition pins down the marginal type as a quantile of the reset distribution $F(\delta)$ at level $1 - s$ and, therefore, establishes the *existence* and *generic uniqueness* of the equilibrium.

This condition also shows that the marginal type does not depend on the contact rate with dealers λ in our semi-centralized market setting. In fact, this requirement is exactly the same as in the frictionless benchmark of Section 4.2 because, at each instant, the group of investors who are matched with dealers participate in a frictionless market. This, however, does *not* imply that search frictions do not matter for prices and allocations in a semi-centralized market. It only means that, over a short time interval, search frictions do not affect the post-contact asset holdings of those investors who contact a dealer. But search frictions do affect both the allocation and the valuation of those investors who do not get matched with a dealer. In particular, search frictions imply that there are *misallocated* investors with utility flow $\delta < \delta^\star$ ($\delta > \delta^\star$) who hold (don't hold) the asset because they have switched utility flow but have not yet been able to contact a dealer with whom to trade.

4.4.2 Asset allocation

To precisely understand the impact of search frictions on the asset allocation, consider the cumulative measure $\Phi_q(\delta)$ of investors who hold q units of the asset and have utility flow in the interval $(\underline{\delta}, \delta]$. In a steady-state equilibrium these measures must satisfy three conditions. First, they must add up to the distribution of utility flows across the whole populations:

$$\Phi_0(\delta) + \Phi_1(\delta) = F(\delta), \qquad \delta \in \mathcal{D}. \tag{4.18}$$

Second, they must enforce market clearing in the sense that the total measure of owners must equal the asset supply:

$$\Phi_1(\bar{\delta}) \equiv \lim_{\delta \uparrow \bar{\delta}} \Phi_1(\delta) = s. \tag{4.19}$$

Third, these measures must be consistent with stationarity, which requires that the inflow to and outflow from any group of investors be equalized at all times. Specifically, let us define $\lambda_q(\delta)$ to be the trading intensity of an investor who holds $q \in \{0, 1\}$ units of the asset and has utility flow δ. With this notation, the stationarity condition can be formally written as

$$\int_B \left(\gamma s dF(\delta) + \lambda_0(\delta)\, d\Phi_0(\delta)\right) = \int_B \left(\gamma d\Phi_1(\delta) + \lambda_1(\delta)\, d\Phi_1(\delta)\right) \tag{4.20}$$

for all Borel sets $B \subseteq \mathcal{D}$. This stationary condition holds in all the models we consider in this book, albeit with different market structures and, thus, different trading intensity functions. The left-hand side is the inflow into the group of owners with utility flow in B, whereas the right-hand side is the outflow from that group. The inflow is itself the sum of two terms. The first term, $\gamma s dF(\delta)$ is the inflow induced by preference shocks. Indeed, at any instant a measure $\gamma s = \gamma \Phi_1(\bar{\delta})$ of owners receive a preference shock that resets their utility flow to some $\delta \in B$ with probability $dF(\delta)$. The second term, $\lambda_0(\delta) d\Phi_0(\delta)$ is the inflow due to asset purchase. The two terms that compose the outflow on the right-hand side have a similar interpretation.

In the semi-centralized market setting of this chapter, the trading intensity function $\lambda_q(\delta)$ takes a very simple form:

$$\lambda_q(\delta) \begin{cases} = \lambda q & \text{if } \delta < \delta^\star \\ \in [0, \lambda] & \text{if } \delta = \delta^\star \\ = \lambda(1-q) & \text{if } \delta > \delta^\star. \end{cases} \tag{4.21}$$

This expression allows us to pin down the equilibrium allocation by stating (4.20) for two well chosen groups of investors. First, consider investors with utility flow in $B = (\underline{\delta}, \delta]$ for some $\delta < \delta^\star$. Equation (4.21) shows that the trading intensities of investors with utility flow in that set are given by $\lambda_1(x) = 0$ and $\lambda_0(x) = \lambda$. Therefore, it follows from (4.20) that

$$\Phi_1(\delta) = \frac{\gamma}{\gamma + \lambda} sF(\delta), \qquad \delta < \delta^\star. \tag{4.22}$$

Next, consider the group of investors with utility flow in $B = (\delta, \overline{\delta})$ for some fixed $\delta \geq \delta^\star$. Equation (4.21) shows that the trading intensities of such investors are given by $\lambda_1(x) = 0$ and $\lambda_0(x) = \lambda$. Therefore, (4.20) implies that

$$\Phi_1(\overline{\delta}) - \Phi_1(\delta) = \frac{\gamma s + \lambda}{\gamma + \lambda} \left(1 - F(\delta)\right), \qquad \delta \geq \delta^\star,$$

and substituting (4.19) we obtain that

$$\Phi_1(\delta) = \frac{\gamma}{\gamma + \lambda} sF(\delta) + \frac{\lambda}{\gamma + \lambda} \left(F(\delta) - (1 - s)\right), \qquad \delta \geq \delta^\star. \tag{4.23}$$

Finally, combining (4.22) and (4.23) while keeping in mind that $\delta^\star$ is a quantile of the distribution of utility flows at the level $1 - s$, we obtain an expression that is valid across all utility flows:

$$\Phi_1(\delta) = \frac{\gamma}{\gamma + \lambda} sF(\delta) + \frac{\lambda}{\gamma + \lambda} \Psi_1(\delta) \tag{4.24}$$

where $\Psi_1(\delta)$ is the frictionless allocation in (4.3). Importantly, this expression depends neither on the choice of a marginal type $\delta^\star$ satisfying (4.17) nor on the specification of the trading intensities of marginal investors (see Exercises 4.8

and 4.9). This shows that the allocation is always unique—even in the non-generic cases in which the marginal type is not.

Equation (4.24) shows that the equilibrium allocation arises from the convex combination of two distributions: the distribution that obtains when randomly allocating assets across the population, $sF(\delta)$, and the efficient distribution that results from allocating assets to those investors who value them most, $\Psi_1(\delta)$. The weights in this combination depend on λ and γ only through their ratio $\phi \equiv \gamma/\lambda$. This constant measures the arrival rate of trading opportunities relative to that of preference shocks and controls the efficiency of the allocation. Indeed, it follows directly from (4.3) and (4.24) that

$$\frac{\partial \Phi_1(\delta)}{\partial \phi} = \frac{sF(\delta) - \Psi_1(\delta)}{(1+\phi)^2} \geq 0$$

and

$$\lim_{\phi \to 0} \Phi_1(\delta) = \Psi_1(\delta) \leq sF(\delta) = \lim_{\phi \to \infty} \Phi_1(\delta).$$

Therefore, the distribution of utility flows among asset owners starts from the efficient allocation of the frictionless benchmark at $\phi = 0$ and then decreases monotonically in the sense of first order stochastic dominance to reach the purely random allocation when $\phi \to \infty$. Intuitively, as ϕ decreases, investors either have more opportunities to trade (if λ increases) or find themselves needing to trade less often (if instead γ decreases). In both cases, the market is better able to channel assets toward investors with high utility flows, which reduces the impact of search frictions.

Note that the efficient allocation can be obtained even in the presence of search frictions ($\lambda < \infty$) provided that the arrival rate of preference shocks γ approaches zero. This is because, with permanent utility flows, non owners with $\delta \geq \delta^\star$ (owners with $\delta < \delta^\star$) will acquire (sell) the asset upon their first meeting with a dealer and then hold on to it (remain asset-free), so that the trading volume is zero in the steady state.

Example 4.4. To illustrate the results above, Figure 4.4 plots the cumulative measures $\Phi_q(\delta)$ in a semi-centralized market where the reset distribution is uniform on the unit interval, the supply is $s = 0.4$, and $\phi = 1$ so that preference shocks and trading opportunities arrive at the same speed.

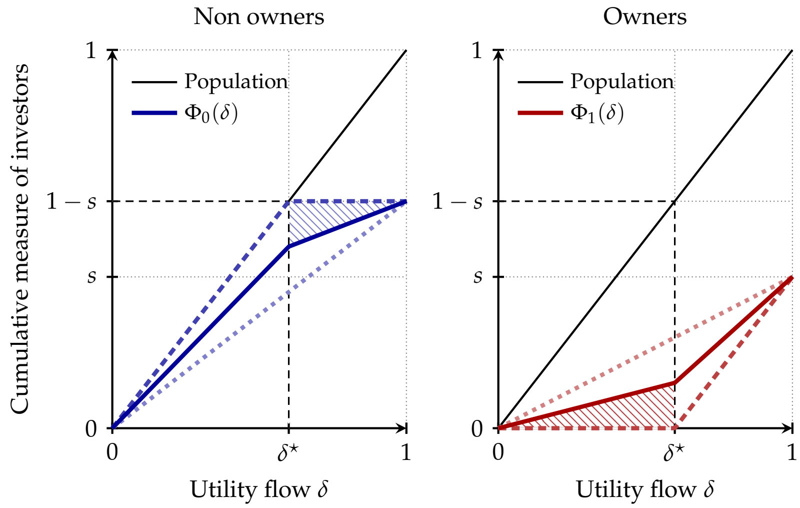

FIGURE 4.4: Equilibrium allocation in the semi-centralized market

This figure illustrates the equilibrium distribution of utility flows among owners (right panel) and non owners (left panel) in a model with $\lambda/\gamma = 1$, asset supply $s = 0.4$, and uniform distribution $F(\delta) = \delta$ on $\mathcal{D} = (0,1)$. In each panel, the solid line represents the population distribution, the thick dashed line represents the frictionless benchmark, the dash-dotted line represents the purely random case where $\phi \to 0$, and the hatched region represents the misallocated investors.

Comparing the distributions to their frictionless counterparts shows that the impossibility to trade immediately after a preference shocks makes the equilibrium less efficient by pushing the distribution among asset owners up and to the left, so that there are investors with $\delta < \delta^\star$ who would not hold the asset in the frictionless benchmark but do so in the semi-centralized market. In this numerical example, the steady state measure of these *misallocated* investors is $\Phi_1(\delta^\star) = 0.12$ which implies that $0.12/s = 30\%$ of existing assets are in the wrong hands at any point in time .

Symmetrically, the left panel shows that search frictions push the distribution among non-owners down and to the right, so that non-owners tend to have higher utility flows in the environment with search frictions. In particular, at any point in time, there is a mass $\Phi_0(\overline{\delta}) - \Phi_0(\delta^\star) = 0.12$ of investors with utility flow $\delta > \delta^\star$ who would be owners in the frictionless market but are not in the semi-centralized market because they have not been able to locate a dealer since the occurence of their last preference shock.

4.4.3 The interdealer price

To complete the construction of the equilibrium, we now determine the price of the asset in the interdealer market. From the characterization of the marginal type, we know that $P \equiv R(\delta^\star)$, i.e., that the price is given by the reservation value of the asset to an investor with current utility flow equal to the marginal type. Therefore, it immediately follows from (4.12) and (4.15) that

$$P = R(\delta^\star) = D(\delta^\star) = \frac{\delta^\star}{r} - \frac{\gamma}{r\,(r+\gamma+\lambda_\theta)}\left(\delta^\star - \mathbf{E}^F\left[\delta\right]\right), \tag{4.25}$$

where $\mathbf{E}^F\left[\cdot\right]$ is defined in (4.14). This expression shows that the equilibrium interdealer price is the sum of the price $\delta^\star/r$ in the frictionless benchmark and a correction term that reflects the frictions at play in the semi-centralized market environment. Specifically, a direct calculation shows that

$$\lim_{\lambda_\theta\to\infty} P = \frac{\delta^\star}{r}$$

and

$$\lim_{\lambda_\theta\to 0} P = \frac{\delta^\star}{r} - \frac{\gamma\left(\delta^\star - \mathbf{E}^F\left[\delta\right]\right)}{r\,(r+\gamma)} = \mathbf{E}_{\delta^\star}\left[\int_0^\infty e^{-rt}\delta_t dt\right].$$

Moreover, one can easily show that

$$\frac{\partial P}{\partial \lambda_\theta} \propto -\frac{\partial P}{\partial \gamma} \propto \left(\delta^\star - \mathbf{E}^F\left[\delta\right]\right),$$

so that the direction of the deviation and its comparative statistics are entirely determined by the difference between the marginal type and the average utility flow in the population of investors, $\delta^\star - \mathbf{E}^F\left[\delta\right]$. Figure 4.5 illustrates for an example with a uniform distribution of utility flows.

To understand these results, recall from Chapter 3 that the reservation value function contains two components related to future trading opportunities. First, part of the value of acquiring an asset derives from the option to sell it at a later date. Frictions, however, reduce this value and make investors less willing

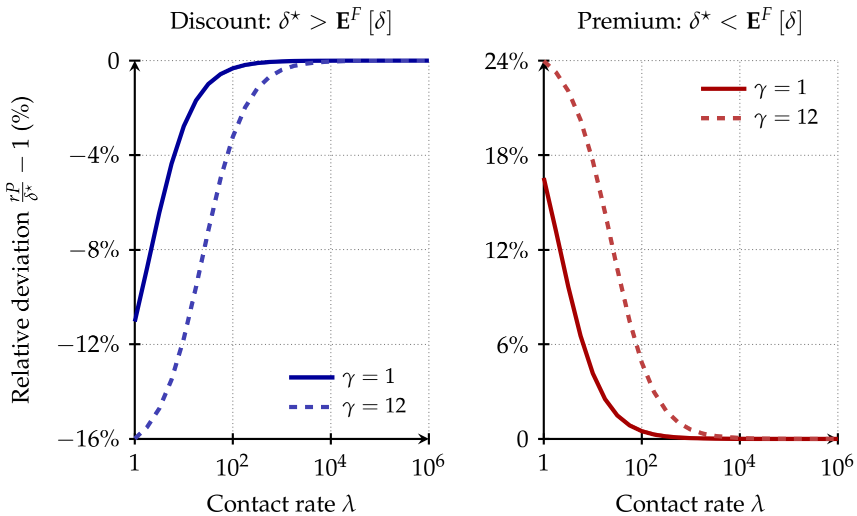

FIGURE 4.5: The interdealer price in the semi-centralized market

This figure illustrates the behavior of the interdealer price in a model with a uniform distribution $F(\delta) = \delta$ on $\mathcal{D} = (0,1)$. The left panel shows the relative deviation from the frictionless price as a function of λ in a market where $\delta^\star = 1 - s = 0.6 > \mathbf{E}^F[\delta]$ and either $\gamma = 1$ (solid) or $\gamma = 12$ (dashed). The right panel plots the same quantities but in a market where the marginal type $\delta^\star = 0.4$ is lower than the average.

to pay for the asset—that is, frictions introduce an *illiquidity discount*. When frictions ease, the investor is able to exercise the option to sell more quickly, the illiquidity discount declines, and his reservation value rises. Second, by acquiring an asset now, the investor derives value from the fact that he *doesn't* have to search for an asset later if he receives a positive shock to his utility flow. Unlike the option value of reselling, the presence of frictions actually increases this second component of the investor's reservation value—that is, frictions introduce what we call a *scarcity premium*. As frictions vanish, this scarcity premium declines, since it becomes less valuable to not have have to search for the asset in the future.

Which of the illiquidity discount or scarcity premium dominates depends on the asset supply. When the average utility flow $\mathbf{E}^F[\delta]$ is smaller than the marginal utility flow $\delta^\star$, which occurs if and only if the supply is large, then the illiquidity discount effect described above dominates, since the investor is more likely to receive a preference shock that makes him want to exercise the option to sell the asset. Hence, in equilibrium, the asset price is below the price in the frictionless benchmark and increases as frictions vanish—either because the meeting rate increases or because the frequency of preference shocks declines. Alternatively, when $\mathbf{E}^F[\delta] > \delta^\star$ then the scarcity premium effect dominates since the investor is more likely to receive a preference shock that makes him want to own the asset. In this case, the presence of frictions make the investor more willing to pay for the asset now, raising his reservation value and hence pushing the equilibrium price above $P = \delta^\star/r$. As frictions vanish, however, the reduction in the scarcity premium outweighs the reduction in the illiquidity discount and the price falls.

In either case, the price approaches the Walrasian price as the bargaining adjusted meeting rate $\lambda_\theta \to \infty$. An intriguing implication of (4.25), however, is that the price can exceed or fall short of the buy and hold price

$$\lim_{\lambda_\theta \to 0} D(\delta) = \mathbf{E}_\delta \left[\int_0^\infty e^{-rt} \delta_t dt \right] = \frac{\delta}{r} + \frac{\gamma}{r(r+\gamma)} \left(\mathbf{E}^F[\delta] - \delta \right) \tag{4.26}$$

for *any* investor in the market. For example, as $s \to 0$ we have that the marginal type converges to the upper bound of the support of the distribution $F(\delta)$, and it thus follows from (4.25) and (4.26) that

$$\min_{\delta \in \mathcal{S}} \left(P - \lim_{\lambda_\theta \to 0} D(\delta) \right) \to \max_{\delta \in \mathcal{S}} \frac{\gamma \lambda_\theta (\delta - \mathbf{E}^F[\delta])}{r(r+\gamma)(r+\gamma+\lambda_\theta)} \geq 0.$$

The opposite conclusion obtains when $s \to 1$, in which case the interdealer price lies below the buy-and-hold value of any investor. To understand these results, recall from Chapter 3 that the reservation value is not determined by the value of buying and holding forever. Instead, it is determined by the value of buying and then *following an optimal strategy* that requires to sell whenever possible if $\delta < \delta^\star$ and to buy whenever possible if $\delta > \delta^\star$.

4.4.4 Sequential representation of reservation values

To close this section, we discuss an insightful representation of equilibrium reservation values. In equilibrium, the interdealer price P must coincide with the reservation value $R(\delta^\star)$ of an investor with current utility flow equal to the marginal type. Substituting this condition into (4.9) shows that, *in equilibrium*, the reservation value function solves

$$\begin{aligned} rR(\delta) &= \delta + \lambda_\theta \left(R(\delta^\star) - R(\delta) \right) + \gamma \int_{\underline{\delta}}^{\overline{\delta}} \left(R(x) - R(\delta) \right) dF(x) \\ &= \delta + (\gamma + \lambda_\theta) \int_{\underline{\delta}}^{\overline{\delta}} \left(R(x) - R(\delta) \right) d\hat{F}(x) \end{aligned}$$

with the auxiliary distribution defined by

$$\hat{F}(\delta) = \frac{\gamma}{\gamma + \lambda_\theta} F(\delta) + \frac{\lambda_\theta}{\gamma + \lambda_\theta} \mathbf{1}_{\{\delta \geq \delta^\star\}}.$$

Combining this equation with Proposition A.37 of Appendix A.6 then delivers the following explicit representation.

Proposition 4.5. *In equilibrium,*

$$R(\delta) = \mathbf{E}_\delta \left[\int_0^\infty e^{-rt} \hat{\delta}_t dt \right] \tag{4.27}$$

where the market valuation process $(\hat{\delta}_t)_{t \geq 0}$ is a Markov jump process with constant jump intensity $\gamma + \lambda_\theta$ and reset distribution $\hat{F}(\delta)$.

The proposition refers to $\hat{\delta}_t$ as the *market valuation* process because it takes into account not only the physical changes of utility flows of an investor, but also his trading opportunities. To describe this process note that, during a small interval of length h, the market valuation can change for two reasons. First, a preference shock arrives with probability γh, in which case a new utility flow is drawn from $F(x)$. Second, a trading opportunity arrives with probability $\lambda_\theta h$, in which case the market valuation jumps to the marginal type so that the reservation value jumps to the interdealer price.

Representations such as (4.27) are standard in classical asset pricing, where private asset values are obtained as the present value of cash flows under an auxiliary probability constructed from marginal rates of substitution. The emergence of such a representation in a decentralized market can be viewed as a generalization of the concept of the marginal investor that accounts for search and bargaining frictions. Indeed, in the frictionless benchmark, we have $\lambda_\theta / \gamma \to \infty$ and it follows that the market valuation is constant and equal to the marginal type. By contrast, in the semi-centralized trading environment, the market valuation process is stochastic and differs from the marginal type because, following a negative preference shock, every asset owner is forced to enjoy his depressed utility flow until a dealer can be contacted.

4.4.5 Summary

The above derivations deliver a complete description of the equilibrium in the semi-centralized market setting. Specifically, the equilibrium is generically unique and comprised of

- The interdealer price P in (4.25)
- The reservation value function $R(\delta)$ in (4.15)
- The asset allocation captured by the measures $\Phi_q(\delta)$ in (4.24)

where the marginal type $\delta^\star$ is the generically unique solution to the market clearing condition (4.17).

4.5 EQUILIBRIUM IMPLICATIONS

We now use the explicit characterization of the equilibrium to study the effects of search and bargaining frictions on several common measures of liquidity: trading volume, the bid-ask spread, and the the yield spread.

To simplify the exposition, we *assume in this section that the distribution of utility flows is continuous*—so that the measure of investors who are indifferent between buying or selling is zero at all times—and guide the reader through the general case in the end-of-chapter exercises. This assumption has two important consequences that we use repeatedly throughout this section: it implies that the cumulative measures $\Phi_q(\delta)$ are continuous and it guarantees that $F(\delta^\star) = 1 - s$ at any marginal type.

4.5.1 Trading volume

In equilibrium, the instantaneous order flows originating from investors who meet a dealer and either sell or buy are equal and given by

$$\lambda \Phi_1(\delta^\star) = \lambda \left(1 - s - \Phi_0(\delta^\star)\right) = \frac{\lambda}{\gamma + \lambda} \gamma s(1 - s), \tag{4.28}$$

where the equalities follows from (4.18), (4.24), and the assumed continuity of $F(\delta)$. In contrast, when the distribution of utility flows is continuous, Example 4.2 shows that the flow of trades required to clear markets in the frictionless benchmark is $\gamma s(1-s)$. Therefore, (4.28) reveals that search frictions reduce trading volume by a factor $\lambda/(\gamma+\lambda)$. Note that this factor approaches 1 as $\phi = \gamma/\lambda \to 0$: search frictions create a bottleneck that slows down asset reallocation relative to the frictionless benchmark unless utility flows are permanent (in which case there is no trade in either models).

4.5.2 Trading costs

One of the most common measures of market liquidity is the bid-ask spread, which is commonly defined as the cost of a *round-trip* transaction, i.e., an asset purchase followed by an asset sale. The semi-centralized framework offers a coherent theory of the bid-ask spread based on search and bargaining frictions. Indeed, with search frictions, dealers generate a surplus by offering investors the opportunity to immediately rebalance their portfolio; and with bargaining power, they extract part of this surplus in the form of the bid-ask spread. In this section, we derive the equilibrium bid-ask spread and study how it depends on the severity of the frictions and the preferences of the investors.

Recall that, upon meeting a dealer, owners with utility flow δ prefer to sell if and only if $\delta \le \delta^\star$, in which case the price they receive is

$$B(\delta) = \theta R(\delta) + (1-\theta)P = \theta R(\delta) + (1-\theta)R(\delta^\star),$$

where we have used (4.25). Moreover, since $F(\delta)$ is continuous, we have that the distribution of utility flows among owners who prefer to sell is

$$\frac{\Phi_1(\min\{\delta,\delta^\star\})}{\Phi_1(\delta^\star)} = \frac{F(\min\{\delta,\delta^\star\})}{F(\delta^\star)} \equiv F\left(\delta|\delta \le \delta^\star\right). \tag{4.29}$$

Hence, the average transaction cost incurred by an investor who sells—or the average *bid-to-interdealer spread*—can be written as

$$\begin{aligned} P - \overline{B} &= \int_{\underline{\delta}}^{\delta^\star} (P - B(\delta)) \frac{d\Phi_1(\delta)}{\Phi_1(\delta^\star)} = \mathbf{E}^F \left[P - B(\delta) \mid \delta \leq \delta^\star \right] \\ &= \theta \mathbf{E}^F \left[R(\delta^\star) - R(\delta) | \delta \leq \delta^\star \right] = \frac{\theta \mathbf{E}^F \left[\delta^\star - \delta | \delta \leq \delta^\star \right]}{r + \gamma + \lambda_\theta}, \end{aligned} \tag{4.30}$$

where $\mathbf{E}^F \left[\cdot \mid \delta \leq \delta^\star \right]$ denotes an integral against the distribution in (4.29) and the last equality follows from (4.16), which shows that the reservation value function is linear in utility flow with slope $1/\left(r + \gamma + \lambda_\theta\right)$.

Notice that the average bid-to-interdealer spread is zero when dealers have no bargaining power, since in that case investors can sell at the interdealer price. When $\theta > 0$, the average bid-to-interdealer spread is strictly positive and decreasing in the adjusted contact rate, λ_θ, so that the cost of selling the asset increases as trading frictions becomes more severe. Also note that the spread gets larger when asset owners experience larger declines in utility flow upon becoming sellers—i.e., more extreme distress costs—as captured by the conditional expectation $\mathbf{E}^F \left[\delta^\star - \delta | \delta \leq \delta^\star \right]$. Finally, note that the average selling cost is decreasing in the arrival rate of preference shocks. Intuitively, as γ increases, utility flows become less persistent and differences in reservation values across investors shrink. As a result, the gains from trade created by reallocating the asset is diminished and spreads decline.

Symmetrically, using (4.6) and (4.24) together with the assumed continuity of $F(\delta)$ shows that the average *ask-to-interdealer spread* is

$$\begin{aligned} \overline{A} - P &= \mathbf{E}^F \left[A(\delta) - P | \delta > \delta^\star \right] \\ &= \theta \mathbf{E}^F \left[R(\delta) - R(\delta^\star) | \delta > \delta^\star \right] = \frac{\theta \mathbf{E}^F \left[\delta - \delta^\star | \delta > \delta^\star \right]}{r + \gamma + \lambda_\theta}. \end{aligned} \tag{4.31}$$

Comparing (4.30) and (4.31) reveals an important empirical implication of this class of model: the bid-to-interdealer spread and the ask-to-interdealer spread are not necessarily symmetric. The reason why the spreads charged to buyers and sellers may be different is that these costs depend on the average gains from trade above and below the marginal type $\delta^\star$, which in turn reflect the (possibly asymmetric) shape of the distribution of utility flows.

The asymmetry of transaction costs on the bid and the ask side of the market is empirically relevant. Indeed, we saw in Chapter 2 that customers pay larger transaction costs when they buy than when they sell. Through the lens of the model considered here, this indicates that the difference $|\delta - \delta^\star|$ tends to be larger above $\delta^\star$ than below. See Chapter 2 for a more detailed discussion.

Finally, combining (4.30) and (4.31) shows in a semi-centralized market setting that the average *bid-ask spread* is explicitly given by

$$\overline{A} - \overline{B} = \frac{\theta}{r + \gamma + \lambda_\theta} \left(\mathbf{E}^F \left[\delta | \delta > \delta^\star \right] - \mathbf{E}^F \left[\delta | \delta \leq \delta^\star \right] \right).$$

Intuitively, this expression shows that the average bid-ask spread is equal to the surplus between the average buyer and seller, and thus proportional to the difference between the utility flows of the average buyer and the average seller. Following the same logic as above, the average bid-ask spread is increasing in the bargaining power of dealers, increasing in the severity of the trading frictions, and decreasing in the arrival rate of preference shocks.

4.5.3 Liquidity yield spreads

The yield spread of a fixed-income instrument is comprised of one component related to risk and one component related to liquidity. The model studied here delivers a simple, intuitive formula for the liquidity component. In particular, we define the liquidity yield spread, ℓ, as the solution to

$$\frac{\delta^\star}{r + \ell} = P = R(\delta^\star).$$

In words, the liquidity yield spread is the adjustment to the interest rate that makes the present value of the cash flows to a perpetually marginal investor equal to the interdealer price. Substituting the above definition into (4.25) and simplifying yields

$$\frac{\ell}{r + \ell} = \frac{\gamma}{r + \gamma + \lambda_\theta} \left(1 - \mathbf{E}^F \left[\delta / \delta^\star \right] \right). \tag{4.32}$$

Note that the liquidity yield spread can be either positive or negative, which follows from our finding that trading frictions have ambiguous effects on the interdealer price. In particular, the yield spread is positive when $\delta^\star > \mathbf{E}^F [\delta]$, which corresponds to the case in which the asset trades at a discount relative to the frictionless benchmark.

Assuming for the sake of discussion that the spread is positive, as received wisdom suggests, we see that several factors contribute to a larger spread. The first is larger values of γ, which correspond to more frequent preference shocks. The second is more severe search frictions or greater bargaining power for dealers, as captured by the adjusted contact rate λ_θ. Finally, the theory predicts that larger liquidity yield spreads are associated with more extreme distress costs, as measured by the difference between the marginal type and the average utility flow across the population of investors, $\delta^\star - \mathbf{E}^F[\delta]$.

Interestingly, (4.32) shows that the liquidity yield spread can be positive even if investors have all the bargaining power and, thus, do not appear to incur a transaction cost when they trade with dealers. The reason is that the price still capitalizes the present value of the distress costs that investors will incur while waiting for an opportunity to trade with dealers. This observation stands in contrast with traditional formulae for asset pricing with transaction costs, such as Amihud and Mendelson [1986], in which yield spreads capitalize the present value of transaction costs incurred by successive buyers.

4.6 TRADING RELATIONSHIPS

The random matching mechanism that underlies the semi-centralized market setting assumes that, after executing a trade, an investor and a dealer part ways and never meet again. This simplifies the analysis but ignores the fact that, in practice, investors tend to form relationships with dealers that allow them to bypass search frictions and obtain better terms of trade. In this section, we briefly outline a variant of the semi-centralized model that incorporates the possibility of such long-term relationships.

4.6.1 The model

We consider the same economic environment as in previous sections but with a different modelling of relationships. Specifically, we assume that instead of parting ways after negotiating a trade, a dealer and an investor remain connected for some period of time. As a result, the population of investors is now comprised of two types: *matched* and *unmatched* investors.

An unmatched investor becomes matched at rate $\lambda \geq 0$. Upon matching, an investor and a dealer bargain over a lump-sum fee, which allows the investor to trade instantly at the interdealer price until the end of the relationship, which occurs at rate $\beta \geq 0$. After the relationship ends, the investors cannot trade until it meets a new dealer. In particular, no transaction can be executed at the breakup time. We assume that a dealer can be simultaneously matched with several investors, but an investor can be matched with at most one dealer at a time. As before, we denote by $\theta \in [0,1]$ the bargaining power of dealers relative to investors and, to simplify the presentation, we assume throughout this section that the distribution of utility flow is *continuous*.

The model we present in this section assumes that every meeting between a dealer and an unmatched investor initiates a long-term relationship, which implies that all transactions occur at the interdealer price. This admittedly stark assumption can be relaxed by assuming that only a fraction of such meetings lead to a relationship, while the remaining fraction generate at most one trade, as in the previous sections. In this case, the model would produce two types of trades: relationship trades that occur at the interdealer price and spot trades (with dispersed prices) that result from one-off meetings between investors and dealers. See Exercise 4.11 for the analysis of such a model.

4.6.2 Asset demand

Let $V(q,\delta)$ denote the value function of an *unmatched* investor with current utility flow δ who holds q units of the asset, and $M(q,\delta)$ denote the value function of an otherwise identical *matched* investor. Using integration by parts as in Section 4.2 shows that

$$\begin{aligned} M(q,\delta) &= \sup_{\vartheta\in\mathcal{Q}_q} \mathbf{E}_\delta \left[\int_0^\sigma e^{-rt} \left(\vartheta_t \delta_t dt - P d\vartheta_t \right) + e^{-r\sigma} V(\vartheta_\sigma, \delta_\sigma) \right] \\ &= qP + \sup_{\vartheta\in\mathcal{Q}_0} \mathbf{E}_\delta \left[\int_0^\sigma e^{-rt} \vartheta_t \left(\delta_t - rP \right) dt + e^{-r\sigma} \left(V(\vartheta_\sigma, \delta_\sigma) - \vartheta_\sigma P \right) \right], \end{aligned}$$

where P is the interdealer price, σ is an exponential random time, with rate β, at which the relationshjp ends, and $\mathcal{Q}_q$ denotes the set of $\{0,1\}$ $-$valued processes that start from q. Integrating against the distribution of σ inside the expectation and maximizing the result pointwise over ϑ_t yields

$$\begin{aligned} M(\delta) &\equiv M(0,\delta) = M(q,\delta) - qP \\ &= \mathbf{E}_\delta \left[\int_0^\infty e^{-(r+\beta)t} \left((\delta_t - rP + \beta(R(\delta_t) - P))^+ + \beta V(0,\delta_t) \right) dt \right] \end{aligned}$$

where

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

denotes the reservation value of an unmatched investor. The key implication of this expression is that a matched investor holds the asset if and only if his utility flow is such that

$$\delta \geq rP + \beta(P - R(\delta)). \tag{4.33}$$

To understand this condition, it suffices to compare the instantaneous costs and benefits of holding an asset for a matched investor with utility flow δ. Over a small time interval of length h, the benefit is approximately δh, whereas the cost is approximately

$$\beta h \left(P - e^{-rh} R(\delta) \right) + (1 - \beta h) \left(P - e^{-rh} P \right) \approx (rP + \beta(P - R(\delta)))\, h.$$

This expression shows that the cost depends on whether the relationship ends or continues. If the relationship ends, with probability βh, the investor pays the inter-dealer price P at the beginning of the time interval, and receives the unmatched reservation value at the end. If the relationship continues, with probability $1 - \beta h$, the investor pays and receives the inter-dealer price at the beginning and at the end of the time.

Importantly, (4.33) shows that the asset demand of matched investors is fully determined by the unmatched reservation value $R(\delta)$. To compute this function, note that the surplus generated by a meeting between a dealer and an unmatched investor is

$$M(q,\delta) - V(q,\delta) = M(\delta) + qP - V(q,\delta).$$

Therefore, the Hamilton-Jacobi-Bellman equation for the value function of an unmatched investor is given by

$$rV(q,\delta) = q\delta + \lambda_\theta \left(M(\delta) + qP - V(q,\delta)\right) + \gamma \mathbf{E}^F \left[V(q,x) - V(q,\delta)\right],$$

where $\lambda_\theta = \lambda(1-\theta)$ denotes the bargaining-adjusted meeting intensity. Using this equation to derive the difference $R(\delta) = V(1,\delta) - V(0,\delta)$ reveals that the unmatched reservation value satisfies

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

Comparing this equation to (4.8) illustrates that the reservation value function in the model with long term relationships is the same as in the model with instantaneous relationships. This is due to the fact that $M(\delta,1) - M(\delta,0) = P$: when an investor contacts a dealer, the pair has instantaneous access to the interdealer market, so the value of owning the asset vs. not is just equal to the price of the asset, P. This was true before when investors and dealers could only engage in spot trade, and continues to be true now that they can form relationships. In particular, proceeding as in Section 4.3.2 shows that

$$R(\delta) = \frac{\lambda_\theta P}{r+\lambda_\theta} + \frac{1}{r+\gamma+\lambda_\theta}\left(\delta + \frac{\gamma \mathbf{E}^F[x]}{r+\lambda_\theta}\right) \tag{4.34}$$

and it follows that the demand of matched investors who take the interdealer price as given can be expressed as

$$\hat{q}(\delta;P) \equiv \mathbf{1}_{\left\{\delta > \hat{\delta}(P)\right\}}$$

with the threshold

$$\hat{\delta}(P) \equiv rP + \frac{\gamma\beta\left(rP - \mathbf{E}^F[x]\right)}{(r+\lambda_\theta)(r+\gamma+\beta+\lambda_\theta)}. \tag{4.35}$$

This expression naturally converges to the asset demand of the semi-centralized market as the average duration of trading relationships $1/\beta$ decreases to zero, and to the frictionless asset demand as the average bargaining adjusted duration of unmatched periods $1/\lambda_\theta$ decreases to zero.

Whether shorter relationships (smaller $1/\beta$) reduce asset demand (larger $\hat{\delta}(P)$) depends on the sign of $rP - \mathbf{E}^F[x]$. If this term is positive then a reduction in the duration of relationships reduces demand, and otherwise it increases it. The intuition is similar to the case of spot trade. An investor's demand is driven by two considerations. First, while she remains in a relationship, the investor compares the rental cost of the asset, rP, with her current utility flow, δ. Second, as she anticipates that the relationship may terminate, the investor compares the rental cost with her expected utility flow over the period between termination and the next opportunity to form a relationship. This expected utility flow depends on both δ and $\mathbf{E}^F[x]$. If $\mathbf{E}^F[x]$ is smaller than rP, then the investor expects a relatively low utility type upon termination, which increases the threshold type $\hat{\delta}(P)$, reduces demand and the more so when the relationship has a shorter duration.

4.6.3 Asset allocation

Since trading relationships are created at rate λ and destroyed at rate β, the distributions of utility flows $(\Phi_m(\delta), \Phi_u(\delta))$ within the population of matched and unmatched investors must satisfy the inflow-outflow equation

$$(\gamma + \lambda)\Phi_u(\delta) = \gamma\Phi_u(\bar{\delta})F(\delta) + \beta\Phi_m(\delta) \tag{4.36}$$

subject to the consistency condition

$$\Phi_u(\delta) + \Phi_m(\delta) = F(\delta).$$

The term on the left of (4.36) captures exits from the group of unmatched investors with utility flow lower than δ that result from preference shocks and matching with dealers. The two terms on the right-hand side of (4.36) capture entrants into the group of unmatched investors that result from preference shocks and relationship break ups, respectively.

Solving this system shows that the distributions of utility flows within the population of matched and unmatched investors are simply scaled versions of the distribution in the overall population:

$$\Phi_m(\delta) = F(\delta) - \Phi_u(\delta) = \frac{\lambda F(\delta)}{\beta + \lambda}.$$

Next, since matched investors participate in a frictionless market, we have that the distribution of utility flows among matched asset owners is given by the efficient allocation:

$$\Phi_{m1}(\delta) = \Phi_m(\bar{\delta})\left(F(\delta) - 1 + s_m\right)^+ \tag{4.37}$$

where the constant $s_m \equiv \Phi_{m1}(\bar{\delta})/\Phi_m(\bar{\delta})$ denotes the asset supply per capita in the population of matched investors. On the other hand, the inflow-outflow equation for the distribution among unmatched owners is given by

$$(\gamma + \lambda)\Phi_{u1}(\delta) = \gamma\Phi_{u1}(\bar{\delta})F(\delta) + \beta\Phi_{m1}(\delta) \tag{4.38}$$

subject to the market clearing condition

$$\Phi_{u1}(\bar{\delta}) + \Phi_{m1}(\bar{\delta}) = s. \tag{4.39}$$

Substituting (4.39) into (4.38) evaluated at $\bar{\delta}$ reveals that $s_m = s$ so that the supply per capita in the populations of matched and unmatched investors are the same as in the overal population.

Plugging this constant back into (4.37), substituting the result into (4.38), and solving then gives

$$\Phi_{u1}(\delta) = \frac{s\beta}{\beta + \lambda}\left(\frac{\gamma}{\gamma + \lambda}F(\delta) + \frac{\lambda}{s(\gamma + \lambda)}\left(F(\delta) - 1 + s\right)^+\right).$$

To understand this expression note that, at any given time the population of unmatched owners can be split into two groups. A fraction $\lambda/(\gamma + \lambda)$ of unmatched owners have the same utility flow as when their match broke up and the distribution of utility flows in that group is the frictionless one. On the other hand, the distribution of utility flows among the remaining unmatched owners coincides with the population distribution since investors in that group have experienced one or more preference shocks since the breakup of their last trading relationship. A similar intuition naturally also applies to the asset allocation (4.24) of the model with instantaneous relationships.

Finally, combining this expression with (4.37) shows that the distribution of utility flows among *all* asset owners is

$$\begin{aligned}\Phi_1(\delta) &\equiv \Phi_{u1}(\delta) + \Phi_{m1}(\delta) \\ &= \frac{\gamma\beta}{(\beta+\lambda)(\gamma+\lambda)} sF(\delta) + \left(1 - \frac{\gamma\beta}{(\beta+\lambda)(\gamma+\lambda)}\right)(F(\delta) - 1 + s)^+\end{aligned}$$

This expression shows that, as in previous sections, the allocation is a convex combination of the purely random allocation that would prevail absent trading and the efficient allocation that would prevail absent search frictions. The weights in this combination reflect the frictions at play in the market and can be easily compared to those of the semi-centralized model with instantaneous relationships. In particular, since

$$\frac{\gamma\beta}{(\beta+\lambda)(\gamma+\lambda)} < \frac{\gamma}{\gamma+\lambda}$$

the above combination puts less weight on the purely random allocation and it follows that lasting relationships improve the efficiency of the allocation. The intuition is simple: With long-term relationships only unmatched investors are subject to misallocation and the asset allocation within that group is the same as in the semi-centralized market with instantaneous relations. Therefore, the market with long-term relationships is less prone to misallocation.

4.6.4 The interdealer price

For a given interdealer price P, the aggregate flow of demand received by the dealer sector at any given point in time is

$$\lambda\left(\Phi_{u0}(\bar{\delta}) - \Phi_{u0}(\hat{\delta}(P))\right) + \gamma\Phi_{m0}(\bar{\delta})\left(1 - F(\hat{\delta}(P))\right),$$

where the first term is the demand from unmatched non-owners with utility flow above $\hat{\delta}(P)$ who have just met a dealer and the second term captures the demand from matched non owners whose utility flow has just been reset to some value above $\hat{\delta}(P)$. Similarly, the aggregate flow of asset supply received by the dealer sector is

$$\lambda\Phi_{u1}(\hat{\delta}(P)) + \gamma\Phi_{m1}(\bar{\delta})F(\hat{\delta}(P)).$$

Equating the two and using

$$\Phi_{n0}(\delta) + \Phi_{n1}(\delta) = \Phi_n(\delta), \qquad (n, \delta) \in \{u, m\} \times \mathcal{D}$$

reveals that the market clearing condition boils down to

$$\hat{\delta}(P) = \delta^\star \equiv F^{-1}(1 - s).$$

Using (4.35) to invert this relation shows that the equilibrium interdealer price is explicitly given by

$$P = \frac{\beta}{r + \beta + \lambda_\theta} P_\infty + \frac{r + \lambda_\theta}{r + \beta + \lambda_\theta} P^\star$$

where

$$P_\infty \equiv \frac{\gamma}{r + \gamma + \lambda_\theta} \mathbf{E}^F[\delta/r] + \frac{r + \lambda_\theta}{r + \gamma + \lambda_\theta} P^\star$$

denotes the interdealer price in the semi-centralized market with instantaneous relationships (see (4.25)) and $P^\star \equiv \delta^\star/r$ denotes the frictionless price. An immediate implication is that the deviation between the interdealer price with long-term relationships and the frictionless price,

$$\begin{aligned} P - P^\star &= \frac{\beta}{r + \beta + \lambda_\theta} \left(P_\infty - P^\star\right) \\ &= \frac{\beta\gamma}{r(r + \beta + \lambda_\theta)(r + \gamma + \lambda_\theta)} \left(\mathbf{E}^F[\delta] - \delta^\star\right), \end{aligned}$$

has the same sign but a smaller magnitude than the deviation that prevails in the market without long-term relationships. In particular, in both models, there is a net liquidity discount if and only if the marginal utility flow $\delta^\star$ exceeds the average, and the magnitude of this discount decreases with the average duration of relationships.

4.7 EXERCISES

Exercise 4.6. Consider the frictionless model of Section 4.2 with a continuous $F(\delta)$ and assume that the initial distribution of utility flows in the population is given by some continuous $F_0(\delta)$.

1. Calculate the distribution $F_t(\delta)$ of utility flows at every $t \geq 0$.
2. Derive the investor's objective function when the price is a bounded and continuously differentiable function $p : \mathbf{R}_+ \to \mathbf{R}$ instead of a constant.
3. Derive the equilibrium equation for the marginal type $\delta_t^\star$ at every $t \geq 0$.
4. Show that the equilibrium price $p_t^\star$ satisfies an ODE and derive an explicit expression for the unique bounded solution to this equation.

Exercise 4.7. Consider the frictionless benchmark of Section 4.2 and assume that investors only trade immediately after receiving a preference shock. Recall that, in equilibrium, owners with utility flow $\delta < \delta^\star$ strictly prefer to sell, non owners with utility flow $\delta > \delta^\star$ strictly prefer to buy, and investors with utility flow $\delta = \delta^\star$ are indifferent between buying and selling.

Denote by $\alpha_q \in [0, 1]$ the fraction of investors holding q units of the asset who trade after drawing utility flow $\delta^\star$.

1. Calculate the sell and buy order flows as functions of α_1 and α_0.
2. Use the market clearing condition (the equality of buy and sell order flow) to derive the equilibrium condition that links α_1 and α_0.
3. Show that this equilibrium condition admits infinitely many solutions in $[0, 1]^2$. Explain why.
4. Derive the solution that minimizes trading volume.

Exercise 4.8. Consider a semi-centralized model with a general distribution in which the marginal type $\delta^\star$ is unique and such that $\Delta F(\delta^\star) \neq 0$. Denote by $\alpha_q \in [0, 1]$ the fraction of investors holding q units of the asset who trade after drawing utility flow $\delta^\star$.

1. Calculate the sell and buy order flows as functions of α_1 and α_0.
2. Use the market clearing condition (the equality of buy and sell order flow) to derive the equilibrium condition that links α_1 and α_0.
3. Show that this condition admits infinitely many solutions in $[0, 1]^2$. Under what condition can we have $\alpha_0 = 0$ or $\alpha_1 = 0$?

4. Derive the solution that minimizes trading volume as well as the induced trading volume v.
5. Derive the solution that maximizes trading volume under the assumption that each meeting results in at most one trade.

Exercise 4.9. Consider a semi-centralized model with a general distribution in which the marginal type $\delta^\star$ is unique and such that $\Delta F(\delta^\star) \neq 0$. Let (α_0, α_1) be as in Exercise 4.8 and assume that we choose these execution probabilities to minimize trading volume.

1. Show that the distribution of utility flows within the groups of owners who would trade upon meeting a dealer is given

$$\frac{1}{m}\left(\mathbf{1}_{\{\delta<\delta^\star\}} dF(\delta) + \mathbf{1}_{\{\delta=\delta^\star\}} \beta \Delta F(\delta^\star)\right)$$

for some constants m and β to be determined.
2. Use the previous result to derive the average bid-to-interdealer spread.
3. Proceed as in the two previous questions to derive the distribution of utility flows in the group of non owners who would trade upon meeting a dealer, the average ask-to-interdealer spread, and the average bid-ask spread

Exercise 4.10. We have shown in the text that an increase in λ_θ increases the interdealer price if and only if $\delta^\star > \mathbf{E}^F[\delta]$. The goal of this exercise is to develop more intuition for this result based on the marginal investor's option value of search. The starting point of the analysis is the equation

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

1. Show that, for the marginal investor:

$$\frac{\partial R}{\partial \lambda_\theta}(\delta^\star) \propto \mathbf{E}^F\left[\frac{\partial R}{\partial \lambda_\theta}(\delta)\right].$$

Argue that this means that the marginal investor does not care about the option value of continuing search while keeping its current type, $\delta = \delta^\star$, but only after its next utility flow reset time. Explain why.

2. Show that

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

Interpret this formula in terms of search options to sell and buy upon the next utility flow re-set time.

3. Use the logic of search options to provide an explanation of the effect of a reduction in frictions on the interdealer price.

Exercise 4.11. Consider the same model of long-term relationships as in Section 4.6 but assume that only a fraction $\pi \in [0, 1]$ of meetings between unmatched investors and dealers lead to a long-term relationship.

1. Show that a matched investor holds the asset if and only if his utility flow is such that (4.33) holds.
2. Show that equation (4.34) continues to hold and use it to determine the asset demands of matched and unmatched investors.
3. Determine the distribution of utility flows $\Phi_{nq}(\delta)$ in the populations of matched $(n = m)$ and unmatched $(n = u)$ investors with holding $q \in \{0, 1\}$.
4. Compute the equilibrium interdealer price and study its dependence on the probability π that a meeting initiates a lasting relation.
5. Compute the trading volume generated by off meetings and that generated by relationships.

4.8 NOTES AND REFERENCES

There is a long history of using search models of decentralized trade to study equilibrium asset pricing and intermediation. A non-exhaustive list includes Rubinstein and Wolinsky [1987], Yavaş [1992], Gehrig [1993], Spulber [1996], Li [1998], LeRoy and Krainer [2002], Rust and Hall [2003], Shevchenko [2004], and Masters [2007], among many others. The specific model studied in this chapter was originally introduced by Duffie, Gârleanu, and Pedersen [2005]. While they assume that investors could have just two possible utility flows, $\delta \in \{\delta_L, \delta_H\}$, in this chapter we generalize their analysis to the case of an arbitrary distribution of utility flows.

The semi-centralized market model studied here takes one element of the market structure as exogenous—namely, that inter-dealer trades are frictionless while customer-to-dealer trades are subject to search frictions. A related market structure emerges endogenously in Farboodi, Jarosch, and Shimer [2022]. In their model, after making a costly choice of search intensity, identical investors contact each other at random in a market subject to search frictions. In equilibrium, they find that ex ante homogeneous investors actually choose different search intensities and, if search costs are linear, the set of equilibrium search intensities is unbounded. Hence, in this case, an infinitesimally small fraction of investors have infinite search intensity, so that the equilibrium behavior of these investors endogenously resembles that of the dealers in this chapter.

Also note that, in this environment with a frictionless inter-dealer market, the assumption that dealers cannot hold inventory is without loss of generality for the analysis of the steady-state equilibrium. Indeed, as long as dealers are not the final holders of the asset—in the sense that their utility flows are sufficiently small relative to that of investors —they find it optimal to hold no inventory. However, as we show later, it is important to allow for dealer inventories in the steady state analysis of environments where the interdealer market is frictional (Chapters 5 and 7) and in the analysis of out-of-steady-state dynamics following an aggregate liquidity shock (Chapter 9).

The semi-centralized market model uses a particular matching function, which determines the mapping between the masses of various types of market participants and the total (flow) number of meetings between them. In the formulation studied in this chapter, this function is simply $\lambda\mu_i$, where μ_i is the mass of investors, so that each investor meets a dealer at rate $\lambda\mu_i/\mu_i = \lambda$. In applications of search models to other markets, it is commonly assumed that the matching function has constant returns to scale. While this assumption is supported by empirical evidence in labour markets [Petrongolo and Pissarides, 2001], it seems less appropriate in the context of financial markets due the presence of scale effects such as the positive relationship between issue size (or asset supply) and various measures of liquidity. For this reason, some of the literature has considered matching functions that exhibit increasing returns [such as Pagano, 1989, Vayanos and Wang, 2007, Weill, 2008, Vayanos and Weill, 2008, Shen, Wei, and Yan, 2020, Sambalaibat, 2022]. A related matching technology with scale effects arises in the so-called "stock-flow" model of Coles and Smith [1992], An and Zheng [2018] and An [2019], in which the inflow of new participants on one side of the market matches with unmatched participants on the other side of the market. In this environment, trading delays do not arise from search frictions per se, but rather because each buyer is only willing to purchase a finite collection of assets, drawn at random according to some point process. As a result, the demand for any particular asset takes time to materialize and, in equilibrium, a stock of buyers and sellers is waiting in the market for suitable counterparties to arrive.

Another difference between search-theoretic models of labor and financial markets is that the former treat matches as being long-term relationships while the latter almost always assume that matches are dissolved immediately after the parties decide whether or not to trade. In some OTC markets, this is clearly counterfactual, as certain investors appear to repeatedly trade with the same counterpary. This observation draws attention to the need for better models of relationships in OTC markets; the simple of Section 4.6 is inspired by Maciocco [2023], who develops a closely related framework with divisible asset holdings to study relationship vs. spot trading of municipal bonds.

Note that this chapter also focuses on the case of a *single* asset trading in an OTC market, which is highly tractable and constitutes an important step in understanding the impact of OTC market frictions on liquidity and asset prices. Korenok, Kospentaris, and Lightle [2024] have recently proposed an experimental evaluation of the model-implied relationship between trading speed and prices. Consistent with the model, they find that an increase in the trading speed causes a rise (fall) in the price when the supply is small (large). However, to further test the theory in a non-experimental setting, one would need to establish closer contact with *cross-sectional* empirical evidence. For example, Amihud, Mendelson, and Pedersen [2005] (Chapter 3) document the cross-sectional relationship between risk-adjusted returns and measures of liquidity such as turnover, volume, or bid-ask spreads, while Amihud and Mendelson [1991] provide evidence that liquidity is related to violations of no-arbitrage relationships. To fully address these patterns, one would ideally want to incorporate multiple assets into the model, to explain why an investor's ability to choose between payoff-equivalent assets does not undo an arbitrage relationship or equalize risk-adjusted return differentials. Search-theoretic models that address these questions include Vayanos and Wang [2007], Weill [2008], Vayanos and Weill [2008], Sambalaibat [2022], Milbradt [2017], Üslü and Velioğlu [2019] and Li and Yu [2021].

In this chapter—as in the remainder of this book—we study OTC markets *in isolation* from the rest of the economy. However, an important strand of the literature analyzes OTC markets *within a broader macroeconomic context*, which is important for at least two reasons. First, it shows that OTC market frictions matter for broader economic questions. Second, it makes normative analyses more meaningful, since it can account for the welfare of all economic agents and not only that of OTC market participants. We highlight several branches of this literature. The new monetarist literature [see Nosal and Rocheteau, 2011, for an overview] considers a general equilibrium framework in which agents exchange assets for money in an OTC market, before exchanging money for goods and services in a separate market; see, for example, Geromichalos and Herrenbrueck [2016], Lagos and Zhang [2020], Lagos and Zhang [2019a], Lagos and Zhang [2019b], Geromichalos, Herrenbrueck, and Salyer [2016], Mattesini and Nosal [2016], and Lebeau [2019]. A large literature focuses on the cost of capital of firms who issue debt in OTC markets; see, for example, He and Milbradt [2014], Hugonnier, Malamud, and Morellec [2015], Chen, Cui, He, and Milbradt [2018], d'Avernas [2017], Arseneau, Rappoport, and Vardoulakis [2017], Bethune, Sultanum, and Trachter [2019], Cui and Radde [2020], Roh [2019], Kozlowski [2018], and Chang [2022]. Bianchi and Bigio [2022], Bigio and Sannikov [2019], and Malamud and Schrimpf [2017] incorporate OTC markets into larger macroeconomic models to analyze the impact of monetary policy on aggregate outcomes. Finally, an emerging literature utilizes some of the models developed in this book to study international implications; see, for example, Geromichalos and Jung [2018], Malamud and Schrimpf [2018] Bianchi, Bigio, and Engel [2018], Passadore and Xu [2018] and Chaumont [2018].