The Emergence of Market Structure

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

December 18, 2018

Abstract

What market structure emerges when market participants can choose the rate at which they contact others? We show that traders who choose a higher contact rate emerge as intermediaries, earning profits by taking asset positions that are misaligned with their preferences. The endogenous distribution of contact rates has no mass points, giving rise to intermediation chains. When search costs are linear, the contact rate distribution has a Pareto tail with tail index two, and middlemen emerge: a positive fraction of meetings are with traders with unboundedly high contact rates. As search costs vanish, traders still make dispersed investments and trade occurs in intermediation chains, yielding a theory of intermediation in frictionless markets. These features arise both in equilibrium and in the optimal allocation.

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

---

*This paper was previously circulated under the title “Meeting Technologies in Decentralized Asset Markets”. We are grateful to Fernando Alvarez, Markus Brunnermeier, Xavier Gabaix, Ricardo Lagos, Pierre-Olivier Weill, Randall Wright, and audiences at various seminars and conferences for their thoughts and comments. We are also grateful for comments from the editor and four anonymous referees.

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

This paper examines an over-the-counter market for assets where traders periodically meet in pairs with the opportunity to trade (Rubinstein and Wolinsky, 1987). We are interested in understanding the origins and implications of the observed heterogeneity in these markets, whereby we mean that some individuals trade much more frequently and with many more partners than others do. In particular, real world trading networks appear to have a core-periphery structure. Traders at the core of the network act as financial intermediaries, earning profits by taking either side of a trade, while traders in the periphery trade less frequently and their trades are more geared towards obtaining an asset position aligned with their portfolio needs.

We develop an infinite horizon overlapping generations model in which ex-ante identical traders make an irreversible decision to invest in their contact rate at the moment that they enter the market and then trade a single asset for an outside good in bilateral meetings. A higher contact rate gives more trading opportunities but it also incurs a higher upfront cost. Following Duffie, Gârleanu and Pedersen (2005) and a large subsequent literature, we assume traders have an intrinsic reason for trade, differences in the flow utility they receive from holding the asset. Moreover, this idiosyncratic valuation changes over time, creating a motive for continual trading and retrading. We add to this a second source of heterogeneity, namely in contact rates. We allow traders to make different investments in the contact rate technology and hence meet other traders at different rates.

Under these assumptions, we show that intermediation arises naturally. When two traders who have the same flow valuation for the asset meet, the trader who has a higher contact rate acts as an intermediary, leaving the meeting with holdings that are further from the intrinsically desired one. This occurs because traders with a faster contact rate expect to have more future trading opportunities and so place less weight on their current flow payoff. Intermediation thus moves misaligned asset holdings towards traders with higher contact rates, which improves future trading opportunities. Thus, equilibrium displays a core-periphery structure where fast traders endogenously form the core, buying and selling largely irrespective of their idiosyncratic preferences, while those with infrequent contacts remain on the periphery, trading only to adjust their holdings after a valuation shock.1

1Recent empirical work documents that bilateral asset markets frequently exhibit a core-periphery network structure where few central institutions account for most of the turnover while the majority of market participants remains at the fringe. For the federal funds market, see Bech and Atalay (2010), Allen and Saunders (1986), and Afonso, Kovner and Schoar (2014). For evidence on international interbank lending, see Boss, Elsinger, Summer and Thurner (2004), Chang, Lima, Guerra and Tabak (2008), Craig and Von Peter (2014), and in ’t Veld and van Lelyveld (2014). For credit default swaps, see Peltonen, Scheicher and Vuilleme (2014), for the corporate bond market Maggio, Kermani and Song (2017), for the municipal bond
The full model recognizes that traders’ contact rates are endogenous. One possible interpretation is that agents may improve their contact rate by investing in their communication capacity, either through improved communication technology or by simply hiring more or more able individuals staffing their trading desk. An alternative interpretation is that agents may invest into relationships with more counterparties. That is, they may invest time and resources to increase the length of their contact list. While we assume that traders are ex-ante identical, we recognize that they may choose different contact rates, as in a mixed strategy equilibrium. We focus throughout on an aggregate steady state with a constant population sustained through exogenous entry and exit.

We begin the theoretical characterization by offering an algorithm to recover the investment cost function consistent with any observed distribution of contact rates. The rest of the paper focuses on characterizing the equilibrium counterparty distribution, the distribution of contact rates among the counterparties that any given trader faces, conditional on the investment cost function.

We first prove that if the cost is a continuously differentiable function of the contact rate, then the contact rate distribution must be continuous at any contact rate, except possibly at the maximum and minimum feasible values. This emphasizes that identical traders make heterogeneous investments. The force pushing towards heterogeneity is the gains from intermediation. If everyone else chooses the same contact rate, a trader who chooses a slightly faster contact rate acts as an intermediary for everyone else, repeatedly buying and selling irrespective of her intrinsic valuation; while a trader who chooses a slightly slower contact rate never trades once her asset position is aligned with her preferences. The marginal returns to additional meetings thus jump discretely at any mass point, inconsistent with equilibrium under a differentiable cost function.

We then offer a representation of equilibrium as a system of three ordinary differential equations for any twice continuously differentiable cost function. This representation is useful in proving further properties of equilibrium and for numerical purposes.

Using this representation, we show that under the assumption that both cost and marginal cost are convex functions of the contact rate, the support of the counterparty distribution is convex. Thus we can rule out a star network, with a few fast traders and many much slower traders, inter alia. Furthermore, we show that the equilibrium rate of misalignment is strictly increasing in the contact rate. That is, a higher contact rate comes with an inferior asset position (relative to fundamentals) and derives its benefits from trading profits. In turn, traders on the fringe of the trading network have well-aligned asset positions but pay for the intermediation services provided by the core through bid-ask spreads.

market Li and Schürhoff (2018), and for asset-backed securities Hollifield, Neklyudov and Spatt (2016).
We then turn to the case where the cost is proportional to the contact rate up to an exogenous upper bound. We show that when the cost per meeting exceeds a threshold, an equilibrium with no trade exists. When the cost falls below a strictly lower threshold, an equilibrium where all traders choose the highest feasible contact rate exists. For all intermediate costs, an equilibrium exists where the distribution of contact rates has a positive lower bound and continuous support up to the upper bound. Further, the counterparty distribution has a mass point at the upper bound and the misalignment rate is again increasing on the support.

The mass point at the upper bound motivates us to define a limiting equilibrium, which relaxes the exogenous upper bound on contact rates. Two key features emerge: First, we prove that the right tail of the contact rate distribution is Pareto with tail index 2. To the best of our knowledge, we are the first to show that a power law is an equilibrium outcome when homogeneous individuals choose their search technology under linear cost. This result carries over to the distribution of trading frequencies and connects our theory tightly with empirical evidence on frictional asset markets.\textsuperscript{2}

Second, we show that a trader has a strictly positive probability of meeting a counterparty with an arbitrarily fast contact rate. We call such traders \textit{middlemen}. To reconcile this with the first result, note that there is a vanishing measure of middlemen in the economy, yet their high meeting rate ensures that they account for a strictly positive fraction of meetings. We stress that ex-ante there is no difference between middlemen (the core of the network) and the periphery; however, they choose to make different investments and so ultimately play a very different role in the market.\textsuperscript{3}

We then demonstrate that these forces remain important even in an environment where the cost per contact goes to zero, so search frictions vanish. To do so, we offer a volume decomposition in the frictionless limit. We show that aggregate trading volume converges to a number more than four times larger than the amount of net asset reallocation induced by

\begin{footnotesize}
\textsuperscript{2}There is ample empirical evidence on concentration of trade among very few financial institutions. The largest sixteen derivatives dealers intermediate more than 80 percent of the global total notional amount of outstanding derivatives (Mengle, 2010; Heller and Vause, 2012). Bech and Atalay (2010) document that the distribution of trading frequencies in the federal funds markets is well-approximated by a power law, while Peltonen, Scheicher and Vuilleumey (2014) find that the degree distribution of the aggregate credit default swap network can be scale-free. For additional financial market variables that are—at least in the tail—well approximated by power laws with tail indices between 1 and 3, see Gabaix, Gopikrishnan, Plerou and Stanley (2006).

\textsuperscript{3}Duffie, Gârleanu and Pedersen (2005) and the ensuing literature frequently assume the existence of marketmakers with access to a frictionless interdealer market. The middlemen which emerge endogenously in our environment share many features: Regular trader stochastically meet them, they bargain over the terms of trade, and middlemen continuously trade with other middlemen. One difference is that our middlemen are atomless which implies that they are also in continuous contact with regular traders and, as a whole, never hold any inventory.
\end{footnotesize}
taste shocks. The fact that trading volume far exceeds the minimal amount of reallocation needed to offset preference shocks is a direct consequence of intermediation. We further show that almost all asset reallocation runs through an intermediation chain that involves middlemen.

We then consider Pareto optimal trading patterns and investments. We prove that the equilibrium trading pattern—passing misalignment to traders with higher contact rates—is optimal. We also show that all the qualitative features of equilibrium carry over to an optimum: Under the same restrictions on the cost function, there is optimally no mass points in the distribution of contact rates, faster traders optimally have asset positions that are increasingly detached from their intrinsic preferences, the optimal contact rate distribution has a Pareto tail with parameter 2, and middlemen emerge.

Equilibrium, however, is inefficient due to search externalities. Pigouvian taxes highlight the inefficiencies: traders only capture half the surplus in each meeting, a force towards underinvestment in contacts; at the same time, they do not internalize a business stealing effect, a force towards overinvestment in contacts. We numerically contrast equilibrium with optimum in the linear cost case and find systematic overinvestment, so the equilibrium contact rate distribution first order stochastically dominates the optimum. We also prove that in the frictionless limit, equilibrium trading volume exceeds the optimum.

Finally, we emphasize the connection between intermediation and dispersion in contact rates. We consider an economy in which traders with the same desired asset holdings never meet, which eliminates the possibility of intermediation in our model economy. Under these conditions, we show that if the cost is a weakly convex function of the contact rate, all traders choose the same contact rate, both in equilibrium and in the optimum. Thus dispersion in contact rates and intermediation are intimately connected: if there is dispersion in contact rates, faster traders act as intermediaries; and if intermediation is permitted, contact rates are naturally dispersed to leverage the gains from intermediation.

**Related Work** This paper is closely related to a growing body of work on trade and intermediation in markets with search frictions. Rubinstein and Wolinsky (1987) were the first to model middlemen 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 their basic framework to accommodate newly available empirical evidence on trade and intermediation in over-the-counter markets.
The decentralized interdealer market in Neklyudov (2014) features dealers with heterogeneous contact rates. The same dimension of heterogeneity is present in Üslü (2018), who also allows for heterogeneity in pricing and inventory holdings in a market where traders have continuously distributed flow payoffs.\textsuperscript{4} As in our framework, fast dealers in these setups are more willing to take on misaligned asset positions, thus endogenously emerging as intermediaries. The marketplace features intermediation chains and a core-periphery trading network with fast traders at the core. We add to this literature by first showing that heterogeneity in meeting technologies arises naturally to leverage the gains from intermediation even with ex-ante homogeneous agents, and second by showing how the endogenous choice of contact rates 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 (2018) also model a market with a continuum of flow payoffs and show that this gives rise to intermediation chains; market participants with extreme flow payoff constitute the periphery and those with moderate payoff value constitute the core. Afonso and Lagos (2015) similarly has endogenous intermediation because banks with heterogeneous asset positions buy and sell depending on their counterparties’ reserve holdings. In contrast to these setups, ours offers a theory where the identity of the individuals at the center of the intermediation chain remains stable over time, a key empirical feature of many decentralized asset markets (see, for instance, Bech and Atalay (2010) for the federal funds market.)

The identity of the institutions at the core is also stable in Chang and Zhang (2018), where agents differ in terms of the volatility of their taste for an asset and those with less volatile valuation act as intermediaries. The same is true in our framework but heterogeneity in the volatility of an agent’s taste arises endogenously since a higher contact rate buffers the impact of the flow payoff on the net valuation of asset ownership.

Nosal, Wong and Wright (2016) extend Rubinstein and Wolinsky (1987) to allow for heterogeneous bargaining power and storage cost. Agents can select between three distinct roles, each with its own exogenous contact rate. Farboodi, Jarosch and Menzio (2018) 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 upon 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 determines the distribution of bargaining power in the population,

\textsuperscript{4}A related literature studies the positive and normative consequences of high-frequency trading in centralized financial markets; see, for instance, Pagnotta and Philippon (2018).
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 features exogenously given middlemen who facilitate trade and have access to a frictionless interdealer market (Duffie, Gârleanu and Pedersen, 2005; Weill, 2008; Lagos and Rocheteau, 2009). We show that such middlemen who are in continuous contact with the market are a natural equilibrium outcome when homogeneous agents invest into a search technology.

Other recent work studies the structure of financial markets using explicit network formation models, which also generate core-periphery network structures (Farboodi, 2017; Wang, 2018) or star networks (Babus and Hu, 2017). In this class of models, agents (traders, banks) form explicit links, over which trade can be executed, at either an explicit cost—the cost of maintaining a relationship as in Babus and Hu (2017) and Wang (2018)—or an implicit cost—the counterparty risk in Farboodi (2017). The cost of acquiring a contact rate in our random search setup is closely related to the price of links in this network formation literature following Jackson (2010). Network models tend to generate a somewhat extreme core-periphery structure, where traders take on one of two roles, the core or the periphery; and traders in the periphery only trade with those in the core. Our model predicts a continuous distribution of trading frequencies.

In summary, the theoretical literature on frictional asset markets has offered a variety of economic mechanisms that give rise to empirically observed intermediation chains and core-periphery trading structures. We share with much of the literature the notion that intermediaries trade against their static holding payoff. Our analysis offers novel insights along four distinct dimensions: (i) time-invariant heterogeneity arises endogenously to leverage the gains from trade; (ii) middlemen with continuous market contact arise endogenously; (iii) the tail of the endogenous distribution of contact and trading rates is Pareto and our theory hence connects with the empirical regularities in a very tight way; (iv) our normative analysis shows that both intermediation and heterogeneity in the search technology are closely interrelated and socially desirable.

Finally, the finding that both the equilibrium and optimal allocations have a Pareto tail relates the paper to a large literature in economics that explores theoretical mechanisms which give rise to endogenous power law distributions (Gabaix, 1999; Eeckhout, 2004; Geerolf, 2017). Many other economically important regularities, such as the distributions of city and firm size and the distributions of income and wealth, are empirically well-approximated by power laws. To the best of our knowledge, the mechanism giving rise to the Pareto tail in our environment is novel and unrelated to the ones that are established in the literature (see Gabaix, 2009, 2016, for an overview).
Outline. The rest of the paper is organized as follows: Section 2 lays out the model. Section 3 defines equilibrium and Section 4 characterizes it. Section 5 discusses the Pareto optimal allocation and how it can be decentralized. Section 6 considers an economy where intermediation is prohibited. Section 7 concludes.

2 Model

We study an economy where time is continuous and extends forever. We focus throughout on an aggregate steady state. A unit measure of traders 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 (‘die’) when hit by an idiosyncratic shock with arrival rate $r > 0$. When a trader dies, she is replaced with a newborn trader so as to keep the population fixed at 1.

Asset Holdings and Preferences. The supply of the asset is fixed at $\frac{1}{2}$ and an individual trader’s asset holding is restricted to be $b \in \{0, 1\}$, so at any point in time half the traders hold the asset and half do not. 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 assume that $\Delta \equiv \frac{1}{2}(\delta_{h,1} + \delta_{l,0} - \delta_{h,0} - \delta_{l,1})$ is 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 $(l, 0)$. 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. 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. Traders discount the future only because of the death probability $r > 0$. When a trader dies holding the asset, it is transferred to a newborn trader with taste $h$, and the dying trader is not compensated.

Contact Rates and Search Technology. Asset trades occur pairwise in a frictional asset market. Newborn traders choose a time-invariant rate $\lambda \in [0, \bar{\lambda}]$ at which they make contact

---

5We 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 $\lambda$. 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 $\lambda$. 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 preferences.
with another trader, where $\lambda$ is an exogenous upper bound. A high contact rate is costly: a trader who chooses a contact rate $\lambda$ pays an ex-ante cost $C(\lambda)$, where $C : [0, \bar{\lambda}] \to \mathbb{R}$ is a continuous function. We allow for the possibility that different traders choose different contact rates. Let $G(\lambda)$ denote the cumulative distribution function of contact rates in the population and let $\Lambda \equiv \int_0^{\bar{\lambda}} \lambda dG(\lambda) \in [0, \bar{\lambda}]$ denote the average contact rate.

Search is random. A trader who chooses a contact rate $\lambda$ meets another trader at rate $\lambda$. Whom the trader meets is independent of her current taste and asset holding, but is proportional to the other trader’s contact rate. That is, conditional on meeting a counterparty, the probability that his contact rate $\lambda'$ is less or equal than $\lambda$ is given by $F(\lambda) \equiv \int_0^{\lambda} \frac{\lambda'}{\Lambda} dG(\lambda')$. 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.

By differentiating this expression and rearranging terms, we get

$$dG(\lambda) = \frac{\Lambda}{\lambda} dF(\lambda).$$

Moreover, since $G$ is a cumulative distribution function, $\int_0^{\bar{\lambda}} dG(\lambda) = 1$, we can recover $\Lambda$ from $F$:

$$\Lambda = \frac{1}{\int_0^{\bar{\lambda}} \frac{1}{\lambda} dF(\lambda)},$$

with $\Lambda = 0$ when the integral is improper. Thus, the function $F$ uniquely determines $G$ and $\Lambda$. Conversely, $G$ uniquely determines $F$ if and only if $\Lambda$ is positive. In particular, for any counterparty distribution $F$ with $F(0)$ strictly positive, $\Lambda = 0$ and $G(0) = 1$. For this reason, it is convenient to define equilibrium in terms of $F$ rather than $G$.

**Trade** When two traders meet, their asset holdings, preferences, and contact rates are observed by each. If both traders hold the asset or neither trader holds the asset, there is no gain from trade because of the binary restriction on asset holdings. If only one trader holds the asset, as will be the case in half of all meetings, the traders may swap the asset in exchange for an endogenous amount of the outside good. Whether trade occurs and what the terms of trade are is determined according to the symmetric Nash bargaining solution.

---

6When the cost function is strictly convex, the upper bound on contact rates $\bar{\lambda}$ is without loss of generality. However, we later pay particular attention to the linear cost case, $C(\lambda) = c\lambda$, in which case we also consider the limit $\bar{\lambda} \to \infty$. 

---

8
3 Equilibrium

This section focuses on defining equilibrium. We start by characterizing the value functions, the Nash bargaining solution, and the flow of traders in and out of different states of asset holdings and preferences. We then turn our focus to symmetric environments, where a trader’s behavior only depends on her contact rate and on whether her asset holdings are well-aligned with her preferences. We next analyze newborn traders’ choice of contact rates and finish with a formal definition of equilibrium.

3.1 Value Functions and Flows

One piece of equilibrium is figuring out who trades with whom and what transfers are made. Towards this end, let $\lambda, \lambda', i, i', b, b' \in [0, \bar{\lambda}]\{h, l\} \{0, 1\}$ denote the probability that a trader with contact rate $\lambda \in [0, \bar{\lambda}]$ in preference state $i \in \{h, l\}$ with asset holdings $b \in \{0, 1\}$ trades when she contacts a trader with contact rate $\lambda' \in [0, \bar{\lambda}]$ in preference state $i' \in \{l, h\}$ with asset holdings $b' \in \{0, 1\}$. Also let $p_{\lambda, i, b}^{\lambda', i', b'}$ denote the transfer of the outside good from $\{\lambda, i, b\}$ to $\{\lambda', i', b'\}$ when such a trade takes place. Feasibility requires that $p_{\lambda, i, b}^{\lambda', i', b'} + p_{\lambda', i', b'}^{\lambda, i, b} \geq 0$, since there are no outside resources available to the trading pair. The trading probability and price are determined by Nash bargaining.

Let $\mu_{\lambda, i, b}$ denote the endogenous fraction of traders with contact rate $\lambda$ who are in preference state $i$ and have asset holding $b$. Let $v_{\lambda, i, b}$ denote the present value of a trader $\{\lambda, i, b\}$. Given $F$ and $\mu$, this is defined recursively by

$$rv_{\lambda, i, b} = \delta_{i, b} + \gamma (v_{\lambda, \sim i, b} - v_{\lambda, i, b}) + \lambda \int_0^\lambda \sum_{i' \in \{h, l\}} \sum_{b' \in \{0, 1\}} \mu_{\lambda', i', b'}^{\lambda, i, b} 1_{\lambda, i, b}^{\lambda', i', b'} (v_{\lambda, i, b} - v_{\lambda, i, b} - p_{\lambda, i, b}^{\lambda', i', b'}) dF(\lambda'). \quad (3)$$

The left hand side of equation (3) is the flow value of the trader, where discounting reflects the death rate. The value comes from three sources, listed in order on the right hand side. First, she receives a flow payoff $\delta_{i, b}$ that depends on her preferences and asset holdings. Second, her preferences shift from $i$ to $\sim i$ at rate $\gamma$, in which case the trader has a capital gain $v_{\lambda, \sim i, b} - v_{\lambda, i, b}$. Third, she meets another trader at rate $\lambda$ with type $\lambda'$ drawn from the counterparty distribution $F$, in which case they may swap asset holdings in return for a payment. Conditional on $\lambda'$, the counterparty’s state is $(i', b')$ with probability $\mu_{\lambda', i', b'}$. If the two trade, with probability $1_{\lambda, i, b}^{\lambda', i', b'}$, the $(\lambda, i, b)$ trader has a capital gain from swapping assets and transferring the outside good, $v_{\lambda, i, b} - v_{\lambda, i, b} - p_{\lambda, i, b}^{\lambda', i', b'}$.

Symmetric Nash bargaining imposes that trade occurs whenever it makes both parties
better off, and that trading prices equate the gains from trade without throwing away any resources. That is, if there are transfers \( p_{\lambda,i,b}' \) and \( p_{\lambda,i,b} \) with nonnegative sum satisfying \( v_{\lambda,i,b}' - v_{\lambda,i,b} - p_{\lambda,i,b}' \) and \( v_{\lambda,i,b}' - v_{\lambda,i,b} - p_{\lambda,i,b} \) both positive, then trade occurs at a price such that

\[
\frac{v_{\lambda,i,b}'}{p_{\lambda,i,b}'} - \frac{v_{\lambda,i,b}}{p_{\lambda,i,b}} = \frac{v_{\lambda,i,b}'}{p_{\lambda,i,b}'} - \frac{v_{\lambda,i,b}}{p_{\lambda,i,b}}.
\]

If any feasible transfer implies a strict loss from trade, there is no trade. It follows immediately that

\[
1_{\lambda,i,b} = \begin{cases} 
1 & \text{if } v_{\lambda,i,b} + v_{\lambda,i,b}' \gtrless v_{\lambda,i,b} + v_{\lambda,i,b}' ; \\
0 & \text{otherwise} \end{cases}
\]

(4)

and that when there is trade, the price satisfies

\[
p_{\lambda,i,b}' = \frac{1}{2} (v_{\lambda,i,b} + v_{\lambda,i,b}' - v_{\lambda,i,b} - v_{\lambda,i,b}') .
\]

(5)

These equations also imply that if \( b = b' \), there is no gain from trade and no transfer.

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

\[
\left( r + \gamma + \lambda \int_0^\lambda \sum_{i' \in \{h,l\}} \mu_{\lambda,i',1-b} 1_{\lambda,i,b} dF(\lambda') \right) \mu_{\lambda,i,b} \\
= \gamma \mu_{\lambda,1-b} + \lambda \left( \int_0^\lambda \sum_{i', \in \{h,l\}} \mu_{\lambda,i',b} 1_{\lambda,i,1-b} dF(\lambda') \right) \mu_{\lambda,1-b} + \frac{r}{2} \mathbb{I}_{(i,b) \in \{(h,1),(l,0)\}}.
\]

(6)

The left hand side of equation (6) measures the outflows from state \((i, b)\) for traders with contact rate \( \lambda \). A trader exits the state either when she dies, at rate \( r \), when she has a preference shock, at rate \( \gamma \), or when she trades with another trader with the opposite asset holding. The right hand side measures the inflows. A trader with contact rate \( \lambda \) enters state \((i, b)\) when she is in the opposite preference state and has a preference 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.

### 3.2 Symmetry

We call traders’ asset holding positions misaligned either when they hold the asset and are in preference state \( l \), or when they do not hold the asset and are in preference state \( h \). We call traders’ asset holding positions well-aligned in the other two states. For the remainder
of the paper, we 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

$$1_{\lambda,i,b}^{\lambda',i',b'} = 1_{\lambda,i,1-b}^{\lambda,i,1-b'},$$

so if a type $\lambda$ trader sells the asset to a type $\lambda'$ trader when both are in state $h$ then it must be that they trade in the opposite direction when both are in state $l$. That such trading patterns may be consistent with equilibrium is a consequence of our symmetric market structure, where half of the traders are in each preference state and half of the traders hold the asset.

When

$$1_{\lambda,i,b}^{\lambda',i',b'} = 1_{\lambda,i,1-b}^{\lambda,i,1-b'},$$

equation (6) implies

$$\mu_{\lambda,i,b} = \mu_{\lambda,i,1-b}$$

for all $\{\lambda, i, b\}$. That is, the fraction of traders with contact rate $\lambda$ 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-$\lambda$ 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 $\lambda$.

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 $1_{\lambda,a}^{\lambda',a'}$ indicate the trading probability between traders $(\lambda, a)$ and $(\lambda', a')$ conditional on them having the opposite asset holdings; there cannot be trade if they have the same asset holdings. Let $m_\lambda \equiv \mu_{\lambda,l,1} + \mu_{\lambda,h,0}$ denote the fraction of traders with contact rate $\lambda$ who are misaligned.

Equation (6) reduces to

$$\left( r + \gamma + \frac{\lambda}{2} \int_0^{\lambda'} \left( 1_{\lambda,0}^{\lambda',0} m_{\lambda'} + 1_{\lambda,0}^{\lambda',1} (1 - m_{\lambda'}) \right) dF(\lambda') \right) m_\lambda$$

$$= \left( \gamma + \frac{\lambda}{2} \int_0^{\lambda'} \left( 1_{\lambda,1}^{\lambda',0} m_{\lambda'} + 1_{\lambda,1}^{\lambda',1} (1 - m_{\lambda'}) \right) dF(\lambda') \right) (1 - m_\lambda). \quad (7)$$

The left hand side is the outflow rate from the misaligned states. This occurs either following death, a preference shock, or a meeting with a trader who has the opposite asset holdings where trade occurs. The right hand side is the inflow rate. This occurs following a preference shock or meeting a trader with the opposite asset holdings where trade occurs. Following the death of a well-aligned trader, she is replaced by another well aligned trader, and so this event does not appear on the right hand side of equation (7).

Let $s(\lambda) \equiv \frac{1}{2}(v_{\lambda,h,1} + v_{\lambda,l,0} - v_{\lambda,l,1} - v_{\lambda,h,0})$, the average surplus from being well-aligned rather than misaligned. Knowing the surplus function is sufficient to tell whether trade occurs. To see this, we prove in appendix B.1 that

$$v_{\lambda,h,1} + v_{\lambda,l,1} - v_{\lambda,h,0} - v_{\lambda,l,0} = \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0}}{r}$$
\[ v_{\lambda,b,1} - v_{\lambda,b,0} + v_{\lambda',l,0} - v_{\lambda',l,1} = s(\lambda) + s(\lambda') \]
\[ v_{\lambda,l,1} - v_{\lambda,l,0} + v_{\lambda',h,0} - v_{\lambda',h,1} = -s(\lambda) + s(\lambda') \]
\[ v_{\lambda,b,1} - v_{\lambda,b,0} + v_{\lambda',h,0} - v_{\lambda',h,1} = s(\lambda) - s(\lambda') \]
\[ v_{\lambda,l,1} - v_{\lambda,l,0} + v_{\lambda',h,0} - v_{\lambda',h,1} = -s(\lambda) - s(\lambda'). \]

Thus condition (4) tells us that the surplus function \( s(\lambda) \) governs the patterns of trade. We note that these four equations reflect the symmetry of the environment and are useful in simplifying the analysis. They show that the gains from trade in a feasible transaction between two traders \((\lambda, \lambda')\) only reflects the change in their respective misalignment status but does not depend on the trader’s actual asset holdings.

Now define \( \delta_0 \equiv \frac{1}{2}(\delta_{h,0} + \delta_{l,1}) \), and \( \delta_1 \equiv \frac{1}{2}(\delta_{h,1} + \delta_{l,0}) = \delta_0 + \Delta > \delta_0 \). Taking advantage of symmetry in the misalignment rates, the Nash bargaining solution (4) and (5), and exploiting the four equations provided in the previous paragraph, equation (3) reduces to

\[ \frac{r}{2} (v_{\lambda,l,1} + v_{\lambda,h,0}) = \delta_0 + \gamma s(\lambda) + \frac{\lambda}{4} \int_{0}^{\lambda} \left((s(\lambda) + s(\lambda'))^\dagger m_{\lambda'} + (s(\lambda) - s(\lambda'))^\dagger (1 - m_{\lambda'})\right) dF(\lambda'), \] (8)

\[ \frac{r}{2} (v_{\lambda,h,1} + v_{\lambda,l,0}) = \delta_1 - \gamma s(\lambda) + \frac{\lambda}{4} \int_{0}^{\lambda} \left((-s(\lambda) + s(\lambda'))^\dagger m_{\lambda'} + (-s(\lambda) - s(\lambda'))^\dagger (1 - m_{\lambda'})\right) dF(\lambda'), \] (9)

where \( z^\dagger \equiv \max\{z, 0\} \) and reflects that meetings result in trade if and only if trade is bilaterally efficient. Again, both equations reflect that flow values are the sum of three terms. The first is the average flow payoff of a misaligned or well-aligned trader. The second is the gain or loss from a preference shock that switches the alignment status. The third is the gain from meetings, reflecting that only half of all meetings are with traders who hold the opposite asset; and in these events each trader walks away with half of the joint surplus, if this is positive.

Subtracting one of these equations from the other gives us a functional equation describing the surplus:

\[ \Delta = (r + 2\gamma)s(\lambda) + \frac{\lambda}{4} \int_{0}^{\lambda} \left(\left(\left((s(\lambda) + s(\lambda'))^\dagger - (s(\lambda') - s(\lambda))^\dagger\right) m_{\lambda'} \right. \right. \]
\[ \left. \left. + \left(\left((-s(\lambda) + s(\lambda'))^\dagger - (-s(\lambda) - s(\lambda'))^\dagger\right)(1 - m_{\lambda'})\right) dF(\lambda') \right. \right. \right. \] (10)
Finally, we can apply the Nash trading rule (4), which states that trade occurs if and only if doing so is bilaterally efficient, to equation (7):

\[
\left( r + \gamma + \frac{\lambda}{2} \right) \int_{\lambda_0}^{\lambda} \left( I_{s(\lambda_0) + s(\lambda') > 0} m_{\lambda'} + I_{s(\lambda) > s(\lambda')} (1 - m_{\lambda'}) \right) dF(\lambda') m_{\lambda'} = \left( \gamma + \frac{\lambda}{2} \right) \int_{\lambda_0}^{\lambda} \left( I_{s(\lambda) < s(\lambda')} m_{\lambda'} + I_{s(\lambda) + s(\lambda') < 0} (1 - m_{\lambda'}) \right) dF(\lambda') (1 - m_{\lambda}). \quad (11)
\]

Here the indicator function \( I \) is equal to 1 if the inequality in the subscript holds and is zero otherwise.

### 3.3 Endogenizing the Distribution of Contact Rates

Newborn traders choose their contact rate \( \lambda \) to maximize their expected present value 
\[
\frac{1}{2} (v_{\lambda,h,1} + v_{\lambda,l,0}) - C(\lambda)
\]
Using equation (9), we can rewrite the choice of contact rates as one of finding \( \lambda \) to maximize

\[
\pi_\lambda \equiv \frac{\delta_1 - \gamma s(\lambda) + \frac{1}{4} \int_{\lambda_0}^{\lambda} ((-s(\lambda) + s(\lambda'))^+ m_{\lambda'} + (-s(\lambda) - s(\lambda'))^+ (1 - m_{\lambda'}) \ dF(\lambda'))}{r} - C(\lambda)
\]

This reflects the fact that traders choose their contact rate when they enter the market, in one of the well-aligned states.\(^7\) In equilibrium, identical traders may make different ex-ante decisions about \( \lambda \). In this case, there must be multiple values of \( \lambda \) that maximize the expected payoff \( \pi_\lambda \).

### 3.4 Definition of Equilibrium

We are now able to define an equilibrium.

**Definition 1** An equilibrium is a nondecreasing, right-continuous counterparty distribution \( F : [0, \bar{\lambda}] \rightarrow [0, 1] \) with \( F(\bar{\lambda}) = 1 \), a misalignment rate function \( m : [0, \bar{\lambda}] \rightarrow [0, 1] \), and a surplus function \( s : [0, \bar{\lambda}] \rightarrow \mathbb{R} \), satisfying:

1. the surplus equation (10);
2. the flow balance equation (11); and

\(^7\)As can be seen from the expression for \( \pi_\lambda \), our 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.
3. optimality of the ex-ante investment decision: $\lambda$ is in the support of $F$ only if $\pi_\lambda = \max_{\lambda'} \pi_{\lambda'}$ where $\pi_\lambda$ is defined in equation (12).

We have already explained all three conditions. Once we have found an equilibrium counterparty distribution $F$, it is straightforward to recover the contact rate distribution $G$ and the average contact rate $\Lambda$ from equations (1) and (2).

4 Equilibrium Characterization

This section develops eight propositions which characterize equilibrium. Proposition 1 focuses on which trades occur for an arbitrary counterparty distribution, Proposition 2 shows how to back out the cost function that generates any equilibrium counterparty distribution, and Propositions 3–8 characterize the counterparty distribution under different restrictions on the cost function $C$.

4.1 Equilibrium Trading Patterns

We start by characterizing equilibrium trading patterns given any distribution $F(\lambda)$.

**Proposition 1** In 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;

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 appendix contains proofs of all our propositions. Once we establish that the surplus function is non-negative and decreasing, the trading patterns follow from Nash bargaining.

The first two parts of the proposition 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 proposition reflects option value considerations and is the key feature of the endogenous trading pattern which arises in this environment, namely intermediation. It states that a faster trader buys the asset from a slower trader if both are in preference state $l$; and she sells the asset to the slower trader if both are in preference state $h$. These trades do not immediately increase the number of well-aligned traders, but they move
direction of trade

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

misalignment towards traders who expect more future trading opportunities. These trades yield 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 preferences 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 frequent meetings. Figure 1 shows the intermediation chain which follows from Proposition 1. Slow traders are at the periphery of the trading chain, not trading once their asset position is aligned with their preferences. In turn, fast traders constitute the endogenous core of the trading network, buying and selling largely irrespective of their preference state. In doing so, they take on misaligned asset positions from types with lower contact rate simply because they are better at locating other traders. That is, they intermediate.

4.2 Reverse-Engineering the Cost Function

We next show how we can use our analysis to recover the set of cost functions consistent with any observed counterparty distribution $F$. Towards that end, let $M(\lambda) \equiv \int_0^\lambda m_\lambda dF(\lambda')$ denote the fraction of meetings that are with a misaligned trader with contact rate less than or equal to $\lambda$, so $dM(\lambda) = m_\lambda dF(\lambda)$ for all $\lambda$. Next, multiply both sides of the inflow-outflow equation (11) by $dF(\lambda)$ and use the fact that the surplus function is decreasing in $\lambda$ (Proposition 1) to get

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

(13)
Since \( F \) is a monotonic function, it is almost everywhere differentiable. At such points, the inflow-outflow equation (13) reduces to

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

(14)
an ordinary differential equation for \( M \). If there is a mass point at some \( \lambda \in [0, \bar{\lambda}] \), equation (13) is a quadratic equation for \( dM(\lambda) \). The larger solution has \( dM(\lambda) > dF(\lambda) \), which is inconsistent with the fact that their ratio, the misalignment rate \( m_\lambda \), is less than 1. The smaller solution, with \( dM(\lambda) \in (0, dF(\lambda)/2) \), determines the jump in \( M \) at \( \lambda \).

Once we have recovered the misalignment distribution \( M \), we compute the surplus function. To do this, it is useful to extend the definition of \( F \) and \( M \) to the positive real line through the convenient normalizations \( F(\lambda) = 1 \) and \( M(\lambda) = M(\bar{\lambda}) \) if \( \lambda > \bar{\lambda} \). Then using the fact that surplus function is non-negative and decreasing (Proposition 1), we prove the following result:

**Lemma 1** In equilibrium, the surplus function is

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

(15)

where

\[
\phi_\lambda \equiv \frac{4(r + 2\gamma)}{\lambda \left( 4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda)) \right)}.
\]

(16)

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

Again taking advantage of the fact that surplus function is non-negative and decreasing, we can rewrite equation (12) as

\[
\pi_\lambda = \frac{\delta_1 - \gamma s(\lambda)}{r} + \frac{\lambda}{4} \int_{0}^{\lambda} (s(\lambda') - s(\lambda)) dM(\lambda') - C(\lambda).
\]

(17)

Part 3 of the definition of equilibrium imposes that \( \lambda \) is in the support of \( F \) only if it maximizes \( \pi_\lambda \). For any number \( \bar{\pi} \), this gives us a lower bound on the cost function,

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

with equality if \( \lambda \) is in the support of the counterparty distribution \( F \). Since for any \( F \), equations (13) and (15) uniquely determine \( M \) and \( s \), the lower bound is unique up to the constant \( \bar{\pi} \). We summarize these results in the following proposition.
Proposition 2 For any counterparty distribution $F$, there exists a cost function $C$ such that $F$ is an equilibrium. Moreover, $C$ is unique on the support of $F$, up to the constant $\bar{\pi}$.

The proof is in the preceding text.

4.3 Dispersion in Contact Rates

We next show that a non-degenerate counterparty distribution $F$, that is the coexistence of traders with different $\lambda$, arises naturally in equilibrium even when market participants are ex-ante homogeneous.

Proposition 3 Assume $C(\lambda)$ is continuously differentiable. Then the equilibrium counterparty and contact rate distributions $F$ and $G$ are continuous on $(0, \bar{\lambda})$.

The proof of this proposition relies heavily on the characterization in Lemma 1. We note that the proposition allows for the possibility that $F$ and $G$ are discontinuous at $\bar{\lambda}$. It also allows for $F(0)$ or $G(0)$ to be strictly positive. Finally, the proposition is consistent with $F$ and $G$ constant on $(0, \bar{\lambda})$; the rest of this section characterizes restrictions on the cost function $C$ which ensures that this is not the case.

Proposition 3 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 $\bar{\lambda}$. Even stronger, almost all traders choose different contact rates, except at those two points. The proof shows that traders’ profit function $\pi_{\lambda}$ would have a convex kink at any interior mass point, inconsistent with that point being profit-maximizing.

To develop an intuition for the result, we suppose that everyone has a common contact rate $\lambda$ and argue that this creates a convex kink in the value function $\pi$ at $\lambda$. The force pushing towards heterogeneity is the gains from intermediation. To understand why, consider what happens to the marginal return to additional contacts at the masspoint. An individual with contact rate just below $\lambda$ never intermediates, so few of any additional meetings would lead to trade. This discreetly changes at the masspoint: If the trader has contact rate just above $\lambda$ she intermediates the entire marketplace trading entirely independently of intrinsic valuation. As a consequence, far more of any additional meetings of such a trader lead to trade which causes a jump in the marginal returns to meeting. The consequence is a convex kink in the value function at the mass point $\lambda$. It follows that $\lambda$ cannot be a profit-maximizing contact rate if the cost function is continuously 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 in the marginal return to contacts at this mass point, inconsistent with equilibrium under a differentiable cost function.
The absence of a pure strategy equilibrium is a common feature of 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 and the associated intermediation services, and we find that the profit function is continuous but not differentiable. Nevertheless, at a basic level, we are dealing with a similar phenomenon. A mass point creates a discontinuity in the probability of trade, although here there is no discontinuity in the gains from trade. As a consequence payoffs do not jump at a mass point, but their derivative jumps. This precludes a mass point with a differentiable cost function.

4.4 Representation of Equilibrium as ODE System

This section offers a characterization of equilibrium as a system of ordinary differential equations when the cost function is twice continuously differentiable. We use this characterization throughout the remainder of the analysis of equilibrium. Proposition 3 states that the counterparty distribution is continuous on the open interval $(0, \bar{\lambda})$. Since it is also monotone, Lebesgue’s theorem implies it is almost everywhere differentiable on $(0, \bar{\lambda})$. Moreover, $M(\lambda) = \int_0^\lambda m_{\lambda'}dF(\lambda')$ inherits the same properties: it is monotone, continuous, and almost everywhere differentiable on $(0, \bar{\lambda})$. Equation (14) gives one linear relationship between $F'(\lambda)$ and $M'(\lambda)$ at points where both exist, based on the balance of flows. Here we use optimality to derive another linear relationship.

As a preliminary step, we can differentiate equation (15) to get expressions for the first and second derivatives of the surplus function. Using the functional form of $\phi$ in equation (16), we get

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

(18)

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

(19)

Equation (19) only holds almost everywhere, at points where $F$ and $M$ are differentiable.

Next, let $A \subseteq [0, \bar{\lambda}]$ denote the support of $F$, the smallest closed set such that $\int_A dF(\lambda) = 1$. Unless $A \subseteq \{0, \bar{\lambda}\}$, continuity of $F$ on $(0, \bar{\lambda})$ implies that $A$ must have a nonempty interior, denoted $A^\circ$. At $\lambda \in A^\circ$, part 3 of the definition of equilibrium states that $\pi_\lambda = \max_{\lambda'} \pi_{\lambda'}$.
and so in particular the first order condition \( \pi'_\lambda = 0 \) holds if \( \pi \) is differentiable:

\[
\pi'_\lambda = \frac{-(4\gamma + \lambda M(\lambda))s'(\lambda) + \int_0^\lambda (s(\lambda') - s(\lambda))dM(\lambda')}{4r} - C''(\lambda) = 0.
\] (20)

Moreover, since the interior points \( A^o \) are not isolated, \( \pi''_\lambda = 0 \) at points \( \lambda' \) in a neighborhood of \( \lambda \). This in turn implies \( \pi''_\lambda = 0 \). Using equation (17) and assuming the cost function \( C \) is twice continuously differentiable, this implies that at any \( \lambda \in A^o \) where \( F \) and \( M \) are differentiable,

\[
\pi''_\lambda = \frac{-(4\gamma + \lambda M(\lambda))s''(\lambda) - (2M(\lambda) + \lambda M'(\lambda))s'(\lambda)}{4r} - C''(\lambda) = 0.
\] (21)

Now eliminate \( s'(\lambda) \) and \( s''(\lambda) \) from equation (21) using equations (18) and (19). This gives us another linear relationship between \( F'(\lambda) \) and \( M'(\lambda) \). We can solve this and equation (14) explicitly for \( F'(\lambda) \) and \( M'(\lambda) \) as functions of \( F(\lambda) \), \( M(\lambda) \), and \( s(\lambda) \); see equations (42) and (43) in the appendix. These two equations, together with equation (18), are an ordinary differential equation system in \( (F, M, s) \) on the set \( A^o \) that must hold in any equilibrium.

When \( C \) twice continuously differentiable, the system is sufficiently well-behaved to ensure that \( F, M, \) and \( s' \) are everywhere differentiable on \( A^o \).

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 \( A \) 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 (20) implies

\[
\frac{-\gamma s'(\underline{\lambda})}{r} = C'(\underline{\lambda}).
\] (22)

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 (13) to find \( M(\bar{\lambda}) \).

We still need to validate the guess of \( \bar{\lambda} \). 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 from equation (15). Taking advantage of the fact that \( F(\lambda) = 1 \) and
\( M(\lambda) = M(\bar{\lambda}) \) for \( \lambda > \bar{\lambda} \), this direct computation gives us

\[
s(\bar{\lambda}) = \frac{2\Delta}{2(r + 2\gamma) + \bar{\lambda}M(\bar{\lambda})}.
\]

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 \( \lambda \). Finding all equilibria may require a grid search on \( \lambda \).

### 4.5 Increasing and Convex Marginal Cost

Proposition 3 shows that there is dispersion in equilibrium contact rates under a relatively weak condition. This naturally leads us to ask how much dispersion. Here we characterize equilibrium when the marginal cost function is increasing and weakly convex. We conclude that dispersion is not too extreme, in the sense that the support of the counterparty distribution is convex:

**Proposition 4** Assume \( C(\lambda) \) is twice continuously differentiable with \( C''(\lambda) \) strictly positive and nondecreasing. In any equilibrium with \( \Lambda > 0 \), the support of the counterparty and contact rate distributions \( F \) and \( G \) are convex.

The proof works by contradiction. If there were a “hole” in the support of \( F(\lambda) \), we can use equation (19) to characterize the second derivative of the surplus function on that hole. We can then use equation (21) to prove that the second derivative of the profit function is strictly negative on the hole. Strict concavity of the profit function implies the extreme points of the hole cannot be profit maximizing, a contradiction. Having established that the support of \( F \) is convex, convexity carries over immediately to the support of \( G \).

Proposition 4 rules out the possibility that most traders choose a low contact rate, while a few traders choose a discretely higher contact rate acting as intermediaries, as in a star network. When the cost function is convex but the second derivative is decreasing, we can construct examples where the support of the counterparty and contact rate distributions consists of two disjoint convex intervals, i.e. a group of heterogeneous slow traders and another group of heterogeneous intermediaries.

We next turn to the connection between contact rates and misalignment rates. A higher contact rate has two opposing effects on a trader’s misalignment rate. First, a trader is more frequently able to offset a misaligned position. However, a trader with a higher contact rate also intermediates more frequently, taking on misalignment from slower traders. Proposition 5 states that the latter force dominates everywhere on the support of \( F \) if the marginal cost function is increasing and weakly convex:
Proposition 5 Assume \( C(\lambda) \) is twice continuously differentiable with \( C''(\lambda) \) strictly positive and nondecreasing. In any equilibrium with \( \Lambda > 0 \), the misalignment rate \( m_\lambda \) is strictly increasing on the support of \( F \).

We stress that Proposition 5 holds in equilibrium, not for an arbitrary distribution of contact rates. For example, if there were a hole in the support of the counterparty distribution, the misalignment rate would be strictly decreasing over this interval;\(^8\) by continuity, this extends to any distribution with sufficiently little support in an interval.

Proposition 5 reflects the fact that traders with faster contact rates are more likely to serve as intermediaries. It implies that traders do not invest in a faster contact rate to reduce their misalignment, but rather to trade more frequently at a favorable bid-ask spread.

A corollary of Proposition 5 considers an extension to our model where traders differ ex-ante in how much they care about having a well-aligned asset position, \( \Delta \). Proposition 5 suggests that faster traders at the core of the network will naturally be those with smaller \( \Delta \), while slower traders at the periphery will be those with larger \( \Delta \). This is the opposite of what one would expect to see without intermediation since, in that case, those with the largest \( \Delta \) would have the highest incentives to invest in a high contact rate.

We conclude this subsection by pointing out that, whenever the assumptions on the cost function in propositions 4 and 5 are satisfied, marginal costs grow without bound. However, the gains from trade in a meeting are always bounded above by \( 2s(0) = 2\Delta / (r + 2\gamma) \) (see equation 15). This implies that if \( \lambda \) is sufficiently large, the upper bound on the contact rate is not binding: \( F(\lambda) = 1 \) at some \( \lambda < \bar{\lambda} \).

### 4.6 Constant Marginal Cost

In this section, we restrict the cost function to be linear, so the cost per meeting is constant, \( C(\lambda) = c\lambda \). We begin by showing how to determine the marginal cost \( c \) that corresponds to a particular counterparty distribution. We then prove existence of equilibrium given a marginal cost \( c \) and characterize its properties.

Since \( C''(\lambda) = 0 \) and \( s'(\lambda) \) is nonzero, we can substitute (19) into (21) and simplify to get

\[
(1 - F(\lambda))(8\gamma - \lambda^2 M'(\lambda)) - \lambda F'(\lambda)(4\gamma + \lambda M(\lambda)) = 4r(2M(\lambda) + \lambda M'(\lambda)).
\]

For a given lower bound on the counterparty distribution \( \tilde{\lambda} \), equations (14) and (23) with terminal condition \( F(\tilde{\lambda}) = M(\tilde{\lambda}) = 0 \) are an initial value problem. We solve this problem

\(^8\)To prove this, note that the types of trades a trader would undertake are identical anywhere on the interval. Since trades reduce the misalignment rate on average, raising the contact rate without changing the types of trades reduces the misalignment rate.
for $\lambda \in [\underline{\lambda}, \bar{\lambda})$. We then set $F(\bar{\lambda}) = 1$ and compute $dM(\bar{\lambda})$ using equation (13). Since $F$ and $M$ are constant outside of the interval $(\underline{\lambda}, \bar{\lambda})$, this completely characterizes a candidate counterparty distribution.

To see if this is an equilibrium, we check whether $\pi'_\lambda = 0$. Using equation (20) along with the expressions for $s(\lambda)$ in equation (15) and $\phi_\lambda$ in equation (16), this reduces to

$$
\frac{c}{\Delta} = \frac{4\gamma}{2r(4r+2\gamma)+\bar{\lambda}} \exp \left( -\int_{\underline{\lambda}}^{\infty} \frac{4(r+2\gamma)}{\lambda(4(r+2\gamma)+\lambda(1-F(\lambda)+2M(\lambda)))} d\lambda \right). \quad (24)
$$

Thus, for any lower bound $\underline{\lambda} \in (0, \bar{\lambda})$, we can use this algorithm to find the cost that is consistent with this counterparty distribution as an equilibrium.

We are interested in solving this in the other direction: For a given cost, we look for an equilibrium. Our main result builds on the characterization above to prove existence of an equilibrium and establish general properties of the counterparty distribution.

**Proposition 6** Assume $C(\lambda) = c\lambda$. Fix $r, \gamma, \Delta,$ and $\bar{\lambda}$. There exists thresholds $\bar{c} > c > 0$ such that

$$
\begin{cases}
  c \geq \bar{c} & \text{if } \Lambda = 0 \\
  c \in (\underline{c}, \bar{c}) & \text{then there is an equilibrium with } \Lambda = 0, \Lambda \in (0, \bar{\lambda}) \text{ and } \Lambda = \bar{\lambda}.
\end{cases}
$$

In any equilibrium with $\Lambda \in (0, \bar{\lambda})$, the supports of the counterparty and contact rate distributions $F$ and $G$ are a convex interval $[\underline{\lambda}, \bar{\lambda}]$ with $\underline{\lambda} \in (0, \bar{\lambda})$; $F$ and $G$ have a mass point at $\bar{\lambda}$; and the misalignment rate $m_\lambda$ is increasing on $[\underline{\lambda}, \bar{\lambda}]$.

The proof gives explicit expressions for $\bar{c}$ and $\underline{c}$ as functions of $r, \gamma, \Delta,$ and $\bar{\lambda}$, and characterizes $F$ for any value of $c$. Once we find $F$, it is straightforward to construct an equilibrium, recovering $M$ and hence the misalignment rate $m_\lambda$ from equation (13) and the surplus function from equation (15).

Note that we do not claim uniqueness of the equilibrium and indeed can construct examples in which an equilibrium with $\Lambda = 0$ and an equilibrium with $\Lambda \in (0, \bar{\lambda})$ coexist for the same parameter values. However, any equilibrium must lie in one of the three classes described in the proposition.

Proposition 6 states that for $c$ sufficiently large, there exists an equilibrium where all trading activity collapses, while for $c$ sufficiently small, there exists an equilibrium without intermediation since all traders choose the highest contact rate $\bar{\lambda}$. More interestingly, for a
nonempty interval of costs, there exists an equilibrium where a non-degenerate counterparty
distribution $F$ (and hence contact rate distribution $G$) and intermediation emerge endoge-
nously. Such an equilibrium has three key properties: First, no trader has a contact rate
below a strictly positive lower bound. Second, a strictly positive fraction of meetings are
with traders choosing $\bar{\lambda}$. Third, the remaining counterparties have a continuously distributed
contact rate on $[\underline{\lambda}, \bar{\lambda})$. Any equilibrium must have one of these three forms.

The strictly positive lower bound $\underline{\lambda}$ reflects the fact that the value of a trader $\pi_{\lambda}$ is a
continuous function, converging to the autarky value $\pi_0 = \frac{\gamma d_0 + (r + \gamma) d_1}{r(r + 2\gamma)}$ as $\lambda$ 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.

The finding that $F$ is continuous on the interior has already been established in the more
general environment of Proposition 3.

Finally, to understand the mass point at $\bar{\lambda}$, it helps to decompose profits $\pi_{\lambda}$ into pure
trading profits and the returns from having a well-aligned asset position. If there was no
mass point at $\bar{\lambda}$, traders with $\lambda$ close to $\bar{\lambda}$ would only meet slower traders, and so their
misalignment rate would converge to $\frac{1}{2}$ in a strictly concave fashion. For such traders to be
better off than under autarky would thus require that their total trading profits, and hence
their profits per trade, are strictly positive. But trading profits would be linear in $\lambda$ for $\lambda$
close to $\bar{\lambda}$ if there were no mass point at $\bar{\lambda}$, and this is inconsistent with constant $\pi_{\lambda}$ near $\bar{\lambda}$.

4.7 Constant Marginal Cost: Limiting Equilibrium

Proposition 6 implies that whenever $\Lambda$ is positive, a positive fraction of contacts are with
traders who choose the maximum possible meeting rate, $\lambda = \bar{\lambda}$; 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:

**Definition 2** Assume $C(\lambda) = c\lambda$. Fix $r$, $\gamma$, $\Delta$, and $c$. The functions $(F, m, s)$, each with
domain $[0, \infty)$, are a limiting equilibrium if there is a sequence of functions \( \{F_n, m_n, s_n\} \),
each with domain $[0, \infty)$, which converge pointwise to $(F, m, s)$, and there is an increasing,
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 $\bar{\lambda}_n$.

Intuitively, a limiting equilibrium is the limit of a sequence of equilibria with a large maximum
contact rate. 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 way to do this would be to set $F_n(\lambda) = 1$
(reflecting that $F_n$ is a cumulative distribution function), $m_n(\lambda) = \frac{2\gamma + \Lambda M(\lambda)}{2(r + 2\gamma + \Lambda M(\lambda))}$ (reflecting
equation 11), and 
\[ s_n(\lambda) = \frac{2\Delta}{2(r+2\gamma)+\lambda M(\lambda)} \] (reflecting equation 10) when \( \lambda > \bar{\lambda}_n \). Our definition allows, but does not impose, this possibility.

Once we find a limiting equilibrium, we can compute \( \Lambda \) and \( G \) using versions of equations (1) and (2) with \( \bar{\lambda} = \infty \):

\[ \Lambda = \frac{1}{\lambda} \int_0^\infty dF(\lambda) \quad \text{and} \quad dG(\lambda) = \frac{\Lambda}{\lambda} dF(\lambda), \]

with \( \Lambda = 0 \) whenever \( \int_0^\infty \frac{1}{\lambda} dF(\lambda) \) is an improper integral. This gives us the following characterization:

**Lemma 2** Assume \( C(\lambda) = c\lambda \). If \( c \geq c^* \equiv \frac{\Delta}{8r(r+\gamma)(r+2\gamma)} \), a limiting equilibrium with \( \Lambda = 0 \) exists. If \( c < c^* \), a limiting equilibrium with \( \Lambda > 0 \) exists.

Building on this, we characterize any limiting equilibrium of this economy:

**Proposition 7** Assume \( C(\lambda) = c\lambda \). Take any limiting equilibrium with \( \Lambda > 0 \). Then 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 6 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 rules out the possibility that this fraction vanishes as \( \bar{\lambda} \) goes to infinity. In the limiting economy, almost every trader has a finite contact rate (\( \lim_{\lambda \to \infty} G(\lambda) = 1 \)), yet a fraction, bounded above zero, of their meetings are with traders who have a still faster contact rate. We call these traders middlemen.

The emergence of middlemen follows the same logic as in the previous proposition. They guarantee that the misalignment rate of traders with very high contact rate does not converge to \( \frac{1}{2} \), since middlemen allow even very fast types to offload misaligned asset positions. That is a necessary feature of equilibrium, since otherwise fast traders would have to have strictly positive net trading profits to avoid being worse off than under autarky. And if net trading profits were strictly positive, they would linearly scale in the tail of the contact rate distribution, inconsistent with indifference of traders across \( \lambda \). The fast traders in the tail, of course, compensate the middlemen for their intermediation services and hence middlemen make strictly positive net flow profits from trade, offsetting the jump in the misalignment rate at \( \bar{\lambda} \).\(^9\)

More broadly, one may wonder how it can be optimal to invest in a potentially very high contact rate \( \lambda \) given that traders’ fundamental tastes for the asset are persistent. The reason

\(^9\)Of course, middlemen are in continuous contact with the market so the profits per trade are negligible but the flow profits from trade stay well behaved.
is that such fast traders specialize in intermediation. They trade far more frequently than their taste changes and cover the cost of each meeting with the profit it generates.

Of course, that is not to say that middlemen only exist for the linear cost case. Indeed, our approach to constructing the cost function in Section 4.2 implies that any counterparty distribution with middlemen is an equilibrium for some cost function. And as long as that counterparty distribution is inconsistent with the differential equations (14) and (23), such a cost function will not be linear.

The remainder of the Proposition establishes that in the limiting economy, the equilibrium contact rate distribution’s tail looks the same as that of a Pareto distribution with tail index 2.\textsuperscript{10} To develop an understanding of this result, note that we have already established that the distribution has no mass points. It thus follows that gross flow values (ignoring the linear cost \( c\lambda \)) must be affine above the lowest contact rate \( \Lambda \). The shape of \( F \) as \( \lambda \to \infty \) implies that increasing a trader’s contact rate does not affect the frequency of meeting a faster finite trading partner. Furthermore, the relative contact rate conditional on meeting a faster trader is also independent of \( \lambda \). On the other hand, increasing \( \lambda \) linearly scales the frequency at which she meets a slower counterparty; and the expected contact rate of a slower counterparty converges to a constant. Jointly, these features guarantee linear gross flow values, and hence a flat \( \pi \) in the tail. The endogenous Pareto tail arises due to reasons unrelated to well-known mechanisms in various other contexts (Gabaix, 2016).

A large body of empirical work documents that the degree distribution of various financial markets is often well described by a power law (see footnote 2). However, what is mapped out empirically is the distribution of trading rates \( \alpha \equiv \lambda p_{\lambda} \), the product of the contact rate \( \lambda \) and the probability of trading in a meeting, \( p_{\lambda} \). Assuming trades occur only if there are strict gains, we can write \( p_{\lambda} \equiv \frac{1}{2}m_{\lambda}(1 - F(\lambda) + dM(\lambda)) + M(\lambda) - dM(\lambda) \): trade occurs when a misaligned trader meets a trader with a strictly higher contact rate or a misaligned trader with the same contact rate; it also occurs when a well- or misaligned trader meets a misaligned trader with a strictly lower contact rate. Let \( \hat{G}(\alpha) \) denote the population distribution of trading rates in a limiting equilibrium. Then a corollary to Proposition 7 connects the results describing the distribution of contact rates to the distribution of trading rates:

\textbf{Corollary 1} Assume \( C(\lambda) = c\lambda \). Take any limiting equilibrium with \( \Lambda > 0 \). The fraction of trades accounted for by middlemen is strictly positive and the trading rate distribution has a Pareto tail with parameter 2, meaning \( \lim_{\alpha \to \infty} \alpha^{2}(1 - \hat{G}(\alpha)) \) is positive and finite.

\textsuperscript{10}Our definition of a tail index follows the definition of a tail index \( \chi \) offered in Gabaix, Gopikrishnan, Plerou and Stanley (2006). It implies that the fraction of traders with contact rate larger than \( \lambda \) is proportional to \( \lambda^{-\chi} \). In this definition, a larger tail index implies a thinner tail.
Intuitively, the trading rate inherits the tail properties of the contact rate distribution, since the trading probability conditional on a meeting converges to a positive constant in the tail. This also ensures that middlemen account for a positive fraction of trades.

Thus, with linear cost, our setup gives rise to a distribution of trading rates that looks like its empirical counterpart. We highlight that the empirical literature frequently finds tail indices close to 2. For instance, Gabaix, Gopikrishnan, Plerou and Stanley (2006) report a tail parameter of 1.5 for the distribution of trading volume in the stock market.

### 4.8 Constant Marginal Cost: Frictionless Limit

For many real world markets, frictions are small and so a natural question is whether intermediation retains its prominent role in the frictionless limit and indeed whether we obtain any insights from studying frictions in markets where frictions are small. This section therefore focuses on the model with a linear cost when the cost of a meeting is 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 that heterogeneity and intermediation are preserved in the limit.\(^{11}\) In particular, we obtain a sharp characterization of volume, that is the rate at which assets are traded, and how that volume is divided between different types of trades.

We measure volume as the rate at which a trader buys the asset. If \( C(\lambda) = c\lambda \), we can decompose volume as \( \mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2 + \mathcal{V}_3 \) where

\[
\mathcal{V}_1 = \frac{\Lambda}{4} \lim_{\lambda \to \bar{\lambda}} \int_{\lambda}^{\bar{\lambda}} \left( m_{\lambda}(1 - F(\lambda')) + M(\lambda') \right) dF(\lambda') \\
\mathcal{V}_2 = \frac{\Lambda}{4} dF(\bar{\lambda}) (M(\bar{\lambda}) - dM(\bar{\lambda})) \\
\mathcal{V}_3 = \frac{\Lambda}{4} dF(\bar{\lambda})^2 m_{\bar{\lambda}}^2.
\]

\( \mathcal{V}_1 \) is the rate at which traders with contact rate \( \lambda < \bar{\lambda} \) purchase the asset. \( \mathcal{V}_2 \) is the rate at which traders with contact rate \( \bar{\lambda} \) buy the asset from slower traders. \( \mathcal{V}_3 \) is the rate at which the fastest traders buy from one another. For a typical trader with contact rate \( \lambda < \bar{\lambda} \), the trading probability conditional on a meeting is \( p_\lambda = \frac{1}{2} (m_\lambda (1 - F(\lambda)) + M(\lambda)) \). To calculate volume, we only count asset purchases, in a fraction \( \frac{1}{2} p_\lambda \) of meetings. For a trader with contact rate \( \bar{\lambda} \), the purchase probability is discretely lower, both because they never meet a faster trader and because we assume traders with equal contact rate do not trade with each other.

\(^{11}\)The order of limits is important. We first take a limiting equilibrium, where \( \bar{\lambda} \) grows without bound, and then take the limit as \( c \) converges to zero. With the opposite order of limits, there is no intermediation. 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.
other if one is misaligned and the other is well-aligned: $\frac{1}{2}p_\lambda = \frac{1}{4}(M(\bar{\lambda}) - (1 - m_\lambda)dM(\bar{\lambda}))$.

We then integrate these buying probabilities using the contact frequencies $\Lambda dF(\lambda)$. We define volume in a limiting equilibrium as the obvious limit; see the proof of Proposition 8 for details.

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

**Proposition 8** Assume $C(\lambda) = c\lambda$. Consider a sequence of limiting equilibria as $c$ converges to zero. The aggregate trading volume $V$ converges to a number between $2\gamma$ and $\frac{5}{2}\gamma$ and can be decomposed as follows: middlemen’s purchases from middlemen account for a volume of $V_3 = \frac{1}{2}\gamma$; middlemen’s purchases from non-middlemen account for a volume of $V_2 = \frac{1}{2}\gamma$; and non-middlemen’s purchases account for the remaining volume $V_1$, which lies between $\gamma$ and $\frac{3}{2}\gamma$. $V_1$ can be further decomposed into $\frac{1}{2}\gamma$ purchases from middlemen and between $\frac{1}{2}\gamma$ and $\gamma$ purchases from other non-middlemen.

The proof of Proposition 8 provides exact expressions for volume (including $V_1$), the fraction of meetings with middlemen, and the ratio between the average contact rate $\Lambda$ and the lower bound on contact rates $\bar{\lambda}$ (both of which converge to infinity in the limit). It also provides an implicit equation for the distribution of the relative contact rate $\rho \equiv \lambda/\bar{\lambda}$.

We contrast Proposition 8 with a naive view of a market without frictions: all traders can trade instantaneously upon receiving a preference shock and only trade with other traders who receive the opposite preference shock at the same instant. That means that volume equals the share of traders in the low preference state times the rate at which they are hit by preference shocks, $\frac{1}{2}\gamma$. Note that this view leaves no role for intermediation or middlemen. In contrast, we obtain four to 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 by both finite traders and middlemen is preserved in the limiting economy.

To understand this result note that we are looking at a frictionless limit where almost no one is misaligned. Whenever a trader (who is almost surely a finite trader) suffers a preference shock, she is very likely to become misaligned and very unlikely to contact another misaligned trader, $\lim_{\lambda \to \infty} M(\lambda) \to 0$. As a consequence, a double-coincidence of wants becomes 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 traders. This piece of the process stops once the misalignment is passed to a middleman. What the volume decomposition shows is that, in the frictionless limit, the reallocation of the asset in response to taste
shocks runs through an intermediation chain that always involves middlemen. Every time there is a preference shock, there is a trade between a middleman and a finite trader. Half these trades are asset purchases and half are asset sales.

When middlemen take on the asset from a finite trader they, too, move into misalignment. Afterwards, they quickly trade away from misalignment, but only by meeting another middleman. As a consequence trade from middlemen to middlemen accounts for a volume of $\frac{1}{2}\gamma$. The reason is that, when $c$ is small, the misalignment rate of finite traders is proportional to the square of the misalignment rate of middlemen. Thus, as misalignment converges to zero, traders are far more likely to contact misaligned middlemen than misaligned finite traders. Taken together, whenever a trader experiences a taste shock the market rapidly reallocates her asset position to a misaligned finite type with opposite preferences. But instead of doing so directly, the position gets traded through an intermediation chain. This chain runs through faster finite types towards middlemen who then first reallocate the position internally before passing it back to misaligned finite types with opposite preferences.

5 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 $\lambda$ at birth and on whether to trade in each future meeting, but who cannot directly alleviate the search frictions in the economy.

5.1 Planner Problem

With transferable utility, any Pareto optimal allocation also solves the problem of a utilitarian planner that equally weights all traders’ welfare. Since traders do not discount the future (except through death, in which case they get replaced by another trader), this corresponds to maximizing undiscounted utility net of the cost of meetings. We therefore focus on a planner who seeks to maximize steady state welfare. Moreover, we impose a symmetric trading pattern on the planner, as in equilibrium, so we only need to keep track of the misalignment rate $m_\lambda$ at each contact rate. The planner’s objective is thus to maximize

$$\delta_0 + \Delta \int_0^{\lambda} (1 - m_\lambda) dG(\lambda) - r \int_0^{\lambda} C(\lambda) dG(\lambda).$$

Each misaligned trader gets a flow payoff of $\delta_0$, while each well-aligned trader gets a flow payoff of $\delta_1 = \Delta + \delta_0$, expressed in the first two terms. In addition, the planner must
repay the search costs at the death rate. Using equation (1), the planner’s objective can alternatively be written as

\[ \delta_1 - \Lambda \int_0^\lambda \frac{\Delta m_\lambda + rC(\lambda)}{\lambda} dF(\lambda). \]  

(26)

The planner has two instruments. First, she chooses the set of admissible trades: In particular, when a trader with contact rate \( \lambda \) and alignment status \( a \in \{0,1\} \) meets a trader with contact rate \( \lambda' \), alignment status \( a' \in \{0,1\} \), and opposite asset position, they trade with time-invariant probability \( 1_{\lambda,a}^{\lambda',a'} \in [0,1] \), chosen by the planner. This implies that the steady state misalignment rate \( m_\lambda \in [0,1] \) satisfies equation (7). Second, the planner chooses the time-invariant counterparty distribution \( F(\lambda) \), a nondecreasing, right-continuous function \( F : [0,\bar{\lambda}] \rightarrow [0,1] \) with \( F(\bar{\lambda}) = 1 \). The solution to the planner’s problem is thus a distribution \( F(\lambda) \) and trading probabilities \( 1_{\lambda,a}^{\lambda',a'} \in [0,1] \) that maximize (26) subject to (7).

In the appendix, we derive the necessary conditions for the planner’s constrained optimization problem. These conditions are closely related to the three conditions satisfied by an equilibrium allocation as defined in Definition 1. First, the first order conditions yield an expression for the social net value of asset ownership, \( S(\lambda) \), which is the Lagrange multiplier on the misalignment constraint (7). The functional form of \( S(\lambda) \) is closely related to the private surplus \( s(\lambda) \). In particular, using the approach developed for decentralized equilibrium, we show that it is decreasing and nonnegative, satisfying

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

(27)

where

\[ \Phi_{\lambda} \equiv \frac{2(r + 2\gamma)}{\lambda(2(r + 2\gamma) + \lambda - F(\lambda) + 2M(\lambda))}. \]  

(28)

The surplus function is scarcely changed from equations (15) and (16) in the decentralized equilibrium.

Second, the planning allocation is subject to the same symmetric flow-balance equation that governs equilibrium.

And third, we prove in equation (78) in the appendix that the planner sets \( dF(\lambda) > 0 \) only if \( \lambda \) maximizes

\[ \Pi_\lambda \equiv -\gamma S(\lambda) + \frac{\lambda}{2} \int_0^\lambda (S(\lambda') - S(\lambda))dM(\lambda') - rC(\lambda) - \theta \lambda. \]  

(29)

Condition (29) is scarcely changed from condition (12), i.e. the third part of the definition of
equilibrium. The new piece is the linear term \( \theta \lambda \), where \( \theta \) is the Lagrange multiplier on the constraint \( \int_0^\lambda dF(\lambda) = 1 \). \( \theta \) is fixed from the perspective of this optimization problem. The general expression for \( \theta \) is cumbersome, and is provided in equation (88) in the appendix. However, if the upper bound on the contact rate distribution is not binding or in the limiting allocation with \( \lambda \to \infty \), we prove that \( \theta \) is simply the annuitized, contact-rate-weighted average marginal cost, \( \theta = r \int_0^\lambda C'(\lambda)dF(\lambda) \).

Thus one key result is that a set of necessary conditions for the planning problem that has an almost-identical mathematical structure to the definition of equilibrium. Henceforth, we refer to any allocation satisfying those conditions as an “optimum” or as an “optimal allocation”.

Our main result is that the equilibrium and optimum allocations are qualitatively similar:

**Proposition 9** The characterization of equilibrium in propositions 1–8 applies to any optimum.

The first piece of this proposition is that the equilibrium trading pattern is optimal. The planner requires trade whenever the joint social surplus of the transaction is positive. Moreover, the social surplus function, just like the private surplus function, is strictly positive and strictly decreasing. Using this observation, the intuition for the result 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. That is, the planner uses faster traders as intermediaries. This does not affect the current misalignment rate, but improves future trading possibilities. It follows that the intermediated trading pattern governing equilibrium is optimal and the allocation would be strictly inferior if traders could only engage in fundamental trades.

The atomless feature of the optimal contact rate distribution allows the planner to leverage the gains from meetings through intermediation. That is, any meeting between two traders with identical contact rates \( \lambda \) is beneficial solely in the double-coincidence case. In contrast, when two traders with different contact rates meet each other, the meeting is socially beneficial not only in the double-coincidence case but also when misalignment can be traded towards the faster trader. An atomless distribution maximizes the fraction of meetings in which there are gains from trade.

Furthermore, the proposition implies that with a linear cost function, the optimal distribution has a fat tail and features middlemen when \( \lambda \to \infty \). The fat tail (and tail parameter

\[12\] We highlight that we do not explicitly check the second order conditions.
2) arises from exactly the same logic discussed above for the equilibrium allocation. Given
the linear cost function, the fat tail equates the marginal social returns to meetings created
by different types \( \lambda \).

The planner introduces middlemen for reasons that likewise mimic the equilibrium case.
A trader’s social value consists of the direct flow valuation she derives from her asset holdings
and from the impact she has on the overall allocation net of the cost of meetings. The latter
component again becomes linearly scalable as \( \lambda \) becomes very high; when a fast trader
becomes even faster, she conducts the same types of trade, just more frequently. It thus
follows that the net social value that arises from trade by fast types converges to zero when
fast types coexist, as is the case with the Pareto tail. But in the absence of middlemen,
fast traders have a misalignment rate converging to \( \frac{1}{2} \), worse than autarky where their
misalignment rate is \( \frac{1}{2} \).

Finally, with a linear cost function and the per-meeting cost \( c \) approaching zero, an
optimal allocation shares the same qualitative features as equilibrium. However, the exact
magnitudes are different. In an optimal allocation, the fraction of meetings that goes to
middlemen is \( e^{-1} \), more than in equilibrium. Volume is \( (e - \frac{1}{2}) \gamma \), which, as we show in the
proof, is less than equilibrium volume. As in equilibrium, middlemen buy from middlemen
at rate \( \frac{1}{2} \gamma \) and from finite traders at rate \( \frac{1}{2} \gamma \), while finite traders buy from middlemen at
rate \( \frac{1}{2} \gamma \). That is, just like in equilibrium, the reallocation of the asset following a taste shock
happens through an intermediation chain that always involves two middlemen. The only
difference from the equilibrium allocation is that the optimal volume of purchases by one
finite trader of the asset held by another is slightly lower, namely \( (e - 2) \gamma \). Equilibrium
features too much trading in the limiting case.

We can also characterize the limiting behavior of the optimal search intensity distribution
in closed form. Let \( \Psi \) denote the distribution of relative contact rates \( \rho \equiv \lambda/\Lambda \), so \( \Psi(\rho) = G(\rho\Lambda) \) for all \( \rho \geq 1 \). When costs converge zero, \( \Lambda \) grows without bound but \( \Psi(\rho) \) converges
to \( (1 - \rho^{-1})e^{\rho^{-1}} \). Compared to the equilibrium contact rate distribution provided in the
appendix, the equilibrium distribution is more concentrated than the optimal solution. These
discrepancies between the optimal and the equilibrium allocation in the frictionless limit are
consistent with the numerical illustrations provided in the next subsection.

### 5.2 Equilibrium vs. Optimum: Pigouvian Taxes and Numerical Illustration

Although the qualitative features of the equilibrium and optimal allocations are identical,
the equilibrium allocation is still inefficient. There are two sources of inefficiency. The first
comes from bargaining. In the decentralized equilibrium, each trader keeps only half of the surplus from every meeting, while the social planner recognizes the value of the entire surplus. This force induces traders to underinvest in meetings in equilibrium. Formally, this is readily observable by contrasting equations (17) and (29): The planner puts twice the weight on the option value of trade compared to the individual trader.

The second source of inefficiency is a business stealing effect. When one trader invests more in meetings, she diverts meetings towards herself. That is, investing more in meetings does not affect the contact rate of the other market participants, but it changes the distribution of whom they meet. The failure to internalize this effect causes traders to overinvest in meetings in equilibrium. This business stealing effect is represented by the last (negative) term in (29). There is no corresponding term in the equilibrium condition (17).

To further illustrate this, note that we can directly correct for each of these externalities. First, assume that whenever two traders meet, an outside agent doubles the surplus from the meeting. Second, assume that whenever a trader meets anyone, the outside agent charges her a tax which is independent of her contact rate, \( \bar{\tau} = \theta \). We show in appendix B.2 how this alters the value functions, the derivation of surplus, and the optimality of the ex-ante investment decision. It follows directly from those derivations that the proposed tax and transfer system decentralizes the optimum.

Importantly, in the linear cost case with no upper bound, appendix B.2 also shows that the optimal tax is \( \bar{\tau} = rc \). To understand this finding, first note that the social value of an average meeting in an optimal allocation is the cost of that meeting, \( 2rc \) for all traders. The subsidy we propose doubles the joint surplus in each meeting, hence adds on average an additional \( 2c \). Since the subsidy is split symmetrically, the average expected subsidy per trader and meeting is \( rc \). To understand the tax, it is easiest to think about the business stealing externality in the following way: When increasing her contact rate, a trader effectively prevents others from meeting each other. In particular, for each two additional meetings a trader “acquires”, she replaces one average meetings between two other traders. In an optimal allocation, that average meeting has value \( 2rc \). Thus, the tax that internalizes the business stealing externality equals \( rc \) in that case.

We proceed by numerically contrasting limiting equilibria with their optimal counterpart along several dimensions for the linear cost case. To start, we impose \( r = \gamma = 1 \) and consider the effect of changes in \( c/\Delta \). The blue lines in Figure 2 summarize the optimal allocation, while the red lines show the equilibrium. For any level of costs, the top panel shows that the lower bound on the optimal contact rate distribution is lower than under the equilibrium.

---

13A similar business stealing effect has been documented in various other frictional environments, see Atkeson, Eisfeldt and Weill (2015) for over-the-counter markets.
Figure 2: Features of the equilibrium and optimal contact rate distribution given relative costs $c/\Delta$ with $r = \gamma = 1$. The vertical dotted line indicates the value of $c/\Delta$ where the market shuts down, $c^*$, the same in both equilibrium and the optimal allocation.
contact rate distribution. The second panel shows that the average contact rate is higher relative to the lower bound in the optimum than the equilibrium. The third panel shows middlemen play a more prominent role in the optimum. And the bottom panel suggests that volume is generally inefficiently high in equilibrium. Each of these last three results is consistent with our analytical results with vanishing costs.

To develop an intuition for these observations, we note that the business stealing externality is identical for all types $\lambda$. However, the positive externality that arises from bargaining is larger for those traders with higher average surplus meetings which are those with lower $\lambda$. As a consequence, there are too few slow types and too many fast types and the undistorted equilibrium displays excessive trading volume. On the other hand, given the lower bound on the contact rate, the rest of the distribution is more compressed in equilibrium than in an optimum.

Figure 3 gives more details of the equilibrium and optimum for a particular value of $c/\Delta = 0.001$, still with $r = \gamma = 1$. The top panel shows that the entire distribution of contact rates, both in optimum and equilibrium, is well approximated by a Pareto distribution with tail parameter 2 (slope -2 in the figure). The middle panel shows that this carries over to the distribution of trading rates, implying 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 constant. In equilibrium and optimum, most traders have an optimal trading rate that far exceeds $\gamma$, the natural benchmark for an economy without intermediation. This makes it clear that intermediation is key to understanding the amount of trades in the optimal allocation.

Finally, the bottom panel in Figure 3 plots the misalignment rate against $\lambda$ in both equilibrium and the optimal allocation. As expected from Propositions 6 and 9, traders with a higher contact rate have a higher misalignment rate in both equilibrium and optimum. The equilibrium distribution of contact rates first order stochastically dominates the optimal one; as a consequence, all traders with finite $\lambda$ are more likely to meet with a faster trader when comparing equilibrium with the optimal case. The reason the misalignment rate crosses is that the planner allocates a larger fraction of meetings to middlemen; see the bottom panel of Figure 2. This disproportionately benefits fast traders since they can offload misalignment.

6 Constrained Economy: The Role of Intermediation

To understand the role of intermediation and its connection with heterogeneity, consider an economy in which intermediation is not allowed. To be concrete, suppose meetings between two traders with the same preference state simply do not occur. It follows that whenever
Figure 3: Equilibrium and optimal contact rate distribution, trading rate distribution, and misalignment rates. We set $c/\Delta = 0.001$ and $r = \gamma = 1$. 

35
a misaligned trader meets a well-aligned trader, they have opposite preference states 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 distribution of contact rates are degenerate as long as the cost function $C(\lambda)$ is weakly convex.

**Constrained Equilibrium**  We start by defining an equilibrium without intermediation analogously to the definition of equilibrium with intermediation.

**Definition 3** An equilibrium is an increasing, right-continuous counterparty distribution $F : [0, \bar{\lambda}] \to [0, 1]$ with $F(\bar{\lambda}) = 1$, a misalignment rate function $m : [0, \bar{\lambda}] \to [0, 1]$, and a surplus function $s : [0, \bar{\lambda}] \to \mathbb{R}$, satisfying:

1. the surplus equation adjusted for the no-intermediation case,

$$\Delta = (r + 2\gamma)s(\lambda) + \frac{\lambda}{4} \int_0^{\bar{\lambda}} \left( (s(\lambda) + s(\lambda'))^+ m_{\lambda'} - (-s(\lambda) - s(\lambda'))^+(1 - m_{\lambda'}) \right) dF(\lambda')$$

2. the balanced inflow-outflow adjusted for the no-intermediation case,

$$\left( r + \gamma + \frac{\lambda}{2} \int_0^{\bar{\lambda}} (\mathbb{1}_{s(\lambda) + s(\lambda') > 0} m_{\lambda'}) dF(\lambda') \right) m_{\lambda} = \left( r + \gamma + \frac{\lambda}{2} \int_0^{\bar{\lambda}} (\mathbb{1}_{s(\lambda) + s(\lambda') < 0} (1 - m_{\lambda'})) dF(\lambda') \right) (1 - m_{\lambda}).$$

3. optimality of the ex-ante investment decision: $\lambda$ is in the support of $F$ only if $\pi_{\lambda} = \max_{\lambda'} \pi_{\lambda'}^{NI}$ where $\pi_{\lambda}^{NI}$ is defined in the no-intermediation case as

$$\pi_{\lambda}^{NI} = \frac{\delta_1 - \gamma s(\lambda) + \frac{\lambda}{4} \int_0^{\bar{\lambda}} ((-s(\lambda) - s(\lambda'))^+(1 - m_{\lambda'})) dF(\lambda')}{r} - C(\lambda).$$

Note that there are two relevant types of meetings in this constrained economy, those between two misaligned traders with the opposite asset holdings, and those between two well-aligned traders with the 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 find it optimal to jointly trade into misalignment.

**Constrained Planner’s Problem**  We turn next to the planner’s problem. The planner again chooses the distribution of contact rates 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 (26). 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,

\[
\left( r + \gamma + \frac{\lambda}{2} \int_0^{\bar{\lambda}} \left( 1_{\lambda',\lambda} \left( 1 - m_{\lambda'} \right) dF(\lambda') \right) \right) m_\lambda = \left( \gamma + \frac{\lambda}{2} \int_0^{\bar{\lambda}} \left( 1_{\lambda',\lambda} \left( 1 - m_{\lambda'} \right) dF(\lambda') \right) \right) (1 - m_\lambda).
\] (33)

The following proposition then summarizes our key findings for the environment where solely fundamental trades between traders with different preferences are allowed.

**Proposition 10** Consider an economy with no intermediation and a weakly convex cost \( C(\lambda) \). In equilibrium all traders choose a common value \( \lambda = \Lambda \). The same holds in the solution to the planner’s problem. With a continuously differentiable cost function and interior solution, the equilibrium contact rate is inefficiently high.

The second part of the proposition highlights that the overinvestment result we prove for the limiting case above also holds in the constrained case, for any equilibrium with \( \Lambda > 0 \). Individual traders do not internalize that investing in additional meetings lowers misalignment in the marketplace thereby reducing the marginal value of meetings acquired by others. This force leads to overinvestment in the aggregate.

More importantly, the first part of the proposition highlights that the heterogeneity that arises in the full economy is an immediate, and socially desirable, consequence of intermediation. When traders are restricted to trades driven by static fundamentals, there is no gains 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, the only distribution where ex-ante homogeneous traders have identical value is a degenerate one.

In summary, intermediation and heterogeneity are closely 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 preferences, misalignment can be transmitted towards the faster trader to facilitate the transfer of the asset to those with a desire for it.
7 Conclusions

We study a model of over-the-counter trading in which ex-ante identical traders invest in a meeting 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 endogenous and optimal distribution of trading rates is governed by a power law, an empirical hallmark of various financial markets. Moreover, middlemen with an infinite 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 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 natural to extend our model to a more complex environment, for example one in which the two misaligned states are not symmetric, or one in which the binary restriction on asset holdings is relaxed. We believe that the basic forces we highlight in this paper will be robust to such extensions. 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


A Appendix

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

To show prove $s$ is positive, suppose to the contrary that $s(\lambda) \leq 0$ for some $\lambda$. Then $(-s(\lambda) + s(\lambda'))^+ \geq (s(\lambda) + s(\lambda'))^+$ and $(-s(\lambda) - s(\lambda'))^+ \geq (s(\lambda) - s(\lambda'))^+$ for all $\lambda'$. Equation (10) then implies $\Delta \leq (r + 2\gamma)s(\lambda)$. Since $\Delta \geq 0$, this is a contradiction, which proves $s$ is positive.

Next, the definition of equilibrium imposes that $m_\lambda$, the fraction of traders with contact rate $\lambda$ who are misaligned, lies between 0 and 1. Then since $s$ is strictly positive, it follows that for all $\lambda$ and $\lambda'$, $\mathbb{I}_{s(\lambda) + s(\lambda') > 0}m_\lambda \geq \mathbb{I}_{s(\lambda) < s(\lambda') m_\lambda}$ and $\mathbb{I}_{s(\lambda) > s(\lambda')} (1 - m_\lambda') \geq \mathbb{I}_{s(\lambda) + s(\lambda') < 0} (1 - m_\lambda')$. Equation (11) then implies $m_\lambda \leq 1/2$ for all $\lambda$.

Next, since $s$ is 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 (10) as

$$(r + 2\gamma)s(\lambda) = \Delta + \frac{\lambda}{4} \int_0^\lambda \left( (\min\{s(\lambda), s(\lambda')\} - s(\lambda)) m_{\lambda'} - (s(\lambda) - \min\{s(\lambda), s(\lambda')\}) (1 - m_{\lambda'}) \right) dF(\lambda').$$

Grouping terms, this gives

$$s(\lambda) = \frac{4\Delta + \lambda \int_0^\lambda \left( \min\{s(\lambda), s(\lambda')\} (1 - 2m_{\lambda'}) \right) dF(\lambda')}{4r + 8\gamma + \lambda} = T(s)_\lambda,$$

(34)

where $T$ maps surplus functions into surplus functions. We claim that for any cumulative distribution function $F$ and misalignment function $m$ with range $[0, 1/2]$, $T$ 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 $s \leq \Delta/(r + 2\gamma)$,

$$T(s)_\lambda \leq \left( \frac{4r + 8\gamma + \lambda \int_0^\lambda (1 - 2m_{\lambda'}) dF(\lambda')}{4r + 8\gamma + \lambda} \right) \left( \frac{\Delta}{r + 2\gamma} \right).$$

Since the misalignment rate is nonnegative, the result follows.

Finally, we prove $T$ is a contraction. If $|s^1_\lambda - s^2_\lambda| \leq \varepsilon$ for all $\lambda$,

$$|T(s^1)_\lambda - T(s^2)_\lambda| \leq \frac{\lambda \varepsilon \int_0^\lambda (1 - 2m_{\lambda'}) dF(\lambda')}{4r + 8\gamma + \lambda} \leq \frac{\bar{\lambda} \varepsilon}{4r + 8\gamma + \lambda}.$$
The second inequality uses $\int_0^\bar{\lambda} (1 - 2m_{\lambda'}) dF(\lambda') \leq 1$ 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 $\frac{\bar{\lambda}}{4r + 8\gamma + \bar{\lambda}} < 1$.

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

$$E(\lambda) \equiv \int_0^\lambda \min\{s(\lambda), s(\lambda')\}(1 - 2m_{\lambda'}) dF(\lambda').$$

Note that $m$ nonnegative and $s(\lambda) \leq \Delta/(r+2\gamma)$ implies $E(\lambda) \leq \Delta/(r+2\gamma)$ as well. Similarly, $s$ nonincreasing implies $E$ is nonincreasing as well. Then

$$T(s)_{\lambda_1} - T(s)_{\lambda_2} = \frac{4\Delta + \lambda_1 E(\lambda_1)}{4r + 8\gamma + \lambda_1} - \frac{4\Delta + \lambda_2 E(\lambda_2)}{4r + 8\gamma + \lambda_2} \geq \frac{4\Delta + \lambda_1 E(\lambda_1)}{4r + 8\gamma + \lambda_1} - \frac{4\lambda_2 - \lambda_1}{(4r + 8\gamma + \lambda_1)(4r + 8\gamma + \lambda_2)} > 0,$$

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. \[\square\]

**Proof of Lemma 1.** Since the surplus function is nonnegative and nonincreasing, we can rewrite equation (10) as

$$\Delta = (r + 2\gamma)s(\lambda) + \frac{\lambda}{4} \int_0^\lambda \left( ((s(\lambda) + s(\lambda')) - \mathbb{I}_{\lambda' < \lambda}(s(\lambda') - s(\lambda))) m_{\lambda'} + (\mathbb{I}_{\lambda' > \lambda}(s(\lambda) - s(\lambda')))(1 - m_{\lambda'}) \right) dF(\lambda').$$

Regrouping terms gives

$$4\Delta = 2\left(2(r + 2\gamma) + \lambda \int_0^\lambda m_{\lambda'} dF(\lambda') \right) s(\lambda) + \lambda \int_0^\lambda \left( s(\lambda) - s(\lambda') \right) (1 - 2m_{\lambda'}) dF(\lambda'). \quad (35)$$

Since $s$ is monotonic, it is almost everywhere differentiable. At such points, we can different-
tiate equation (35) with respect to $\lambda$ to get

$$0 = \left( 4(r + 2\gamma) + \lambda \int_{\lambda}^{\bar{\lambda}} dF(\lambda') + 2\lambda \int_{0}^{\lambda} m_{\lambda'} dF(\lambda') \right) s'(\lambda)$$

$$+ 2 \int_{0}^{\lambda} m_{\lambda'} dF(\lambda') s(\lambda) + \int_{\lambda}^{\bar{\lambda}} (s(\lambda) - s(\lambda'))(1 - 2m_{\lambda'}) dF(\lambda')$$

Replace the last two terms using equation (35) and simplify the first term using the definitions of $F$ and $M$:

$$s'(\lambda) = x(\lambda, s(\lambda)) \equiv \frac{4((r + 2\gamma)s(\lambda) - \Delta)}{\lambda \left( 4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda)) \right)}.$$  \hspace{1cm} (36)

Also evaluate equation (35) at $\lambda = \bar{\lambda}$ and simplify using the definition of $M$:

$$s(\bar{\lambda}) = \frac{2\Delta}{2(r + 2\gamma) + \lambda M(\lambda)}.$$  \hspace{1cm} (37)

Equations (36) and (37) are an initial value problem. Since $F$ and $M$ are monotone, they have only countably many discontinuities. Since also $0 \leq F(\lambda) \leq 1$ and $0 \leq M(\lambda) \leq 1/2$, $x$ on all sets $(\lambda, \bar{\lambda}]$ with $\lambda > 0$. It follows that the initial value problem has a unique solution; see the discussion around equation (5.4) in chapter 1 of Hale (1980). The solution is

$$s(\lambda) = \frac{\Delta}{r + 2\gamma} \left( 1 - \frac{\bar{\lambda}M(\bar{\lambda})e^{-\int_{\lambda}^{\bar{\lambda}} \phi_{\lambda', \alpha} d\lambda'}}{2(r + 2\gamma) + \lambda M(\lambda)} \right),$$

where $\phi_{\lambda}$ is defined in equation (16). Now use the fact that $F(\lambda') = 1$ and $M(\lambda') = M(\bar{\lambda})$ for all $\lambda' > \bar{\lambda}$, as well as the definition of $\phi$ in equation (16), to get

$$e^{-\int_{\lambda}^{\infty} \phi_{\lambda', \alpha} d\lambda'} = \frac{\bar{\lambda}M(\bar{\lambda})}{2(r + 2\gamma) + \lambda M(\lambda)}.$$  \hspace{1cm} (15)

Substituting this into the previous equation gives us equation (15).

Finally, note that this argument does not apply when $\lambda = 0$, but one can verify that $s(0) = \Delta/(r + 2\gamma)$ directly from equation (35).

Proof of Proposition 3. We proceed by contradiction. Assume $F$ has a mass point at
some $\lambda^* \in (0, \bar{\lambda})$. Rewrite equation (13) as

$$m_\lambda = \frac{2\gamma + \lambda M(\lambda)}{2r + 4\gamma + \lambda(1 - F(\lambda) + (1 - m_\lambda)dF(\lambda) + 2M(\lambda))}$$

The right hand side is strictly positive and it is strictly less than $\frac{1}{2}$ since $1 - F(\lambda) \geq 0$, $M(\lambda) \geq 0$, and $(1 - m_\lambda)dF(\lambda) \geq 0$. Thus $0 < dM(\lambda) < \frac{1}{2}dF(\lambda)$.

The definition of $F$ and $M$ imply that $1 - F(\lambda) + 2M(\lambda)$ is continuous in $\lambda$, except at mass points, where it jumps down. It follows from equation (16) that $\phi_\lambda$ is continuous in $\lambda$ except at mass points, where it jumps up. Next, differentiating equation (15), we have that

$$s'(\lambda) = \frac{-\Delta \phi_\lambda e^{-\int_0^\infty \phi_{\lambda'} d\lambda'}}{r + 2\gamma}.$$ 

This is again continuous in $\lambda$ except at mass points, where it jumps down (becomes more negative). Thus $s(\lambda)$ has a concave kink at $\lambda^*$.

Now consider part 3 of the definition of equilibrium. The choice of $\lambda$ must maximize $\pi_\lambda$ in (17). The first derivative with respect to $\lambda$ is

$$- (\gamma + \frac{1}{4}M(\lambda)) s'(\lambda) + \frac{1}{2} \int_0^\lambda (s'(\lambda') - s(\lambda))m_\lambda dF(\lambda') - C''(\lambda).$$

This function is again continuous except at mass points, where it jumps up. That is, the objective function $\pi_\lambda$ has a convex kink at $\lambda^*$. This point cannot be a maximum of the traders’ objective function, contradicting the assumption that $F$ has a mass point at $\lambda^*$.

**Proof of Proposition 4.** Let $A$ denote the support of $F(\lambda)$, the smallest closed set such that $\int_A dF(\lambda) = 1$. To find a contradiction, suppose that there is a hole in $A$. That is, there are contact rates $\lambda_1, \lambda_2 \in A$ with $0 \leq \lambda_1 < \lambda_2 \leq \bar{\lambda}$ and for all $\lambda \in (\lambda_1, \lambda_2)$, $F(\lambda) = F(\lambda_1)$.

If $\lambda_1 = 0$, the assumption that $\lambda_1 \in A$ implies $F(0) > 0$. It follows that $\int_0^\lambda \frac{1}{\lambda} dF(\lambda)$ does not converge and so equation (2) implies $\Lambda = 0$. Thus $\lambda_1 > 0$.

Since $0 < \lambda_1 < \bar{\lambda}$, Proposition 3 implies $\lambda_1$ is not a mass point of $F$. The fact $\lambda_1 \in A$ and $F$ is almost everywhere differentiable then implies that there is a $\lambda_0 < \lambda_1$ with $F(\lambda)$ strictly increasing on $\lambda \in (\lambda_0, \lambda_1)$. It follows that $\pi_\lambda$ must be constant on this interval, and in particular $\pi'_\lambda = \pi''_\lambda = 0$ for almost all $\lambda \in (\lambda_0, \lambda_1)$.

We turn next to evaluating $\pi''_\lambda$. Start with equation (21). Eliminate $s''_\lambda$ using equation (19), then eliminate $s'_\lambda$ by differentiating equation (15), and finally eliminate $\phi_\lambda$ using
its definition in equation (16):

\[
\pi''_\lambda = \frac{\Delta e^{-\int_\lambda^\infty \phi d\lambda'} r M(\lambda) - \gamma(1 - F(\lambda))}{r \lambda (4(r + 2\gamma) + \lambda (1 - F(\lambda) + 2M(\lambda)))^2} \left( 8(r M(\lambda) - \gamma(1 - F(\lambda))) + \lambda \left(M'(\lambda)(4r + \lambda(1 - F(\lambda))) + F'(\lambda)(4\gamma + \lambda M(\lambda))\right) \right) - C''(\lambda).
\]

This defines the second derivative of the profit function almost everywhere, at all points where \( F \) and \( M \) are differentiable.

\( F' \) and \( M' \) may have discontinuities at \( \lambda_1 \): \( \lim_{\lambda \to \lambda_1} F'(\lambda) \geq 0 = \lim_{\lambda \to \lambda_1} F'(\lambda) \) and \( \lim_{\lambda \to \lambda_1} M'(\lambda) \geq 0 = \lim_{\lambda \to \lambda_1} M'(\lambda) \). Using equation (38), this implies \( \lim_{\lambda \to \lambda_1} \pi''_\lambda \geq \lim_{\lambda \to \lambda_1} \pi''_\lambda \). Since \( \lim_{\lambda \to \lambda_1} \pi''_\lambda = 0 \), this implies \( \lim_{\lambda \to \lambda_1} \pi''_\lambda \leq 0 \).

Next, suppose that \( \pi''_\lambda \geq 0 \) for some \( \lambda \in (\lambda_1, \lambda_2) \). Since \( F(\lambda) = F(\lambda_1) \) for all \( \lambda \in (\lambda_1, \lambda_2) \), \( M(\lambda) = M(\lambda_1) \) as well. Substituting this into equation (38) implies that

\[
\pi''_\lambda = \frac{8\Delta e^{-\int_\lambda^\infty \phi d\lambda'} r M(\lambda) - \gamma(1 - F(\lambda))}{r \lambda (4(r + 2\gamma) + \lambda (1 - F(\lambda) + 2M(\lambda)))^2} - C''(\lambda) \geq 0.
\]

Since \( C''(\lambda) > 0 \), this implies

\[
r M(\lambda_1) - \gamma(1 - F(\lambda_1)) > 0.
\]

Next, since the second derivative of the cost function is increasing, it is almost everywhere differentiable. Thus we can differentiate (39) and use equation (16) to eliminate \( \phi' \):

\[
\pi''''_\lambda = \frac{-24\Delta e^{-\int_\lambda^\infty \phi d\lambda'} r M(\lambda) - \gamma(1 - F(\lambda))}{r \lambda (4(r + 2\gamma) + \lambda (1 - F(\lambda) + 2M(\lambda)))^3} - C'''(\lambda).
\]

Since \( r M(\lambda_1) - \gamma(1 - F(\lambda_1)) > 0 \), the first term is strictly negative. By assumption, \( C'''(\lambda) \geq 0 \). This proves that at \( \lambda \in (\lambda_1, \lambda_2) \), if \( \pi''_\lambda \geq 0 \), \( \pi''_\lambda \) is also strictly decreasing. This contradicts \( \lim_{\lambda \to \lambda_1} \pi''_\lambda \leq 0 \). The contradiction implies \( \pi''_\lambda < 0 \) for all \( \lambda \in (\lambda_1, \lambda_2) \).

Finally, \( \lambda_1, \lambda_2 \in \mathcal{A} \) implies \( \pi_{\lambda_1} = \pi_{\lambda_2} \geq \pi_{\lambda} \) for all \( \lambda \) from part (3) of the definition of equilibrium. But this is inconsistent with continuity and strict concavity of the profit function on \([\lambda_1, \lambda_2]\). This proves that there cannot be a hole in \( \mathcal{A} \). \( \blacksquare \)

**Proof of Proposition 5.** Since \( F \) and \( M \) are increasing, they are almost everywhere
differentiable. We can thus differentiate equation (14) to get an expression for \( m'_\lambda \).

\[
m'_\lambda = \frac{2(rM(\lambda) - \gamma(1 - F(\lambda))) + \lambda(2r + \gamma(1 - F(\lambda))) + F'(\lambda)(2\gamma + \lambda M(\lambda))}{(2r + 4\gamma + \lambda(1 - F(\lambda) + M(\lambda))^2}
\]  

Under our assumptions, \( \mathcal{A} \), the support of \( F \), is a convex set (Proposition 4). Thus we need to prove that \( m'_\lambda \) is positive on \( \mathcal{A} \).

For any profit-maximizing \( \lambda \in (0, \bar{\lambda}) \), the logic in the proof of Proposition 4 implies \( \pi''_\lambda = 0 \), where for almost all \( \lambda, \pi''_\lambda \) is given in equation (38). Rewrite this as

\[
8(rM(\lambda) - \gamma(1 - F(\lambda))) + \lambda \left( M'(\lambda)(4r + \gamma(1 - F(\lambda))) + F'(\lambda)(4\gamma + \lambda M(\lambda)) \right) = \frac{r\lambda(4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda)))^2 C''(\lambda)}{\Delta - (r + 2\gamma)s(\lambda)} \equiv \zeta(\lambda) > 0. \tag{41}
\]

Solve equations (14) and (41) for \( M'(\lambda) \) and \( F'(\lambda) \) in terms of \( F(\lambda), M(\lambda), \zeta(\lambda), \) and parameters:

\[
F'(\lambda) = \frac{(2r + 4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda)))(8\gamma(1 - F(\lambda)) - 8rM(\lambda) + \zeta(\lambda))}{2\lambda(\gamma r + 8\gamma + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda))))} \tag{42}
\]

\[
M'(\lambda) = \frac{(2\gamma + \lambda M(\lambda))(8\gamma(1 - F(\lambda)) - 8rM(\lambda) + \zeta(\lambda))}{2\lambda(\gamma r + 8\gamma + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda))))}. \tag{43}
\]

We have shown that these equations hold for almost every \( \lambda \). Since \( F \) and \( M \) are continuous (Proposition 3), as is the surplus function \( s \), and equations (18), (42), and (43) define \( s'(\lambda), F'(\lambda), \) and \( M'(\lambda) \) as real numbers for any \( F(\lambda) \in [0, 1], M(\lambda) > 0, \) and \( s(\lambda) < \Delta/(r + 2\gamma) \), the equations must hold at all \( \lambda \). Using these expressions, we can rewrite equation (40) as

\[
m'_\lambda = \frac{\xi(\lambda)M'(\lambda) + \zeta(\lambda)}{4(2r + 4\gamma + \lambda(1 - F(\lambda) + M(\lambda)))^2}
\]

where

\[
\xi(\lambda) \equiv \frac{2\lambda(\gamma r + 8\gamma + 5\lambda(1 - F(\lambda) + M(\lambda))) + \lambda M(\lambda)(5r + 5\gamma + 3\lambda(1 - F(\lambda) + M(\lambda))))}{2\gamma + \lambda M(\lambda)}
\]

is positive. Then \( M'(\lambda) \geq 0 \) implies \( m'_\lambda > 0 \) for all \( \lambda \in \mathcal{A} \).}

**Proof of Proposition 6.** 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 \( F(0) \in \left[ \frac{r + 2\gamma}{2(r + \gamma)}, 1 \right] \) and \( 1 - F(0) \), respectively. For this to be an equilibrium, we require \( \pi_0 = \pi_{\bar{\lambda}} \geq \pi_\lambda \) for all \( \lambda \in [0, \bar{\lambda}] \). Note that \( F(0) > 0 \) implies \( \Lambda = 0 \) by equation (2).

In such an equilibrium, \( F(\lambda) = F(0) \) and \( M(\lambda) = \frac{\gamma}{r + 2\gamma} F(0) \) for all \( \lambda \in [0, \bar{\lambda}] \); the constant \( \frac{\gamma}{r + 2\gamma} \) follows from evaluating equation (13) at \( \lambda = 0 \). Substitute this into equations (19) and then (21) and use \( C''(\lambda) = 0 \) to get

\[
\pi''_\lambda = \frac{\gamma (r + 2\gamma - 2(r + \gamma) F(0))}{r (4(r + 2\gamma)^2 + \lambda (2\gamma + r(1 - F(0))))} s'(\lambda).
\]

Since \( s'(\lambda) < 0 \), \( F(0) \geq \frac{r + 2\gamma}{2(r + \gamma)} \) implies \( \pi''_\lambda \geq 0 \), i.e. the profit function is globally convex. Then \( \pi_0 = \pi_{\bar{\lambda}} \) implies that these two points are profit maximizing.

Next we find the cost that makes \( \pi_0 = \pi_{\bar{\lambda}} \). Equation (17) implies

\[
\pi_{\bar{\lambda}} - \pi_0 = \frac{(4\gamma + \bar{\lambda} M(0))(s(0) - s(\bar{\lambda}))}{4r} - c\bar{\lambda}.
\]

Eliminate \( s(0) - s(\bar{\lambda}) \) using equation (15):

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

(44)

For a given \( F(0) \geq \frac{r + 2\gamma}{2(r + \gamma)} \), we have \( M(0) = \frac{\gamma}{r + 2\gamma} F(0) \) and pin down \( M(\bar{\lambda}) = M(0) + dM(\bar{\lambda}) \) by finding the unique solution to equation (13) with \( dM(\bar{\lambda}) \in (0, (1 - F(0))/2) \). Notably this solution is continuous in \( F(0) \). By setting the right hand side of equation (45) to zero, we then find the cost \( c \) that gives us an equilibrium with a given value of \( F(0) \). Let \( \bar{c} \) be the solution to this when \( F(0) = \frac{r + 2\gamma}{2(r + \gamma)} \). Let \( \bar{\bar{c}} \) be the solution to this when \( F(0) = 1 \) and so \( M(0) = M(\bar{\lambda}) = \frac{\gamma}{r + 2\gamma} \); in general we cannot order \( \bar{\lambda} \) and \( \bar{\bar{c}} \). The intermediate value theorem implies that for any \( c \in [\min\{\bar{c}, \bar{\bar{c}}\}, \max\{\bar{c}, \bar{\bar{c}}\}] \), there exists an \( F(0) \in \left[ \frac{r + 2\gamma}{2(r + \gamma)}, 1 \right] \) 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 $c = \bar{c}$. Raising the cost further does not change the equilibrium, i.e. we still have $F(0) = 1$, but now $\pi_0 > \pi_\lambda$ and $\pi$ is globally convex.

Finally, given $F$ and $M$, it is straightforward to find the surplus and misalignment rates from equations (15) and (13). 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 $c$ such that if $c \leq c$, there is an equilibrium with $F(\lambda) = 0$ for $\lambda \in [0, \bar{\lambda})$ and $\Lambda = \bar{\lambda}$. This proof is again constructive.

In such an equilibrium, equation (2) implies $\Lambda = \bar{\lambda}$. $F(\lambda) = 0$ implies $M(\lambda) = 0$ for $\lambda \in [0, \bar{\lambda})$. We uniquely recover $M(\bar{\lambda})$ from equation (13). Then equation (17) 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 (10), we can solve explicitly for $s(\lambda) - s(\bar{\lambda})$. This gives

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

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

$$c \leq \frac{4\gamma \Delta M(\bar{\lambda})}{r(4r + 8\gamma + \lambda)(2r + 4\gamma + \lambda M(\bar{\lambda}))} \equiv \zeta; \quad (46)$$

since this implies $c < \frac{4\gamma \Delta M(\bar{\lambda})}{r(4r + 8\gamma + \lambda)(2r + 4\gamma + \lambda M(\bar{\lambda}))}$ for all $\lambda < \bar{\lambda}$. We can again find the surplus and misalignment rates from equations (15) and (13). Thus we have constructed an equilibrium for any $c \leq \zeta$.

**Intermediate Cost.** Finally, we turn to the case where $c \in (\zeta, \bar{c})$. This case is the most complicated. 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 determine $F(\lambda)$ and $M(\lambda)$ and hence the right hand side of equation (24) as continuous functions of $\Lambda$. The third and fourth steps characterize the limiting behavior of $F$ and $M$, and hence the right hand side of equation (24), when $\Lambda \to 0$ and when $\Lambda \to \bar{\lambda}$. Finally, we use the intermediate value theorem to prove that for any $c \in (\zeta, \bar{c})$, there is a $\bar{\lambda} \in (0, \bar{\lambda})$ such that equilibrium is described by the $F$ and $M$ that solve the initial value problem (47).
Step 1: The Initial Value Problem  Consider equations (42) and (43) with \( C''(\lambda) = 0 \). Equation (41) implies \( \zeta(\lambda) = 0 \) and so we can write this as

\[
F'(\lambda) = x_F(\lambda), \quad M'(\lambda) = x_M(\lambda), \quad \text{and} \quad F(\bar{\lambda}) = M(\bar{\lambda}) = 0, \quad (47)
\]

where

\[
x_F(\lambda) \equiv \frac{4(2r + 4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda)))(\gamma(1 - F(\lambda)) - rM(\lambda))}{\lambda(\gamma(8r + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda))))},
\]

\[
x_M(\lambda) \equiv \frac{4(2\gamma + \lambda M(\lambda))(\gamma(1 - F(\lambda)) - rM(\lambda))}{\lambda(\gamma(8r + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda))))}.
\]

Let \((\lambda, F, M) \in \Omega = (0, \bar{\lambda}) \times (-\varepsilon, 1) \times (-\varepsilon, \frac{1}{2})\), where we pick \( \varepsilon > 0 \) but small enough that the denominators of \( x_F \) and \( x_M \) are positive for all \((\lambda, F, M) \in \Omega\). This is our initial value problem.

It is straightforward to verify that the functions \( x_F \) and \( x_M \) are continuous on \( \Omega \), 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 \( \underline{\lambda} < \lambda < \lambda + \delta \) (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 \( F(\bar{\lambda}) = M(\bar{\lambda}) = 0 \), \( \gamma(1 - F(\bar{\lambda})) > rM(\bar{\lambda}) \). 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 \( \gamma(1 - F(\lambda)) > rM(\lambda) \).

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 spcile 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 \( M(\lambda) = \lim_{\lambda \to \bar{\lambda}} M(\lambda') + dM(\bar{\lambda}) \), where the first term is determined by the initial value problem and the second term satisfies equation (13):

\[
\left( r + \gamma + \frac{\bar{\lambda}}{2} \left( \lim_{\lambda \to \bar{\lambda}} M(\lambda) + dM(\bar{\lambda}) \right) \right) dM(\bar{\lambda}) = \left( \gamma + \frac{\bar{\lambda}}{2} \lim_{\lambda \to \bar{\lambda}} M(\lambda) \right) (dF(\bar{\lambda}) - dM(\bar{\lambda})). \quad (48)
\]
This is a quadratic equation for \( dM(\bar{\lambda}) \). Only the smaller root, which satisfies \( dM(\bar{\lambda}) \in (0, dF(\bar{\lambda})/2) \), is valid.

Once we have computed \( F \) and \( M \), we have \( m_\lambda = dM(\lambda)/dF(\lambda) \) at \( \lambda \geq \bar{\lambda} \). At smaller values of \( \lambda \), equation (11) implies \( m_\lambda = \frac{2\gamma}{2r + 4\gamma + \lambda} \). Equation (15) 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 \( \lambda \) but constant; and profits must be weakly lower at other values of \( \lambda \). We turn to those next.

Equation (24) captures the requirement that the profit function is constant for \( \lambda \in [\underline{\lambda}, \bar{\lambda}) \). It states that \( \pi'_\lambda = 0 \) and hence pins down \( c \) for a given \( \underline{\lambda} \). Since the solution to the initial value problem has \( \pi''_\lambda = 0 \) for all \( \lambda \in [\underline{\lambda}, \lambda) \), 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 18) and hence so is \( \pi'_\lambda \) (equation 20). On the other hand, \( F' \) and \( M' \) jump up discontinuously at \( \underline{\lambda} \) and so \( \pi''_\lambda \) can be discontinuous at that point. In fact, plugging equation (19) into equation (21) 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 \( \pi'_\lambda = 0 \) 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 (24) is continuous in \( \underline{\lambda} \) on the same interval.
Step 3: Limit as $\lambda \to 0$. We characterize the solution to the initial value problem (47) when $\lambda$ is small. We prove that for all $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\lambda \in (0, \delta)$,

$$
\frac{r + 2\gamma}{2(r + \gamma)} + \varepsilon \geq F(\lambda; \Delta) \geq \frac{r + 2\gamma}{2(r + \gamma)} - \varepsilon, \tag{49}
$$

$$
\frac{\gamma}{2(r + \gamma)} \geq M(\lambda; \Delta) \geq \frac{\gamma}{2(r + \gamma)} - \varepsilon \tag{50}
$$

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 $Y(\lambda) \equiv \log \left( \gamma(1 - F(\lambda)) - rM(\lambda) \right)$, suppressing dependence on $\Delta$ in the remainder of this step for notational convenience. Differentiating this and using the initial value problem (47) to eliminate $F'$ and $M'$, we have

$$
Y'(\lambda) \equiv \frac{-4(\gamma(4r + 4\gamma + \lambda(1 - F(\lambda))) + (r + 2\gamma)\lambda M(\lambda))}{\lambda(\gamma(8r + 8\gamma + 3\lambda(1 - F(\lambda))) + \lambda M(\lambda)(3r + 6\gamma + \lambda(1 - F(\lambda) + M(\lambda))))}. \tag{51}
$$

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

$$
F(\lambda) + \frac{\gamma}{\gamma} M(\lambda) \geq 1 - (\Delta/\lambda)^\kappa \tag{52}
$$

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

$$M(\lambda) \leq m_0 F(\lambda) = \frac{\gamma}{r + 2\gamma} F(\lambda). \quad (53)$$

Combining inequalities (52) and (53) gives us

$$F(\lambda) \geq \frac{r + 2\gamma}{2(r + \gamma)} (1 - (\Delta/\lambda)^\kappa) \quad (54)$$

This is a lower bound on the contact rate distribution, close to $\frac{r + 2\gamma}{2(r + \gamma)}$ whenever $\Delta/\lambda$ is close to zero. Using this, it is straightforward to use this to establish the lower bound in inequality (49) through an appropriate choice of $\delta$ for each $\varepsilon$.

To find a lower bound on $M$, use equation (14) to get

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

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

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 (54), we get

$$M(\lambda) \geq \frac{\gamma(r + 2\gamma)}{2(r + \gamma)(r + 2\gamma + \lambda/2)} (1 - (\Delta/\lambda)^\kappa) \quad (55)$$

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

$$M(\lambda) > \frac{\gamma(r + 2\gamma)}{2(r + \gamma)(r + 2\gamma + \lambda/2)} - \frac{1}{2}\varepsilon$$

$$= \frac{\gamma}{2(r + \gamma)} - \frac{\gamma\lambda/2}{(2(r + \gamma))(r + 2\gamma + \lambda/2)} - \frac{1}{2}\varepsilon.$$

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 $\delta'$ such that if $\lambda < \delta'$, $M(\lambda') > \frac{\gamma}{2(r + \gamma)} - \varepsilon$. Since $M$ is monotonically increasing, this implies $M(\lambda) > \frac{\gamma}{2(r + \gamma)} - \varepsilon$ as well. This establishes the lower bound in inequality (50).

We now turn to upper bounds. First, combine $\gamma(1 - F(\lambda)) - rM(\lambda) \geq 0$ with inequal-
ity (53) to get \( \gamma \geq (2(r + \gamma))M(\lambda) \). This is the upper bound in inequality (50).

Lastly, combine \( \gamma(1 - F(\lambda)) - rM(\lambda) \geq 0 \) with the lower bound in inequality (50) to get \( F(\lambda) \leq \frac{r + 2\gamma}{2(r + \gamma)} + \frac{\epsilon}{\gamma} \). The upper bound in inequality (49) then follows from an appropriate rescaling of \( \epsilon \).

In summary, in the limit as \( \lambda \) 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 (24).

**Step 4: Limit as \( \lambda \to \bar{\lambda} \).** It is easier to characterize the solution to the initial value problem (47) when \( \lambda \) 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 as long as \( F(\lambda) \in [0, 1] \) and \( M(\lambda) \in [0, 1/2] \). This implies that for all \( \epsilon > 0 \), there exists a \( \delta > 0 \) such that if \( \lambda \in (\bar{\lambda} - \delta, \bar{\lambda}), 0 \leq 2M(\lambda) \leq F(\lambda) < \epsilon \) for all \( \lambda < \bar{\lambda} \). Of course, we always have \( F(\bar{\lambda}) = 1 \) and we can construct \( M(\bar{\lambda}) \) using equation (48). This implies that \( F \) and \( M \) converge exactly to the counterparty distribution associated with the cost \( \bar{c} \), as defined earlier in the proof.

**Step 5: Intermediate Value Theorem.** We have shown that the right hand side of equation (24) is a continuous function of \( \lambda \), converging to \( \bar{c} \) when \( \lambda \to 0 \) and to \( \bar{c} \) when \( \lambda \to \bar{\lambda} \). The intermediate value theorem therefore implies that for any \( c \in (\bar{c}, \bar{c}) \), there is a \( \lambda \in (0, \bar{\lambda}) \) such that the solution to the initial value problem (47) satisfies equation (24) and hence \( F \) is associated with an equilibrium.

**Necessity of Full Support.** We now prove that in any equilibrium with \( \Lambda \in (0, \bar{\lambda}) \), the support of \( F \) is an interval \([\lambda, \bar{\lambda}]\).

First, to find a contradiction, suppose that there is a hole in the support. That is, there are contact rates \( \lambda_1, \lambda_2 \) in the support with \( 0 \leq \lambda_1 < \lambda_2 \leq \bar{\lambda} \) and for all \( \lambda \in (\lambda_1, \lambda_2) \), \( F(\lambda) = F(\lambda_1) \). The proof follows the one for Proposition 4. \( \Lambda > 0 \) implies \( \lambda_1 > 0 \). Moreover, we can rederive equation (39) with \( C''(\lambda) = 0 \) to get a necessary condition for a hole in the support,

\[
\pi''_\lambda = \frac{8\Delta e^{-f_\lambda} \phi_\lambda \cdot dM(\lambda) M(\lambda_1) - \gamma(1 - F(\lambda_1))}{r\lambda(4(r + 2\gamma) + \lambda(1 - F(\lambda_1) + 2M(\lambda_1)))^2} \geq 0.
\]

This implies \( rM(\lambda_1) \geq \gamma(1 - F(\lambda_1)) \). But this implies \( x_F(\lambda_1) \leq 0 \) and \( x_M(\lambda_1) \leq 0 \). Then the initial value problem (47) implies that there is no interval \((\lambda_0, \lambda_1)\) with \( F(\lambda) \) strictly increasing for \( \lambda \in (\lambda_0, \lambda_1) \), contradicting the hypothesis that \( \lambda_1 \) was in the support of \( F \).

With no holes in the support and \( \Lambda \in (0, \bar{\lambda}) \), 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 \( \lambda \), precluding this possibility.

**Remaining Details.** 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 \( dF(\bar{\lambda}) > 0 \), 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 \( dF(\bar{\lambda}) \). Finally, since the support of \( F \) is an interval \( [\underline{\lambda}, \bar{\lambda}] \) with \( \underline{\lambda} > 0 \) and \( dF(\bar{\lambda}) < 1 \), equation (2) implies \( \Lambda \in (0, \bar{\lambda}) \).

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

**Degenerate Autarky Equilibrium.** Suppose \( c \geq \bar{c}^* \equiv \frac{\gamma \Delta}{4r(r + 2\gamma)^2} \). 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 (45) and the argument around it implies that there is an equilibrium with \( F_n(0) = 1 \) if and only if

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

Simple algebra implies \( \bar{c}_n < \bar{c}^* \), and hence such an equilibrium exists whenever \( c \geq \bar{c}^* \). In such an equilibrium, we can also construct \( m_n \) from steady state equation (11), which gives a version of equation (13):

\[
\left( r + \gamma + \frac{\lambda}{2} (1 - F_n(\lambda) + M_n(\lambda)) \right) m_{n, \lambda} = \left( \gamma + \frac{\lambda}{2} (M_n(\lambda) - dM_n(\lambda)) \right) (1 - m_{n, \lambda}),
\]

where \( m_{n, \lambda} \) is the misalignment rate of traders with contact rate \( \lambda \) when the upper bound on contact rates is \( \bar{\lambda}_n \). Similarly we can construct \( s_n \) from equation (15). 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 \( \bar{\lambda}_n \), \( \bar{c}_n \) converges to \( \bar{c}^* \) and so a limiting equilibrium of this type does not exist when \( c < \bar{c}^* \).

**Two-Point Autarky Equilibrium.** Suppose \( \bar{c}^* > c \geq \bar{c}^* \equiv \frac{\gamma \Delta}{8r(r + \gamma)(r + 2\gamma)} \), where we can confirm algebraically that \( \bar{c}^* > \bar{c}^* \). 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 6 defines the threshold for such an equilibrium to exist, $\bar{c}_n$, as the value that makes the right hand side of equation (45) equal to zero when $F(0) = \frac{r + 2\gamma}{2(r + \gamma)}$ and $M(0) = \frac{\gamma}{2(r + \gamma)}$:

$$
\bar{c}_n = \frac{\gamma \Delta (8(r + \gamma) + \bar{\lambda}_n) M(\bar{\lambda}_n)}{8r(r + \gamma)(r + 2\gamma)(2r + 2\gamma + \bar{\lambda}_n M(\bar{\lambda}_n))} = \left( \frac{(8(r + \gamma) + \bar{\lambda}_n) M(\bar{\lambda}_n)}{2(r + 2\gamma) + \bar{\lambda}_n M(\bar{\lambda}_n)} \right) \bar{c}^* \quad (57)
$$

Note that $M(\bar{\lambda}_n) = M(0) + m_{\lambda_n}(1 - F(0)) < M(0) + \frac{1}{2} (1 - F(0)) = \frac{r + 2\gamma}{4(r + \gamma)}$, 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 $\bar{c}_n$ converges to $\bar{c}^*$ and so it follows that if $\bar{c}^* > c \geq \bar{c}^*$, 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 (45) 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 (56). This gives us an equation that implicitly defines $F_n(0)$:

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

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

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

for all $\lambda \geq 0$. Again, we can recover $m_n$ and $s_n$ from equations (56) and (15). 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, $\Lambda > 0$.

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

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

Applying this inequality to the integrand in equation (24) and solving the integral explicitly
gives
\[ c \leq \frac{4\gamma \Delta}{r(4(r + 2\gamma) + \bar{\lambda} n)^2}, \]
where \( \bar{\lambda} n \) is the lower bound for a given \( \bar{\lambda}_n \). This equation gives us an upper bound on \( \bar{\lambda}_n \) for a given value of \( c \). Equivalently, if
\[ \bar{\lambda}_n > 2\sqrt{\frac{\gamma \Delta}{rc}} - 4(r + 2\gamma), \]
\( \bar{\lambda}_n > \Delta_n \). This proves that when \( \bar{\lambda}_n \) is sufficiently large, there is an equilibrium in the intermediate range, with \( \Lambda \in (0, \bar{\lambda}_n) \). From Proposition 6, any such equilibrium is completely characterized by a positive lower bound \( \lambda n \). Moreover, the preceding argument implies \( \Delta_n < 2\sqrt{\frac{\gamma \Delta}{rc}} - 4(r + 2\gamma) \).

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 \( \Delta_n \in [0, 2\sqrt{\frac{\gamma \Delta}{rc}} - 4(r + 2\gamma)] \), with associated functions \( F_n(\lambda) \) and \( M_n(\lambda) \) that solve the initial value problem (47). By the Bolzano-Weierstrass Theorem, the sequence of lower bounds has a convergent subsequence, \( \Delta_n \rightarrow \Delta^* \). For notational convenience, we assume that the sequence \( \bar{\lambda}_n \) is chosen such that \( \Delta_n \) itself converges to \( \lambda^* \).

\( F_n(\lambda) \) and \( M_n(\lambda) \) solve the initial value problem (47) with lower bound \( \lambda n \). Since \( \Delta_n \) converges to \( \Delta^* \) and the solution to (47) is continuous in \( \lambda \) (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 (56) and (15). Since each depends continuously on the functions \( F \) and \( M \), they converge as well. Thus we have found a limiting equilibrium with \( \Lambda > 0 \).

Proof of Proposition 7.

Middlemen. Fix \( r \) and \( \gamma \). We prove that there exists a \( F^* > 0 \), independent of \( c \), \( \Delta \), and \( \bar{\lambda} \), such that in any equilibrium with \( \Lambda > 0 \), \( dF(\bar{\lambda}) \geq F^* \). From lemma 2, for any \( c < \bar{c}^* \), a limiting economy with \( \Lambda > 0 \) exists. It then follows that in any limiting economy there are middlemen, \( F(\lambda) \leq 1 - F^* \) for all \( \lambda \).

We work with the initial value problem (47). If \( \Lambda = \bar{\lambda} \), \( F(\lambda) = 0 \) for all \( \lambda < \bar{\lambda} \) (equation 2), so \( dF(\bar{\lambda}) = 1 \geq F^* \) and we are done. We thus assume \( \Lambda \in (0, \bar{\lambda}) \) and look at equilibria characterized by \( \Delta \in (0, \bar{\lambda}) \).

We make two preliminary observations. First, the initial value problem (47) indicates that \( c \) and \( \Delta \) only affect equilibrium through the value of \( \Delta \). We will therefore prove that there is a number \( F^* > 0 \), independent of \( \lambda \) and \( \bar{\lambda} \), such that in any equilibrium with \( \Lambda > 0 \),
\[ dF(\bar{\lambda}) \geq F^*, \text{ or equivalently } F(\lambda) < 1 - F^* \text{ for } \lambda < \bar{\lambda}. \] This implies that the same \( F^* \) works for all \( c \) and \( \Delta \) such that \( \Lambda \in (0, \bar{\lambda}) \). Second, the initial value problem (47) also indicates that \( \bar{\lambda} \) only affects equilibrium through the value of \( \lambda \) 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 \( \lambda \), such that in the solution to the initial value problem (47), \( \lim_{\lambda \to \infty} F(\lambda) < 1 - F^* \).

As in the proof of Proposition 6, let \( Y(\bar{\lambda}) \equiv \log \left( \gamma (1 - F(\lambda)) - rM(\lambda) \right) \). We already proved that in the solution to the initial value problem, this is finite for any finite \( \lambda \). Here we prove that it is bounded below for fixed \( \lambda \). Evaluating equation (51) at \( \lambda = \bar{\lambda} \) using \( F(\bar{\lambda}) = M(\bar{\lambda}) = 0 \) gives

\[ Y'(\bar{\lambda}) = -\frac{4(4r + 4\gamma + \bar{\lambda})}{\bar{\lambda}(8r + 8\gamma + 3\bar{\lambda})}. \]

Moreover, simple algebra takes us from equation (51) to

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

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

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

Integrating up this lower bound on the slope gives us that for \( \lambda > \bar{\lambda} \),

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

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 \( \lambda \). In particular, since \( M(\lambda) \geq 0 \), \( \lim_{\lambda \to \infty} F(\lambda) < 1 \).

As we vary \( \lambda \), \( \lim_{\lambda \to \infty} F(\lambda) \) 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 \( \lambda \). Continuity implies that this cannot happen at an interior value of \( \lambda \): 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 \( \lambda \to 0 \) or \( \lambda \to \infty \).

First recall the solution to the initial value problem in the limit as \( \lambda \to 0 \). We showed in the proof of Proposition 6 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 \( \lambda \to \infty \). In this
case, $M(\lambda) \to 0$ for all $\lambda$, and so the argument above breaks down. Instead, let $\rho \equiv \lambda/\Delta$, $H(\rho) \equiv F(\rho \Delta)$, and $L(\rho) \equiv \rho \Delta M(\rho \Delta)$. Rewrite the initial value problem as

\[
H'(\rho) = \frac{4(2r + 4\gamma + \rho \Delta(1 - H(\rho)) + 2L(\rho))(\gamma \rho \Delta(1 - H(\rho)) - rL(\rho))}{\rho^2 \Delta(\gamma(8r + 8\gamma + 3\rho \Delta(1 - H(\rho)) + L(\rho)(3r + 6\gamma + \rho \Delta(1 - H(\rho)) + L(\rho))}) + \frac{L(\rho)}{\rho},
\]

\[
L'(\rho) = \frac{4(2\gamma + L(\rho))}{\rho(\gamma(8r + 8\gamma + 3\rho \Delta(1 - H(\rho)) + L(\rho)(3r + 6\gamma + \rho \Delta(1 - H(\rho)) + L(\rho)))} + \frac{L(\rho)}{\rho},
\]

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

\[
H'(\rho) = \frac{4\gamma(1 - H(\rho))}{\rho(3\gamma + L(\rho))}, \quad (58)
\]

\[
L'(\rho) = \frac{4\gamma(2\gamma + L(\rho))}{\rho(3\gamma + L(\rho))} + \frac{L(\rho)}{\rho}, \quad (59)
\]

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

\[
\eta(L) \equiv \left(\frac{16\gamma + (7 - \sqrt{17})L}{16\gamma + (7 + \sqrt{17})L}\right)^{\frac{1}{\sqrt{17}}},
\]

\[
q(L) \equiv \eta(L)^{\frac{1}{3}} \sqrt{1 + \frac{7L}{8\gamma} + \frac{L^2}{8\gamma^2}}. \quad (61)
\]

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

**Fat 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 \{\lambdatn\} such that for each $n$, $(F_n, m_n, s_n)$ restricted to the domain $[0, \lambdatn]$ is an equilibrium when the maximum contact rate is $\lambdatn$.

Use the initial value problem (47) to get that for fixed $n$,

\[
\lim_{\lambda \to \lambdatn} \lambda^2 F'_n(\lambda) = \frac{4\lambdatn(2r + 4\gamma + \lambdatn(1 - \bar{F}_n + 2\bar{M}_n))}{\gamma(8r + 8\gamma + 3\lambdatn(1 - \bar{F}_n)) + \lambdatnM_n(3r + 6\gamma + \lambdatn(1 - \bar{F}_n + \bar{M}_n))),}.
\]
where \( \bar{F}_n \equiv \lim_{\lambda \to \bar{\lambda}_n} F_n(\lambda) \) and \( \bar{M}_n \equiv \lim_{\lambda \to \bar{\lambda}_n} M_n(\lambda) \). Then we take the limit as \( n \) gets large:

\[
\lim_{n \to \infty} \lim_{\lambda \to \bar{\lambda}_n} \lambda^2 F'_n(\lambda) = 4 \lim_{n \to \infty} \frac{1 - \bar{F}_n + 2\bar{M}_n}{M_n} (\gamma(1 - \bar{F}_n) - r\bar{M}_n).
\] (62)

Since \( \bar{\lambda}_n \) converges to \( \lambda^* \) and \( F \) and \( M \) are continuous in \( \lambda \) 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 \( \lambda^{-2} \).

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

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

Equation (1) implies \( G'_n(\lambda) = \Lambda_n F'_n(\lambda)/\lambda \), where \( \Lambda_n \) is the average contact rate when the upper bound is \( \bar{\lambda}_n \). This implies

\[
\lim_{n \to \infty} \lim_{\lambda \to \bar{\lambda}_n} \lambda^2 (1 - G_n(\lambda)) = \lim_{n \to \infty} \lim_{\lambda \to \bar{\lambda}_n} \frac{\Lambda_n}{2} \lambda^2 F'_n(\lambda)
= 2 \lim_{n \to \infty} \frac{1 - \bar{F}_n + 2\bar{M}_n}{M_n(1 - \bar{F}_n + \bar{M}_n)} \int_{\lambda_n}^{\bar{\lambda}_n} \frac{1}{\lambda} dF_n(\lambda)
\]

The second equation pulls \( \Lambda_n \) out of the inner limit, since it does not depend on \( \lambda \). It then rewrites \( \Lambda_n \) using equation (2) and replaces the inner limit using equation (62). The result follows because as \( \bar{\lambda}_n \) grows, the limit of each term exists and is a positive number.

**Proof of Corollary 1.** To show that middlemen account for a strictly positive fraction of trades, it is sufficient to show that their trading probability is strictly positive and that they account for a strictly positive fraction of meetings. Their trading probability is equal to the misalignment rate of finite traders, \( \lim_{\lambda \to \infty} M(\lambda) \), plus the fraction of time that they are misaligned and meet a misaligned middleman. Both probabilities are positive. Proposition 7 proved that they account for a strictly positive fraction of meetings.

Next, use differentiation to show that \( p_\lambda = \frac{1}{2} (m_\lambda(1 - F(\lambda)) + M(\lambda)) \) is strictly increasing for finite \( \lambda \). This uses the fact that \( m'_\lambda > 0 \) (Proposition 5) and \( m_\lambda F'(\lambda) = M'(\lambda) \). This in turn implies that \( \lambda p_\lambda \) 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 \( p_\lambda \) converges to a positive
constant as $\lambda \to \infty$, $G$ and $\hat{G}$ share a common tail parameter. □

**Proof of Proposition 8.**

**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 7 that in any limiting equilibrium, $F(\lambda) < 1 - F^*$ for all $\lambda$, where $F^*$ is strictly positive. We also have $M(\lambda) \geq 0$.

This implies
\[
\frac{4(r + 2\gamma)}{\lambda(4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda)))} \leq \frac{4(r + 2\gamma)}{\lambda(4(r + 2\gamma) + \lambda F^*)}.
\]

Integrating this implies
\[
\exp\left(-\int_{\lambda}^{\infty} \frac{4(r + 2\gamma)}{\lambda(4(r + 2\gamma) + \lambda(1 - F(\lambda) + 2M(\lambda)))} d\lambda\right) \geq \frac{\lambda F^*}{4(r + 2\gamma) + \lambda F^*}.
\]

Using equation (24), this implies that if $\Delta > 0$,
\[
\frac{c}{\Delta} \geq \frac{4\gamma F^*}{r(4(r + 2\gamma) + \Delta)(4(r + 2\gamma) + \lambda F^*)}.
\]

or
\[
\Delta \geq 2 \left( \sqrt{\frac{\gamma\Delta}{rc}} + \left(\frac{(1 - F^*)(r + 2\gamma)}{F^*}\right)^2 - \frac{(1 + F^*)(r + 2\gamma)}{F^*} \right).
\]

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

**Volume in Limiting Equilibrium.** Next, recalling that $m_\lambda = M'(\lambda)/F'(\lambda)$, we can rewrite volume in a limiting equilibrium as
\[
V = \frac{\Lambda}{4} \left(\int_{\Delta}^{\infty} ((1 - F(\lambda))M'(\lambda) + M(\lambda)F'(\lambda)) d\lambda + \lim_{\lambda \to \infty} \left((1 - F(\lambda))M(\lambda) + (1 - F(\lambda))^2 m_\infty^2\right)\right),
\]

where $m_\infty$ is the misalignment rate of middlemen, solving a version of equation (13) with $\lambda \to \infty$:
\[
\lim_{\lambda \to \infty} (M(\lambda) + m_\infty(1 - F(\lambda))) m_\infty = \lim_{\lambda \to \infty} M(\lambda)(1 - m_\infty).
\]
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 \( \Lambda \) going to infinity, we use the same change in variables as in the proof of Proposition 7 to have well-behaved objects in this limit: \( \rho \equiv \lambda/\Lambda \), \( H(\rho) \equiv F(\rho \Lambda) = F(\lambda) \), and \( L(\rho) \equiv \rho \lambda M(\rho \Lambda) \). Written in terms of these variables, volume is

\[
V = \frac{\Lambda}{4\lambda} \left( \int_1^\infty \left( 1 - H(\rho) \right) \left( \frac{L'(\rho)}{\rho} - \frac{L(\rho)}{\rho} \right) + L(\rho) H'(\rho) \right) d\rho \\
+ \lim_{\rho \to \infty} \left( \frac{(1 - H(\rho)) L(\rho)}{\rho} + \lambda(1 - H(\rho))^2 m_\infty^2 \right),
\]

where

\[
\lim_{\rho \to \infty} \left( \frac{L(\rho)}{\rho \Lambda} + m_\infty (1 - H(\rho)) \right) m_\infty = \lim_{\rho \to \infty} \frac{L(\rho)}{\rho \Lambda} (1 - m_\infty).
\]

The terms in (63) have the same interpretation as in equation (25) discussed above.

**Mean Contact Rate Relative to Lower Bound.** We next calculate the mean contact rate relative to the lower bound, \( \Lambda/\lambda \). Using equation (2) for the limiting equilibrium, this is

\[
\frac{\Lambda}{\lambda} = \frac{1}{\lambda} \int_1^\infty \frac{1}{\lambda} dF(\lambda) = \frac{1}{\lambda} \int_1^\infty \frac{1}{\rho} dH(\rho) = \frac{1}{\lambda} \int_0^\infty \frac{-\eta'(L)}{q(L)} dL.
\]

The last equality rewrites the previous one using the inverse functions \( q(L(\rho)) = \rho \) and \( \eta(L) \equiv 1 - H(q(L)) \). This hold for arbitrary \( \Lambda \). In the limit as \( c \to 0 \), we have \( \Lambda \to \infty \). We can therefore apply equations (60) and (61) 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:

\[
\lim_{c \to 0} \frac{\Lambda}{\lambda} = \frac{1}{\sqrt{2}} \left( \frac{7 + \sqrt{17}}{7 - \sqrt{17}} \right)^{\frac{5}{2\sqrt{17}}} \approx 2.23.
\]

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

\[
\Gamma'(L) = q'(L)G'(q(L)\Lambda) = \frac{\Lambda q'(L) F'(q(L)\Lambda)}{q(L)} = \frac{\Lambda q'(L) H'(q(L))}{2q(L)} = -\frac{\Lambda \eta'(L)}{2q(L)}.
\]
The first equation differentiates $\Gamma(L) = G(q(L)\lambda)$, the second uses equation (1) to rewrite this in terms of $F$, the third differentiates $F(\rho_L) = H(\rho)$, and the fourth differentiates the definition $\eta(L) \equiv 1 - H(q(L))$.

In the limit as $c \to 0$, we again apply equations (60) and (61) and integrate $\Gamma$ explicitly using the initial condition $\Gamma(0) = 0$. We simplify this further using the limiting behavior of $\Lambda/\lambda$ from equation (65):

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

(67)

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

**Volume of Purchases by Non-Middlemen.** Next consider the first term in the trading rate (63), the rate at which a non-middleman buys the asset from another trader, either a middleman or a non-middleman.

$$\mathcal{V}_1 \equiv \frac{\Lambda}{4\lambda} \int_1^\infty \frac{(1 - H(\rho))(L'(\rho) - L(\rho)/\rho) + L(\rho)H'(\rho)}{\rho} d\rho = \frac{\Lambda}{2\lambda} \int_1^\infty \frac{(\gamma + L(\rho))H'(\rho)}{\rho} d\rho$$

$$= \frac{1}{2} \int_1^\infty (\gamma + L(\rho))\Gamma'(L(\rho)) L'(\rho) d\rho = \frac{1}{2} \left( \gamma + \int_0^\infty L\Gamma'(L) dL \right).$$

The first equation eliminates $L'(\rho)$ using equation (59) and then simplifies with equation (58). The second eliminates $H'(\rho) = H'(q(L(\rho)))$ using equation (66) and then uses $q'(L(\rho)) = 1/L'(\rho)$, since the functions are inverses. The last equation is a change of the variable of integration from $\rho$ to $L(\rho)$. Once again, we now take the limit as $c \to 0$ and so apply the functional form in equation (66) to get

$$\lim_{c \to 0} \mathcal{V}_1 = \left( \sqrt{2} \left( \frac{7 + \sqrt{17}}{7 - \sqrt{17}} \right)^{\frac{7}{2\sqrt{17}}} - 3 \right) \gamma \approx 1.46\gamma.$$  

(68)

**Volume of Purchases by Middlemen from Non-Middlemen.** Next consider the second term in the trading rate (63), the rate at which a middleman buys the asset from a non-middleman:

$$\mathcal{V}_2 \equiv \frac{\Lambda}{4\lambda} \lim_{\rho \to \infty} \frac{(1 - H(\rho))L(\rho)}{\rho} = \frac{\Lambda}{4\lambda} \lim_{\rho \to \infty} (1 - H(\rho))L'(\rho) = \frac{\Lambda}{4\lambda} \lim_{L \to \infty} \frac{\eta(L)}{q'(L)}$$

(69)

The first equation uses L'Hôpital's rule, since $H(\rho)$ converges to a number less than 1, while $L(\rho)$ and $\rho$ 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 (60) and (61) as well as equation (65). This gives

\[
\lim_{c \to 0} V_2 = \frac{\gamma}{2}.
\]

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 \(\frac{1}{2}\gamma\). Subtracting this from equation (68), we get that finite traders buy from finite traders at rate

\[
\left(\sqrt{2} \left(\frac{7 + \sqrt{17}}{7 - \sqrt{17}}\right)^{\frac{7}{2 \sqrt{17}}} - \frac{7}{2}\right) \gamma \approx 0.96\gamma,
\]

a number between \(\frac{1}{2}\gamma\) and \(\gamma\).

**Volume of Purchases by Middlemen from Middlemen.** Finally consider the third term in equation (63), the rate that middlemen buy from middlemen:

\[
V_3 = \frac{\Lambda}{4} \lim_{\rho \to \infty} (1 - H(\rho))^2 m_\infty^2
\]

where \(m_\infty\) solves equation (64). Multiply both sides of equation (64) by \(\frac{\Lambda}{4} \left(1 - \lim_{\rho \to \infty} H(\rho)\right)\) to get

\[
V_3 = \frac{\Lambda}{4} \lim_{\rho \to \infty} (1 - H(\rho))^2 m_\infty^2 = \frac{\Lambda(1 - 2m_\infty)}{4} \lim_{\rho \to \infty} \frac{(1 - H(\rho))L(\rho)}{\rho\lambda} = (1 - 2m_\infty)V_2,
\]

where \(V_2\) is defined in equation (69). When \(c \to 0, \lambda \to \infty, m_\infty \to 0,\) and the results in the previous paragraph imply \(\lim_{c \to 0} V_3 = \lim_{c \to 0} V_2 = \frac{\gamma}{2}\). This completes the proof.

**Proof of Proposition 9.** We break the proof into steps. For the most part, these correspond to the proofs of the related equilibrium propositions, and so we label them accordingly.

**The Lagrangian** The planner chooses symmetric, time-invariant trading probabilities

\(1_{\lambda,a'}^{\lambda,a} = 1_{\lambda,a'}^{\lambda,a'} \in [0, 1]\) and the time-invariant counterparty distribution \(F(\lambda)\), a nondecreasing, right-continuous function \(F : [0, \lambda] \to [0, 1]\) subject to the constraints that \(\int_0^\lambda dF(\lambda) = 1\) and that the misalignment rate \(m_\lambda : [0, \lambda] \to [0, 1]\) satisfies symmetric flow-balance constraint (7).
The planner’s Lagrangian is

\[ L = \delta_1 + \frac{1}{\int_0^\lambda \frac{1}{\lambda} dF(\lambda)} \left( -\int_0^\lambda \frac{\Delta m_\lambda + r C(\lambda)}{\lambda} dF(\lambda) + \theta \left( 1 - \int_0^\lambda dF(\lambda) \right) \right) \]

\[ + \int_0^\lambda S(\lambda) \left( \left( \frac{r + \gamma}{\lambda} + \frac{1}{2} \int_0^\lambda \left( \mathbf{1}_{\lambda,0}^{\lambda',0} m_{\lambda'} + \mathbf{1}_{\lambda,0}^{\lambda',1} (1 - m_{\lambda'}) \right) dF(\lambda') \right) m_\lambda \]

\[ - \left( \frac{\gamma}{\lambda} + \frac{1}{2} \int_0^\lambda \left( \mathbf{1}_{\lambda,1}^{\lambda',0} m_{\lambda'} + \mathbf{1}_{\lambda,1}^{\lambda',1} (1 - m_{\lambda'}) \right) dF(\lambda') (1 - m_\lambda) \right) dF(\lambda) \].

The Lagrange multiplier \( \theta \) represents the shadow cost of the constraint that the counterparty distribution \( F \) must be a proper cumulative distribution function. The Lagrange multiplier \( S(\lambda) \) represents the shadow value of switching a trader from misaligned to well-aligned.

Since these are continuous objects, the planner has a continuum of control variables. To use Lagrangian techniques, we follow Üslü (2018) in appealing to van Imhoff (1982) and Hutson, Pym and Cloud (2005) so as to interpret the integrals in the Lagrangian as summations over discrete intervals with length \( d\lambda \) converging to 0.

Note that the value of the Lagrangian is \( \delta_1 - \Delta m_0 - r C(0) \) when \( \Lambda = \frac{1}{\int_0^\lambda \frac{1}{\lambda} dF(\lambda)} \) is zero. In this case, the choice of trading probabilities, counterparty distribution, and misalignment rates is irrelevant to the planner because they do not affect her objective. As a consequence, we are free to impose any additional restrictions on those objects in the case where \( \Lambda = 0 \). In what follows, we write the necessary first order conditions when \( \Lambda \) is positive and then, in order to mimic the equilibrium case as closely as possible, we also impose them when \( \Lambda = 0 \).

**Planner’s Proposition 1: Trading Patterns**

Take the first order conditions with respect to the trading probabilities \( \mathbf{1}_{\lambda,a}^{\lambda',a'} \). Using the constraint \( 0 \leq \mathbf{1}_{\lambda,a}^{\lambda',a'} = \mathbf{1}_{\lambda',a'}^{\lambda,a} \leq 1 \), we get that when \( dF(\lambda) \) and \( dF(\lambda') \) are both positive,

\[ S(\lambda) + S(\lambda') \geq 0 \Rightarrow \mathbf{1}_{\lambda,0}^{\lambda',0} = \begin{cases} 1 \\
0 \end{cases} , \quad S(\lambda) \geq S(\lambda') \Rightarrow \mathbf{1}_{\lambda,0}^{\lambda',1} = \begin{cases} 1 \\
0 \end{cases} \]

\[ S(\lambda) + S(\lambda') \leq 0 \Rightarrow \mathbf{1}_{\lambda,1}^{\lambda',1} = \begin{cases} 1 \\
0 \end{cases} , \quad S(\lambda') \geq S(\lambda) \Rightarrow \mathbf{1}_{\lambda,1}^{\lambda',0} = \begin{cases} 1 \\
0 \end{cases} . \]
Next, take the first order condition with respect to $m_\lambda$. Canceling redundant terms, we get that when $dF(\lambda)$ is positive,

$$\Delta = (r + 2\gamma)S(\lambda) + \lambda \left( \int_0^{\lambda} \left( 1_{\lambda,0}^{\gamma,0}(S(\lambda) + S(\lambda')) m_{\lambda'} + 1_{\lambda,1}^{\gamma,1}(S(\lambda) - S(\lambda'))(1 - m_{\lambda'}) \right) dF(\lambda') \right).$$

(72)

Using (71), rewrite the first order condition (72) as

$$\Delta = (r + 2\gamma)S(\lambda) + \lambda \left( \int_0^{\lambda} \left( (S(\lambda) + S(\lambda'))^+ - (S(\lambda') - S(\lambda))^+ \right) m_{\lambda'} + \left( (S(\lambda) - S(\lambda'))^+ - (-S(\lambda) - S(\lambda'))^+ \right)(1 - m_{\lambda'}) dF(\lambda'). \right)$$

(73)

Comparing this to equation (10), the only difference is that the factor multiplying the integral is twice as large for the planner. Thus, the proof of Proposition 1 immediately implies that the surplus function is uniquely defined by this equation and moreover is decreasing and nonnegative. The optimal trading pattern (the planner’s version of Proposition 1) follows.

**Planner’s Lemma 1: Surplus Function** Use monotonicity of $S$ to rewrite equation (73) as

$$2\Delta = 2 \left( r + 2\gamma + \lambda \int_0^{\lambda} m_{\lambda'} dF(\lambda') \right) S(\lambda) + \lambda \int_0^{\lambda} (S(\lambda) - S(\lambda'))(1 - 2m_{\lambda'}) dF(\lambda').$$

(74)

This is similar to equation (35), except for the numerical value of the coefficients. We can then replicate the remainder of the proof of Lemma 1 to obtain equations (27) and (28).

**Characterization of Optimal $F$** Now return to the Lagrangian (70). The first order condition with respect to $dF(\lambda)$ implies that $dF(\lambda) \geq 0$ and

$$\theta \geq -\frac{\Delta m_\lambda + rC(\lambda)}{\lambda} + \frac{\Lambda}{\lambda} \int_0^{\lambda} \frac{\Delta m_{\lambda'} + rC(\lambda')}{\lambda'} dF(\lambda') + \frac{1}{2} \int_0^{\lambda} S(\lambda') \left( \left( 1_{\lambda',0}^{\lambda,0} m_\lambda + 1_{\lambda',1}^{\lambda,1}(1 - m_\lambda) \right) m_{\lambda'} - \left( 1_{\lambda',1}^{\lambda,1} m_\lambda + 1_{\lambda',1}^{\lambda,1}(1 - m_\lambda) \right) (1 - m_{\lambda'}) \right) dF(\lambda').$$

(75)
with complementary slackness. Next, rewrite equation (72) as

$$ \frac{\Delta m_\lambda}{\lambda} = \frac{(r + 2\gamma)S(\lambda)m_\lambda}{\lambda} + \frac{1}{2} \left( S(\lambda)m_\lambda \int_{0}^{\bar{\lambda}} \left( \left( 1_{\lambda,0}^{X,0} + 1_{\lambda,1}^{X,0} \right) m_\lambda + \left( 1_{\lambda,0}^{X,1} + 1_{\lambda,1}^{X,1} \right) \left( 1 - m_\lambda \right) \right) dF(\lambda') \right) $$

$$ - \int_{0}^{\bar{\lambda}} S(\lambda') \left( 1_{\lambda,0}^{X,0} m_\lambda - 1_{\lambda,1}^{X,1} \left( 1 - m_\lambda \right) \right) dF(\lambda') $$

$$ + \int_{0}^{\bar{\lambda}} S(\lambda') \left( \left( 1_{\lambda,0}^{X,0} m_\lambda + 1_{\lambda,1}^{X,0} \left( 1 - m_\lambda \right) \right) m_\lambda - \left( 1_{\lambda,0}^{X,1} m_\lambda + 1_{\lambda,1}^{X,1} \left( 1 - m_\lambda \right) \right) \left( 1 - m_\lambda \right) \right) dF(\lambda') $$

Use this to rewrite condition (75) as $dF(\lambda) \geq 0$ and

$$ \theta \geq - \frac{m_\lambda \left( r + 2\gamma + \frac{\lambda}{2} \int_{0}^{\bar{\lambda}} \left( \left( 1_{\lambda,0}^{X,0} + 1_{\lambda,1}^{X,0} \right) m_\lambda + \left( 1_{\lambda,0}^{X,1} + 1_{\lambda,1}^{X,1} \right) \left( 1 - m_\lambda \right) \right) dF(\lambda') \right) S(\lambda)}{\lambda} $$

$$ - \frac{rC(\lambda)}{\lambda} + \frac{\Lambda}{\lambda} \int_{0}^{\bar{\lambda}} \Delta m_\lambda + rC(\lambda') \frac{dF(\lambda)}{\lambda'} $$

$$ + \frac{1}{2} \int_{0}^{\bar{\lambda}} S(\lambda') \left( 1_{\lambda,0}^{X,0} m_\lambda - 1_{\lambda,1}^{X,1} \left( 1 - m_\lambda \right) \right) dF(\lambda') $$

with complementary slackness. Simplify the first term using the steady state equation (7) to get $dF(\lambda) \geq 0$ and

$$ \theta \geq - \left( \gamma + \frac{\lambda}{2} \int_{0}^{\bar{\lambda}} \left( 1_{\lambda,1}^{X,0} m_\lambda + 1_{\lambda,1}^{X,1} \left( 1 - m_\lambda \right) \right) dF(\lambda') \right) S(\lambda) $$

$$ - \frac{rC(\lambda)}{\lambda} + \frac{\Lambda}{\lambda} \int_{0}^{\bar{\lambda}} \Delta m_\lambda + rC(\lambda') \frac{dF(\lambda)}{\lambda'} $$

$$ + \frac{1}{2} \int_{0}^{\bar{\lambda}} S(\lambda') \left( 1_{\lambda,1}^{X,0} m_\lambda - 1_{\lambda,1}^{X,1} \left( 1 - m_\lambda \right) \right) dF(\lambda') $$

with complementary slackness. The optimal trading pattern (71) and monotonicity of the surplus function gives us $dF(\lambda) \geq 0$ and

$$ \theta \geq - \frac{\gamma S(\lambda)}{\lambda} + \frac{1}{2} \int_{0}^{\bar{\lambda}} \left( S(\lambda') - S(\lambda) \right) dM(\lambda') - \frac{rC(\lambda)}{\lambda} + \frac{\Lambda}{\lambda} \int_{0}^{\bar{\lambda}} \Delta m_\lambda + rC(\lambda') \frac{dF(\lambda)}{\lambda'} $$

(76)
with complementary slackness. Multiply both sides of inequality (76) by \( \lambda \) and rewrite this as 

\[
dF(\lambda') \geq -\Lambda \int_0^\lambda \frac{\Delta m \lambda' + rC(\lambda')}{\lambda'} dF(\lambda') \geq \Pi_\lambda
\]

with complementary slackness, where

\[
\Pi_\lambda \equiv -\gamma S(\lambda) + \frac{\lambda^2}{2} \int_0^\lambda (S(\lambda') - S(\lambda)) dM(\lambda') - rC(\lambda) - \theta \lambda.
\]

This is a key representation of the solution to the planner’s problem and shares a similar mathematical structure to equation (17). In particular, it implies that if \( \lambda \) is in the support of \( F \), \( \lambda = \arg \max_{\lambda'} \Pi_{\lambda'} \equiv \bar{\Pi} \).

**Planner’s Proposition 2: Reverse Engineering the Cost Function**  
As in equilibrium, the misalignment rate \( m \) and the cumulative misalignment function \( M \) are uniquely determined from \( F \) using steady state conditions. We can then recover the surplus function from equation (27). We next find \( \theta \) by multiplying both sides of equation (75) by \( dF(\lambda) \), integrating, and canceling the first two terms on the right hand side:

\[
\theta = \frac{1}{2} \int_0^\lambda S(\lambda') \left( \int_0^\lambda \left( 1_{\lambda',0}^\lambda m_{\lambda'} + 1_{\lambda',1}^\lambda (1 - m_{\lambda'}) \right) m_{\lambda'} \right) dF(\lambda) dF(\lambda')
\]

\[
= \frac{1}{2} \int_0^\lambda S(\lambda) \left( \int_0^\lambda \left( 1_{\lambda',0}^\lambda m_{\lambda'} + 1_{\lambda',1}^\lambda (1 - m_{\lambda'}) \right) dF(\lambda') m_{\lambda} \right.
\]

\[
- \int_0^\lambda \left( 1_{\lambda',1}^\lambda m_{\lambda'} + 1_{\lambda',0}^\lambda (1 - m_{\lambda'}) \right) dF(\lambda')(1 - m_{\lambda}) dF(\lambda)
\]

\[
= \int_0^\lambda \frac{\gamma - (r + 2\gamma)m_{\lambda}}{\lambda} S(\lambda) dF(\lambda).
\]

The second equation swaps the role of \( \lambda \) and \( \lambda' \) in the integrals and then regroups terms. The last equation follows from the inflow-outflow equation (7). Thus we recover \( \theta \) from \( F \), \( m \), and \( S \). Finally substituting these objects into equation (78) gives us the cost function up to an additive constant on the support of \( F \).

**Planner’s Proposition 3: No Mass Points**  
The replication of the proof of Proposition 3 follows immediately using the functional form of \( S \) in equation (27) and \( \Pi \) in equation (78).
**Representation of Optimum as ODE System**  For the remainder of the proof, we assume $C$ is twice continuously differentiable. As in equilibrium, this implies $\Pi'' = 0$ almost everywhere on the interior of the support. Using the definition of $\Pi$ in equation (78) gives

$$-(2\gamma + \lambda M(\lambda))S''(\lambda) - (2M(\lambda) + \lambda M'(\lambda))S'(\lambda) - C''(\lambda) = 0.$$  

(80)

This is analogous to equation (21). Next, use the known functional form of $S(\lambda)$ and $\Phi(\lambda)$ to express $S'$ and $S''$ as a function of $F$, $M$, $F'$ and $M'$ almost everywhere on the support:

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

(81)

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

(82)

We then substitute these back into equation (80) to get one linear relationship between $F'(\lambda)$ and $M'(\lambda)$ in the optimal solution. As in the equilibrium case, we do not write it out because it is very messy. Still, we can solve this and equations (14) and (81) explicitly for $F'(\lambda)$, $M'(\lambda)$, and $S'(\lambda)$ as functions of $F(\lambda)$, $M(\lambda)$, and $S(\lambda)$. This is an ordinary differential equation system that is valid on the interior of the support of $F$.

As in the equilibrium case, we now have an initial value problem for $F$, $M$, and $S$. If we know $\bar{\lambda}$, we then use the terminal conditions $F(\bar{\lambda}) = M(\bar{\lambda}) = 0$ as well as

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

(83)

which we obtain by evaluating equation (74) at $\lambda = \bar{\lambda}$.

Finally, to verify that the choice of $\bar{\lambda}$ is a solution, we use the first order condition $\Pi'_{\lambda} = 0$, where $\Pi_{\lambda}$ is defined in equation (78):

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

(84)

and $\theta$ is defined in equation (79). This is the equivalent of equation (22), or (24) when cost function is linear, for equilibrium.

**The Lagrange Multiplier $\theta$ Redux**  To further characterize the optimum, we find it useful to relate the Lagrange multiplier $\theta$ to the cost function. First, multiply equation (76)
by $dF(\lambda)$ and integrate up to but not including $\bar{\lambda}$:

$$(1 - dF(\bar{\lambda}))\theta = \lim_{\lambda \to \bar{\lambda}} \left( - \int_{0}^{\hat{\lambda}} \frac{\gamma S(\lambda)}{\lambda} dF(\lambda) + \frac{1}{2} \int_{0}^{\hat{\lambda}} \int_{0}^{\lambda} (S(\lambda') - S(\lambda)) dM(\lambda') dF(\lambda) \\
- \int_{0}^{\hat{\lambda}} \frac{rC(\lambda)}{\lambda} dF(\lambda) + \Lambda \int_{0}^{\hat{\lambda}} \frac{\Delta m_{\lambda'} + rC(\lambda')}{\lambda'} dF(\lambda') \int_{0}^{\hat{\lambda}} \frac{1}{\lambda} dF(\lambda) \right).$$

Note from equation (2) that

$$\lim_{\lambda \to \bar{\lambda}} \int_{0}^{\hat{\lambda}} \frac{1}{\lambda} dF(\lambda) = \frac{1}{\Lambda} - \frac{1}{\bar{\lambda}} dF(\bar{\lambda}).$$

Therefore we have

$$(1 - dF(\bar{\lambda}))\theta = \lim_{\lambda \to \bar{\lambda}} \left( \int_{0}^{\hat{\lambda}} \frac{\Delta m_{\lambda} - \gamma S(\lambda)}{\lambda} dF(\lambda) + \frac{1}{2} \int_{0}^{\hat{\lambda}} \int_{0}^{\lambda} (S(\lambda') - S(\lambda)) dM(\lambda') dF(\lambda) \\
+ \left( \Delta m_{\lambda} + rC(\bar{\lambda}) - \Lambda \int_{0}^{\hat{\lambda}} \frac{\Delta m_{\lambda'} + rC(\lambda')}{\lambda'} dF(\lambda') \right) \frac{dF(\bar{\lambda})}{\bar{\lambda}}. \quad (85)$$

This is the first key equation.

Second, on the interior of the support of $F$, we have that $\Pi_{\lambda}$ is constant and hence $\Pi'_{\lambda} = 0$. Differentiating equation (78) gives

$$\theta = - \left( \gamma + \frac{\lambda M(\lambda)}{2} \right) S'(\lambda) + \frac{1}{2} \int_{0}^{\lambda} (S(\lambda') - S(\lambda)) dM(\lambda') - rC'(\lambda)$$

Equations (81) and (14) imply that on the interior of the support,

$$\left( \gamma + \frac{\lambda}{2} M(\lambda) \right) S'(\lambda) = \frac{((r + 2\gamma) S(\lambda) - \Delta) m_{\lambda}}{\lambda},$$

and so the previous equation becomes

$$\theta = - \frac{((r + 2\gamma) S(\lambda) - \Delta) m_{\lambda}}{\lambda} + \frac{1}{2} \int_{0}^{\lambda} (S(\lambda') - S(\lambda)) dM(\lambda') - rC'(\lambda).$$
Again multiply this equation by \( dF(\lambda) \) and integrate up to but not including \( \tilde{\lambda} \):

\[
(1 - F(\tilde{\lambda}))\theta = \lim_{\tilde{\lambda} \to \lambda} \left( - \int_0^{\tilde{\lambda}} \frac{(r + 2\gamma)S(\lambda) - \Delta}{\lambda} m_\lambda dF(\lambda) \right.

\[
+ \frac{1}{2} \int_0^{\tilde{\lambda}} \int_0^{\lambda} (S(\lambda') - S(\lambda))dM(\lambda')dF(\lambda) - \int_0^{\tilde{\lambda}} rC'(\lambda)dF(\lambda) \bigg) \].

(86)

This is the second key equation.

Now equate the right hand sides of (85) and (86) to get

\[
\int_0^{\tilde{\lambda}} \frac{\gamma - (r + 2\gamma)m_\lambda}{\lambda} S(\lambda) dF(\lambda) = r \int_0^{\tilde{\lambda}} C'(\lambda)dF(\lambda)

\[
+ \left( \Delta m_\lambda + r(C(\tilde{\lambda}) - \tilde{\lambda}C'(\tilde{\lambda})) \right) - \Lambda \int_0^{\tilde{\lambda}} \frac{\Delta m_{\lambda'} + rC'(\lambda')}{\lambda'} dF(\lambda') + (\gamma - (r + 2\gamma)m_\lambda)S(\tilde{\lambda}) \frac{dF(\tilde{\lambda})}{\lambda}.

(87)

Equivalently, using equation (79),

\[
\theta = r \int_0^{\tilde{\lambda}} C'(\lambda)dF(\lambda) + \left( \Delta m_\lambda + r(C(\tilde{\lambda}) - \tilde{\lambda}C'(\tilde{\lambda})) \right)

\[
- \Lambda \int_0^{\tilde{\lambda}} \frac{\Delta m_{\lambda'} + rC'(\lambda')}{\lambda'} dF(\lambda') + \left( \gamma - (r + 2\gamma)m_\lambda \right)S(\tilde{\lambda}) \frac{dF(\tilde{\lambda})}{\lambda}. \]

(88)

We can simplify this a bit further, regrouping terms using equation (83) to eliminate \( S(\tilde{\lambda}) \):

\[
\theta = r \int_0^{\tilde{\lambda}} C'(\lambda)dF(\lambda) + \left( \Pi + \Delta \frac{\gamma + \tilde{\lambda}M(\tilde{\lambda})m_{\tilde{\lambda}}}{r + 2\gamma + \lambda M(\tilde{\lambda})} + r(C(\tilde{\lambda}) - \tilde{\lambda}C'(\tilde{\lambda})) \right) \frac{dF(\tilde{\lambda})}{\lambda}, \]

(89)

where \( \Pi = -\Lambda \int_0^{\tilde{\lambda}} \frac{\Delta m_{\lambda'} + rC'(\lambda')}{\lambda'} dF(\lambda') \), the maximum value of \( \Pi_; \) see condition (77). This is an explicit equation for \( \theta \). Notably when either \( \tilde{\lambda} \rightarrow 0 \) or \( dF(\tilde{\lambda}) = 0 \), it implies \( \theta = r \int_0^{\tilde{\lambda}} C'(\lambda)dF(\lambda) \), as we noted in the text.

**Planner’s Propositions 4–5: Convex Support and Increasing Misalignment Rate**

Given the preceding argument it is now straightforward to replicate Propositions 4 and 5 for the social planner. We use equations (14), (80), and (81) to re-derive versions of equations (42) and (43) for an optimum. Only the numerical coefficients change.

**Planner’s Proposition 6: Existence** We first redefine the cost thresholds \( \bar{c} \) and \( \bar{c} \) for the planner’s problem. When \( F(\lambda) \) is constant for all \( \lambda \in [0, \bar{\lambda}) \), we can find the surplus function explicitly using equations (27) and (28) and substitute this into equation (78) to
get that if $F(0) < (>) \frac{r+2\gamma}{2r+2\gamma}$, the profit function is strictly concave (convex) on the interval $[0, \bar{\lambda}]$. The first is inconsistent with the planner allocation with a linear cost function, while the latter is consistent with a two point distribution for $F$, 0 and $\bar{\lambda}$. In this case, we have

$$\Pi_\lambda - \Pi_0 = \left( \frac{(2\gamma + \bar{\lambda}M(0)) M(\bar{\lambda}) \Delta}{2(r+2\gamma)(r+2\gamma + \lambda M(\lambda))} - (rc + \theta) \right) \bar{\lambda},$$

analogous to equation (45). We can further simplify by eliminating $\theta$ using equation (88) and thus writing the indifference condition as

$$2rc = \frac{(2\gamma + \bar{\lambda}M(0)) M(\bar{\lambda}) \Delta}{2(r+2\gamma)(r+2\gamma + \lambda M(\lambda))} - \left( m_0 + \frac{\gamma + \bar{\lambda}M(\bar{\lambda})m_\lambda}{r+2\gamma + \lambda M(\lambda)} \right) \frac{\Delta dF(\bar{\lambda})}{\lambda}. \quad (89)$$

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 $M(\bar{\lambda}) = M(0) + m_\lambda dF(\bar{\lambda})$, with $m_\lambda$ defined in equation (13), 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 $\bar{c}$, above which setting $F(\lambda) = 1$ for all $\lambda$ is an optimal allocation.

As in equilibrium, we can also find a threshold $\underline{c}$, below which there is an optimal allocation with $\Lambda = \bar{\lambda}$.

At intermediate values of $c \in (\underline{c}, \bar{c})$, we look for an optimal allocation described by the solution to an initial value problem. We use equations (14), (80), and (81) to derive versions of equations (42) and (43) for an optimum, and evaluate with $C''(\lambda) = 0$, $\forall \lambda$. This gives

$$F'(\lambda) = X_F(\lambda), \quad M'(\lambda) = X_M(\lambda), \quad \text{and} \quad F(\lambda) = M(\lambda) = 0, \quad (90)$$

where

$$X_F(\lambda) \equiv \frac{2(2r+4\gamma + \lambda(1 - F(\lambda) + 2M(\lambda))) \left( \gamma(1 - F(\lambda)) - rM(\lambda) \right)}{\lambda(2r+2\gamma + \lambda(1 - F(\lambda) + M(\lambda)))(2\gamma + \lambda M(\lambda))}, \quad X_M(\lambda) \equiv \frac{2(\gamma(1 - F(\lambda)) - rM(\lambda))}{\lambda(2r+2\gamma + \lambda(1 - F(\lambda) + M(\lambda)))},$$

analogous to equation (47). Moreover, the lower bound must satisfy the first order condition $\Pi_\lambda' = 0$ or

$$2rc = -\gamma S'(\bar{\lambda}) - \left( \Pi + \Delta \frac{\gamma + \bar{\lambda}M(\bar{\lambda})m_\lambda}{r+2\gamma + \lambda M(\lambda)} \right) \frac{dF(\bar{\lambda})}{\lambda}. \quad (91)$$

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.
Planner’s Lemma 2: Limiting Optimum  Here we introduce an analogous notion to limiting equilibrium, as defined in definition 2:

Definition 4  Assume $C(\lambda) = c\lambda$. Fix $r$, $\gamma$, $\Delta$, and $c$. The functions $(F, m, S)$, each with domain $[0, \infty)$, are a limiting optimum if there is a sequence of functions $\{F_n, m_n, S_n\}$, each with domain $[0, \infty)$, which converge pointwise to $(F, m, S)$, and there is an increasing, unbounded sequence $\{\bar{\lambda}_n\}$ such that for each $n$, $(F_n, m_n, S_n)$ restricted to the domain $[0, \bar{\lambda}_n]$ are an optimal allocation when the maximum contact rate is $\bar{\lambda}_n$.

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 (89), the threshold $\bar{c}_n$ satisfies

$$\bar{c}_n = \frac{(2 + \frac{\bar{\lambda}}{2r + 2\gamma}) M(\bar{\lambda}) \gamma \Delta}{4r(r + 2\gamma)(r + 2\gamma + \lambda M(\bar{\lambda}))} - \left(\frac{m_0 + \gamma + \bar{\lambda}M(\bar{\lambda})m_\lambda}{r + 2\gamma + \lambda M(\bar{\lambda})}\right) \frac{\Delta}{4(r + \gamma)\bar{\lambda}}.$$

The second term is positive and converges to zero when $\bar{\lambda} \to \infty$. We can also confirm algebraically that since $m_\lambda \leq 1/2$, $M(\bar{\lambda}) \leq \frac{r + 2\gamma}{4(r + \gamma)}$ and so

$$\frac{(2 + \frac{\bar{\lambda}}{2r + 2\gamma}) M(\bar{\lambda}) \gamma \Delta}{4r(r + 2\gamma)(r + 2\gamma + \lambda M(\bar{\lambda}))} \leq \frac{\gamma \Delta}{8r(r + \gamma)(r + 2\gamma)} = \bar{c}^*, \tag{90}$$

again with equality in the limit as $\bar{\lambda} \to \infty$. 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 from the equilibrium.

Planner’s Proposition 7: Middlemen and Tail Behavior of $G$  We first prove the existence of middlemen. Similar to the proof for equilibrium, we prove that for fixed $r$ and $\gamma$, there exists a $F^* > 0$, independent of $c$, $\Delta$, and $\bar{\lambda}$, such that in any allocation that satisfies first order conditions to the planner’s problem, with $\Lambda > 0$, $dF(\bar{\lambda}) \geq F^*$. It follows that there are middlemen in any limiting optimum.

The proof is analogous to that of equilibrium. We focus here on the proof that $\lim_{\lambda \to \infty} F(\lambda)$ is bounded below 1 even when $\bar{\lambda} \to \infty$. We work with the initial value problem (90) and apply the same transformation of variables as in equilibrium: let $\rho \equiv \lambda/\bar{\lambda}$, $H(\rho) \equiv F(\rho \bar{\lambda})$ and $L(\rho) \equiv \rho \bar{\lambda} M(\rho \bar{\lambda})$, with $H(1) = L(1) = 0$. In the limit as $\bar{\lambda} \to \infty$, the initial value problem becomes

$$H'(\rho) = \frac{2 \gamma(1 - H(\rho))}{\rho(2 \gamma + L(\rho))} \quad \text{and} \quad L'(\rho) = \frac{2 \gamma + L(\rho)}{\rho}, \tag{91}$$

74
with \( L(1) = H(1) = 0 \). These differential equations can be solved in closed form,

\[
L(\rho) = 2\gamma(\rho - 1) \quad \text{and} \quad H(\rho) = 1 - e^{\rho^{-1} - 1}.
\]  (92)

Thus the unique limit of \( H(\rho) \) is \( 1 - e^{-1} \approx 0.632 < 1 \) which, similar to equilibrium, establishes the bound in the limit.

Finally, the argument for the fat tail is exactly analogous to equilibrium. We find that

\[
\lim_{n \to \infty} \lim_{\lambda \to \bar{\lambda}_n} \lambda^2 F_n' = 2 \lim_{n \to \infty} \frac{(1 - \bar{F}_n + 2\bar{M}_n)(\gamma(1 - \bar{F}_n) - r\bar{M}_n)}{\bar{M}_n(1 - \bar{F}_n + \bar{M}_n)},
\]

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 (62). We can again prove that the right hand side has a well-behaved limit and so the density \( F' \) is asymptotically proportional to \( \lambda^{-2} \).

**Planner’s Proposition 8: Frictionless Limit**  In a limiting optimum, condition (91) reduces to \( 2rc = -\gamma S'(\Lambda) \). We can use this to obtain a lower bound on \( \Lambda \) for fixed \( c \), and prove that the lower bound converges to infinity as \( c \) goes to zero, as in the limiting equilibrium.

We thus focus on the behavior of the functions \( L \) and \( H \), corresponding to an allocation with \( \Lambda \to \infty \), defined in equation (92). The ratio of the average contact rate to the lower bound converges to:

\[
\frac{\Lambda}{\bar{\lambda}} = \frac{1}{\Lambda \int_\Lambda^\infty \frac{1}{\lambda} dF(\lambda)} = \frac{1}{\int_1^\infty \frac{1}{\rho} dH(\rho)} = e \approx 2.718,
\]

where we do the usual change in variables and then take advantage of the known functional form of \( H \).

Next, let \( \Psi'(\rho) \equiv G(\rho \Lambda) \) denote the cumulative distribution of relative contact rates. Using equation (1) and the known functional form of \( H \), we have

\[
\Psi'(\rho) = \rho^{-3}e^{\rho^{-1}} \Rightarrow \Psi(\rho) = (1 - \rho^{-1})e^{\rho^{-1}}, \quad (93)
\]

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 (68) in equilibrium, we get that this is \( \lim_{c \to 0} V_1 = (e - 3/2)\gamma \approx 1.22\gamma \).

The logic behind the other trading rates is unchanged. In particular, middlemen buy from non-middlemen \( \lim_{c \to 0} V_2 = \gamma/2 \). They sell to them at the same rate, and so finite
traders buy from finite traders at rate \((e - 2)\gamma \approx 0.72\gamma \in (\frac{1}{2}\gamma, \gamma)\) and from middlemen at rate \(\frac{1}{2}\gamma\). Finally, the expression for the rate that middlemen buy from middlemen is unchanged, giving us \(\lim_{\epsilon \to 0} V_3 = \gamma / 2\).

**Proof of Proposition 10.**

We begin with equilibrium. Solve equation (30) to get

\[
s(\lambda) = \frac{4\Delta - \lambda \int_{0}^{\lambda} s(\lambda')dM(\lambda')}{4(r + 2\gamma) + \lambda M(\lambda)},
\]

(94)

a decreasing and convex function. Condition 3 then implies that if \(C(\lambda)\) is weakly convex, all traders choose the same value of \(\lambda = \Lambda\). That is, \(\int_{0}^{\Lambda} s(\lambda')dM(\lambda') = s(\Lambda)m_\Lambda\) and \(M(\bar{\lambda}) = m_\Lambda\).

In an interior solution with \(\Lambda < \bar{\lambda}\), to find the equilibrium contact rate, we solve explicitly for \(s(\lambda)\). First, simplify (94) when \(\lambda = \Lambda\):

\[
s(\Lambda) = \frac{2\Delta}{2(r + 2\gamma) + \Lambda m_\Lambda}.
\]

Then rewrite the expression for a general value of \(\lambda\):

\[
s(\lambda) = \frac{\Delta \left(8(r + 2\gamma) + 4\Lambda m_\Lambda - 2\lambda m_\Lambda\right)}{(2(r + 2\gamma) + \Lambda m_\Lambda)(4(r + 2\gamma) + \lambda m_\Lambda)},
\]

(95)

which is continuously differentiable. Finally, using condition 3, it follows that if \(C(\lambda)\) is weakly convex, the equilibrium choice of \(\Lambda\) solves \(-\gamma s'(\Lambda) = rC'(\Lambda)\), thus

\[
\frac{4\gamma \Delta m_\Lambda}{(2(r + 2\gamma) + \Lambda m_\Lambda)(4(r + 2\gamma) + \lambda m_\Lambda)} = rC'(\Lambda).
\]

Finally, use the steady state misalignment equation, \((r + \gamma + \frac{\Lambda}{2}m_\Lambda) m_\Lambda = \gamma (1 - m_\Lambda)\), to rewrite this as

\[
\frac{\Delta m_\Lambda^3}{\gamma + (r + 2\gamma)m_\Lambda} = rC'(\Lambda)
\]

(96)

Note that the left hand side is increasing in \(m_\Lambda\) and hence decreasing in \(\Lambda\). This therefore uniquely defines the equilibrium choice of \(\Lambda\) if it is interior.

We next turn to the planner’s problem. As in equilibrium, it is straightforward to show that only meetings between two misaligned traders result in trade. Replicating the proof of
Proposition 9, we get that the optimal surplus function satisfies

\[ \Delta = (r + 2\gamma)S(\lambda) + \frac{\lambda}{2} \int_0^{\bar{\lambda}} (S(\lambda) + S(\lambda'))dM(\lambda') \Rightarrow S(\lambda) = \frac{2\Delta - \lambda \int_0^{\bar{\lambda}} S(\lambda')dM(\lambda')}{2(r + 2\gamma) + \lambda M(\lambda)}, \]

decreasing and convex.

We also obtain that the planner has \( dF(\lambda) > 0 \) only if \( \lambda \) maximizes

\[ -\gamma S(\lambda) - rC(\lambda) - \lambda \theta, \]

analogous to condition (29). Convexity of \( S \) implies that if the cost function is convex, the planner places all weight on a single value of \( \lambda \).

Replicating the arguments for equilibrium, when \( \lambda = \Lambda \):

\[ S(\Lambda) = \frac{\Delta}{r + 2\gamma + \Lambda m_\Lambda} \]

Use this to get

\[ S(\lambda) = \frac{\Delta(2(r + 2\gamma) + 2\Lambda m_\Lambda - \lambda m_\Lambda)}{(r + 2\gamma + \Lambda m_\Lambda)(2(r + 2\gamma) + \lambda m_\Lambda)}. \]  

(97)

It follows that the optimal choice of \( \Lambda \) satisfies the first order condition

\[ \frac{2\gamma \Delta m_\Lambda}{(r + 2\gamma + \Lambda m_\Lambda)(2(r + 2\gamma) + \lambda m_\Lambda)} = rC'(\Lambda) + \theta. \]

(98)

Analogous to equation (87), at an interior single mass point \( \Lambda < \bar{\lambda}, \theta = rC'(\Lambda) \), which implies the first order condition for the planner reduces to

\[ \frac{\gamma \Delta m_\Lambda}{(r + 2\gamma + \Lambda m_\Lambda)(2(r + 2\gamma) + \lambda m_\Lambda)} = rC'(\Lambda). \]

Again use the steady state misalignment equation, \( (r + \gamma + \frac{\Lambda}{2} m_\Lambda) m_\Lambda = \gamma(1 - m_\Lambda) \), to rewrite this as

\[ \frac{\Delta m_\Lambda^3}{2(2\gamma - (r + 2\gamma)m_\Lambda)} = rC'(\Lambda). \]

(98)

Again, the left hand side is increasing in \( m_\Lambda \), hence decreasing in \( \Lambda \), so this uniquely defines the optimal choice of \( \Lambda \) at an interior optimum.

To prove that the equilibrium contact rate is inefficiently high, note that for any \( \Lambda > 0 \), \( 0 \leq m_\Lambda < \frac{\gamma}{r + 2\gamma} \) and therefore \( \gamma + (r + 2\gamma)m_\Lambda < 4\gamma - 2(r + 2\gamma)m_\Lambda \). It follows that the left hand side of equation (96) is always bigger than the left hand side of equation (98) at a fixed value of \( \Lambda > 0 \). Since both left hand sides are increasing in \( m_\Lambda \), it follows that the equilibrium
misalignment rate must be weakly lower than the optimal one, strictly so if the equilibrium rate is less than \( \frac{\gamma}{r+2\gamma} \) and the cost function is continuously differentiable. Again using the steady state misalignment equation, the same condition ensures that the equilibrium contact rate strictly exceeds the optimal one whenever the rate is positive. ■

B Additional Material

B.1 Deriving Symmetric Value Functions

Start with the value function (3). Nash bargaining implies the trading rule (4) and trading prices (5). Also use \( m_\lambda = 2\mu_{\lambda,1} = 2\mu_{\lambda,0} \) and \( 1 - m_\lambda = 2\mu_{\lambda,1} = 2\mu_{\lambda,0} \). For traders in the high state, this gives us

\[
rv_{\lambda,h,1} = \delta_{h,1} + \gamma(v_{\lambda,l,1} - v_{\lambda,h,1}) + \frac{\lambda}{4} \int_0^{\bar{\lambda}} \left( m_\lambda (v_{\lambda,h,0} + v_{\lambda',h,1} - v_{\lambda,h,1} - v_{\lambda',h,0})^+ \right. \\
+ \left. (1 - m_\lambda) (v_{\lambda,h,0} + v_{\lambda',l,1} - v_{\lambda,h,1} - v_{\lambda',l,0})^+ \right) dF(\lambda'),
\]

(99)

\[
rv_{\lambda,h,0} = \delta_{h,0} + \gamma(v_{\lambda,l,0} - v_{\lambda,h,0}) + \frac{\lambda}{4} \int_0^{\bar{\lambda}} \left( m_\lambda (v_{\lambda,h,1} + v_{\lambda',h,0} - v_{\lambda,h,0} - v_{\lambda',h,1})^+ \right. \\
+ \left. (1 - m_\lambda) (v_{\lambda,h,1} + v_{\lambda',l,0} - v_{\lambda,h,0} - v_{\lambda',l,1})^+ \right) dF(\lambda').
\]

(100)

Since \( (v_{\lambda,h,0} + v_{\lambda',h,1} - v_{\lambda,h,1} - v_{\lambda',h,0}) = -(v_{\lambda,h,1} + v_{\lambda',h,0} - v_{\lambda,h,0} - v_{\lambda',h,1}) \),

\[
(v_{\lambda,h,0} + v_{\lambda',h,1} - v_{\lambda,h,1} - v_{\lambda',h,0})^+ - (v_{\lambda,h,1} + v_{\lambda',h,0} - v_{\lambda,h,0} - v_{\lambda',h,1})^+ \\
= v_{\lambda,h,0} + v_{\lambda',h,1} - v_{\lambda,h,1} - v_{\lambda',h,0}.
\]

Symmetrically,

\[
(v_{\lambda,h,0} + v_{\lambda',l,1} - v_{\lambda,h,1} - v_{\lambda',l,0})^+ - (v_{\lambda,h,1} + v_{\lambda',l,0} - v_{\lambda,h,0} - v_{\lambda',l,1})^+ \\
= v_{\lambda,h,0} + v_{\lambda',l,1} - v_{\lambda,h,1} - v_{\lambda',l,0}.
\]
This allows us to subtract equation (100) from (99):

\[
r(v_{\lambda,h,1} - v_{\lambda,h,0}) = \delta_{h,1} - \delta_{h,0} + \gamma(v_{\lambda,f,1} + v_{\lambda,h,0} - v_{\lambda,h,1} - v_{\lambda,f,0}) \\
+ \frac{\lambda}{4} \int_0^\lambda \left( m_{\lambda'}(v_{\lambda,h,0} + v_{\lambda',h,1} - v_{\lambda,h,1} - v_{\lambda',h,0}) \\
+ (1 - m_{\lambda'})(v_{\lambda,h,0} + v_{\lambda',f,1} - v_{\lambda,h,1} - v_{\lambda',f,0}) \right) dF(\lambda')
\]

(101)

The same logic using the Bellman equations for traders in the low preference state gives us

\[
r(v_{\lambda,l,0} - v_{\lambda,l,1}) = \delta_{l,0} - \delta_{l,1} + \gamma(v_{\lambda,f,1} + v_{\lambda,h,0} - v_{\lambda,h,1} - v_{\lambda,f,0}) \\
\frac{\lambda}{4} \int_0^\lambda \left( m_{\lambda'}(v_{\lambda,l,0} - v_{\lambda,l,0} - v_{\lambda',l,1}) \\
+ (1 - m_{\lambda'})(v_{\lambda,l,1} + v_{\lambda',h,0} - v_{\lambda,l,0} - v_{\lambda',h,1}) \right) dF(\lambda').
\]

(102)

Next, subtract equation (102) from (101):

\[
r(v_{\lambda,h,1} + v_{\lambda,l,1} - v_{\lambda,h,0} - v_{\lambda,l,0}) = \delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0} \\
\frac{\lambda}{4} \int_0^\lambda \left( v_{\lambda,h,0} + v_{\lambda,l,0} - v_{\lambda,h,1} - v_{\lambda,l,1} + v_{\lambda',h,1} + v_{\lambda',f,1} - v_{\lambda',h,0} - v_{\lambda',f,1} \right) dF(\lambda').
\]

Rewrite this as

\[
v_{\lambda,h,1} + v_{\lambda,l,1} - v_{\lambda,h,0} - v_{\lambda,l,0} = \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0} + \frac{\lambda}{4} z}{r + \frac{\lambda}{4}},
\]

(103)

where \( z \equiv \int_0^\lambda \left( v_{\lambda',h,1} + v_{\lambda',l,1} - v_{\lambda',h,0} - v_{\lambda',l,0} \right) dF(\lambda') \). Integrating over \( dF(\lambda) \), we can rewrite this as

\[
z = \int_0^\lambda \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0} + \frac{\lambda}{4} z}{r + \frac{\lambda}{4}} dF(\lambda)
\]

or subtracting \( z \) from both sides,

\[
0 = \int_0^\lambda \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0} - rz}{r + \frac{\lambda}{4}} dF(\lambda).
\]

The unique solution to this equation is \( z = (\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0})/r \). Plugging this into (103) gives

\[
v_{\lambda,h,1} + v_{\lambda,l,1} - v_{\lambda,h,0} - v_{\lambda,l,0} = \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0}}{r}.
\]

(104)
This is one of the key observations.

Next, using \( s(\lambda) \equiv \frac{1}{2}(v_{\lambda,h,1} + v_{\lambda,l,0} - v_{\lambda,l,1} - v_{\lambda,h,0}) \), we have

\[
s(\lambda) + s(\lambda') = \frac{1}{2}(v_{\lambda,h,1} + v_{\lambda,l,0} - v_{\lambda,l,1} - v_{\lambda,h,0} + v_{\lambda',h,1} + v_{\lambda',l,0} - v_{\lambda',l,1} - v_{\lambda',h,0})
\]

\[
= v_{\lambda,h,1} - v_{\lambda,h,0} + v_{\lambda',l,0} - v_{\lambda',l,1},
\]

where the second line simplifies the first using equation (104), which implies \( v_{\lambda,l,0} - v_{\lambda,l,1} = v_{\lambda,h,1} - v_{\lambda,h,0} - \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0}}{r} \) and \( v_{\lambda',h,1} - v_{\lambda',h,0} = v_{\lambda',l,0} - v_{\lambda',l,1} + \frac{\delta_{h,1} + \delta_{l,1} - \delta_{h,0} - \delta_{l,0}}{r} \). A similar proof establishes that

\[
v_{\lambda,l,1} - v_{\lambda,l,0} + v_{\lambda',l,0} - v_{\lambda',l,1} = -s(\lambda) + s(\lambda')
\]

\[
v_{\lambda,h,1} - v_{\lambda,h,0} + v_{\lambda'h,0} - v_{\lambda',h,1} = s(\lambda) - s(\lambda')
\]

\[
v_{\lambda,l,1} - v_{\lambda,l,0} + v_{\lambda',l,0} - v_{\lambda',l,1} = -s(\lambda) - s(\lambda').
\]

**B.2 Optimal Investment with Taxes**

We repeat the first step of the derivation of the joint surplus \( s(\lambda) \) under the Pigouvian tax and transfer system that implements the optimum, as discussed in section 5.2. Specifically, we propose to tax each trader a fixed amount \( \bar{\tau} \) whenever she meets another trader and, at the same time, to subsidize each meeting by doubling its joint surplus. For traders in the high state, this then gives us the counterparts to equations (99) and (100) with taxes and transfers,

\[
r v_{\lambda,h,1} = \delta_{h,1} + \gamma(v_{\lambda,l,1} - v_{\lambda,h,1}) + \frac{\lambda}{2} \int_0^\lambda \left( m_{\lambda'}(v_{\lambda,h,0} + v_{\lambda',h,1} - v_{\lambda',h,0})^+ + (1 - m_{\lambda'})(v_{\lambda,h,0} + v_{\lambda',l,1} - v_{\lambda',l,0})^+ \right) dF(\lambda') - \lambda \bar{\tau}, \quad (105)
\]

\[
r v_{\lambda,h,0} = \delta_{h,0} + \gamma(v_{\lambda,l,0} - v_{\lambda,h,0}) + \frac{\lambda}{2} \int_0^\lambda \left( m_{\lambda'}(v_{\lambda,h,1} + v_{\lambda',h,0} - v_{\lambda',h,1})^+ + (1 - m_{\lambda'})(v_{\lambda,h,1} + v_{\lambda',l,0} - v_{\lambda',l,1})^+ \right) dF(\lambda') - \lambda \bar{\tau}. \quad (106)
\]

It is then immediate that \( \bar{\tau} \) does not affect \( s(\lambda) \), the private surplus from holding the asset. Furthermore, the only difference between the value functions with and without the tax and subsidy is that in the former case, the option value of trade is multiplied by \( \frac{\lambda}{2} \) rather than \( \frac{\lambda}{4} \), because of the subsidy. Applying the analogous adjustment to the value functions in the
low state, equations (8) and (9) with taxes can be written as

\[ \frac{r}{2}(v_{\lambda,1} + v_{\lambda,0}) = \delta_0 + \gamma s(\lambda) \]

\[ + \frac{\lambda}{2} \int_0^{\lambda} ((s(\lambda) + s(\lambda'))^+ m_{\lambda'} + (s(\lambda) - s(\lambda'))^+(1 - m_{\lambda'})) dF(\lambda') - \lambda \bar{\tau}, \]  

(107)

\[ \frac{r}{2}(v_{\lambda,1} + v_{\lambda,0}) = \delta_1 - \gamma s(\lambda) \]

\[ + \frac{\lambda}{2} \int_0^{\lambda} ((-s(\lambda) + s(\lambda'))^+ m_{\lambda'} + (-s(\lambda) - s(\lambda'))^+(1 - m_{\lambda'})) dF(\lambda') - \lambda \bar{\tau}, \]  

(108)

while equation (10) reads

\[ \Delta = (r + 2\gamma)s(\lambda) + \frac{\lambda}{2} \int_0^{\lambda} \left( ((s(\lambda) + s(\lambda'))^+ - (s(\lambda) - s(\lambda'))^+) m_{\lambda'} + (s(\lambda) - s(\lambda'))^+ - (s(\lambda) - s(\lambda'))^+ (1 - m_{\lambda'}) \right) dF(\lambda'). \]

(109)

This is identical to equation (73) and it follows that \( s(\lambda) = S(\lambda) \) under this tax and transfer system. Finally, the optimal ex-ante investment decision under this system is such that newborn traders choose their contact rate \( \lambda \) to maximize their expected present value \( \frac{1}{2}(v_{\lambda,1} + v_{\lambda,0}) - C(\lambda) \). Using equation (108), we can restate the choice of contact rates, previously given in (12), as one of choosing \( \lambda \) to maximize

\[ r\pi_\lambda \equiv \delta_1 - \gamma s(\lambda) \]

\[ + \frac{\lambda}{2} \int_0^{\lambda} \left( (-s(\lambda) + s(\lambda'))^+ m_{\lambda'} + (-s(\lambda) - s(\lambda'))^+(1 - m_{\lambda'}) \right) dF(\lambda') - \lambda \bar{\tau} - rC(\lambda). \]  

(110)

Since \( s(\lambda) = S(\lambda) \), it follows from comparison with equation (29) that the conditions characterizing socially and privately optimal ex-ante investment are identical when \( \bar{\tau} = \theta \), where \( \theta \) is defined in equation (88). We have already established that \( \theta = rc \) in the linear cost case with no upper bound.