The Emergence of Market Structure*

Maryam Farboodi† MIT
Gregor Jarosch‡ Princeton University
Robert Shimer§ University of Chicago

August 18, 2021

Abstract

We study a model of over-the-counter trading in which ex-ante identical traders invest in a contact technology and participate in bilateral trade. We show that a rich market structure emerges both in equilibrium and in an optimal allocation. There is continuous heterogeneity in market access under weak regularity conditions. If the cost per contact is constant, heterogeneity is governed by a power law and there are middlemen, market participants with unboundedly high contact rates who account for a positive fraction of meetings. Externalities lead to overinvestment in equilibrium, and policies that reduce investment in the contact technology can improve welfare. We relate our findings to important features of real-world trading networks.

Keywords: Over-the-Counter Markets, Intermediation, Middlemen, Random Matching, Endogenous Search Intensity, Network Formation, Pareto Distribution, Welfare

*This paper was previously circulated under the title "Meeting Technologies in Decentralized Asset Markets." We are grateful to Rebekah Dix for research assistance, and to Fernando Alvarez, Markus Brunnermeier, Xavier Gabaix, Ricardo Lagos, Moritz Lenel, Hugo Lhuillier, Pierre-Olivier Weill, four anonymous referees and the editor, and audiences at various seminars and conferences for their thoughts and comments.

†farboodi@mit.edu
‡gjarosch@princeton.edu
§shimer@uchicago.edu

1 Introduction

In many over-the-counter markets, some participants trade much more frequently and with many more partners than others do (see, e.g., Bech and Atalay, 2010; Craig and Von Peter, 2014; Fricke and Lux, 2015; Hollifield, Neklyudov and Spatt, 2017). We are interested in understanding why such a market structure exists and in assessing its normative properties.

We do this by examining an over-the-counter market for assets where market participants periodically meet in pairs with the opportunity to trade (Rubinstein and Wolinsky, 1987). We assume that ex-ante identical market participants make a costly investment which governs how often they are in bilateral contact with others. Whenever two participants are in contact, they may trade an asset for an outside good, as in Duffie, Gârleanu and Pedersen (2005) and a large ensuing literature. We find that a market structure with rich heterogeneity emerges naturally in this environment, creating specialized intermediaries who mitigate trading frictions while economizing on aggregate investment costs.

We verify that if market participants have heterogeneous contact rates, intermediation arises naturally and participants who are more often in contact with others act as intermediaries (Üslü, 2019). By intermediation, we mean that fast market participants buy assets with the sole intent to quickly resell them and sell assets with the sole intent to quickly repurchase them. While there are no static efficiency gains from intermediation, these trades move assets towards fast traders, which improves the economy's future trading opportunities.

Our novel contribution is that we treat participants' contact rates as an investment choice and characterize the resulting market structure in a decentralized equilibrium and a Pareto optimal allocation. In equilibrium, market participants choose their contact rate to maximize future profits net of the investment cost. In an optimal allocation, the contact rate distribution maximizes the social value of trade net of the investment cost. Although the two allocations are qualitatively similar, we find that equilibrium is inefficient due to congestion and thick-market externalities.

We highlight several key results on market structure in both the equilibrium and optimal allocations which connect our theory to stylized facts about real world trading networks. First, we prove that heterogeneity in contact rates arises endogenously, with ex-ante homogeneous participants making heterogeneous investments. Specifically, we show that *any* distribution of contact rates can be rationalized through an appropriate investment cost function (Propositions 1 for equilibrium and 1-P for Pareto optimum). Conversely, we show that traders make dispersed investment decisions such that the contact rate distribution is continuous without interior mass when the investment cost function is differentiable (Propositions 2 and 2-P).

The force pushing towards heterogeneity is the gains from intermediation. If everyone else chooses the same contact rate, a trader who sets a slightly higher contact rate acts as an intermediary for everyone else. The (private or social) value of each intermediation trade is proportional to the difference in contact rates, and hence linearly increasing in the deviating trader's contact rate. Conversely, when choosing a slightly lower contact rate, all other traders intermediate for the slower trader. Again, the (private or social) value from each intermediation trade is proportional to the difference in contact rates, hence linearly *decreasing* in the slower trader's contact rate. The gains from intermediation thus generates a convex kink in the value function at the mass point, which creates an incentive to choose a different contact rate from everyone else. Consequently, our theory aligns with the observed heterogeneity in trading activity we mentioned in the first sentence.

Second, we characterize the equilibrium and optimal contact rate distribution when there is an exogenous upper bound on contacts, up to which the cost of each contact is constant. If the cost of a contact is neither too high nor too low, we prove that the contact rate distribution is continuous on a convex support extending from a strictly positive endogenous lower bound to the exogenous upper bound (Propositions 3 and 3-P).

We then take the limit of equilibrium and optimal allocations as the exogenous upper bound on contacts goes to infinity. We prove that the limiting distribution of contact rates is Pareto with a tail index of 2 (Propositions 4 and 4-P). Additionally, we call someone a *middleman* if their contact rate is infinite. Although almost no one is a middleman in either an equilibrium or optimal allocation, we prove that middlemen account for a positive fraction of meetings and trades. More precisely, there is a strictly positive probability that a counterparty in any meeting or trade has a contact rate exceeding any finite threshold (still Propositions 4 and 4-P and Corollaries 1 and 1-P).[1]

We argue that both the Pareto tail and middlemen connect with evidence. A host of empirical work documents a multi-tiered network structure with a small core and many layers of intermediaries, as well as a power law, namely a Pareto tail in the distribution of the number of trading partners. See, for example, Craig and Von Peter (2014) for the German interbank market and Bech and Atalay (2010) for the federal funds market.

Third, we show that when the cost of each contact is constant but arbitrarily small, the average contact rate is arbitrarily large. Despite this, heterogeneity and intermediation

[1]Duffie, Gârleanu and Pedersen (2005) and the ensuing literature frequently assume the existence of marketmakers with access to a frictionless interdealer market. The middlemen who emerge endogenously in our environment share many features with these marketmakers: other traders stochastically meet middlemen and bargain over the terms of trade, while there is continuous trade among middlemen. One difference is that our middlemen have the same preferences as other traders and in particular care about their asset position. In contrast, the literature following Rubinstein and Wolinsky (1987) and Duffie, Gârleanu and Pedersen (2005) assumes that marketmakers only value their trading profits.

survive in this limit. Each shock to a market participant's idiosyncratic valuation for an asset leads to more than four trades on average as the asset gets reallocated through intermediation chains that typically involve a middleman (Propositions 5 and 5-P). This is consistent with the observation that, despite recent advances in information technology which have reduced search frictions, intermediation remains a hallmark of many decentralized marketplaces (see, for example, Kuprianov, 1993; Philippon, 2015; Biais and Green, 2019).

Finally, we stress that heterogeneity and intermediation are intimately connected. Prior research, e.g. Üslü (2019) and Nekludyov (2019), established that heterogeneity in market access creates a role for intermediation. We show that if intermediation is prohibited, all market participants choose the same contact rate, both in equilibrium and in the optimum (Proposition 6). That is, without heterogeneity there is no intermediation, and without intermediation there is no heterogeneity.

Our main contribution is to endogenize a market structure with rich heterogeneity. After establishing the relevant propositions, we relate the results to an empirical literature which identifies a core of a few highly connected traders, a larger number of peripheral traders that still frequently intermediate, and heterogeneity in market access that is well-described by a Pareto distribution.

Second, our approach offers a natural way to predict the endogenous response of market structure to technological change, such as a reduction in the cost of contacts. We capture this most clearly in our limiting economy where the cost of each contact is arbitrarily small. This allows us to speak to the increasing prominence of financial intermediation, including in decentralized asset markets, despite improving information technologies.

Our approach also offers a direct assessment of the efficiency of endogenous market structure and how those vary with technological parameters. For instance, we show formally that the equilibrium market structure is inefficient. Using Pigouvian taxes, we show that optimal policy taxes frequent traders more per trade than peripheral ones. Numerically, we show that trading volume is excessive in equilibrium and connect these observations with the discussion on financial transaction taxes and regulatory intervention (Tobin, 1978; Burman, Gale, Gault, Kim, Nunns and Rosenthal, 2016).

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes our model. Section 4 defines equilibrium while Sections 5 and 6 characterize it. Section 7 analyzes the Pareto optimal allocation and the nature of search externalities. Section 8 establishes that contact rate dispersion disappears if intermediation is prohibited. Section 9 concludes.

2 Related Work

Our paper is a particular development of stochastic network formation models in a frictional trading context. As such, it is related to both the large literature on trade and intermediation in frictional asset markets and to stochastic network formation models in other settings such as disease transmission or social networks. We will review each in turn.

Rubinstein and Wolinsky (1987) were the first to model intermediation in a frictional goods market. We share with them the notion that intermediaries have access to a superior search technology. In two important papers, Duffie, Gârleanu and Pedersen (2005, 2007) study an over-the-counter asset market where time-varying taste leads to trade. This is also the fundamental force giving rise to gains from trade in our setup.

Much of the more recent theoretical work extends the Duffie et al. framework to accommodate newly-available evidence on trade and intermediation in over-the-counter markets; see Weill (2020) for a recent survey. Üslü (2019) allows for rich heterogeneity in contact rates, pricing, and inventory holdings in a market where traders have continuously distributed flow payoffs.[2] As in our framework, fast dealers are more willing to take on misaligned asset positions, thus emerging as intermediaries. The marketplace features intermediation chains and a core-periphery trading network. Our contribution to this literature is to show that heterogeneous contact rates arise endogenously to leverage the gains from intermediation even with ex-ante homogeneous traders. We further show how the endogenous choice of contact rates given a cost function disciplines key features of the contact rate distribution. Additionally, our normative analysis shows that both technological heterogeneity and intermediation by those with a high contact rate are socially desirable.

Hugonnier, Lester and Weill (2020) model a market with separate dealer and costumer sectors, where dispersion in flow payoffs gives rise to intermediation chains among dealers. Afonso and Lagos (2015) similarly have endogenous intermediation because banks with heterogeneous asset positions buy and sell depending on their counterparties' reserve holdings. In contrast to these setups, we offer a theory of endogenous heterogeneity which is rooted in the gains from intermediation.

Farboodi, Jarosch, Menzio and Wiriadinata (2019) model an environment where some traders have superior bargaining power and emerge as middlemen due to dynamic rent extraction motives which are, at best, neutral for welfare. In contrast, intermediation in our setup improves the allocation since misaligned asset positions are traded toward those who are more efficient at offsetting them. They also study an initial investment stage which de-

[2] A related literature studies the positive and normative consequences of high-frequency trading in centralized financial markets; see, for instance, Pagnotta and Philippon (2018). The decentralized interdealer market in Nekludyov (2019) also features dealers with exogenously given heterogeneous contact rates.

termines the distribution of bargaining power in the population, but restrict the distribution to two points. We allow for a continuous distribution of contact rates and prove that this is consistent with both equilibrium and optimum.

Furthermore, some of the theoretical work on intermediation in over-the-counter markets assumes the existence of middlemen who facilitate trade through their continuous access to an interdealer market (Duffie, Gârleanu and Pedersen, 2005; Weill, 2008; Lagos and Rocheteau, 2009). We show that middlemen are a natural outcome when homogeneous traders invest in contact rates and the marginal cost of contacts is constant.

Three recent papers endogenize market structure using a search framework. Hendershott, Li, Livdan and Schürhoff (2020) model a client-dealer network in which clients cannot act as intermediaries themselves, but choose the number of dealers they contact. They find that clients choose homogeneous contact rates. In Chang and Zhang (2019) there are gains from concentrating misallocated positions, which in turn gives rise to endogenous intermediation and market structure. Dugast, Üslü and Weill (2019) ask which agents prefer to trade in a centralized and multilateral versus in a decentralized and bilateral fashion. We assume all trade takes place in decentralized markets, but endogenize contact rates, which are exogenous in Dugast, Üslü and Weill (2019).

Second, the paper relates to a literature on stochastic network formation models where agents need to make an upfront investment decision. These models share with ours the notion that agents make a costly contact intensity choice in a frictional, bilateral setting. Currarini, Jackson and Pin (2009) develop a frictional model of friendship formation, applied to homophily and segregation. Indviduals decide on how long to partake in a costly stochastic matching process which is similar to our assumption of a costly contact rate. A key difference, however, is that agents of the same type in that model make symmetric choices and ex-post heterogeneity is driven solely by the presence of different types and shocks.

Cabrales, Calvó-Armengol and Zenou (2011) consider a framework where agents choose random social interactions and investment simultaneously. The payoff is quadratic, depending on both partners' investments, differing from our linear meeting technology. More importantly, while individuals differ in their private returns to investment and hence make different choices the paper considers only symmetric equilibria where identical agents make identical choices. This also applies to the remaining papers discussed in this section.

Kremer (1996) is an early contribution to the literature on disease transmission that integrates behavior into an epidemiological model. Agents choose a rate of partner change that gives utility yet comes with a higher risk of HIV infection. Our paper Farboodi, Jarosch and Shimer (2021) similarly models the choice of social activity in the context of the Covid-19 pandemic. These papers have in common that they model endogenous contact intensity

that leads to bilateral transmission (see also Quercioli and Smith, 2006).

Similarly, Duffie, Malamud and Manso (2009) study a search setting where agents make costly effort choices in their search for information which percolates via bilateral meetings. Cross-sectional heterogeneity in search effort arises due to heterogeneity in current information.

In Galeotti and Merlino (2014), workers choose how much information to obtain about job opprtunities. They do so by making a costly investement decision into connections with others along the lines of Cabrales, Calvó-Armengol and Zenou (2011).

Finally, our results on endogenous heterogeneity in contact rates superficially resemble a literature showing the absence of a pure strategy equilibrium in search models (Butters, 1977; Burdett and Judd, 1983; Burdett and Mortensen, 1998; Duffie, Dworczak and Zhu, 2017). These papers have in common that if all firms charge the same price (or offer the same wage), firms that offer a slightly lower price (higher wage) earn discontinuously higher profits. Our results concern a different object, the contact rate distribution, and we find that the profit function is continuous but not differentiable. More fundamentally, all of the papers referenced in this paragraph show that equilibria are asymmetric. We demonstrate that both the equilibrium and the socially optimal allocation are asymmetric. Thus our results do not reflect a particular assumption about price formation, but rather demonstrate how the possibility of providing or using intermediation services creates a reason for ex ante identical traders to make heterogeneous investments.

3 Model

We study an economy where time is continuous and extends forever. We focus throughout on an aggregate steady state. There is a unit measure of market participants, hereafter traders, who each have preferences defined over their holdings of an indivisible asset in fixed supply and their consumption or production of an outside good. Traders exit the market when hit by an idiosyncratic shock with arrival rate $r > 0$. When a trader exits, she is replaced with a newborn trader so as to keep the population fixed at 1.[3]

3.1 Asset Holdings and Preferences

Traders' asset holdings and preferences follow Duffie, Gârleanu and Pedersen (2005). An individual trader's asset holding is restricted to be $b \in \{0, 1\}$. Traders have time-varying taste $i \in \{h, l\}$ for the asset and receive flow utility $\delta_{i,b}$ when they are in state (i, b). We

[3]Vayanos and Wang (2007) is an early example of an asset market with stochastic exit and entry.

assume that $\Delta \equiv \frac{1}{2}\big(\delta_{h,1} + \delta_{l,0} - \delta_{h,0} - \delta_{l,1}\big)$ is strictly positive, which implies that traders in the high state are the natural asset owners.

Half of all traders are born in state $(h, 1)$ and half in state $(l, 0)$.[4] Thereafter, a trader's taste switches from l to h when hit by an idiosyncratic shock with arrival rate $\gamma > 0$ and back again at the same rate. Since this shock is idiosyncratic, half the traders are in state h and half are in state l in the stationary distribution. We similarly fix the supply of the asset at $\frac{1}{2}$, so at any point in time half the traders hold the asset and half do not. Thus, in a frictionless environment, the supply of assets is exactly enough to satiate the demand from traders with taste h.

Preferences over net consumption of the outside good are linear, so the outside good effectively serves as transferable utility when trading the asset. We assume that whether trade occurs and what the terms of trade are is determined according to the symmetric Nash bargaining solution. Traders discount the future only because of the exit probability r. When a trader exits holding the asset, it is transferred to a newborn trader with taste h, and the dying trader is not compensated.

3.2 Contact Technology

Asset trades occur pairwise in a frictional asset market. Newborn traders choose a time-invariant rate $\lambda \in \mathcal{X} \equiv [0, \bar{\lambda}]$ at which they make contact with another trader, where $\bar{\lambda}$ is an exogenous upper bound. A high contact rate is costly: a trader who chooses a contact rate λ pays an ex-ante cost $C(\lambda)$, where $C : \mathcal{X} \to \mathbb{R}$ is nondecreasing.[5] We allow different traders to choose different contact rates.[6]

A trader who chooses a contact rate λ meets a counterparty at rate λ. Search is random, so whom a trader meets is independent of her contact rate, taste, and asset holding. Let $\mathcal{B}$ be the Borel σ-algebra generated by $\mathcal{X}$ and μ_F be a probability measure on the measurable space $(\mathcal{B}, \mathcal{X})$ which gives the probability that, conditional on a meeting, the counterparty's contact rate is some $\lambda' \in S \subset \mathcal{B}$, with $\mu_F(\mathcal{X}) = 1$. This probability measure is a key equilibrium

[4] We can relax the assumption that all newborn traders are born in one of these two states, but it is convenient to assume that they do not know their state when they choose λ. We view this as reasonable because we do not think that the short-run desire to trade is an important determinant of the irreversible investment in λ. Preference shifts occur at a much higher frequency than exit, while trading opportunities in many markets occur at a higher frequency still. This implies that the initial state of new entrants will have little impact on the steady state distribution of asset holdings and tastes.

[5] In Section 6, we focus on the linear cost case, $C(\lambda) = c\lambda$. There we take a limit with $\bar{\lambda} \to \infty$.

[6] One possible interpretation of this assumption is that participants may increase their contact rate by investing in their communication capacity, either through improved information technology or by simply hiring more or more able individuals to staff their trading desk. An alternative interpretation is that market participants may invest into relationships with more counterparties. That is, they may invest time and resources to increase the length of their contact list.

object in our environment. For notational convenience, let F with $F(\lambda) \equiv \mu_F([0,\lambda])$ denote the cumulative distribution function associated with the measure μ_F. Conditional on meeting a counterparty with a given contact rate, the counterparty's taste and asset holding are drawn from the population distribution with that contact rate, independent of the trader's contact rate, taste, and asset holding.

One can think of this contact technology as the continuous-time, continuous-type limit of the following physical environment with n traders and discrete time periods of length $dt < 1/\bar{\lambda}$: Each period, each type λ trader accesses a market with i.i.d. probability λdt. If this process results in an odd number of traders accessing the market, one additional trader, chosen uniformly at random, is selected to also access the market. The traders who thus access the market are then randomly matched in pairs (see, e.g., Shimer and Smith, 2001). In a large economy, the chance of any trader matching because an odd number of traders accessed the market becomes negligible, and so a trader matches each period with probability λdt. Moreover, for the same reason, traders meet counterparties in proportion to the counterparties' contact rate.[7] We return to this at the end of Section 4 and then offer additional detail in Appendix A.2.

We will show that individual behavior depends only on the counterparty measure μ_F. Nevertheless, to connect the model to data, we also characterize the distribution of contact rates in the population. Let $\mu_G(S)$ denote the measure of traders whose contact rate is some $\lambda' \in S \subset \mathcal{B}$, with $\mu_G(\mathcal{X}) = 1$. Again, let G with $G(\lambda) \equiv \mu_G([0,\lambda])$ denote the associated cumulative distribution function. The measures μ_F and μ_G are related through the following transformation,[8]

$$\mu_F(S) \equiv \frac{\int_S \lambda d\mu_G(\lambda)}{\int_{\mathcal{X}} \lambda d\mu_G(\lambda)}. \tag{1}$$

This captures our assumption that traders meet counterparties in proportion to their contact rate. In words, the conditional probability of drawing a counterparty from a particular group of traders is given by the fraction of meetings that accrues to that group. Finally, we let the average contact rate in the population be denoted by $\Lambda \equiv \int_{\mathcal{X}} \lambda d\mu_G(\lambda)$.

[7] This is the natural counterpart to a frictional labor market where an employer meets job seekers in proportion to the job seekers' search intensity (see Petrongolo and Pissarides, 2001). An alternative is a "telephone-line" matching function where traders initiate contacts at some chosen rate λ and can also be contacted otherwise. The distribution of λ' among the counterparties then depends on who initiated the contact. Such a technology is inconsistent with our assumption that the distribution of whom a trader meets is independent of her contact rate.

[8] Appendix A.1 discusses technical details on how we move between the two probability measures, in particular how we deal with cases where equation (1) is not invertible. Here and throughout the paper, $\int_S \Omega(\lambda) d\mu_G(\lambda) = \int_S \Omega d\mu_G$ is the integral of Ω on $S \subset \mathcal{B}$ with respect to the measure μ_G. We use $\int_{\lambda_1}^{\lambda_2} \Omega(\lambda) d\lambda$ to denote the integral on $[\lambda_1, \lambda_2]$ with respect to the Lebesgue measure.

3.3 Individual Trader Behavior

We next turn to the behavior of individual traders. To do so, we first need some notation. Let $p_{\lambda,i,b}^{\lambda',i',b'} = p_{\lambda',i',b'}^{\lambda,i,b}$ denote the endogenous probability that a trader with contact rate $\lambda \in [0, \bar{\lambda}]$, taste $i \in \{h, l\}$, and asset holdings $b \in \{0, 1\}$ trades when she contacts a trader with contact rate $\lambda' \in [0, \bar{\lambda}]$, taste $i' \in \{l, h\}$, and asset holdings $b' \in \{0, 1\}$. Let $t_{\lambda,i,b}^{\lambda',i',b'} = -t_{\lambda',i',b'}^{\lambda,i,b}$ denote the endogenous transfer of the outside good from (λ, i, b) to (λ', i', b') when such a trade takes place. The trading probability and price are determined by Nash bargaining, as we discuss further in Section 3.4. Finally, let $\sigma_{\lambda,i,b}$ denote the endogenous fraction of traders with contact rate λ who have taste i and asset holding b. This is determined from a steady state condition which we discuss in Section 3.5.

Now let $v_{\lambda,i,b}$ denote the present value of a trader (λ, i, b). Given μ_F, p, t, and σ, the value function satisfies

$$rv_{\lambda,i,b} = \delta_{i,b} + \gamma\big(v_{\lambda,\sim i,b} - v_{\lambda,i,b}\big) + \lambda \int_{\mathcal{X}} \sum_{i' \in \{h,l\}} \sum_{b' \in \{0,1\}} \sigma_{\lambda',i',b'} p_{\lambda,i,b}^{\lambda',i',b'} \big(v_{\lambda,i,b'} - v_{\lambda,i,b} - t_{\lambda,i,b}^{\lambda',i',b'}\big) d\mu_F(\lambda'). \quad (2)$$

The left hand side of equation (2) is the flow value of the trader, where discounting reflects the exit rate. The value comes from three sources, listed in order on the right hand side. First, she receives a flow payoff $\delta_{i,b}$ which depends on her tastes and asset holdings. Second, her tastes shift from i to $\sim i$ at rate γ. Third, she meets another trader at rate λ with type λ' drawn from the counterparty measure μ_F, in which case they may swap asset holdings in return for a payment. Conditional on λ', the counterparty's state is (i', b') with probability $\sigma_{\lambda',i',b'}$. If there is trade, the trader has a capital gain from swapping assets and transferring the outside good, $v_{\lambda,i,b'} - v_{\lambda,i,b} - t_{\lambda,i,b}^{\lambda',i',b'}$.

Armed with the value function, we can formally describe the restriction which the model imposes on the contact rate distribution. Recall that traders choose their contact rate when they enter the market in order to maximize their expected value. This implies that any contact rate chosen in equilibrium must maximize

$$\pi_\lambda \equiv \tfrac{1}{2}(v_{\lambda,l,0} + v_{\lambda,h,1}) - C(\lambda). \quad (3)$$

This formulation reflects the fact that traders are equally likely to enter in the low taste state without the asset or the high taste state with the asset, and they choose λ accordingly subject to the investment cost C.

3.4 Terms of Trade

The terms of trade are set in accordance with a symmetric Nash bargaining solution. This means that trade occurs whenever doing so can make both parties better off, and that transfers equate the gains from trade without throwing away any resources. That is, if there is a transfer $t_{\lambda,i,b}^{\lambda',i',b'} = -t_{\lambda',i',b'}^{\lambda,i,b}$ satisfying $v_{\lambda,i,b'} - v_{\lambda,i,b} - t_{\lambda,i,b}^{\lambda',i',b'}$ and $v_{\lambda',i',b} - v_{\lambda',i',b'} - t_{\lambda',i',b'}^{\lambda,i,b}$ both positive, then trade occurs with a transfer such that $v_{\lambda,i,b'} - v_{\lambda,i,b} - t_{\lambda,i,b}^{\lambda',i',b'} = v_{\lambda',i',b} - v_{\lambda',i',b'} - t_{\lambda',i',b'}^{\lambda,i,b}$. If any feasible transfer implies a strict loss from trade, there is no trade.

With a little bit of algebra, this means that the trading probability is given by

$$p_{\lambda,i,b}^{\lambda',i',b'} = \begin{cases} 1 & \\ 0 & \end{cases} \quad \text{if } v_{\lambda,i,b'} + v_{\lambda',i',b} \gtrless v_{\lambda,i,b} + v_{\lambda',i',b'}; \tag{4}$$

and that when there is trade, the transfer satisfies

$$t_{\lambda,i,b}^{\lambda',i',b'} = \frac{1}{2}\left(v_{\lambda,i,b'} + v_{\lambda',i',b'} - v_{\lambda,i,b} - v_{\lambda',i',b}\right). \tag{5}$$

When $v_{\lambda,i,b'} + v_{\lambda',i',b} = v_{\lambda,i,b} + v_{\lambda',i',b'}$, trade may be probabilistic. If trade does occur, the transfer is still given by equation (5).

3.5 Stationary Distribution

The steady state fraction of type λ traders in different states, $\sigma_{\lambda,i,b}$, depends on the trading probabilities through the balance of inflows and outflows:

$$\begin{aligned}\left(r + \gamma + \lambda \int_{\mathcal{X}} \sum_{i' \in \{h,l\}} \sigma_{\lambda',i',1-b} p_{\lambda,i,b}^{\lambda',i',1-b} d\mu_F(\lambda')\right) \sigma_{\lambda,i,b} \\ = \gamma \sigma_{\lambda,\sim i,b} + \lambda \left(\int_{\mathcal{X}} \sum_{i' \in \{h,l\}} \sigma_{\lambda',i',b} p_{\lambda,i,1-b}^{\lambda',i',b} d\mu_F(\lambda')\right) \sigma_{\lambda,i,1-b} + \frac{r}{2} \mathbb{I}_{(i,b) \in \{(h,1),(l,0)\}}.\end{aligned} \tag{6}$$

The left hand side of equation (6) measures the outflows from state (i, b) for traders with contact rate λ. A trader leaves the state either when she exits, when she has a taste shock, or when she trades with another trader with the opposite asset holding. The right hand side measures the inflows. A trader with contact rate λ enters state (i, b) when she has the opposite taste and has a taste shock, when she has the opposite asset holding and trades, or, if (i, b) is equal to either $(h, 1)$ or $(l, 0)$, half the time when she is newborn. Here the indicator function $\mathbb{I}$ is equal to 1 if the condition in the subscript holds and is zero otherwise.

3.6 Summary

The natural notion of equilibrium at this stage is given by the following *primitive* objects: a counterparty measure, value function, trading probability function, price function, and share function which jointly satisfy equations (2)–(6). However, we henceforth focus on a particular type of equilibrium, imposing an additional symmetry restriction that we spell out at the start of the next section. Such an equilibrium is far easier to characterize and depends only on three objects: the counterparty measure and two additional *reduced-form* functions. We verify below, however, that it is possible to recover the primitive objects which we have described thus far from these reduced-form objects.

4 Equilibrium

This section focuses on defining a symmetric equilibrium. To that end, we impose a restriction on the trading pattern which we keep in place for the remainder of the paper. We relegate formal derivations and details to Online Appendix C.

To begin with, we call traders' asset holding positions *misaligned* either when they hold the asset and have taste l, or when they do not hold the asset and have taste h. We call traders' asset holding positions *well-aligned* in the other two states. Let $m_\lambda \equiv \sigma_{\lambda,l,1} + \sigma_{\lambda,h,0}$ denote the fraction of traders with contact rate λ who are misaligned.

Symmetry means we restrict attention to equilibria in which the two misaligned states and the two well-aligned states are treated symmetrically. That is, we impose $p_{\lambda,i,0}^{\lambda',i',1} = p_{\lambda,\sim i,1}^{\lambda',\sim i',0}$ for all λ, $i \neq \sim i$, and $i' \neq \sim i'$. This means that if a type λ trader with taste i would buy the asset from a type λ' trader with taste i', then a type λ trader with the opposite taste $\sim i$ would sell the asset to a type λ' trader with the opposite taste $\sim i'$.

In Online Appendix C.1 we prove that when trading probabilities are symmetric in this sense, misalignment rates are also symmetric, $\sigma_{\lambda,l,b} = \sigma_{\lambda,h,1-b}$ for all λ and b. Moreover, we prove the existence of a *surplus function* $s(\lambda) = v_{\lambda,h,1} - v_{\lambda,h,0} - q = v_{\lambda,l,0} - v_{\lambda,l,1} + q$ where

$$q \equiv \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0}}{2r}, \tag{7}$$

independent of λ. The surplus function tells us the value of being well-aligned, up to the additive constant q, which equals the price of the asset in the frictionless limit. Manipulating the value function (2) as well as the Nash bargaining solution (4) and (5), we derive in Online

Appendix C.1 a Bellman equation for s:

$$(r+2\gamma)s(\lambda) = \Delta + \frac{\lambda}{4}\int_{\mathcal{X}} \Big(\big((s(\lambda')-s(\lambda))^+ - (s(\lambda)+s(\lambda'))^+\big) m_{\lambda'} + \big((-s(\lambda)-s(\lambda'))^+ - (s(\lambda)-s(\lambda'))^+\big)(1-m_{\lambda'})\Big) d\mu_F(\lambda') \quad (8)$$

where $z^+ \equiv \max\{z, 0\}$ and reflects that meetings result in trade if and only if trade is bilaterally efficient. The $r + 2\gamma$ on the left hand side reflects discounting due to exit and taste shocks. Δ is the average difference in flow payoffs between the well-aligned and misaligned states. The remaining terms capture how the option value of trade changes through alignment. At rate $\frac{\lambda}{2}$ traders meet others with opposite asset holdings. When trade occurs, the gains are split equally, giving us $\frac{\lambda}{4}$. The trader meets both misaligned (fraction $m_{\lambda'}$) and well-aligned (fraction $1-m_{\lambda'}$) counterparties of type λ'. Whenever a trade makes this trader well-aligned (misaligned), there is a gain (loss) $s(\lambda)$. Similarly, whenever a trade makes the partner λ' well-aligned (misaligned), there is a gain (loss) $s(\lambda')$. Trade occurs only if the sum of the two gains is positive.

We also use the steady state equation (6) and the Nash bargaining solution (4) to derive in Online Appendix C.1 the symmetric flow balance equation governing the stationary distribution of misalignment:

$$\left(r+\gamma+\frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{s(\lambda)+s(\lambda')>0} m_{\lambda'} + \mathbb{I}_{s(\lambda)>s(\lambda')}(1-m_{\lambda'})\right) d\mu_F(\lambda')\right) m_\lambda = \left(\gamma+\frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{s(\lambda)<s(\lambda')} m_{\lambda'} + \mathbb{I}_{s(\lambda)+s(\lambda')<0}(1-m_{\lambda'})\right) d\mu_F(\lambda')\right)(1-m_\lambda). \quad (9)$$

The indicator function $\mathbb{I}$ is equal to 1 if the inequality in the subscript holds and is zero otherwise. Misaligned traders become well-aligned if they exit and are replaced, if they have a taste shock, or if they trade. The latter event can occur with both well-aligned and misaligned traders, so long as the joint surplus of the transaction is positive. The right hand side captures the inflow into misalignment from well-aligned traders who experience a taste shock or who trade.

Finally, building on equation (3), we prove in Online Appendix C.1 that symmetry and Nash bargaining imply that the choice of contact rate maximizes

$$r\pi_\lambda = \delta_1 - \gamma s(\lambda) + \frac{\lambda}{4}\int_{\mathcal{X}} \big((s(\lambda')-s(\lambda))^+ m_{\lambda'} + (-s(\lambda)-s(\lambda'))^+(1-m_{\lambda'})\big) d\mu_F(\lambda') - rC(\lambda), \quad (10)$$

where $\delta_1 \equiv \frac{1}{2}(\delta_{h,1} + \delta_{l,0})$ is the average flow payoff in the well-aligned state. The structure of this equation is similar to equation (8). Flow profits is equal to the average flow payoff δ_1,

minus the loss of surplus following a preference shock, plus terms involving the gains from trade, minus the flow cost of contacts $rC(\lambda)$.[9]

Equation (10) reflects each trader's cost-benefit analysis when she chooses her contact rate. The first term is the flow payoff from being well aligned. The second term is the cost of losing her well-aligned status following a taste shock. This cost turns out to be decreasing in the trader's contact rate, reflecting the fact that a higher contact rate enables a trader to more quickly realign her asset position with her tastes. The third term captures the profit that a trader earns from her contacts when she is in the well-aligned state. We show below that this comes from providing intermediation services. Finally, the last term captures the exogenous cost associated with choosing a contact rate, parallel to the cost of forming a link in Jackson and Wolinsky (1996) and the subsequent literature on network formation.

We are now able to define a (symmetric) equilibrium.

Definition 1 *An equilibrium is a counterparty measure μ_F, a misalignment rate function $m : \mathcal{X} \to [0, 1]$, and a surplus function $s : \mathcal{X} \to \mathbb{R}$, satisfying:*

1. *the surplus equation* (8)*;*
2. *the flow balance equation* (9)*; and*
3. *optimal investment: $\mu_F(\mathcal{Y}) = 1$, where $\mathcal{Y} = \arg\max_{\lambda \in \mathcal{X}} \pi_\lambda$ and π_λ is defined given μ_F, m, and s in equation* (10)*.*

We have already explained all three conditions.

Given an equilibrium tuple of reduced-form objects (μ^F, m, s), we can recover the primitive objects described in Section 3; see Appendix C.2 for more details. First, we use all three objects, most notably the surplus function, to recover the value function $v_{\lambda,i,b}$. From that we get the trading probability $p_{\lambda,i,b}^{\lambda',i',b'}$ and the transfer $t_{\lambda,i,b}^{\lambda',i',b'}$ via equations (4) and (5). Next, using symmetry, the misalignment rate encodes the fraction of type λ traders in different states, $\sigma_{\lambda,i,b}$. Putting this together, we obtain a value function satisfying equation (2) and steady state shares satisfying equation (6). Finally, the choice of contact rates is equivalent to what is in equation (3).

To understand the definition of equilibrium, it may help to comment briefly on the possibility of an *autarky equilibrium*, where the average contact rate is $\Lambda = 0$, or equivalently $\mu_G(\{0\}) = 1$. We emphasize that such an equilibrium does not necessarily exist in this environment. This is because our contact technology allows each trader to choose any contact

[9]The assumption that $C(\lambda)$ is paid upfront is isomorphic to one where traders pay $rC(\lambda)$ per unit of time, or one where traders pay $rC(\lambda)/\lambda$ per meeting.

rate $\lambda \in \mathcal{X}$, regardless of what others are doing.[10] In particular, a trader who chooses a contact rate λ will meet some other trader drawn from the counterparty measure μ_F at rate λ. In Section 3.2, we already discussed a physical matching process which justifies this assumption, as we show in more detail in Appendix A.2. There, we explain how a single trader can contact others at an arbitrary strictly positive rate, even in the case where everyone else chooses a zero contact rate. We also show that this matching process is consistent with autarky, $\mu_G(\{0\}) = 1$, and *simultaneously* with $\mu_F(\{0\})$ taking an arbitrary value between 0 and 1, depending on details.

The flexibility of the counterparty distribution in an autarky equilibrium reflects the fact that we cannot use equation (1), i.e. Bayes rule, to recover μ_F from μ_G when $\mu_G(\{0\}) = 1$. But we stress that our definition of equilibrium insists that μ_F puts all its weight on value-maximizing contact rates, including possibly a zero contact rate. If 0 is the unique profit-maximizing contact rate, this implies $\mu_F(\{0\}) = 1$, but we find that an autarky equilibrium sometimes requires a more complicated counterparty measure; see the proof of Proposition 3 for an example.

5 Characterization with General Cost Functions

This section develops two main results characterizing equilibrium. As a stepping stone, Lemma 1 establishes which trades occur for an arbitrary counterparty distribution. Proposition 1 then shows that any counterparty distribution is an equilibrium for some cost function and shows how to construct such a cost function; and Proposition 2 shows that rich dispersion in contact rates arises under general conditions. See Online Appendix C.3 for technical details, including all formal proofs in this section.

5.1 Equilibrium Trading Patterns

We start by characterizing equilibrium trading patterns given any counterparty measure μ_F.

Lemma 1 *In any equilibrium, the surplus function $s(\lambda)$ is positive-valued and strictly decreasing. When two traders with opposite asset positions meet they*

1. *always trade the asset if both are misaligned;*

2. *never trade the asset if both are well-aligned;*

[10] One could instead impose that when the contact rate distribution is degenerate at zero, $\mu_G(\{0\}) = 1$, it is impossible to meet other traders. Such an alternative assumption would imply that an autarky equilibrium always exists.

3. *trade the asset if one is misaligned and the other is well-aligned and the well-aligned trader has the higher contact rate.*

The Nash bargaining solution (4) and the definition of the surplus function $s(\lambda) = v_{\lambda,h,1} - v_{\lambda,h,0} - q = v_{\lambda,l,0} - v_{\lambda,l,1} + q$ jointly imply

$$p_{\lambda,h,0}^{\lambda',l,1} = \begin{cases} 1 & \text{if } s(\lambda) + s(\lambda') > 0, \\ 0 & \end{cases} \qquad p_{\lambda,h,0}^{\lambda',h,1} = \begin{cases} 1 & \text{if } s(\lambda) > s(\lambda'), \\ 0 & \end{cases}$$

$$p_{\lambda,l,0}^{\lambda',h,1} = \begin{cases} 1 & \text{if } 0 > s(\lambda) + s(\lambda'), \\ 0 & \end{cases} \qquad p_{\lambda,l,0}^{\lambda',l,1} = \begin{cases} 1 & \text{if } s(\lambda') > s(\lambda). \\ 0 & \end{cases}$$

The bulk of the proof establishes that the surplus function is positive-valued and strictly decreasing, from which the trading patterns follow immediately.

Üslü (2019) and Nekludyov (2019) both derive similar results in richer settings with an exogenous counterparty distribution. Thus, Lemma 1 is a special case of the key findings in these papers, a stepping stone to our novel result on equilibrium dispersion in contact rates.

The first two parts of Lemma 1 reflect fundamentals. Trade between two misaligned traders turns both into well-aligned traders, thus creating gains in a direct, static fashion. Trade between two well-aligned traders turns both misaligned and never happens for the same static reason.

The third part of the Lemma reflects option value considerations and is the key endogenous trading pattern that arises in this environment. It states that a faster trader buys the asset from a slower trader if both have taste l; and she sells the asset to the slower trader if both have taste h. We label trades *intermediation* when both traders have the same taste for the asset. Intermediation does not immediately increase the number of well-aligned traders, but it moves misalignment towards traders who expect more future trading opportunities. Intermediation yields gains in equilibrium because traders with higher contact rates are faster at offloading misaligned positions in future trades.

The possibility of intermediation implies that a trader's buying and selling decisions become increasingly detached from her idiosyncratic tastes as her contact rate increases. In other words, a high contact rate moderates the impact of the idiosyncratic taste component on a trader's valuation of the asset. It follows that those who become intermediaries, positioned at the center of the trading chain, are traders with a high contact rate. Figure 1 shows the intermediation chain which follows from Lemma 1. Slow traders are at the periphery of the trading chain, not trading once their asset position is aligned with their tastes. In turn, fast traders constitute the endogenous core of the trading network, buying and selling largely

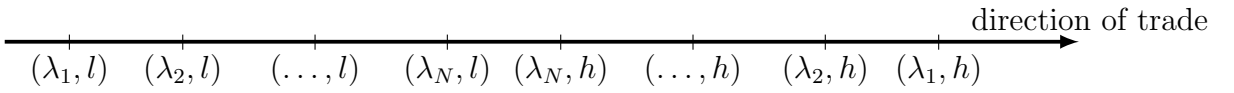

Figure 1: Direction of trade across traders with contact rate $\lambda \in \{\lambda_1, .., \lambda_N\}$ with $\lambda_1 < \lambda_2 < .. < \lambda_N$ and current taste $i \in \{l, h\}$.

irrespective of their tastes. In doing so, they take on misaligned asset positions from types with lower contact rates simply because they are better at locating other traders. That is, they intermediate.

5.2 Recovering the Cost Function

We next prove that our model can rationalize any observed counterparty measure μ_F:

Proposition 1 *For any counterparty measure μ_F, there exists a cost function C such that μ_F is an equilibrium. Moreover, C is unique on support of μ_F, up to an additive constant.*

The formal proof proceeds in three steps. First, we show that the counterparty distribution uniquely determines the misalignment rate. Then, we derive the functional form for the surplus equation, thereby proving that the counterparty distribution uniquely determines the surplus function. Finally, we show how to recover the cost function from these three objects.

Since every contact rate measure μ_G is associated with some counterparty measure μ_F, a corollary is that our model can rationalize any contact rate measure through the choice of an appropriate cost function C. As such, our framework offers a novel way of modeling and rationalizing real-world trading networks, complementing the existing literature on network formation. In particular, we can rationalize the coexistence of traders with very different rates of market access despite them being ex-ante identical. In the next subsection as well as Section 6, we invert the analysis in Proposition 1 to characterize the counterparty distribution under some natural assumptions on the shape of the cost function.

5.3 Heterogeneity in Contact Rates

We next show our second main result, that the coexistence of traders with different contact rates arises naturally in equilibrium even when market participants are ex-ante homogeneous.

Proposition 2 *Assume C is differentiable and C' is Lipschitz continuous. Then any equilibrium counterparty distribution F and contact rate distribution G are absolutely continuous*

on $[0, \bar{\lambda})$*. If additionally* C *is weakly convex,* $C'(0) < \frac{\Delta\gamma^2}{2r(r+2\gamma)^3}$*, and* $C'(\bar{\lambda}) \geq \frac{4\gamma\Delta}{r\bar{\lambda}^2}$*, then a positive measure of traders choose a contact rate in the interval* $(0, \bar{\lambda})$ *in any equilibrium.*

Recall that $F(\lambda) \equiv \mu_F([0, \lambda])$ is the counterparty distribution and $G(\lambda) \equiv \mu_G([0, \lambda])$ is the contact rate distribution.

Proposition 2 implies that although all traders are ex-ante identical, there is no symmetric equilibrium in which all traders choose identical actions, except possibly at the boundaries of the choice set 0 and $\bar{\lambda}$. In particular, under mild restrictions on the cost function, ex-ante identical traders choose to be continuously heterogeneous, meaning $\mu_F((0, \bar{\lambda})) > 0 = \mu_F(\{\lambda\})$ for all $\lambda \in (0, \bar{\lambda})$. That is, a positive measure of traders choose an interior contact rate, but a zero measure choose the same interior contact rate.

The critical force underlying Proposition 2 is intermediation. To develop an intuition, we argue that if everyone has a common contact rate λ, there is a convex kink in the profit function π_λ at λ, reflecting the gains from intermediation. To see this, consider the marginal return to a change in the contact rate at the mass point λ. A trader with contact rate λ only engages in fundamental trades, but this changes discretely at slightly different contact rates. A trader who chooses a contact rate $\lambda + \varepsilon$, $\varepsilon > 0$, intermediates for the entire marketplace, trading independently of her intrinsic valuation whenever the counterparty is misaligned and trading is feasible (Lemma 1). The profits from intermediation are proportional to the difference in surplus functions, $s(\lambda) - s(\lambda + \varepsilon)$. Since s is strictly decreasing, this is locally linearly increasing in the trader's contact rate when it exceeds λ.

Conversely, consider the intermediation returns of a trader who chooses a contact rate $\lambda + \varepsilon$, $\varepsilon < 0$. Now the trader benefits from others intermediating for her, with profits still proportional to the difference in surplus functions, $s(\lambda + \varepsilon) - s(\lambda)$. We still have s strictly decreasing, so now as ε increases to zero from below, the trading profits again shrink to zero from above. That is, intermediation profits are locally linearly decreasing in the trader's contact rate when it is smaller than λ.

The intermediation benefits of moving away from the mass points are hence positive in both directions. On top of that, there are fundamental trading benefits to a higher contact rate, which are locally linear. This leads to a convex kink, which in turn means that choosing the mass point cannot be optimal if the cost function is differentiable. This logic carries over to any interior mass point. As soon as a positive measure of traders has the same contact rate, there is a discrete jump up in the marginal return to contacts at this mass point, inconsistent with equilibrium under a differentiable cost function. A similar intuition rules out counterparty distributions that are not absolutely continuous, but for this we require a stronger condition, that marginal cost is Lipschitz.

The theoretical finding that intermediation in this frictional setting leads to a market

structure with rich heterogeneity connects closely with empirical evidence. Fricke and Lux (2015) estimate a network model that allows for a continuous notion of "coreness" for the Italian interbank market. Their Figure 9 strongly suggest a vast amount of heterogeneity, inconsistent with a simple binary classification of market participants. in 't Veld and van Lelyveld (2014) find very similar results for the Dutch interbank market (see their Figure A.13). Di Maggio, Kermani and Song (2017) similarly show that various measures of market participation in the US corporate bond market are continuously distributed (see their Figure 1).[11] These markets thus have a tiered network structure with rich heterogeneity.

Our theory further posits that almost all market participants at least occasionally intermediate. This likewise connects with data. Craig and Von Peter (2014) only classify 2.4 percent of the banks in ther German interbank market as the core, yet find that 92.7 percent of banks occasionally intermediate in that market. Bech and Atalay (2010) group banks in the US federal funds market into multiple tiers and show that there are many links within tiers, so intermediation is frequently done in a decentralized fashion (see their Figure 4). For similar evidence from the US market for asset-backed securities, see Hollifield, Neklyudov and Spatt (2017), Figure 5. These authors also show that peripheral dealers are frequently part of long intermediation chains that involve multiple dealers; see their Table 4.

Jointly, this suggests that frictional asset markets frequently feature traders with various degrees of market access, most of whom at least occasionally engage in intermediation activities. This closely aligns with the endogenous market structure which arises in our setting.

5.4 Ordinary Differential Equation System

Absolute continuity, as established in Proposition 2, implies that the cumulative distribution function is described by its derivative, $F(\lambda) - F(0) = \int_0^\lambda F'(\lambda')d\lambda'$ for all λ, and similarly for G. In Online Appendix C.3, we define $M(\lambda) \equiv \int_0^\lambda m_{\lambda'} dF(\lambda')$ to be the fraction of meetings that are with a misaligned trader with contact rate below some $\lambda \in \mathcal{X}$. We then characterize equilibrium using a first order ordinary differential equation system in F, M, and s with known boundary values. We show the differential equations in Appendix B. The simplicity of this differential equation system enables us to solve the model numerically for arbitrary cost functions that satisfy our smoothness requirement. We also use the system throughout Section 6 to obtain an analytical characterization with a linear cost function.

[11] Similarly, see Boss, Elsinger, Summer and Thurner (2004) for the Austrian interbank market, Cont, Moussa and Santos (2010) for the Brazilian interbank network, Martinez-Jaramillo, Alexandrova-Kabadjova, Bravo-Benitez and Solórzano-Margain (2014) for the Mexican interbank network.

5.5 Robustness

Proposition 2 is the most general result in this paper, showing that dispersion in contact rates arises under weak conditions. To underscore its generality, we next offer some conjectures on the robustness of the result to our modeling assumptions.

Symmetry: We have assumed that the asset endowment is equal to $1/2$, exactly equal to the measure of traders in the high state at any point in time. This makes the restriction to symmetric equilibrium natural. Without symmetry, we would have to deal with four value functions and with the share of traders in each of four taste-asset holding states, making notation more cumbersome. Moreover, it is no longer ex ante obvious which trades take place. For example, if the asset is scarce, the market may shut out the slowest traders, never selling the asset to them, even if they are in the high state. Still, there will always be a role for intermediation among the traders fast enough to hold the asset. Since intermediation is what drives heterogeneity, as we discussed above, we expect a version of Proposition 2 to be robust to such an extension.

Restricted asset holdings: We have restricted asset holdings to be either zero or one, which we view as the limit of an extremely convex inventory cost. Allowing for unrestricted asset holdings, as for instance in Üslü (2019), would preserve the connection between intermediation and heterogeneous contact rates. In fact, we believe it would amplify the force creating a kink in the profit function at a mass point, since a slightly faster trader would be unrestricted in her ability to intermediate for the mass of traders.

Contact technology: We have assumed that a trader's contact rate does not depend on the choices others make, but who she meets depends on these choices. We could have made other choices. For example, in footnote 7, we discuss the telephone matching function, where a trader chooses how often to call others but also receives calls from others at a rate that is independent of her choice of contact rate. Thus even a trader who chooses not to contact anyone will be able to buy and sell assets. We believe that this does not affect the forces pushing towards dispersion in contact rates. In particular, if everyone else chooses a common contact rate λ, one trader's choice of contact rate does not affect whom she meets in either model. It just leads to gains from acting as an intermediary if the trader chooses a faster contact rate, or to gains from being intermediated for if the trader chooses a slower contact rate.

Time-invariant λ: If they could do so costlessly, traders would want to adjust their contact rate in response to their time-varying taste and asset holding. They do not do so because we have assumed that the choice of contact rate is irreversible. If we allow identical traders to do so at no cost, then a trader's current contact rate no longer affects future trading opportunities and so the motive for intermediation disappears. Thus some irreversibility in contact rates is important for our results. But as long as changing the contact rate either takes time or incurs an irreversible cost, the current contact rate is relevant for future trading opportunities and so some trades will involve intermediation. And because there is intermediation, there is an incentive to choose a different contact rate than others. That is, each trader's contact rate may move around over time, but the associated stationary distribution of contact rates will still not feature any mass points.

6 Characterization with a Linear Cost Function

This section characterizes equilibrium under the assumption that the cost function C is linear, $C(\lambda) = c\lambda$. After analyzing the baseline model, we extend our analysis to a limiting economy with no upper bound on contacts, $\bar{\lambda} \to \infty$. For this case, we also consider what happens in the frictionless limit, when the marginal cost of contacts c converges to zero. We relegate technical details, including all proofs, to Online Appendix C.4.

6.1 Equilibrium Characterization

We start by proving existence of equilibrium and characterizing its properties when the cost function is linear:

Proposition 3 *Assume $C(\lambda) = c\lambda$. Fix r, γ, Δ, and $\bar{\lambda}$. There exist thresholds $\bar{c} > \underline{c} > 0$ such that*

$$\text{if } \begin{cases} c \geq \bar{c} \\ c \in (\underline{c}, \bar{c}) \\ c \leq \underline{c}, \end{cases} \quad \text{then there is a } \begin{cases} \textit{autarky equilibrium} \\ \textit{intermediated trade equilibrium} \\ \textit{degenerate trade equilibrium,} \end{cases}$$

and any equilibrium takes one of these three forms. In an autarky equilibrium, the average contact rate is $\Lambda = 0$. In an intermediated trade equilibrium, the average contact rate is $\Lambda \in (0, \bar{\lambda})$; the support of the counterparty distribution is a convex interval $[\underline{\lambda}, \bar{\lambda}]$ with $\underline{\lambda} \in (0, \bar{\lambda})$ and $dF(\bar{\lambda}) > 0$; and the misalignment rate m_λ is increasing on $[\underline{\lambda}, \bar{\lambda}]$. In a degenerate trade equilibrium, the average contact rate is $\Lambda = \bar{\lambda}$.

Recall that $\Lambda \equiv \int_{\mathcal{X}} \lambda d\mu_G(\lambda)$. The proof gives explicit expressions for $\bar{c}$ and $\underline{c}$ and characterizes the contact rate and counterparty distributions for any value of c.

We do not claim uniqueness of the equilibrium and indeed can construct examples in which an autarky equilibrium and an intermediated trade equilibrium coexist for the same parameter values. However, any equilibrium must lie in one of the three classes described in the proposition.

Proposition 3 states that for $c \geq \bar{c}$, there exists an equilibrium where all trading activity collapses, while for $c \leq \underline{c}$, there exists an equilibrium without intermediation since all traders choose the highest contact rate $\bar{\lambda}$. More interestingly, for a nonempty interval of costs $(\underline{c}, \bar{c})$, there exists an equilibrium where a non-degenerate contact rate distribution G and intermediation emerge endogenously. Such an equilibrium has four key properties: First, no trader has a contact rate below a strictly positive lower bound $\underline{\lambda}$. Second, a strictly positive fraction of traders choose $\bar{\lambda}$. Third, the remaining counterparties have a continuously distributed contact rate on $[\underline{\lambda}, \bar{\lambda})$. And finally, traders who choose a faster contact rate are misaligned more often.

The strictly positive lower bound $\underline{\lambda}$ in the intermediated trade equilibrium reflects the fact that the profits of a trader are a continuous function, converging to the autarky value as λ converges to 0. With $c < \bar{c}$, traders in the non-degenerate equilibrium do strictly better than autarky and so it must be the case that no one chooses a contact rate too close to zero.

We postpone the discussion of the second feature, mass at $\bar{\lambda}$, to the next subsection. To understand the third finding, suppose there was a "hole" in the support, with no trader choosing contacts inside a strictly positive interval on $[0, \bar{\lambda}]$. In this case, the proof shows that the profits must be unequal at the two endpoints. Why? Because trading profits over that range would be linear in λ since trading opportunities would not be changing, while improvements in a trader's asset position show diminishing returns to scale. This implies the profit function must be concave on the interval, which is inconsistent with both extreme points yielding a higher value than any intermediate point.

The finding that faster contact rates are associated with higher misalignment rates might be counterintuitive. A higher contact rate has two opposing effects on a trader's misalignment rate. On the one hand, a trader is more frequently able to offset a misaligned position. On the other hand, a trader with a higher contact rate intermediates more frequently, taking on misalignment from slower traders. The proposition states that the latter force dominates everywhere on the support of F. That is, traders do not invest in a faster contact rate to reduce their misalignment, but rather to trade more frequently.

Intuitively, doubling a trader's contact rate from λ to 2λ more than doubles his opportunities for intermediation, since he can also intermediate for traders with contact rate

$\lambda' \in [\lambda, 2\lambda)$. An increase in the misalignment rate is then needed to offset this increase in intermediation profits, leaving the trader's profits unchanged. We stress that this result holds in equilibrium, not for an arbitrary distribution of contact rates. Still, we find that this result is more general than the linear cost case. For example, in Farboodi, Jarosch and Shimer (2017) we show that misalignment rate is increasing in the contact rate when the marginal cost of contacts is increasing and convex.

6.2 Limiting Equilibrium: $\bar{\lambda} \to \infty$

Proposition 3 implies that whenever there is trade, a positive fraction of contacts are with traders who choose the maximum permissible contact rate, $dF(\bar{\lambda}) > 0$, and so the choice of $\bar{\lambda}$ affects equilibrium. We next examine what happens when $\bar{\lambda}$ is large. To do this, we define a *limiting equilibrium* as the limit of equilibria of a sequence of economies n which are identical except for their upper bounds $\bar{\lambda}_n$, with $\bar{\lambda}_n \to \infty$:

Definition 2 *Assume $C(\lambda) = c\lambda$. Fix r, γ, Δ, and c. For any $\bar{\lambda}$, let $(\mu_{F,\bar{\lambda}}, m_{\bar{\lambda}}, s_{\bar{\lambda}})$ be an equilibrium when the maximum contact rate is $\bar{\lambda}$ and as usual let $F_{\bar{\lambda}}(\lambda) = \mu_{F,\bar{\lambda}}([0,\lambda])$ for all $\lambda \leq \bar{\lambda}$. Also extend the definition of $(F_{\bar{\lambda}}, m_{\bar{\lambda}}, s_{\bar{\lambda}})$ to the positive reals in an arbitrary way. (F, m, s) with domain $[0,\infty)^3$ is a* limiting equilibrium *if there exists an increasing unbounded sequence $\{\bar{\lambda}_n\}$ with associated $(F_{\bar{\lambda}_n}, m_{\bar{\lambda}_n}, s_{\bar{\lambda}_n})$ which converges pointwise to (F, m, s).*

Intuitively, a limiting equilibrium is the limit of a sequence of equilibria as we increase $\bar{\lambda}$. The only subtle point is that we need to extend the range of the functions (F_n, m_n, s_n) above the upper bound $\bar{\lambda}_n$. A natural, but not necessary, way to do this is to impose that $F_{\bar{\lambda}}(\lambda) = 1$ (reflecting that $F_{\bar{\lambda}}$ is a cumulative distribution function), $m_{\bar{\lambda}}(\lambda) = \frac{2\gamma + \lambda M(\bar{\lambda})}{2(r+2\gamma+\lambda M(\bar{\lambda}))}$ (reflecting equation 9), and $s_{\bar{\lambda}}(\lambda) = \frac{2\Delta}{2(r+2\gamma)+\lambda M(\bar{\lambda})}$ (reflecting equation 8) when $\lambda > \bar{\lambda}$.

To get some intuition for how the limit of equilibria behaves, it is useful for us to briefly discuss the mathematical structure we use to characterize equilibrium. To begin, recall that if the marginal cost function is Lipschitz continuous, we can represent any equilibrium as the solution to a system of three ordinary differential equations in F, M, and s. In the linear cost case, we prove that we can reduce this to a pair of ordinary differential equations, $(F', M') = X_1(\lambda, F, M)$ on $[\underline{\lambda}, \bar{\lambda})$, with boundary condition $F(\underline{\lambda}) = M(\underline{\lambda}) = 0$. The function X_1 depends on the parameters r and γ but not on c, Δ, or $\bar{\lambda}$. We show that there is a discontinuous increase in F and M at $\bar{\lambda}$, the $dF(\bar{\lambda}) > 0$ result in Proposition 3. In a limiting equilibrium, we simply solve the same differential equations on $[\underline{\lambda}, \infty)$.

If we knew the lower bound on contact rates $\underline{\lambda}$, we would be done, but $\underline{\lambda}$ is endogenous. To find it, we use a second equation, expressed succinctly as $c = X_2(\underline{\lambda})$, telling us the cost c which makes $\underline{\lambda}$ the lower bound on contact rates. The function X_2 depends on r, γ, and

Δ both directly and indirectly through the solution to the ordinary differential equation system $(F', M') = X_1(\lambda, F, M)$, and on $\bar{\lambda}$ indirectly through the discontinuity in F and M at $\bar{\lambda}$. Moreover, if c is not too big, we can always find a value of $\underline{\lambda} > 0$ that makes this equation hold. As we approach a limiting equilibrium, the indirect effect of $\bar{\lambda}$ on $\underline{\lambda}$ vanishes, which ensures that $\underline{\lambda}$ has a well-behaved limit when we take the upper bound on contacts to infinity. This value of $\underline{\lambda}$, combined with the solution to the differential equation system on the unbounded interval $[\underline{\lambda}, \infty)$, yields our characterization of limiting equilibrium.

Using this approach, we obtain the following characterization of limiting equilibrium:

Proposition 4 *Assume* $C(\lambda) = c\lambda$ *with* $c < \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)}$. *Then in a limiting equilibrium, there are* middlemen*, meaning* $\lim_{\lambda\to\infty} F(\lambda) < 1$*; and the contact rate distribution has a* Pareto tail with tail index 2*, meaning* $\lim_{\lambda\to\infty} \lambda^2(1 - G(\lambda))$ *is positive and finite.*

Proposition 3 showed that a strictly positive fraction of meetings are with traders at any finite upper bound $\bar{\lambda}$. The existence of middlemen is a stronger result, because it implies that this fraction does not vanish as $\bar{\lambda}$ goes to infinity. In order to show the existence of middlemen, we show that the counterparty distribution remains bounded away from one as $\lambda \to \infty$. In the limiting economy, almost every trader has a finite contact rate ($\lim_{\lambda\to\infty} G(\lambda) = 1$), yet a positive fraction of their counterparties has a higher contact rate. We refer to these counterparties as *middlemen*.

Why do middlemen emerge? Suppose there were none, so $\lim_{\lambda\to\infty} F(\lambda) = 1$. Then a trader with a high contact rate has almost all her meetings with slower traders. Such a trader will therefore trade irrespective of her intrinsic valuation and so will have a misalignment rate close to $\frac{1}{2}$. In particular, her misalignment rate would be higher than that of a trader living in autarky, for whom equation (9) implies $m_0 = \frac{\gamma}{r+2\gamma}$. To justify the choice of a large value of λ, it must then be the case that the fast trader earns strictly positive profits from trading. But trading profits scale linearly in the tail of the distribution, since trading opportunities are effectively the same and a trader can choose any multiple of λ. This is inconsistent with a large finite value of λ being optimal, contradicting the hypothesis that $\lim_{\lambda\to\infty} F(\lambda) = 1$. In short, what middlemen do is ensure that even very fast traders have a misalignment rate strictly below $\frac{1}{2}$, obviating the need for them to earn profits from intermediation.[12]

We next note that the contact rate distribution has support $[\underline{\lambda}, \infty)$, a natural extension of the support $[\underline{\lambda}, \bar{\lambda}]$ in Proposition 3. The reason for the unbounded support with a linear cost function is that a trader's misalignment rate converges to a constant, while her trading profits scale linearly with her contact rate, since her trading opportunities no longer change.

[12]Correspondingly, with a finite $\bar{\lambda}$, mass at the upper bound guarantees that traders remain indifferent across λ as $\lambda \to \bar{\lambda}$.

In fact, linear scaling of the benefits of contacts tells us that the cost function must be asymptotically linear for an open tail to emerge. While we have strong intuition for the open tail, we do not have a clear understanding of why the tail turns out to be a Pareto tail with a tail index of 2. Nonetheless, this finding connects closely with empirical evidence as we document below.

Before discussing the evidence, however, we note that what is mapped out empirically is the distribution of trading rates $\alpha \equiv \lambda p_\lambda$, the product of the contact rate λ and the probability of trading in a meeting, p_λ. Assuming trades occur only if there are strict gains, p_λ is uniquely defined. In particular, trade only occurs when a misaligned trader meets a trader with a strictly higher contact rate or a misaligned trader with the same contact rate; or when a trader, well-aligned or misaligned, meets a misaligned trader with a lower contact rate. Let $\hat{G}(\alpha)$ denote the population distribution of trading rates in a limiting equilibrium. Then a corollary to Proposition 4 connects the results describing the distribution of contact rates to the distribution of trading rates:

Corollary 1 *Assume $C(\lambda) = c\lambda$ with $c < \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)}$. In a limiting equilibrium, the fraction of trades with middlemen is strictly positive; and the trading rate distribution has a Pareto tail with tail index 2, meaning $\lim_{\alpha\to\infty} \alpha^2(1-\hat{G}(\alpha))$ is positive and finite.*

Since a positive fraction of meetings are with middlemen and there is a positive probability of trade in one of these meetings, the first part of the result is immediate. The trading rate inherits the tail properties of the contact rate distribution, because the trading probability conditional on a meeting converges to a positive constant at high contact rates.

We turn now to the empirical content of these results. Our finding that there are middlemen is an extension of the result in Proposition 3 that there is a mass of traders at the upper bound $\bar{\lambda}$, showing that this is not an artefact of a finite upper bound. We view both results as indicating the existence of a "core" of the market, traders who are highly connected both to each other and to the rest of the market, and who intermediate for all other traders. The empirical literature frequently identifies a core of highly connected entities. Craig and Von Peter (2014) define the core as the top tier of banks which constitute a complete graph among themselves. Applying this to the German interbank market, they classify 2.7 percent of banks as the core. in 't Veld and van Lelyveld (2014), applying the same methodology to the Dutch interbank market, group 13 percent of banks in the core. Hollifield, Neklyudov and Spatt (2017) identify a core of 6 to 10 percent of dealers in the inter-dealer derivatives market, accounting for 60 to 70 percent of trades (see their Table 4). Di Maggio, Kermani and Song (2017), for the corporate bond market, think of the core as the top 50 dealers, who account for 80 percent of transactions.

We turn next to the Pareto tail in $\hat{G}$. The empirical literature, motivated by network models, typically measures a trader's *degree*, i.e. the number of counterparties during some interval of time. In our model, a trader's expected degree in a unit time interval is simply equal to the number of trading partners α per unit of time, and hence the degree distribution inherits the Pareto tail of the trading rate distribution.[13] Many papers document that the degree distribution has a Pareto tail in different over-the-counter markets. Examples include Li and Schürhoff (2019) for the municipal bonds market, Hollifield, Neklyudov and Spatt (2017) for derivatives, Peltonen, Scheicher and Vuillemey (2014) for the credit default swap market, Bech and Atalay (2010) for the out-degree of banks in the federal funds market, and Boss, Elsinger, Summer and Thurner (2004), De Masi, Iori, Precup, Gabbi and Caldarelli (2008) and De Masi, Iori and Caldarelli (2006) for different European interbank markets. We also note that, in the numerical illustration in Section 7.2, we show that the entire trading rate distribution, not just the tail, is well-approximated by a Pareto distribution.

Overall, we thus argue that the key features of the endogenous market structure which arises in our setting connects tightly with a set of stylized facts on over-the-counter markets. It features traders with vastly different amounts of activity, many of whom at least occasionally intermediate for others. It also features a core of a few, highly connected traders who account for a substantial amount of overall activity. And it features a Pareto tail in the degree distribution.

6.3 Frictionless Limit: $c \to 0$

In many real world markets, trading frictions are small and so one might question the value of modeling frictions in such markets. Furthermore, advances in information technologies are likely to reduce frictions over time so one might wonder whether this leads to a diminished role of intermediation and heterogeneity as emphasized here.

This section uses our model to show that intermediation retains its prominent role in the frictionless limit, which we capture through an assumption that the marginal cost of contacts becomes negligible, $c \to 0$. In this case, everyone chooses a fast contact rate and so the aggregate misalignment rate converges to zero. Still, we demonstrate a clear sense in which heterogeneity and intermediation are preserved in the limit.[14] In particular, we obtain

[13] For a trader with trading rate α, the realized degree during a unit time interval is a Poisson random variable with mean α. Since the standard deviation of a Poisson distribution is equal to the square root of the mean, uncertainty about the realized degree becomes irrelevant when α is large. As a result, a Pareto mixture of Poisson distributions inherits the tail properties of the Pareto distribution.

[14] We first take a limiting equilibrium, where $\bar{\lambda}$ grows without bound, and then take the limit as c converges to zero. The order of limits is important. With the opposite order of limits, there is no intermediation when c is small. We find this order of limits to be more interesting since in our view the upper bound $\bar{\lambda}$ is present only for technical reasons.

a sharp characterization of trading volume, measured as the amount of asset purchases per unit of time.[15]

The following proposition characterizes the overall trading volume along with its decomposition in the frictionless limit.

Proposition 5 *Assume $C(\lambda) = c\lambda$. Consider a sequence of limiting equilibria as c converges to zero. The aggregate trading volume $\mathcal{V}$ converges to approximately 2.46γ and can be decomposed as follows: middlemen's purchases from other middlemen account for a volume $\mathcal{V}_{mm} = \frac{1}{2}\gamma$; middlemen's purchases from non-middlemen account for a volume of $\mathcal{V}_{mn} = \frac{1}{2}\gamma$; non-middlemen's purchases from middlemen account for a volume $\mathcal{V}_{nm} = \frac{1}{2}\gamma$; and non-middlemen's purchases from non-middlemen account for a volume $\mathcal{V}_{nn} \approx 0.96\gamma$.*

To prove this proposition, we first compute trading probabilities in an economy with finite $\bar{\lambda}$, then construct a limiting equilibrium, and then take the limit of trade volume as c converges to zero. See the proof of Proposition 5 for details. The proof also provides the exact expression for volume and the fraction of meetings with middlemen.

We contrast Proposition 5 with a naïve view of a market without frictions: all traders can trade instantaneously upon receiving a taste shock and only trade with other traders who receive the opposite taste shock at the same instant. That means that volume equals the share of traders with taste l times the rate at which they are hit by taste shocks, $\frac{1}{2}\gamma$. Note that this view leaves no role for intermediation or middlemen. In contrast, we obtain nearly five times as much trading volume in the frictionless limit. Furthermore, the proposition highlights that a meaningful role for heterogeneity in contact rates and intermediation is preserved in the limiting economy.

To understand this result, note that we are looking at a frictionless limit so almost no one is misaligned. Whenever a trader (who is almost surely not a middleman) suffers a taste shock, she is very likely to become misaligned and very unlikely to contact another misaligned trader. As a consequence, "fundamental" trades between two misaligned traders become exceedingly rare. Instead, the market passes the asset towards faster traders whenever possible. Since the faster trader is still very unlikely to be misaligned, this trade does not reduce misalignment, but simply moves it towards the core. The volume decomposition shows that, in the frictionless limit, the reallocation of the asset in response to taste shocks runs through an intermediation chain that always involves middlemen. That is, middlemen purchase the asset from (sell the asset to) non-middlemen at exactly the same rate at which asset owners (non-owners) get moved into misalignment by a taste shock.

[15] We maintain that agents with identical λ and different misalignment status do not trade since the transaction has zero value. With minimum curvature in the utility function they might well do so and volume would be higher then. In this sense these results can be viewed as a lower bound on volume.

When a middleman purchases the asset from a slower trader, they too move into misalignment. Afterwards, they find another misaligned middleman with the opposite asset position, both becoming well-aligned. As a consequence trade between middlemen accounts for a volume of $\frac{1}{2}\gamma$. The reason that reallocation always involves middlemen when c is small is that the average misalignment rate of traders with a finite contact rate is proportional to the square of the misalignment rate of middlemen. Thus, as misalignment converges to zero, a misaligned counterparty is almost surely a misaligned middleman, although even misaligned middlemen are scarce.

Taken together, whenever a trader experiences a taste shock, the market rapidly reallocates her asset position. But instead of doing so directly, the position gets traded through an intermediation chain. This chain runs through increasingly faster types towards middlemen, who then first reallocate the position internally before passing it back to slower misaligned traders who actually desire the asset position.

Taken together, intermediaries retain their prominent role in an almost frictionless setting. This finding connects naturally with the ever-increasing prominence of financial intermediation services despite the massive advances in information technologies in recent decades (Kuprianov, 1993; Philippon, 2015; Biais and Green, 2019).

7 Optimal Allocation

This section examines which trading patterns and contact rate distributions are Pareto optimal. We imagine a hypothetical social planner who can instruct traders both on their choice of λ at birth and on whether to trade in each future meeting, but who cannot directly alleviate the search frictions in the economy.

The section first sets up the planner problem and establishes several formal result that mimic the equilibrium characterization. We then contrast equilibrium and optimum numerically and then close with two exercises that document that equilibrium displays excessive trade.

7.1 Planner Problem

The hypothetical social planner chooses a contact rate measure μ_G as well as symmetric trading patterns in order to maximize steady state utility net of the future cost of meetings:[16]

$$\delta_1 - \Delta \int_{\mathcal{X}} m_\lambda d\mu_G(\lambda) - r \int_{\mathcal{X}} C(\lambda) d\mu_G(\lambda). \tag{11}$$

The first two terms gives the flow payoffs from alignment. The average well-aligned trader has a flow payoff of $\delta_1 \equiv \frac{1}{2}(\delta_{h,1} + \delta_{l,0})$, the average flow payoff in the well-aligned state. The average misaligned trader has a flow payoff of $\delta_1 - \Delta$. In addition, the planner must pay the search costs when a trader exits and is replaced by a newborn one. These are integrated using the contact rate measure μ_G. The planner also recognizes the misalignment m_λ is endogenous and depends both on the contact rate measure and on the choice of who trades with whom, both of which are under the planner's control.

In Online Appendix D.1, we use calculus of variations to derive necessary conditions characterizing optimality. First, we show that there is a social surplus function $S(\lambda)$, which tells us the gain the planner enjoys by moving a trader from the misaligned state to the well-aligned state. We prove that this satisfies

$$\begin{aligned}(r+2\gamma)S(\lambda) = \Delta + \frac{\lambda}{2}\int_{\mathcal{X}} \Big(&\big((S(\lambda') - S(\lambda))^+ - (S(\lambda) + S(\lambda'))^+\big) m_{\lambda'} \\ &+ \big((-S(\lambda) - S(\lambda'))^+ - (S(\lambda) - S(\lambda'))^+\big)(1 - m_{\lambda'})\Big) d\mu_F(\lambda'). \end{aligned} \tag{12}$$

This is the direct counterpart to equation (8). The only difference is that the planner internalizes the full gains from trade in each transaction, whereas an individual trader only internalizes her own half.

Second, we prove that trade occurs if and only if it results in an increase in total social surplus. This gives us the flow balance equation governing the stationary distribution of misalignment:

$$\begin{aligned}\left(r + \gamma + \frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{S(\lambda)+S(\lambda')>0} m_{\lambda'} + \mathbb{I}_{S(\lambda)>S(\lambda')}(1 - m_{\lambda'})\right) d\mu_F(\lambda')\right) m_\lambda \\ = \left(\gamma + \frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{S(\lambda)<S(\lambda')} m_{\lambda'} + \mathbb{I}_{S(\lambda)+S(\lambda')<0}(1 - m_{\lambda'})\right) d\mu_F(\lambda')\right)(1 - m_\lambda). \end{aligned} \tag{13}$$

[16] With transferable utility, any Pareto optimal allocation also solves the problem of a utilitarian planner who weights all traders' welfare equally. Since traders do not discount the future (except through exit, in which case they get replaced by another trader), this is equivalent to maximizing undiscounted, i.e. steady state, utility. Thus the problem we solve here characterizes any symmetric Pareto optimal allocation.

This is exactly analogous to equation (9), with social surplus playing the role of private surplus.

Third, we define the social value function

$$r\Pi_\lambda = -\gamma S(\lambda) + \frac{\lambda}{2}\int_{\mathcal{X}} \big((S(\lambda') - S(\lambda))^+ m_{\lambda'} + (-S(\lambda) - S(\lambda'))^+ (1 - m_{\lambda'})\big) d\mu_F(\lambda') - \bar{\theta}\lambda - rC(\lambda), \quad (14)$$

where

$$\bar{\theta} = \int_{\mathcal{X}} \frac{\gamma - (r + 2\gamma)m_\lambda}{\lambda} S(\lambda) d\mu_F(\lambda). \quad (15)$$

We prove that $\mu_F(\mathcal{Y}^P) = 1$ where $\mathcal{Y}^P = \arg\max_{\lambda \in \mathcal{X}} \Pi_\lambda$ is the set of contact rates that maximize the social value function. Thus the planner only utilizes contact rates that maximize this measure of social value.

There are two differences between equations (10) and (14). First, the planner values the whole surplus from a trade, while in equilibrium each party only gets half the surplus. This is similar to the difference between equations (8) and (12). And second, the planner internalizes that there is a cost $\bar{\theta}$ from each meeting, approximately equal to the annuitized marginal cost of meetings for an average counterparty, $\bar{\theta} \approx r \int_{\mathcal{X}} C'(\lambda) d\mu_F(\lambda)$.[17] This reflects the fact that an increase in one trader's contact rate diverts meetings from other traders, something individuals do not internalize in equilibrium. We return to these differences when discussing efficiency in Section 7.3.

Putting this together, the set of necessary conditions for the planning problem has an almost-identical mathematical structure to the definition of equilibrium. In the remainder of Online Appendix D, we leverage this to show through a series of propositions, structured to mimic the equilibrium propositions 1–5, that the equilibrium and optimum allocations are qualitatively similar. We summarize those results here.

We first show in **Lemma 1-P** that the equilibrium trading pattern is optimal. Trade occurs whenever two misaligned traders with the opposite asset holdings meet. It also occurs whenever a slower misaligned trader meets a faster well-aligned trader with the opposite asset holding. The intuition is straightforward. The planner's objective function boils down to minimizing the average rate of misalignment for a given distribution of contact rates. The planner therefore demands trade if it reduces static misalignment and rejects it if it raises static misalignment. In the case where only one trader is misaligned, the planner moves the misalignment towards the trader with more future trading opportunities, since this does not affect the current misalignment rate, but improves future trading possibilities. That is, the

[17] See Online Appendix D.2 for details on this approximation.

planner uses faster traders as intermediaries.

Next, in **Proposition 1-P** we show that any counterparty distribution satisfies the necessary conditions for optimality for some cost function C. This means that we cannot make any inference about the efficiency of the observed contact rate distribution unless we have independent knowledge of the cost function.

In **Proposition 2-P**, we prove that when marginal cost is continuous, the optimal contact rate distribution is continuous on $[0, \bar{\lambda})$. The atomless feature of the optimal contact rate distribution allows the planner to leverage the gains from meetings through intermediation. Any meeting between two traders with identical contact rates λ is beneficial solely when both are misaligned. In contrast, when two traders with different contact rates meet each other, the meeting is socially beneficial even if only the slower trader is misaligned, since there are gains from intermediation. An atomless distribution maximizes the fraction of meetings in which there are gains from trade. Similarly, if marginal cost is Lipschitz, the planner does not want to place too much mass in a neighborhood of any interior contact rate, ensuring that the counterparty distribution is absolutely continuous on $[0, \bar{\lambda})$.

Proposition 3-P shows that with a linear cost function, it is optimal to have one of three configurations, depending on the level of marginal cost. Autarky is optimal when marginal cost is high. A degenerate trade allocation, with everyone at $\bar{\lambda}$, is optimal when marginal cost is low. And an intermediated trade allocation is optimal for intermediate values. In this allocation, the support of the contact rate distribution is an interval $[\underline{\lambda}, \bar{\lambda}]$ and the misalignment rate is increasing on this support.

We then extend this to the limiting case with $\bar{\lambda} \to \infty$. In **Proposition 4-P**, we prove that with a linear cost function, the optimal distribution has a Pareto tail with parameter 2 and features middlemen. The planner introduces middlemen for reasons that mimic the equilibrium case. If there were no middlemen, then the misalignment rate of the fastest traders would be $\frac{1}{2}$, higher than the misalignment rate in autarky. To compensate, it would have to be the case that fast traders' meetings generate social value in excess of their cost rc. But that would imply that the planner would want to increase their contact rate, thereby creating middlemen.

Finally, we consider the frictionless limit, where the marginal cost of contacts converges to zero. In **Proposition 5-P**, we calculate trading volume. While the volume of trades involving middlemen is unchanged from equilibrium, we prove that there are optimally fewer trades between pairs of non-middlemen than occur in equilibrium. This reflects the fact that the planner relies to a greater extent on middlemen, making them account for a larger fraction of meetings than would occur in equilibrium.

7.2 Equilibrium vs. Optimum: Numerical Illustration

We next contrast limiting equilibria with the optimal allocation in the linear cost case. The resulting model only has four parameters, the exit rate r, the arrival rate of preference shocks γ, the cost of contacts c, and the average benefit from alignment Δ. It is straightforward to prove that both equilibrium and optimal allocations are homogeneous in c and Δ. Doubling both doubles the surplus s and S but affects neither the counterparty distribution nor the misalignment rate. Effectively these determine the size of a unit of payoff. Similarly, the equilibrium and optimal allocations are homogeneous in r, γ, Δ, and $1/c$. Doubling the first three parameters and cutting c in half leads to new equilibrium and optimal allocations in which everyone chooses twice as high a contact rate without changing their surplus or misalignment. Effectively this scaling determines the length of a unit of time. Putting these two obsevations together, we conclude that only two of the four parameters can qualitatively affect equilibrium or optimal allocations.

With this in mind, we impose $r = 0.05$ and $\gamma = 2.75$, consistent with Duffie, Gârleanu and Pedersen (2007). To start, we fix $c = 0.001\Delta$, but later consider the robustness of our results to other values of the cost. We take the limit of equilibria as $\bar{\lambda} \to \infty$.[18]

The red lines in Figure 2 summarize the optimal allocation, while the blue lines show the equilibrium. The top left panel shows that the equilibrium contact rate distribution first order stochastically dominates the optimal one. That is, the equilibrium displays excessive investment in contacts across the board. We revisit this observation in the next subsection where we complement it with two theoretical exercises that relate overinvestment to the model's externalities. We also note that equilibrium and optimal contact rate distributions are both well-approximated by a Pareto distribution with tail parameter 2, a line with slope -2 in the figure.

The bottom left panel shows that these features carry over to the distribution of trading rates. In particular, although equilibrium trading rates are too high, both the equilibrium and optimal trading rate distributions have a Pareto tail with parameter 2. This implies that the empirically-documented scale-free nature of many financial networks is also a feature of a market that optimally leverages the gains from intermediation when the cost per meeting is

[18] As mentioned previously, we do not have a uniqueness proof. Nevertheless, our numerical calculations are consistent with the depicted allocations, both in equilibrium and optimum, being unique. In particular, recall that with a linear cost function, the equilibrium and optimal allocations are both described by a lower bound $\underline{\lambda}$ and a pair of well-behaved ordinary differential equations in F and M. To verify that this is an equilibrium or optimal allocation, we then compute the implied marginal cost of contacts c. That is, we have a mapping from $\underline{\lambda}$ to c. Numerically, this relationship appears to be monotonically decreasing, so higher cost is associated with a smaller lower bound on contacts; see the top left panel of Figure 3. This would imply that the equilibrium and optimal allocations are unique for arbitrary c. We therefore refer to *the* equilibrium/optimum in this section.

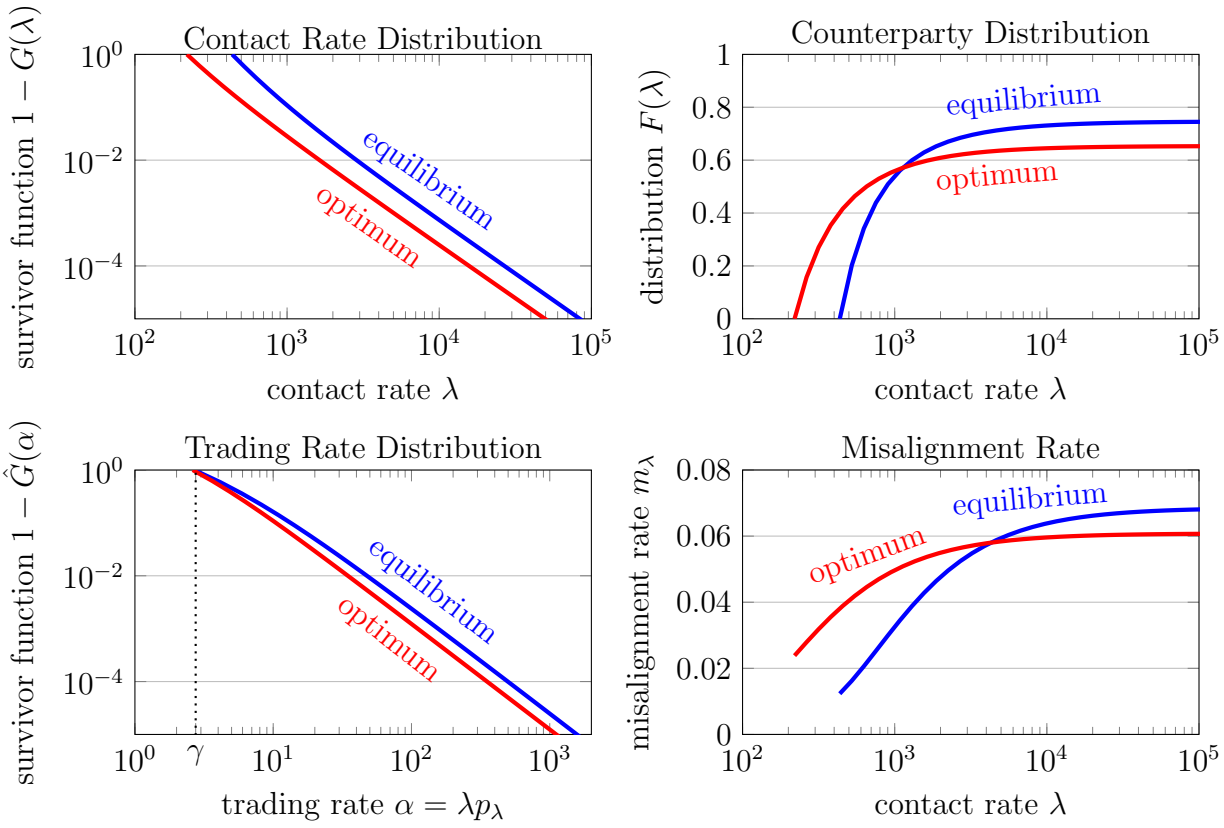

Figure 2: Equilibrium and optimal counterparty distribution, contact rate distribution, trading rate distribution, and misalignment rates. We set $c/\Delta = 0.001$, $r = 0.05$, and $\gamma = 2.75$. The dotted line in the last panel indicates the value of γ, the maximum trading rate needed for fundamental trades.

constant. We also note in the bottom left panel that 97 percent of traders in the equilibrium allocation and 93 percent in the optimal allocation have a trading rate that exceeds γ, a natural upper bound for an economy without intermediation. The explanation for these additional trades is intermediation.

If the socially optimal contact rate distribution were just a proportionately shifted version of the equilibrium contact rate distribution (e.g. everyone had half as many contacts), the counterparty distributions would be shifted by the same proportion. The top right panel in Figure 2 shows that this is not the case. The equilibrium counterparty distribution does not first order stochastically dominate the optimal one. Instead, we find that in the socially optimal allocation, a larger fraction of meetings are with middlemen and other fast traders compared with the equilibrium. This ensures that fast traders more often encounter other fast traders. Since faster traders are more frequently misaligned, this facilitates fundamental trades, reducing their misalignment rate.

Finally, the bottom right panel in Figure 2 plots the misalignment rate as a function of the contact rate in both equilibrium and the optimal allocation. As expected from Propositions 3 and 3-P, traders with a higher contact rate have a higher misalignment rate. Moreover, we can see that the equilibrium misalignment rate of fast traders is too high. This reflects the relative scarcity of meetings with middlemen and other fast traders in the decentralized equilibrium.

Figure 3 shows the robustness of these results to the level of cost c/Δ. The top left panel shows that the lower bound on the optimal contact rate distribution is lower than under the equilibrium contact rate distribution. The top right panel shows that the average contact rate is higher relative to the lower bound in the optimum than the equilibrium. The bottom left panel shows that middlemen play a more prominent role in the optimum, as we found in the example with $c/\Delta = 0.001$. And the bottom right panel shows that both total and intermediation volume is inefficiently high in equilibrium. These last three results are all consistent with our analytical results with vanishing costs. We show here that they hold more generally. We find qualitatively similar results with other values of r and γ

7.3 Equilibrium vs. Optimum: Excessive Trade

Although the qualitative features of the equilibrium and optimal allocations are nearly identical, the equilibrium allocation is still inefficient. This section discusses the externalities and subsequently shows, using two separate formal arguments, that the planner discourages investment in contacts, consistent with our numerical results.

The inefficiency is rooted in externalities in the decentralized contact technology. When

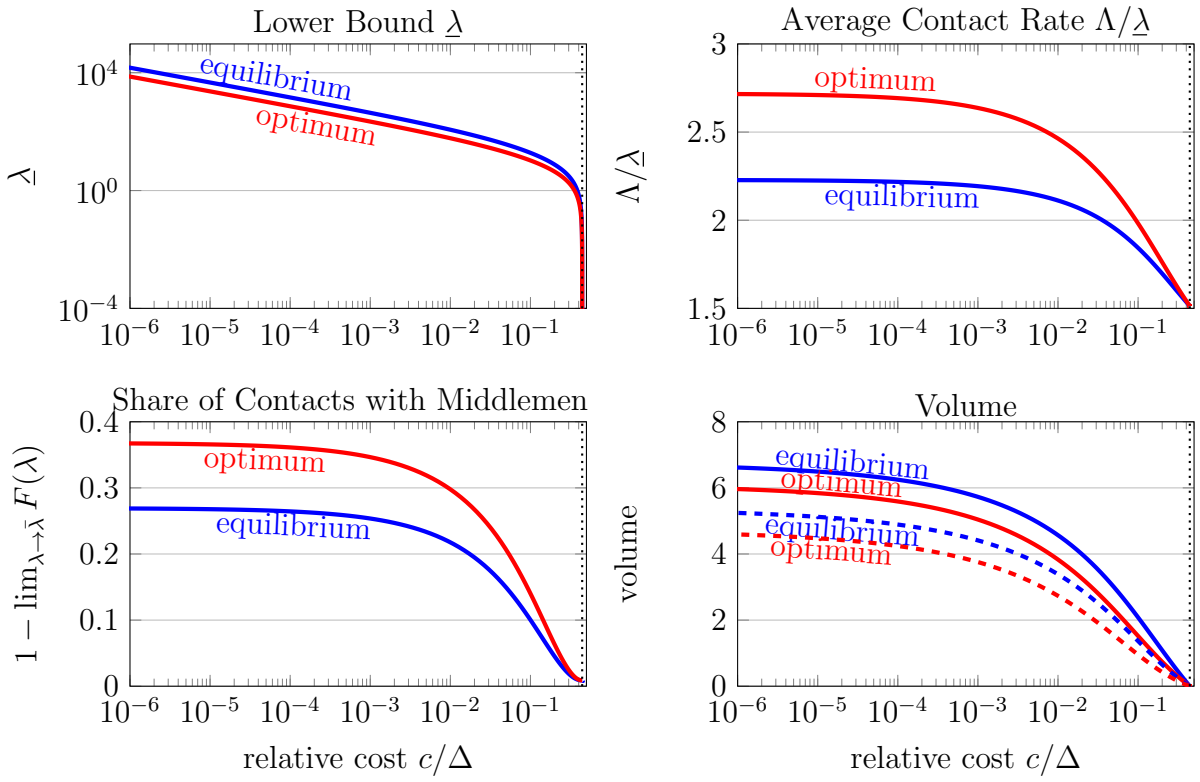

Figure 3: Features of the equilibrium and optimal allocation given relative costs c/Δ with $r = 0.05$ and $\gamma = 2.75$. The vertical dotted line indicates the value of c/Δ where the market shuts down, $\bar{c}^*$, the same in both equilibrium and the optimal allocation. The solid lines in the bottom right panel show total volume, while the dashed lines show intermediation volume.

a trader invests in more contacts, she diverts contacts towards herself and away from other traders. This reflects the fact that investing more in meetings does not affect the contact rate of the other traders, but it changes the distribution of whom they meet.

The externalities can be seen by comparing equation (8) with (12) and equation (10) with (14). The external cost of one trader increasing her contact rate is that doing so reduces the rate that other traders meet each other. We capture this *congestion externality* in our characterization of the social optimum by imposing a constant cost $\bar{\theta}$ on meetings in equation (14). The external benefit of the trader increasing her contact rate is that other traders value meeting her and she only captures half of this through Nash bargaining. We capture this *thick-market externality* in our characterization by doubling the value of trade in equations (12) and (14). Hence, the social surplus captures the full joint value of these exchanges. This is in contrast to the private surplus, which disregards the half that accrues to the counterparty.

The surplus in a typical trade, and hence the thick-market externality, is generally higher for slower traders, reflecting the fact that the social surplus function is globally decreasing (Lemma 1-P). Conversely, the congestion externality is constant for all traders. We can directly correct for each of these externalities. In Online Appendix E.1, we show that a simple tax and subsidy scheme, where traders get type-specific payments which depend on their alignment status, decentralizes the planning allocation. We show that the scheme subsidizes the misaligned state relative to the well-aligned state. Interestingly, when the cost function C is linear, the marginal subsidy, averaging across the misaligned and well-aligned states, is zero, so Pigouvian taxes do not distort investment choices by transferring resources directly between traders. Instead, the tax and subsidy scheme works by manipulating the threat points in bargaining through subsidies to misaligned traders and taxes on well-aligned ones. This shifts the terms of trade in favor of slower traders, which in turn discourages investment.

We offer a complementary perspective on overinvestment in Online Appendix E.2. We again study the linear cost case and consider a situation where the counterparty distribution F is exogenously given at its socially optimal level, but prices are set through decentralized bargaining without taxes or subsidies. In other words, we drop the third part of the definition of equilibrium. We then examine the incentives of a single trader who enters such a marketplace and can choose her contact rate to maximize her expected revenue net of investment costs, π_λ in equation (10). While the social planner is indifferent across all values $\lambda \geq \underline{\lambda}$, we show that private payoffs are not constant. Instead, an individual trader confronted with the optimal distribution F has a strictly increasing and unbounded profit function π. Aligning with the intuition above, this shows that private incentives lead to excessive investment in

equilibrium. Put differently, optimal policy must discourage private incentives to overinvest.

In closing, we connect the difference between the equilibrium and optimal allocation with ongoing debates about financial or securities transaction taxes (FTT or STT) (Tobin, 1978; Burman, Gale, Gault, Kim, Nunns and Rosenthal, 2016; Hemmelgarn, Nicodème, Tasnadi and Vermote, 2016). The Unites States currently implements an FTT, set at roughly 2 cents per $1,000 traded (SEC, 2019; Klein, 2020). Many European countries impose FTTs at varying levels covering stocks, bonds, and derivatives.

Our results show that there is overinvestment and excessive trade in equilibrium, which broadly makes the case for policies that discourage trade. More specifically, our model implies that low-value transactions, e.g. intermediated trades between two fast traders, should be taxed more heavily than high-value transactions, e.g. fundamental trades between two slow misaligned traders. Trades between a slow trader and a middleman should face an intermediate tax. Thus our model offers a novel and natural rationale for policies that selectively tax traders in the core of the financial network while going easy on infrequent market participants with low volume. We stress, however, that the goal of the tax should not be the elimination of intermediation. On the contrary, we have seen that an optimal policy reduces the number of meetings across the board but, in relative terms, redirects meetings towards middlemen and other fast traders.

8 Constrained Economy: The Role of Intermediation

Without intermediation, our model would not generate dispersed contact rates. To prove this, we consider an economy in which meetings between two traders with the same tastes do not occur, ending the scope for intermediation. It follows that whenever a misaligned trader meets a well-aligned trader, they have opposite tastes and hence the same asset holdings, and so there is no scope for trade. We show in this section that without intermediation, the equilibrium and optimal distributions of contact rates are degenerate as long as the cost function $C(\lambda)$ is weakly convex.

In Online Appendix F, we first offer the adjusted definition of equilibrium along with the adjusted planner problem that correspond to this setting. We then prove the following result:

Proposition 6 *Consider an economy with no intermediation and a weakly convex cost function $C : \mathcal{X} \to \mathbb{R}$. In equilibrium, all traders choose a common value λ. The same holds in the solution to the planner's problem.*

The proposition highlights that the heterogeneity that arises in the full economy is an im-

mediate, and socially desirable, consequence of intermediation. When traders are restricted to trades driven by static fundamentals, there is no gain from heterogeneity in the contact rate. This result reflects that, without intermediation, there is effectively decreasing returns to contacts at the individual level. The misalignment rate is strictly decreasing in meetings; as a trader becomes increasingly well-aligned, fewer meetings lead to gainful trading opportunities. As a consequence, an unequal distribution of meetings comes with first-order losses and the optimal distribution is degenerate. The same is true in equilibrium; with a weakly convex cost function but decreasing returns on the individual level, all individuals choose the same contact rate.

In summary, intermediation and heterogeneity are interconnected in a market with search frictions. Without heterogeneity there is no intermediation, and without intermediation there is no heterogeneity. Heterogeneity is useful because in meetings where both sides have identical tastes, misalignment can be transmitted towards the faster trader to facilitate the transfer of the asset to those who desire it.

9 Conclusions

We study a model of over-the-counter trading in which ex-ante identical traders invest in a contact technology and participate in bilateral trade. We show that when traders have heterogeneous search efficiencies, fast traders intermediate for slow traders: they trade against their desired position and take on misalignment from slower traders. Moreover, we characterize how, starting with ex-ante homogeneous traders, the distribution of contact rates is determined endogenously in equilibrium, and how it compares with the corresponding Pareto optimal distribution. We argue that an economy with homogeneous contact rates is neither an equilibrium nor socially desirable when the cost of meetings is differentiable. Under a linear cost function, the equilibrium and optimal distributions of trading rates are governed by a power law, an empirical feature of various financial markets. Moreover, middlemen with the highest possible contact rate account for a positive fraction of meetings. We also characterize the transfer scheme which decentralizes the optimal allocation, offsetting the forces that lead to overinvestment in the undistorted equilibrium. Finally, we argue that when intermediation is prohibited, dispersion in contact rates disappears both in equilibrium and in the optimal allocation, which illustrates the interplay between heterogeneity and intermediation in a frictional marketplace.

We close by highlighting areas for future research. An important one is a general existence result, beyond the linear cost case. Furthermore, we have kept our model as simple as possible in order to show how intermediation and middlemen naturally arise in over-the-counter

markets. It would be interesting to extend our model to a more complex environment, for example one in which traders differ ex-ante in how much they care about having a well-aligned asset position. This might "purify" the mixed strategy equilibrium we study here. Recall that under natural restrictions on the cost function, slow traders' asset positions are well-aligned with their taste more often than the faster traders who intermediate for them. We therefore conjecture that traders who care the least about their alignment status are the natural intermediaries and have the highest incentives to invest in a high contact rate. Likewise, we believe that the random matching model with endogenous contact rates may be useful for understanding other issues in financial markets, such as the percolation of information (Duffie and Manso, 2007). We hypothesize that middlemen may serve a useful role in this process as well.

References

Afonso, Gara and Ricardo Lagos, "Trade Dynamics in the Market for Federal Funds," *Econometrica*, 2015, *83* (1), 263–313.

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

Biais, Bruno and Richard Green, "The microstructure of the bond market in the 20th century," *Revew of Economic Dynamics*, 2019.

Boss, Michael, Helmut Elsinger, Martin Summer, and Stefan Thurner, "Network Topology of the Interbank Market," *Quantitative finance*, 2004, *4* (6), 677–684.

Burdett, Kenneth and Dale T. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, pp. 257–273.

_ **and Kenneth L. Judd**, "Equilibrium Price Dispersion," *Econometrica*, 1983, *51* (4), 955–969.

Burman, Leonard E, William G Gale, Sarah Gault, Bryan Kim, Jim Nunns, and Steve Rosenthal, "Financial transaction taxes in theory and practice," *National Tax Journal*, 2016, *69* (1), 171–216.

Butters, Gerard R., "Equilibrium Distributions of Sales and Advertising Prices," *The Review of Economic Studies*, 1977, *44* (3), 465–491.

Cabrales, Antonio, Antoni Calvó-Armengol, and Yves Zenou, "Social interactions and spillovers," *Games and Economic Behavior*, 2011, *72* (2), 339–360.

Chang, Briana and Shengxing Zhang, "Endogenous market making and network formation," 2019.

Cont, Rama, Amal Moussa, and Edson B. Santos, "Network structure and systemic risk in banking systems," *Edson Bastos e, Network Structure and Systemic Risk in Banking Systems (December 1, 2010)*, 2010.

Craig, Ben and Goetz Von Peter, "Interbank tiering and money center banks," *Journal of Financial Intermediation*, 2014, *23* (3), 322–347.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin, "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 2009, *77* (4), 1003–1045.

Duffie, Darrell and Gustavo Manso, "Information Percolation in Large Markets," *American Economic Review*, 2007, *97* (2), 203–209.

_ **, Nicolae Gârleanu, and Lasse Heje Pedersen**, "Over-the-Counter Markets," *Econometrica*, 2005, *73* (6), 1815–1847.

_ **,** _ **, and** _ , "Valuation in Over-the-Counter Markets," *Review of Financial Studies*, 2007, *20* (6), 1865–1900.

_ **, Piotr Dworczak, and Haoxiang Zhu**, "Benchmarks in search markets," *The Journal of Finance*, 2017, *72* (5), 1983–2044.

_ **, Semyon Malamud, and Gustavo Manso**, "Information Percolation With Equilibrium Search Dynamics," *Econometrica*, 2009, *77* (5), 1513–1574.

Dugast, Jérôme, Semih Üslü, and Pierre-Olivier Weill, "A theory of participation in OTC and centralized markets," Technical Report, National Bureau of Economic Research 2019.

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer, "The Emergence of Market Structure," NBER WP 23234 March 2017.

_ **,** _ **, and** _ , "Internal and External Effects of Social Distancing in a Pandemic," 2021.

_ **,** _ **, Guido Menzio, and Ursula Wiriadinata**, "Intermediation as Rent Extraction," 2019.

Fricke, Daniel and Thomas Lux, "Core–periphery structure in the overnight money market: evidence from the e-mid trading platform," *Computational Economics*, 2015, *45* (3), 359–395.

Galeotti, Andrea and Luca Paolo Merlino, "Endogenous job contact networks," *International Economic Review*, 2014, *55* (4), 1201–1226.

Hemmelgarn, Thomas, Gaëtan Nicodème, Bogdan Tasnadi, and Pol Vermote, "Financial transaction taxes in the European Union," *National Tax Journal*, 2016, *69* (1), 217.

Hendershott, Terrence, Dan Li, Dmitry Livdan, and Norman Schürhoff, "Relationship Trading in Over-the-Counter Markets," *The Journal of Finance*, 2020, *75* (2), 683–734.

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

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill, "Frictional intermediation in over-the-counter markets," *The Review of Economic Studies*, 2020, *87* (3), 1432–1469.

in 't Veld, Daan and Iman van Lelyveld, "Finding the Core: Network structure in interbank markets," *Journal of Banking & Finance*, 2014, *49*, 27–40.

Jackson, Matthew O. and Asher Wolinsky, "A strategic model of social and economic networks," *Journal of Economic Theory*, 1996, *71* (1), 44–74.

Klein, Aaron, "What is a financial transaction tax?," *Brookings Policy*, 2020.

Kremer, Michael, "Integrating behavioral choice into epidemiological models of AIDS," *The Quarterly Journal of Economics*, 1996, *111* (2), 549–573.

Kuprianov, Anatoli, "Over-the-counter interest rate derivatives," *FRB Richmond Economic Quarterly*, 1993, *79* (3), 65–94.

Lagos, Ricardo and Guillaume Rocheteau, "Liquidity in Asset Markets with Search Frictions," *Econometrica*, 2009, *77*, 403–426.

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

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

Martinez-Jaramillo, Serafin, Biliana Alexandrova-Kabadjova, Bernardo Bravo-Benitez, and Juan Pablo Solórzano-Margain, "An empirical study of the Mexican banking system's network and its implications for systemic risk," *Journal of Economic Dynamics and Control*, 2014, *40*, 242–265.

Masi, Giulia De, Giulia Iori, and Guido Caldarelli, "Fitness model for the Italian interbank money market," *Physical Review E*, 2006, *74* (6), 066112.

_ , _ , Ovidiu Vasile Precup, Giampaolo Gabbi, and Guido Caldarelli, "A network analysis of the Italian overnight money market," *Journal of Economic Dynamics and Control*, 2008, *32* (1), 259–278.

Nekludyov, Artmem, "Bid-ask spreads and the over-the-counter interdealer markets: Core and peripheral dealers," *Review of Economic Dynamics*, 2019, *33*, 57 – 84.

Pagnotta, Emiliano S. and Thomas Philippon, "Competing on Speed," *Econometrica*, 2018, *86* (3), 1067–1115.

Peltonen, Tuomas A., Martin Scheicher, and Guillaume Vuillemey, "The network structure of the CDS market and its determinants," *Journal of Financial Stability*, 2014, *13*, 118–133.

Petrongolo, Barbara and Christopher A. Pissarides, "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, June 2001, *39* (2), 390–431.

Philippon, Thomas, "Has the US finance industry become less efficient? On the theory and measurement of financial intermediation," *American Economic Review*, 2015, *105* (4), 1408–38.

Quercioli, Elena and Lones Smith, "Contagious Matching Games," Technical Report, Technical Report. Working paper 2006.

Rubinstein, Ariel and Asher Wolinsky, "Middlemen," *The Quarterly Journal of Economics*, 1987, pp. 581–594.

SEC, "Fee Rate Advisory #2 for Fiscal Year 2019," 2019.

Shimer, Robert and Lones Smith, "Matching, Search, and Heterogeneity," *Advances in Macroeconomics*, 2001, *1* (1).

Sideris, Thomas C., *Ordinary Differential Equations and Dynamical Systems*, Vol. 2 of *Atlantis Studies in Differential Equations*, Atlantis Press, 2013.

Tobin, James, "A proposal for international monetary reform," *Eastern economic journal*, 1978, *4* (3/4), 153–159.

Üslü, Semih, "Pricing and liquidity in decentralized asset markets," *Econometrica*, 2019, *87* (6), 2079–2140.

Vayanos, Dimitri and Tan Wang, "Search and Endogenous Concentration of Liquidity in Asset Markets," *Journal of Economic Theory*, 2007, *136* (1), 66–104.

Weill, Pierre-Olivier, "Liquidity Premia in Dynamic Bargaining Markets," *Journal of Economic Theory*, 2008, *140* (1), 66–96.

_ , "The Search Theory of Over-the-Counter Markets," *Annual Review of Economics*, 2020, *12* (1), 747–773.

Appendix

A Contact Rate Distribution

A.1 Additional Details

We would like to invert equation (1) to recover μ_G from μ_F, but unfortunately this is impossible without further restrictions. To see why, fix a measure μ_G and a number $\alpha \in (0,1)$ and then define a new measure $\tilde{\mu}_G \equiv \alpha\mu_G(S) + (1-\alpha)\mathbb{I}_{0\in S}$, where the indicator function is 1 if $0 \in S$ and 0 otherwise. Since $\int_S \lambda d\tilde{\mu}_G(\lambda) = \alpha \int_S \lambda d\mu_G(\lambda)$ for all $S \subset \mathcal{B}$, equation (1) implies the two contact rate distributions have the same counterparty distribution. Intuitively, they differ only in the fraction of the population with a zero contact rate. Since these traders never meet anyone, the fraction does not affect the counterparty distribution.

We mitigate this issue by imposing a natural restriction, beyond equation (1), on the share of traders with a zero contact rate, $\mu_G(\{0\})$. Recall that $\mathcal{Y}$ is the set of utility maximizing contact rates. If $0 \notin \mathcal{Y}$, we impose $\mu_G(\{0\}) = 0$. When this is the case, we have $\Lambda \equiv \int_{\mathcal{X}} \lambda d\mu_G(\lambda) > 0$. Then we can use the Radon-Nikodym theorem and equation (1) to move between the probability measures $\mu_G(\lambda)$ and $\mu_F(\lambda)$:

$$\frac{d\mu_F(\lambda)}{d\mu_G(\lambda)} = \frac{\lambda}{\Lambda}. \tag{16}$$

Multiply both sides by $\frac{1}{\lambda}$ and integrate both sides under the measure μ_G to get

$$\int_{\mathcal{X}} \frac{1}{\lambda}\frac{d\mu_F(\lambda)}{d\mu_G(\lambda)} d\mu_G(\lambda) = \int_{\mathcal{X}} \frac{1}{\Lambda} d\mu_G(\lambda) = \frac{1}{\Lambda}$$

or

$$\Lambda = \frac{1}{\int_{\mathcal{X}} \frac{1}{\lambda} d\mu_F(\lambda)}. \tag{17}$$

Together equations (16) and (17) define Λ and μ_G given μ_F whenever $\frac{1}{\lambda}$ is Lebesgue integrable under the measure μ_F and $0 \notin \mathcal{Y}$.

If $\frac{1}{\lambda}$ is Lebesgue integrable under the measure μ_F but $0 \in \mathcal{Y}$, then the right hand side of equation (17) is an upper bound on Λ. For any value of $\Lambda \leq 1/\int_{\mathcal{X}} \frac{1}{\lambda} d\mu_F(\lambda)$ and any set S with $0 \notin S$, we can then find $\mu_G(S)$ using equation (16). We then set $\mu_G(\{0\}) = 1 - \mu_G((0, \bar{\lambda}])$. This is a valid contact rate measure associated with the counterparty measure μ_F.

Finally, if $\frac{1}{\lambda}$ is not Lebesgue integrable under the measure μ_F, then $\Lambda = 0$ and $\mu_G(\{0\}) = 1$. This is the case whenever $\mu_F(\{0\}) > 0$.

A.2 Physical Matching Process

Consider an economy with a (large) finite number of traders $n \geq 2$ and (short) finite periods of length dt. Each trader has a type λ satisfying $0 \leq \lambda \leq \bar{\lambda}$, with $dt < 1/\bar{\lambda}$. The type determines the probability of matching during each period. Matching proceeds in two stages. First, a trader with type λ draws a binary outcome, 0 or 1, with probability $1 - \lambda dt$ and λdt, respectively. Let $k \leq n$ denote the number of traders who draw 1. If k is even, all traders who drew 1 match in pairs, with all such pairings equally likely. If k is odd, uniformly randomly select one trader who drew 0 and switch that outcome to 1. Then match all 1's in pairs, with all such pairings equally likely.

We are interested in the limit where $n \to \infty$. In such an economy, the per-period matching probability for a type λ trader converges to λdt. This is because $\bar{\lambda} dt < 1$ ensures that many traders draw 0. Since at most one trader draws 0 and gets switched to 1, the likelihood of this happening to any particular trader is infinitesimal.

Now consider an economy in autarky, where everyone sets $\lambda = 0$. A single trader can deviate and set a strictly positive value for λ, in which case he matches with a counterparty with contact rate 0 with probability λdt in each period. In this case, $\mu_F(\{0\}) = \mu_G(\{0\}) = 1$.

More generally, this environment can accommodate a limit where $\mu_G(\{0\}) = 1$ but $\mu_F(\{0\}) < 1$. To see how, fix a number $p > 0$ and consider an economy with $n > p$ traders. Assume each trader has contact rate $\lambda > 0$ with probability p/n and 0 otherwise. The traders then draw 0 or 1 as described above. For finite n, the number of traders who draw a 1 each period is a Binomial random variable with parameters $(n, p\lambda dt/n)$. As n grows without bound, this converges to a Poisson random variable with mean $p\lambda dt$ and hence density $e^{-p\lambda dt}(p\lambda dt)^k/k!$ for $k = 0, 1, 2, \ldots$.

Using this, one can compute the counterparty distribution for those who have chosen a contact rate $\lambda > 0$. (For those who chose a zero contact rate, all counterparties have a contact rate λ, but those matches almost never occur.) For example, if in a given period $k = 1$, then that trader matches with a zero-contact-rate trader. If $k = 2$, the two traders match with each other, i.e. with a λ-contact-rate trader. If $k = 3$, it is a mix of the previous cases, so the counterparty distribution for a λ-contact-rate trader puts weight 1/3 on a zero-contact-rate counterparty. Summing across k and weighting implies that the fraction of traders who match with a zero contact rate partner is

$$\frac{\sum_{k \text{ odd}} \frac{e^{-p\lambda dt}(p\lambda dt)^k}{k(k!)}}{p\lambda dt},$$

where $e^{-p\lambda dt}(p\lambda dt)^k/k!$ is the density of the Poisson distribution, $1/k$ is the odds of matching

with a zero partner when k is odd, and $p\lambda dt$ is the expected number of traders who choose $\lambda > 0$ and match. Through an appropriate choice of p, this sum can take any value between 0 and 1.

This example describes a physical matching process where as $n \to \infty$, $\mu_G(\{0\}) = 1 > \mu_F(\{0\})$. The key is that in a large economy, almost everyone has a zero contact rate, but counterparties are very different than the typical trader.

B Differential Equation System

In Online Appendix C.3, we prove that on the interior of its support, (F, m, s) solves the following differential equation system:

$$F'(\lambda) = \frac{\big(2r + 4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda))\big)(8\gamma(1 - F(\lambda)) - 8rM(\lambda) + \zeta(\lambda))}{2\lambda\big(\gamma(8r + 8\gamma + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda)))\big)}, \quad (18)$$

$$M'(\lambda) = \frac{\big(2\gamma + \lambda M(\lambda)\big)(8\gamma(1 - F(\lambda)) - 8rM(\lambda) + \zeta(\lambda))}{2\lambda\big(\gamma(8r + 8\gamma + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda)))\big)}, \quad (19)$$

$$s'(\lambda) = \frac{4\left((r + 2\gamma)s(\lambda) - \Delta\right)}{\lambda\big(4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))\big)}, \quad (20)$$

where

$$\zeta(\lambda) \equiv \frac{r\lambda\big(4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))\big)^2 C''(\lambda)}{\Delta - (r + 2\gamma)s(\lambda)}; \quad (21)$$

see equations (34) and (50)–(52). Moreover, we have terminal conditions $F(\underline{\lambda}) = M(\underline{\lambda}) = 0$ and $s(\bar{\lambda}) = \frac{2\Delta}{2r+4\gamma+\bar{\lambda}M(\bar{\lambda})}$, as well as the requirement that F and M are nondecreasing.

If the cost function is linear, $C''(\lambda) = \zeta(\lambda) = 0$ and so we can solve the differential equations (18) and (19) for F and M alone.

Online Appendix

C Equilibrium — Details and Proofs

C.1 Symmetry

We now restrict attention to equilibria in which the two misaligned states and the two well-aligned states are treated symmetrically. That is, we look only at equilibria where $p_{\lambda,i,b}^{\lambda',i',b'} = p_{\lambda,\sim i,1-b}^{\lambda',\sim i',1-b'}$, so if a type λ trader sells the asset to a type λ' trader when both are in state h, then a type λ trader must buy the asset from a type λ' trader when both are in state l.

We proceed in two steps. First, we show that symmetry of the trading probabilities, $p_{\lambda,i,b}^{\lambda',i',b'} = p_{\lambda,\sim i,1-b}^{\lambda',\sim i',1-b'}$, implies symmetry of the steady state shares, $\sigma_{\lambda,i,b} = \sigma_{\lambda,\sim i,1-b}$. Second, we show that this implies we can use a surplus function to characterize who matches with whom and use the surplus function to derive equations (8)–(10) in the body of the paper.

Step 1: Trading Probabilities When $p_{\lambda,i,b}^{\lambda',i',b'} = p_{\lambda,\sim i,1-b}^{\lambda',\sim i',1-b'}$, equation (6) implies $\sigma_{\lambda,i,b} = \sigma_{\lambda,\sim i,1-b}$ for all (λ, i, b). That is, the fraction of traders with contact rate λ in the high state, $i = h$, who hold the asset, $b = 1$, is equal to the fraction of traders with the same contact rate who are in the low state, $i = l$, and do not hold the asset, $b = 0$. Equivalently, the fractions of type-λ traders in either well-aligned state are equal. The remaining traders are misaligned, and again there are equal shares of the two misaligned states for each λ. Thus we define the misalignment rate $m_\lambda \equiv \sigma_{\lambda,l,1} + \sigma_{\lambda,h,0} = 2\sigma_{\lambda,l,1} = 2\sigma_{\lambda,h,0}$ and $1 - m_\lambda = \sigma_{\lambda,l,0} + \sigma_{\lambda,h,1} = 2\sigma_{\lambda,l,0} = 2\sigma_{\lambda,h,1}$.

In a symmetric equilibrium, it is convenient to refer to traders only by their alignment status a, where $a = 0$ indicates misaligned and $a = 1$ indicates well-aligned. Let $p_{\lambda,a}^{\lambda',a'}$ indicate the trading probability between traders (λ, a) and (λ', a') conditional on them having the opposite asset holdings; there cannot be trade if they have the same asset holdings. Equation (6) reduces to

$$\left(r + \gamma + \frac{\lambda}{2}\int_{\mathcal{X}} \left(p_{\lambda,0}^{\lambda',0} m_{\lambda'} + p_{\lambda,0}^{\lambda',1}(1 - m_{\lambda'})\right) d\mu_F(\lambda')\right) m_\lambda$$
$$= \left(\gamma + \frac{\lambda}{2}\int_{\mathcal{X}} \left(p_{\lambda,1}^{\lambda',0} m_{\lambda'} + p_{\lambda,1}^{\lambda',1}(1 - m_{\lambda'})\right) d\mu_F(\lambda')\right)(1 - m_\lambda). \quad (22)$$

Step 2: Surplus Function Next, we prove that when $p_{\lambda,i,b}^{\lambda',i',b'} = p_{\lambda,\sim i,1-b}^{\lambda',\sim i',1-b'}$, there is a surplus function $s : [0, \bar{\lambda}] \to \mathbb{R}$ which tells us whether trade occurs. First, condition (5)

implies

$$v_{\lambda,h,1} - v_{\lambda,h,0} - t_{\lambda,h,0}^{\lambda',h,1} = v_{\lambda',h,0} - v_{\lambda',h,1} + t_{\lambda,h,0}^{\lambda',h,1} = \frac{v_{\lambda,h,1} + v_{\lambda',h,0} - v_{\lambda,h,0} - v_{\lambda',h,1}}{2}, \tag{23a}$$

$$v_{\lambda,h,1} - v_{\lambda,h,0} - t_{\lambda,h,0}^{\lambda',l,1} = v_{\lambda',l,0} - v_{\lambda',l,1} + t_{\lambda,h,0}^{\lambda',l,1} = \frac{v_{\lambda,h,1} + v_{\lambda',l,0} - v_{\lambda,h,0} - v_{\lambda',l,1}}{2}, \tag{23b}$$

$$v_{\lambda,l,1} - v_{\lambda,l,0} - t_{\lambda,l,0}^{\lambda',h,1} = v_{\lambda',h,0} - v_{\lambda',h,1} + t_{\lambda,l,0}^{\lambda',h,1} = \frac{v_{\lambda,l,1} + v_{\lambda',h,0} - v_{\lambda,l,0} - v_{\lambda',h,1}}{2}, \tag{23c}$$

$$v_{\lambda,l,1} - v_{\lambda,l,0} - t_{\lambda,l,0}^{\lambda',l,1} = v_{\lambda',l,0} - v_{\lambda',l,1} + t_{\lambda,l,0}^{\lambda',l,1} = \frac{v_{\lambda,l,1} + v_{\lambda',l,0} - v_{\lambda,l,0} - v_{\lambda',l,1}}{2} \tag{23d}$$

whenever the last term on each line is positive. Next, rewrite equation (2) explicitly as

$$rv_{\lambda,h,1} = \delta_{h,1} + \gamma\big(v_{\lambda,l,1} - v_{\lambda,h,1}\big) + \lambda \int_{\mathcal{X}} \sum_{i' \in \{h,l\}} \sigma_{\lambda',i',0} p_{\lambda,h,1}^{\lambda',i',0} \big(v_{\lambda,h,0} - v_{\lambda,h,1} - t_{\lambda,h,1}^{\lambda',i',0}\big) d\mu_F(\lambda'),$$

$$rv_{\lambda,h,0} = \delta_{h,0} + \gamma\big(v_{\lambda,l,0} - v_{\lambda,h,0}\big) + \lambda \int_{\mathcal{X}} \sum_{i' \in \{h,l\}} \sigma_{\lambda',i',1} p_{\lambda,h,0}^{\lambda',i',1} \big(v_{\lambda,h,1} - v_{\lambda,h,0} - t_{\lambda,h,0}^{\lambda',i',1}\big) d\mu_F(\lambda'),$$

$$rv_{\lambda,l,1} = \delta_{l,1} + \gamma\big(v_{\lambda,h,1} - v_{\lambda,l,1}\big) + \lambda \int_{\mathcal{X}} \sum_{i' \in \{l,h\}} \sigma_{\lambda',i',0} p_{\lambda,l,1}^{\lambda',i',0} \big(v_{\lambda,l,0} - v_{\lambda,l,1} - t_{\lambda,l,1}^{\lambda',i',0}\big) d\mu_F(\lambda'),$$

$$rv_{\lambda,l,0} = \delta_{l,0} + \gamma\big(v_{\lambda,h,0} - v_{\lambda,l,0}\big) + \lambda \int_{\mathcal{X}} \sum_{i' \in \{l,h\}} \sigma_{\lambda',i',1} p_{\lambda,l,0}^{\lambda',i',1} \big(v_{\lambda,l,1} - v_{\lambda,l,0} - t_{\lambda,l,0}^{\lambda',i',1}\big) d\mu_F(\lambda').$$

Using symmetry and the Nash bargaining solution as summarized in equation (23), rewrite these as

$$\begin{aligned} rv_{\lambda,h,1} = \delta_{h,1} + \gamma\big(v_{\lambda,l,1} - v_{\lambda,h,1}\big) &+ \frac{\lambda}{4} \int_{\mathcal{X}} m_{\lambda'} \big(v_{\lambda',h,1} + v_{\lambda,h,0} - v_{\lambda',h,0} - v_{\lambda,h,1}\big)^+ d\mu_F(\lambda') \\ &+ \frac{\lambda}{4} \int_{\mathcal{X}} (1 - m_{\lambda'}) \big(v_{\lambda',l,1} + v_{\lambda,h,0} - v_{\lambda',l,0} - v_{\lambda,h,1}\big)^+ d\mu_F(\lambda'), \end{aligned} \tag{24a}$$

$$\begin{aligned} rv_{\lambda,h,0} = \delta_{h,0} + \gamma\big(v_{\lambda,l,0} - v_{\lambda,h,0}\big) &+ \frac{\lambda}{4} \int_{\mathcal{X}} m_{\lambda'} \big(v_{\lambda,h,1} + v_{\lambda',l,0} - v_{\lambda,h,0} - v_{\lambda',l,1}\big)^+ d\mu_F(\lambda') \\ &+ \frac{\lambda}{4} \int_{\mathcal{X}} (1 - m_{\lambda'}) \big(v_{\lambda,h,1} + v_{\lambda',h,0} - v_{\lambda,h,0} - v_{\lambda',h,1}\big)^+ d\mu_F(\lambda'), \end{aligned} \tag{24b}$$

$$\begin{aligned} rv_{\lambda,l,1} = \delta_{l,1} + \gamma\big(v_{\lambda,h,1} - v_{\lambda,l,1}\big) &+ \frac{\lambda}{4} \int_{\mathcal{X}} m_{\lambda'} \big(v_{\lambda',h,1} + v_{\lambda,l,0} - v_{\lambda',h,0} - v_{\lambda,l,1}\big)^+ d\mu_F(\lambda') \\ &+ \frac{\lambda}{4} \int_{\mathcal{X}} (1 - m_{\lambda'}) \big(v_{\lambda',l,1} + v_{\lambda,l,0} - v_{\lambda',l,0} - v_{\lambda,l,1}\big)^+ d\mu_F(\lambda'), \end{aligned} \tag{24c}$$

$$\begin{aligned} rv_{\lambda,l,0} = \delta_{l,0} + \gamma\big(v_{\lambda,h,0} - v_{\lambda,l,0}\big) &+ \frac{\lambda}{4} \int_{\mathcal{X}} m_{\lambda'} \big(v_{\lambda,l,1} + v_{\lambda',l,0} - v_{\lambda,l,0} - v_{\lambda',l,1}\big)^+ d\mu_F(\lambda') \\ &+ \frac{\lambda}{4} \int_{\mathcal{X}} (1 - m_{\lambda'}) \big(v_{\lambda,l,1} + v_{\lambda',h,0} - v_{\lambda,l,0} - v_{\lambda',h,1}\big)^+ d\mu_F(\lambda'), \end{aligned} \tag{24d}$$

where again $z^+ \equiv \max\{z, 0\}$. Now, conjecture that $v_{\lambda,h,1} - v_{\lambda,h,0} = s(\lambda) + q$ and $v_{\lambda,l,0} - v_{\lambda,l,1} = s(\lambda) - q$ for some function $s(\lambda)$ and number q. Subtracting equation (24b) from equation (24a) gives

$$\begin{aligned} r(s(\lambda)+q) &= \delta_{h,1} - \delta_{h,0} - 2\gamma s(\lambda) \\ &\quad + \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}\big(-s(\lambda)+s(\lambda')\big)^+ + (1-m_{\lambda'})\big(-s(\lambda)-s(\lambda')\big)^+ \Big) d\mu_F(\lambda') \\ &\quad\quad - \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}\big(s(\lambda)+s(\lambda')\big)^+ + (1-m_{\lambda'})\big(s(\lambda)-s(\lambda')\big)^+ \Big) d\mu_F(\lambda'). \end{aligned} \tag{25a}$$

Subtracting equation (24c) from equation (24d) gives

$$\begin{aligned} r(s(\lambda)-q) &= \delta_{l,0} - \delta_{l,1} - 2\gamma s(\lambda) \\ &\quad + \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}\big(-s(\lambda)+s(\lambda')\big)^+ + (1-m_{\lambda'})\big(-s(\lambda)-s(\lambda')\big)^+ \Big) d\mu_F(\lambda') \\ &\quad\quad - \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}\big(s(\lambda)+s(\lambda')\big)^+ + (1-m_{\lambda'})\big(s(\lambda)-s(\lambda')\big)^+ \Big) d\mu_F(\lambda'). \end{aligned} \tag{25b}$$

The sum of equations (25a) and (25b) gives

$$\begin{aligned} 2rs(\lambda) &= 2\Delta - 4\gamma s(\lambda) \\ &\quad + \frac{\lambda}{2}\int_{\mathcal{X}} \Big(m_{\lambda'}\big(-s(\lambda)+s(\lambda')\big)^+ + (1-m_{\lambda'})\big(-s(\lambda)-s(\lambda')\big)^+ \Big) d\mu_F(\lambda') \\ &\quad\quad - \frac{\lambda}{2}\int_{\mathcal{X}} \Big(m_{\lambda'}\big(s(\lambda)+s(\lambda')\big)^+ + (1-m_{\lambda'})\big(s(\lambda)-s(\lambda')\big)^+ \Big) d\mu_F(\lambda'), \end{aligned}$$

from which equations (7) and (8) in the text follows immediately. Conversely, the difference between equations (25a) and (25b) gives us the value of q, consistent with its definition in equation (7). This validates our conjecture that $v_{\lambda,h,1} - v_{\lambda,h,0} = s(\lambda) + q$ and $v_{\lambda,l,0} - v_{\lambda,l,1} = s(\lambda) - q$.

Next, we simplify the flow balanced equation (22). Using $v_{\lambda,h,1} - v_{\lambda,h,0} = s(\lambda) + q$ and $v_{\lambda,l,0} - v_{\lambda,l,1} = s(\lambda) - q$, we get

$$v_{\lambda,h,1} - v_{\lambda,h,0} + v_{\lambda',l,0} - v_{\lambda',l,1} = s(\lambda) + s(\lambda') \tag{26a}$$

$$v_{\lambda,l,1} - v_{\lambda,l,0} + v_{\lambda',l,0} - v_{\lambda',l,1} = -s(\lambda) + s(\lambda') \tag{26b}$$

$$v_{\lambda,h,1} - v_{\lambda,h,0} + v_{\lambda',h,0} - v_{\lambda',h,1} = s(\lambda) - s(\lambda') \tag{26c}$$

$$v_{\lambda,l,1} - v_{\lambda,l,0} + v_{\lambda',h,0} - v_{\lambda',h,1} = -s(\lambda) - s(\lambda'). \tag{26d}$$

Then equation (4) implies that we can use the surplus function to tell when trade occurs. In particular, applying this to equation (22) gives us equation (9).[19]

In the last step, we obtain equation (10) by defining $\pi_\lambda = \frac{v_{\lambda,h,1}+v_{\lambda,l,0}}{2} - C(\lambda)$, replacing $v_{\lambda,h,1}$ and $v_{\lambda,l,0}$ using equations (24a) and (24d), respectively, and then simplifying with $v_{\lambda,h,1} - v_{\lambda,h,0} = s(\lambda) + q$ and $v_{\lambda,l,0} - v_{\lambda,l,1} = s(\lambda) - q$.

C.2 Recovering Primitive Objects

Given μ_F, m, and s, we can recover the primitive value function v, trading probability p, transfer t, and shares σ. To do this, we start with equation (24), using $s(\lambda) = v_{\lambda,h,1} - v_{\lambda,h,0} - q = v_{\lambda,l,0} - v_{\lambda,l,1} + q$ to write this as

$$\begin{aligned}
rv_{\lambda,h,1} &= \delta_{h,1} + \gamma\big(v_{\lambda,l,1} - v_{\lambda,h,1}\big) + \frac{\lambda}{4}\int_{\mathcal{X}} m_{\lambda'}\big(s(\lambda') - s(\lambda)\big)^+ d\mu_F(\lambda') \\
&\qquad + \frac{\lambda}{4}\int_{\mathcal{X}} (1 - m_{\lambda'})\big(-s(\lambda') - s(\lambda)\big)^+ d\mu_F(\lambda'), \\
rv_{\lambda,h,0} &= \delta_{h,0} + \gamma\big(v_{\lambda,l,0} - v_{\lambda,h,0}\big) + \frac{\lambda}{4}\int_{\mathcal{X}} m_{\lambda'}\big(s(\lambda) + s(\lambda')\big)^+ d\mu_F(\lambda') \\
&\qquad + \frac{\lambda}{4}\int_{\mathcal{X}} (1 - m_{\lambda'})\big(s(\lambda) - s(\lambda')\big)^+ d\mu_F(\lambda'), \\
rv_{\lambda,l,1} &= \delta_{l,1} + \gamma\big(v_{\lambda,h,1} - v_{\lambda,l,1}\big) + \frac{\lambda}{4}\int_{\mathcal{X}} m_{\lambda'}\big(s(\lambda) + s(\lambda')\big)^+ d\mu_F(\lambda') \\
&\qquad + \frac{\lambda}{4}\int_{\mathcal{X}} (1 - m_{\lambda'})\big(s(\lambda) - s(\lambda')\big)^+ d\mu_F(\lambda'), \\
rv_{\lambda,l,0} &= \delta_{l,0} + \gamma\big(v_{\lambda,h,0} - v_{\lambda,l,0}\big) + \frac{\lambda}{4}\int_{\mathcal{X}} m_{\lambda'}\big(s(\lambda') - s(\lambda)\big)^+ d\mu_F(\lambda') \\
&\qquad + \frac{\lambda}{4}\int_{\mathcal{X}} (1 - m_{\lambda'})\big(-s(\lambda') - s(\lambda)\big)^+ d\mu_F(\lambda'),
\end{aligned}$$

Since $r > 0$, we can solve the first and third equations for $v_{\lambda,h,1}$ and $v_{\lambda,l,1}$ given μ_F, m, and s, and similarly solve the second and fourth equations for $v_{\lambda,h,0}$ and $v_{\lambda,l,0}$.

Having recovered the value function, we then get the trading probabilities from equation (4) and the transfers from equation (5). We then set $\sigma_{\lambda,l,1} = \sigma_{\lambda,h,0} = m_\lambda/2$ and $\sigma_{\lambda,l,0} = \sigma_{\lambda,h,1} = (1 - m_\lambda)/2$, consistent with a symmetric equilibrium. Finally, using these, we can reverse the steps in Online Appendix C.1 to verify that the value functions solve equation (2). We can also verify the steady state equation (6) as well as equation (3) governing the choice of contact rate. Thus, armed with a symmetric equilibrium satisfying Definition 1,

[19] We assume that trade does not occur when the gains from trade is exactly zero. Alternative assumptions would not affect our characterization of equilibrium, although it could lead to a higher trading volume.

we can find primitive objects satisfying equations (2)–(6).

C.3 Characterization with General Cost Functions

Proof of Lemma 1. We prove that $s(\lambda)$ is strictly positive and strictly decreasing for all $\lambda > 0$. The remainder of the lemma then immediately follows from the Nash bargaining assumption.

To prove s is strictly positive, suppose to the contrary that $s(\lambda) \leq 0$ for some λ. Then $(-s(\lambda) + s(\lambda'))^+ \geq (s(\lambda) + s(\lambda'))^+$ and $(-s(\lambda) - s(\lambda'))^+ \geq (s(\lambda) - s(\lambda'))^+$ for all λ'. Equation (8) then implies $\Delta \leq (r + 2\gamma)s(\lambda) \leq 0$. Since $\Delta > 0$, this is a contradiction, which proves s is strictly positive.

Next, write equation (9) as

$$m_\lambda = \frac{\gamma + \lambda p_{1,\lambda}}{r + 2\gamma + \lambda(p_{0,\lambda} + p_{1,\lambda})} \tag{27}$$

where

$$\begin{aligned} p_{0,\lambda} &\equiv \frac{1}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{s(\lambda)+s(\lambda')>0} m_{\lambda'} + \mathbb{I}_{s(\lambda)>s(\lambda')}(1 - m_{\lambda'})\right) d\mu_F(\lambda'), \\ p_{1,\lambda} &\equiv \frac{1}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{s(\lambda)<s(\lambda')} m_{\lambda'} + \mathbb{I}_{s(\lambda)+s(\lambda')<0}(1 - m_{\lambda'})\right) d\mu_F(\lambda') \end{aligned}$$

define the probability that a trader's meetings result in trade as a function of her alignment status. Since s is positive and $m_\lambda \in [0,1]$, $\mathbb{I}_{s(\lambda)+s(\lambda')>0} m_{\lambda'} \geq \mathbb{I}_{s(\lambda)<s(\lambda')} m_{\lambda'}$ and $\mathbb{I}_{s(\lambda)>s(\lambda')}(1 - m_{\lambda'}) \geq \mathbb{I}_{s(\lambda)+s(\lambda')<0}(1 - m_{\lambda'})$. This implies $1/2 \geq p_{0,\lambda} \geq p_{1,\lambda} \geq 0$.

Now minimize the expression for m_λ in equation (27) subject to $1/2 \geq p_{0,\lambda} \geq p_{1,\lambda} \geq 0$. The minimum is $\frac{2\gamma}{2r+4\gamma+\lambda}$, achieved when $p_{0,\lambda} = 1/2$ and $p_{1,\lambda} = 0$. Similarly, the maximum value of m_λ subject to $1/2 \geq p_{0,\lambda} \geq p_{1,\lambda} \geq 0$ is $\frac{2\gamma+\lambda}{2r+4\gamma+2\lambda}$, achieved with $p_{0,\lambda} = p_{1,\lambda} = 1/2$. Since $\lambda \leq \bar{\lambda}$, $\frac{2\gamma}{2r+4\gamma+\lambda}$ is decreasing in λ, and $\frac{2\gamma+\lambda}{2r+4\gamma+2\lambda}$ is increasing in λ, this proves that in any equilibrium,

$$\underline{m} \equiv \frac{2\gamma}{2r + 4\gamma + \bar{\lambda}} \leq m_\lambda \leq \frac{2\gamma + \bar{\lambda}}{2r + 4\gamma + 2\bar{\lambda}} \equiv \bar{m}. \tag{28}$$

We stress that $0 < \underline{m} < \bar{m} < 1/2$.

Next, since s is strictly positive, we have $(s(\lambda) + s(\lambda'))^+ = s(\lambda) + s(\lambda')$ and $(-s(\lambda) - s(\lambda'))^+ = 0$. Moreover, $(s(\lambda') - s(\lambda))^+ = s(\lambda') - \min\{s(\lambda), s(\lambda')\}$ and $(s(\lambda) - s(\lambda'))^+ = s(\lambda) - \min\{s(\lambda), s(\lambda')\}$ regardless of the behavior of s. This allows us to rewrite equation (8)

as

$$(r+2\gamma)s(\lambda) = \Delta + \\ \frac{\lambda}{4}\int_{\mathcal{X}} \Big(\big(-\min\{s(\lambda), s(\lambda')\} - s(\lambda)\big) m_{\lambda'} - \big(s(\lambda) - \min\{s(\lambda), s(\lambda')\}\big)(1-m_{\lambda'})\Big) d\mu_F(\lambda').$$

Grouping terms, this gives

$$s(\lambda) = \frac{4\Delta + \lambda \int_{\mathcal{X}} \big(\min\{s(\lambda), s(\lambda')\}(1-2m_{\lambda'})\big) d\mu_F(\lambda')}{4r + 8\gamma + \lambda} \equiv T(s(\lambda)), \tag{29}$$

where T maps surplus functions into surplus functions.

We claim that for any measure μ_F and misalignment function m with range $[\underline{m}, \bar{m}]$, T is a contraction, mapping continuous functions on $[0, \Delta/(r+2\gamma)]$ into the same set of functions. Continuity is immediate. Similarly, if s is nonnegative, $T(s)$ is nonnegative. If $s \leq \Delta/(r+2\gamma)$,

$$T(s(\lambda)) \leq \left(\frac{4r + 8\gamma + \lambda \int_{\mathcal{X}} (1-2m_{\lambda'}) d\mu_F(\lambda')}{4r + 8\gamma + \lambda} \right) \left(\frac{\Delta}{r+2\gamma} \right).$$

Since the misalignment rate lies between $\underline{m} > 0$ and $\bar{m} < 1/2$ (inequality (28)), the result follows.

Finally, we prove T is a contraction. Consider two functions $s^1(\lambda)$ and $s^2(\lambda)$. If $|s^1(\lambda) - s^2(\lambda)| \leq \varepsilon$ for all λ,

$$|T(s^1(\lambda)) - T(s^2(\lambda))| \leq \frac{\lambda\varepsilon \int_{\mathcal{X}} (1-2m_{\lambda'}) d\mu_F(\lambda')}{4r + 8\gamma + \lambda} \leq \frac{\bar{\lambda}\varepsilon}{4r + 8\gamma + \bar{\lambda}}.$$

The second inequality uses $\int_{\mathcal{X}} (1-2m_{\lambda'}) d\mu_F(\lambda') \leq 1$ and $\frac{\lambda}{4r+8\gamma+\lambda} \leq \frac{\bar{\lambda}}{4r+8\gamma+\bar{\lambda}}$. This proves that T is a contraction in the sup-norm, with modulus $\frac{\bar{\lambda}}{4r+8\gamma+\bar{\lambda}} < 1$. It follows that the equilibrium surplus function is uniquely determined by μ_F.

Next we prove that the mapping T takes nonincreasing functions s and maps them into strictly decreasing functions. Take $\lambda_1 < \lambda_2$ and for notational convenience let

$$E(\lambda) \equiv \int_{\mathcal{X}} \min\{s(\lambda), s(\lambda')\}(1-2m_{\lambda'}) d\mu_F(\lambda').$$

Note that $m \geq \underline{m} > 0$ and $s(\lambda) \leq \frac{\Delta}{r+2\gamma}$ implies $E(\lambda) < \Delta/(r+2\gamma)$. Similarly, s nonincreasing

implies E is nonincreasing as well. Then for any $\lambda_1 < \lambda_2$,

$$\begin{aligned} T(s(\lambda_1)) - T(s(\lambda_2)) &= \frac{4\Delta + \lambda_1 E(\lambda_1)}{4r + 8\gamma + \lambda_1} - \frac{4\Delta + \lambda_2 E(\lambda_2)}{4r + 8\gamma + \lambda_2} \\ &\geq \frac{4\Delta + \lambda_1 E(\lambda_1)}{4r + 8\gamma + \lambda_1} - \frac{4\Delta + \lambda_2 E(\lambda_1)}{4r + 8\gamma + \lambda_2} \\ &= \frac{4(\lambda_2 - \lambda_1)\big(\Delta - (r + 2\gamma)E(\lambda_1)\big)}{(4r + 8\gamma + \lambda_1)(4r + 8\gamma + \lambda_2)} > 0, \end{aligned}$$

The first equality is the definition of T. The first inequality uses $E(\lambda_2) \leq E(\lambda_1)$. The second equality groups the two fractions over a common denominator. And the second inequality uses $E(\lambda) < \Delta/(r + 2\gamma)$. This proves the result. It follows that the equilibrium surplus function is decreasing. ■

Proof of Proposition 1. The proof proceeds in three steps. First, we show that the counterparty distribution uniquely determines the misalignment rate. Then, we derive the functional form for the surplus equation, thereby proving that the counterparty distribution uniquely determines the surplus function. Finally, we show how to recover the cost function from these three objects.

Step 1: Recovering misalignment For any set $S \subseteq \mathcal{X}$, let $\mu_M(S) \equiv \int_S m_\lambda d\mu_F(\lambda)$ denote the fraction of meetings that are with a misaligned trader whose contact rate is some $\lambda \in S$. It follows that the misalignment rate m is the equal to the Radon-Nikodym derivative $d\mu_M/d\mu_F$. Also let $M(\lambda) \equiv \mu_M([0, \lambda])$ denote the fraction of meetings that are with a misaligned trader whose contact rate is less than or equal to λ.

Using the fact that the surplus function is decreasing in λ, and again appealing to the Radon-Nikodym theorem, the inflow-outflow equation (9) reduces to

$$\left(r + \gamma + \frac{\lambda}{2}\big(\mu_F((\lambda, \bar{\lambda}]) + \mu_M([0, \lambda])\big)\right) \frac{d\mu_M(\lambda)}{d\mu_F(\lambda)} = \left(\gamma + \frac{\lambda}{2}\mu_M([0, \lambda))\right)\left(1 - \frac{d\mu_M(\lambda)}{d\mu_F(\lambda)}\right) \quad (30)$$

If $\mu_F(\{\lambda\}) > 0$, equation (30) is a quadratic equation for $d\mu_M(\lambda)/d\mu_F(\lambda) = \mu_M(\{\lambda\})/\mu_F(\{\lambda\})$. The smaller solution has $d\mu_M(\lambda) < 0$, which is inconsistent with the fact that this is equal to the misalignment rate m_λ, and hence must be positive. The larger solution, with $d\mu_M(\lambda)/d\mu_F(\lambda) \in [\underline{m}, \bar{m}] \subset (0, 1/2)$, determines $\mu_M(\{\lambda\})$. Alternatively, if $\mu_F(\{\lambda\}) = 0$, the solution is unique and still implies $d\mu_M(\lambda)/d\mu_F(\lambda) \in [\underline{m}, \bar{m}]$.

If F is differentiable at λ, equation (30) simplifies further:

$$\left(r+\gamma+\frac{\lambda}{2}(1-F(\lambda)+M(\lambda))\right)M'(\lambda)=\left(\gamma+\frac{\lambda}{2}M(\lambda)\right)(F'(\lambda)-M'(\lambda)), \tag{31}$$

an ordinary differential equation for M given F.

Step 2: Explicit solution for the surplus function Once we have recovered the misalignment measure μ_M and associated function M, we compute the surplus function. Recall that the surplus function solves equation (29), a contraction, which implies that it is the unique such solution. We look for a differentiable solution. Towards that end, rewrite equation (29) as

$$\frac{(4r+8\gamma+\lambda)s(\lambda)-4\Delta}{\lambda}=\int_{\mathcal{X}}\min\{s(\lambda'),s(\lambda)\}(1-2m_{\lambda'})d\mu_F(\lambda').$$

Since s is decreasing, we can rewrite this as

$$\frac{(4r+8\gamma+\lambda)s(\lambda)-4\Delta}{\lambda}=s(\lambda)(F(\lambda)-2M(\lambda))+\int_{(\lambda,\bar{\lambda}]}s(\lambda')(1-2m_{\lambda'})d\mu_F(\lambda'). \tag{32}$$

Evaluating at $\lambda=\bar{\lambda}$ and solving for $s(\bar{\lambda})$ gives

$$s(\bar{\lambda})=\frac{2\Delta}{2r+4\gamma+\bar{\lambda}M(\bar{\lambda})}. \tag{33}$$

Moreover, assuming s is differentiable, we can differentiate equation (32) to get

$$s'(\lambda)=\frac{4\left((r+2\gamma)s(\lambda)-\Delta\right)}{\lambda\big(4(r+2\gamma)+\lambda(1-F(\lambda)+2M(\lambda))\big)} \tag{34}$$

Integrate equation (34) using equation (33) as a boundary condition. By extending the definition of F and M to the positive real line through the convenient normalizations $F(\lambda)=1$ and $M(\lambda)=M(\bar{\lambda})$ for all $\lambda>\bar{\lambda}$, we find that the unique solution is

$$s(\lambda)=\frac{\Delta}{r+2\gamma}\left(1-e^{-\int_{\lambda}^{\infty}\phi_{\lambda'}d\lambda'}\right) \tag{35}$$

where

$$\phi_\lambda\equiv\frac{4(r+2\gamma)}{\lambda\Big(4(r+2\gamma)+\lambda(1-F(\lambda)+2M(\lambda))\Big)}. \tag{36}$$

Thus the surplus function is uniquely determined by the counterparty distribution F.

Step 3: Recovering the cost function Again taking advantage of the fact that the surplus function is non-negative and decreasing, we can rewrite equation (10) as

$$\pi_\lambda = \frac{\delta_1 - \gamma s(\lambda) + \frac{\lambda}{4}\int_{[0,\lambda]}(s(\lambda') - s(\lambda))d\mu_M(\lambda')}{r} - C(\lambda). \tag{37}$$

Part 3 of the definition of equilibrium imposes can then be viewed as a providing a lower bound on the cost function. Fix any $\bar{\pi}$. Then we must have

$$C(\lambda) \geq \frac{\delta_1 - \gamma s(\lambda) + \frac{\lambda}{4}\int_{[0,\lambda]}(s(\lambda') - s(\lambda))d\mu_M(\lambda')}{r} - \bar{\pi},$$

for all $\lambda \in \mathcal{X}$ with equality for all $\lambda \in \mathcal{Y}$, where $\mu_F(\mathcal{Y}) = 1$. Since for any μ_F, equations (30) and (35) uniquely determine μ_M and s, the lower bound is unique up to the additive constant $\bar{\pi}$. ■

Proof of Proposition 2. We proceed in three stages. First we show that $F(\lambda)$ is continuous on $[0, \lambda)$, then we show that it is absolutely continuous, and finally we show that a positive measure of traders choose $\lambda \in (0, \bar{\lambda})$.

Step 1: No Discontinuities in F Suppose there is a discontinuity in F at some point $\lambda_0 \in (0, \bar{\lambda})$, i.e. $\mu_F(\{\lambda_0\}) > 0$. We will find a contradiction. Start with the denominator of ϕ_λ in equation (36). From the definition of M and the bound $m_\lambda \leq \bar{m} < 1/2$ (inequality (28)), it follows that $1 - F(\lambda) + 2M(\lambda)$ is nonincreasing, and that it decreases discontinuously at λ_0. Thus, ϕ increases discontinuously at λ_0.

Differentiating equation (35), we have

$$s'(\lambda) = -\frac{\Delta\phi_\lambda e^{-\int_\lambda^\infty \phi_{\lambda'}d\lambda'}}{r + 2\gamma}. \tag{38}$$

Because ϕ_λ increases discontinuously at λ_0 we have that $s'(\lambda)$ decreases discontinuously (becomes more negative) at this point.

Next, differentiating equation (37) gives

$$\pi'_\lambda = \frac{-\left(\gamma + \frac{\lambda}{4}M(\lambda)\right)s'(\lambda) + \frac{1}{4}\int_{[0,\lambda]}\left(s(\lambda') - s(\lambda)\right)d\mu_M(\lambda')}{r} - C'(\lambda).$$

Because $s'(\lambda)$ decreases discontinuously at λ_0 and C is differentiable, π'_λ increases discontinuously at this point. That is, the profit function is locally convex and so λ_0 is not profit maximizing. But then part 3 of the definition of equilibrium implies that since λ_0 does not

maximize π_λ, $\mu_F(\{\lambda_0\}) = 0$, a contradiction. This proves $\mu_F(\{\lambda\}) = 0$ at any $\lambda \in (0, \bar{\lambda})$. Since F is right-continuous by definition, it is continuous on $[0, \bar{\lambda})$.

Step 2: Absolute continuity of F We again break this part into two stages. We first prove that $\lambda F(\lambda)$ is Lipschitz continuous. We then establish absolute continuity of F.

Step 2.1: $\lambda F(\lambda)$ Lipschitz continuous We proceed by contradiction. We will use part 1 here, i.e. we specifically assume F is continuous on $[0, \bar{\lambda})$, but $\lambda F(\lambda)$ is not Lipschitz continuous. It follows that for all $K \geq 0$, there exists $\lambda_1, \lambda_2 \in [0, \bar{\lambda})$ with $\lambda_2 > \lambda_1$ subject to

$$\frac{\lambda_2 F(\lambda_2) - \lambda_1 F(\lambda_1)}{\lambda_2 - \lambda_1} > K.$$

Part 3 of the definition of equilibrium implies F is constant at points that do not maximize π_λ. It follows that for all $K \geq 0$, there are profit maximizing $\lambda_1, \lambda_2 \in [0, \bar{\lambda})$ with $\lambda_2 > \lambda_1$ subject to

$$\frac{\lambda_2 F(\lambda_2) - \lambda_1 F(\lambda_1)}{\lambda_2 - \lambda_1} > K. \tag{39}$$

Next, since $m_\lambda < \bar{m} < \frac{1}{2}$ (inequality (28)), $F(\lambda) \geq 2M(\lambda)$. Let $Z(\lambda) \equiv -\lambda(1 - F(\lambda) + 2M(\lambda))$. Then,

$$\begin{aligned}
\frac{Z(\lambda_2) - Z(\lambda_1)}{\lambda_2 - \lambda_1} &= \frac{\lambda_2 F(\lambda_2) - \lambda_1 F(\lambda_1) - 2\lambda_1(M(\lambda_2) - M(\lambda_1))}{\lambda_2 - \lambda_1} - \big(1 + 2M(\lambda_2)\big) \\
&> \frac{F(\lambda_2)\,(\lambda_2 - 2\lambda_1 \bar{m}) - \lambda_1 F(\lambda_1)(1 - 2\bar{m})}{\lambda_2 - \lambda_1} - 2 \\
&\geq \frac{\lambda_2 F(\lambda_2) - \lambda_1 F(\lambda_1)}{\lambda_2 - \lambda_1}(1 - 2\bar{m}) - 2.
\end{aligned}$$

The first line is algebra. The second line uses inequality (28), which implies both that $M(\lambda_2) < \frac{1}{2}$ and for all $\lambda_2 > \lambda_1$,

$$M(\lambda_2) - M(\lambda_1) = \int_{[\lambda_1, \lambda_2]} m_\lambda d\mu_F(\lambda') \leq \bar{m}(F(\lambda_2) - F(\lambda_1)).$$

The third line uses $F(\lambda_2)\bar{m} \geq 0$. Since $\bar{m} < 1/2$, if λF is not Lipschitz, Z is also not Lipschitz.

Next, let $\varrho \equiv 4r + 8\gamma$. Since $Z(\lambda) \geq -\lambda$, we have

$$\phi_\lambda = \frac{\varrho}{\lambda(\varrho - Z(\lambda))} \geq \frac{\varrho}{\lambda(\varrho + \lambda)}.$$

Integrating and exponentiating gives

$$e^{\int_{\lambda_1}^{\lambda_2}\phi_\lambda d\lambda} \geq \frac{\lambda_2(\varrho+\lambda_1)}{\lambda_1(\varrho+\lambda_2)}. \tag{40}$$

This implies

$$\begin{aligned}\phi_{\lambda_1} - \phi_{\lambda_2} e^{\int_{\lambda_1}^{\lambda_2}\phi_\lambda d\lambda} &\leq \frac{\varrho}{\lambda_1(\varrho - Z(\lambda_1))} - \frac{\varrho}{\lambda_2(\varrho - Z(\lambda_2))}\frac{\lambda_2(\varrho+\lambda_1)}{\lambda_1(\varrho+\lambda_2)} \\ &= \frac{\varrho(\lambda_2-\lambda_1)}{\lambda_1(\varrho - Z(\lambda_1))(\varrho - Z(\lambda_2))}\left(\frac{\varrho - Z(\lambda_1)}{\varrho+\lambda_2} - \frac{Z(\lambda_2)-Z(\lambda_1)}{\lambda_2-\lambda_1}\right). \end{aligned} \tag{41}$$

The inequality uses $\phi_\lambda = \frac{\varrho}{\lambda(\varrho - Z(\lambda))}$ and inequality (40). The equality is algebra. Note that Z is increasing. If it is not Lipschitz, then we can find values of $\lambda_2 > \lambda_1$ such that this expression is negative.

Now, let $\vartheta \equiv \frac{\varrho}{\inf_{\lambda\in[0,\bar{\lambda}]}(-Z(\lambda)/\lambda)} > 0$. Then

$$\phi_\lambda = \frac{\varrho}{\lambda(\varrho - Z(\lambda))} \leq \frac{\vartheta}{\lambda(\vartheta+\lambda)}.$$

Again, we can integrate and exponentiate this to get

$$e^{-\int_{\lambda_1}^{\infty}\phi_\lambda d\lambda} \geq \frac{\lambda_1}{\vartheta+\lambda_1}. \tag{42}$$

Now take any values of $\lambda_1 < \lambda_2$ such that $\frac{\varrho - Z(\lambda_1)}{\varrho+\lambda_2} < \frac{Z(\lambda_2)-Z(\lambda_1)}{\lambda_2-\lambda_1}$. Then

$$\begin{aligned}\frac{s'(\lambda_2)-s'(\lambda_1)}{\lambda_2-\lambda_1} &= \frac{4\Delta}{\varrho(\lambda_2-\lambda_1)} e^{-\int_{\lambda_1}^{\infty}\phi_\lambda d\lambda}\left(\phi_{\lambda_1} - \phi_{\lambda_2} e^{\int_{\lambda_1}^{\lambda_2}\phi_\lambda d\lambda}\right) \\ &\leq \frac{4\Delta}{(\vartheta+\lambda_1)(\varrho - Z(\lambda_1))(\varrho - Z(\lambda_2))}\left(\frac{\varrho - Z(\lambda_1)}{\varrho+\lambda_2} - \frac{Z(\lambda_2)-Z(\lambda_1)}{\lambda_2-\lambda_1}\right).\end{aligned}$$

The inequality uses (41) and (42). Since all the other terms are positive and bounded away from both zero and infinity, it follows that if Z is not Lipschitz, s' is also not Lipschitz, and in particular $\frac{s'(\lambda_2)-s'(\lambda_1)}{\lambda_2-\lambda_1}$ is unbounded below.

Now consider part 3 of the definition of equilibrium. The choice of λ must maximize π_λ in (37) and so for both λ_1 and λ_2 to be optimal, we must have that $\pi'_{\lambda_1} = \pi'_{\lambda_2} = 0$, which

implies

$$r\big(C'(\lambda_2) - C'(\lambda_1)\big) = -\gamma\big(s'(\lambda_2) - s'(\lambda_1)\big) - \frac{1}{4}\big(\lambda_2 M(\lambda_2)s'(\lambda_2) - \lambda_1 M(\lambda_1)s'(\lambda_1)\big) + \frac{1}{4}\int_{[0,\lambda_2]}(s(\lambda') - s(\lambda_2))d\mu_M(\lambda') - \frac{1}{4}\int_{[0,\lambda_1]}(s(\lambda') - s(\lambda_1))d\mu_M(\lambda').$$

Since $\lambda_2 > \lambda_1$ and $s(\lambda)$ is decreasing, both

$$-\frac{1}{4}(\lambda_2 M(\lambda_2) - \lambda_1 M(\lambda_1))s'(\lambda_2) \geq 0$$

and

$$\frac{1}{4}\int_{[0,\lambda_2]}(s(\lambda') - s(\lambda_2))d\mu_M(\lambda') - \frac{1}{4}\int_{[0,\lambda_1]}(s(\lambda') - s(\lambda_1))d\mu_M(\lambda') \geq 0.$$

Combining with the previous equation, this implies

$$\frac{4r}{4\gamma + \lambda_1 M(\lambda_1)}\big(C'(\lambda_2) - C'(\lambda_1)\big) \geq -\big(s'(\lambda_2) - s'(\lambda_1)\big).$$

Since $\frac{s'(\lambda_2)-s'(\lambda_1)}{\lambda_2-\lambda_1}$ is unbounded below, this requires that $\frac{C'(\lambda_2)-C'(\lambda_1)}{\lambda_2-\lambda_1}$ is unbounded above, i.e. marginal cost is not Lipschitz, a contradiction.

Step 2.2: $F(\lambda)$ absolutely continuous We have thus far established that F is nonnegative, continuous, and nondecreasing on $[0, \bar{\lambda})$ and $\lambda F(\lambda)$ is Lipschitz on $[0, \bar{\lambda})$. The final part of the proof shows that it follows that F is absolutely continuous on $[0, \bar{\lambda})$.

To proceed, take any $\varepsilon > 0$. Since F is continuous, there exists a $\delta \in (0, \bar{\lambda})$ such that $F(\delta) - F(0) \leq \varepsilon/2$. Moreover, since F is nondecreasing, for any constant K_1 and finite sequence of pairwise disjoint subintervals $(\lambda_{1,k}, \lambda_{2,k}) \subset [0, \delta]$, $k \in \{1, \ldots, K_1\}$,

$$\sum_{k=1}^{K_1} |F(\lambda_{2,k}) - F(\lambda_{1,k})| = \sum_{k=1}^{K_1}(F(\lambda_{2,k}) - F(\lambda_{1,k})) \leq F(\delta) - F(0) \leq \frac{\varepsilon}{2}.$$

Next, let L denote the Lipschitz constant for $\lambda F(\lambda)$. We have that for any $\lambda_2 > \lambda_1 \geq \delta$,

$$(\lambda_2 - \lambda_1)L \geq \lambda_2 F(\lambda_2) - \lambda_1 F(\lambda_1) \geq \lambda_1\big(F(\lambda_2) - F(\lambda_1)\big) \geq \delta\big(F(\lambda_2) - F(\lambda_1)\big).$$

The first inequality is the definition of Lipschitz continuity for a nondecreasing function. The second uses $\lambda_2 F(\lambda_2) \geq \lambda_1 F(\lambda_2)$, which holds since F is nonnegative. The third uses $\lambda_1 \geq \delta$. Since F is nondecreasing, this proves that $F(\lambda)$ is Lipschitz with constant L/δ on $[\delta, \bar{\lambda})$.

Next, $F(\lambda)$ Lipschitz with constant L/δ implies that for any constant K and finite sequence of pairwise disjoint subintervals $(\lambda_{1,k}, \lambda_{2,k}) \subset [\delta, \bar{\lambda}]$, $k \in \{K_1 + 1, \ldots, K\}$, if

$$\sum_{k=K_1+1}^{K} (\lambda_{2,k} - \lambda_{1,k}) \leq \frac{\varepsilon\delta}{2L},$$

then

$$\sum_{k=K_1+1}^{K} |F(\lambda_{2,k}) - F(\lambda_{1,k})| = \sum_{k=K_1+1}^{K} (F(\lambda_{2,k}) - F(\lambda_{1,k})) \leq \frac{\varepsilon}{2}.$$

Now we put these pieces together. Take any finite sequence of pairwise disjoint subintervals $(\lambda_{1,k}, \lambda_{2,k}) \subset [0, \bar{\lambda}]$ with

$$\sum_{k} (\lambda_{2,k} - \lambda_{1,k}) \leq \min\left\{\delta, \frac{\varepsilon\delta}{2L}\right\}.$$

Without loss of generality, order the intervals so $\lambda_{2,k} \leq \lambda_{1,k+1}$. If $\delta \in (\lambda_{1,k}, \lambda_{2,k})$ for some k, break this into two separate intervals at this threshold. Let K denote the resulting number of subintervals and let $K_1 \in \{0, 1, \ldots, K\}$ satisfy $\lambda_{2,K_1} \leq \delta \leq \lambda_{1,K_1+1}$, with $\lambda_{2,0} = 0$ and $\lambda_{1,K+1} = \bar{\lambda}$. By construction, we have that $\bigcup_{k=1}^{K_1} (\lambda_{1,k}, \lambda_{2,k}) \subset [0, \delta]$ and $\sum_{k=K_1+1}^{K} (\lambda_{2,k} - \lambda_{1,k}) \leq \frac{\varepsilon\delta}{2L}$. Then the arguments above ensure $\sum_{k=1}^{K_1} |F(\lambda_{2,k}) - F(\lambda_{1,k})| \leq \frac{\varepsilon}{2}$ and $\sum_{k=K_1+1}^{K} |F(\lambda_{2,k}) - F(\lambda_{1,k})| \leq \frac{\varepsilon}{2}$. Adding these together gives

$$\sum_{k} |F(\lambda_{2,k}) - F(\lambda_{1,k})| \leq \varepsilon,$$

which proves F is absolutely continuous.

Step 3: Positive measure of traders in the interval $(0, \bar{\lambda})$ We again break this part into two stages. We first use the Lipschitz continuity of $F(\lambda)$ to derive some useful preliminaries. Then we establish the claim by contradiction.

Step 3.1: Implication of absolute continuity of $F(\lambda)$ Absolute continuity of F implies F is almost everywhere differentiable and $F(\lambda) = F(0) + \int_0^\lambda F'(\lambda')d\lambda'$. From equation (30), M inherits the same properties and so $M(\lambda) = M(0) + \int_0^\lambda M'(\lambda')d\lambda'$.

Differentiate equation (34) to get an expression for the first and second derivatives of the surplus function:

$$s''(\lambda) = -\frac{2(1 - F(\lambda) + 2M(\lambda)) - \lambda(F'(\lambda) - 2M'(\lambda))}{4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))} s'(\lambda). \tag{43}$$

Equation (43) holds at points where F and M are differentiable. Let $R(\lambda) \equiv \pi_\lambda + C(\lambda)$ denote the revenue of a newborn trader. Use equation (37) and then twice differentiate to get

$$R''(\lambda) = \frac{-(4\gamma + \lambda M(\lambda))s''(\lambda) - (2M(\lambda) + \lambda M'(\lambda))s'(\lambda)}{4r}. \tag{44}$$

Step 3.2: Positive measure of traders with interior contact rates To prove that a positive measure of trader choose $\lambda \in (0, \bar{\lambda})$, we proceed by contradiction. First, suppose there is an equilibrium in which a fraction $\psi \in (0, 1)$ of contacts are with traders with $\lambda = 0$, while the remaining contacts are with traders with $\lambda = \bar{\lambda}$. Such an equilibrium requires that 0 and $\bar{\lambda}$ are both optimal contact rates, $\pi_0 = \pi_{\bar{\lambda}} \geq \pi_\lambda$ for all $\lambda \in (0, \lambda)$. We prove that convexity of the cost function imposes a lower bound on ψ. In such an equilibrium, $F(\lambda) = \psi$ and $M(\lambda) = \frac{\gamma\psi}{r+2\gamma}$ for all $\lambda \in [0, \bar{\lambda})$; the constant $\frac{\gamma}{r+2\gamma}$ follows from evaluating equation (30) at $\lambda = 0$. Substitute this into equations (34) and (43), and then (44) to get

$$R''(\lambda) = \frac{4\gamma(r+\gamma)\big(\bar{\psi} - \psi\big)s'(\lambda)}{r\big(4(r+2\gamma)^2 + \lambda(2\gamma + r(1-\psi))\big)}, \tag{45}$$

where $\bar{\psi} \equiv \frac{r+2\gamma}{2(r+\gamma)}$. Since $s'(\lambda) < 0$, $\psi < \bar{\psi}$ implies $R''(\lambda) < 0$. Along with weak convexity of the cost function, this implies that the profit function is strictly concave, contradicting $\pi_0 = \pi_{\bar{\lambda}} \geq \pi_\lambda$ for all $\lambda \in (0, \lambda)$. Thus an equilibrium with a two point distribution at 0 and $\bar{\lambda}$ must have $F(0) = \psi \geq \bar{\psi}$.

Now assume $F(\lambda) = \psi$ for all $\lambda < \bar{\lambda}$, where $1 \geq \psi \geq \bar{\psi}$. This covers both the case of a two point distribution and also a degenerate distribution at 0. We use equation (30) to find $M(\bar{\lambda})$ when $F(\lambda) = \psi$ and $M(\lambda) = \frac{\gamma\psi}{r+2\gamma}$ for $\lambda \in [0, \bar{\lambda})$:

$$\left(r + \gamma + \frac{\bar{\lambda}}{2}M(\bar{\lambda})\right)(M(\bar{\lambda}) - M(0)) = \left(\gamma + \frac{\bar{\lambda}}{2}M(0)\right)\left(1 - F(0) - M(\bar{\lambda}) + M(0)\right).$$

This is a quadratic equation where the larger root is valid. We can verify algebraically that for any $1 \geq \psi \geq \bar{\psi}$,

$$M(\bar{\lambda}) \geq \frac{4\gamma(r+2\gamma)}{4(r+2\gamma)^2 + r\bar{\lambda}(1-\psi)}$$

or equivalently

$$\frac{\Delta\gamma\Big((r+2\gamma)\big(4(r+2\gamma) + \bar{\lambda}\big) - r\bar{\lambda}\psi\Big)M(\bar{\lambda})}{4r(r+2\gamma)^3\big(2(r+2\gamma) + \bar{\lambda}M(\bar{\lambda})\big)} \geq \frac{\Delta\gamma^2}{2r(r+2\gamma)^3} > C'(0), \tag{46}$$

where the second inequality is the assumed upper bound on the marginal cost function at 0.

We use this inequality in the next step.

Still assume $F(\lambda) = \psi$ for all $\lambda < \bar{\lambda}$, where $1 \geq \psi \geq \bar{\psi}$. Using $F(\lambda) = \psi$ and $M(\lambda) = \frac{\gamma\psi}{r+2\gamma}$ for $\lambda \in [0, \bar{\lambda})$, as well as $F(\lambda) = 1$ and $M(\lambda) = M(\bar{\lambda})$ for $\lambda \geq \bar{\lambda}$, we can explicitly solve the integral in equation (35) to get that for $\lambda \in [0, \bar{\lambda}]$,

$$s(\lambda) = \frac{2\Delta\Big(4(r+2\gamma)^2 + \lambda\big(2\gamma + r(1-\psi)\big) - 2(r+2\gamma)(\lambda - \bar{\lambda})M(\bar{\lambda})\Big)}{\Big(4(r+2\gamma)^2 + \lambda\big(2\gamma + r(1-\psi)\big)\Big)\big(2(r+2\gamma) + \bar{\lambda}M(\bar{\lambda})\big)}.$$

Differentiating this and evaluating at $\lambda = 0$ gives us

$$s'(0) = \frac{-\Delta\Big((r+2\gamma)\big(4(r+2\gamma) + \bar{\lambda}\big) - r\bar{\lambda}\psi\Big)M(\bar{\lambda})}{4(r+2\gamma)^3\big(2(r+2\gamma) + \bar{\lambda}M(\bar{\lambda})\big)}.$$

Equation (48) implies $\pi_0' = -\gamma s'(0)/r - C'(0)$, so equation (46) implies $\pi_0' > 0$. This means that choosing a slightly positive contact rate yields strictly higher profits, contradicting the optimality of $\lambda = 0$. This proves that there is no equilibrium with $\psi > 0$.

In the final step, we suppose there is an equilibrium in which all traders set $\lambda = \bar{\lambda}$. Then equation (30) implies

$$M(\bar{\lambda}) = \frac{-(r+2\gamma) + \sqrt{(r+2\gamma)^2 + 2\gamma\bar{\lambda}}}{\bar{\lambda}}. \tag{47}$$

Moreover, we can again explicitly solve the integral in equation (35) to compute $s(\lambda)$ for $\lambda < \bar{\lambda}$. Since equation (37) implies $\pi_\lambda = \frac{\delta_1 - \gamma s(\lambda)}{r} - C(\lambda)$ for $\lambda < \bar{\lambda}$, we can then compute the left derivative of π at $\bar{\lambda}$:

$$\lim_{h \nearrow 0} \frac{\pi_{\bar{\lambda}} - \pi_{\bar{\lambda}-h}}{h} = \frac{4\gamma\Delta\left(\sqrt{(r+2\gamma)^2 + 2\gamma\bar{\lambda}} - (r+2\gamma)\right)}{r\bar{\lambda}(4(r+2\gamma) + \bar{\lambda})\left(\sqrt{(r+2\gamma)^2 + 2\gamma\bar{\lambda}} + (r+2\gamma)\right)} - C'(\bar{\lambda}).$$

It is straightforward to prove that

$$\frac{4\gamma\Delta\left(\sqrt{(r+2\gamma)^2 + 2\gamma\bar{\lambda}} - (r+2\gamma)\right)}{r\bar{\lambda}(4(r+2\gamma) + \bar{\lambda})\left(\sqrt{(r+2\gamma)^2 + 2\gamma\bar{\lambda}} + (r+2\gamma)\right)} < \frac{4\gamma\Delta}{r\bar{\lambda}^2}$$

for any finite $\bar{\lambda}$, and thus

$$\lim_{h \nearrow 0} \frac{\pi_{\bar{\lambda}} - \pi_{\bar{\lambda}-h}}{h} < \frac{4\gamma\Delta}{r\bar{\lambda}^2} - C'(\bar{\lambda}) \leq 0,$$

where the second inequality is the assumed lower bound on the marginal cost function at $\bar{\lambda}$. It follows that the profit from choosing a contact rate slightly below $\bar{\lambda}$ exceeds the profit at $\bar{\lambda}$, contradicting the possibility of an equilibrium where everyone sets $\lambda = \bar{\lambda}$. ■

Representation of Equilibrium as an ODE System We derive a first order differential equation system in (F, M, s) to characterize equilibrium when the cost function is twice continuously differentiable.

Absolute continuity of F implies F is almost everywhere differentiable and $F(\lambda) = F(0) + \int_0^\lambda F'(\lambda')d\lambda'$. From equation (30), M inherits the same properties and so $M(\lambda) = M(0) + \int_0^\lambda M'(\lambda')d\lambda'$. In particular, equation (31) gives one linear relationship between $F'(\lambda)$ and $M'(\lambda)$ at points where both exist, based on the balance of flows.

Next, let $\mathcal{A} \subseteq \mathcal{X}$ denote the support of F, the smallest closed set such that $\mu_F(\mathcal{A}) = 1$. Unless $\mathcal{A} \subseteq \{0, \bar{\lambda}\}$, absolute continuity of F on $(0, \bar{\lambda})$ implies that $\mathcal{A}$ must have a nonempty interior, denoted $\mathcal{A}^o$. At $\lambda \in \mathcal{A}^o$, part 3 of the definition of equilibrium states that $\pi_\lambda = \bar{\pi}$ and so in particular the first order condition $\pi'_\lambda = 0$ holds if π is differentiable. Using equation (37) this implies:

$$\pi'_\lambda = \frac{-(4\gamma + \lambda M(\lambda))s'(\lambda) + \int_0^\lambda (s(\lambda') - s(\lambda))M'(\lambda')d\lambda'}{4r} - C'(\lambda) = 0. \tag{48}$$

Moreover, since the interior points $\mathcal{A}^o$ are not isolated, $\pi'_{\lambda'} = 0$ at points λ' in a neighborhood of λ. This in turn implies $\pi''_\lambda = 0$. Using equation (37) and twice differentiability of C, this implies that at any $\lambda \in \mathcal{A}^o$ where F and M are differentiable,

$$\pi''_\lambda = \frac{-(4\gamma + \lambda M(\lambda))s''(\lambda) - (2M(\lambda) + \lambda M'(\lambda))s'(\lambda)}{4r} - C''(\lambda) = 0. \tag{49}$$

Use equations (34) and (43) to eliminate $s'(\lambda)$ and $s''(\lambda)$ from equation (49). This gives us the second linear relationship between $F'(\lambda)$ and $M'(\lambda)$. We can solve this and equation (31) explicitly for $F'(\lambda)$ and $M'(\lambda)$ as functions of $F(\lambda)$, $M(\lambda)$, and $s(\lambda)$:

$$F'(\lambda) = \frac{\big(2r + 4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda))\big)(8\gamma(1 - F(\lambda)) - 8rM(\lambda) + \zeta(\lambda))}{2\lambda\big(\gamma(8r + 8\gamma + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda)))\big)}, \tag{50}$$

$$M'(\lambda) = \frac{\big(2\gamma + \lambda M(\lambda)\big)(8\gamma(1 - F(\lambda)) - 8rM(\lambda) + \zeta(\lambda))}{2\lambda\big(\gamma(8r + 8\gamma + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda)))\big)}, \tag{51}$$

where

$$\zeta(\lambda) \equiv \frac{r\lambda\big(4(r+2\gamma)+\lambda(1-F(\lambda)+2M(\lambda))\big)^2 C''(\lambda)}{\Delta-(r+2\gamma)s(\lambda)}. \tag{52}$$

These two equations, together with equation (34), are an ordinary differential equation system in (F, M, s) on the set $\mathcal{A}^o$ that must hold in any equilibrium.

To solve this ordinary differential equation system, we must find the appropriate boundary condition. For this, we focus on the case where the support of $\mathcal{A}$ is a convex set with $F(0) = 0$, but the logic can handle more complicated cases. First guess a lower bound of the support, $\underline{\lambda}$, so $F(\underline{\lambda}) = 0$. By the definition of M, $M(\underline{\lambda}) = 0$ as well. Then equation (48) implies

$$\frac{-\gamma s'(\underline{\lambda})}{r} = C'(\underline{\lambda}). \tag{53}$$

This gives us a third terminal condition. We then solve this initial value problem for F, M, and s. We stop either when we find a $\hat{\lambda} \leq \bar{\lambda}$ such that $F(\hat{\lambda}) = 1$ or when we hit the upper bound $\bar{\lambda}$. Note that in the former case, the value of $\bar{\lambda}$ does not affect F, M, or s for $\lambda \leq \hat{\lambda}$. For expositional convenience, we therefore redefine $\bar{\lambda} = \hat{\lambda}$ in this case. In the latter case, we impose $F(\bar{\lambda}) = 1$, so there is a mass point in F and hence M at $\bar{\lambda}$. We then use equation (30) to find $M(\bar{\lambda})$.

We still need to validate the guess of $\underline{\lambda}$. To do so, we compare the surplus at the upper bound, $s(\bar{\lambda})$, which comes from the solution to the initial value problem, with the value implied directly in equation (33). If the two methods of computing the surplus function give us the same answer, we have found an equilibrium, while otherwise we need to change the initial guess $\underline{\lambda}$.

C.4 Characterization with a Linear Cost Function

Proof of Proposition 3. We divide the proof into three pieces, depending on whether $c \geq \bar{c}$ (high cost), $\bar{c} > c > \underline{c}$ (intermediate cost), or $c \leq \underline{c}$ (low cost). We characterize equilibrium in each region of the parameter space, starting with high cost, then low cost, and finally intermediate cost, since that case builds on the other two.

High Cost. We start by proving that there is a threshold $\bar{c}$ such that if $c \geq \bar{c}$, there is an equilibrium with $\Lambda = 0$. This proof is constructive.

First we look for an equilibrium in which the support of the counterparty distribution is 0 and $\bar{\lambda}$ with weights $\psi \in (0, 1)$ and $1 - \psi$, respectively. For this to be an equilibrium, we require $\pi_0 = \pi_{\bar{\lambda}} \geq \pi_\lambda$ for all $\lambda \in [0, \bar{\lambda}]$. A linear cost implies that the second derivative of the profit function and revenue function are equal. We can then replicate the derivation of

equation (45) to get

$$\pi''_\lambda = \frac{4\gamma(r+\gamma)\big(\bar{\psi}-\psi\big)s'(\lambda)}{r\big(4(r+2\gamma)^2+\lambda(2\gamma+r(1-\psi))\big)},$$

where $\bar{\psi} \equiv \frac{r+2\gamma}{2(r+\gamma)}$. Since $s'(\lambda) < 0$, $\psi < \bar{\psi}$ implies $\pi''_\lambda < 0$, i.e. the profit function is globally concave. This is inconsistent with $\pi_0 = \pi_{\bar{\lambda}} \geq \pi_\lambda$ for all $\lambda \in [0, \bar{\lambda}]$. But if $\psi \geq \bar{\psi}$, the profit function is globally convex. We just need to find the value of c that makes $\pi_0 = \pi_{\bar{\lambda}}$.

Equation (37) implies

$$\pi_{\bar{\lambda}} - \pi_0 = \frac{\big(4\gamma + \bar{\lambda}M(0)\big)\big(s(0) - s(\bar{\lambda})\big)}{4r} - c\bar{\lambda}. \tag{54}$$

Eliminate $s(0) - s(\bar{\lambda})$ using equation (35):

$$\pi_{\bar{\lambda}} - \pi_0 = \left(\frac{\big(4\gamma + \bar{\lambda}M(0)\big)M(\bar{\lambda})\Delta}{4r(r+2\gamma)\big(2(r+2\gamma) + \bar{\lambda}M(\bar{\lambda})\big)} - c\right)\bar{\lambda}. \tag{55}$$

For a given $F(0) = \psi \geq \bar{\psi}$, we have $M(0) = \frac{\gamma\psi}{r+2\gamma}$. We then pin down $M(\bar{\lambda}) = M(0) + \mu_M(\{\bar{\lambda}\})$ by finding the unique solution to equation (30) with $\mu_M(\{\bar{\lambda}\}) \in (0, (1-\psi)/2)$. Notably this solution is continuous in ψ. By setting the right hand side of equation (55) to zero, we then find the cost c that gives us an equilibrium with a given value of ψ. Let $\bar{c}$ be the solution to this when $\psi = \bar{\psi}$. Let $\bar{\bar{c}}$ be the solution to this when $\psi = 1$ and so $M(0) = M(\bar{\lambda}) = \frac{\gamma}{r+2\gamma}$; in general we cannot order $\bar{c}$ and $\bar{\bar{c}}$. The intermediate value theorem implies that for any $c \in \big[\min\{\bar{c}, \bar{\bar{c}}\}, \max\{\bar{c}, \bar{\bar{c}}\}\big]$, there exists an $F(0) \in \big[\bar{\psi}, 1\big]$ such that there is an equilibrium in which the support of the counterparty distribution is 0 and $\bar{\lambda}$ with weights $F(0)$ and $1 - F(0)$, respectively.

Now consider $c \geq \bar{\bar{c}}$. The previous paragraph proved that there is an equilibrium with $F(0) = 1$ when $c = \bar{\bar{c}}$. Raising the cost further does not change the equilibrium, i.e. we still have $F(0) = 1$, but now $\pi_0 > \pi_{\bar{\lambda}}$ and π is globally convex.

Finally, given F and M, it is straightforward to find the misalignment rate and surplus function from equations (30) and (35). Thus regardless of whether $\bar{c} \gtrless \bar{\bar{c}}$, we have constructed an equilibrium for any $c \geq \bar{c}$.

Low Cost. We next prove that there is a threshold $\underline{c}$ such that if $c \leq \underline{c}$, there is an equilibrium with $F(\lambda) = 0$ for $\lambda \in [0, \bar{\lambda})$ and $\mu_F(\{\bar{\lambda}\}) = 1$. This proof is again constructive.

In such an equilibrium, equation (17) implies $\Lambda = \bar{\lambda}$. $F(\lambda) = 0$ implies $M(\lambda) = 0$ for all $\lambda \in [0, \bar{\lambda})$. We uniquely recover $M(\bar{\lambda})$ from equation (30). Then equation (37) implies that

for any $\lambda < \bar{\lambda}$,

$$\pi_{\bar{\lambda}} - \pi_{\lambda} = \frac{\gamma(s(\lambda) - s(\bar{\lambda}))}{r} - c(\bar{\lambda} - \lambda).$$

Using equation (35), we can solve explicitly for $s(\lambda) - s(\bar{\lambda})$. This gives

$$\pi_{\bar{\lambda}} - \pi_{\lambda} = \left(\frac{4\gamma \Delta M(\bar{\lambda})}{r(4r + 8\gamma + \lambda)(2r + 4\gamma + \bar{\lambda} M(\bar{\lambda}))} - c \right) (\bar{\lambda} - \lambda).$$

Thus $\pi_{\bar{\lambda}} > \pi_{\lambda}$ for all $\lambda < \bar{\lambda}$ if and only if

$$c \leq \frac{4\gamma \Delta M(\bar{\lambda})}{r(4r + 8\gamma + \bar{\lambda})(2r + 4\gamma + \bar{\lambda} M(\bar{\lambda}))} \equiv \underline{c}, \tag{56}$$

since this implies $c < \frac{4\gamma \Delta M(\bar{\lambda})}{r(4r+8\gamma+\lambda)(2r+4\gamma+\bar{\lambda} M(\bar{\lambda}))}$ for all $\lambda < \bar{\lambda}$. We can again find the misalignment rate and surplus function from equations (30) and (35). Thus we have constructed an equilibrium for any $c \leq \underline{c}$.

Intermediate Cost. Finally, we turn to the case where $c \in (\underline{c}, \bar{c})$. We first show how to construct an equilibrium as the solution to an initial value problem. In the second step, we prove that the initial value problem determines $F(\lambda)$ and $M(\lambda)$ and hence the right hand side of equation (58) as continuous functions of $\underline{\lambda}$. The third and fourth steps characterize the limiting behavior of F and M, and hence the right hand side of equation (58), when $\underline{\lambda} \to 0$ and when $\underline{\lambda} \to \bar{\lambda}$. Finally, we use the intermediate value theorem to prove that for any $c \in (\underline{c}, \bar{c})$, there is a $\underline{\lambda} \in (0, \bar{\lambda})$ such that equilibrium is described by the F and M that solve the initial value problem (57).

Step 1: The Initial Value Problem Since $C''(\lambda) = 0$ and $s(\lambda) < \Delta/(r + 2\gamma)$, equation (52) implies $\zeta(\lambda) = 0$. Then equations (50) and (51) reduce to a two variable initial value problem:

$$F'(\lambda) = x_F(\lambda, F(\lambda), M(\lambda)), \quad M'(\lambda) = x_M(\lambda, F(\lambda), M(\lambda)), \quad \text{and } F(\underline{\lambda}) = M(\underline{\lambda}) = 0, \tag{57}$$

where

$$x_F(\lambda, F, M) \equiv \frac{4\big(2r + 4\gamma + \lambda(1 - F + 2M)\big)(\gamma(1 - F) - rM)}{\lambda\big(\gamma(8r + 8\gamma + 3\lambda(1 - F)) + \lambda M(3r + 6\gamma + \lambda(1 - F + M))\big)},$$
$$x_M(\lambda, F, M) \equiv \frac{4\big(2\gamma + \lambda M\big)(\gamma(1 - F) - rM)}{\lambda\big(\gamma(8r + 8\gamma + 3\lambda(1 - F)) + \lambda M(3r + 6\gamma + \lambda(1 - F + M))\big)}.$$

This is the differential equation system $(F', X') = X_1(\lambda, F, M)$ discussed in the text. If we knew $\underline{\lambda}$, we could easily solve this initial value problem. Moreover, we could verify that it is an equilibrium by checking whether $\pi'_{\underline{\lambda}} = 0$. Using equation (48) along with the expressions for $s(\lambda)$ in equation (35) and ϕ_λ in equation (36) implies that in equilibrium

$$\frac{c}{\Delta} = \frac{4\gamma}{\underline{\lambda} r \left(4(r+2\gamma) + \underline{\lambda}\right)} \exp\left(- \int_{\underline{\lambda}}^{\infty} \frac{4(r+2\gamma)}{\lambda\Big(4(r+2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))\Big)} d\lambda \right). \quad (58)$$

This is the equation $c = X_2(\underline{\lambda})$ discussed in the text. Note that X_2 depends on the path of F and M, but since those are functions $\bar{\lambda}$, r, and γ through equation (57), we suppress that dependence. Thus, for any lower bound $\underline{\lambda} \in (0, \bar{\lambda})$, we can use this algorithm to find the marginal cost such that this is an equilibrium.

We are interested in using this characterization in the other direction: For a given cost, we look for an equilibrium. To do this, fix a small, positive value of ε and let $\Omega \equiv (0, \bar{\lambda}) \times (-\varepsilon, 1) \times (-\varepsilon, \frac{1}{2})$. Our initial value problem (57) maps $(\lambda, F, M) \in \Omega$ into (F', M'). If ε is sufficiently small, x_F and x_M are positive for all $(\lambda, F, M) \in \Omega$.

It is straightforward to verify that the functions x_F and x_M are continuous on Ω, and hence they are locally Lipschitz continuous in F and M (Lemma 3.1 in Sideris, 2013). This implies that for any $\underline{\lambda} \in (0, \bar{\lambda})$, there exists a $\delta > 0$ such that our initial value problem has a unique solution when $\underline{\lambda} < \lambda < \underline{\lambda} + \delta$ (Theorems 3.2 and 3.3 in Sideris, 2013).

Next, note that for $(\lambda, F, M) \in \Omega$, x_F and x_M have the same sign as $\gamma(1-F(\lambda)) - rM(\lambda)$. Because of the initial condition $F(\underline{\lambda}) = M(\underline{\lambda}) = 0$, $\gamma(1-F(\underline{\lambda})) > rM(\underline{\lambda})$. Moreover, we claim that $\gamma(1 - F(\lambda)) > rM(\lambda)$ and hence $x_F(\lambda)$ and $x_M(\lambda)$ are both positive for all $\lambda \in (\underline{\lambda}, \bar{\lambda})$. To prove this, suppose to the contrary that there were a $\lambda > \underline{\lambda}$ with $\gamma(1 - F(\lambda)) = rM(\lambda)$. We could write an initial value problem with this boundary condition instead of the one at $\underline{\lambda}$. For the reasons in the previous paragraph, local Lipschitz continuity of x_M and x_F implies a unique solution to this problem, the constant solution. But this contradicts $\gamma(1 - F(\underline{\lambda})) > rM(\underline{\lambda})$.

Now $x_M(\lambda) > 0$ for all $\lambda \in (\underline{\lambda}, \bar{\lambda})$ implies $M(\lambda) > 0$ for all $\lambda \in (\underline{\lambda}, \bar{\lambda})$. Hence $\gamma(1-F(\lambda)) > rM(\lambda)$ implies $F(\lambda) < 1$ for all $\lambda \in (\underline{\lambda}, \bar{\lambda})$ as well.

Also note that for $(\lambda, F, M) \in \Omega$, $x_F > 2x_M > 0$ and hence in the solution to the initial value problem, $F(\lambda) \geq 2M(\lambda)$ for all $\lambda \geq \underline{\lambda}$. Then $F(\lambda) < 1$ implies $0 \leq M(\lambda) < \frac{1}{2}$ at all $\lambda \geq \underline{\lambda}$. It follows that the maximal existence interval for the initial value problem includes $[\underline{\lambda}, \bar{\lambda})$. We splice this solution together with two conditions: For $\lambda < \underline{\lambda}$, $F(\lambda) = M(\lambda) = 0$; and for $\lambda \geq \bar{\lambda}$, $F(\lambda) = 1$ and $M(\lambda) = \mu_M([0, \bar{\lambda})) + \mu_M(\{\bar{\lambda}\})$, where the first term is determined by the initial value problem and the second terms satisfies equation (30). This is a quadratic

equation for $\mu_M(\{\bar{\lambda}\})$. Only the larger root, which satisfies $\mu_M(\{\bar{\lambda}\}) \in (0, \mu_F(\{\bar{\lambda}\})/2)$, is valid.

Once we have computed F and M, we have $m_\lambda = d\mu_M(\lambda)/d\mu_F(\lambda)$ at $\lambda \geq \underline{\lambda}$. At smaller values of λ, equation (30) implies $m_\lambda = \frac{2\gamma}{2r+4\gamma+\lambda}$. Equation (35) gives us the surplus function. Thus we have determined the three functions in the definition of equilibrium. We now verify that this is an equilibrium.

The initial value problem captures two requirements of an equilibrium: the misalignment rate is in steady state for each $\lambda \in [\underline{\lambda}, \bar{\lambda})$; and the profit function is linear on $[\underline{\lambda}, \bar{\lambda})$, $\pi''_\lambda = 0$. Equilibrium imposes two other restrictions: the level of the cost must ensure that the profit function is not only linear in λ but constant; and profits must be weakly lower at other values of λ. We turn to those next.

Equation (58) captures the requirement that the profit function is constant for $\lambda \in [\underline{\lambda}, \bar{\lambda})$. It states that $\pi'_{\underline{\lambda}} = 0$ and hence pins down c for a given $\underline{\lambda}$. Since the solution to the initial value problem has $\pi''_\lambda = 0$ for all $\lambda \in [\underline{\lambda}, \bar{\lambda})$, it follows that $\pi'_\lambda = 0$ as well. Of course, we are interested in understanding how c determines $\underline{\lambda}$ and so still need to invert this requirement, i.e. to find $\underline{\lambda}$ given c. That is the purpose of remaining steps in this proof.

Turn now to profits at values of $\lambda < \underline{\lambda}$, where $F(\lambda) = M(\lambda) = 0$. Since F and M are continuous at $\underline{\lambda}$, s' is continuous as well (equation 34) and hence so is π'_λ (equation 48). On the other hand, F' and M' jump up discontinuously at $\underline{\lambda}$ and so π''_λ can be discontinuous at that point. In fact, plugging equation (43) into equation (49) and imposing $F(\lambda) = M(\lambda) = F'(\lambda) = M'(\lambda) = 0$ for $\lambda < \underline{\lambda}$ gives

$$\pi''_\lambda = \frac{2\gamma}{r(4r + 8\gamma + \lambda)} s'(\lambda) < 0.$$

Combining with $\pi'_{\underline{\lambda}} = 0$ gives $\pi'_\lambda > 0$ and hence $\pi_\lambda < \pi_{\underline{\lambda}}$ at all $\lambda < \underline{\lambda}$. This verifies the third part of the definition of equilibrium: profits are maximized on the support of F.

Step 2: Continuity of Solution to Initial Value Problem. Let $F(\lambda; \underline{\lambda})$ and $M(\lambda; \underline{\lambda})$ denote the unique solution to the initial value problem for a given $\underline{\lambda} \in (0, \bar{\lambda})$. Theorem 3.5 in Sideris (2013) implies that F and M are continuous in $\underline{\lambda}$ on the interval $(\underline{\lambda}, \bar{\lambda})$. It follows immediately that the right hand side of equation (58) is continuous in $\underline{\lambda}$ on the same interval.

Step 3: Limit as $\underline{\lambda} \to 0$. We characterize the solution to the initial value problem (57) when $\underline{\lambda}$ is small. We prove that for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\underline{\lambda} \in (0, \delta)$,

$$\frac{r+2\gamma}{2(r+\gamma)} + \varepsilon \geq F(\lambda; \underline{\lambda}) \geq \frac{r+2\gamma}{2(r+\gamma)} - \varepsilon, \tag{59}$$

$$\frac{\gamma}{2(r+\gamma)} \geq M(\lambda; \underline{\lambda}) \geq \frac{\gamma}{2(r+\gamma)} - \varepsilon \tag{60}$$

for all $\lambda \in (\varepsilon, \bar{\lambda})$. That is, both F and M converge to step functions, the same step function as applies in the case of $c = \bar{c}$.

First, let $Y(\lambda) \equiv \log\big(\gamma(1-F(\lambda)) - rM(\lambda)\big)$, suppressing dependence on $\underline{\lambda}$ in the remainder of this step for notational convenience. Differentiating this and using the initial value problem (57) to eliminate F' and M', we have

$$Y'(\lambda) \equiv \frac{-4\big(\gamma(4r+4\gamma+\lambda(1-F(\lambda))) + (r+2\gamma)\lambda M(\lambda)\big)}{\lambda\big(\gamma(8r+8\gamma+3\lambda(1-F(\lambda))) + \lambda M(\lambda)(3r+6\gamma+\lambda(1-F(\lambda)+M(\lambda)))\big)}. \tag{61}$$

Since $\lambda \geq 0$, $M(\lambda) \geq 0$, and $F(\lambda) \leq 1$,

$$\gamma(4r+4\gamma+\lambda(1-F(\lambda))) + (r+2\gamma)\lambda M(\lambda) \geq \gamma(4r+4\gamma).$$

And since $\lambda \leq \bar{\lambda}$, $F(\lambda) \geq 0$, and $M(\lambda) \leq \frac{1}{2}$,

$$\gamma(8r+8\gamma+3\lambda(1-F(\lambda))) + \lambda M(\lambda)(3r+6\gamma+\lambda(1-F(\lambda)+M(\lambda))) \\ \leq \gamma(8r+8\gamma+3\bar{\lambda}) + \frac{\bar{\lambda}}{2}\left(3r+6\gamma+\frac{3\bar{\lambda}}{2}\right).$$

Putting this together gives us

$$Y'(\lambda) \leq \frac{-4\gamma(4r+4\gamma)}{\lambda\left(\gamma(8r+8\gamma+3\bar{\lambda}) + \frac{\bar{\lambda}}{2}\left(3r+6\gamma+\frac{3\bar{\lambda}}{2}\right)\right)} \equiv \frac{-\kappa}{\lambda},$$

where κ is a positive constant. In addition, we have the terminal condition $Y(\underline{\lambda}) = \log\gamma$. This implies that if $\lambda > \underline{\lambda}$, $Y(\lambda)$ is smaller than the value of a curve with slope $-\kappa/\lambda$ through the point $Y(\underline{\lambda}) = \log\gamma$. That is, $Y(\lambda) \leq \log\gamma + \kappa\log(\underline{\lambda}/\lambda)$ for all $\lambda \in [\underline{\lambda}, \bar{\lambda})$. Equivalently, using the definition of Y,

$$F(\lambda) + \frac{r}{\gamma}M(\lambda) \geq 1 - (\underline{\lambda}/\lambda)^{\kappa} \tag{62}$$

for all $\lambda \in [\underline{\lambda}, \bar{\lambda})$. This implies that when $\underline{\lambda}/\lambda$ is sufficiently close to zero, $F(\lambda) + \frac{r}{\gamma}M(\lambda)$ must be close to 1.

We now use this to get a lower bound on F alone. Autarky gives an upper bound on the misalignment rate,

$$M(\lambda) \leq m_0 F(\lambda) = \frac{\gamma}{r+2\gamma} F(\lambda). \tag{63}$$

Combining inequalities (62) and (63) gives us

$$F(\lambda) \geq \frac{r+2\gamma}{2(r+\gamma)} \left(1 - (\underline{\lambda}/\lambda)^{\kappa}\right) \tag{64}$$

This is a lower bound on the contact rate distribution, close to $\frac{r+2\gamma}{2(r+\gamma)}$ whenever $\underline{\lambda}/\lambda$ is close to zero. Using this, it is straightforward to use this to establish the lower bound in inequality (59) through an appropriate choice of δ for each ε.

To find a lower bound on M, use equation (31) to get

$$\begin{aligned} M'(\lambda') &= \frac{2\gamma + \lambda' M(\lambda')}{2r + 4\gamma + \lambda'(1 - F(\lambda') + 2M(\lambda'))} F'(\lambda') \\ &\geq \frac{\gamma}{r + 2\gamma + \lambda'/2} F'(\lambda') \geq \frac{\gamma}{r + 2\gamma + \lambda/2} F'(\lambda') \end{aligned}$$

for all $\lambda' \leq \lambda$. The first inequality follows because the fraction in the first line is increasing in F and M. The second follows because $\lambda' \leq \lambda$. Integrating up gives

$$M(\lambda) \geq \frac{\gamma}{r + 2\gamma + \lambda/2} F(\lambda).$$

Then using equation (64), we get

$$M(\lambda) \geq \frac{\gamma(r+2\gamma)}{2(r+\gamma)(r+2\gamma+\lambda/2)} \left(1 - (\underline{\lambda}/\lambda)^{\kappa}\right) \tag{65}$$

This implies that for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\underline{\lambda} < \delta$ and $\lambda > \varepsilon$,

$$\begin{aligned} M(\lambda) &> \frac{\gamma(r+2\gamma)}{2(r+\gamma)(r+2\gamma+\lambda/2)} - \tfrac{1}{2}\varepsilon \\ &= \frac{\gamma}{2(r+\gamma)} - \frac{\gamma\lambda/2}{(2(r+\gamma))(r+2\gamma+\lambda/2)} - \tfrac{1}{2}\varepsilon. \end{aligned}$$

The second equality uses simple algebra. Now take any $\lambda' \in (0, \lambda)$ with $\frac{\gamma\lambda'/2}{(2(r+\gamma))(r+2\gamma+\lambda'/2)} < \frac{1}{2}\varepsilon$. Then the preceding logic implies that there exists a δ' such that if $\underline{\lambda} < \delta'$, $M(\lambda') > \frac{\gamma}{2(r+\gamma)} - \varepsilon$. Since M is monotonically increasing, this implies $M(\lambda) > \frac{\gamma}{2(r+\gamma)} - \varepsilon$ as well. This establishes the lower bound in inequality (60).

We now turn to upper bounds. First, combine $\gamma(1 - F(\lambda)) - rM(\lambda) \geq 0$ with inequal-

ity (63) to get $\gamma \geq (2(r+\gamma))M(\lambda)$. This is the upper bound in inequality (60).

Lastly, combine $\gamma(1-F(\lambda)) - rM(\lambda) \geq 0$ with the lower bound in inequality (60) to get $F(\lambda) \leq \frac{r+2\gamma}{2(r+\gamma)} + \frac{r\varepsilon}{\gamma}$. The upper bound in inequality (59) then follows from an appropriate rescaling of ε.

In summary, in the limit as $\underline{\lambda}$ converges to 0, F converges to a step function $F(\lambda) = \frac{r+2\gamma}{2(r+\gamma)}$ for $\lambda < \bar{\lambda}$ and the associated misalignment rate. This is exactly the counterparty distribution associated with the cost $\bar{c}$, as defined earlier in the proof and as can be verified directly from equation (58).

Step 4: Limit as $\underline{\lambda} \to \bar{\lambda}$. It is easier to characterize the solution to the initial value problem (57) when $\underline{\lambda}$ is large. The key observation is that $\lim_{\lambda \to \bar{\lambda}} x_F(\lambda)$ and $\lim_{\lambda \to \bar{\lambda}} x_M(\lambda)$ are both finite since $F(\lambda) \in [0,1]$ and $M(\lambda) \in [0, 1/2]$. This implies that for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\underline{\lambda} \in (\bar{\lambda} - \delta, \bar{\lambda})$, $0 \leq 2M(\lambda) \leq F(\lambda) < \varepsilon$ for all $\lambda < \bar{\lambda}$. Of course, we always have $F(\bar{\lambda}) = 1$ and we can construct $M(\bar{\lambda})$ using equation (30). This implies that F and M converge exactly to the counterparty distribution associated with the cost $\underline{c}$, as defined earlier in the proof.

Step 5: Intermediate Value Theorem. We have shown that the right hand side of equation (58) is a continuous function of $\underline{\lambda}$, converging to $\bar{c}$ when $\underline{\lambda} \to 0$ and to $\underline{c}$ when $\underline{\lambda} \to \bar{\lambda}$. The intermediate value theorem therefore implies that for any $c \in (\underline{c}, \bar{c})$, there is a $\underline{\lambda} \in (0, \bar{\lambda})$ such that the solution to the initial value problem (57) satisfies equation (58) and hence F is associated with an equilibrium.

Necessity of Full Support. We now prove that in any equilibrium with $\mu_F((0,\bar{\lambda})) = 1 - \mu_F(\{\bar{\lambda}\}) \in (0,1]$, the support of F is an interval $[\underline{\lambda}, \bar{\lambda}]$.

First, to find a contradiction, suppose that there is a hole in the support. That is, there are contact rates λ_1, λ_2 in the support with $0 \leq \lambda_1 < \lambda_2 \leq \bar{\lambda}$ and for all $\lambda \in (\lambda_1, \lambda_2)$, $F(\lambda) = F(\lambda_1) > 0$. Since by assumption $\mu_F((0,\bar{\lambda}]) = 1$, it must be the case that $\lambda_1 > 0$.

Eliminate $s''(\lambda)$ from equation (49) using equation (43). Then simplify with $C''(\lambda) = F'(\lambda) = M'(\lambda) = 0$, where the first restriction uses linearity of the cost function and the others use the assumption that there is a hole in the support of the distribution. This implies that for all $\lambda \in (\lambda_1, \lambda_2)$,

$$\pi''_\lambda = \frac{8\Delta e^{-\int_\lambda^\infty \phi_{\lambda'} d\lambda'} \big(rM(\lambda_1) - \gamma(1-F(\lambda_1))\big)}{r\lambda\big(4(r+2\gamma) + \lambda(1-F(\lambda_1)+2M(\lambda_1))\big)^2}.$$

Now if $rM(\lambda_1) < \gamma(1-F(\lambda_1)$, the profit function would be strictly concave on the interval

(λ_1, λ_2). This is inconsistent with λ_1 and λ_2 both being profit maximizing.

Thus it must be the case that $rM(\lambda_1) \geq \gamma(1 - F(\lambda_1))$. But then equation (57) implies $x_F(\lambda_1) \leq 0$ and $x_M(\lambda_1) \leq 0$. Then the initial value problem (57) implies that there is no interval (λ_0, λ_1) with $F(\lambda)$ strictly increasing for $\lambda \in (\lambda_0, \lambda_1)$, contradicting the hypothesis that λ_1 was in the support of F.

With no holes in the support and $\mu_F((0, \bar{\lambda})) \in (0, 1]$, the only other possibility is that the upper bound of the support is less than $\bar{\lambda}$. However, $\gamma(1 - F(\lambda)) > rM(\lambda) > 0$, established in Step 1, implies $F(\lambda) < 1$ for all λ, precluding this possibility.

Support at the Upper Bound. In the intermediate range, we have already proved in Step 1 above that F and M are continuous and strictly increasing on $[\underline{\lambda}, \bar{\lambda})$. G inherits the same support. To prove that $\mu_F(\{\bar{\lambda}\}) > 0$, we use $\gamma(1 - F(\lambda)) > rM(\lambda) > 0$ for all $\lambda \in [\underline{\lambda}, \bar{\lambda})$; again we established this in step 1 above. This puts an upper bound on $F(\lambda)$ and hence a lower bound on $\mu_F(\{\bar{\lambda}\})$. Finally, since the support of F is an interval $[\underline{\lambda}, \bar{\lambda}]$ with $\underline{\lambda} > 0$ and $\mu_F(\{\bar{\lambda}\}) < 1$, equation (17) implies $\mu_F(\{\bar{\lambda}\}) = 1 - \mu_F((0, \bar{\lambda})) \in (0, 1)$.

Increasing misalignment rate. Since F is absolutely continuous, we can differentiate equation (31) to get an expression for m'_λ.

$$m'_\lambda = \frac{2(rM(\lambda) - \gamma(1 - F(\lambda))) + \lambda\big(M'(\lambda)(2r + \lambda(1 - F(\lambda))) + F'(\lambda)(2\gamma + \lambda M(\lambda))\big)}{\big(2r + 4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda))\big)^2} \quad (66)$$

We have shown that the support of F is a convex set. Thus we need to prove that m'_λ is positive on $[\underline{\lambda}, \bar{\lambda}]$. We claim that

$$m'_\lambda = \frac{2\lambda\big(\gamma(8r + 8\gamma + 5\lambda(1 - F(\lambda) + M(\lambda))) + \lambda M(\lambda)(5r + 5\gamma + 3\lambda(1 - F(\lambda) + M(\lambda)))\big)M'(\lambda)}{4(2\gamma + \lambda M(\lambda))(2r + 4\gamma + \lambda(1 - F(\lambda) + M(\lambda)))^2}, \quad (67)$$

which is clearly positive. The proof is brute force: eliminate $F'(\lambda)$ and $M'(\lambda)$ from equations (66) and (67) using equation (57) and show that the two expressions are equivalent. ■

We next turn to a limiting equilibrium. Before proving Proposition 4, we prove the following Lemma, which characterizes limiting equilibrium in a manner similar to Proposition 3:

Lemma 2 *Assume $C(\lambda) = c\lambda$. If $c \geq \bar{c}^* \equiv \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)}$, a limiting (autarky) equilibrium with $F(\lambda) = F(0) > 0$ for all λ exists. If $c < \bar{c}^*$, a limiting (intermediated trade) equilibrium*

with $F(\underline{\lambda}) = 0$ and F strictly increasing on $[\underline{\lambda}, \infty)$ for some $\underline{\lambda} > 0$ exists. Moreover, any limiting equilibrium takes one of these forms.

Proof of Lemma 2. We construct the limiting equilibrium (F, m, s), as well as the corresponding convergent sequence, (F_n, m_n, s_n), for different ranges of the constant marginal cost, c, separately.

Degenerate Autarky Equilibrium. Suppose $c \geq \bar{\bar{c}}^* \equiv \frac{\gamma\Delta}{4r(r+2\gamma)^2}$. We prove that for any finite $\bar{\lambda}_n$, there is an equilibrium with $F_n(\lambda) = 1$ for all $\lambda \geq 0$ and hence this is true in the limiting equilibrium as well. To prove this, note that if $F_n(0) = 1$, $M_n(0) = M_n(\bar{\lambda}) = \frac{\gamma}{r+2\gamma}$. Then equation (55) and the argument around it implies that there is an equilibrium with $F_n(0) = 1$ if and only if

$$c \geq \bar{\bar{c}}_n \equiv \frac{\gamma\Delta\big(4\gamma(r+2\gamma) + \gamma\bar{\lambda}_n\big)}{4r(r+2\gamma)^2\big(2(r+2\gamma)^2 + \gamma\bar{\lambda}_n\big)} = \left(\frac{4\gamma(r+2\gamma) + \gamma\bar{\lambda}_n}{2(r+2\gamma)^2 + \gamma\bar{\lambda}_n}\right)\bar{\bar{c}}^*$$

Simple algebra implies $\bar{\bar{c}}_n < \bar{\bar{c}}^*$, and hence such an equilibrium exists whenever $c \geq \bar{\bar{c}}^*$. In such an equilibrium, we can also construct m_n from steady state equation (9), which gives a version of equation (30):

$$\left(r + \gamma + \frac{\lambda}{2}(1 - F_n(\lambda) + M_n(\lambda))\right) m_{n,\lambda} = \left(\gamma + \frac{\lambda}{2}(M_n(\lambda) - \mu_n(\{\lambda\}))\right)(1 - m_{n,\lambda}), \quad (68)$$

where $m_{n,\lambda}$ is the misalignment rate of traders with contact rate λ when the upper bound on contact rates is $\bar{\lambda}_n$. Similarly we can construct s_n from equation (35). Since F_n and M_n do not depend on n, m_n and s_n are also independent of n and hence their limits are trivial. Note also from this equation that for large $\bar{\lambda}_n$, $\bar{\bar{c}}_n$ converges to $\bar{\bar{c}}^*$ and so a limiting equilibrium of this type does not exist when $c < \bar{\bar{c}}^*$.

Two-Point Autarky Equilibrium. Suppose $\bar{\bar{c}}^* > c \geq \bar{c}^* \equiv \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)}$, where we can confirm algebraically that $\bar{\bar{c}}^* > \bar{c}^*$. In this case, we claim that for sufficiently large $\bar{\lambda}_n$, the contact rate distribution has two points in its support, 0 and $\bar{\lambda}_n$. The proof of Proposition 3 defines the threshold for such an equilibrium to exist, $\bar{c}_n$, as the value that makes the right hand side of equation (55) equal to zero when $F(0) = \frac{r+2\gamma}{2(r+\gamma)}$ and $M(0) = \frac{\gamma}{2(r+\gamma)}$:

$$\bar{c}_n = \frac{\gamma\Delta\big(8(r+\gamma) + \bar{\lambda}_n\big)M(\bar{\lambda}_n)}{8r(r+\gamma)(r+2\gamma)\big(2(r+2\gamma) + \bar{\lambda}_n M(\bar{\lambda}_n)\big)} = \left(\frac{\big(8(r+\gamma) + \bar{\lambda}_n\big)M(\bar{\lambda}_n)}{2(r+2\gamma) + \bar{\lambda}_n M(\bar{\lambda}_n)}\right)\bar{c}^* \quad (69)$$

Note that $M(\bar{\lambda}_n) = M(0) + m_{\bar{\lambda}_n}(1 - F(0)) < M(0) + \frac{1}{2}(1 - F(0)) = \frac{r+2\gamma}{4(r+\gamma)}$, since the misalignment rate is always less than $\frac{1}{2}$. This implies $\frac{(8(r+\gamma)+\bar{\lambda}_n)M(\bar{\lambda}_n)}{2(r+2\gamma)+\bar{\lambda}_n M(\bar{\lambda}_n)} < 1$ and hence $\bar{c}_n < \bar{c}^*$. On the other hand, as $\bar{\lambda}_n$ converges to infinity, we know $\bar{\bar{c}}_n$ converges to $\bar{\bar{c}}^*$ and so it follows that if $\bar{\bar{c}}^* > c \geq \bar{c}^*$, an equilibrium with this two point contact distribution exists for sufficiently large $\bar{\lambda}_n$.

To find equilibrium with finite $\bar{\lambda}_n$, set the right hand side of (55) to zero and use $M_n(0) = \frac{\gamma}{r+2\gamma}F_n(0)$ and $M_n(\bar{\lambda}_n) = M_n(0) + m_{n,\bar{\lambda}_n}(1 - F_n(0))$, with the misalignment rate at the upper bound, $m_{n,\bar{\lambda}_n}$, defined by equation (68). This gives us an equation that implicitly defines $F_n(0)$:

$$c = \frac{\big(4\gamma + \bar{\lambda}_n \frac{\gamma}{r+2\gamma}F_n(0)\big)\big(\frac{\gamma}{r+2\gamma}F_n(0) + m_{n,\bar{\lambda}_n}(1 - F_n(0))\big)\Delta}{4r(r+2\gamma)\big(2(r+2\gamma) + \bar{\lambda}_n\big(\frac{\gamma}{r+2\gamma}F_n(0) + m_{n,\bar{\lambda}_n}(1 - F_n(0))\big)\big)}$$

Although we cannot solve this explicitly for $F_n(0)$, this equation implies it is continuous in $\bar{\lambda}_n$ and so for sufficiently large $\bar{\lambda}_n$, the right hand side converges to $\frac{\gamma\Delta F_n(0)}{4r(r+2\gamma)^2}$. Inverting this implies that F_n converges pointwise to F satisfying

$$F(\lambda) = \frac{4r(r+2\gamma)^2 c}{\gamma\Delta}$$

for all $\lambda \geq 0$. Again, we can recover m_n and s_n from equations (68) and (35). Since each depends continuously on the functions F and M, they converge as well.

Trading Equilibrium. Finally suppose $c < \bar{c}^*$. We first prove that for fixed c and sufficiently large $\bar{\lambda}_n$, the equilibrium counterparty distribution is not degenerate. We have already shown that it must have trade, $F(0) = 0$.

Since $F(\lambda) \geq 2M(\lambda)$ for all λ,

$$\frac{4(r+2\gamma)}{\lambda\big(4(r+2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))\big)} \geq \frac{4(r+2\gamma)}{\lambda\big(4(r+2\gamma) + \lambda\big)}.$$

Applying this inequality to the integrand in equation (58) and solving the integral explicitly gives

$$c \leq \frac{4\gamma\Delta}{r\left(4(r+2\gamma) + \underline{\lambda}_n\right)^2},$$

where $\underline{\lambda}_n$ is the lower bound for a given $\bar{\lambda}_n$. This equation gives us an upper bound on $\underline{\lambda}_n$ for a given value of c. Equivalently, if

$$\bar{\lambda}_n > 2\sqrt{\frac{\gamma\Delta}{rc}} - 4(r+2\gamma),$$

$\bar{\lambda}_n > \underline{\lambda}_n$. This proves that when $\bar{\lambda}_n$ is an intermediated trade equilibrium. From Proposition 3, any such equilibrium is completely characterized by a positive lower bound $\underline{\lambda}_n$. Moreover, the preceding argument implies $\underline{\lambda}_n < 2\sqrt{\frac{\gamma\Delta}{rc}} - 4(r+2\gamma)$.

Now take any increasing and unbounded sequence $\{\bar{\lambda}_n\}$. The preceding argument implies that for sufficiently large $\bar{\lambda}_n$, equilibrium is characterized by a bounded sequence $\underline{\lambda}_n \in \left[0, 2\sqrt{\gamma\Delta/rc} - 4(r+2\gamma)\right]$, with associated functions $F_n(\lambda)$ and $M_n(\lambda)$ that solve the initial value problem (57). By the Bolzano-Weierstrass Theorem, the sequence of lower bounds has a convergent subsequence, $\underline{\lambda}_n \to \underline{\lambda}^*$. For notational convenience, we assume that the sequence $\bar{\lambda}_n$ is chosen such that $\underline{\lambda}_n$ itself converges to $\underline{\lambda}^*$.

$F_n(\lambda)$ and $M_n(\lambda)$ solve the initial value problem (57) with lower bound $\underline{\lambda}_n$. Since $\underline{\lambda}_n$ converges to $\underline{\lambda}^*$ and the solution to (57) is continuous in $\underline{\lambda}$ (Theorem 3.5 in Sideris, 2013), (F_n, M_n) converges pointwise to (F, M). Once more, we can recover m_n and s_n from equations (68) and (35). Since each depends continuously on the functions F and M, they converge as well. Thus we have found a limiting equilibrium with $F(0) = 0$.

That any limiting equilibrium takes one of these two forms then follows from the arguments made in the proof of proposition 3. ■

Proof of Proposition 4. We first show there are middlemen and then establish the Pareto tail of the contact rate distribution.

Middlemen. Fix r and γ. From Lemma 2, for any $c < \bar{c}^*$, a limiting equilibrium with $F(0) = 0$ exists. We prove that there exists a $F^* > 0$, independent of c, Δ, and $\bar{\lambda}_n$, such that in any equilibrium with $F_n(0) = 0$, $\mu_{F_n}(\{\bar{\lambda}_n\}) \geq F^*$. It then follows that in any limiting equilibrium there are middlemen, $F(\lambda) \leq 1 - F^*$ for all λ.

We now prove that there exists a $F^* > 0$, independent of c, Δ, and $\bar{\lambda}_n$, such that in any equilibrium with $F_n(0) = 0$, $\mu_{F_n}(\{\bar{\lambda}_n\}) \geq F^*$. For expositional convenience, we suppress n in what follows. We work with the initial value problem (57). If $\mu_F(\{\bar{\lambda}\}) = 1 \geq F^*$ then the claim is true. We thus consider intermediated trade equilibria.

We make two preliminary observations. First, the initial value problem (57) indicates that c and Δ only affect equilibrium through the value of $\underline{\lambda}$. We will therefore prove that there is a number $F^* > 0$, independent of $\underline{\lambda}$ and $\bar{\lambda}$, such that in any equilibrium with $F(0) = 0$, $\mu_F(\{\bar{\lambda}\}) \geq F^*$, or equivalently $F(\lambda) < 1 - F^*$ for $\lambda < \bar{\lambda}$. This implies that the same F^* works for all c and Δ such that $\mu_F(\{\bar{\lambda}\}) = 1 - \mu_F((0, \bar{\lambda})) \in (0, 1)$. Second, the initial value problem (57) also indicates that $\bar{\lambda}$ only affects equilibrium through the value of $\underline{\lambda}$ and through the values of F and M at $\bar{\lambda}$. Since F is nondecreasing, it suffices to prove that there is a number F^*, independent of $\underline{\lambda}$, such that in the solution to the initial value

problem (57), $\lim_{\lambda\to\infty} F(\lambda) < 1 - F^*$.

As in the proof of Proposition 3, let $Y(\lambda) \equiv \log\big(\gamma(1-F(\lambda))-rM(\lambda)\big)$. We already proved that in the solution to the initial value problem, this is finite for any finite λ. Here we prove that it is bounded below for fixed $\underline{\lambda}$. Evaluating equation (61) at $\underline{\lambda}$ using $F(\underline{\lambda}) = M(\underline{\lambda}) = 0$ gives

$$Y'(\underline{\lambda}) = \frac{-4(4r + 4\gamma + \underline{\lambda}))}{\underline{\lambda}(8r + 8\gamma + 3\underline{\lambda})}.$$

Moreover, simple algebra takes us from equation (61) to

$$Y'(\lambda) \geq \frac{-4(r + 2\gamma)}{\lambda^2 M(\lambda)}.$$

Now fix $\tilde{\lambda} > \underline{\lambda}$ and note that $M(\tilde{\lambda}) > 0$. For all $\lambda > \tilde{\lambda}$, $M(\lambda) \geq M(\tilde{\lambda})$ and hence

$$Y'(\lambda) \geq \frac{-4(r + 2\gamma)}{\lambda^2 M(\tilde{\lambda})}.$$

Integrating up this lower bound on the slope gives us that for $\lambda > \tilde{\lambda}$,

$$Y(\lambda) \geq Y(\tilde{\lambda}) - \frac{4(r + 2\gamma)}{M(\tilde{\lambda})}\left(\frac{1}{\tilde{\lambda}} - \frac{1}{\lambda}\right) \geq Y(\tilde{\lambda}) - \frac{4(r + 2\gamma)}{\tilde{\lambda}M(\tilde{\lambda})},$$

where the second inequality is algebra. This proves that $Y(\lambda)$ is bounded below and hence $\lim_{\lambda\to\infty} \gamma(1 - F(\lambda)) - rM(\lambda) > 0$ for any fixed $\underline{\lambda}$. In particular, since $M(\lambda) \geq 0$, $\lim_{\lambda\to\infty} F(\lambda) < 1$.

As we vary $\underline{\lambda}$, $\lim_{\lambda\to\infty} F(\lambda)$ changes continuously (Theorem 3.5 in Sideris, 2013), but is always strictly less than 1. To prove the existence of the uniform upper bound $1 - F^*$, we still need to prove that $\lim_{\lambda\to\infty} F(\lambda)$ does not converge to 1 for some value of $\underline{\lambda}$. Continuity implies that this cannot happen at an interior value of $\underline{\lambda}$: in the interior, the supremum of $\lim_{\lambda\to\infty} F(\lambda)$ is equal to the maximum, which we already proved is strictly less than 1. We next show that the supremum is less than 1 even if it occurs at either the limit as $\underline{\lambda} \to 0$ or $\underline{\lambda} \to \infty$.

First recall the solution to the initial value problem in the limit as $\underline{\lambda} \to 0$. We showed in the proof of Proposition 3 that for all $\lambda \in (0, \bar{\lambda})$, $F(\lambda) \to \frac{r+2\gamma}{2(r+\gamma)} < 1$.

Next turn to the solution to the initial value problem in the limit as $\underline{\lambda} \to \infty$. In this case, $M(\lambda) \to 0$ for all λ, and so the argument above breaks down. Instead, let $\rho \equiv \lambda/\underline{\lambda}$,

$H(\rho) \equiv F(\rho\underline{\lambda})$, and $L(\rho) \equiv \rho\underline{\lambda}M(\rho\underline{\lambda})$. Rewrite the initial value problem as

$$H'(\rho) = \frac{4\big(2r + 4\gamma + \rho\underline{\lambda}(1 - H(\rho)) + 2L(\rho)\big)\big(\gamma\rho\underline{\lambda}(1 - H(\rho)) - rL(\rho)\big)}{\rho^2\underline{\lambda}\big(\gamma(8r + 8\gamma + 3\rho\underline{\lambda}(1 - H(\rho))) + L(\rho)(3r + 6\gamma + \rho\underline{\lambda}(1 - H(\rho)) + L(\rho))\big)},$$
$$L'(\rho) = \frac{4\big(2\gamma + L(\rho)\big)\big(\gamma\rho\underline{\lambda}(1 - H(\rho)) - rL(\rho)\big)}{\rho\big(\gamma(8r + 8\gamma + 3\rho\underline{\lambda}(1 - H(\rho))) + L(\rho)(3r + 6\gamma + \rho\underline{\lambda}(1 - H(\rho)) + L(\rho))\big)} + \frac{L(\rho)}{\rho},$$

with $H(1) = L(1) = 0$. Although we cannot solve these equations explicitly for arbitrary $\underline{\lambda}$, we know the solution is continuous in $\underline{\lambda}$ (Theorem 6.2 in Sideris, 2013) and we can therefore take the limit of the differential equations as $\underline{\lambda} \to \infty$ and then solve the equations to find the limits of the H and L. The initial value problem becomes

$$H'(\rho) = \frac{4\gamma(1 - H(\rho))}{\rho(3\gamma + L(\rho))}, \tag{70}$$
$$L'(\rho) = \frac{4\gamma\big(2\gamma + L(\rho)\big)}{\rho(3\gamma + L(\rho))} + \frac{L(\rho)}{\rho}, \tag{71}$$

still with $H(1) = L(1) = 0$. The solution to these equations is $\rho = q(L(\rho))$ and $H(\rho) = 1 - \eta(L(\rho))$ where

$$\eta(L) \equiv \left(\frac{16\gamma + (7 - \sqrt{17})L}{16\gamma + (7 + \sqrt{17})L}\right)^{\frac{4}{\sqrt{17}}}, \tag{72}$$
$$q(L) \equiv \eta(L)^{\frac{1}{8}}\sqrt{1 + \frac{7L}{8\gamma} + \frac{L^2}{8\gamma^2}}. \tag{73}$$

This implies $\lim_{L\to\infty}\eta(L) = \left(\frac{7-\sqrt{17}}{7+\sqrt{17}}\right)^{\frac{4}{\sqrt{17}}}$ and $\lim_{L\to\infty} q(L) = \infty$. Using these, we find that the unique limit as ρ converges to infinity of $L(\rho)$ is infinite; and the unique limit of $H(\rho)$ is $1 - \left(\frac{7-\sqrt{17}}{7+\sqrt{17}}\right)^{\frac{4}{\sqrt{17}}} \approx 0.731 < 1$. This establishes the bound in this limit.

Pareto Tail. For a limiting equilibrium (F, m, s), fix the sequence of functions (F_n, m_n, s_n) that converge to (F, m, s) and the increasing and unbounded sequence $\{\bar{\lambda}_n\}$ such that for each n, (F_n, m_n, s_n) restricted to the domain $[0, \bar{\lambda}_n]$ is an equilibrium when the maximum contact rate is $\bar{\lambda}_n$.

Use the initial value problem (57) to get that for fixed n,

$$\lim_{\lambda\to\bar{\lambda}_n}\lambda^2 F_n'(\lambda) = \frac{4\bar{\lambda}_n\big(2r + 4\gamma + \bar{\lambda}_n(1 - \bar{F}_n + 2\bar{M}_n)\big)(\gamma(1 - \bar{F}_n) - r\bar{M}_n)}{\big(\gamma(8r + 8\gamma + 3\bar{\lambda}_n(1 - \bar{F}_n)) + \bar{\lambda}_n\bar{M}_n(3r + 6\gamma + \bar{\lambda}_n(1 - \bar{F}_n + \bar{M}_n))\big)},$$

where $\bar{F}_n \equiv \lim_{\lambda\to\bar{\lambda}_n} F_n(\lambda)$ and $\bar{M}_n \equiv \lim_{\lambda\to\bar{\lambda}_n} M_n(\lambda)$. Then we take the limit as n gets large:

$$\lim_{n\to\infty}\lim_{\lambda\to\bar{\lambda}_n} \lambda^2 F_n'(\lambda) = 4\lim_{n\to\infty} \frac{(1-\bar{F}_n+2\bar{M}_n)(\gamma(1-\bar{F}_n)-r\bar{M}_n)}{\bar{M}_n(1-\bar{F}_n+\bar{M}_n)}. \tag{74}$$

Since $\underline{\lambda}_n$ converges to $\underline{\lambda}^*$ and F and M are continuous in $\underline{\lambda}$ for fixed $\bar{\lambda}_n$ (again, Theorem 3.5 in Sideris, 2013), $\bar{F}_n$ and $\bar{M}_n$ have well-behaved limits. Recalling that $\gamma(1-F_n(\lambda))-rM_n(\lambda)$ is bounded above zero, it follows that the right hand side is a positive number. That is, the density F' is asymptotically proportional to λ^{-2}.

We are interested in characterizing the contact rate distribution G rather than the counterparty density F'. To do this, first use L'Hopital's rule to get

$$\lim_{n\to\infty}\lim_{\lambda\to\bar{\lambda}_n} \lambda^2(1-G_n(\lambda)) = \frac{1}{2}\lim_{n\to\infty}\lim_{\lambda\to\bar{\lambda}_n} \lambda^3 G_n'(\lambda).$$

Since F is absolutely continuous, it follows from equation (16) that $G_n'(\lambda) = \Lambda_n F_n'(\lambda)/\lambda$, where Λ_n is the average contact rate when the upper bound is $\bar{\lambda}_n$. This implies

$$\begin{aligned}\lim_{n\to\infty}\lim_{\lambda\to\bar{\lambda}_n} \lambda^2(1-G_n(\lambda)) &= \lim_{n\to\infty}\lim_{\lambda\to\bar{\lambda}_n} \frac{\Lambda_n}{2}\lambda^2 F_n'(\lambda) \\ &= 2\lim_{n\to\infty} \frac{(1-\bar{F}_n+2\bar{M}_n)(\gamma(1-\bar{F}_n)-r\bar{M}_n)}{\bar{M}_n(1-\bar{F}_n+\bar{M}_n)\int_0^{\bar{\lambda}_n}\frac{1}{\lambda}dF_n(\lambda)}\end{aligned}$$

The second equation pulls Λ_n out of the inner limit, since it does not depend on λ. It then rewrites Λ_n using equation (17) and replaces the inner limit using equation (74). The result follows because as $\bar{\lambda}_n$ grows, the limit of each term exists and is a positive number. ■

Proof of Corollary 1. Proposition 4 proved that a strictly positive fraction of meetings is with middlemen. To show that a strictly positive fraction of trades are with middlemen, it is thus sufficient to show that a positive fraction of meetings with middlemen result in trade. This immediately follows from the fact that the misalignment rate of non-middlemen, $\lim_{\lambda\to\infty} M(\lambda)$, is strictly positive.

Next $p_\lambda = \frac{1}{2}\left(m_\lambda(1-F(\lambda))+M(\lambda)\right)$ is the fraction of meetings that result in trade for a trader with contact rate λ. Differentiate this to prove that p_λ is strictly increasing. This uses the fact that $m_\lambda' > 0$ (proposition 3) and $m_\lambda F'(\lambda) = M'(\lambda)$ by absolute continuity of F. This in turn implies that λp_λ is increasing and so the trading rate and contact rate distributions in the limiting economy are related by $\hat{G}(\lambda p_\lambda) = G(p_\lambda)$. Since p_λ converges to

a positive constant as $\lambda \to \infty$, G and $\hat{G}$ share a common tail parameter. ■

Measuring Volume We measure volume as the rate at which some trader buys the asset or, equivalently, the rate at which an asset gets bought. We do this first in an economy with finite $\bar{\lambda}$ but later take limits.

It is useful to decompose volume into four terms, $\mathcal{V} = \mathcal{V}_{mm} + \mathcal{V}_{mn} + \mathcal{V}_{nm} + \mathcal{V}_{nn}$. For expositional convenience, we call traders with $\lambda = \bar{\lambda}$ "middlemen" and traders with slower contact rates "non-middlemen". The four volumes are defined as follows: First, the rate that middlemen buy from other middlemen is

$$\mathcal{V}_{mm} = \frac{\bar{\lambda}}{4}\mu_G(\{\bar{\lambda}\})\mu_F(\{\bar{\lambda}\})m_{\bar{\lambda}}^2 = \frac{\Lambda}{4}\mu_F(\{\bar{\lambda}\})^2 m_{\bar{\lambda}}^2. \tag{75a}$$

This is equal to product of several terms: the measure of middlemen $\mu_G(\{\bar{\lambda}\})$, the fraction of those middlemen who wish to buy the asset $\frac{m_{\bar{\lambda}}}{2}$, the rate that they meet another middleman $\bar{\lambda}\mu_F(\{\bar{\lambda}\})$, and the fraction of those potential trading partners who wish to sell the asset $\frac{m_{\bar{\lambda}}}{2}$. The second equation uses the relationship between the counterparty and contact rate measures in Appendix A.1 to obtain $\Lambda\mu_F(\{\bar{\lambda}\}) = \bar{\lambda}\mu_G(\{\bar{\lambda}\})$.

Second, the rate that middlemen buy from non-middlemen, is

$$\mathcal{V}_{mn} = \frac{\bar{\lambda}}{4}\mu_G(\{\bar{\lambda}\})\big(M(\bar{\lambda}) - \mu_M(\{\bar{\lambda}\})\big) = \frac{\Lambda}{4}\mu_F(\{\bar{\lambda}\})\big(M(\bar{\lambda}) - \mu_M(\{\bar{\lambda}\})\big). \tag{75b}$$

This is again the product of several terms: the measure of middlemen $\mu_G(\{\bar{\lambda}\})$, the fraction of those middlemen who are able to buy the asset $\frac{1}{2}$ (whether or not they want to buy), and the rate that they meet a non-middleman who wants to sell the asset $\frac{M(\bar{\lambda})-\mu_M(\{\bar{\lambda}\})}{2}$. We again simplify this using $\Lambda\mu_F(\{\bar{\lambda}\}) = \bar{\lambda}\mu_G(\{\bar{\lambda}\})$.

Third, non-middlemen buy from middlemen at the same rate. This follows from the symmetry of the model, since it is equal to the rate that middlemen sell to non-middlemen

$$\mathcal{V}_{nm} = \frac{\Lambda}{4}\mu_F(\{\bar{\lambda}\})\big(M(\bar{\lambda}) - \mu_M(\{\bar{\lambda}\})\big). \tag{75c}$$

Finally, a trader with contact rate $\lambda < \bar{\lambda}$ purchases the asset when he does not own it and meets a slower misaligned trader who wants to sell it, at rate $\frac{\lambda M(\lambda)}{4}$, or when he wants to buy it and meets a faster trader who owns it, at rate $\frac{\lambda m_\lambda (1-F(\lambda)}{4}$. The latter event includes the possibility of buying from a middleman. Thus integrating over the contact rate distribution

gives

$$\mathcal{V}_{nn} + \mathcal{V}_{nm} = \frac{1}{4}\int_{\underline{\lambda}}^{\bar{\lambda}} \lambda\big(M(\lambda) + m_\lambda(1 - F(\lambda))\big)dG(\lambda) \tag{75d}$$

$$= \frac{\Lambda}{4}\int_{\underline{\lambda}}^{\bar{\lambda}} \big(M(\lambda)F'(\lambda) + (1 - F(\lambda))M'(\lambda)\big)d\lambda. \tag{75e}$$

We can subtract V_{nm} from this to get V_{nn}.

Proof of Proposition 5. We proceed in steps. We first describe the limit of limiting equilibria as costs become small before computing volume in this double limit.

Limiting Equilibrium with Small Costs We first prove that in a limiting equilibrium, the lower bound on the support of the contact rate distribution converges to infinity when the cost of a meeting goes to zero. Recall from the proof of Proposition 4 that in any limiting equilibrium, $F(\lambda) < 1 - F^*$ for all λ, where F^* is strictly positive. We also have $M(\lambda) \geq 0$. This implies

$$\frac{4(r+2\gamma)}{\lambda\Big(4(r+2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))\Big)} \leq \frac{4(r+2\gamma)}{\lambda\Big(4(r+2\gamma) + \lambda F^*\Big)}.$$

Integrating this implies

$$\exp\left(-\int_{\underline{\lambda}}^{\infty} \frac{4(r+2\gamma)}{\lambda\Big(4(r+2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))\Big)}d\lambda\right) \geq \frac{\underline{\lambda}F^*}{4(r+2\gamma) + \underline{\lambda}F^*}.$$

Using equation (58), this implies that if $\underline{\lambda} > 0$,

$$\frac{c}{\Delta} \geq \frac{4\gamma F^*}{r\big(4(r+2\gamma) + \underline{\lambda}\big)\big(4(r+2\gamma) + \underline{\lambda}F^*\big)}.$$

or

$$\underline{\lambda} \geq 2\left(\sqrt{\frac{\gamma\Delta}{rc} + \left(\frac{(1-F^*)(r+2\gamma)}{F^*}\right)^2} - \frac{(1+F^*)(r+2\gamma)}{F^*}\right).$$

As c converges to zero, the lower bound on $\underline{\lambda}$ goes to infinity, and hence $\underline{\lambda}$ must as well.

Volume in Limiting Equilibrium. Next, recalling that $m_\lambda = M'(\lambda)/F'(\lambda)$, we can take limits of the expressions in equations (75) to express volume in a limiting equilibrium as

$$\mathcal{V} = \frac{\Lambda}{4}\left(\int_{\underline{\lambda}}^{\infty} \big(M(\lambda)F'(\lambda)+(1-F(\lambda))M'(\lambda)\big)d\lambda + \lim_{\lambda\to\infty}\Big((1-F(\lambda))M(\lambda)+(1-F(\lambda))^2 m_\infty^2\Big)\right), \tag{76}$$

where m_∞ is the misalignment rate of middlemen, solving a version of equation (30) with $\lambda \to \infty$:

$$\lim_{\lambda\to\infty}\big(M(\lambda) + m_\infty(1-F(\lambda))\big)m_\infty = \lim_{\lambda\to\infty} M(\lambda)(1-m_\infty).$$

Misaligned middlemen become well-aligned when they meet any other misaligned trader. Well-aligned middlemen trade only when they meet slower misaligned traders.

Since the limiting equilibrium with small costs has $\underline{\lambda}$ going to infinity, we use the same change in variables as in the proof of Proposition 4 to have well-behaved objects in this limit: $\rho \equiv \lambda/\underline{\lambda}$, $H(\rho) \equiv F(\rho\underline{\lambda}) = F(\lambda)$, and $L(\rho) \equiv \rho\underline{\lambda}M(\rho\underline{\lambda})$. Written in terms of these variables, volume is

$$\mathcal{V} = \frac{\Lambda}{4\underline{\lambda}}\left(\int_1^{\infty} \frac{(1-H(\rho))(L'(\rho) - L(\rho)/\rho) + L(\rho)H'(\rho)}{\rho} d\rho \right.$$
$$\left. + \lim_{\rho\to\infty}\left(\frac{(1-H(\rho))L(\rho)}{\rho} + \underline{\lambda}(1-H(\rho))^2 m_\infty^2\right)\right), \tag{77}$$

where

$$\lim_{\rho\to\infty}\left(\frac{L(\rho)}{\rho\underline{\lambda}} + m_\infty(1-H(\rho))\right)m_\infty = \lim_{\rho\to\infty}\frac{L(\rho)}{\rho\underline{\lambda}}(1-m_\infty). \tag{78}$$

The terms in (77) have the same interpretation as in equation (76).

Mean Contact Rate Relative to Lower Bound. We next calculate the mean contact rate relative to the lower bound, $\Lambda/\underline{\lambda}$. Using equation (17) for the limiting equilibrium, this is

$$\frac{\Lambda}{\underline{\lambda}} = \frac{1}{\underline{\lambda}\int_{\mathcal{X}} \frac{1}{\lambda} d\mu_F(\lambda)} = \frac{1}{\int_1^\infty \frac{1}{\rho} H'(\rho) d\rho} = \frac{1}{\int_0^\infty \frac{-\eta'(L)}{q(L)} dL}.$$

The last equality rewrites the previous one using the inverse functions $q(L(\rho)) = \rho$ and $\eta(L) \equiv 1 - H(q(L))$. This holds for arbitrary $\underline{\lambda}$. In the limit as $c \to 0$, we have $\underline{\lambda} \to \infty$. We can therefore apply equations (72) and (73) to the previous equation and solve the integral explicitly. This gives us an exact solution for the mean-min contact ratio in the limiting

economy:

$$\lim_{c\to 0}\frac{\Lambda}{\underline{\lambda}}=\frac{1}{\sqrt{2}}\left(\frac{7+\sqrt{17}}{7-\sqrt{17}}\right)^{\frac{7}{2\sqrt{17}}}\approx 2.23. \tag{79}$$

Probability Distribution over L. Let $\Gamma(L)$ be the population distribution of L. Since L is an increasing function of ρ, which in turn is increasing in λ, this is a transformation of the contact rate distribution, $\Gamma(L)=G(q(L)\underline{\lambda})$. Differentiating this gives us

$$\Gamma'(L)=q'(L)G'(q(L)\underline{\lambda})\underline{\lambda}=\frac{\Lambda q'(L)F'(q(L)\underline{\lambda})}{q(L)}=\frac{\Lambda q'(L)H'(q(L))}{\underline{\lambda}q(L)}=-\frac{\Lambda\eta'(L)}{\underline{\lambda}q(L)}. \tag{80}$$

The first equation differentiates $\Gamma(L)=G(q(L)\underline{\lambda})$, the second uses equation (16) to rewrite this in terms of F, the third differentiates $F(\rho\underline{\lambda})=H(\rho)$, and the fourth differentiates the definition $\eta(L)\equiv 1-H(q(L))$.

In the limit as $c\to 0$, we again apply equations (72) and (73) and integrate Γ explicitly using the initial condition $\Gamma(0)=0$. We simplify this further using the limiting behavior of $\Lambda/\underline{\lambda}$ from equation (79):

$$\lim_{c\to 0}\Gamma(L)=\frac{L}{\sqrt{8\gamma^2+7\gamma L+L^2}}\left(\frac{2L+\gamma(7+\sqrt{17})}{2L+\gamma(7-\sqrt{17})}\right)^{\frac{7}{2\sqrt{17}}} \tag{81}$$

The fraction of traders with a relative contact rate less than ρ is then $\Gamma(L(\rho))$.

Volume of Purchases by Non-Middlemen. Next consider the first term in the trading rate (77), the rate at which a non-middleman buys the asset from another trader, either a middleman or a non-middleman.

$$\begin{aligned}\mathcal{V}_{nn}+\mathcal{V}_{nm}&\equiv\frac{\Lambda}{4\underline{\lambda}}\int_1^\infty\frac{(1-H(\rho))(L'(\rho)-L(\rho)/\rho)+L(\rho)H'(\rho)}{\rho}d\rho\\&=\frac{\Lambda}{2\underline{\lambda}}\int_1^\infty\frac{(\gamma+L(\rho))H'(\rho)}{\rho}d\rho\\&=\frac{1}{2}\int_1^\infty(\gamma+L(\rho))\Gamma'(L(\rho))L'(\rho)d\rho=\frac{1}{2}\left(\gamma+\int_0^\infty L\Gamma'(L)dL\right).\end{aligned}$$

The first equation eliminates $L'(\rho)$ using equation (71) and then simplifies with equation (70). The second eliminates $H'(\rho)=H'(q(L(\rho))$ using equation (80) and then uses $q'(L(\rho))=1/L'(\rho)$, since tho functions are inverses. The last equation is a change of the variable of integration from ρ to $L(\rho)$. Once again, we now take the limit as $c\to 0$ and so apply the

functional form in equation (80) to get

$$\lim_{c\to 0}(\mathcal{V}_{nn}+\mathcal{V}_{nm}) = \left(\sqrt{2}\left(\frac{7+\sqrt{17}}{7-\sqrt{17}}\right)^{\frac{7}{2\sqrt{17}}} - 3\right)\gamma \approx 1.46\gamma. \tag{82}$$

Volume of Purchases by Middlemen from Non-Middlemen. Next consider the second term in the trading rate (77), the rate at which a middleman buys the asset from a non-middleman:

$$\mathcal{V}_{mn} \equiv \frac{\Lambda}{4\underline{\lambda}}\lim_{\rho\to\infty}\frac{(1-H(\rho))L(\rho)}{\rho} = \frac{\Lambda}{4\underline{\lambda}}\lim_{\rho\to\infty}(1-H(\rho))L'(\rho) = \frac{\Lambda}{4\underline{\lambda}}\lim_{L\to\infty}\frac{\eta(L)}{q'(L)} \tag{83}$$

The first equation uses L'Hôpital's rule, since $H(\rho)$ converges to a number less than 1, while $L(\rho)$ and ρ both grow without bound. The second equation uses $\rho = q(L)$, $H(q(L)) = 1-\eta(L)$, and $q'(L) = 1/L'(q(L))$. Again, we take the limit as $c \to 0$ and so apply the functional forms in equations (72) and (73) as well as equation (79). This gives

$$\lim_{c\to 0}\mathcal{V}_{mn} = \frac{\gamma}{2}.$$

The same argument implies middlemen sell to traders with finite contact rate at rate $\frac{1}{2}\gamma$ and an immediate corollary is that the reverse trade occurs at the same rate, $\lim_{c\to 0}\mathcal{V}_{nm} = \frac{\gamma}{2}$. Subtracting this from equation (82), we get that non-middlemen buy from non-middlemen at rate

$$\lim_{c\to 0}\mathcal{V}_{nn} = \left(\sqrt{2}\left(\frac{7+\sqrt{17}}{7-\sqrt{17}}\right)^{\frac{7}{2\sqrt{17}}} - \frac{7}{2}\right)\gamma \approx 0.96\gamma.$$

Volume of Purchases by Middlemen from Middlemen. Finally consider the third term in equation (77), the rate that middlemen buy from middlemen:

$$\mathcal{V}_{mm} = \frac{\Lambda}{4}\lim_{\rho\to\infty}(1-H(\rho))^2 m_\infty^2,$$

where m_∞ solves equation (78). Multiply both sides of equation (78) by $\frac{\Lambda}{4}\big(1-\lim_{\rho\to\infty}H(\rho)\big)$ to get

$$\mathcal{V}_{mm} = \frac{\Lambda}{4}\lim_{\rho\to\infty}(1-H(\rho))^2 m_\infty^2 = \frac{\Lambda(1-2m_\infty)}{4}\lim_{\rho\to\infty}\frac{(1-H(\rho))L(\rho)}{\rho\underline{\lambda}} = (1-2m_\infty)\mathcal{V}_{mn},$$

where $\mathcal{V}_{mn}$ is defined in equation (83). When $c \to 0$, $\underline{\lambda} \to \infty$, $m_\infty \to 0$, and the results in

the previous paragraph imply $\lim_{c\to 0}\mathcal{V}_{mm}=\lim_{c\to 0}\mathcal{V}_{mn}=\frac{\gamma}{2}$. ■

D Optimum – Details and Proofs

D.1 Planning Problem and General Characterization

Using equations (16) and (17), we can rewrite the planner's objective in (11) as

$$\delta_1-\frac{1}{\int_{\mathcal{X}}\frac{1}{\lambda}d\mu_F(\lambda)}\int_{\mathcal{X}}\frac{\Delta m_\lambda+rC(\lambda)}{\lambda}d\mu_F(\lambda), \tag{84}$$

now expressed in terms of the counterparty measure μ_F. The planner has two instruments. First, she chooses the set of permissible trades: In particular, when a trader with contact rate λ and alignment status $a\in\{0,1\}$ meets a trader with contact rate λ', alignment status $a'\in\{0,1\}$, and the opposite asset position, they trade with time-invariant probability $\mathbb{I}_{\lambda,a}^{\lambda',a'}=\mathbb{I}_{\lambda',a'}^{\lambda,a}\in[0,1]$, chosen by the planner. This implies that the steady state misalignment rate $m_\lambda\in[0,1]$ satisfies equation (22). Second, the planner chooses $\mu_F(S)$, the time-invariant probability that conditional on a meeting, the counterparty's contact rate is some $\lambda\in S$ with $\mu_F(\mathcal{X})=1$.

The Lagrangian The solution to the planner's problem is a probability measure $\mu_F(\lambda)$, misalignment rate m_λ, and trading probabilities $\mathbb{I}_{\lambda,a}^{\lambda',a'}\in[0,1]$ that maximize (84) subject to (22). Denote by $S(\lambda)d\mu_G(\lambda)$ the respective multiplier. We express this as a Lagrangian:

$$\begin{aligned}\mathcal{L}=\delta_1+\frac{1}{\int_{\mathcal{X}}\frac{1}{\lambda}d\mu_F(\lambda)}\Bigg(&-\int_{\mathcal{X}}\frac{\Delta m_\lambda+rC(\lambda)}{\lambda}d\mu_F(\lambda)\\&+\int_{\mathcal{X}}S(\lambda)\Bigg(\Bigg(\frac{r+\gamma}{\lambda}+\frac{1}{2}\int_{\mathcal{X}}\left(\mathbb{I}_{\lambda,0}^{\lambda',0}m_{\lambda'}+\mathbb{I}_{\lambda,0}^{\lambda',1}(1-m_{\lambda'})\right)d\mu_F(\lambda')\Bigg)m_\lambda\\&-\left(\frac{\gamma}{\lambda}+\frac{1}{2}\int_{\mathcal{X}}\left(\mathbb{I}_{\lambda,1}^{\lambda',0}m_{\lambda'}+\mathbb{I}_{\lambda,1}^{\lambda',1}(1-m_{\lambda'})\right)d\mu_F(\lambda')\right)(1-m_\lambda)\Bigg)d\mu_F(\lambda)\Bigg).\end{aligned} \tag{85}$$

The trading probabilities, counterparty measures, and misalignment rates must all be either extreme points that maximize the Lagrangian or interior stationary points of the Lagrangian.[20]

[20]The planner may want to choose autarky, $\mu_G(\{0\})=1$. In this case, any μ_F with $\frac{1}{\lambda}$ not Lebesgue integrable under μ_F yields the same value of the objective function. For expositional simplicity, we write the planner's first order necessary conditions for optimality when autarky is not optimal and impose the same

Now consider the following variational functions:

- the trading probabilities are $\mathbb{I}_{\lambda,a}^{\lambda',a'} + \varepsilon_{\mathbb{I}} \nu_{\mathbb{I},\lambda,a}^{\lambda',a'}$ for some nonnegative number $\varepsilon_{\mathbb{I}}$ and some function $\nu_{\mathbb{I},\lambda,a}^{\lambda',a'}$ satisfying $\nu_{\mathbb{I},\lambda,a}^{\lambda',a'} = \nu_{\mathbb{I},\lambda',a'}^{\lambda,a}$ and $\mathbb{I}_{\lambda,a}^{\lambda',a'} + \nu_{\mathbb{I},\lambda,a}^{\lambda',a'} \in [0,1]$ for all λ, a, λ', a';
- the counterparty measure on any set $S \subset \mathcal{X}$ is $(1-\varepsilon_F)\mu_F(S) + \varepsilon_F \nu_F(S)$ for some nonnegative number ε_F and some measure ν_F on X; and
- the misalignment rate is $m_\lambda + \varepsilon_m \nu_{m,\lambda}$ for arbitrary ε_m and functions $\nu_{m,\lambda}$.

A necessary condition for optimality is that for any such deviation functions $(\nu_{\mathbb{I},\lambda,a}^{\lambda',a'}, \nu_F, \nu_{m,\lambda})$, the first derivative of the Lagrangian with respect to $\varepsilon_{\mathbb{I}}$ and ε_F should be non-positive when evaluated at $(\varepsilon_{\mathbb{I}}, \varepsilon_F, \varepsilon_m) = 0$, while the first derivative with respect to ε_m should be zero.

To see the implications of these optimality conditions, set $\nu_{\mathbb{I},\lambda,0}^{\lambda',1} = \nu_{\mathbb{I},\lambda,1}^{\lambda',0} = \nu_{\mathbb{I},\lambda,1}^{\lambda',1} = 0$ and consider the derivative of the Lagrangian with respect to $\varepsilon_{\mathbb{I}}$ evaluated at 0. This gives

$$\left.\frac{\partial L}{\partial \varepsilon_{\mathbb{I}}}\right|_{\varepsilon_{\mathbb{I}}=0} = \frac{1}{2\int_{\mathcal{X}} \frac{1}{\lambda} d\mu_F(\lambda)} \int_{\mathcal{X}^2} S(\lambda) \nu_{\mathbb{I},\lambda,0}^{\lambda',0} m_\lambda m_{\lambda'} d(\mu_F(\lambda) \times \mu_F(\lambda')). \tag{86}$$

Now consider the following deviation function

$$\nu_{\mathbb{I},\lambda,0}^{\lambda',0} = \begin{cases} 1 - \mathbb{I}_{\lambda,0}^{\lambda',0} \\ 0 \qquad\qquad \text{if } S(\lambda) + S(\lambda') \gtreqless 0 \\ -\mathbb{I}_{\lambda,0}^{\lambda',0} \end{cases}$$

This deviation is feasible since it respects the symmetry of the trading probability function and keeps it a probability. Then the right hand side of equation (86) is strictly positive if Λ is positive and there is a $\mu_F \times \mu_F$ positive measure of (λ, λ') with $S(\lambda) + S(\lambda') > 0$ and $\mathbb{I}_{\lambda,0}^{\lambda',0} < 1$ or with $S(\lambda) + S(\lambda') < 0$ and $\mathbb{I}_{\lambda,0}^{\lambda',0} > 0$ (or both). This implies such a configuration is not optimal. This gives us the first of the following necessary conditions:

$$\begin{aligned} S(\lambda) + S(\lambda') \gtrless 0 \Rightarrow \mathbb{I}_{\lambda,0}^{\lambda',0} &= \begin{cases} 1 \\ 0 \end{cases}, \qquad S(\lambda) \gtrless S(\lambda') \Rightarrow \mathbb{I}_{\lambda,0}^{\lambda',1} = \begin{cases} 1 \\ 0 \end{cases}, \\ S(\lambda) + S(\lambda') \lessgtr 0 \Rightarrow \mathbb{I}_{\lambda,1}^{\lambda',1} &= \begin{cases} 1 \\ 0 \end{cases}, \qquad S(\lambda') \gtrless S(\lambda) \Rightarrow \mathbb{I}_{\lambda,1}^{\lambda',0} = \begin{cases} 1 \\ 0 \end{cases}. \end{aligned} \tag{87}$$

almost everywhere under the measure $\mu_F \times \mu_F$. The remaining necessary conditions in condition (87) follow from similar deviations. Substituting this into the constraint (22) gives

conditions when autarky is optimal.

equation (13) in the text.

Next, turn to the misalignment rate. When $\Lambda > 0$, the first derivative of the Lagrangian with respect to ε_m is

$$\begin{aligned}\left.\frac{\partial L}{\partial \varepsilon_m}\right|_{\varepsilon_m=0} = \int_{\mathcal{X}} \Bigg(& -\frac{\Delta}{\lambda} + \frac{r+2\gamma}{\lambda} S(\lambda) \\ & + \frac{1}{2}\Bigg(\int_{\mathcal{X}} \Big(\mathbb{I}_{\lambda,0}^{\lambda',0}(S(\lambda)+S(\lambda'))m_{\lambda'} + \mathbb{I}_{\lambda,0}^{\lambda',1}(S(\lambda)-S(\lambda'))(1-m_{\lambda'}) \\ & + \mathbb{I}_{\lambda,1}^{\lambda',0}(S(\lambda)-S(\lambda'))m_{\lambda'} + \mathbb{I}_{\lambda,1}^{\lambda',1}(S(\lambda)+S(\lambda'))(1-m_{\lambda'}) \Big) d\mu_F(\lambda') \Bigg) \Bigg) \nu_{m,\lambda} d\mu_F(\lambda).\end{aligned}$$

Let $\nu_{m,\lambda} = 1$ whenever the integrand is positive and -1 whenever it is negative. Unless the integrand is zero almost everywhere under the measure μ_F, this gives $\left.\frac{\partial L}{\partial \varepsilon_m}\right|_{\varepsilon_m=0} > 0$, a contradiction with optimality. It follows that whenever $\Lambda > 0$, a necessary condition for optimality is that almost everywhere under the measure μ_F,

$$\begin{aligned}\Delta = (r+2\gamma)S(\lambda) & \\ & + \frac{\lambda}{2}\Bigg(\int_{\mathcal{X}} \Big(\mathbb{I}_{\lambda,0}^{\lambda',0}(S(\lambda)+S(\lambda'))m_{\lambda'} + \mathbb{I}_{\lambda,0}^{\lambda',1}(S(\lambda)-S(\lambda'))(1-m_{\lambda'}) \\ & \quad + \mathbb{I}_{\lambda,1}^{\lambda',0}(S(\lambda)-S(\lambda'))m_{\lambda'} + \mathbb{I}_{\lambda,1}^{\lambda',1}(S(\lambda)+S(\lambda'))(1-m_{\lambda'}) \Big) d\mu_F(\lambda') \Bigg). \quad (88)\end{aligned}$$

Using (87), rewrite the first order condition (88) as equation (12) in the text. This leads to our first result:

Lemma 1-P *In an optimum, the social surplus function $S(\lambda)$ is positive-valued and strictly decreasing. When two traders with opposite asset positions meet they*

1. *always trade the asset if both are misaligned;*
2. *never trade the asset if both are well-aligned;*
3. *trade the asset if one is misaligned and the other is well-aligned and the well-aligned trader has the higher contact rate.*

The proof of this result mimics the proof of Lemma 1, showing that the social surplus function is uniquely defined by equation (12) and moreover is decreasing and nonnegative. The optimal trading pattern follows immediately. Given the similarity of the mathematical structure of these lemmas, we omit the proof.

We next find the cost function for which a particular counterparty measure is optimal.

Proposition 1-P *For any counterparty measure μ_F, there exists a cost function C such that μ_F satisfies the necessary conditions for optimality. Moreover, C is unique on support of μ_F, up to an additive constant.*

Proof. The first two steps of the proof proceed as in the equilibrium proof, skipping details where the two proofs are identical. We proceed more slowly through the last step, where we recover the cost function, because it is a bit different than in equilibrium.

Step 1: Recovering misalignment Since both the social surplus function and equilibrium surplus function are positive and decreasing, equation (87) implies that the inflow-outflow equation (22) reduces to equations (30) and (31). Using these, we can recover μ_M from the counterparty measure μ_F.

Step 2: Explicit Solution for the Surplus Function Use monotonicity of S to rewrite equation (12) as

$$\frac{(2r+4\gamma+\lambda)S(\lambda)-2\Delta}{\lambda} = S(\lambda)(F(\lambda)-2M(\lambda)) + \int_{(\lambda,\bar{\lambda}]} S(\lambda')(1-2m_{\lambda'})d\mu_F(\lambda'). \tag{89}$$

This is identical to equation (32), except for the numerical value of the coefficients. We can then replicate this step in the proof of Proposition 1 to prove that the unique solution to equation (89) is

$$S(\lambda) = \frac{\Delta}{r+2\gamma}\left(1-e^{-\int_\lambda^\infty \Phi_{\lambda'}d\lambda'}\right), \tag{90}$$

where

$$\Phi_\lambda \equiv \frac{2(r+2\gamma)}{\lambda\left(2(r+2\gamma)+\lambda(1-F(\lambda)+2M(\lambda))\right)}. \tag{91}$$

Step 3: Recovering the Cost Function This step differs from the equilibrium. To describe the optimal choice of contact rates, again return to the Lagrangian (85). The derivative with respect to ε_F is

$$\left.\frac{\partial \mathcal{L}}{\partial \varepsilon_F}\right|_{\varepsilon_F=0} = \Lambda\left(\int_{\mathcal{X}} \theta_\lambda d\nu_F(\lambda) - \int_{\mathcal{X}} \theta_\lambda d\mu_F(\lambda)\right)$$

where

$$\theta_\lambda \equiv -\frac{\Delta m_\lambda + rC(\lambda)}{\lambda} + \frac{\Lambda}{\lambda}\int_{\mathcal{X}} \frac{\Delta m_{\lambda'} + rC(\lambda')}{\lambda'} d\mu_F(\lambda') + \frac{1}{2}\int_{\mathcal{X}} S(\lambda')\Big(\big(\mathbb{I}^{\lambda,0}_{\lambda',0} m_\lambda + \mathbb{I}^{\lambda,1}_{\lambda',0}(1-m_\lambda)\big) m'_\lambda - \big(\mathbb{I}^{\lambda,0}_{\lambda',1} m_\lambda + \mathbb{I}^{\lambda,1}_{\lambda',1}(1-m_\lambda)\big)(1-m'_\lambda)\Big) d\mu_F(\lambda'). \quad (92)$$

Now assume $\Lambda > 0$ and let $\mathcal{Y}^P \equiv \arg\max_{\lambda\in\mathcal{X}} \theta_\lambda$. If $\mu_F(\mathcal{Y}^P) < 1$, there exists a measure ν_F with $\int_{\mathcal{X}} \theta_\lambda d\nu_F(\lambda) > \int_{\mathcal{X}} \theta_\lambda d\mu_F(\lambda)$ and hence $\frac{\partial \mathcal{L}}{\partial \varepsilon_F} > 0$, a contradiction. Thus assuming $\Lambda > 0$, a necessary condition for optimality is $\mu_F(\mathcal{Y}^P) = 1$.

Let $\bar{\theta} \equiv \max_{\lambda\in\mathcal{X}} \theta_\lambda$. Since $\mathcal{Y}^P = \arg\max_{\lambda\in\mathcal{X}} \theta_\lambda$ and $\mu_F(\mathcal{Y}^P) = 1$, it follows that $\bar{\theta} = \int_{\mathcal{X}} \theta_\lambda d\mu_F(\lambda)$. We compute this by integrating equation (92) and canceling the first two terms on the right hand side:

$$\begin{aligned}
\bar{\theta} &= \frac{1}{2}\int_{\mathcal{X}} S(\lambda') \int_{\mathcal{X}} \Bigg(\Big(\mathbb{I}^{\lambda,0}_{\lambda',0} m_\lambda + \mathbb{I}^{\lambda,1}_{\lambda',0}(1-m_\lambda)\Big) m_{\lambda'} \\
&\qquad - \Big(\mathbb{I}^{\lambda,0}_{\lambda',1} m_\lambda + \mathbb{I}^{\lambda,1}_{\lambda',1}(1-m_\lambda)\Big)(1-m_{\lambda'})\Bigg) d\mu_F(\lambda)\, d\mu_F(\lambda') \\
&= \frac{1}{2}\int_{\mathcal{X}} S(\lambda)\Bigg(\int_{\mathcal{X}} \Big(\mathbb{I}^{\lambda',0}_{\lambda,0} m_{\lambda'} + \mathbb{I}^{\lambda',1}_{\lambda,0}(1-m_{\lambda'})\Big) d\mu_F(\lambda') m_\lambda \\
&\qquad - \int_{\mathcal{X}} \Big(\mathbb{I}^{\lambda',0}_{\lambda,1} m_{\lambda'} + \mathbb{I}^{\lambda',1}_{\lambda,1}(1-m_{\lambda'})\Big) d\mu_F(\lambda')(1-m_\lambda)\Bigg) d\mu_F(\lambda)
\end{aligned}$$

The second equation swaps the role of λ and λ' in the integrals and then regroups terms. Using the inflow-outflow equation (22), we can then derive equation (15) in the text.

Next, rewrite equation (88) as

$$\begin{aligned}
\frac{\Delta m_\lambda}{\lambda} &= \frac{(r+2\gamma)S(\lambda)m_\lambda}{\lambda} \\
&\quad + \frac{1}{2}\Bigg(S(\lambda)m_\lambda \int_{\mathcal{X}} \Big(\Big(\mathbb{I}^{\lambda',0}_{\lambda,0} + \mathbb{I}^{\lambda',0}_{\lambda,1}\Big) m_{\lambda'} + \Big(\mathbb{I}^{\lambda',1}_{\lambda,0} + \mathbb{I}^{\lambda',1}_{\lambda,1}\Big)(1-m_{\lambda'})\Big) d\mu_F(\lambda') \\
&\qquad - \int_{\mathcal{X}} S(\lambda')\Big(\mathbb{I}^{\lambda',0}_{\lambda,1} m'_\lambda - \mathbb{I}^{\lambda',1}_{\lambda,1}(1-m_{\lambda'})\Big) d\mu_F(\lambda') \\
&+ \int_{\mathcal{X}} S(\lambda')\Big(\Big(\mathbb{I}^{\lambda',0}_{\lambda,0} m_\lambda + \mathbb{I}^{\lambda',0}_{\lambda,1}(1-m_\lambda)\Big) m_{\lambda'} - \Big(\mathbb{I}^{\lambda',1}_{\lambda,0} m_\lambda + \mathbb{I}^{\lambda',1}_{\lambda,1}(1-m_\lambda)\Big)(1-m_{\lambda'})\Big) d\mu_F(\lambda')\Bigg)
\end{aligned}$$

Use this to rewrite equation (92) as

$$\theta_\lambda = -\frac{m_\lambda\Big(r+2\gamma+\frac{\lambda}{2}\int_{\mathcal{X}}\Big(\Big(\mathbb{I}_{\lambda,0}^{\lambda',0}+\mathbb{I}_{\lambda,1}^{\lambda',0}\Big)m_{\lambda'}+\Big(\mathbb{I}_{\lambda,0}^{\lambda',1}+\mathbb{I}_{\lambda,1}^{\lambda',1}\Big)(1-m_{\lambda'})\Big)d\mu_F(\lambda')\Big)S(\lambda)}{\lambda}$$
$$-\frac{rC(\lambda)}{\lambda}+\frac{\Lambda}{\lambda}\int_{\mathcal{X}}\frac{\Delta m_{\lambda'}+rC(\lambda')}{\lambda'}d\mu_F(\lambda')$$
$$+\frac{1}{2}\int_{\mathcal{X}}S(\lambda')\Big(\mathbb{I}_{\lambda,1}^{\lambda',0}m'_\lambda-\mathbb{I}_{\lambda,1}^{\lambda',1}(1-m_{\lambda'})\Big)d\mu_F(\lambda').$$

Simplify the first term using the steady state equation (22) to get

$$\theta_\lambda = -\frac{\Big(\gamma+\frac{\lambda}{2}\int_{\mathcal{X}}\Big(\mathbb{I}_{\lambda,1}^{\lambda',0}m_{\lambda'}+\mathbb{I}_{\lambda,1}^{\lambda',1}(1-m_{\lambda'})\Big)d\mu_F(\lambda')\Big)S(\lambda)}{\lambda}$$
$$-\frac{rC(\lambda)}{\lambda}+\frac{\Lambda}{\lambda}\int_{\mathcal{X}}\frac{\Delta m_{\lambda'}+rC(\lambda')}{\lambda'}d\mu_F(\lambda')$$
$$+\frac{1}{2}\int_{\mathcal{X}}S(\lambda')\Big(\mathbb{I}_{\lambda,1}^{\lambda',0}m'_\lambda-\mathbb{I}_{\lambda,1}^{\lambda',1}(1-m_{\lambda'})\Big)d\mu_F(\lambda').$$

The optimal trading patterns in equation (87) give us

$$\theta_\lambda = -\frac{\gamma S(\lambda)}{\lambda}+\frac{1}{2}\int_{\mathcal{X}}\Big(\big(S(\lambda')-S(\lambda)\big)^+m'_\lambda+\big(-S(\lambda)-S(\lambda')\big)^+(1-m_{\lambda'})\Big)d\mu_F(\lambda')$$
$$-\frac{rC(\lambda)}{\lambda}+\frac{\Lambda}{\lambda}\int_{\mathcal{X}}\frac{\Delta m_{\lambda'}+rC(\lambda')}{\lambda'}d\mu_F(\lambda'). \quad (93)$$

Now define

$$\bar{\Pi}\equiv-\frac{\Lambda}{r}\int_{\mathcal{X}}\frac{\Delta m_{\lambda'}+rC(\lambda')}{\lambda'}d\mu_F(\lambda'), \quad (94)$$

so the value of the planner's objective function is $\delta_1+r\bar{\Pi}$. Then $\bar{\theta}\geq\theta_\lambda$ is equivalent to $\bar{\Pi}\geq\Pi_\lambda$ where Π_λ is defined in equation (14). Moreover, $\bar{\Pi}=\Pi_\lambda$ exactly when $\bar{\theta}=\theta_\lambda$. This means that the set $\mathcal{Y}^P\equiv\arg\max_{\lambda\in\mathcal{X}}\theta_\lambda$ also satisfies $\mathcal{Y}^P=\arg\max_{\lambda\in\mathcal{X}}\Pi_\lambda$.

Using the fact that the surplus function is positive-valued and decreasing, we can simplify equations (93) and (14) to get

$$\theta_\lambda = -\frac{\gamma S(\lambda)}{\lambda}+\frac{1}{2}\int_{[0,\lambda]}(S(\lambda')-S(\lambda))d\mu_M(\lambda')-\frac{rC(\lambda)}{\lambda}-\frac{r\bar{\Pi}}{\lambda}. \quad (95)$$

and

$$\Pi_\lambda=\frac{-\gamma S(\lambda)+\frac{\lambda}{2}\int_{[0,\lambda]}(S(\lambda')-S(\lambda))d\mu_M(\lambda')-\bar{\theta}\lambda}{r}-C(\lambda). \quad (96)$$

These will be useful later.

Finally, we can invert the requirement that $\bar{\Pi} \geq \Pi_\lambda$ with equality for $\lambda \in \mathcal{Y}^P$ to find a lower bound on the cost function:

$$C(\lambda) \geq \frac{-\gamma S(\lambda) + \frac{\lambda}{2}\int_{[0,\lambda]}(S(\lambda') - S(\lambda))d\mu_M(\lambda') - \bar{\theta}\lambda}{r} - \bar{\Pi}$$

with equality for $\lambda \in \mathcal{Y}^P$. We have already expressed all the objects on the right hand side of this equation, except the additive constant $\bar{\Pi}$, as functions of μ_F. Thus the counterparty distribution pins down the cost function up to an additive constant on the support of the counterparty distribution. ■

Proposition 2-P *Assume C is differentiable and C' is Lipschitz continuous. Then the optimum counterparty distribution F and contact rate distribution G are absolutely continuous on $[0, \bar{\lambda})$. If additionally $C'(0) < \frac{\Delta\gamma^2}{2r(r+2\gamma)^3}$, and $C'(\bar{\lambda}) \geq \Delta\frac{2\gamma^2\left(4(r+2\gamma)^2+8\gamma\bar{\lambda}-\bar{\lambda}^2\right)}{r\psi(\bar{\lambda}+2\psi)(r+2\gamma+\psi)^2(2r+4\gamma+\bar{\lambda})}$ where $\psi \equiv \sqrt{(r+2\gamma)^2+2\gamma\bar{\lambda}}$, then there is a positive measure of traders with contact rate in the interval $(0, \bar{\lambda})$ in the optimal counterparty distribution.*

Proof. Analogous to Proposition 2, we proceed in three steps. First we show that $F(\lambda)$ is continuous on $[0, \lambda)$, then we show that it is absolutely continuous, and finally characterize conditions that ensure that a positive measure of traders choose $\lambda \in (0, \bar{\lambda})$.

Step 1: No Discontinuities in F The first step of the proof immediately follows using the functional form of S and Φ in equations (90) and (91) and Π in equation (96).

Step 2: Absolute continuity of F This step also follows the same logic using the modified functional form of the (social) surplus equation with the following minor adjustment: using (91), we need to define $\varrho \equiv 2r + 4\gamma$. Step 2 of Proposition 2 then follows immediately.

Step 3: Positive measure of traders in the interval $(0, \bar{\lambda})$ We use a direct approach to find conditions which ensure $\mu_F(\{0\}) = 0$. Then we follow a similar approach to Proposition 2 to find conditions which ensure $\mu_F(\{\bar{\lambda}\}) < 1$.

First, assume there is an optimal allocation in which $\mu_F(\{0\}) > 0$. In any such allocation $\mu_G(\{0\}) = 1$. So generally consider an allocation with $\mu_G(\{\lambda\}) = 1$ for some $\lambda \in \mathcal{X}$. Equation (9) implies

$$m_\lambda = \frac{\sqrt{(r+2\gamma)^2+2\gamma\lambda} - (r+2\gamma)}{\lambda} \tag{97}$$

Then using equation (11), the value of a marginal increase in everyone's contact rate from $\mu_G(\{0\}) = 1$ to $\mu_G(\{\varepsilon\}) = 1$ is

$$\lim_{\varepsilon \to 0} \frac{\Delta(m_0 - m_\varepsilon) + r(C(0) - C(\varepsilon))}{\varepsilon} = \frac{\Delta\gamma^2}{2(r+2\gamma)^2} - rC'(0).$$

Thus, as in equilibrium, $C'(0) < \frac{\Delta\gamma^2}{2r(r+2\gamma)^2}$ rules out $\mu_G(\{0\}) = 1$ and hence rules out optimality of any allocation with $\mu_F(\{0\}) > 0$.

Now, suppose $\mu_F(\{\bar{\lambda}\}) = 1$. Then equation (89) implies that for $\lambda \in \mathcal{X}$,

$$S(\lambda) = \frac{\Delta\big(2r + 4\gamma + \lambda(1 - 2m_{\bar{\lambda}}) + 2\bar{\lambda}m_{\bar{\lambda}}\big)}{(2r + 4\gamma + \lambda)(r + 2\gamma + \bar{\lambda}m_{\bar{\lambda}})} \tag{98}$$

where $m_{\bar{\lambda}}$ satisfies equation (97). Moreover, equation (15) implies

$$\bar{\theta} = \frac{\Delta(\gamma - (r+2\gamma)m_{\bar{\lambda}})}{\bar{\lambda}(r + 2\gamma + \bar{\lambda}m_{\bar{\lambda}})}. \tag{99}$$

Finally, equation (96) implies that for $\lambda \in \mathcal{X}$,

$$\Pi_\lambda \equiv \frac{-\gamma S(\lambda) - \bar{\theta}\lambda}{r} - C(\lambda).$$

Using the previous expressions, we have that the left derivative of the profit function at $\bar{\lambda}$ satisfies

$$\Pi'_{\bar{\lambda}} = \frac{-\gamma S'(\bar{\lambda}) - \bar{\theta}}{r} - C'(\bar{\lambda}) = \Delta \frac{2\gamma^2\big(4(r+2\gamma)^2 + 8\gamma\bar{\lambda} - \bar{\lambda}^2\big)}{r\psi(\bar{\lambda} + 2\psi)(r + 2\gamma + \psi)^2(2r + 4\gamma + \bar{\lambda})} - C'(\bar{\lambda})$$

where $\psi \equiv \sqrt{(r+2\gamma)^2 + 2\gamma\bar{\lambda}}$. Thus if

$$C'(\bar{\lambda}) > \Delta \frac{2\gamma^2\big(4(r+2\gamma)^2 + 8\gamma\bar{\lambda} - \bar{\lambda}^2\big)}{r\psi(\bar{\lambda} + 2\psi)(r + 2\gamma + \psi)^2(2r + 4\gamma + \bar{\lambda})}, \tag{100}$$

$\Pi'_{\bar{\lambda}} < 0$. But this means that $\bar{\lambda}$ does not maximize Π_λ, hence that it is not optimal to set $\mu_F(\{\bar{\lambda}\}) = 1$. ■

Note that when $\bar{\lambda} > 4\gamma + 2\sqrt{r^2 + 4r\gamma + 8\gamma^2}$, the right hand side of condition (100) is negative. This means that if the upper bound on contact rates is sufficiently large, it is not optimal for everyone to set their contact rate at the upper bound even if doing so is free. This is a (perhaps unexpected) feature of our search technology. By reducing the contact rate for a few traders, the remaining traders may still efficiently intermediate for the slower

traders while concentrating their meetings among themselves. Nevertheless, we show below that in a limiting allocation with $\bar{\lambda} \to 0$ and a constant marginal cost of contacts, all contact rates exceed a lower bound; and when the marginal cost vanishes, the lower bound grows without bound.

Representation of Optimum as ODE System We again derive a first order ordinary differential equation system in (F, M, S) to characterize any optimal allocation when the cost function is twice continuously differentiable. We omit details when the logic is similar to the corresponding equilibrium system.

As a preliminary step, differentiate equation (90) and use equation (91) to get

$$S'(\lambda) = \frac{2\left((r+2\gamma)S(\lambda) - \Delta\right)}{\lambda\Big(2(r+2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))\Big)}, \tag{101}$$

$$S''(\lambda) = -\frac{2(1 - F(\lambda) + 2M(\lambda)) - \lambda(F'(\lambda) - 2M'(\lambda))}{2(r+2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda))} S'(\lambda). \tag{102}$$

Next, since $\Pi'_\lambda = 0$ on the support of the contact rate distribution, twice differentiating the definition of Π in equation (96) gives us

$$\Pi''_\lambda = \frac{-\big(2\gamma + \lambda M(\lambda)\big)S''(\lambda) - \big(2M(\lambda) + \lambda M'(\lambda)\big)S'(\lambda)}{2r} - C''(\lambda) = 0 \tag{103}$$

on the support if F. Substitute $S'(\lambda)$ and $S''(\lambda)$ from equations (101) and (102) into equation (103) to get one linear relationship between $F'(\lambda)$ and $M'(\lambda)$ in the optimal solution. As in equilibrium, equation (31) gives us a second such relationship. We can solve these to get

$$F'(\lambda) = \frac{\big(2r + 4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda))\big)\big(4\big(\gamma(1 - F(\lambda)) - rM(\lambda)\big) + Z(\lambda)\big)}{2\lambda\big(2r + 2\gamma + \lambda(1 - F(\lambda) + M(\lambda))\big)(2\gamma + \lambda M(\lambda))}, \tag{104}$$

$$M'(\lambda) = \frac{4\big(\gamma(1 - F(\lambda)) - rM(\lambda)\big) + Z(\lambda)}{2\lambda\big(2r + 2\gamma + \lambda(1 - F(\lambda) + M(\lambda))\big)} \tag{105}$$

where

$$Z(\lambda) \equiv \frac{r\lambda\big(2r + 4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda))\big)^2 C''(\lambda)}{\Delta - (r+2\gamma)S(\lambda)} \tag{106}$$

Equations (101), (104), and (105) are an ordinary differential equation system, valid on the interior of the support of F.

If we know $\underline{\lambda}$, we pin down the level of F, M, and S using the terminal conditions

$F(\underline{\lambda}) = M(\underline{\lambda}) = 0$ as well as

$$S(\bar{\lambda}) = \frac{\Delta}{r + 2\gamma + \bar{\lambda}M(\bar{\lambda})}, \tag{107}$$

which we obtain by evaluating equation (89) at $\lambda = \bar{\lambda}$.

Finally, to verify the choice of $\underline{\lambda}$, we use the first order condition $\Pi'_{\underline{\lambda}} = 0$,

$$-\gamma S'(\underline{\lambda}) = rC'(\underline{\lambda}) + \bar{\theta}, \tag{108}$$

where Π_λ is defined in equation (96) and $\bar{\theta}$ is defined in equation (15). This is the equivalent of equation (53).

D.2 Simpler expression for $\bar{\theta}$

To further characterize the optimum, we find it useful to relate $\bar{\theta}$ to the cost function. First, note that $\theta_\lambda = \bar{\theta}$ for all $\lambda \in \mathcal{Y}^P$ and integrate equation (95) under the absolutely continuous measure μ_F, up to but not including $\bar{\lambda}$:

$$(1 - dF(\bar{\lambda}))\bar{\theta} = \lim_{\lambda \to \bar{\lambda}} \Bigg(-\int_0^\lambda \frac{\gamma S(\lambda')}{\lambda'} dF(\lambda') + \frac{1}{2}\int_0^\lambda \int_0^{\lambda'} (S(\lambda'') - S(\lambda'))dM(\lambda'')dF(\lambda') \\ - \int_0^\lambda \frac{rC(\lambda')}{\lambda'} dF(\lambda') - r\bar{\Pi}\int_0^\lambda \frac{1}{\lambda'} dF(\lambda') \Bigg).$$

Note from equation (17) that

$$\lim_{\lambda \to \bar{\lambda}} \int_0^\lambda \frac{1}{\lambda'} dF(\lambda') = \frac{1}{\Lambda} - \frac{dF(\bar{\lambda})}{\bar{\lambda}}.$$

Then using equation (94), we have

$$(1 - dF(\bar{\lambda}))\bar{\theta} = \lim_{\lambda \to \bar{\lambda}} \left(\int_0^\lambda \frac{\Delta m_{\lambda'} - \gamma S(\lambda')}{\lambda'} dF(\lambda') + \frac{1}{2}\int_0^\lambda \int_0^{\lambda'} (S(\lambda'') - S(\lambda'))dM(\lambda'')dF(\lambda') \right) \\ + \left(\Delta m_{\bar{\lambda}} + rC(\bar{\lambda}) + r\bar{\Pi}\right) \frac{dF(\bar{\lambda})}{\bar{\lambda}} \tag{109}$$

This is the first key equation.

Second, on the interior of the support of F, we have that Π_λ is constant and hence $\Pi'_\lambda = 0$. Differentiating equation (96) gives

$$\bar{\theta} = -\left(\gamma + \frac{\lambda M(\lambda)}{2}\right) S'(\lambda) + \frac{1}{2}\int_0^\lambda (S(\lambda') - S(\lambda))dM(\lambda') - rC'(\lambda)$$

Equations (101) and (31) imply that on the interior of the support,

$$\left(\gamma + \frac{\lambda}{2}M(\lambda)\right) S'(\lambda) = \frac{\left((r+2\gamma)S(\lambda) - \Delta\right) m_\lambda}{\lambda},$$

and so the previous equation becomes

$$\bar{\theta} = -\frac{\left((r+2\gamma)S(\lambda) - \Delta\right) m_\lambda}{\lambda} + \frac{1}{2}\int_0^\lambda (S(\lambda') - S(\lambda))dM(\lambda') - rC'(\lambda)$$

Again integrate this under the measure $\mu_F(\lambda)$ up to but not including $\bar{\lambda}$:

$$\begin{aligned}(1 - dF(\bar{\lambda}))\bar{\theta} = \lim_{\lambda \to \bar{\lambda}} \Bigg(&- \int_0^\lambda \frac{\left((r+2\gamma)S(\lambda') - \Delta\right) m_{\lambda'}}{\lambda'} dF(\lambda') \\ &+ \frac{1}{2}\int_0^\lambda \int_0^{\lambda'} (S(\lambda'') - S(\lambda'))dM(\lambda'')dF(\lambda') - r\int_0^\lambda C'(\lambda')dF(\lambda')\Bigg). \quad (110)\end{aligned}$$

This is the second key equation.

Now equate the right hand sides of (109) and (110) to get

$$\begin{aligned}\int_0^{\bar{\lambda}} \frac{\gamma - (r+2\gamma)m_\lambda}{\lambda} S(\lambda)dF(\lambda) = &\, r\int_0^{\bar{\lambda}} C'(\lambda')dF(\lambda') \\ &+ \left(r\bar{\Pi} + \Delta m_{\bar{\lambda}} + r\big(C(\bar{\lambda}) - \bar{\lambda}C'(\bar{\lambda})\big) + \big(\gamma - (r+2\gamma)m_{\bar{\lambda}}\big)S(\bar{\lambda})\right)\frac{dF(\bar{\lambda})}{\bar{\lambda}}\end{aligned}$$

Equation (15) implies that the left hand side is $\bar{\theta}$. We can then regroup terms using equation (107) to eliminate $S(\bar{\lambda})$:

$$\bar{\theta} = r\int_0^{\bar{\lambda}} C'(\lambda')dF(\lambda') + \left(r\bar{\Pi} + \Delta\frac{\gamma + \bar{\lambda}M(\bar{\lambda})m_{\bar{\lambda}}}{r + 2\gamma + \bar{\lambda}M(\bar{\lambda})} + r\big(C(\bar{\lambda}) - \bar{\lambda}C'(\bar{\lambda})\big)\right)\frac{dF(\bar{\lambda})}{\bar{\lambda}}. \quad (111)$$

This is an explicit equation for $\bar{\theta}$. Notably when either $\bar{\lambda} \to \infty$ or $dF(\bar{\lambda}) = 0$, the last term vanishes and $\bar{\theta} = r\int_{\mathcal{X}} C'(\lambda)d\mu_F(\lambda)$, the average counterparty's marginal cost of contacts, as mentioned in the text.

D.3 Characterization with a Linear Cost Function

We now characterize the optimum with a linear cost function $C(\lambda) = c\lambda$ for $c > 0$.

Proposition 3-P *Assume $C(\lambda) = c\lambda$. Fix r, γ, Δ, and $\bar{\lambda}$. There exists thresholds $\bar{c} > \underline{c}$*

such that

$$\text{if } \begin{cases} c \geq \bar{c} \\ c \in (\underline{c}, \bar{c}) \\ c \leq \underline{c}, \end{cases} \quad \text{then there is a } \begin{cases} \textit{autarky allocation} \\ \textit{intermediated trade allocation} \\ \textit{degenerate trade allocation} \end{cases}$$

satisfying the necessary conditions for optimality. Moreover, the optimal allocation takes one of these three forms. In an autarky allocation, the average contact rate is $\Lambda = 0$*. In an intermediated trade allocation, the average contact rate is* $\Lambda \in (0, \bar{\lambda})$*; the support of the counterparty distribution is a convex interval* $[\underline{\lambda}, \bar{\lambda}]$ *with* $\underline{\lambda} \in (0, \bar{\lambda})$ *and* $dF(\bar{\lambda}) > 0$*; and the misalignment rate* m_λ *is increasing on* $[\underline{\lambda}, \bar{\lambda}]$*. In a degenerate trade allocation, the average contact rate is* $\Lambda = \bar{\lambda}$*.*

If $\bar{\lambda} \geq 4\gamma + 2\sqrt{r^2 + 4r\gamma + 8\gamma^2}$*,* $\underline{c} \leq 0$ *and hence the optimal allocation must be either autarky or intermediated trade.*

Proof. We follow the structure of the proof of Proposition 3, omitting repetitive details.

High Cost. We first redefine the cost thresholds $\bar{c}$ and $\bar{\bar{c}}$ for the planner's problem. Eliminate $S''(\lambda)$ from equation (103) using equation (102). When $F(\lambda)$ (and hence $M(\lambda)$ is constant for all $\lambda \in [0, \bar{\lambda})$, this implies that the profit function is globally convex if and only if $F(0) \geq \frac{r+2\gamma}{2(r+\gamma)}$. Smaller values of $F(0)$ are therefore inconsistent with optimality of an autarky allocation.

Next, in an autarky allocation, we have

$$r\left(\Pi_{\bar{\lambda}} - \Pi_0\right) = \left(\frac{\big(2\gamma + \bar{\lambda}M(0)\big)M(\bar{\lambda})\Delta}{2(r+2\gamma)\big(r + 2\gamma + \bar{\lambda}M(\bar{\lambda})\big)} - (rc + \bar{\theta}) \right) \bar{\lambda},$$

analogous to equation (55). We can further simplify by eliminating $\bar{\theta}$ using equation (111). The planner is willing to put support on both 0 and $\bar{\lambda}$ if

$$2rc = \frac{\big(2\gamma + \bar{\lambda}M(0)\big)M(\bar{\lambda})\Delta}{2(r+2\gamma)\big(r + 2\gamma + \bar{\lambda}M(\bar{\lambda})\big)} - \left(m_0 + \frac{\gamma + \bar{\lambda}M(\bar{\lambda})m_{\bar{\lambda}}}{r + 2\gamma + \bar{\lambda}M(\bar{\lambda})} \right) \frac{\Delta}{\bar{\lambda}} dF(\bar{\lambda}). \tag{112}$$

Setting $F(0) = \frac{r+2\gamma}{2r+2\gamma}$, $M(0) = m_0 F(0) = \frac{\gamma}{2r+2\gamma}$, $dF(\bar{\lambda}) = 1 - F(0) = \frac{r}{2r+2\gamma}$, and $M(\bar{\lambda}) = M(0) + m_{\bar{\lambda}}\mu_F(\{\bar{\lambda}\})$, with $m_{\bar{\lambda}}$ defined in equation (30), gives us the threshold $\bar{c}$, above which $\Lambda = 0$ is an optimal allocation. Setting $F(0) = 1$ and $M(0) = M(\bar{\lambda}) = m_0 = \frac{\gamma}{r+2\gamma}$ gives us the threshold $\bar{\bar{c}}$, above which setting $F(\lambda) = 1$ for all λ is an optimal allocation.

Low Cost. We next characterize an intermediated trade optimum, $\mu_F(\{\bar{\lambda}\}) = 1$. For such an allocation to be optimal, it must be the case that $\Pi_\lambda \leq \Pi_{\bar{\lambda}}$ for all $\lambda \in \mathcal{X}$. Following the proof of Proposition 2-P, this is equivalent to

$$c \leq \frac{\gamma \frac{S(\lambda) - S(\bar{\lambda})}{\bar{\lambda} - \lambda} - \bar{\theta}}{r}$$

for all λ. Eliminate $S(\lambda)$ using equation (98) and $\bar{\theta}$ using equation (99), both applicable to the case with $\mu_F(\{\bar{\lambda}\}) = 1$. It is easy to show that the right hand side of the previous equation is decreasing in λ. This means that the inequality is most strict in the limit as $\lambda \to \bar{\lambda}$. That is, an intermediated trade allocation is optimal if and only if

$$c \leq \frac{-\gamma S'(\bar{\lambda}) - \bar{\theta}}{r} = \Delta \frac{2\gamma^2\big(4(r+2\gamma)^2 + 8\gamma\bar{\lambda} - \bar{\lambda}^2\big)}{r\psi(\bar{\lambda} + 2\psi)(r + 2\gamma + \psi)^2(2r + 4\gamma + \bar{\lambda})} = \underline{c},$$

where again $\psi \equiv \sqrt{(r+2\gamma)^2 + 2\gamma\bar{\lambda}}$. This is the same as condition (100), adapted to an environment with a linear cost.

Note that if $\bar{\lambda} \geq 4\gamma + 2\sqrt{r^2 + 4r\gamma + 8\gamma^2}$, $\underline{c} \leq 0$, so this region vanishes when the cost function is nondecreasing.

Intermediate Cost. Now assume $c \in (\underline{c}, \bar{c})$. We look for an optimal allocation described by the solution to an initial value problem. With $C''(\lambda) = Z(\lambda) = 0$, equations (104) and (105) are a pair of ordinary differential equations in F and M. Given a lower bound $\underline{\lambda}$, we can solve these and verify that the solutions are increasing with $\lim_{\lambda \to \bar{\lambda}} F(\lambda) < 1$.

Optimality dictates that the lower bound must satisfy equation (108). Eliminating $\bar{\theta}$ using equation (111) gives

$$2rc = -\gamma S'(\underline{\lambda}) - \left(r\bar{\Pi} + \Delta \frac{\gamma + \bar{\lambda} M(\bar{\lambda}) m_{\bar{\lambda}}}{r + 2\gamma + \bar{\lambda} M(\bar{\lambda})} \right) \frac{\mu_F(\{\bar{\lambda}\})}{\bar{\lambda}}. \tag{113}$$

We again use continuity of the right hand side of this expression, as well as the intermediate value theorem, to prove the existence of an optimal allocation in the same fashion as equilibrium. ■

Definition 2-P *Assume $C(\lambda) = c\lambda$. Fix r, γ, Δ, and c. For any $\bar{\lambda}$, let $(\mu_{F,\bar{\lambda}}, m_{\bar{\lambda}}, S_{\bar{\lambda}})$ satisfy the necessary conditions for an optimal allocation when the maximum contact rate is $\bar{\lambda}$ and as usual let $F_{\bar{\lambda}}(\lambda) = \mu_{F,\bar{\lambda}}([0, \lambda])$ for all $\lambda \leq \bar{\lambda}$. Also extend the definition of $(F_{\bar{\lambda}}, m_{\bar{\lambda}}, S_{\bar{\lambda}})$ to the positive reals in an arbitrary way. (F, m, S) with domain $[0, \infty)^3$ is a* limiting optimum

if there exists an increasing unbounded sequence $\{\bar{\lambda}_n\}$ *with associated* $(F_{\bar{\lambda}_n}, m_{\bar{\lambda}_n}, S_{\bar{\lambda}_n})$ *which converges pointwise to* (F, m, S)*.*

Lemma 2-P *Assume* $C(\lambda) = c\lambda$*. If* $c \geq \bar{c}^* \equiv \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)}$*, a limiting (autarky) optimum with* $F(\lambda) = F(0) > 0$ *for all* λ *exists. If* $c < \bar{c}^*$*, a limiting (intermediated trade) optimum with* $F(\underline{\lambda}) = 0$ *and* F *strictly increasing on* $[\underline{\lambda}, \infty)$ *for some* $\underline{\lambda} > 0$ *exists. Moreover, any limiting optimum takes one of these forms.*

Proof. To prove the planner's version of Lemma 2, the key step we need to reproduce is that $\bar{c}_n$ converges to $\bar{c}^*$ from below, where $\bar{c}^*$ is the cost threshold for the existence of a limiting optimum with $\Lambda > 0$. Expanding equation (112), the threshold $\bar{c}_n$ satisfies

$$\bar{c}_n = \frac{\left(2 + \frac{\bar{\lambda}}{2r+2\gamma}\right) M(\bar{\lambda})\gamma\Delta}{4r(r+2\gamma)\left(r + 2\gamma + \bar{\lambda}M(\bar{\lambda})\right)} - \left(m_0 + \frac{\gamma + \bar{\lambda}M(\bar{\lambda})m_{\bar{\lambda}}}{r + 2\gamma + \bar{\lambda}M(\bar{\lambda})}\right)\frac{\Delta}{4(r+\gamma)\bar{\lambda}}.$$

The second term is positive and converges to zero when $\bar{\lambda} \to \infty$. Moreover,

$$\frac{\left(2 + \frac{\bar{\lambda}}{2r+2\gamma}\right) M(\bar{\lambda})\gamma\Delta}{4r(r+2\gamma)\left(r + 2\gamma + \bar{\lambda}M(\bar{\lambda})\right)} \leq \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)} = \bar{c}^*,$$

as can be confirmed algebraically. The inequality also binds in the limit as $\bar{\lambda} \to \infty$ (since $M(\bar{\lambda}) \to M(0) > 0$). This proves that $\bar{c}_n$ converges from below to $\bar{c}^*$, the same cost threshold as in equilibrium. The remainder of the proof of this Lemma is unchanged. ■

Proposition 4-P *Assume* $C(\lambda) = c\lambda$ *with* $c < \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)}$*. Then in a limiting optimum there are* middlemen*, meaning* $\lim_{\lambda\to\infty} F(\lambda) < 1$*; and the contact rate distribution has a* Pareto tail with tail index 2*, meaning* $\lim_{\lambda\to\infty} \lambda^2(1 - G(\lambda))$ *is positive and finite.*

Proof. In a limiting optimum, condition (113) reduces to $2rc = -\gamma S'(\underline{\lambda})$. We can use this to obtain a lower bound on $\underline{\lambda}$ for fixed c, and prove that the lower bound converges to infinity as c goes to zero, as in the limiting equilibrium.

We first prove the existence of middlemen. Similar to the proof for equilibrium, we prove that for fixed r and γ, there exists $F^* > 0$, independent of c, Δ, and $\bar{\lambda}$, such that in any allocation that satisfies the necessary first order conditions for optimality, with $F(0) = 0$, $\mu_F(\{\bar{\lambda}\}) \geq F^*$. It follows that there are middlemen in any limiting optimum.

We focus here on the proof that $\lim_{\lambda\to\infty} F(\lambda)$ is bounded below 1 even when $\underline{\lambda} \to \infty$. We work with the initial value problem (104)–(105) with $Z(\lambda) = 0$ and apply the same

transformation of variables as in equilibrium: let $\rho \equiv \lambda/\underline{\lambda}$, $H(\rho) \equiv F(\rho\underline{\lambda})$ and $L(\rho) \equiv \rho\underline{\lambda}M(\rho\underline{\lambda})$, with $H(1) = L(1) = 0$. In the limit as $\underline{\lambda} \to \infty$, the initial value problem becomes

$$H'(\rho) = \frac{2\gamma(1 - H(\rho))}{\rho(2\gamma + L(\rho))} \text{ and } L'(\rho) = \frac{2\gamma + L(\rho)}{\rho},$$

with $L(1) = H(1) = 0$. These differential equations can be solved in closed form,

$$L(\rho) = 2\gamma(\rho - 1) \text{ and } H(\rho) = 1 - e^{\rho^{-1}-1}. \tag{114}$$

Thus $\lim_{\rho\to\infty} H(\rho) = 1 - e^{-1} \approx 0.632 < 1$ which, similar to equilibrium, establishes the bound in the limit.

Finally, the argument for the Pareto tail is exactly analogous to equilibrium. We find that

$$\lim_{n\to\infty} \lim_{\lambda\to\bar{\lambda}_n} \lambda^2 F_n'(\lambda) = 2 \lim_{n\to\infty} \frac{(1 - \bar{F}_n + 2\bar{M}_n)(\gamma(1 - \bar{F}_n) - r\bar{M}_n)}{\bar{M}_n(1 - \bar{F}_n + \bar{M}_n)},$$

where $\bar{F}_n \equiv \lim_{\lambda\to\bar{\lambda}_n} F_n(\lambda)$ and $\bar{M}_n \equiv \lim_{\lambda\to\bar{\lambda}_n} M_n(\lambda)$, analogous to equation (74). We can again prove that the right hand side has a well-behaved limit and so the density F' is asymptotically proportional to λ^{-2}.

The argument translating this from the density of counterparties to the cumulative distribution of contact rates follows the logic in the proof of Proposition 4. ■

Corollary 1-P *Assume $C(\lambda) = c\lambda$ with $c < \frac{\gamma\Delta}{8r(r+\gamma)(r+2\gamma)}$. In a limiting optimum, the fraction of trades with middlemen is strictly positive; and the trading rate distribution has a Pareto tail with tail index 2, meaning $\lim_{\alpha\to\infty} \alpha^2(1 - \hat{G}(\alpha))$ is positive and finite.*

We omit the proof, since it is unchanged from Corollary 1.

Proposition 5-P *Assume $C(\lambda) = c\lambda$. Consider a sequence of limiting optima as c converges to zero. The aggregate trading volume $\mathcal{V}$ converges to approximately 2.22γ and can be decomposed as follows: middlemen's purchases from other middlemen account for a volume $\mathcal{V}_{mm} = \frac{1}{2}\gamma$; middlemen's purchases from non-middlemen account for a volume of $\mathcal{V}_{mn} = \frac{1}{2}\gamma$; non-middlemen's purchases from middlemen account for a volume $\mathcal{V}_{nm} = \frac{1}{2}\gamma$; and non-middlemen's purchases from non-middlemen account for a volume $\mathcal{V}_{nn} \approx 0.72\gamma$.*

Proof. We focus on the behavior of the functions L and H, corresponding to an allocation with $\underline{\lambda} \to \infty$, defined in equation (114). The ratio of the average contact rate to the lower

bound converges to:

$$\frac{\Lambda}{\underline{\lambda}} = \frac{1}{\underline{\lambda}\int_{\mathcal{X}} \frac{1}{\lambda} d\mu_F(\lambda)} = \frac{1}{\int_1^\infty \frac{1}{\rho} H'(\rho) d\rho} = e \approx 2.718,$$

where we do the usual change in variables and then take advantage of the known functional form of H.

Next, let $\Psi(\rho) \equiv G(\rho\underline{\lambda})$ denote the cumulative distribution of relative contact rates. Using equation (16) and the known functional form of H, we have

$$\Psi'(\rho) = \rho^{-3} e^{\rho^{-1}} \Rightarrow \Psi(\rho) = (1 - \rho^{-1}) e^{\rho^{-1}}, \tag{115}$$

where the result follows by integrating the density function. This is an explicit solution for the distribution of relative contact rates in the limiting economy.

Next, we turn to volume. We start with the volume of purchases by non-middlemen. Analogous to equation (82) in equilibrium, we get that this is $\mathcal{V}_{nn}+\mathcal{V}_{nm} \to (e-3/2)\gamma \approx 1.22\gamma$.

The logic behind the other trading rates is unchanged. In particular, middlemen buy from non-middlemen a volume $\mathcal{V}_{mn} \to \gamma/2$. They sell to them at the same rate, and so non-middlemen buy from middlemen a volume $\mathcal{V}_{nm} \to \gamma/2$. This implies $\mathcal{V}_{mm} \to (e-2)\gamma$. Finally, the volume of trades between middlemen is unchanged, $\mathcal{V}_{mm} \to \gamma/2$. ■

E Comparison of Equilibrium and Optimum

E.1 Pigouvian Subsidies

Proposed Mechanism We propose that a trader who chooses a contact rate λ receives a net subsidy

$$\sigma_1(\lambda) \equiv \frac{\lambda}{4} \int_{\mathcal{X}} \Big(m_{\lambda'} (S(\lambda') - S(\lambda))^+ + (1 - m_{\lambda'})(-S(\lambda') - S(\lambda))^+ \Big) d\mu_F(\lambda') - \bar{\theta}\lambda \tag{116}$$

per unit of time spent in a well-aligned state and a net subsidy

$$\sigma_0(\lambda) \equiv \frac{\lambda}{4} \int_{\mathcal{X}} \Big(m_{\lambda'} (S(\lambda') + S(\lambda))^+ + (1 - m_{\lambda'})(-S(\lambda') + S(\lambda))^+ \Big) d\mu_F(\lambda') - \bar{\theta}\lambda \tag{117}$$

per unit of time in a misaligned state. If $\sigma_a(\lambda) < 0$, we interpret $-\sigma_a(\lambda)$ as a state-contingent net tax. The social surplus function S as well as $\bar{\theta}$ are defined by the solution to the planner's

problem.

To see how this affects trading and investments, we extend equation (24) to incorporate this subsidy:

$$\begin{aligned}
rv_{\lambda,h,1} =&\delta_{h,1} + \gamma(v_{\lambda,l,1} - v_{\lambda,h,1}) + \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}(v_{\lambda,h,0} + v_{\lambda',h,1} - v_{\lambda,h,1} - v_{\lambda',h,0})^+ \\
&+ (1-m_{\lambda'})(v_{\lambda,h,0} + v_{\lambda',l,1} - v_{\lambda,h,1} - v_{\lambda',l,0})^+\Big) d\mu_F(\lambda') + \sigma_1(\lambda), \qquad (118a)\\
rv_{\lambda,h,0} =&\delta_{h,0} + \gamma(v_{\lambda,l,0} - v_{\lambda,h,0}) + \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}(v_{\lambda,h,1} + v_{\lambda',h,0} - v_{\lambda,h,0} - v_{\lambda',h,1})^+ \\
&+ (1-m_{\lambda'})(v_{\lambda,h,1} + v_{\lambda',l,0} - v_{\lambda,h,0} - v_{\lambda',l,1})^+\Big) d\mu_F(\lambda') + \sigma_0(\lambda), \qquad (118b)\\
rv_{\lambda,l,1} =&\delta_{l,1} + \gamma(v_{\lambda,h,1} - v_{\lambda,l,1}) + \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}(v_{\lambda,l,0} + v_{\lambda',h,1} - v_{\lambda,l,1} - v_{\lambda',h,0})^+ \\
&+ (1-m_{\lambda'})(v_{\lambda,l,0} + v_{\lambda',l,1} - v_{\lambda,l,1} - v_{\lambda',l,0})^+\Big) d\mu_F(\lambda') + \sigma_0(\lambda), \qquad (118c)\\
rv_{\lambda,l,0} =&\delta_{l,0} + \gamma(v_{\lambda,h,0} - v_{\lambda,l,0}) + \frac{\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}(v_{\lambda,l,1} + v_{\lambda',h,0} - v_{\lambda,l,0} - v_{\lambda',h,1})^+ \\
&+ (1-m_{\lambda'})(v_{\lambda,l,1} + v_{\lambda',l,0} - v_{\lambda,l,0} - v_{\lambda',l,1})^+\Big) d\mu_F(\lambda') + \sigma_1(\lambda). \qquad (118d)
\end{aligned}$$

We conjecture that $v_{\lambda,h,1} - v_{\lambda,h,0} = S(\lambda) + q$ and $v_{\lambda,l,0} - v_{\lambda,l,1} = S(\lambda) - q$ for some number q. This ensures that trades occur in exactly the same situation as the planner would like. Moreover, by adding equations (118a) and (118d) and then subtracting equations (118b) and (118c), we obtain equation (12). Thus the tax ensures that the equilibrium and optimal surplus functions are identical. Moreover, averaging equations (118a) and (118d) to compute $\pi_\lambda = \frac{v_{\lambda,h,1}+v_{\lambda,l,0}}{2} - C(\lambda)$, we get that the choice of contact rate maximizes $r\Pi_\lambda - \delta_1$, where Π_λ is defined in equation (14). That is, the optimal distribution of contact rates is an equilibrium of the model with the proposed subsidy.

In closing, we note that the subsidies in equations (116)–(117) are isomorphic to doubling the surplus in every meeting but also charging a tax $\bar{\theta}$ for each meeting.

Terms of Trade Consider the difference between the subsidy to a misaligned trader and the subsidy to a well-aligned trader, $\sigma_0(\lambda) - \sigma_1(\lambda)$. Use equations (116) and (117) to write this out explicitly and then simplify using equation (12):

$$\sigma_0(\lambda) - \sigma_1(\lambda) = \frac{\Delta - (r+2\gamma)S(\lambda)}{2}.$$

This is non-negative and strictly increasing in λ. That is, for all traders, the Pigouvian subsidy provides more support to traders in the misaligned compared to the well-aligned

state. Further, for faster traders, the difference in subsidies across the two states is larger.

The state-contingent subsidies have an impact on the prices due to Nash bargaining. In particular, many trades involve a fast trader intermediating for a slow trader. The Pigouvian subsidies pay the fast trader for taking on misalignment. The slower trader then extracts some of that subsidy during Nash bargaining. Conversely, subsidies fall when the fast trader becomes well-aligned, so the Pigouvian subsidies shift the terms of trade towards the faster trader in a fundamental trade involving two misaligned traders. Still, since most traders are well-aligned, the former effect outweighs the latter, and so Pigouvian subsidies on average improve the terms of trade enjoyed by the slowest traders. That is, Pigouvian subsidies manipulate trading prices in manner that discourages traders from investing in their contact rate.

Average Subsidy We can also compute the average subsidy per unit of time going to a type λ trader, $\sigma(\lambda) \equiv m_\lambda \sigma_0(\lambda) + (1-m_\lambda)\sigma_1(\lambda)$, where σ_1 and σ_0 are defined in equations (116) and (117) and m_λ is the type-specific misalignment rate:

$$\begin{aligned}\sigma(\lambda) = \frac{\lambda}{4}\int_{\mathcal{X}} \Big(& m_{\lambda'}(S(\lambda') - S(\lambda))^+ + (1-m_{\lambda'})(-S(\lambda') - S(\lambda))^+ \Big) d\mu_F(\lambda') \\ & \frac{\lambda m_\lambda}{4}\int_{\mathcal{X}} \Big(m_{\lambda'}\big((S(\lambda') + S(\lambda))^+ - (S(\lambda') - S(\lambda))^+\big) \\ & \quad + (1-m_{\lambda'})\big((-S(\lambda') + S(\lambda))^+ - (-S(\lambda') - S(\lambda))^+\big)\Big) d\mu_F(\lambda') - \bar{\theta}\lambda.\end{aligned}$$

Use equation (14) to eliminate the first integral and equation (12) to eliminate the second integral. The optimal subsidy for such contact rates is

$$\sigma(\lambda) = \frac{r\Pi_\lambda + \Delta m_\lambda + rC(\lambda) + \big(\gamma - (r+2\gamma)m_\lambda\big)S(\lambda) - \bar{\theta}\lambda}{2}. \tag{119}$$

Now differentiate this at any point $\lambda \in \mathcal{Y}$. At points selected by a measure zero of traders, we first eliminate $m_\lambda = M'(\lambda)/F'(\lambda)$ using equation (31). We then differentiate the resulting expression and eliminate F' and M' using equations (104) and (105) and S' using equation (101). This gives

$$\sigma'(\lambda) = \frac{r(C'(\lambda) + \lambda C''(\lambda)) - \bar{\theta}}{2}. \tag{120}$$

In the linear cost case, $C(\lambda) = c\lambda$, and so equation (120) reduces to $\sigma'(\lambda) = (rc - \bar{\theta})/2$. Since $\bar{\theta} - rc$ converges to zero when $\bar{\lambda} \to \infty$ (equation 111), it follows that the marginal subsidy $\sigma'(\lambda)$ converges to zero in this limit. That is, the Pigouvian subsidy scheme does not

manipulate investments by transferring resources across traders with different contact rates. Instead, it does so exclusively by manipulating the terms of trade through the differential subsidy in the mis- and well-aligned states.

Finally, consider the population average subsidy, $\bar{\sigma} \equiv \int_{\mathcal{X}} \sigma_\lambda d\mu_G(\lambda) = \Lambda \int_{\mathcal{X}} (\sigma(\lambda)/\lambda) d\mu_F(\lambda)$. Recall that $\int_{\mathcal{X}} \Pi_\lambda d\mu_F(\lambda) = \bar{\Pi}$ since the planner only uses contact rates that maximize Π_λ. We can then use equation (94) to eliminate $\bar{\Pi}$ and equation (15) to eliminate $\bar{\theta}$ from equation (119), so the Pigouvian subsidy is consistent with a balance budget.

E.2 Distortions to Investment

Assume the cost function is linear, $C(\lambda) = c\lambda$, and consider a limiting optimal counterparty distribution F. We prove that with this allocation and cost function, the equilibrium profit function π_λ is strictly increasing in λ and unbounded above, so traders prefer an unboundedly large contact rate. Put differently, the optimal allocation suppresses investment in contacts.

First differentiate equation (96) to get

$$r\Pi'_\lambda = -\left(\gamma + \frac{\lambda M(\lambda)}{2}\right) S'(\lambda) + \frac{1}{2}\int_0^\infty (S(\lambda') - S(\lambda)) dM(\lambda') - \bar{\theta} - rc = 0,$$

where we use the necessary condition for optimality $\Pi'_\lambda = 0$. Equation (111) implies $\bar{\theta} = rc$ when $\bar{\lambda} \to \infty$. Additionally, for sufficiently large λ, equation (90) implies $S(\lambda) \to 0$ and equation (101) implies $\lambda S'(\lambda) \to 0$. Thus we get

$$\int_0^\infty S(\lambda') dM(\lambda') = 4rc \tag{121}$$

in a limiting optimal allocation.

Similarly, differentiate equation (37) to get

$$r\pi'_\lambda = -\left(\gamma + \frac{\lambda M(\lambda)}{4}\right) s'(\lambda) + \frac{1}{4}\int_0^\lambda (s(\lambda') - s(\lambda)) dM(\lambda') - rc.$$

Again, for sufficiently large λ, equation (35) implies $s(\lambda) \to 0$ and equation (34) implies $\lambda s'(\lambda) \to 0$. Using equation (121) to eliminate c, we obtain

$$\lim_{\lambda\to\infty} \pi'_\lambda = \frac{1}{4r}\int_0^\infty (s(\lambda') - S(\lambda')) dM(\lambda').$$

A comparison of equations (36) and (91) implies $\phi_\lambda > \Phi_\lambda$ for all $\lambda > 0$. Hence equations (35) and (90) imply $s(\lambda) > S(\lambda)$ for all $\lambda > 0$. This proves $\lim_{\lambda\to\infty} \pi'_\lambda > 0$.

Next twice differentiate equation (37) at any $\lambda \geq \underline{\lambda}$. Eliminate s'' using equation (43) and F' and M' using equation (104) and (105). We obtain

$$r\pi''_\lambda = \frac{\lambda\big(\gamma(1-F(\lambda)-rM(\lambda))\big)\big(\gamma(1-F(\lambda))+(r+2\gamma)M(\lambda)+\lambda M(\lambda)(1-F(\lambda)+M(\lambda))\big)s'(\lambda)}{\big(2\gamma+\lambda M(\lambda)\big)\big(2r+2\gamma+\lambda(1-F(\lambda)+M(\lambda))\big)\big(4r+8\gamma+\lambda(1-F(\lambda)+2M(\lambda))\big)}.$$

Since s is decreasing and the other terms are all positive (recall, in particular, that we have already shown that $\gamma(1-F(\lambda)) > rM(\lambda)$), this implies that the profit function is concave on $[\underline{\lambda}, \infty)$.

Finally, consider $\lambda \leq \underline{\lambda}$, where $F(\lambda) = M(\lambda) = 0$. Equation (37) implies $\pi_\lambda = -\gamma s(\lambda) - rc\lambda$. Additionally, equation (43) implies $s''(\lambda) = -\frac{2}{4(r+2\gamma)+\lambda}s'(\lambda) > 0$. Convexity of s implies concavity of π_λ for $\lambda \leq \underline{\lambda}$. Thus the profit function is globally concave.

Global concavity of the profit function implies that for all $\lambda \geq 0$, $\pi'_\lambda > \lim_{\lambda' \to \infty} \pi'_{\lambda'}$. Since we have shown that this limit is strictly positive, it follows that the profit function is strictly increasing with slope bounded above 0.

F Constrained Economy: The Role of Intermediation

We start by defining an equilibrium without intermediation analogously to the definition of equilibrium with intermediation.

Definition 3 *An equilibrium is a measure $\mu_F(S)$ which gives the probability that a counterparty's contact rate falls into the set $S \subseteq \mathcal{X}$, a misalignment rate function $m : \mathcal{X} \to [0,1]$, and a surplus function $s : \mathcal{X} \to \mathbb{R}$, satisfying:*

1. *the surplus equation adjusted for the no-intermediation case,*

$$\Delta = (r+2\gamma)s(\lambda) + \frac{\lambda}{4}\int_{\mathcal{X}} \Big((s(\lambda)+s(\lambda'))^+ m_{\lambda'} - (-s(\lambda)-s(\lambda'))^+(1-m_{\lambda'})\Big) d\mu_F(\lambda') \quad (122)$$

2. *the balanced inflow-outflow adjusted for the no-intermediation case,*

$$\left(r+\gamma+\frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{s(\lambda)+s(\lambda')>0} m_{\lambda'}\right) d\mu_F(\lambda')\right) m_\lambda = \left(\gamma + \frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{s(\lambda)+s(\lambda')<0}(1-m_{\lambda'})\right) d\mu_F(\lambda')\right)(1-m_\lambda). \quad (123)$$

3. *optimality of the ex-ante investment decision: $\mu_F(\mathcal{Y}^{NI}) = 1$, where $\mathcal{Y}^{NI} \equiv \arg\max_{\lambda \in \mathcal{X}} \pi^{NI}_\lambda$*

and

$$\pi_\lambda^{NI} = \frac{\delta_1 - \gamma s(\lambda) + \frac{\lambda}{4}\int_{\mathcal{X}} \left((-s(\lambda) - s(\lambda'))^+ (1 - m_{\lambda'})\right) d\mu_F(\lambda')}{r} - C(\lambda). \tag{124}$$

Note that there are two relevant types of meetings in this constrained economy, those between two misaligned traders with opposite asset holdings, and those between two well-aligned traders with opposite asset holdings. We can extend our earlier results to prove that the first type of meeting results in trade while the second does not. That is, in equilibrium, two well-aligned agents never jointly trade into misalignment.

We turn next to the planner's problem. The planner again chooses $\mu_F(S)$, the time-invariant probability that conditional on a meeting, the counterparty's contact rate is some $\lambda \in S$, along with the admissible set of trades; as in equilibrium, the planner is subject to the constraint that intermediation is not allowed. The objective of the planner is unchanged, given by equation (84). Since we are interested in the case where intermediation is not allowed the planner is subject to an adjusted constraint on the evolution of the misalignment rate,

$$\left(r + \gamma + \frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{\lambda,0}^{\lambda',0} m_{\lambda'}\right) d\mu_F(\lambda')\right) m_\lambda = \left(\gamma + \frac{\lambda}{2}\int_{\mathcal{X}} \left(\mathbb{I}_{\lambda,1}^{\lambda',1}(1 - m_{\lambda'})\right) d\mu_F(\lambda')\right)(1 - m_\lambda). \tag{125}$$

Proof of Proposition 6.

Equilibrium. We first prove that equation (122) defines the surplus function to be positive valued, as in the proof of Lemma 1. Then we can solve the equation to get

$$s(\lambda) = \frac{4\Delta - \lambda\int_{\mathcal{X}} s(\lambda') d\mu_M(\lambda')}{4(r + 2\gamma) + \lambda M(\bar{\lambda})}, \tag{126}$$

a decreasing and convex function. Condition 3 then implies traders choose λ to maximize $-\gamma s(\lambda)/r - C(\lambda)$. If $C(\lambda)$ is weakly convex, all traders choose the same value of $\lambda = \Lambda$.

Planner. Replicating the argument in Online Appendix D.1, we can write down the Lagrangian and find the social surplus function. It is again positive-valued, hence trade only occurs between misaligned traders. Moreover, the optimal surplus function satisfies

$$\Delta = (r + 2\gamma)S(\lambda) + \frac{\lambda}{2}\int_{\mathcal{X}} (S(\lambda) + S(\lambda')) d\mu_M(\lambda') \Rightarrow S(\lambda) = \frac{2\Delta - \lambda\int_{\mathcal{X}} S(\lambda') d\mu_M(\lambda')}{2(r + 2\gamma) + \lambda M(\bar{\lambda})},$$

decreasing and convex. Using the same variational argument as in Online Appendix D.1, we also obtain that the planner's puts weight on λ only if it maximizes

$$\frac{-\gamma S(\lambda) - \lambda\bar{\theta}}{r} - C(\lambda),$$

analogous to condition (96). Convexity of S implies that if the cost function is convex, the planner places all weight on a single value of $\lambda = \Lambda$. ■