The Formation of Exchanges:
Risk Sharing and Information Aggregation*

Kei Kawakami†

June 1st, 2010

Abstract

Why are some financial markets segmented and opaque? I propose a two-stage game framework to study how such a market structure arises as a result of interaction among traders and exchanges. In the first stage, traders and exchanges play a market formation game, which determines a number and size of exchanges. In the second stage, motivated by both risk sharing and speculation, traders in each exchange play a trading game. I show that the gain from trade is hump-shaped in the number of traders. When many traders speculate, prices become so informative that the ex ante gain from risk sharing is reduced. I show that this can endogenously constrain the exchange size in the market formation game. With free entry of exchanges, the number and size of exchanges will be such that prices don’t reveal much information, and thus each trader’s gain from trade is maximized.

Keywords: Asymmetric information, CARA-normal, Imperfect competition, Information sharing, Market segmentation.

*I thank Andrew Atkeson, Antonio Bernardo, Francisco Buera, Bruce Carlin, Roger Farmer, Mark Garmais, Shogo Hamasaki, Christian Hellwig, Hugo Hopenhayn, Ichiro Obara, Lee Ohanian, Marek Pycia, Avanidhar Subrahmanyam, Pierre-Olivier Weill, Mark Wright, and participants in seminars at DBJ and UCLA for comments and suggestions. All errors are mine. The latest version is downloadable at http://kei.bol.ucla.edu/

†Department of Economics, University of California Los Angeles, e-mail: kei@ucla.edu


1 Introduction

In many financial markets, trades occur within a restricted, small group of traders, rather than in an open-access, large auction. Examples of such segmented markets include the markets for corporate bonds, foreign exchange, derivatives, and inter-bank lending. These markets are often described as opaque since only a small amount of trade-relevant information is shared among market participants. Empirical studies suggest that sometimes traders prefer segmentation and opaqueness to openness and transparency, and that some markets have indeed evolved in that direction (Pagano and Roell (1992); Board and Sutcliffe (1996); Gemmill (1996); Bias and Green (2007); Ready (2008)).

These observations raise the following questions: Why are segmentation and opaqueness attractive to traders? How does such a market structure arise? What is its welfare implication? I develop a theoretical market formation model to address these questions. The model shows that the emergence of segmented and opaque financial markets is consistent with traders’ desire for risk sharing, which would be difficult in an open and transparent exchange. Existing empirical evidence offers support for the importance of risk sharing in financial markets (Reiss and Werner (1998, 2004)).

I analyze a two-stage game framework, where exchanges and traders play a market formation game at the first stage, and traders in each exchange play a trading game at the second stage. At time zero, referred to as the ex ante stage, all traders are identical. Exchanges provide costly trading service for traders and charge fixed entry fees. Traders decide in which exchange to participate. The equilibrium of market formation game determines the number and size of exchanges. At time one, referred to as the interim stage, traders in each exchange play a trading game. Traders trade a risky asset given their private information. The equilibrium of the trading game determines allocation of the asset. At time two, referred to as the ex post stage, all the uncertainty is resolved and traders enjoy realized profits.

In the trading game at the second stage, traders are motivated by both risk sharing and speculation. Each trader comes to the exchange with two pieces of private information.
First, his risky asset endowment determines his own risk sharing needs and provides risk sharing opportunities for other traders. If this were the only factor each trader brings into the exchange, the gain from trade would increase as the exchange size increases. Second, each trader has a private signal about the risky asset return, which motivates informed speculation. In the trading game, each trader submits an order that is contingent on the market-clearing price (“limit order”). This allows each trader to condition his order in part on other traders’ signals. Thus, the price not only clears the market but also serves as a communication tool among traders. Importantly, this information sharing decreases the benefit of risk sharing, because the risk to be shared has decreased due to information revealed through prices.

The second mechanism mentioned above is one variant of the Hirshleifer Effect, that is, that the revelation of information can prevent risk sharing and decrease welfare. In the trading game, the trade-off between the size of risk sharing pool and the Hirshleifer Effect produces a hump-shape gain. On the one hand, a large exchange increases risk sharing possibilities, and thus, trading volume. On the other hand, increased speculation makes prices more informative and decreases the benefit of risk sharing. The latter effect works independently of the quantity traded, because it is an informational force. Therefore, in a large exchange, gain from trade approaches zero even though each trader’s trading volume is large. The hump-shape is illustrated in Figure 1.

![Figure 1](image-url)
The hump-shape influences how exchanges form at the first stage. I begin my analysis of the market formation game by assuming that there is only one exchange. In this case, since traders will participate as long as the gain from trade exceeds the entry fee, the exchange can extract all the gain from trade as its profit. At the fee level that maximizes the monopoly exchange’s profit, the exchange could further expand its size by lowering its fee, but there is no incentive to do so. This is the case because each additional trader creates negative externality and lowers the gain from trade of all other traders inside the exchange. At the monopoly optimum, the gain from each additional trader and the loss from a decreased gain for current traders in the exchange are equalized. Hence, the monopoly exchange limits the entry of traders at some finite size.

I then study how competition between two exchanges changes the analysis. At first glance, it would seem that Bertrand-like competition would induce the two exchanges to bid down the entry fee to marginal cost. Interestingly, this is not necessarily the case. If the potential number of traders is large enough, both exchanges set monopoly fees, and there is no incentive to lower the fee, just like in the monopoly case. When the potential number of traders is not large enough, the exchanges cannot reach monopoly size, and competition starts to matter. Still, as long as traders create negative externality at the shared market size, the exchanges make profits. This occurs because each exchange knows that setting a higher fee will not drive all of the traders to their competitor.

Competition between two exchanges becomes intense when the shared market size is small enough that the negative externality among traders disappears. In this case, there is strong incentive to cut fees, since a slightly lower fee allows each exchange to steal all of the traders from another exchange. In sum, competition has bite when it forces the exchanges to have small size. This creates the incentives to expand exchange sizes by cutting fees and, thus, decreases the exchanges’ profits. I show that free entry increases the number of exchanges until each exchange size attains the peak of the hump-shaped gain derived in the
trading game (i.e., until each trader’s gain from trade is maximized).

I briefly review the related literature next. Following that, Sections 2 and 3 analyze the trading game for a given number of traders, which I take as an exchange size. Section 2 describes the environment of the game. Section 3 characterizes equilibrium and derives hump-shaped gain from trade. The exchange size is endogenized through a market formation game in Section 4. In particular, I study the implication of the hump-shape for competition among exchanges. Section 5 provides two extensions to the model. Section 6 concludes. Appendix A contains all proofs. Appendix B compares endogenous asset supply in the trading game with exogenous supply.

1.1 Related literature

It is well documented that some financial markets are highly segmented and opaque. The over-the-counter (OTC) market for bonds is one extreme example where trades are often conducted as a bilateral negotiation between two traders. Biais and Green (2007) study the history of the U.S. bond market and show that bond trading was active on the New York Stock Exchange (NYSE) as a transparent limit-order market until the 1940s. Then it declined as trading migrated to the OTC market. Today, only a small fraction of bond trading is done through a centralized market on the NYSE.

Such migration has also been observed for stocks. For example, Pagano and Roell (1992) argue that when an opaque dealer market for cross-listed stocks was created in London in 1986, it drew “volume away from the auction markets of continental Europe.” Another example of preference for opaqueness was found in the London Stock Exchange (LSE). According to Board and Sutcliffe (1996), delayed publication of trades was introduced in 1989 in response to pressure from market makers. Gemmill (1996) also studies the LSE and argues that “there will be a tendency for large transactions to gravitate toward the least transparent available market.”1

---

1Gemmill (1996) also states that, recognizing this tendency, the Paris Bourse introduced a delayed publication in 1994 and there have been suggestions that Frankfurt do the same. Comerton-Forde and Tang
More recently, private trading networks that match orders without routing them to a public exchange — sometimes called “dark pools” — have been developed and gained popularity.\(^2\) In a study of three major dark pools, Ready (2008) states that their trading systems cater to a specific group of institutions and eliminate small orders to control information leakage.

Reiss and Werner (1998, 2004) study inter-dealer trades in the LSE to test whether risk sharing is an important motivation for trading. They conclude that dealers are risk-averse and use inter-dealer trades to share inventory risk. They also study anonymous brokered trading systems in the LSE, and state that the LSE rules restricted access to the anonymous systems, making the system a risk sharing device for registered market makers.

The trading game in this paper uses the noisy rational expectation equilibrium (REE) setup, which was initiated by three papers: Grossman and Stiglitz (1980), Hellwig (1980), and Diamond and Verrecchia (1981). My model environment is similar to Diamond and Verrecchia (1981), except that I study imperfect competition among traders. Following Kyle (1989), I characterize a Nash equilibrium in demand functions. As Reny and Perry (2006) criticize, most noisy REE models have the feature that no trade occurs without the presence of irrational agents. To get around this problem and facilitate welfare analysis, I replace noise trading in Kyle (1989) with rational traders’ private endowments, as in Diamond and Verrecchia (1981). A similar environment is studied in Madhavan (1992), but he does not study the implication of the model for traders’ ex ante participation, which is a main focus of my work.\(^3\) Kovalenkov and Vives (2009) is another closely related work, which studies the convergence property of the Kyle model. However, they maintain noise trading and focus on how closely a competitive model can approximate a strategic model as noise trading increases.

\(^2\)See the following article for dark pools:
www.businessweek.com/investor/content/oct2007/pi2007102_394204.htm.

\(^3\)In Madhavan (1992), prior variance of the asset value is assumed to be infinite. Hence, ex ante utility is not well defined.
Reny and Perry (2006) study the strategic foundation of an REE by considering a double auction with risk-neutral agents. They show that as the market size increases equilibrium converges to a fully revealing and efficient REE. Although their analysis is much more general than this paper in many dimensions, one important difference is that traders are risk-averse in my model. I show that when price aggregates private information, risk-averse traders’ ex ante welfare may be reduced. This force, missing in many models with risk-neutral agents, has an important implication for the efficiency of large markets.

Welfare reducing information sharing in my model is one variant of the Hirshleifer Effect (Hirshleifer (1971)). Pithyachariyakul (1986) compares the social welfare of an exchange economy under two market structures: Walrasian and monopolistic market makers. In the Walrasian system, the asymmetry of information disappears because of the communication through prices. This allows for an efficient risk sharing against the remaining risk, while some insurance opportunities are destroyed. In the market maker system, preserved asymmetric information means that more risk sharing is possible but the allocation is less efficient. The relative efficiency of two systems depends on the nature of uncertainty. Hatchondo, Krusell and Schneider (2008) consider an asset market in which investors have private information about both payoffs and their own exposure to an aggregate risk. Although they do not relate their results to the Hirshleifer Effect, they show that there is less risk sharing than in a symmetric information benchmark similar to the result in my model.

The Hirshleifer Effect has drawn more attention in finance literature. Marin and Rahi (1999, 2000) analyze the trade-off between the adverse selection effect and the Hirshleifer Effect in the security design problem. They show that traders may be better off with the set of securities that leaves prices noisy. Bernardo and Judd (1997) examine welfare of traders in the Grossman and Stiglitz (1980) exchange economy by replacing noise traders with rational traders whose risk aversions are not common knowledge. They show that information gathering leads to suboptimal risk sharing. Lyons (1996) and Viswanathan and Wang (2002) also study the implication of the Hirshleifer Effect in different trading
arrangements. Compared with these works featuring the Hirshleifer Effect, the contribution of this paper is to analyze the Hirshleifer Effect under a fully strategic environment with a closed-form solution. In particular, the model identifies the mechanism through which the Hirshleifer Effect can shape a market structure.

My model also yields a version of no-trade result due to adverse selection. Glosten (1989), Bhattacharaya and Spiegel (1991), Subrahmanym (1991), Bhattacharaya, Reny and Spiegel (1995), and Spiegel and Subrahmanyam (1992, 2000) all show no-trade result under various circumstances. Mailath and Noeldeke (2008) analyze two-dimensional asymmetric information, and provide necessary and sufficient conditions for market breakdowns. Compared with these works, no-trade result in my model is much weaker since I focus on the linear equilibrium. However, the papers cited above all assume a nested information structure, i.e., information asymmetry only between a group of symmetrically informed agents and uninformed agents. In contrast, I analyze an arbitrary number of traders with heterogeneous information. In this environment, if the asset supply is exogenous noise trading as in Kyle (1989), it is well known that in the limit as noise trading vanishes, markets become so illiquid that profit-maximizing quantities are driven to zero. In my model, asset supply is endogenous, and I show that the same symptom can arise away from the zero-noise limit.4

Competition among exchanges studied in this paper is related to a different strand of literature. In my model, multiple exchanges competing by setting prices can make positive profits through endogenous capacity constraint. This result is related to Kreps and Scheinkman (1983)’s finding that Bertrand competition can yield Cournot outcome if competing firms precommit to capacity constraint. The difference of my model from theirs is that exchanges do not choose capacity in my model. Nevertheless, capacity constraint endogenously arises through informational interaction in the trading game. Finally, I borrowed the view that exchanges compete to get traders “on board” from the literature on two-sided markets (Rochet and Tirole (2006)). Although there is no side in my model, exchanges’

---

4 Also, since interim allocation is Pareto inefficient due to dispersed endowments, a simple no-trade theorem (e.g. Milgrom and Stokey (1982)) does not apply.
objective and pricing strategy are modeled in a similar way that “platforms” are modeled in this literature. I show that a monopoly exchange does not perfectly internalizes externality among traders. This is because utilities of marginal rather than all participating traders determine the marginal revenues generated by an additional trader, as explained by Weyl (2009) in a more general context.

2 Trading game environment

Consider $N$ traders indexed by $i = 1, \ldots, N$ in a single exchange, who have CARA preferences and trade a risky asset with liquidation value $v$. The asset value $v$ is not known but traders have a common prior that it is a realization of a normal random variable with mean zero and variance $\tau_v^{-1}$. Each trader has two types of private information: (i) risky asset endowment $x_i$, and (ii) a private signal $s_i$ about $v$.

Endowment $x_i$ can be interpreted as follows: Suppose that traders are professional and their endowments are created as a result of preceding trades with their customers. The professional traders then come to an inter-dealer market to adjust their positions. An implicit assumption here is that customers do not have direct access to the inter-dealer market. The sum of endowments $\sum_{i=1}^N x_i$ is the total amount of the risky asset in the market. Though each trader does not know the other traders’ endowments, traders have a common prior that each trader’s endowment is a realization of an independent normal random variable with mean zero and variance $\tau_x^{-1}$.

It is assumed that the private signal takes the form $s_i = v + \varepsilon_i$, where $\varepsilon_i$ is unobserved noise in the signal, and follows a normal distribution with mean zero and variance $\tau_{\varepsilon}^{-1}$. To summarize, $2N + 1$ random variables $v, x_1, \ldots, x_N, \varepsilon_1, \ldots, \varepsilon_N$ are assumed to be normally and independently distributed with zero means, and variances

$$Var(v) = \tau_v^{-1}, \ Var(x_i) = \tau_x^{-1}, \ Var(\varepsilon_i) = \tau_{\varepsilon}^{-1}.$$
After observing the realization of private information \((s_i, x_i)\), each trader chooses her order \(q_i(p; s_i, x_i)\) which is explicitly conditioned on the market-clearing price \(p\). The utility function of traders is

\[
U(\pi_i) = -\exp(-\rho \pi_i),
\]

where \(\pi_i\) is trader \(i\)'s profit and \(\rho\) is the constant absolute risk aversion. Profit is the sum of return on the new position \(q_i + x_i\) and payment or receipt for net trading \(q_i\), and thus,

\[
\pi_i = v(q_i + x_i) - pq_i.
\]

Following Kyle (1989), a demand schedule \(q_i(p; s_i, x_i)\) is allowed to be any convex-valued, upper-hemicontinuous correspondence which maps prices \(p\) into non-empty subsets of the closed infinite interval \([-\infty, \infty]\). The following market-clearing rule gives a well-defined market price for all strategy choices and a well-defined allocation given a finite price. An auctioneer calculates the set of market-clearing prices and quantity allocation which satisfy

\[
\sum_{i=1}^{N} q_i(p; s_i, x_i) = 0. \tag{1}
\]

An allocation with infinite trade is assumed to be market-clearing if and only if there is at least one positive and one negative infinite quantity at that price. If a market-clearing price exists, the auctioneer chooses the price with minimum absolute value and the market-clearing quantity allocation that minimizes the sum of squared quantities traded. If there is positive excess demand at all prices, \(p = \infty\) is announced and all buyers receive negative infinite utility. If there is negative excess demand at all prices, \(p = -\infty\) is announced and all sellers receive negative infinite utility.

Suppose that the set of market-clearing prices that satisfy (1) is empty. From the restriction on the correspondence, it must be either: (i) there is positive excess demand at all prices, or (ii) that there is negative excess demand at all prices. Since positive (negative) infinite price will be announced in the first (second) case, some traders will receive negative infinite utility. Similarly, the situation with both positive and negative infinite quantity at
some price level will give at least one trader negative infinite utility. Thus, infinite prices and quantities do not occur in equilibrium.

This market-clearing rule specifies how a market-clearing price depends on the strategies of traders. To make this dependence explicit, write

\[ p = p(q), q_i = q_i(q), \]

where \( q = (q_1, ..., q_N) \) is a vector of strategies. A rational expectations equilibrium with \textit{imperfect competition} is defined as a \( q \) that satisfies

\[
E [U ((v - p(q)) q_i(q) + vx_i)] \geq E [U ((v - p(q')) q_i(q') + vx_i)]
\]

(2)

for all \( i = 1, ..., N \) and for any \( q' \) differing from \( q \) only in the \( i \)-th component. For comparison, a rational expectations equilibrium with \textit{price-taking} is defined as a \( q \) that satisfies

\[
E [U ((v - p(q)) q_i(q) + vx_i)] \geq E [U ((v - p(q)) q_i(q') + vx_i)]
\]

(3)

for all \( i = 1, ..., N \) and for any \( q' \) differing from \( q \) only in the \( i \)-th component. In (2), the expected utility under deviation from \( q \) is calculated by taking into account the effect on the price \( p(q') \). In (3) trader \( i \) assumes that the price is not affected by his deviation.

An important feature of the trading game is illustrated in the market-clearing condition (1). In most models that use a CARA-normal framework, an independent random variable, often named noise trading, is added in (1). Noise trading is also referred to as a liquidity trader, since it offers a random trading opportunity for the other traders. In the same sense, each rational trader in this model is also a liquidity trader because each trader’s private endowment creates a random trading opportunity for the other traders. However, while the randomness of noise trading is exogenously fixed, the impact of random endowments depends on how much traders change their initial positions to reduce the risk associated
with the asset value. This in turn affects the amount of noise in the price, and thus the traders’ risk assessments. This mutual dependence between the traders’ risk sharing needs and price informativeness endogenizes liquidity supply in the trading game.

The absence of irrational traders facilitates welfare analysis. Here, I introduce a welfare measure based on each trader’s ex ante utility. Let $E_i[.]$ denote trader $i$’s interim expectation, $E[|s_i, x_i, p]$, which is conditional on the information that he has at the time of order submission.

**Definition 1** *Interim certainty equivalent (ICE) profit $\Pi_{i}^{\text{ice}}$ is defined by*

$$E_i[\exp (-\rho \pi_i)] = \exp (-\rho \Pi_{i}^{\text{ice}}).$$

*Ex ante certainty equivalent (ACE) profit $\Pi_{i}^{\text{ace}}$ is defined by*

$$E[\exp (-\rho \Pi_{i}^{\text{ice}})] = \exp (-\rho \Pi_{i}^{\text{ace}}).$$

**Definition 2** *Gain from trade (GFT) is defined by the difference in ACE profit between trading and not trading.*

When there is a trade, the profit is given by $\pi_i = v(q_i + x_i) - pq_i$, where quantity traded $q_i$ and the price are determined in equilibrium. When there is no trade, profit is $\pi_i = vx_i$. Note that each trader can always achieve this no-trade profit by submitting zero demand. In fact, no trade is one equilibrium outcome. Let $Var_i[.]$ denote trader $i$’s conditional variance, conditional on the same information used to form $E_i[.]$. Recognizing that conditional variance of $v$ will be independent of $i$ due to normality, I define the precision of each trader $\tau \equiv (Var_i[v])^{-1}$. Further, I assume $1 - \frac{\rho^2}{\tau x \tau x} > 0$ throughout the paper so that ex ante utility is well defined.

**Lemma 1 (no-trade)** A no-trade equilibrium always exists. In this equilibrium, $E_i[v] = \ldots$
\[
\frac{\gamma + \delta}{\tau} s_i \text{ and } \tau = \tau_v + \tau_\varepsilon. \quad ICE \text{ profit is } \Pi_{iNT}^{\text{ice}} = E_i[v]x_i - \frac{\rho^2}{2\tau} x_i^2. \quad ACE \text{ profit is } \Pi_{iNT}^{\text{ace}} = \frac{1}{2\rho} \log \left( 1 - \frac{\rho^2}{\tau_v \tau_x} \right).
\]

This lemma serves two purposes. First, the option of not trading, \( \Pi_{iNT}^{\text{ace}} \), sets a participation constraint for any trading arrangements. Note that neither ICE nor ACE profit depends on the number of traders \( N \) due to the autarkic nature. The next section derives ACE profit in equilibrium with trading. Gain from trade in Definition 2 is then given by the derived ACE profit with trading minus \( \Pi_{iNT}^{\text{ace}} \) in Lemma 1.

Second, the form of ACE profit is useful to understand an informational property of the trading game. No-trade ACE profit does not depend on precision of private signals \( \tau_\varepsilon \), while ICE profit does. In a no-trade equilibrium, private signal affects the interim stage valuation of the asset, but does not change the allocation. Informational difference at the interim stage can not affect ex ante utility, unless it changes trading behavior and thus allocation. On the other hand, \( \tau_v \) and \( \tau_x \) are relevant for no-trade ACE profit, since the former affects ex ante valuation and the latter directly affects interim and ex post allocation.

### 3 Trading game equilibrium

This section contains three subsections. The first subsection establishes the existence of equilibrium. As a benchmark, three special cases, price-taking equilibrium, no-speculation equilibrium, and public information equilibrium are also presented. Subsection 3.2 analyzes hump-shaped gain from trade. The last subsection provides numerical evaluation of the hump-shape.

#### 3.1 Equilibrium characterization

I use a guess-and-verify method and start from the following conjecture.
Conjecture There exists a symmetric linear equilibrium, in which the strategies are

\[ q_i(p; s_i, x_i) = \beta_s s_i - \beta_x x_i - \beta_p p \tag{4} \]

for some positive coefficients \((\beta_s, \beta_x, \beta_p)\).

Given a conjecture that the other traders use a strategy of the form (4), a maximization problem of trader \(i\) is solved. The solution, which is the best response to the conjectured strategy, must have the same form by symmetry. This defines a set of equations in coefficients \(\beta_s, \beta_x\) and \(\beta_p\), which defines a fixed point in \(\mathbb{R}^3\).

The key feature of the trading game is information sharing through prices. On the one hand, how much information is shared among traders and how much is kept private affect trading, because trading is partially driven by informational differences among traders. On the other hand, trading reveals information and changes the information distribution among traders. To capture this interaction between information and trading, I formally define the amount of information sharing. Recall that precision \(\tau\) for any trader is defined by

\[ \tau \equiv (Var_i[v])^{-1}. \]

This is bounded below by its value in the no-trade equilibrium \(\tau_v + \tau_\varepsilon\). At the other extreme, if traders can share all the available information about \(v\), \(\tau\) will be \(\tau_v + N\tau_\varepsilon\). Thus, there exists a constant \(\varphi \in [0, 1]\) such that

\[ \tau = \tau_v + \tau_\varepsilon + \varphi (N - 1) \tau_\varepsilon. \tag{5} \]

The parameter \(\varphi\) measures the share of the other \(N-1\) traders’ private signals that are revealed through prices. If \(\varphi\) is zero, prices do not reveal any information about \(v\). If \(\varphi\) is one, prices reveal all the private signals in the market.

The following lemma characterizes \(\varphi\) given conjecture (4), and its relation to the valuation
Lemma 2 (information sharing) Given (4), the amount of information sharing is

\[
\varphi = \frac{1}{1 + \left(\frac{\beta_s}{\beta_p}\right)^2},
\]

and

\[
E_i[v] = \frac{\tau s - \varphi \beta_s x_i}{\tau} + \frac{\tau (1 - \varphi)}{\tau} s_i + \frac{\tau \varphi \beta x_i}{\tau} x_i + \frac{\tau \varphi N}{\tau} \beta_p p.
\]

First, it is useful to see how \(\varphi\) depends on \(\beta_s\) and \(\beta_x\). Since \(\beta_s\) measures the sensitivity of orders to a private signal \(s_i\), higher \(\beta_s\) implies more information sharing in the market. The relation between \(\varphi\) and \(\beta_x\) is negative, because as orders become more responsive to endowments, a market-clearing price becomes more noisy (recall that endowment contains no information about \(v\)).

Second, the conditional mean of the asset return is increasing in \(s_i\), \(x_i\) and \(p\), provided that \(\beta_s\), \(\beta_x\) and \(\beta_p\) are positive in equilibrium. The signs of \(s_i\) and \(p\) are positive, since a higher private signal or market-clearing price implies a higher return \(v\). The sign of \(x_i\) is positive for the following reason: Since large endowments lower prices through the market-clearing condition, a trader with high endowment rationally expects (for a given market-clearing price) that the other traders’ private signals indicate a high return.

Finally, Lemma 2 illustrates the interaction between information and trading. On the one hand, trading behavior captured by \(\beta_s\) and \(\beta_x\) affects information sharing \(\varphi\). On the other hand, \(\varphi\) affects the traders’ valuation of the asset \(E_i[v]\) and \(Var_i[v] = \tau^{-1}\). Since the traders maximize a mean-variance criterion due to CARA-normal set-up, \(E_i[v]\) and \(\tau\) characterize the optimal order \(q_i\).

To derive the best response of trader \(i\), I start from the market-clearing condition \(\sum_{j \neq i} q_j + q_i = 0\). Given conjecture (4),

\[
\sum_{j \neq i} q_j = \beta_s \sum_{j \neq i} s_j - \beta_x \sum_{j \neq i} x_j - (N - 1) \beta_p p.
\]

Solving for price, we obtain

\[
p = c_i + \lambda q_i,
\]

where \(c_i = \frac{\beta_s}{\beta_p} \bar{s}_{-i} - \frac{\beta_x}{\beta_p} \bar{x}_{-i}\), \(\bar{s}_{-i}\) and \(\bar{x}_{-i}\) are the average of private signals and endowments held by all traders except \(i\), and \(\lambda \equiv \frac{1}{(N - 1) \beta_p}\). Equation (6) is trader \(i\)'s residual supply curve with slope \(\lambda\) and an intercept \(c_i\), where \(\lambda\) measures each trader’s price impact. Although
trader $i$ can not directly condition her order on $c_i$, $(s_i, x_i, p)$ provides the same information as $(s_i, x_i, c_i)$ through the linear relationship (6).

The objective function at the interim stage is $E_i[-\exp(-\rho \pi_i)]$. Because of the normality of $v$ conditional on each trader’s information, the objective of the trader is to maximize the mean-variance criterion

$$E_i[v](q_i + x_i) - \frac{\rho}{2} Var_i [v](q_i + x_i)^2 - pq_i$$

subject to (6).

The first-order condition and the second-order condition are

$$E_i[v] - \rho Var_i [v](q_i + x_i) = c_i + 2\lambda q_i = p + \lambda q_i,$$  \hspace{1cm} (8)$$

$$2\lambda + \rho Var_i [v] > 0.$$  \hspace{1cm} (9)$$

The first-order condition (8) equates the trader’s marginal valuation of the asset (left-hand side) to his marginal cost of trading (right-hand side). Marginal valuation is decreasing in the trader’s new position $q_i + x_i$ because of risk aversion $\rho$ and conditional variance $Var_i [v]$. Marginal cost is increasing in his order $q_i$ due to price impact $\lambda$. The second-order condition (9) needs to be verified since $\lambda \equiv \frac{1}{(N-1)\beta_p}$ is determined in equilibrium.\(^5\)

From (8) and $Var_i [v] = \tau^{-1}$, we obtain

$$q_i^* = \frac{E_i[v] - p - \frac{\rho}{2}x_i}{\lambda + \frac{\rho}{2}}.$$  \hspace{1cm} (10)$$

Using $E_i[v] = k_1 s_i + k_2 x_i + k_3 p$, where $k_1, k_2, k_3$ are given in Lemma 2, we obtain

$$q_i^* = \frac{k_1 s_i - \left(\frac{\rho}{2} - k_2\right)x_i - (1 - k_3) p}{\lambda + \frac{\rho}{2}}.$$  \hspace{1cm} (11)$$

\(^5\)The second-order condition (9) must hold with strict inequality. With equality, traders would like to order infinite quantity because $E_i[v] - \rho Var_i [v] x_i \neq c_i$ almost surely.
This is the best response of trader $i$ when the other traders use (4). By equating coefficients of (4) and (11), we have three equations

$$\beta_s = \frac{k_1}{\lambda + \frac{\varphi}{\tau}} = \frac{\tau \varphi}{\lambda \tau + \rho} (1 - \varphi), \quad (12)$$

$$\beta_x = \frac{\varphi - k_2}{\lambda + \frac{\varphi}{\tau}} = \frac{\rho}{\lambda \tau + \rho} \left( 1 - \varphi \frac{\tau \beta_x}{\rho \beta_s} \right), \quad (13)$$

$$\beta_p = \frac{1 - k_3}{\lambda + \frac{\varphi}{\tau}} = \frac{\tau}{\lambda \tau + \rho} \left( 1 - \varphi N \frac{\tau \beta_p}{\beta_s} \right), \quad (14)$$

where $\lambda \equiv \frac{1}{(N-1)\beta_p}$ and $\varphi \equiv \frac{1}{1 + (\frac{\tau \beta_x}{\rho \beta_s})^\frac{1}{\tau \beta_x}}$.

The following proposition establishes the existence and uniqueness of a symmetric linear equilibrium with trading. For the rest of the paper, I use variables with upper bar to denote the average across traders. For example, $\bar{s} \equiv \frac{1}{N} \sum_{i=1}^{N} s_i$ and $\bar{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i$.

**Proposition 1 (trade equilibrium)** (a) A symmetric linear equilibrium with trading exists if and only if $\frac{\tau \beta_x}{\rho \beta_s} < 1 - \frac{2}{N}$. If it exists, it is unique. (b) Optimal order is

$$q_i = \left( 1 - 2\varphi - \frac{1}{N-1} \right) \left\{ \frac{\tau \varphi}{\rho} s_i - x_i - \frac{\tau \varphi}{\rho} \frac{(\tau_\varphi + \tau_x + (\tau_\varphi + \tau_x) N) \chi}{\tau_\varphi + \tau_x + (\tau_\varphi + \tau_x) \chi} p \right\},$$

where $\varphi = \frac{\chi}{1 + \chi}$ and an information parameter $\chi \equiv \frac{\tau \beta_x}{\rho \beta_s}$. (c) Equilibrium price is

$$p = \frac{\tau \beta_x (1 + N \chi)}{\tau_\varphi + \tau_x + (\tau_\varphi + \tau_x) \chi} (\bar{s} - \frac{\rho}{\tau \beta_s} \bar{x})$$

and corresponding quantity traded is

$$q_i = \frac{\tau \varphi}{\rho} \left( 1 - 2\varphi - \frac{1}{N-1} \right) \left\{ (s_i - \bar{s}) - \frac{\rho}{\tau \beta_s} (x_i - \bar{x}) \right\}.$$

Henceforth, a symmetric linear equilibrium with trading shall be called a *trade equilibrium*. A number of important observations follow from **Proposition 1**. First, a trade equilibrium is a partially revealing REE. In the equilibrium, not all of signals are revealed through prices, because $\varphi = \frac{\chi}{1 + \chi}$ is smaller than one. The amount of information sharing is increasing in the information parameter $\chi \equiv \frac{\tau \beta_x}{\rho \beta_s}$. When $\tau_\varphi$ and $\tau_x$ are higher, or when risk aversion $\rho$ is lower, more information is shared in equilibrium.

Second, the amount of information sharing is independent of $N$. This implies that a trade equilibrium is partially revealing even in the limit where $N$ goes to infinity. However, since $\tau = \tau_\varphi + \tau_\varphi + \varphi (N - 1) \tau_\varphi$, conditional variance of $v$ goes to zero in the limit.
Third, Proposition 1 shows that \( \frac{\beta_s}{\beta_x} = \frac{\rho}{\tau_\varepsilon} \). This ratio represents the balance between risk sharing and speculation. The more risk averse traders are, the more weight they put on their endowments in order to adjust their positions. The more precise their private signals are, the more weight they put on private signals.

Finally, three coefficients \( \beta_s, \beta_x, \beta_p \) have well-defined limit as \( N \) goes to infinity. This suggests that there will be non-trivial trade in a large market. In fact, it will be shown later that trading volume per trader is increasing in \( N \). This is in part because price impact disappears in the large market, since \( \lambda \equiv \frac{1}{(N-1)\beta_p} \) goes to zero as \( N \) goes to infinity.

To gain more insight, I study three special cases of the trade equilibrium. The first case is a price-taking equilibrium. This is obtained by setting \( \lambda = 0 \) in (8) and (9). This shuts down price impact for any market size. The second case is a no-speculation equilibrium, where a private signal is assumed to be useless and traders trade only for a risk sharing motive. Setting \( \tau_\varepsilon = 0 \) shuts down the speculation by making common prior the only information about the asset value. The third case is a public signal equilibrium, where average of private signals is publicly announced before trading. This also eliminates asymmetric information, but allows traders to have maximum amount of information available in the original environment.

**Corollary 1 (price-taking)** Assume \( \lambda = 0 \) in (8) and (9). In equilibrium, the amount of information sharing \( \varphi \) is same as in the strategic case. Optimal order is obtained by replacing \( 1 - 2\varphi - \frac{1}{N-1} \) in the strategic case with \( 1 - \varphi \).

The result is immediate by setting \( \lambda = 0 \) in (12)–(14), so the proof is not provided. Notice that the price impact \( \lambda \) affects the level of coefficients through (12)–(14), but not the ratio. Recall from Lemma 2 that \( \varphi \) depends only on \( \frac{\beta_s}{\beta_x} \). Since this ratio is independent of price impact, so is information sharing. Moreover, from (5) and Lemma 2, neither \( \tau \) nor \( E_i[v] \) is affected by price-taking assumption. Where does the strategic behavior matter then? Corollary 1 shows that three coefficients are larger under the price-taking
assumption. Therefore, the strategic behavior reduces trading volume relative to the price-taking behavior, while still allowing each trader to have the same amount of information. The next two symmetric information cases illustrate the consequence of information sharing.

**Corollary 2 (no-speculation)** Assume \( \tau_\varepsilon = 0 \). (a) \( E_i[v] = 0 \), \( \tau = \tau_v \), and traders submit \( q_i = -(1 - \frac{1}{N-1}) x_i - \frac{\tau_v}{\rho} (1 - \frac{1}{N-1}) p \) in equilibrium. (b) Equilibrium price is \( p = \frac{-\rho}{\tau_v} \overline{x} \) and corresponding quantity traded is \( q_i = -(1 - \frac{1}{N-1}) (x_i - \overline{x}) \). (c) ACE profit is increasing in \( N \) with a finite upper bound.

**Corollary 2** shows that, without speculation, price contains no information about the asset value. Traders trade only for risk sharing purpose and a larger exchange provides larger trading volume per trader. In the limit where the market size approaches infinity, after the trade everyone holds average positions that are the ex ante mean \( (q_i + x_i = \overline{x} = 0) \). Moreover, a larger exchange provides larger gain from trade per trader. This positive externality comes from the fact that each additional trader creates more risk sharing opportunity for the other traders.

**Corollary 3 (public signal)** Assume \( \overline{s} \equiv \frac{1}{N} \sum_{i=1}^{N} s_i \) is publicly announced before trading. (a) \( E_i[v] = \frac{N \tau_x}{\tau} \overline{s}, \tau = \tau_v + N \tau_\varepsilon \), and traders submit \( q_i = \frac{\tau_x N}{\rho} \left(1 - \frac{1}{N-1}\right) \overline{s} - \left(1 - \frac{1}{N-1}\right) x_i - \frac{\tau_v}{\rho} (1 - \frac{1}{N-1}) p \) in equilibrium. (b) Equilibrium price is \( p = \frac{\tau_x N}{\tau} \overline{s} - \frac{\rho}{\tau_v} \overline{x} \) and corresponding quantity traded is \( q_i = -(1 - \frac{1}{N-1}) (x_i - \overline{x}) \). (c) As \( \tau_\varepsilon \to 0 \), ACE profit approaches ACE profit for the no-speculation case. (d) For any \( \tau_\varepsilon > 0 \), as \( N \to \infty \), ACE profit approaches no-trade ACE profit.

Note that in this equilibrium, the equilibrium price contains information about the asset value, but traders do not learn anything from it. Because the equilibrium price is only a noisy signal of public signal \( \overline{s} \), it does not contain any additional information on top of \( \overline{s} \). **Corollary 3** shows that, with public signal, trading volume is same as in **Corollary 2**. However, the perfect risk sharing limit is associated with zero gain from trade. This implies that compared to the no-speculation case, introduction of the public signal significantly
decreases traders’ welfare in a large exchange. The reason is the Hirshleifer Effect: public signal reduces ex ante gain from insurance. This case tells us that trading volume may not be a good measure to assess traders’ welfare from ex ante perspective. In particular, if we would like to consider traders’ participation decision at the ex ante stage, the amount of information sharing \textit{at the interim stage} must be carefully analyzed.

\textbf{Figure 2} illustrates gain from trade in the no-speculation case, which is monotonically increasing and approaching the perfect risk sharing upper bound as the exchange size increases. By definition, the no-trade case has zero gain.

![Graph of gain from trade per trader vs. number of traders in one exchange](image)

\textbf{Figure 2}

What happens if traders can use private signals? \textbf{Figure 3} illustrates the consequence of speculation in a trade equilibrium in \textbf{Proposition 1}. Introduction of private signals decreases gain from trade, and importantly, the negative effect does not disappear in a large market. The next subsection derives the hump-shape shown in \textbf{Figure 3}.

![Graph of gain from trade per trader vs. number of traders in one exchange](image)

\textbf{Figure 3}
3.2 Hump-shaped gain from trade

In this subsection, I first establish the non-monotonicity of gain from trade by studying two polar cases: a small and large exchange. After that, I discuss the source of the hump-shape. Finally, the hump-shape is characterized.

I start with a small exchange. Proposition 1 states that there is a parameter region where a trade equilibrium does not exist. The condition in Proposition 1 can be written as,

$$N(\chi) \equiv \frac{2}{1-\chi} < N \text{ for } \chi < 1. \quad (15)$$

To focus on a linear equilibrium, $\chi < 1$ is fixed for the rest and (15) is assumed to be satisfied. Combined with the maintained assumption $1 - \frac{\mu^2}{\tau_v \tau_x} > 0$, this implies $\frac{\tau_x}{\tau_v} < 1$, i.e., one private signal contains less information than prior information. The next lemma shows what happens when the exchange size approaches the lower bound (15).

**Lemma 3 (small exchange)** As condition (15) becomes close to equality, trading volume approaches zero and $\lambda$ goes to infinity.

This establishes the left end of Figure 3. When the exchange size is small, each trader’s price impact becomes huge, and so does marginal cost of trading (right-hand side of (8)). Price impact can become arbitrarily large because $\lambda \equiv \frac{1}{(N-1)\beta_p}$, $\beta_p$ is proportional to $1 - 2\phi - \frac{1}{N-1}$, and thus, when (15) approaches equality, the denominator of $\lambda$ approaches zero. When the exchange size is small, each trader anticipates a large price impact, and thus makes his order less sensitive to the price (smaller $\beta_p$). Moreover, when other traders’ orders become less sensitive to price, each trader has a bigger price impact. Hence, traders’ price impacts are mutually reinforcing through the market-clearing rule. Due to this mechanism, price impact and marginal cost can be so large that profit maximizing trading is driven down to zero. This argument does not hold under price-taking assumption.
It is well known in the literature that in models with exogenous noise trading, the symptom in Lemma 3 arises when noise trading, and thus noise in prices, disappears. In contrast, as condition (15) becomes close to equality, price becomes more informative, but remains noisy. Nevertheless, trading volume goes down to zero and price impact goes to infinity. The key to this difference is that information sharing is independent of \( N \) in this model, while it positively depends on \( N \) in the model with noise trading. Note that \( \beta_p \) may become smaller either: (i) due to smaller \( N \) or (ii) due to larger \( \varphi \). When \( \varphi \) is independent of \( N \), the small market increases price impact through the first channel. When \( \varphi \) is positively related to \( N \), the small market decreases \( \varphi \). This increases \( \beta_p \) through the second channel and partially offsets the increased price impact through the first channel. In the model with noise trading, the impact of noise trading becomes large relative to rational traders’ private signals in a small market, thus making the price less informative. This occurs because noise trading is exogenously fixed independent of \( N \). In the current model, the impact of speculation and that of risk sharing are both proportional to \( N \). This leaves \( \varphi \) independent of \( N \) in equilibrium.\(^6\)

The following proposition expresses the properties of a large exchange.

**Proposition 2 (large exchange)** (a) Let \( E[v|p] \) be expected return without private information. It follows that \( \frac{E[v|p]-E[v]}{p} = \zeta - 1 \), where \( \zeta < 1 \). Also, \( \zeta \) is increasing in \( N \) and approaches 1 as \( N \) goes to infinity. (b) Expected per-trader trading volume is increasing in \( N \) and approaches a finite upper bound as \( N \to \infty \). (c) As \( N \to \infty \), ACE profit approaches no-trade ACE profit.

Proposition 2 characterizes the right end of Figure 3 and establishes non-monotonicity of gain from trade. The first and second parts of Proposition 2 show that as the exchange size increases, price becomes more informationally efficient and trading volume increases. In the meantime, however, the last part shows that gain from trade decreases to zero.\(^6\)

\(^6\)Appendix B provides further discussion on this point.
When the exchange size is finite, price is only a noisy signal of \( v \). This price is not informationally efficient in the following sense. Note that \( E[v|p] - p \) is the excess return forecast by an econometrician outside the exchange, who knows the rules of the trading game, observes prices, but cannot directly observe the traders’ private information.\(^7\) According to **Proposition 2**, the expected excess return is negative. This implies that if an uninformed trader without endowments enters the exchange, he has an incentive to short sell the asset.\(^8\) In other words, there is arbitrage opportunity for outsiders *without risk sharing needs*, if there is no restriction on market entry and short-selling for such traders. **Proposition 2** shows that the expected excess return approaches zero when \( N \to \infty \). This limiting exchange achieves the informationally efficient price that leaves no arbitrage opportunity for outsiders.

**Proposition 2** also shows that expected trading volume *per trader* is increasing in \( N \) with a finite limit. Hence, the model provides a rationale for “liquidity externality,” where liquidity here is measured by trading volume per trader. This result may appear to support a large exchange. However, this is not consistent with traders’ welfare because trading volume does not go with gain from trade.

The last part of **Proposition 2** states that the large exchange with the informative price and large trading volume does not provide gain from trade. Why is there no gain from trade even though trading volume is large? It is easy to show that \( p = \frac{\beta}{\beta_p} \left( \bar{s} - \frac{\sigma^2}{\tau^2} \right) \) converges to \( v \) almost surely as \( N \) goes to infinity. In other words, the price fully reveals the risky asset value. This necessarily decreases expected gain from trade, since (i) there is no return on private signal, and (ii) the asset will be traded at a price of indifference. Recall that the marginal valuation of the asset (left-hand side of (8)) is \( E_i[v] - \rho \text{Var}_i[v] (q_i + x_i) \). We know that conditional variance of \( v \) goes to zero in the limit. It is easy to verify \( E_i[v] \) converges to \( v \) almost surely. Hence, information sharing in a large exchange destroys dispersion of marginal valuation created by different endowment levels. In the limit, even though the

---

\(^7\) Hatchondo, Krusell and Schneider (2008) provide a similar argument.

\(^8\) Of course, equilibrium behavior of informed traders would be different with uninformed traders in the market. This analysis is left for future work.
physical allocation is still dispersed at interim stage, there is no dispersion in valuation.

To show where the hump-shape comes from, I analyze information aggregation through prices in more detail. In particular, I show that there is a sense in which risk sharing can be “most efficiently” done in the exchange of the intermediate size. Recall that each trader’s profit is \( \pi_i = v(q_i + x_i) - pq_i = vx_i + (v - p)q_i \). When traders submit orders, they face a constrained portfolio allocation problem: Given a fixed position \( x_i \), which provides a risky return \( v \), choose \( q_i \), which provides a different risky return \( v - p \).

What matters for this portfolio choice is correlation between two returns \( v \) and \( v - p \). From the proof of Proposition 1, \( v - p = (1 - \frac{\beta_s}{\beta_p})v - \frac{\beta_s}{\beta_p} (\bar{\pi} - \frac{\rho}{\sigma} \bar{\pi}) \), with \( \frac{\beta_s}{\beta_p} < 1 \) and \( \lim_{N \to \infty} \frac{\beta_s}{\beta_p} = 1 \). Therefore, when the exchange size increases, there are two forces that affect correlation between \( v \) and \( v - p \). On the one hand, variance of \( \bar{\pi} - \frac{\rho}{\sigma} \bar{\pi} \) decreases, which makes correlation stronger. On the other hand, price becomes more unbiased \( (1 - \frac{\beta_s}{\beta_p} \to 0) \), which makes correlation weaker. The second force dominates in a large exchange and the marginal benefit of trading the “new security” with return \( v - p \) decreases. Thus, a noisy price in a small exchange can be interpreted as a financial asset that serves traders’ risk sharing needs.

This discussion is summarized in the following lemma.

**Lemma 4 (correlation)** \( \text{Corr} [v, v - p] \) is increasing in \( N \) for \( N \leq \frac{1}{\chi} \), decreasing in \( N \) for \( N \geq \frac{1}{\chi} \), and approaches zero as \( N \to \infty \). Correlation is decreasing in \( \chi \) for all \( N \).

**Lemma 4** implies that if we draw this ex ante correlation as a function of exchange size, it has a single peak at \( N = \frac{1}{\chi} \) and approaches zero in the right tail. It can be shown that the slope of this hump-shape is positive and decreasing for small \( N \), negative and decreasing for intermediate \( N \), and negative and approaching zero for large \( N \). The last part of **Lemma 4** states that higher \( \chi \) lowers this hump-shape. In the proof it is shown that only \( \frac{\beta_s}{\beta_p} \) is relevant for this correlation pattern. Hence, the result holds both under the strategic and price-taking cases.

To fully characterize gain from trade for a different exchange size, it is convenient to
relate the strategic case to the price-taking case. As explained above, the hump-shape comes from the correlation between the price and return of the asset, regardless of traders being strategic or not. In the strategic case, however, it will be shown that the hump-shape is shifted downward and to the right relative to the price-taking case. The following lemma relates the price-taking case to the strategic case.

**Lemma 5 (ICE)** In a trade equilibrium, ICE profit is

\[
\Pi_{i}^{ice} = \tilde{\lambda}\Pi_{iH}^{ice} + \left(1 - \tilde{\lambda}\right)\Pi_{iPT}^{ice} = \tilde{\lambda}\left\{ E_i[v]x_i - \frac{\rho}{2\tau}x_i^2 \right\} + \left(1 - \tilde{\lambda}\right)\left\{ px_i + \frac{\tau}{2\rho}(E_i[v] - p)^2 \right\},
\]

where \( \tilde{\lambda} \equiv \left(\frac{\lambda\tau}{\lambda\tau + \rho}\right)^2 = \left(\frac{N}{N-1}\chi + \frac{1}{N-1}\right)^2. \)

**Lemma 5** decomposes ICE profit in an intuitive way. The first term is related to the value of holding entire endowment, denoted by \( \Pi_{iH}^{ice} \). This appears to be exactly the same as a no-trade ICE \( (E_i[v]x_i - \frac{\rho}{2\tau}x_i^2) \), with the only difference in the information contained in \( E_i[v] \) and \( \tau \) due to information sharing through prices. Recall that a price-taking assumption is captured by setting \( \lambda = 0 \), which implies \( \tilde{\lambda} = 0 \). Hence, the second term in ICE profit is related to the value under the price-taking assumption, denoted by \( \Pi_{iPT}^{ice} \). This is the sum of the value of endowments evaluated at a market-clearing price and the value of speculative trading.\(^9\)

Actual ICE profit \( \Pi_{i}^{ice} \) is a weighted sum of \( \Pi_{iH}^{ice} \) and \( \Pi_{iPT}^{ice} \), where the weight is given by \( \tilde{\lambda} \equiv \left(\frac{\lambda\tau}{\lambda\tau + \rho}\right)^2 \). As the exchange size decreases, price impact \( \lambda \) increases, and ICE profit approaches \( E_i[v]x_i - \frac{\rho}{2\tau}x_i^2 \). This is the case because traders in a small exchange trade less aggressively and hold more of initial endowments. On the other hand, as the exchange size increases, the weight \( \tilde{\lambda} \) goes down toward \( \chi^2 \).

The decomposition is useful to study ex ante gain from trade. From the proof of Proposition 2(c), ACE profit is \( \Pi^{ace} = \frac{1}{2\rho} \log (\det(I_3 + 2\rho\Sigma A)) \), where \( I_3 \) is a 3-by-3 identity matrix.

\(^9\)px\(_i\) appears in \( \Pi_{iPT}^{ice} \) since, because of CARA utility, an initial position is irrelevant for the choice of new position except its informational use in \( E_i[v] \).
matrix and $A \equiv \tilde{\lambda} A_H + (1 - \tilde{\lambda}) A_{PT}$ is also a 3-by-3 matrix. Hence, ACE profit is monotonic transformation of $I_3 + 2\rho \Sigma A = \tilde{\lambda}(I_3 + 2\rho \Sigma A_H) + (1 - \tilde{\lambda})(I_3 + 2\rho \Sigma A_{PT})$. The 3-by-3 matrices $A_H$ and $A_{PT}$ contain coefficients of quadratic representation of ICE profit $\Pi_{iH}^{ice}$ and $\Pi_{iPT}^{ice}$.

First, note that ACE profit of holding endowment is same with no-trade ACE profit, because the allocation is same. Therefore, $\det(I_3 + 2\rho \Sigma A_H) = 1 - \frac{\rho^2}{\sigma_x \sigma_z}$, which corresponds to no-trade case. Second, the value associated with price-taking case $I_3 + 2\rho \Sigma A_{PT}$ is greater than no-trade case. Hence, as long as $\tilde{\lambda} > 0$, the gain from trade is lower in the strategic case than in the price-taking case. As $\tilde{\lambda}$ approaches 1, ACE profit approaches no-trade ACE profit, while as $\tilde{\lambda}$ decreases, ACE profit approaches price-taking ACE profit. Since the weight $\tilde{\lambda}$ is decreasing in $N$, the shape of gain from trade is the down-and-rightward shift of the price-taking gain from trade.

Finally, the shape of gain from trade is summarized in the following proposition:

**Proposition 3 (hump-shape)** (a) Gain from trade is increasing in $N$ for $N \leq N^*(\chi)$, decreasing in $N$ for $N \geq N^*(\chi)$, and decreasing in $\chi$ for any $N$. (b) The exchange size that maximizes gain from trade $N^*(\chi)$ is decreasing in $\chi$ for small $\chi$, and increasing in $\chi$ for $\chi$ close to 1. (c) In the limit $N \rightarrow \infty$, gain from trade converges to zero by the speed of $\frac{1}{N}$.

Properties presented in Proposition 3 are graphically shown in the next subsection.

### 3.3 Numerical evaluation

This subsection provides numerical evaluation of gain from trade. Figure 4 shows each trader’s gain from trade $GFT \equiv \Pi^{ace} - \Pi^{ace}_{NT}$ as a function of $N$ for different levels of $\chi < 1$. Figure 5 shows a total gain from trade $TG \equiv N(\Pi^{ace} - \Pi^{ace}_{NT})$ similarly.
Circle markers represent $N^*(\chi)$ in Figures 4 and 5. From Proposition 3(c), total gain from trade has an upper bound, which is illustrated in Figure 5. The exchange size which maximizes an individual gain from trade may seem “too small” since a total gain from trade
is still increasing at that size. I study the welfare implication of exchange size in the next section.

**Figure 6** compares the four cases for $\chi = 0.7$. Relative to no-speculation case, three other cases have lower gain from trade.\(^{10}\) Since no-speculation case also suffers from price impact, the difference between no-speculation and public signal can be attributed to the Hirshleifer Effect due to a public signal. Price-taking case shows that information sharing through prices can create the Hirshleifer Effect of a similar magnitude that a public signal creates. As shown in the previous subsection, gain from trade in the strategic case is lower and distorted to the right relative to the price-taking case. It is interesting that public signal case can outperform price-taking case in a small exchange, but may not do so in a large exchange. Although this is not a general property and depends on parameter configuration, it illustrates two roles of public signal. Public signal eliminates a problem associated with asymmetric information, but creates a stronger Hirshleifer Effect. The former effect can be dominant in a small exchange, while the latter effect may become more significant in a large exchange. In the next section, I focus on strategic case and study how the exchange size is determined at the ex ante stage.

\(^{10}\) Except for price-taking case at $N = 2$. Under this trading arrangement, bilateral trading is feasible only with price-taking assumption.
4 Market formation game

This section endogenizes a number and size of exchanges through a market formation game. A market formation game is played at the ex ante stage, i.e., before realization of any private information. Hence, all traders are identical. The game can be interpreted as follows. Suppose that a new financial asset is created. Traders, probably dealers, know that they can potentially make profit by trading the asset with their customers. However, to the extent that the amount of trading with customers is uncertain, dealers will be exposed to risk associated with the random position of the risky asset. Since there is potential for mutually beneficial risk sharing among these dealers, exchanges may make profits by providing a service that facilitates trading (i.e., a system that conducts order submission, market-clearing, and settlement).

To focus on the study of competition among exchanges, I abstract from ownership structure of exchanges, and assume that exchanges are independent, for-profit institutions. Each
exchange must cover an operation cost that is proportional to the number of traders in the exchange. Each exchange charges a fixed entry fee for each trader in order to maximize profit. I assume that each exchange can charge only a fixed fee, and can not charge a variable fee per usage (e.g., per trading volume or transaction value). These exchanges are the only places where traders can play the trading game analyzed in the preceding sections. Also, it is assumed that each trader can enter only one exchange, and that the trading game in each exchange occurs simultaneously so that there is no information leakage across exchanges. Note that, for exchanges to be non-trivial players in the market formation game, it must be that traders cannot directly interact to realize gains from trading. The validity of this assumption depends on who, in reality, owns the technology that implements trading. For example, see Reedy (2008) for the type of technology that typical dark pools provide.

Let $N$ be the potential number of traders in the economy, and $K$ be the number of exchanges. The market formation game proceeds in two steps. First, each exchange simultaneously sets an entry fee taking other exchanges’ fees as given. Second, each trader simultaneously decides which exchange to join, or not to join any exchanges, taking all fees and other traders’ choices as given. It is common knowledge that, after this market formation game, the trading game is played independently in each exchange.

Since no exchange will open if $GFT(N^*) \leq c$, $GFT(N^*) > c$ is assumed for the remaining analysis. Also, it is assumed that $\mathcal{N}(\chi) < \mathcal{N}$ for the same reason.

First, I introduce a measure of total surplus for a given number of exchanges.

**Definition 3** Total surplus with $K$ exchanges is $\sum_{j=1}^{K} \{GFT(N_j) - c\} N_j$, where $N_j$ is the size of exchange $j$ and $c$ is the operation cost of each exchange per trader.

Given this definition, I ask the following questions: Given $\mathcal{N}$ potential traders in the economy, how many exchanges should be set up to maximize total surplus? How large should each exchange be? Can exchange competition achieve the maximum total surplus?

\[11\] A variable fee is studied in Section 5.
In the subsections that follow, I analyze three cases: (i) a monopoly exchange \((K = 1)\), (ii) two exchanges \((K = 2)\), and (iii) free entry of exchanges (endogenous \(K\)).

### 4.1 Monopoly exchange

Consider the monopoly exchange which chooses entry fee \(\phi\) to maximize its profit. A monopoly exchange solves the following problem:

\[
\max_{\phi} (\phi - c)N \tag{16}
\]

s.t. \(GFT(N) - \phi \geq 0, \tag{17}\)

\[N \leq \overline{N}. \tag{18}\]

where \(N\) is the number of traders who join the exchange. The notation for gain from trade makes it explicit that it depends on the size of the exchange. When there is only one exchange, it may seem natural that traders come to the exchange as long as the expected gain from trade exceeds an entry fee. However, it is also an equilibrium at the second stage that no trader joins the exchange, since the exchange requires a certain number of traders \(N(\chi)\) in (15)) to create positive gain from trade. Using this no-participation equilibrium as a “threat”, any fee level \(\phi \in [c, GFT(N^*)]\) can be sustained at the first stage, followed by traders’ participation at the second stage. The next subsection discusses this issue in more detail.

For now, suppose that traders cannot use this threat. Since \(GFT(N^*) > c\) is assumed, the constraint (17) binds in equilibrium. Hence, solution is

\[N_m = \arg\max_{N \leq \overline{N}} (GFT(N) - c)N, \tag{19}\]

\[\phi_m = GFT(N_m). \tag{20}\]
Figure 7 illustrates the monopoly exchange size (19) and fee (20), assuming that $c$ is 5% of maximum gain from trade and that $\bar{N}$ is large enough to avoid corner solution $N_m = \bar{N}$.

![Figure 7](image)

The monopoly exchange’s profit for each pair of fee and market size is represented by a rectangle on a marginal cost line (black dashed line), keeping its top-right corner on a gain from trade curve (blue hump-shape). Hence, maximum profit is achieved when a rectangular hyperbola (red dotted line) defined above the marginal cost line is tangent to gain from trade. From Proposition 3(c), this monopoly size exists and finite for any $c > 0$. Also, the monopoly exchange size is greater than $N^\ast \equiv \arg \max_N GFT(N)$, since 

$$
(GFT(N_m) - c)N^\ast \leq (GFT(N^\ast) - c)N^\ast \leq (GFT(N_m) - c)N_m.
$$

To gain more insight for the monopoly exchange size, note that unless it is corner solution $N_m = \bar{N}$, it satisfies the first order condition $c - GFT'(N_m)N_m = GFT(N_m)$. In words, a monopoly exchange equates marginal revenue, which is gross utility of a marginal participating trader, with the marginal cost of having an additional trader. Marginal cost exceeds $c$, if a marginal participating trader decreases gain from trade for other participating traders.
For the later use, let $N_c \in (N^*, \infty)$ be the larger solution to $GFT(x) = c$, and let $N_1 \in (N^*, N_c)$ be a unique solution to $GFT(x) = c - GFT'(x)x$. The monopoly exchange size is characterized by $N_m = N_1$ if $N_1 \leq N$ and otherwise $N_m = N$. The monopoly fee is $\phi_m = GFT(N_m)$.

The monopoly exchange extracts all the surplus as its profit. Traders are made indifferent between paying the entry fee and not entering the exchange. Notice that the objective in (19) is total surplus for $K = 1$. Hence, under the restriction that there can be only one exchange, total surplus is maximized. A lower or higher fee decreases total surplus compared to the monopoly exchange’s profit.

However, there are potential problems with the monopoly exchange. If there are more traders than the monopoly exchange is willing to accept, some traders are excluded from trading. Also, as long as $N^* < N$, traders who join the monopoly exchange create negative externality, because the exchange size is at a downward sloping part of the gain from trade curve. This is because the monopoly exchange sets a fee equal to a marginal trader’s gain from trade, and does not internalize his negative externality imposed on all other traders on the exchange. These problems point to the possibility that competition among multiple exchanges may provide a more socially beneficial outcome.

## 4.2 Two exchanges

To simplify the analysis, I assume that two exchanges must be separate in the following sense: Once traders join one exchange, they can submit orders only in that exchange and cannot observe the other exchange’s activity. This may seem unrealistic if you think about major exchanges such as the NYSE. However, independent market organizers such as Instinet actually provide services that (i) limit the membership to a certain group of institutional traders, (ii) keep the information within their members, and (iii) restrict information sharing among the members. The model should be interpreted as competition among such trading service providers.
At the first stage, two exchanges $A$ and $B$ compete by each setting fee $\phi_A$ and $\phi_B$. At the second stage, given $\phi_A$ and $\phi_B$, trader $i$ chooses $r_i = (r_i^A, r_i^B)$ to solve

$$\max_{r_i} \sum_{k \in \{A,B\}} r_i^k \left\{ GFT \left( \sum_{j=1}^N r_j^k \right) - \phi_k \right\}$$

s.t. $r_i^k \in \{0, 1\}$ for all $k$ and $\sum_{k \in \{A,B\}} r_i^k \in \{0, 1\}$.

The strategy of trader $i$ is a mapping $r_i(\phi_A, \phi_B) : \mathbb{R}_+^2 \mapsto \{0, 1\}^2$. Equilibrium at the second stage for given $\{\phi_k\}_{k \in \{A,B\}}$ is $\{r_i^*\}_{i=1}^N$ such that for all $i$, $r_i^*$ solves (21) subject to (22). Due to constraint (22), each trader can participate in only one exchange.

Given $\{\phi_k\}_{k \in \{A,B\}}$, the equilibrium strategy $\{r_i^*\}_{i=1}^N$ determines the exchange sizes $\sum_{i=1}^N r_i^* = (N_A, N_B)$. Hence, the size of exchange $k$ is a mapping $N_k(\phi_A, \phi_B) : \mathbb{R}_+^2 \mapsto \{0, 1, \ldots, N\}$. This is a demand function from exchanges' perspective. The constraint (22) implies that $\sum_{k \in \{A,B\}} N_k \in \{0, 1, \ldots, N\}$.

In the first stage, exchanges set fees given demand functions $(N_A, N_B)$ which depend on fees. Exchange $A$’s problem is

$$\max_{\phi_A} (\phi_A - c)N_A,$$

s.t. $\sum_{i=1}^N r_i^* = (N_A, N_B)$.

Exchange $B$’s problem is symmetric. Constraint (24) reflects traders’ choice at the second stage. Traders are valuable “resources” for exchanges to generate profit. In this view, gain from trade curve can be interpreted as a production technology induced by a trading game, which transforms resources into profit. Competition among exchanges imposes further restriction (24). Equilibrium at the first stage is $\{\phi_k^*\}_{k \in \{A,B\}}$ such that for all $k$, $\phi_k^*$ solves
(23) subject to (24).

Note that at the second stage there is always a trivial equilibrium where no trader participates in any exchange no matter what the fees are. Also, asymmetric equilibria where all traders concentrate in the one exchange may exist. Finally, there may be an equilibrium where some traders join one exchange while the other traders join the other exchange, and all traders are indifferent between two exchanges. Thus, given \( N \) and \((\phi_A, \phi_B)\), there can be three types of equilibria at the second stage. The following lemma characterizes the first stage equilibrium. Let \( \phi \equiv \min\{\phi_A, \phi_B\} > 0 \) and denote two solutions to \( GFT(x) = \phi \) by \( (N^{-}_{\phi}, N^{+}_{\phi}) \) if they exist.

**Lemma 6 (a) No-participation equilibrium:** \( r^*_i = (0, 0) \) for all \( i \). This always exists.

**Lemma 6 (b) One-exchange equilibrium:** \( \sum_{i=1}^{N} r^*_i = (N_A, 0) \) or \( \sum_{i=1}^{N} r^*_i = (0, N_B) \) and \( N_k(GFT(N_k) - \phi_k) \geq 0 \) for all \( k \), with strict inequality if \( N_k = N \in (N^{-}_{\phi}, N^{+}_{\phi}) \) and equality if \( N_k = 0 \) or \( N_k = N^{+}_{\phi} = N \). This exists if and only if \( GFT(N^*) \geq \phi \) and \( N^{-}_{\phi} \leq N \). (c) Balanced-participation equilibrium: \( \sum_{i=1}^{N} r^*_i = (N_A, N_B) \) such that \( N_A N_B > 0 \) and \( GFT(N_k) - \phi_k \geq \max\{GFT(N_l + 1) - \phi_l, 0\} \) for \( k \neq l \). If this exists, \( GFT(N_k) \geq GFT(N_k + 1) \) for at least one \( k \).

It is clear that no trader has an incentive to change an exchange in all three cases in **Lemma 6.** In **Lemma 6(a) and (b),** this is due to the self-fulfilling belief of traders that “I would be the only one in that exchange”. The existence of these types of equilibria depends only on the property of trading game that there is no gain from trade for \( N \leq N(\chi) \), and not on the precise shape of gain from trade curve.

In the balanced-participation equilibrium, the one group of traders chooses exchange \( A \) and the other group of traders chooses exchange \( B \). This case is different from the other two cases since it’s existence requires that the gain from trade is decreasing in the number of traders for at least one exchange. Thus, negative externality among traders is a key for the balanced-participation equilibrium.
In this paper, I focus on the balanced-participation equilibrium. However, before further characterizing this equilibrium, I characterize the first stage equilibrium associated with no-participation and one-exchange equilibria at the second stage.

**Lemma 7 (a)** Given that no-participation equilibrium is played at the second stage, any \((\phi_A, \phi_B) \in \mathbb{R}^2\) is equilibrium at the first stage. **(b)** Suppose that one-exchange equilibrium \(N_k > 0\) is played at the second stage if \(GFT(N^*) \geq \phi_k\), with the threat to play no-participation equilibrium in response to deviations of exchanges from the following fees: The exchange \(k\) sets \(\phi_k \in [c, \phi_m]\) and the other exchange sets \(\phi_l \in \mathbb{R}^2\). This \((\phi_k, \phi_l)\) is equilibrium at the first stage. **(c)** Consider the following traders’ strategy: (i) If at least one exchange sets fee lower than \(GFT(N^*)\), play one-exchange equilibrium with the exchange with the lower fee. If two fees are the same and lower than \(GFT(N^*)\), play one-exchange equilibrium with exchange \(A\) with probability \(\frac{1}{2}\), and with exchange \(B\) with probability \(\frac{1}{2}\). (ii) Otherwise, play no-participation equilibrium. Given this strategy, \(\phi_A = \phi_B = c\) is the unique equilibrium at the first stage.

**Lemma 7** shows that by using no-participation equilibrium as a threat, wide range of fees can be supported in equilibrium at the first stage. **Lemma 7(a)** is trivial since no-participation is the worst possible outcome for exchanges.

**Lemma 7(b)** shows that a continuum of one-exchange equilibrium indexed by \(\phi_k \in [c, \phi_m]\) exists. Each equilibrium differs with respect to surplus division between the exchange and traders and exchange size. In the equilibrium with \(\phi_k = \phi_m\), the result is same with monopoly case and all the surplus goes to the exchange \(k\). In the equilibrium with \(\phi_k = c\), the second stage equilibrium is as follows. If \(\bar{N} \geq N_c\), exchange size is \(N_c\) and there is no surplus. Otherwise, exchange size is \(\bar{N}\) and all the surplus \((GFT(\bar{N}) - c)\bar{N}\) goes to traders.

In the equilibrium with intermediate fee \(\phi_k \in (c, \phi_m)\), let \(N_\phi\) be the larger solution to

\[12\text{There may also be a mixed strategy equilibrium where traders randomly choose one exchange. This analysis is complicated due to the hump-shaped gain from trade, and is left for future work.}\]
\[ \text{GF} T(x) = \phi_k. \] This satisfies \( N_{\phi} \in (N_m, N_c) \) and it is decreasing in \( \phi_k \). If \( N \geq N_{\phi} \), exchange size is \( N_{\phi} \) and all the surplus \((\text{GF} T(N_{\phi}) - c)N_{\phi}\) goes to one exchange. If \( N_{\phi} > N \), exchange size is \( N \) and the part of surplus \((\text{GF} T(N) - \phi_k)N\) goes to traders while \((\phi_k - c)N\) goes to the exchange \( k \). Note that although the equilibrium with fee level lower than monopoly fee may provide surplus for traders, it can be socially inefficient because total surplus may be reduced relative to \((\phi_m, N_m)\).\(^{13}\)

The reason why one exchange can enjoy profit in this equilibrium is because traders are not sensitive to fee difference between two exchanges. Since no trader has an incentive to go to the other exchange no matter what the fees are, even monopoly fee can be supported as an equilibrium.

**Lemma 7(c)** provides one way to resolve indeterminacy of surplus division in **Lemma 7(b)**. First, since traders are coordinating to go to the exchange with the lower fee, monopoly fee set by one exchange creates incentive for the other exchange to undercut its fee. Thus, monopoly allocation cannot be equilibrium.

Second, consider the case \( N < N_c \). In (b), whether there is surplus in equilibrium and how it is divided depends on the equilibrium indexed by \( \phi_k \in [c, \phi_m) \). In (c), there is unique equilibrium where all the surplus goes to traders. The difference between (b) and (c) comes from the incentive for price competition which exists only in (c).

Finally, consider the case \( N \geq N_c \). In (b), there is social surplus except equilibrium with \( \phi_k = c \). This is because one exchange’s market power helps realize gain from trade as its profit by restricting its size. On the other hand, there is no social surplus for (c). Due to the incentive for price competition introduced by the way traders coordinate on one-exchange equilibrium, exchanges cannot use entry fee to restrict its size. Therefore, positive surplus would invite additional trader’s participation until there is no surplus.

Equilibria in **Lemma 7** all share the feature that potentially socially beneficial exchange

\(^{13}\) If \( N_m = N \), setting \( \phi_k \in (c, \phi_m) \) only changes the division of surplus but does not affect total surplus relative to \((\phi_m, N_m)\). If \( N_m = N_1 \), setting \( \phi_k \in (c, \phi_m) \) makes exchange size \( N_{\phi} \), which is greater than \( N_1 \). This decreases total surplus.
does not open due to self-fulfilling belief. Also, as I discuss later, these equilibria are not robust to collective deviation of traders and exchange when \( N \) is large. To investigate competition between two active exchanges, I characterize the balanced-participation equilibrium.

From Lemma 6(c), the balanced-participation equilibrium is characterized by

\[
GFT(N_k) - \phi_k \geq \max\{GFT(N_l + 1) - \phi_l, 0\} \text{ for } k \neq l. \tag{25}
\]

The condition (25) is implied by

\[
GFT(N_A) - \phi_A = GFT(N_B) - \phi_B \geq 0, \tag{26}
\]

and

\[
GFT'(x) \leq 0 \text{ for } x \geq \min\{N_A, N_B\}. \tag{27}
\]

Therefore, one candidate equilibrium is that two exchanges operating at downward sloping part of the gain from trader curve, such that (26) is satisfied.\(^{14}\)

There is another candidate equilibrium. Note that (25) is also implied by

\[
GFT(N_B + 1) - \phi_B \leq GFT(N_A) - \phi_A < GFT(N_A + 1) - \phi_A < GFT(N_B) - \phi_B. \tag{28}
\]

This condition implies that the gain from trade curve is increasing at \( N_A \) while it is decreasing at \( N_B \). Hence, this must be an asymmetric equilibrium where exchange \( B \) is larger than exchange \( A \). Moreover, (28) also implies that \( GFT(N_B) - GFT(N_B + 1) > GFT(N_A + 1) - GFT(N_A) \). Thus, slope of the gain from trade curve must be greater in absolute value at \( N_B \) than at \( N_A \). In general, there may exist a trading game with a gain from trade curve which satisfies this property. However, this is unlikely to be satisfied by the trading game studied in this paper (see Figure 7). Therefore, I focus on a candidate\(^{14}\)

\(^{14}\)This may not uniquely pin down equilibrium due to discreetness of exchange size. I ignore issues with discreetness as long as it does not change qualitative implications from the model.
equilibrium characterized by (26) and (27).

The indifference condition (26) implicitly defines demand function for exchanges as a function of fees. First, if the last inequality in (26) is equality, then some traders must be excluded from exchanges and are indifferent between participating and not participating. In this case, two exchanges behave like monopoly and extract all the surplus from traders. Next, suppose the last inequality in (26) is strict. In this case \( N_A + N_B = N \) and demand function for exchange \( A \) is defined by

\[
GFT (N_A) - \phi_A - GFT (N - N_A) + \phi_B = 0. \tag{29}
\]

By implicit function theorem,

\[
\frac{\partial N_A(\phi_A, \phi_B)}{\partial \phi_A} = \frac{1}{GFT' (N_A) + GFT' (N - N_A)} \leq 0. \tag{30}
\]

By substituting (29) into (23), exchange \( A \) solves

\[
\max_{\phi_A} \{GFT (N_A(\phi_A, \phi_B)) - GFT (N - N_A(\phi_A, \phi_B)) + \phi_B - c\} N_A(\phi_A, \phi_B). \tag{31}
\]

First order condition is

\[
\left[\{GFT' (N_A) + GFT' (N - N_A)\} N_A + GFT (N_A) - GFT (N - N_A) + \phi_B - c\right] \frac{\partial N_A}{\partial \phi_A} = 0. \tag{32}
\]

Demand function and first order condition for exchange \( B \) are symmetric. The first order condition (32) implicitly defines the best response function \( \phi^*_A(\phi_B) \). Two best response functions \{\( \phi^*_A(\phi_B), \phi^*_B(\phi_A) \}\} characterize the equilibrium at the first stage.

When \( \frac{\partial N_A}{\partial \phi_A} < 0 \), it is easy to verify that the second order condition is satisfied. Then, \( \{GFT' (N_A) + GFT' (N - N_A)\} N_A + GFT (N_A) - GFT (N - N_A) + \phi_B - c = 0 \) defines
By using implicit function theorem, it can be shown that
\[
\frac{\partial \phi^*_A(\phi_B)}{\partial \phi_B} = \frac{GFT'(N_A) + GFT'((N - N_A)) + N_A\{GFT''(N_A) - GFT''(N - N_A)\}}{2\{GFT'(N_A) + GFT'(N - N_A)\} + N_A\{GFT''(N_A) - GFT''(N - N_A)\}}.
\]  

(33)

Note that even if trivial asymmetric equilibria at the first stage are assumed away, there may still be non-trivial asymmetric equilibria at the second stage where each exchange has a different fee and size. For simplicity, I focus on a symmetric equilibrium where two exchanges set the same fee and have the same size.

Imposing symmetry, two exchange must have the same exchange size \( \frac{N}{2} \) in equilibrium. As long as \( \frac{\partial N_A}{\partial \phi_A} = \frac{1}{2GFT'(\frac{N}{2})} < 0 \), first order condition boils down to \( 2GFT'\left(\frac{N}{2}\right) \frac{N}{2} + \phi_B - c = 0 \).

Therefore, two best response functions intersect at \( \phi^*_A = \phi^*_B = c + GFT'\left(\frac{N}{2}\right) \frac{N}{2}, \) where slope of best response functions is \( \frac{1}{2} \) from (33). Thus, the symmetric equilibrium is locally stable.

The following lemma summarizes competition between two active exchanges. Let \( N_2 \in (N^*, N_1) \) be a unique solution to \( GFT(x) = c - 2GFT'(x)x. \)

**Lemma 8 (two exchanges)** [Case 1. \( N_1 \leq \frac{N}{2} \)] Both exchanges set monopoly fee and each has \( N_1 \) traders. \( N - 2N_1 \) traders are excluded from trading. [Case 2a. \( N_2 \leq \frac{N}{2} < N_1 \)] Both set \( \phi = GFT\left(\frac{N}{2}\right) \) and have \( \frac{N}{2} \) traders. [Case 2b. \( N^* \leq \frac{N}{2} < N_2 \)] Both set \( \phi = c - GFT'(\frac{N}{2})N \) and have \( \frac{N}{2} \) traders. Exchanges and traders share the total surplus. [Case 3. \( \frac{N}{2} < N^* \)] There is no symmetric equilibrium with traders’ participation.

**Lemma 8** shows that two exchanges do not necessarily compete away profits even though the competition is of Bertrand type. When the number of potential traders is large enough, two exchanges can afford each being a monopoly. Competition has some bite when the number of potential traders is not large enough, such that both exchanges cannot enjoy a monopoly size. Still, exchanges can make profits as long as a shared market size is on the downward sloping part of the gain from trade curve.

Exchanges’ market power comes from negative externality among traders. Knowing that
raising its fee will not lose all the traders due to negative externality, each exchange sets fee
strategically. This can be seen in (30), where demand becomes less sensitive to change in fee
as negative externality among traders becomes severe.

In Case 3, competition becomes so severe that two exchanges cannot survive (at least as
a symmetric equilibrium). Positive externality among traders creates strong incentive to cut
the fee, since a slightly lower fee allows each exchange to steal all of the traders from the
other exchange. Figures 8 and 9 illustrate Cases 1 and 2.

In Figures 8 and 9, the length between the two vertical axes represents the number of
potential traders $N$. The size of exchange $A$ is measured from the left axis to the right, while
the size of exchange $B$ is measured to the opposite direction. In Figure 8, the number of
potential traders is large enough that each exchange can behave as if it is a monopoly. In

Figure 8

Figure 9

41
equilibrium, each exchange could expand its size by lowering its fee, but does not have such incentive.

In Figure 9, since the number of potential traders is not large enough, the shared market size is smaller than the monopoly size. In this case, there may be an incentive to set the fee lower than $GFT(\frac{N}{2})$, because expanding market size by attracting more traders might increase profit. However, Lemma 8 shows that as long as the slope of gain from trade curve at the shared market size is negative, the equilibrium fee is higher than marginal cost, and thus exchanges earn profit.

4.3 Free entry of exchanges

The analysis with two exchanges extends to a general case with $K$ exchanges. I focus on the case corresponding to Case 2 in Lemma 8, where each exchange sets $\phi = \min\{c - 2GFT'(\frac{N}{K})\frac{N}{K}, GFT(\frac{N}{K})\}$ and has $\frac{N}{K}$ traders. When the number of exchanges $K$ satisfies $N^* < \frac{N}{K}$, the fee exceeds marginal cost and thus exchanges make profits. Entry of exchanges seeking profits will make a shared market size $\frac{N}{K}$ smaller, and there will be no profit when $\frac{N}{K} = N^*$.\(^{15}\)

It is shown below that free entry of exchanges, which achieves $\frac{N}{K} = N^*$, also achieves the maximum total surplus.

**Proposition 4 (free entry)** For given $N$, free entry induces $K = \frac{N}{N^*}$ exchanges. All $N$ traders participate in one of the exchanges and the exchanges earn zero profit. This achieves the maximum total surplus $\{GFT(N^*) - c\}N$.

**Proposition 4** shows that, ignoring discreteness, free entry of exchanges can achieve the maximum total surplus. At the social optimum, there are $K = \frac{N}{N^*}$ exchanges. All exchanges set $\phi = c$ and compete away profit.

\(^{15}\)Strictly speaking, entry could stop after $K = \left\lfloor \frac{N}{N^*} \right\rfloor$ exchanges have entered, where $\left\lfloor \frac{N}{N^*} \right\rfloor$ denotes the largest integer that does not exceed $\frac{N}{N^*}$. A potential entrant could rationally expect that its entry would trigger fierce competition in Case 3.

42
First, the resulting market structure from free entry is characterized by segmentation and opaqueness. Each exchange size $N^*$ is small relative to the size of a monopoly exchange, and each preserves a high degree of asymmetric information with respect to the asset return. Moreover, none of the exchanges and traders has an incentive to create a larger exchange.

Second, as the number of exchanges increases, the equilibrium fee does not necessarily decrease. It can be shown that $\phi = c - 2GFT'(\frac{N}{K})\frac{N}{K}$ decreases in $K$ if and only if $-\frac{GFT''(\frac{N}{K})\frac{N}{K}}{GFT'(\frac{N}{K})} < 1$.\(^{16}\) If $GFT''(\frac{N}{K}) > 0$, the equilibrium fee might increase with more exchanges. This could happen when entry moves a shared market size from a relatively flat part of the gain from trade curve to a steep part. However, a sufficiently large number of exchanges eventually drives the fee down to the marginal cost.

Finally, although equilibria with smaller number of exchanges also exist (for example, one-exchange equilibrium), the symmetric balanced-participation equilibrium with free entry is the only robust equilibrium for collective deviation. In the equilibrium in Proposition 4, it is impossible to open a new exchange where some traders are better off.\(^{17}\) On the other hand, if the number of exchanges is much smaller than $K = \frac{N}{N^*}$, $N^*$ traders can always be made better off by opening a new exchange.

### 4.4 Who organizes exchanges?

So far it has been assumed that exchanges are distinct agents from traders. Exchanges are simply technological veil characterized by per trader operational cost $c$ and gain from trade curve. In a free entry equilibrium, a zero profit condition maximizes social welfare. However, technology that implements trading rule may not be inalienable from some people, who could also be a trader. For example, someone may have to supervise and regulate trading activity in an exchange. If organizing and running exchanges require certain type of human capital that are shared by traders, occupational choice between exchange organizers and traders

---

\(^{16}\)Discrete version is $\frac{GFT'(\frac{N}{K})}{GFT'(\frac{N}{K})} \frac{K}{K+1} < 1$.

\(^{17}\)In a directed search literature, a similar condition sometimes called *submarket completeness* is imposed as an equilibrium condition. See Rocheteau and Wright (2005).
starts to matter. In this case, free entry does not lead to zero profit of exchanges, because an agent with technology can also choose to be a trader using another agent’s technology, if there is freedom of occupational choice.

Lemma 9 (Occupational choice) Suppose \( m \geq 1 \) exchange organizer must be present to implement one trading game, and each agent can choose either to organize an exchange or to trade in one exchange. In a symmetric equilibrium with occupational choice, each exchange size is characterized by the solution to \( GFT(x) - c = -2GFT'(x)x(1 + x) \) and the number of exchanges is \( \frac{N}{x + m} \).

Lemma 9 implies that if those who organize exchanges have an outside option of becoming traders themselves, exchanges make positive profits even if there is no explicit entry cost, and simple efficiency result in Proposition 4 may not be obtained.

5 Extensions

In this section, I study two types of extensions to the model. First, the trading game where traders’ endowments are correlated is analyzed. Second, I study the situation where an exchange uses a variable fee which depends on trading volume.

5.1 Correlated endowments

In the baseline model, the endowments of traders were assumed to be uncorrelated. This may not be a satisfactory description if we interpret traders’ endowments as the outcome of trading with their customers. If different traders have overlapped customer bases, this introduces a positive correlation among traders’ endowments when they come to an exchange. Even if customer bases are distinct, there may still be a positive correlation if customers respond to a macroeconomic shock. In the presence of such a correlation, traders come to the exchange with biased positions, and the exchange tends to face more selling or buying
pressure depending on the direction of macro economic shocks. To investigate how this
changes the workings of exchange trading, I introduce a new notation $e_i$ for trader $i$’s initial
endowments, and assume

$$e_i = \eta + x_i,$$  \hspace{1cm} (34)

where $\eta$ is normally distributed with zero means and variances $\tau_\eta^{-1}$, and is independent from
all other random variables. Heterogenous factor $x_i$ is same as in the baseline model. Common
factor $\eta$ can be interpreted as a macro economic shock, which creates correlation in traders’
endowments. Importantly, each trader observes $e_i$ but does not separately observe $\eta$ and $x_i$.
Hence, no trader is sure about the size of $\eta$, but private knowledge about one’s own position
$e_i$ provides some idea about $\eta$. Note that as the variance of $\eta$ increases, correlation among
endowments becomes larger. Formally, for $i \neq j$

$$C_e \equiv Corr[e_i, e_j] = \frac{\tau_x}{\tau_\eta + \tau_x}. \hspace{1cm} (35)$$

In the extreme case where endowments are entirely driven by common factor, correlation
among traders’ endowments is close to one. If variance of common factor is small, endow-
ments are mainly determined by heterogenous factor and correlation is close to zero. Zero
variance of common factor ($\tau_\eta = \infty$) nests the baseline model without correlation as a special
case.

First, correlation changes the information sharing through prices. To characterize a linear
equilibrium, let us start with a conjectured strategy

$$q_i (p; s_i, e_i) = \beta_s s_i - \beta_e e_i - \beta_p p$$  \hspace{1cm} (36)

for some positive coefficients ($\beta_s, \beta_e, \beta_p$). The coefficient on endowment is now denoted by
$\beta_e$ rather than $\beta_x$. Introduce the notation $N_c \equiv 1 + C_e(N - 1)$. The next lemma generalizes
Lemma 10 Given (36), the amount of information sharing is \[ \varphi_c = \frac{1}{1 + \left( \frac{\tau_s}{\tau_x} \right)^\frac{1}{N_c}} , \] and \[ E_i[v] = \frac{\tau_x(1-\varphi)}{\tau} s_i + \frac{\tau_x \beta_{s} N_c}{\tau} e_i + \frac{\tau_x \varphi N_c}{\tau} p \] and \( \tau = \tau_v + \tau_e + \tau_e (N - 1) \varphi_c \).

To see that this is a generalization of Lemma 2, notice that \( N_c \) takes the value between 1 and \( N \). When there is no correlation, setting \( N_c = 1 \) in Lemma 10 provides Lemma 2. This lemma shows how the correlation among endowments affect information sharing. Other things fixed, higher correlation makes \( N_c \) larger in the denominator of \( \varphi_c \). Also, with correlation, the number of traders directly shows up in \( \varphi_c \).

Derivation of the best response of trader \( i \) when the other traders use (36) is same as before. By equating coefficients, we have three equations

\[ \beta_s = \frac{\tau_e}{\lambda \tau + \rho} (1 - \varphi_c) ; \]
\[ \beta_e = \frac{\rho}{\lambda \tau + \rho} \left( 1 - \varphi_c \frac{\tau \beta_e}{\rho} \right) ; \]
\[ \beta_p = \frac{\tau}{\lambda \tau + \rho} \left( 1 - \varphi_c \frac{N \tau_e \beta_p}{\tau} \right) ; \]

where \( \lambda \equiv \frac{1}{(N-1) \beta_p} \) and \( \varphi_c = \frac{1}{1 + \left( \frac{\tau_e}{\tau_x} \right)^\frac{1}{N_c}} \). Observe that (37) and (39) are same as in the case without correlation. Two equations lead to the solution \( \beta_s = \frac{\tau_e}{\rho} (1 - 2 \varphi_c - \frac{1}{N-1}) \) and \( \beta_p = \frac{\tau}{\tau_e} \beta_s \). These solutions look the same as before, except the difference between \( \varphi \) and \( \varphi_c \). Solutions of \( \frac{\beta_s}{\beta_p} \) and hence \( \beta_e \) are not as simple as before. Recall that in the baseline case \( N_c = 1 \) and we obtained \( \beta_s = \frac{\rho}{\tau_e} \) as a unique solution. When \( N_c > 1 \), we have to carefully analyze the cubic equation defined by (38). In the proof to the next lemma, I show that \( \frac{\beta_s}{\beta_p} \) is a solution to

\[ G(x) \equiv x^3 - \frac{\rho}{\tau_e} x^2 + \frac{\tau_x}{\tau_e} x - \frac{\rho \tau_e}{\tau_x N_c} \]
\[ = (x - \frac{\rho}{\tau_e})(x^2 + \frac{\tau_x}{\tau_e N_c}) + \frac{\tau_e}{\tau_x} (1 - \frac{1}{N_c}) = 0 \]

Again, when \( N_c = 1 \) we recover the baseline case solution. Since \( G(x) < 0 \) for all
\( x \leq 0 \) and \( G(x) > 0 \) for all \( x \geq \frac{\rho}{\tau_e} \), all the roots lie in the range \((0, \frac{\rho}{\tau_e})\) and there must be at least one such root. Before analyzing general solution, I examine the limit as \( N \) goes to infinity. The following lemma establishes the existence and multiplicity of a symmetric linear equilibrium with trading in this case.

**Lemma 11** Let \( N \) go to infinity. (a) Two symmetric linear equilibria with trading exist if and only if \( \chi < \frac{1}{4} \). If \( \chi = \frac{1}{4} \), two equilibria are the same. If \( \chi > \frac{1}{4} \), there is no symmetric linear equilibrium with trading. (b) Optimal order is \( q_i = \frac{\tau_e}{\rho} s_i - \frac{1\pm\sqrt{1-4\chi}}{2} x_i - \frac{\tau}{\tau - \tau_e} \frac{\tau_e}{\rho} p \), with \( \varphi_c = 0 \) but \( \tau = \tau_v + \tau_e + \frac{\tau_e}{2\chi \epsilon_c (1-2\chi \pm \sqrt{1-4\chi})} \). (c) Equilibrium price is \( p = v - \frac{1\pm\sqrt{1-4\chi}}{2} \frac{\rho}{\tau_e} \eta \) and corresponding quantity traded is \( q_i = \frac{\tau_e}{\rho} \left( \epsilon_i - \frac{1\pm\sqrt{1-4\chi}}{2} \frac{\rho}{\tau_e} x_i \right) \).

First, compared to **Proposition 1**, parameter region for the existence of equilibrium becomes smaller even in the limit. Second, information sharing is quite different from the baseline case. Without correlation, \( \varphi_c \) had a finite limit and total amount of information \( \tau = \tau_v + \tau_e + \varphi_c \tau_e (N - 1) \) increased to infinity along with \( N \). With correlated endowments, \( \varphi_c \) approaches zero but \( \tau \) has a finite limit. Finally, equilibrium price is a noisy signal of asset value, and the only source of noise is macro shock \( \eta \), since all the heterogenous shocks disappear in a large exchange. Since price does not perfectly reveal the asset value in the limit, there is a gain from trade in a large exchange. The remained macro shock in the aggregate endowment plays the role of noise traders and prevents the Hirshleifer Effect from fully destroying insurance opportunity.

In the limit where \( N \) goes to infinity, the model environment is almost equivalent to that of Ganguli and Yang (2009) except that traders are strategic here. Under price-taking assumption, they show that there exist two equilibria under the same condition in **Lemma 11(a)**. The one equilibrium, which they call COM-REE and has more informative price than the other equilibrium, corresponds to the equilibrium with \( 1 - \sqrt{1-4\chi} \) in **Lemma 11**. Therefore, this lemma serves as a strategic foundation to their analysis.

Multiplicity result in a large exchange may not carry over to an exchange of finite size.
In general, there may be at most three roots to (40), but the small root typically violates the second order condition and thus does not constitute an equilibrium. The following lemma characterizes the cubic equation (40).

**Lemma 12 (a)** Equation (40) has three positive real roots if and only if \( \chi \in \Phi_{N_c} \), where \( \Phi_{N_c} \equiv (\chi^-, \chi^+) \) and \( \chi^\pm = \frac{N_c^2 + 18N_c - 27 \pm \sqrt{(N_c^2 + 18N_c - 27)^2 - 64N_c^3}}{8N_c^2} \). (b) \( \Phi_{N_c} \neq \emptyset \) if and only if \( N_c > 9 \). (c) Given \( N_c \geq 9 \), if \( \chi = \chi^- \) or \( \chi = \chi^+ \), there exist two positive real roots. (d) If \( \chi \notin \Phi_{N_c} \), there exists only one real positive root. (e) For a solution of (40) to constitute an equilibrium, it must satisfy \( x \in \left( \frac{\rho}{r_c}, \frac{1}{r_c(N_c - 2)}, \frac{\rho}{r_c} \right) \).

Note that in the limit as \( N \) goes to infinity, \( \lim_{N \to \infty} \Phi_{N_c} = (0, \frac{1}{4}) \). In this case, zero is always the solution to (40), but this violates the second order condition since \( \varphi_c = 1 \) if \( \frac{\beta e}{\beta_s} = 0 \). More generally, the second order condition imposes additional restriction on the solution characterized in Lemma 12(a)-(d). Lemma 12(e) sets lower bound on \( \frac{\beta e}{\beta_s} \). If \( \frac{\beta e}{\beta_s} \) is too low, from Lemma 10, too much information is shared and this results in violation of the second order condition. In a large exchange, this condition is not an additional restriction, since it becomes \( x \in \left( 0, \frac{\rho}{r_c} \right) \), which is already implied by (40). In a finite exchange, the solution \( x \) is a function of \( C_e \) and \( N \), therefore the condition in Lemma 12(e) imposes a joint restriction on the primitive parameters of the model. Although general characterization is not available, brief inspection of (40) and Lemma 12(e) implies that the larger the size of exchange becomes, the less restrictive the restriction on the parameters will be.

### 5.2 Exchange’s pricing strategy

In the section 4, it was assumed that exchanges use only a fixed entry fee, which does not affect the property of trading game equilibrium except the size of exchanges. In practice, financial intermediaries often use variable fees which depend on trading activity. This allows for additional channel through which exchanges can make profits. In general, variable fees can take a complicated form that depends arbitrarily on trading activity and traders’ identity.
To keep the analysis tractable, I assume that the fee structure is symmetric across traders, and can depend only on trading volume $q_i$. Denote the variable fee collected from trader $i$ by $\phi_v(q_i)$. Total fee collected from $N$ traders is $\sum_{i=1}^{N} \phi_v(q_i)$. Assuming differentiability of $\phi_v(q_i)$, trader $i$’s first and second order conditions are modified as follows.

$$E_i[v] - \rho \text{Var}_i [v] (q_i + x_i) = p + \lambda q_i + \phi'_v(q_i), \quad \quad (41)$$

$$2\lambda + \rho \text{Var}_i [v] + \phi''_v(q_i) > 0. \quad \quad (42)$$

Variable fee $\phi_v(q_i)$ affects trading behavior by changing both the first-order condition (41) and the second-order condition (42). First, linear fee of the form $\phi_v(q_i) = a q_i$ does not collect any fee in the aggregate since $\sum_{i=1}^{N} \phi_v(q_i) = a \sum_{i=1}^{N} q_i = 0$ by a market-clearing rule. This may appear to be a zero-sum transfer (from buyers to sellers if $a > 0$ and from sellers to buyers if $a < 0$), but it does not affect the allocation. It is easy to check that the equilibrium price drops by $a$ and all the other equilibrium objects remain the same.

To collect fee from traders, $\phi_v(q_i)$ must be non-linear. To preserve the linearity of equilibrium, I focus on the fee of the form $\phi_v(q_i) = \frac{1}{2} b q_i^2$ and study the choice of $b$. With this fee, larger trading volume is charged more fee. Given $b$, traders’ optimal order can be characterized similarly as before. Starting from the same linear conjecture, from (41), trader $i$’s best response is

$$q^*_i = \frac{E_i[v] - p - \frac{\rho}{\lambda} x_i}{\lambda + b + \frac{\rho}{\lambda}}. \quad \quad (43)$$

Since this fee structure does not distort traders’ learning process, we can still use $E_i[v] = k_1 s_i + k_2 x_i + k_3 p$ with $k_1, k_2, k_3$ in Lemma 2.

$$q^*_i = \frac{k_1 s_i - \left( \frac{\rho}{\lambda} - k_2 \right) x_i - (1 - k_3) p}{\lambda + b + \frac{\rho}{\lambda}}. \quad \quad (44)$$
This defines a fixed point problem

$$\beta_s = \frac{\tau_{\varepsilon}}{(\lambda + b)\tau + \rho}(1 - \varphi), \quad (45)$$

$$\beta_x = \frac{\rho}{(\lambda + b)\tau + \rho} \left( 1 - \varphi \frac{\tau_{\varepsilon} \beta_x}{\rho \beta_s} \right), \quad (46)$$

$$\beta_p = \frac{\tau}{(\lambda + b)\tau + \rho} \left( 1 - \varphi \frac{N\tau_{\varepsilon} \beta_p}{\tau \beta_s} \right), \quad (47)$$

where \( \lambda \equiv \frac{1}{(N-1)\beta_p} \) and \( \varphi = \frac{1}{1 + (\frac{\beta_p}{\beta_s})^2} \). From (45) through (47), it is clear that the ratios of coefficients are not affected, but traders trade less as if price impact increased by \( b \). This should be an expected result from the form of fee and the linearity of equilibrium.

**Lemma 13** Introduce tax of the form \( \phi_v(q_i) = \frac{1}{2}bq_i^2 \) with \( b > 0 \). (a) All the properties in Proposition 1 remain the same, except that the optimal order is rescaled by the factor \( \frac{\rho}{br+p} < 1 \). (b) Tax revenue is proportional to \( b(\frac{\rho}{br+p})^2 \), which is maximized by setting \( b = \frac{\rho}{\tau} \). (c) With \( b = \frac{\rho}{\tau} \), ICE profit is \( \Pi_{i,tax}^{ICE} = \lambda_{tax}\Pi_{i,H}^{ICE} + (1 - \lambda_{tax})\Pi_{i,FT}^{ICE} \), where \( \lambda_{tax} = \frac{(\lambda\tau + \rho)^2 + \rho^2}{(\lambda\tau + 2\rho)^2} = \frac{1 + \lambda^2}{2} \). For all \( N \), the weight \( \lambda_{tax} \) is larger than \( \tilde{\lambda} \) in Lemma 5.

Since higher tax is equivalent to increased price impact from traders’ perspective, traders trade less aggressively, thus decreasing trading volume. This trade-off between tax rate and size of tax base determines the revenue-maximizing tax level. With this tax level, \( \frac{\rho}{br+p} = \frac{1}{2} \). Hence, all the coefficients on the order decrease by half. This implies that price impact is twice as large and trading volume decreases by half compared to no-tax case.

Note that this tax level depends on the size of exchange \( N \). Therefore, this should be interpreted as the tax level which for-profit exchanges would choose after the exchange size is determined. This tax scheme may look vulnerable to competition among exchanges. However, if exchanges cannot commit to other tax schemes ex ante, traders always expect that exchanges would use this tax scheme for a given exchange size. Without such a commitment, the market formation game in the section 4 should be reconsidered by using a gain from trade
curve under the profit maximizing tax scheme.

**Lemma 13(c)** shows how the gain from trade curve is distorted by this tax scheme. Decomposition of ICE profit takes the same form as the one without tax, but a larger weight $\lambda_{tax}$ is put on the no-trade value. Thus, gain from trade curve is further distorted downward. This is an expected result given a formal equivalence between price impact and the tax scheme considered here.

Competition among exchanges at the ex ante stage affects the tax level chosen by for-profit exchanges. Since the revenue-maximizing tax level depends negatively on the exchange size, a monopoly exchange tends to have a lower tax. As more exchanges enter, each exchange size gets smaller and tax level increases. With free entry of exchanges, exchanges may set entry fee lower than a marginal cost, since now the cost can be recovered by taxing on trading activity. Therefore, a symmetric equilibrium would be that all exchanges operate at size $N^*$, set entry fee at the level lower than marginal cost of operation $c$, at which they can break even (in expectation) by using a tax scheme with $\bar{b} = \frac{\rho}{\tau_e + \tau_z + \tau_z (N^*-1) \phi}$.

The argument above suggests that a tax scheme can be used to implement different allocation. What tax scheme should a social planner choose? If exchanges can commit to different tax schemes than the one analyzed above, do they have an incentive to do so? With such a commitment ability, do for-profit exchanges implement more efficient allocation? As a first step to address these issues, consider the subsidy in stead of tax. I study the consequence of choosing negative $b < 0$, and in particular, whether the price-taking allocation can be implemented by this form of subsidy.

For the analysis that follows, I introduce a few more notations. Let $\lambda_{PT}$ be the price impact that would prevail under the price-taking assumption. From **Corollary 1**, we know that $\lambda_{PT} = \frac{\frac{\rho}{(N-1)\tau}}{1 + \frac{\phi(N-1)}{1-\phi}} < \frac{\rho}{\tau}$. Let $\lambda(b)$ be the price impact in the strategic case with tax/subsidy level $b$. The value of $\lambda(0)$ corresponds to the size of price impact without any tax/subsidy, and can be derived by plugging the value of $\beta_p$ in **Proposition 1** into $\lambda \equiv \frac{1}{(N-1)\beta_p}$.
First, assume small subsidy in the range \( b \in (-\frac{\tau}{\rho}, 0) \). Equilibrium characterization is same as the analysis following from (41) through (47), except that now we need to deal with a possible case where \( \lambda + b = 0 \). Since \( \lambda \) is determined in equilibrium for each \( b \), we use notation \( \lambda(b) \). Next, we conjecture that \( \lambda(b) + b > 0 \) for \( b \) close to zero, and verify this conjecture. Given that \( \lambda(b) + b > 0 \), all the analysis leading to Lemma 11 is valid, hence

\[
\lambda(b) = \frac{1}{(N-1)\beta_{\rho}} = \frac{b\tau + \rho}{(N-1)\tau} \frac{1 + \varphi(N-1)}{(1 - 2\varphi - \frac{1}{N-1})}.
\] (48)

Note that \( \lambda(b) \) is increasing in \( b \) linearly, and takes value zero at \( b = -\frac{\tau}{\rho} \). This implies that as subsidy size increases (\( b \) decreases), there is a unique level of subsidy \( b^* \in (-\frac{\tau}{\rho}, 0) \) such that \( \lambda(b^*) + b^* = 0 \). Therefore, as long as \( b \neq b^* \), conjecture is verified in equilibrium. What happens at \( b = b^* \)? The next lemma completes this argument by showing that setting \( b = b^* \) implements a price-taking allocation.

**Lemma 14 (a)** For \( b \in (-\frac{\tau}{\rho}, 0) \) and \( b \neq -\lambda_{PT} \), all the properties in Proposition 1 remain the same in the presence of subsidy except that the optimal order is rescaled by the factor \( \frac{\rho}{b\tau + \rho} > 1 \). Equilibrium price impact is given by (48). **(b)** There exists a unique \( b^* \in (-\frac{\tau}{\rho}, 0) \) such that \( \lambda(b^*) + b^* = 0 \). Price-taking allocation is implemented by setting \( b = b^* = -\lambda_{PT} \). **(c)** Under the subsidy \( b^* \), ICE profit is \( \Pi_{i,sub}^{ice} = \Pi_{i,PT}^{ice} + \frac{\tau}{\rho} \lambda_{PT}(\Pi_{i,PT}^{ice} - \Pi_{i,H}^{ice}) \), where \( \frac{\tau}{\rho} \lambda_{PT} = \frac{1 + \chi N}{N-1} \), which is smaller than 1.

**Lemma 14(b)** shows that larger subsidy increases trading volume. The effect of subsidy is just the opposite of tax. As subsidy increases, trading volume gradually increases. However, **Lemma 14(b)** shows that there exists a level of subsidy which drastically changes the nature of trading game. At this point, trading volume discontinuously jumps up. The reason for this odd behavior of the equilibrium is that this level of subsidy helps traders get out of prisoners’ dilemma situation. Price impact is a negative externality which each trader imposes one another. If everyone takes price as given, then everyone gets better off, but each has an incentive to take price impact into account when the others do not. In this sense, strategic
equilibrium is like a prisoner’s dilemma situation. **Lemma 14(b)** shows that this negative externality can be removed by using Pigouvian subsidy. This can be seen from (43). Negative externality emerges through $\lambda$ in the denominator. Taking $\lambda$ into account, each trader trades less amount, which imposes negative impact on other traders. Properly designed subsidy ($b^* = -\lambda_{PT}$) can undo this negative externality, making each trader’s private benefit and social benefit well aligned.

Under this subsidy, **Lemma 14(c)** shows that a trader’s ICE profit is a sum of benefit of having price-taking allocation and the value corresponding to subsidy, which is a fraction of gain from trade in price-taking case. This also shows that overall gain from trade including subsidy is larger than gain from trade in price-taking case ($\Pi_{i,sub}^{ice} - \Pi_{iH}^{ice}$ is scaled up by $1 + \frac{\tau}{\rho} \lambda_{PT}$ from $\Pi_{iPT}^{ice} - \Pi_{iH}^{ice}$).

Although there is a subsidy level that implements price-taking allocation, it is not clear if it is socially efficient to do so. Also, do for-profit exchanges have an incentive to provide subsidy, rather than taxing trading? Since the amount of total subsidy necessary for this arrangement is stochastic, risk aversion of subsidy providers matters. If those who organize exchanges are assumed to be risk-neutral, they only care the expected value of total subsidy, which can be shown to be $NE[\frac{1}{2} b^* q_i^2] = \frac{1}{2} \frac{\rho}{\tau + \tau_c} q_i^2 - \frac{\tau_c q_i}{\tau}$. This is increasing in $N$ and goes to $\frac{1}{2} \frac{\rho}{\tau + \tau_c}$ in the limit. Per trader expected subsidy is decreasing in $N$ and goes to zero in the limit. Therefore, exchange competition analysis should be modified by adding per trader expected subsidy to marginal cost $c$ (or equivalently by adjusting gain from trade curve by subtracting per trader expected subsidy). However, in general certainty equivalent value of total subsidy must be considered in a market formation game.

Whether price-taking allocation can be implemented by for-profit exchanges depends on a number of factors. Here I provide only a brief discussion. First, the adjusted gain from trade (gain from trade in **Lemma 14(c)** minus certainty equivalent value of per trader subsidy) must exceed the gain from trade curve without tax/subsidy. If exchanges are very risk averse, certainty equivalent value of total subsidy may be so large that adjusted gain from trade can
be lower than the gain from trade without subsidy. In this case, even a monopoly exchange has no incentive to use subsidy. Second, competition among exchanges must be limited. If exchanges cannot earn enough profit from fixed entry fee, subsidy in a trading game cannot be supported.

6 Conclusion

This paper presented a two-stage game framework to study the formation of exchanges. The second stage trading game captured the three aspects of financial markets: (i) strategic behavior, (ii) private knowledge regarding the uncertain return of assets, and (iii) private endowments. It was shown that gain from trade is hump-shaped in the number of traders. Each trader affects the trading game in two ways: he (i) increases the size of the risk sharing pool, and (ii) increases the informativeness of prices. The key for the hump-shape is that the second force decreases the benefit of the first force through the Hirshleifer Effect. In the small exchange, the first factor dominates and gain from trade increases in the number of traders. On the other hand, the Hirshleifer Effect becomes dominant in the large exchange and gain from trade decreases in the number of traders. In other words, traders face a trade-off between risk sharing and information sharing. Since the latter tends to decrease the former, either one must be sacrificed for the other.

The first stage marker formation game formalized the interaction among exchanges and traders. It was shown that the negative externality among traders due to the Hirshleifer Effect may enable exchanges to enjoy market powers. The increase in the number of exchanges makes each exchange size smaller and mitigates the negative externality. With free entry of exchanges, each trader’s gain from trade is maximized, because prices don’t reveal much information when each exchange is small. The prediction of the model is consistent with historical evidence that some traders have preferred less transparency, and that market structures have evolved in that direction. In particular, the model may provide a rationale
for the recent development of “dark pools”.

Another extension of the model is to incorporate heterogeneity among traders, such as different risk aversions, introduction of uninformed traders, and unequal access to some trading technology (for example, some traders can be restricted to use market orders). This will clarify who benefits from an exchange and who does not.

More generally, this raises the interesting issue of how different types of trading games attract different types of traders. As documented in Biais and Green (2007), trading did not just disappear when a centralized bond market declined—it migrated to the OTC market, where much less information is shared. Thus, the OTC market structure could also be an endogenous reaction to avoid welfare-reducing information sharing. To further investigate this issue, one could explore a unified framework in which different trading games can be compared. A mechanism design approach might be useful in exploring this possibility.

Finally, it would be interesting to apply this model in a macroeconomic context. While the current paper focused on the welfare of an exchange economy, the welfare implication for a production economy is not obvious. On the one hand, reduced risk sharing due to information sharing may slow down production and innovation. On the other hand, informative prices may guide better productive decisions. Another application of the model is to study a central bank’s market operations by adding the central bank as a trader in financial markets. In the model environment, a central bank can change not only the mean of return (interest rate) but also the variance of return. The current analysis suggests that this market operation may have a significant impact on the viability of financial markets. The nature of the central bank’s private information, asset position, and public announcement may be important in such an analysis. These are all left for future work.

APPENDIX A

Proof of Lemma 1
When each trader submits zero quantity for all price levels (i.e., \(q_i(p) = 0\)), any price can clear the market. Following the market-clearing rule, \(p = 0\) is announced. Given that other traders submit \(q_i(p) = 0\), each trader is indifferent between submitting \(q_i(p) = 0\) and any price-contingent order. If he submits a price-contingent order, it will only change the price so that his net trading is zero, and he still receives \(\pi_i = v x_i\). Also, each trader strictly prefers submitting \(q_i(p) = 0\) to any non-zero market order, because the latter deviation would give him negative infinite utility.\(^{18}\) \(E_i[v]\) and \(\tau\) are obtained by Bayes’ rule. Without trading, each trader’s interim utility is \(\pi_i = v x_i\). Following the market-clearing rule, \(\tau\) is announced. Given that \(x_i > 0\) and \(\tau\) is announced, each trader is indifferent between submitting \(q_i(p) = 0\) and any price-contingent order. If he submits a price-contingent order, it will only change the price so that his net trading is zero, and he still receives \(\pi_i = v x_i\). Also, each trader strictly prefers submitting \(q_i(p) = 0\) to any non-zero market order, because the latter deviation would give him negative infinite utility.

\[
E_i[v] = E_i[\exp(-\rho p x_i)],
\]

Using a characteristic function of normal distribution, \(\Pi_{iNT}^{\infty} = E_i[v] x_i - \frac{\rho}{2 \tau_v} x_i^2\). Write this in a matrix form

\[
\Pi_{iNT}^{\infty} = [s_i, x_i] A_{NT}[s_i, x_i]',
\]

where \(A_{NT} \equiv \begin{bmatrix} 0 & \frac{\tau_p}{2 \tau_v} \\ -\frac{\rho}{2 \tau_v} & 0 \end{bmatrix} \).\(^{19}\) To obtain ACE profit, apply the following formula: if \(X\) follows normal distribution \(\mathcal{N}(0, \Sigma)\), where \(\Sigma\) is \(n\)-by-\(n\) covariance matrix, 
\[
E[\exp(-\rho X'AX)] = \frac{1}{\sqrt{\det(I_n + 2\rho \Sigma A)}},
\]

where \(I_n\) denotes a \(n\)-by-\(n\) identity matrix. Substituting \(\Sigma_{NT} = \begin{bmatrix} \frac{1}{\tau_v} + \frac{1}{\tau_x} & 0 \\ 0 & \frac{1}{\tau_x} \end{bmatrix}\) and \(A_{NT}\) yields 
\[
\det(I_2 + 2\rho \Sigma_{NT} A_{NT}) = 1 - \frac{\rho^2}{\tau_v \tau_x}.
\]

**Proof of Lemma 2**

From the market-clearing condition, \(0 = \sum_{j \neq i} q_j + q_i = \beta_s \sum_{j \neq i} s_j - \beta_x \sum_{j \neq i} x_j - (N-1) \beta_p p + q_i\). Observe that \((s_i, x_i, p)\) provides the same amount of information about \(v\) as \((s_i, x_i, h_i)\) does, where 
\[
h_i \equiv \frac{(N-1) \beta_p p - q_i}{(N-1) \beta_s} = v + \frac{1}{N-1} \sum_{j \neq i} \varepsilon_j - \frac{\beta_x}{\beta_s} \frac{1}{N-1} \sum_{j \neq i} x_j.
\]

By Bayes’ rule, \(\tau = \tau_v + \tau_x + (\text{Var}[v|h_i])^{-1}\), where 
\[
(\text{Var}[v|h_i])^{-1} = (N-1) \left( \tau_{\varepsilon}^{-1} + \left( \frac{\beta_x}{\beta_s} \right) \tau_x^{-1} \right)^{-1} = (N-1) \tau_{\varepsilon} \phi.
\]

Plug this and 
\[
h_i = \frac{(N-1) \beta_p p - q_i}{(N-1) \beta_s} = \frac{N \beta_p}{(N-1) \beta_s} p - \frac{1}{N-1} s_i + \frac{\beta_x}{(N-1) \beta_s} x_i\]

into \(E_i[v] = \tau_v s_i + \frac{(\text{Var}[v|h_i])^{-1}}{\tau} h_i\) to get the last result.

\(^{18}\)The market-clearing rule has some bite here. If traders are not punished by infinite prices, there may be incentive to throw away positive endowments. This happens if \(x_i > 0\) and \(\frac{\tau_v}{\tau} s_i < x_i\). If an alternative market-clearing rule is used for this “free disposal” to be allowed in equilibrium, no-trade utility will be higher, and the level of gain from trade will be lower, but the main results in this paper remain unchanged.

\(^{19}\)Throughout the paper, a symmetric matrix is represented by its upper triangular part.
Proof of Proposition 1

(a) From \( \frac{\beta_s}{\beta_s} = \frac{1}{1 - \varphi \tau_x} - \frac{\varphi}{1 - \varphi \beta_s} \) \( K \equiv \frac{\beta_s}{\beta_s} = \frac{1}{1 - \varphi (K) \tau_x} = \frac{\varphi (K)}{1 - \varphi (K) \beta_s} \), where \( \varphi (K) \equiv \frac{1}{1 + K^2 \tau_x} \). This defines a cubic equation \( \{ K^2 + \frac{\tau_x}{\tau_x} \left( 1 + \frac{1}{(N-1) \beta_s} \right) \} \left( K - \frac{\rho}{\tau_x} \right) = 0 \) with a unique solution \( K = \frac{\rho}{\tau_x} \). It follows that \( \varphi = \frac{1}{1 + \rho^2 \tau_x} = \frac{\chi}{1 + \chi} \). Similarly, solving \( \frac{\beta_p}{\beta_s} = \frac{1}{1 - \varphi \beta_s} \) for \( \frac{\beta_p}{\beta_s} \) shows that \( \frac{\beta_p}{\beta_s} = \tau_\sigma + \varphi (N-1) \tau_x \). Define \( C_s \equiv corr(s_i, s_j) = \frac{\tau_x}{\tau_x + \rho} \) and \( N_s \equiv 1 + C_s (N-1) \). Note that \( \tau_\sigma + \varphi (N-1) \tau_x \), \( N_s \equiv 1 + C_s (N-1) \). The result for \( \frac{\beta_p}{\beta_s} \) obtained above.

(b) Use \( \beta_s, \beta_x \) and \( \beta_p \) obtained above.

(c) From a market-clearing condition, \( p = \frac{\beta_x}{\beta_p} \left( \bar{s} - \frac{\rho}{\tau_x} \bar{x} \right) \).

Then \( q_i = \beta_s \left\{ s_i - \bar{s} - \frac{\rho}{\tau_x} (x_i - \bar{x}) \right\} \). Use \( \frac{\beta_p}{\beta_s} \) and \( \beta_s \) above.

Proof of Corollary 2

The result for \( E_i[v] \) and \( \tau \) is trivial. A trader’s objective is \( E_i[v] (q_i + x_i) - \frac{\rho}{2 \tau_v} (q_i + x_i)^2 - pq_i = -\frac{\rho}{2 \tau_v} (q_i + x_i)^2 - pq_i \). After taking price impact into account, it can be shown that \( q_i = -\left(1 - \frac{1}{N-1}\right) x_i - \frac{\tau_p}{\rho} \left(1 - \frac{1}{N-1}\right) p \). This derivation is similar to the proof of Proposition 1, so it is omitted here. ICE profit is \( [x_i, p] A_{NS} [x_i, p] \), where \( A_{NS} \equiv \frac{1}{2(N-1)^2} \left[ -\frac{\rho}{\tau_v} \frac{(N-2)N}{\tau_v (N-2)N} \right] \).

Using \( A_{NS} \) and \( \Sigma_{NS} = \left[ \begin{array}{c} \frac{1}{\tau_x} - \frac{\rho}{\tau_x} \frac{N}{\tau_v} \\
\frac{\rho^2}{\tau_x} \frac{1}{\tau_v} \end{array} \right] \), obtain \( det(I_2 + 2 \rho \Sigma_{NS} A_{NS}) = 1 - \frac{\rho^2}{\tau_v \tau_x} K(N, \frac{\rho^2}{\tau_v \tau_x}) \), where \( K(N, \frac{\rho^2}{\tau_v \tau_x}) \equiv \frac{1}{N-1}(1 + \frac{\rho^2}{\tau_v \tau_x} \frac{N-2}{N}) \). Note that \( K(2, \frac{\rho^2}{\tau_v \tau_x}) = 1 \). It is straightforward to show that \( K(N, \frac{\rho^2}{\tau_v \tau_x}) \) is decreasing in \( N \). Finally, \( \lim_{N \to \infty} K(N, \frac{\rho^2}{\tau_v \tau_x}) = 0 \).
Proof of Corollary 3

(a) The result for $E_i[v]$ and $\tau$ is trivial. A trader’s objective is $E_i[v](q_i + x_i) - \frac{\rho}{2\tau} (q_i + x_i)^2 - pq_i = -\frac{\rho}{2\tau} (q_i + x_i)^2 - pq_i$. Conjecture $q_i = \tilde{\beta}_s s - \beta_x x_i - \beta_x p$. Proceeding similarly as before, obtain the best response with $\tilde{\beta}_s = \frac{N\tau}{\tau + \rho}, \beta_x = \frac{\rho}{\tau + \rho}$ and $\beta_p = \frac{\tau}{(N-1)\tau + \rho}$. This yields $\frac{\tilde{\beta}_s}{\beta_p} = \frac{N\tau}{\tau}, \beta_p = \frac{\rho}{\tau}$ and $\beta_p = \frac{\tau}{(N-1)\tau + \rho}$. The solution to the last provides $\beta_p = \frac{\tau}{\rho} (1 - \frac{1}{N-1})$ and hence $q_i = \frac{\tau s}{\rho} (1 - \frac{1}{N-1}) s - (1 - \frac{1}{N-1}) x_i - \frac{\tau}{\rho} (1 - \frac{1}{N-1}) p$.

(b) ICE profit is $[s, x_i, p]A_P[s, x_i, p]'$, where

$$A_P = \frac{1}{2p(N-1)^2} \begin{bmatrix} \frac{\tau^2}{N^3} (N-2) & \frac{\tau s}{N} & -\tau \xi N^2 (N-2) \\ -\frac{\rho^2}{\tau} & \rho N (N-2) & \tau N (N-2) \end{bmatrix}.$$  

Using $A_P$ and $\Sigma_P = \frac{1}{N-1} \begin{bmatrix} N-1 & 0 & \frac{N}{\tau_x} \frac{N-1}{N} \\ \frac{N}{\tau_x} & \frac{N-1}{\tau_x} & \frac{\rho}{\tau_x} \frac{N-1}{N} \end{bmatrix}$, obtain

$$I_3 + 2\rho \Sigma_P A_P = \begin{bmatrix} \frac{N}{N-1} \frac{\rho^2}{\tau} & \frac{N}{N-1} \frac{\rho}{\tau} & \frac{N}{N-1} \frac{\rho}{\tau} \\ \frac{\rho^2}{N-1} \tau_x & 1 - \frac{\rho^2}{(N-1)\tau_x} & -\frac{\rho^2}{N-1} \frac{\tau_x}{\tau} \\ \frac{\rho^2}{N-1} \frac{\tau_x}{\tau} & \frac{\rho}{(N-1)\tau_x} \left\{ \frac{N^2\tau_x (N-\tau_x)}{\tau_x} + \frac{\rho^2}{\tau_x} \frac{N-1}{N} \right\} & 1 \end{bmatrix}.$$  

Hence $\det(I_3 + 2\rho \Sigma_P A_P) = 1 - \frac{\rho^2}{\tau_x} \left( \frac{N}{N+\tau_x} \right)^2 - \left\{ \frac{\rho^2}{\tau_x} + \left( \frac{\rho}{\tau_x} \right)^2 \right\} \left( \frac{N}{N+\tau_x} \right)^2 \frac{1}{N-1} - \frac{\rho^2}{\tau_x} \left\{ \frac{1}{\tau} \frac{N}{N+\tau_x} \right\}^2 - \left( \frac{\rho}{\tau_x} \right)^2 \left\{ \frac{N-2}{\tau_x^2 (N\tau_x)} + \frac{\rho^2}{\tau_x^2} \frac{N-1}{N} \right\}.$

Note that by setting $\tau_x = 0$, this becomes $1 - \frac{\rho^2}{\tau_x} K(N, \frac{\rho^2}{\tau_x})$. For $\tau_x > 0$, this approaches $1 - \frac{\rho^2}{\tau_x} \tau_x$ as $N \to \infty$.

Proof of Lemma 3

Recall that $q_i = \beta_s \left\{ s_i - \bar{s} - \frac{\rho}{\tau_x} (x_i - \bar{x}) \right\} = \beta_s \left\{ \xi_i - \frac{\rho}{\tau_x} x_i - (\bar{s} - \frac{\rho}{\tau_x} \bar{x}) \right\}$ and $TV = \frac{\beta_s}{2} \sum \left| \xi_i - \frac{\rho}{\tau_x} x_i - (\bar{s} - \frac{\rho}{\tau_x} \bar{x}) \right|$. Since $\beta_s = \frac{\tau_x}{\rho} (1 - 2\varphi - \frac{1}{N-1})$ and $\varphi = \frac{\tau_x}{\rho}$, $\beta_s$ approaches
zero from above as $\chi < 1 - \frac{2}{N}$ approaches equality. Also, $\lambda = \frac{1}{(N-1)\tau_p} = \frac{\tau - \tau_v}{(N-1)\tau_s}$ goes to infinity.

**Proof of Proposition 2**

(a) Recall that $p = \frac{\beta_s}{\beta_p} \left( \frac{s}{\tau_s} - \frac{\rho}{\tau_v} \right)$. Hence, $\frac{\beta_p}{\beta_s} p = v + \bar{z} - \frac{\rho}{\tau_v} \bar{z}$ is an unbiased signal of $v$ with precision $N\tau_v \varphi$. By Bayes rule, $E[v|p] = \frac{N\tau_v \varphi}{\tau_v + N\tau_v \varphi} \beta_p p$, where $\frac{\beta_p}{\beta_s} = \frac{\tau}{\tau_s + \varphi(N-1)\tau_v}$ from the proof of Proposition 1. Substituting $\tau = \tau_v + \tau_v (1 + (N-1) \varphi)$ and simplifying it yields $\zeta = \frac{\tau_v + \tau_v (1 + (N-1) \varphi)}{\tau_v + \tau_v (1 + (N-1) \varphi)} < 1$. With $\varphi = \frac{N\tau_v \varphi}{\tau_v + N\tau_v \varphi} = \left(1 + \frac{1-C_s}{C_s} \frac{1+\chi}{N \chi}\right)^{-1} = \frac{C_s N \chi}{1 + \chi + C_s (N \chi - 1 \chi)} = \frac{C_s N \chi}{1 - C_s + N \chi}$, Recall that $\frac{\tau}{\tau_v + \varphi(N-1)\tau_v} = \frac{1 + N \chi}{C_s + N \chi}$. Hence, $E[v|p] = \frac{N \chi}{1 + N \chi} \frac{1 + N \chi}{1 - C_s + N \chi} p$. Finally, 

$$\lim_{N \to \infty} \frac{N \chi}{1 + N \chi} \frac{1 + N \chi}{1 - C_s + N \chi} = 1.$$

(b) From the proof of Lemma 3, $\frac{T_X}{N} = \frac{\beta_s}{2N} \sum \left| \varepsilon_i - \frac{\rho}{\tau_v} x_i - \frac{1}{N} \sum \left( \varepsilon_i - \frac{\rho}{\tau_v} x_i \right) \right|$. Note that 

$$\varepsilon_i - \frac{\rho}{\tau_v} x_i - \frac{1}{N} \sum \left( \varepsilon_i - \frac{\rho}{\tau_v} x_i \right) = \frac{N-1}{N} \left( \varepsilon_i - \frac{\rho}{\tau_v} x_i \right) - \frac{1}{N} \sum \left( \varepsilon_j - \frac{\rho}{\tau_v} x_j \right).$$

Hence, $E\left[\frac{T_X}{N}\right] = \frac{\beta_s}{2N} \frac{N-1}{N} \sum \left[ \left( \varepsilon_i - \frac{\rho}{\tau_v} x_i \right) - \frac{1}{N} \sum \left( \varepsilon_j - \frac{\rho}{\tau_v} x_j \right) \right] = \frac{\beta_s}{2N} \frac{N-1}{N} \sum \left( \varepsilon_j - \frac{\rho}{\tau_v} x_j \right).$ Here, 

$$Var \left[ \left( \varepsilon_i - \frac{\rho}{\tau_v} x_i \right) - \frac{1}{N} \sum \left( \varepsilon_j - \frac{\rho}{\tau_v} x_j \right) \right] = \frac{N}{N-1} \left( \frac{\rho^2}{\tau_v^2} \left( \frac{2}{\pi} \right)^2 \right) = \frac{\rho^2}{\tau_v^2} \left( \frac{2}{\pi} \right)^2 \left( \frac{N}{N-1} \right) \left( 1 + \chi \right).$$

Apply the following fact: $E[|X|] = \sqrt{\frac{2}{\pi} \frac{1}{\tau_X}}$, where $X$ is a normal random variable with mean zero and variance $\tau_X^{-1}$, to obtain $E \left[ \left( \varepsilon_i - \frac{\rho}{\tau_v} x_i \right) - \frac{1}{N} \sum \left( \varepsilon_j - \frac{\rho}{\tau_v} x_j \right) \right] = \frac{\rho}{\tau_v} \sqrt{\frac{2}{\pi} \frac{N}{N-1} \frac{1}{\tau_s}} \left( 1 + \chi \right).$

Substituting $\beta_s = \frac{\tau}{\rho} \left(1 - 2\varphi - \frac{1}{N-1}\right)$ and $1 - 2\varphi - \frac{1}{N-1} = \frac{(1-\chi)N-2}{(1+\chi)(N-1)}$ yields 

$$E \left[ \frac{T_X}{N} \right] = \frac{(1-\chi)N-2}{\sqrt{N(N-1)2\pi\tau_x(1+\chi)}} \frac{1}{\sqrt{2\pi\tau_x(1+\chi)}}.$$

This is increasing in $N$ and 

$$\lim_{N \to \infty} \frac{(1-\chi)N-2}{\sqrt{N(N-1)2\pi\tau_x(1+\chi)}} = \frac{1}{\sqrt{2\pi\tau_x(1+\chi)}}.$$

(c) Write ICE profit given in Lemma 5 in a matrix form. Let $X_i \equiv [s_i, x_i, p]'$. ICE profit of holding endowment is $\Pi_{iH}^{ice} = X_i' A_H X_i$ and that of price-taking case is $\Pi_{iPT}^{ice} = X_i' A_{PT} X_i$.

$$A_H \equiv \begin{bmatrix} C_s & 0 & 0 \\ \frac{-\rho}{\tau_v} (1 - I) & \frac{N \chi}{C_s} + N \chi & \frac{N \chi}{C_s} + N \chi \end{bmatrix}, \quad A_{PT} \equiv \begin{bmatrix} \frac{C_s}{2(1+N\chi)} & \frac{1 - \rho}{\tau_v} \frac{1+N\chi}{1+N\chi} \frac{1}{C_s} \\ \frac{-\rho}{\tau_v} (1 - I) & \frac{N \chi}{C_s} + N \chi & \frac{N \chi}{C_s} + N \chi \end{bmatrix}$$

and

59
\[ \Sigma = \begin{bmatrix} \frac{1}{\tau_v C_s} & 0 & \frac{1}{\tau_v N_s^{1+N_N}} \\ \frac{1}{\tau_x} & \frac{1}{\tau_x} & \frac{1}{\tau_x N_s^{1+N_N}} \\ -\frac{\rho}{\tau_x \tau_v} C_s & -\frac{\rho}{\tau_x \tau_v} N_s^{1+N_N} C_s & \left( \frac{1+N_N}{N+1} \right) \left( C_s^{1+N_N} \right)^2 \end{bmatrix} \]. First, \( A_H \) is derived by evaluating coefficients of \( \Pi_{iH}^{ice} = E_i[v|x] = \frac{\tau_v (1-\rho)}{\tau_x} s_i x_i + \frac{\tau_x}{\tau_v} x_i^2 \). Derivation of the coefficient on \( s_i x_i \) as given in the proof of \textbf{Proposition 1}: \( \frac{\tau_v (1-\rho)}{\tau_x} = \frac{C_s}{1+N_N} \). The other coefficients can be evaluated similarly. Next, \( A_{PT} \) is derived by evaluating coefficients of \( \Pi_{iPT}^{ice} = px_i + \frac{\tau}{2\rho} (E_i[v]-p)^2 = \frac{\tau}{2\rho} k_1^2 s_i^2 + \frac{\tau}{\rho} k_1 k_2 s_i x_i + \frac{\tau}{\rho} k_1 (k_3-1) s_i p + \frac{\tau}{2\rho} k_2^2 x_i^2 + \{ \frac{\tau}{\rho} k_2 (k_3-1) + 1 \} x_i p + \frac{\tau}{2\rho} (k_3-1)^2 p^2 \). Substitute \( k_1, k_2, k_3 \) in \textbf{Lemma 2} and algebra yields the result.

Finally, \( \Sigma_{11} = Var[s_i], \Sigma_{22} = Var[x_i], \Sigma_{33} = Var[p], \Sigma_{12} = cov[s_i, x_i], \Sigma_{13} = cov[s_i, p], \Sigma_{23} = cov[x_i, p] \). \( \Sigma_{11}, \Sigma_{12}, \Sigma_{22} \) are trivial. I show the derivation of \( \Sigma_{13} \). \( \Sigma_{23} \) and \( \Sigma_{33} \) are evaluated similarly. \( cov[s_i, p] = \frac{\beta_p}{N\beta_p} (Var[s_i] + (N-1) Cov[s_i, s_j]) = \frac{\beta_p}{N\beta_p} (\tau_v^{-1} + \frac{1}{N} \tau_e^{-1}) \). From the proof of \textbf{Proposition 1}, \( \frac{\beta_p}{\beta_p} = \frac{C_s}{1+N_N} \). Since \( \left( \frac{1}{\tau_v} \right)^{-1} = \frac{1}{N \tau_v} \left( \frac{1}{C_s} + N - 1 \right) = \frac{N}{C_s \tau_v N} \), the desired result is obtained.

ICE profit is a weighted sum of these two, \( \Pi_{i}^{ice} = \lambda X'(A_H + (1-\lambda) A_T) X_i \), where \( \lambda \) is given in \textbf{Lemma 5}. Finally, it is straightforward to show that \( \lim_{N \to \infty} A_H = \lim_{N \to \infty} A_{PT} = \begin{bmatrix} 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \\ 0 & 0 & 1/2 \end{bmatrix} \) and \( \lim_{N \to \infty} \Sigma = \begin{bmatrix} 1/\tau_v C_s & 0 & 1/\tau_v \\ 0 & 1/\tau_x & 0 \\ 1/\tau_v & 0 & 1/\tau_x \end{bmatrix} \). Also, \( \lim_{N \to \infty} \lambda = \chi^2 \). Calculating \( \Pi_{i}^{ace} = \frac{1}{2\rho} \log (\det(\Sigma \det(\Sigma^{-1} + 2\rho A))) \) with these matrices shows that \( E \left[ \exp \left( -\rho \lim_{N \to \infty} \Pi_{i}^{ace} \right) \right] = \exp (\rho \Pi_{i}^{ace} \left( \frac{N}{N_T} \right) ) \). By dominated convergence theorem, the result holds.

**Proof of Lemma 4**

First, \( v - p = (1 - \frac{\beta_p}{\beta_p}) v - \frac{\beta_p}{\beta_p} \left( \tau - \frac{\rho}{\tau} \tau \right) \). Hence, \( Var[v-p] = \left( 1 - \frac{\beta_p}{\beta_p} \right)^2 \tau v^{-1} + \frac{1}{N} \left( \frac{\beta_p}{\beta_p} \right)^2 \left( \tau v^{-1} + (\frac{\rho}{\tau} \tau)^{-1} \right) \). From \textbf{Proposition 1}, \( \frac{\beta_p}{\beta_p} - 1 = \frac{C_s}{C_v} \frac{1+N_N}{1+N_N} - 1 = \frac{1-C_s}{C_v} \frac{1+N_N}{1+N_N} = \frac{\tau_v}{\tau_x} \frac{1+N_N}{1+N_N} \). This

60
establishes \( \text{Corr}[v,v-p] = \frac{1}{\sqrt{1 + \frac{\chi}{\tau}} H(N)} \), where \( H(N) \equiv \frac{(1+\chi N)^2}{(1+\chi N)^p} \). Correlation is monotonically and inversely related to \( H(N) \). The first part follows form the fact that

\[
H'(N) = \frac{1}{1+\chi} \frac{2(1+\chi N)\chi N-(1+\chi N)^2}{N^2} = \frac{1}{1+\chi} \frac{(\chi N)^2-1}{N^2} \text{ and } \lim_{N \to \infty} H'(N) = \frac{\chi^2}{1+\chi}. \]

Finally, \( \frac{\partial H}{\partial \chi} = \frac{(1+\chi N)(2+\chi-\frac{1}{N})}{(1+\chi)^2} > 0. \)

**Proof of Lemma 5**

First, \( \Pi_{i}^{\text{inc}} = (E_{i}[v] - p)(q_{i} + x_{i}) - \frac{\rho}{\tau}(q_{i} + x_{i})^2 + px_{i} \). Substitute \( q_{i} = \frac{\tau}{\lambda \tau + \rho} (E_{i}[v] - p - \frac{\rho}{\tau}x_{i}) \) from (10) and simplify it to obtain the result. To show \( \frac{\lambda \tau}{\lambda \tau + \rho} = \frac{\rho}{\tau + \rho (N-1) \tau \varepsilon} \) from the definition of \( \lambda \). Next, \( \beta_{s} = \frac{\tau}{\rho} (1 - 2 \rho - \frac{1}{N-1}) \) from **Proposition 1**. This implies \( \rho \beta_{s} = \frac{\tau (1 - 2 \rho - \frac{N-1}{N})}{1 + \rho (N-1)} \). Hence,

\[
\frac{\lambda \tau}{\lambda \tau + \rho} = \frac{\rho}{\frac{1}{1+\varphi(N-1)} \frac{1+\varphi(N-1)}{1+\varphi(N-1)}} = \frac{1+\varphi(N-1)}{1+\varphi(N-1)+(1-2\varphi)(N-1)-1} = \frac{1+\varphi(N-1)}{(N-1)(1-\varphi)} = \frac{\varphi+\frac{1}{N-1}}{1-\varphi} = \chi + \frac{1+\chi}{N-1} = \frac{N}{N-1} \chi + \frac{1}{N-1}.
\]

**Proof of Proposition 3**

(a) From **Corollary 2**, \( GFT \) is monotonically increasing in \( N \) if \( \chi = 0 \). Note that as \( N \) increases, the only factor that can decrease ex ante utility is informative price. If \( GFT \) were not uni-modal, there must be some range of \( N \) where \( \tau \) is decreasing in \( N \). This cannot be the case in equilibrium with \( \chi > 0 \), since \( \tau = \tau_{v} + \tau_{\varepsilon} + \varphi(N-1) \tau_{\varepsilon} \). This establishes uni-modal shape of \( GFT \). From **Lemma 5**, larger \( \chi \) increases \( \tilde{\lambda} = \left( \frac{N}{N-1} \chi + \frac{1}{N-1} \right)^2 \). As \( \chi \) increases, higher \( \tilde{\lambda} \) lowers \( GFT \) for given \( N \).

(b) \( \chi \) affects shape of \( GFT \) through two channels. First, as \( \chi \) increases, higher \( \tilde{\lambda} \) shifts hump-shape to the right, thus tends to make \( N^* \) larger. Second, from **Lemma 4**, \( \text{Corr}[v,v-p] \) is maximized at \( N = \frac{1}{\chi} \). Thus, larger \( \chi \) shifts the peak of hump shape for price-taking case to the left. For small \( \chi \), the second factor \( (N = \frac{1}{\chi}) \) is sensitive to the change in \( \chi \), while adjustment through the first factor \( (\tilde{\lambda} = \left( \frac{N}{N-1} \chi + \frac{1}{N-1} \right)^2 \) becomes negligible. Hence \( N^* \) is decreasing in \( \chi \) for small \( \chi \). Recall that lower bound \( \frac{2}{1-\chi} \) in (15) are increasing in \( \chi \). As \( \chi \) approaches \( 1 \) from below, \( \frac{2}{1-\chi} \) goes to infinity and so must do \( N^*(\chi) \).
(c) To show the convergence speed of GFT, it suffices to show that \(\text{det}(I_3 + 2\rho \Sigma A_{PT}) = 1 - \frac{\rho^2}{\tau_v \tau_x} + o(\frac{1}{N})\), where \(o(\frac{1}{N})\) denotes the terms converging to zero at the rate not slower than \(\frac{1}{N}\).

Element-by-element evaluation of \(\Sigma A_{PT}\) yields \(\Sigma A_{PT} = \begin{bmatrix}
o(\frac{1}{N}) & c_{12} + o(\frac{1}{N}) & -o(\frac{1}{N}) 
o(\frac{1}{N}) & o(\frac{1}{N}) & c_{23} - o(\frac{1}{N}) 
o(-o(\frac{1}{N}) & c_{32} + o(\frac{1}{N}) & -o(\frac{1}{N})
\end{bmatrix}\).

I show the derivation of (2,3) and (3,2) elements. The other elements can be derived similarly:

\[
(\Sigma A_{PT})_{(2,3)} = \frac{\varphi}{2} \frac{C_\varphi}{1+N_\varphi \chi C_\chi} \left(\frac{1+N_\varphi N_\chi}{1+N_\chi} \right) = \frac{\varphi}{2} \frac{C_\varphi}{1+N_\varphi \chi C_\chi} \left(\frac{1+N_\varphi N_\chi}{1+N_\chi} \right)
\]

\[
(\Sigma A_{PT})_{(3,2)} = \frac{\varphi}{2} \frac{C_\varphi}{1+N_\varphi \chi C_\chi} \left(\frac{1+N_\varphi N_\chi}{1+N_\chi} \right) = \frac{\varphi}{2} \frac{C_\varphi}{1+N_\varphi \chi C_\chi} \left(\frac{1+N_\varphi N_\chi}{1+N_\chi} \right)
\]

\[
= \left(1 + \frac{1+N_\varphi N_\chi}{1+N_\varphi \chi}\right) \left(\frac{C_\varphi^2}{2\tau_v} + \frac{1-C_\varphi}{\tau_v} + \frac{C_\varphi}{\tau_v (1+N_\chi) N(1+N_\chi)}\right)
\]

\[
= \left(1 + \frac{1+N_\varphi N_\chi}{1+N_\varphi \chi}\right) \left(\frac{C_\varphi^2}{2\tau_v} + \frac{1-C_\varphi}{\tau_v} + \frac{C_\varphi}{\tau_v (1+N_\chi) N(1+N_\chi)}\right)
\]

\[
= \frac{1}{2\tau_v} + o(\frac{1}{N}).
\]

Hence, \(\text{det}(I_3 + 2\rho \Sigma A_{PT}) = 1 - \frac{\rho^2}{\tau_v \tau_x} + o(\frac{1}{N})\).

**Proof of Lemma 6**

(a) If other traders do not join any exchanges, there is no incentive to join any exchanges.

(b) Consider second stage equilibrium \(\sum_{i=1}^{N} r_i^* = (N_A, 0)\). First, \(GFT(N_A) - \phi_A \geq 0\) implies \(GFT(N^*) - \phi \geq 0\). Also, \(N_A - \phi \leq N_A \leq N\). Conversely, given \(GFT(N^*) - \phi \geq 0\) and \(N_A - \phi \leq N\), there exists equilibrium where only one exchange which offers \(\phi\) is active.

(c) Suppose \(GFT(N_k) < GFT(N_k + 1)\) for all \(k\). Then \(0 \leq GFT(N_A) - \phi_A < GFT(N_A + 1) - \phi_A \leq GFT(N_B) - \phi_B\). But since \(GFT(N_B) - \phi_B \leq GFT(N_B + 1) - \phi_B\), this contradicts \(GFT(N_B + 1) - \phi_B \leq GFT(N_A) - \phi_A\).

**Proof of Lemma 7**

(a) This is trivial.

(b) Given \(\phi_k \in [c, \phi_m]\), playing one-exchange equilibrium provides non negative payoff for traders. Setting \(\phi_k \in [c, \phi_m]\) gives non negative profit for this exchange. Hence there is no incentive to deviate which only invokes zero profit.
(c) Both exchanges setting $\phi_k > GFT(N^*)$ cannot be equilibrium since there is profitable deviation by setting $\phi_k \in (c, GFT(N^*)]$. Both setting $\phi_k \in (c, GFT(N^*)]$ cannot be equilibrium since there is profitable deviation by setting a fee that is arbitrarily close to, but lower than the other exchange and higher than $c$. When both setting $\phi_k = c$, there is no incentive to change fee. Given $\phi_A = \phi_B = c$, if $N_c > \overline{N}$, all traders have positive surplus by following specified strategy. If $N_c \leq \overline{N}$, all traders have zero surplus by following specified strategy.

**Proof of Lemma 8**

Case 1 is trivial. In Case 2, from the first order condition derived in the main text, obtain $\phi = c - GFT'\left(\frac{\overline{N}}{2}\right)\overline{N}$. If this fee is greater than $GFT(\frac{\overline{N}}{2})$, traders’ participation constraint binds and $\phi = GFT(\frac{\overline{N}}{2})$, and each exchange extracts all the surplus. This is the case if $GFT(\frac{\overline{N}}{2}) \leq c - GFT'(\frac{\overline{N}}{2})\overline{N} \iff \frac{\overline{N}}{2} \geq N_2$. If $\phi = c - GFT'(\frac{\overline{N}}{2})\overline{N} < GFT(\frac{\overline{N}}{2})$, total surplus $GFT(\frac{\overline{N}}{2})\overline{N}$ is split between the exchanges and traders, since $\phi \in [c, GFT(\frac{\overline{N}}{2}))$. In Case 3, two exchanges with the same size would have to operate at the size where gain from trade is increasing in the number of traders. This cannot be equilibrium from Lemma 6.

**Proof of Proposition 4**

Focus on the case $N^* \leq \frac{\overline{N}}{K} < N_m$. Similarly to the two exchanges case, a typical exchange’s first order condition, combined with symmetry, yields $2GFT'(\frac{\overline{N}}{K})\frac{\overline{N}}{K} + \phi - c = 0 \iff \phi = c - 2GFT'(\frac{\overline{N}}{K})\frac{\overline{N}}{K}$. If this is greater than $GFT(\frac{\overline{N}}{K})$, traders’ participation constraint binds, exchanges set $\phi = GFT(\frac{\overline{N}}{K})$ and extract all the surplus. If $\phi = c - 2GFT'(\frac{\overline{N}}{K})\frac{\overline{N}}{K} < GFT(\frac{\overline{N}}{K})$, total surplus $GFT(\frac{\overline{N}}{K})\overline{N}$ is split between $K$ exchanges and traders, since $\phi \in [c, GFT(\frac{\overline{N}}{K}))$. Profit is zero only if $K = \frac{\overline{N}}{N_c}$. When $K$ exchanges equally split $\overline{N}$ traders, total surplus is given by $K \left\{ GFT\left(\frac{\overline{N}}{K}\right) - c \right\} \frac{\overline{N}}{K}$. Maximize this with respect to $K$: $\max K \left\{ GFT\left(\frac{\overline{N}}{K}\right) - c \right\} \frac{\overline{N}}{K}$. This is equivalent to $\max GFT\left(\frac{\overline{N}}{K}\right)$. Hence, $\frac{\overline{N}}{K} = N^* \iff K = \frac{\overline{N}}{N_c}$. Maximized total surplus is $\{GFT(N^*) - c\}\overline{N}$. 

63
Proof of Lemma 9

If \( mK \) agents out of \( N \) agents choose to organize exchanges, the total number of traders is \( N - mK \). From proof of Proposition 4, exchanges’ pricing in a symmetric equilibrium is given by \( \phi = c - 2GFT'\left(\frac{N-mK}{K}\right)\frac{N-mK}{K} \). In a free entry without occupational choice, \( \phi - c = 0 \) determined exchange size. This cannot be the case with occupational choice since those who earn zero profit by organizing exchanges would have incentive to become a trader.

As \( K \) decreases from \( \frac{N}{N+m} \), each exchange’s profit \( (\phi - c)\frac{N-mK}{K} = -2GFT'\left(\frac{N-mK}{K}\right)\left(\frac{N-mK}{K}\right)^2 \) increases, since \( GFT'\left(\frac{N-mK}{K}\right) \) becomes smaller and \( \left(\frac{N-mK}{K}\right)^2 \) becomes larger. This should continue until exchange’s profit equals each trader’s surplus \( GFT\left(\frac{N-mK}{K}\right) - \phi \). Therefore, the number of exchanges should adjust such that there is no incentive to switch occupation: \( GFT\left(\frac{N-mK}{K}\right) - \phi = (\phi - c)\frac{N-mK}{K} \). Combined with pricing equation, \( GFT(x) - c = -2GFT'(x)x(1+x) \) characterizes exchange size \( x = \frac{N-mK}{K} \), and \( K = \frac{N}{x+m} \).

Proof of Lemma 10

From the market-clearing condition,

\[
0 = \sum_{j \neq i} q_j + q_i = \beta_s \sum_{j \neq i} s_j - \beta_c \sum_{j \neq i} x_j - \beta_c (N-1) \eta - (N-1) \beta_p p + q_i.
\]

Observing \((s_i, e_i, p)\) provides the same amount of information about \( v \) as observing \((s_i, e_i, h_i)\), where \( h_i \equiv \frac{(N-1)\beta_p q_i}{(N-1)\beta_s} + \frac{\beta_c}{\beta_s} \frac{\tau_x}{\tau_x + \tau_q} e_i = v + \frac{1}{N-1} \sum_{j \neq i} \varepsilon_j - \frac{\beta_c}{\beta_s} \frac{1}{N-1} \sum_{j \neq i} x_j - \frac{\beta_c}{\beta_s} \left( \frac{\tau_s}{\tau_s + \tau_q} \eta - \frac{\tau_x}{\tau_x + \tau_q} x_i \right) \). By Bayes’ rule,

\[
\tau = \tau_v + \tau_e + \left(Var\left[v|h_i\right]\right)^{-1}, \text{ where } (Var\left[v|h_i\right])^{-1} = (N-1) \left[ \frac{\tau_x}{\tau_x + \tau_q} \right]^{-1} + \left( \frac{\beta_c}{\beta_s} \frac{\tau_x}{\tau_x + \tau_q} \right)^2 \left( \frac{\tau_x}{\tau_x + \tau_q} \right)^{N-1} = (N-1) \tau_x \varphi_c. \]

Plug this and \( h_i = \frac{(N-1)\beta_p q_i}{(N-1)\beta_s} + \frac{\beta_c}{\beta_s} \frac{\tau_x}{\tau_x + \tau_q} e_i = \frac{N\beta_p}{(N-1)\beta_s} p - \frac{1}{N-1} s_i + \frac{\beta_c(1+(N-1)C_e)}{(N-1)\beta_s} e_i \) into \( E_i[v] = \frac{\tau_x}{\tau} s_i + \frac{(Var[v|h_i])^{-1}}{\tau} h_i \) to get the last result.

Proof of Lemma 11

64
(a) From (37) and (38), \( \frac{\beta}{\beta_e} = \frac{\rho - \varphi_e \tau_e}{\tau_e (1 - \varphi_e)} \). Denoting \( x \equiv \frac{\beta}{\beta_e} \), \((1 - \varphi_e)x = \frac{\rho}{\tau_e} - \varphi_e \tau_e N_c x \).

Plug in \( \varphi_c = \frac{1}{1 + \frac{1}{3} \frac{N_c}{N_e} x^2} \) and simplify it to obtain (40). Since solutions to (40) lie in \((0, \frac{\rho}{\tau_e})\), normalize \( x = \frac{\delta}{\tau_e} \) with \( \delta \in (0, 1) \). Then \( G(\delta \frac{\rho}{\tau_e}) = 0 \Leftrightarrow \tilde{G}(\delta) = \delta^3 - \delta^2 + \chi \delta - \frac{x}{N_c} = 0 \). As \( N \to \infty \), this becomes \( \delta(\delta^2 - \delta + \chi) = 0 \). Thus, there are three real roots \( \delta \in \{0, \frac{1 + \sqrt{1 - 4\chi}}{2}, \frac{1 - \sqrt{1 - 4\chi}}{2}\} \) if \( 1 - 4\chi > 0 \). For any positive \( \delta \), \( \varphi_c \) goes to zero as \( N \to \infty \), so the second order condition \( \varphi_c < \frac{1}{2} \left( 1 - \frac{1}{N-1} \right) \) is satisfied. For \( \delta = 0 \), \( \varphi_c = 1 \) for any \( N \). So the second order condition cannot be satisfied in the limit.

(b) Combine \( \frac{\beta}{\beta_e} = \delta \frac{\rho}{\tau_e} = \frac{\rho}{\tau_e} \frac{1 + \sqrt{1 - 4\chi}}{2} \), \( \beta_s = \frac{\tau_e}{\rho} (1 - 2 \varphi_c - \frac{1}{N-1}) \) and \( \beta_p = \frac{\tau_e}{\tau-e} \beta_s \) to obtain the optimal order. Since \( \varphi_c = \frac{1}{1 + \left( \frac{\varphi_c}{\beta_e} \right)^2 N_c} \), \( \tau = \tau_v + \tau_e + \tau_c \frac{N-1}{2 \chi^2 N_c} \to \tau_v + \tau_e + \frac{\tau_e}{2 \chi^2 (1 - 2 \chi \pm \sqrt{1 - 4\chi})} \).

(c) As in the proof of Proposition 1, use

\[ p = \frac{\beta}{\beta_e} \left( \overline{s} - \frac{\rho}{\tau_e} \overline{x} \right) \quad \text{and} \quad q_i = \beta_s \left\{ s_i - \overline{s} - \frac{\rho}{\tau_e} (x_i - \overline{x}) \right\} . \]

Proof of Lemma 12

(a) The roots of \( \tilde{G}(\delta) \) is characterized by the sign of \( \Delta_\delta \equiv -4 \chi^3 + \chi^2 - 4 \frac{N}{N_c} + 18 \frac{N^2}{N_c} - 27 \left( \frac{\chi}{N_c} \right)^2 = -\frac{N^2}{N_c} \left\{ 4 N^2 \chi^2 - (N_c^2 + 18 N_c - 27) \chi + 4 N_c \right\} \). This is positive if \( \chi \in (\chi^-, \chi^+) \) with \( \chi^\pm = \frac{N^2 + 18 N_c - 27 \pm \sqrt{(N^2 + 18 N_c - 27)^2 - 64 N^3}}{8 N^2} \).

(b) \( \Phi_{N_c} \neq \emptyset \Leftrightarrow (N_c^2 + 18 N_c - 27)^2 - 64 N_c^3 = (N_c - 1)(N_c - 9)^3 > 0 \Leftrightarrow N_c > 9 \).

(c) If \( \Delta_\delta = 0 \Leftrightarrow N_c \geq 9 \) and \( \chi = \chi^- \) or \( \chi = \chi^+ \), then there exist two positive real roots. This cannot collapse to one root since the condition for three identical real roots is

\[ -2(-1)^3 + 9(-1)\chi - 27(-\frac{\chi}{N_c}) = 0 \Leftrightarrow N_c = \frac{27\chi}{9\chi-2}, \]

but this cannot be the case since \( \frac{27\chi}{9\chi-2} \leq 3 \).

(d) This is the remaining case.

(e) From the second order condition, \( 1 - 2 \varphi_c - \frac{1}{N-1} > 0 \Leftrightarrow \varphi_c = \frac{1}{1 + \frac{1}{\chi} \frac{N}{N_c} \delta^2} < \frac{1}{2} \left( 1 - \frac{1}{N-1} \right) \). This is equivalent to \( \frac{N-1}{N^2} < 1 + \frac{1}{\chi} \delta^2 N_c \). If \( \delta \) is a solution to (40), \( \tilde{G}(\delta) = 0 \Leftrightarrow N_c (1 - \delta) (\delta^2 + \chi) = \chi (N_c - 1) \Leftrightarrow N_c (\frac{\delta}{\chi} + 1) = \frac{1}{1 - \delta} (N_c - 1) \)

\[ \Leftrightarrow 1 + \frac{1}{\chi} \delta^2 N_c = \left( \frac{1}{1 - \delta} - 1 \right) (N_c - 1) = \frac{\delta}{1 - \delta} (N_c - 1) . \] Hence, \( \frac{N-1}{N^2} < \frac{\delta}{1 - \delta} (N_c - 1) \Leftrightarrow \delta > \{ 1 + \frac{C_c(N-2)}{2} \}^{-1} \). This implies \( x \in \left( \frac{\rho}{\tau_e \left[ 1 + \frac{C_c(N-2)}{2} \right]} \frac{1}{2} \right)^{-1} \).
Proof of Lemma 13

(a) The first result is immediately obtained by replacing $\lambda$ with $\lambda + b$ in the proof of Proposition 1.

(b) Tax revenue is $\frac{1}{2}b \sum q_i^2 = \frac{1}{2}b\beta^2 \sum \left\{ s_i - \bar{x} - \frac{b}{\tau} (x_i - \bar{x}) \right\}^2$, where $b\beta^2 = b\left(\frac{\tau}{br+\rho}\right)^2 (1 - 2\varphi - \frac{1}{N-1})^2$. It is easy to verify that $\frac{b}{(br+\rho)^2}$ is maximized by setting $b = \frac{\rho}{\tau}$.

(c) Substitute $q_i = \frac{\tau}{(\lambda+b)\tau+\rho}(E_i[v] - p - \frac{\rho}{\tau}x_i) = \frac{\tau}{\lambda\tau+2\rho}(E_i[v] - p - \frac{\rho}{\tau}x_i)$ in $\Pi = (E_i[v] - p)(q_i + x_i) - \frac{\rho}{2\tau}(q_i + x_i)^2 + px_i - \frac{1}{2}E_i^2 q_i^2$ and simplify it to obtain the result. To show $\lambda_{\text{tax}} \equiv \frac{(\lambda\tau+\rho)^2+\rho^2}{(\lambda\tau+2\rho)^2} = \frac{1+\lambda^2}{2}$, use the fact that $\lambda$ is now twice as large as the price impact without tax. Finally, $\lambda^2 < 1$ implies $\frac{1+\lambda^2}{2} > \lambda^2$.

Proof of Lemma 14

(a) Since $\lambda + b > 0$ in this case, following the proof of Proposition 1 yields the result.

(b) Since $\lambda(b)$ is continuous and increasing in $b$ and $\lambda(0) > 0$ and $\lambda(-\frac{\rho}{\tau}) = 0$, it follows $\lambda(b^*) + b^* = 0$ for some $b^* \in \left(-\frac{\rho}{\tau}, 0\right)$. It is easy to verify that $-\lambda_{PT} \in (-\frac{\rho}{\tau}, 0)$ and that $\lambda(-\lambda_{PT}) = \lambda_{PT}$. Since $\lambda(b^*) + b^* = 0$, the first order condition (41) becomes $E_i[v] - \rho Var_i[v] (q_i + x_i) = p$, which is the first order condition under price-taking assumption.

(c) Substitute $q_i = \frac{\tau}{(\lambda+b)\tau+\rho}(E_i[v] - p - \frac{\rho}{\tau}x_i) = \frac{\tau}{\lambda\tau+2\rho}(E_i[v] - p - \frac{\rho}{\tau}x_i)$ in $\Pi = (E_i[v] - p)(q_i + x_i) - \frac{\rho}{2\tau}(q_i + x_i)^2 + px_i + \frac{1}{2}\lambda_{PT} q_i^2$ and simplify it to obtain the result. Finally, $\frac{\tau}{\rho} \lambda_{PT} = \frac{1+\varphi(N-1)}{(1-\varphi)(N-1)} = \frac{1+\chi N}{N-1}$, which is smaller than 1 since $\chi < 1 - \frac{2}{N}$.

APPENDIX B

In this appendix, I compare the trading game in this chapter with an alternative game with exogenous noise trading. The key distinction between the two models is whether a random supply of the asset (liquidity supply) is endogenous or exogenous. In this paper, the source of the random supply is the traders’ private endowments. In the alternative
model, it is noise trading $z$ and endowments are known to be zero ($x_i = 0$ for all $i$, $\tau_x = \infty$). Accordingly, a normal random variable $z$ with $\tau_z < \infty$ appears in a market-clearing condition: $\sum_i q_i(p) + z = 0$. These two models are compared under two environments: strategic and price-taking. Table B.1 summarizes the four cases.

<table>
<thead>
<tr>
<th></th>
<th>Endowments</th>
<th>Noise trading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic</td>
<td>Case A</td>
<td>Case B</td>
</tr>
<tr>
<td>Price-taking</td>
<td>Case C</td>
<td>Case D</td>
</tr>
</tbody>
</table>

Case A corresponds to the model in this paper. In Section 3, I showed a comparative statics where trading decreases to zero and the price impact goes to infinity with noisy prices. In Case B, these symptoms arise asymptotically as noise trading disappears.\(^{20}\) It can be shown that these symptoms are possible also only asymptotically in a price-taking environment. In Case C, these are possible only in the limit where endowment uncertainty vanishes. In Case D, these symptoms happen only in the limit where noise trading vanishes. Thus, it is a combination of strategic behavior and endogenous supply of the asset that makes the condition (15) relevant.

To see why (15) is relevant only in Case A, it is useful to compare the four cases in terms of endogenous parameters ($\beta_s, \varphi$). The difference between a strategic model and a price-taking model is in $\beta_s$:

$$\beta_s^S = \frac{\tau \varepsilon}{\rho} \left( 1 - 2 \varphi - \frac{1}{N-1} \right)$$

$$\beta_s^{PT} = \frac{\tau \varepsilon}{\rho} \left( 1 - \varphi \right).$$

\(^{20}\)See Kyle (1989) for details.
The difference between a model with endowments and a model with noise trading is in \( \varphi \);

\[
\varphi^E = \frac{1}{1 + \frac{\beta^2}{\tau_x \tau_x}}. \tag{51}
\]

\[
\varphi^N = \frac{1}{1 + \frac{1}{(N-1)\beta^2_x \tau_x}}. \tag{52}
\]

Superscripts \( S, PT, E \) and \( N \) denote strategic, price taking, endowments, and noise trading, respectively. Condition (15) is relevant only with a combination \((\beta^S_x, \varphi^E)\).

First, \( \beta^S_x \) is necessary since \( \beta^S_x \) can take negative value with sufficiently high information sharing (\( \varphi \) being close to 1), while \( \beta^{PT}_x \) is always non-negative for any \( \varphi \in [0, 1] \). Second, \( \varphi^E \) is necessary since it mutes a feedback effect through noise trading in the following sense. As \( \beta_x \) becomes large, price becomes more informative, i.e., \( \varphi^N \) becomes large. However, there is a feedback effect from \( \varphi^N \) which makes \( \beta_x \) small. Because the amount of noise trading is not affected by the traders’ choice of \( \beta_x \), large \( \beta_x \) means the market-clearing price provides more information. Since the informational value of prices is higher, the traders put more weight on prices, and thus, less weight on private signals. This partially offsets the initial increase in \( \beta_x \).

In a model with endowments, there is no such feedback. The absence of feedback is due to the balance between risk sharing and speculation: \( \frac{\beta_x}{\beta^S_x} = \frac{\varphi_x}{\tau_x} \). When \( \varphi^E \) becomes small, \( \beta_x \) becomes large, but so does \( \beta_x \). The first effect is the same as the information-based effect described above, while the second effect is based on the risk sharing motive—the less informative a market-clearing price is, the greater the need for sharing risk against an uncertain payoff. More risk-sharing-based trades imply less informative prices. In equilibrium, the two effects exactly cancel out each other, which leaves \( \varphi^E \) independent of \( \beta_x \) and \( \beta_x \).

The following simulation illustrates a comparison of the four cases A through D. Note that total supply of the asset is \( \sum_{i=1}^{N} x_i \) in Case A and C, while it is \( z \) in Case B and D. In the simulation, I set \( \tau_x = (N-1)\tau_z \) such that ex ante variance of total supply of the asset is the same from each trader’s perspective. In the simulation, \( N \) and \( \tau_x \) are increased,
keeping $\frac{\tau_e}{N-1} = 1$ for the model with endowments. For the model with noise trading, $N$ is increased, keeping $\tau_s = 1$. Other parameters are set equally in the two settings: $\tau_e = \tau_v = 1$, and $\rho = 5$. Given these numbers, small $N$ satisfies (15) but large $N$ does not. Therefore we observe for Case $A$ that trade decreases to zero and price impact goes to infinity as $N$ approaches some finite value. In the other three cases, $\beta_s$ goes to zero in the limit as $N$ goes to infinity, but $\lambda$ also decreases and is finite in the limit. Figures B1 through B3 show the behavior of $(\beta_s, \varphi, \lambda)$ as $N$ increases.

![Figure B1](image-url)
Figure B1 compares $\beta_s$ in the four cases. It shows that $\beta_s$ can be negative in Case A for finite $N$. This is when the second-order condition (9) is violated. For the other three cases, $\beta_s$ is always positive. Note also that $\beta_s$ is smaller in the strategic cases (A and B) than in price taking cases (C and D), and more significantly so in a model with endowments.
(A compared with C) than in a model with noise trading (B compared with D).

Figure B2 compares $\varphi$. First it shows that Case A and C have the same $\varphi$, which approaches 1 as $\tau_x$ goes to infinity (recall that $\frac{\tau_x}{N-1} = 1$ is fixed). This causes problem in Case A when $\varphi$ approaches value close to $\frac{1}{2}$, violating the second order condition (9), while in Case C the same value of $\varphi$ never violates (9). The figure also shows that information sharing is greater in a model with endowments than in a model with noise trading, even though ex ante total quantity uncertainty is the same in two markets. As mentioned above, the hedging motive endogenously decreases quantity uncertainty in a model with endowments, while there is no such mechanism with exogenous noise trading.\footnote{Theorem 6.1. and 7.2. in Kyle (1989) show that $\varphi$ approaches $\frac{1}{2}$ in Case B and 1 in Case D.}

Finally, Figure B3 shows the behavior of $\lambda$. In Case A, the market becomes illiquid for finite $N$. It can be proved that $\lambda$ approaches $\lim_{x} \frac{\tau_x}{\rho(N-1)} = 0.2$ in Case C and zero in Case B and D. This shows that, even though ex ante aggregate uncertainty is the same, the liquidity implication can be quite different between a model with endowments and a model with noise trading.
References


