FINANCE RESEARCH SEMINAR SUPPORTED BY UNIGESTION

"Dynamic Trading: Price Inertia, Front-Running and Relationship Banking"

Prof. Yuliy SANNIKOV
Princeton University, Dept. of Economics

Abstract
We build a linear-quadratic model to analyze trading in a market with private information and heterogeneous agents. Agents receive private endowment shocks and trade continuously. Agents differ in their need for trade as well as size, i.e. the ability to stay away from their ideal positions. In equilibrium, trade is gradual, its speed depends on the size of the market, and trade among large market participants is slower than that among small investors. Price has momentum due to the actions of large traders: it drifts up if the sellers are fewer and larger and the buyers are smaller and more competitive, and vice versa. The model captures welfare: it can answer questions about the social costs and benefits of high-frequency traders, the welfare consequences of market consolidation, and many others.

Friday, October 9, 2015, 10:30-12:00
Room 126, Extranef building at the University of Lausanne

Dynamic Trading: Price Inertia, Front-Running and Relationship Banking[*]
Preliminary and Incomplete.

Yuliy Sannikov

Department of Economics

Princeton University

`sannikov@princeton.edu`

Andrzej Skrzypacz

Graduate School of Business

Stanford University

`skrz@stanford.edu`

April 10, 2014

Abstract

We build a linear-quadratic model to analyze trading in a market with private information and heterogeneous agents. Agents receive private endowment shocks and trade continuously. Agents differ in their need for trade as well as size, i.e. the ability to stay away from their ideal positions. In equilibrium, trade is gradual, its speed depends on the size of the market, and trade among large market participants is slower than that among small investors. Price has momentum due to the actions of large traders: it drifts up if the sellers are fewer and larger and the buyers are smaller and more competitive, and vice versa. The model captures welfare: it can answer questions about the social costs and benefits of high-frequency traders, the welfare consequences of market consolidation, and many others.

1   Introduction.

A market with heterogeneous investors - large institutions, small retail investors, liquidity providers and high-frequency traders - presents many puzzles. What determines the speed of trading? What determines price momentum? When a large seller sends his flow to the market, how does the price react at the inception, as the flow continues, and when the flow stops? How detrimental are traders that try to discover large buyers and sellers, and front-run them? Should large institutions be protected from traders that try to front-run them? Do high-frequency traders enhance welfare

---

[*]We are grateful to seminar participants at Yale and the New York Fed for helpful comments.

by providing liquidity or do they harm investors? Why do certain market makers pay to execute retail flow?

To address these questions in a unified manner, a theoretical framework is needed. We need to understand strategic interactions in a market, in which agents who want to buy and sell choose their flows strategically, anticipating the price impact and the execution risk. When the strategic considerations of individual traders are put together, what do they imply about the speed of trading, liquidity, price dynamics and inefficiencies? We need a model that is suitable for analyzing welfare: the costs from the time delay in the execution of large orders and the costs of price impact. Importantly, it has to be a model that captures the utilities of all agents explicitly, i.e. it cannot ignore a part of the market by designating it to be the "noise traders."

We build a model to address these issues and to also be able to analyze welfare. Market participants are heterogeneous: they differ in their trading needs and also in their capacity to wait and absorb risk. There can be many motivations for trade: investors may want to rebalance portfolios, hedge risk exposures, trade to accommodate client needs, etc. Some participants may demand liquidity and have large trading needs, while other participants may have the capacity to make profit while providing liquidity. However, the key source of heterogeneity in the model is not the need to trade, but rather the capacity to wait. Large players are willing to stay further from their ideal positions when the price is not right. They have a greater risk capacity than small traders. It turns out that the concentration of the market, and its composition in terms of participants of different sizes, matters crucially for market dynamics.

Because large and small market participants trade differently, it pays to know the source of trade. Indeed, large players - those with the capacity to wait to minimize price impact - trade slowly. Small players trade fast. When the source of sales is a large trader, then we know that these sales are just a tip of the iceberg: the large player will hide most his desired quantity. Thus, sales by a large player have put a persistent downward pressure on the market price. In contrast, when sales are initiated by small traders, perhaps a group of small traders that decide to sell at the same time, then desired quantities are traded fast. Small traders have no incentives to wait, especially if sales by other small traders are pushing the price down. Thus, while sales of small traders push the price down at the moment, they do not imply a continuous downward pressure on the price.

The knowledge of the source of trades provides important information about price momentum. In practice, market participants can have various strategies to learn the source of trades: they can observe whether most recent orders are executed near the bid or the ask, they can watch the size of orders, they can try to identify active market participants through trading, and they can even pay for the flow that they know comes from small retail investors. Theoretically, if market participants observe only the price, the may try to learn the source of trade from price momentum. This can lead to a very complicated filtering problem, in which players' beliefs about the

trading needs of other market participants, and even higher-order beliefs, may affect strategies.

To highlight important properties of markets with heterogeneous participants in a clean way, throughout this paper we assume that players observe the source of trade, so they do not need to worry about the filtering problem. This creates a clear benchmark, which nevertheless leads to highly non-trivial dynamics. Different traders choose to trade towards their desired positions at different rates. Consequently, different players have different price impact. When some market participants want to buy and others want to sell, trade is not immediate but slow. The rate of trading, and convergence to desired allocations, depends on the composition of both sides of the market. Importantly, prices exhibit momentum: prices drift down if the segment of the market that wants to sell is more concentrated (i.e. consists of fewer and larger traders) than the segment of the market that wants to buy. The welfare of traders depends on whether they provide or take liquidity - liquidity providers can make profit while other traders generally pay costs through price impact and due to delayed execution. The welfare of traders also depends on their size, i.e. their pricing power.

While our model is first to capture many of these features, we build upon a lot of important market microstructure literature. Papers such as Kyle (1985) and Back (1992) capture price impact and gradual trading in a model that features an insider that has private information and noise traders. In these models, from the point of view of market participants who have no inside information, prices have no drift. From the point of insiders, of course, prices drift towards fundamentals known only to insiders. Welfare analysis using these models, however, is restricted by the fact that noise traders provide exogenous flows and have no utility functions. Much closer to our model are the papers of Vayanos (1999) and Du and Zhu (2013). Those papers model a market with finitely many symmetric traders. In those models, even though prices reflect the efficient allocation of assets, allocations themselves are inefficient. Players trade slowly to the efficient allocation: like in our model, they signal their private information by the rate of selling. Vayanos (1999) shows that the speed of trading increases in the number of market participants, and the equilibrium converges to efficiency as the number of players grows to infinity. Du and Zhu (2013), in addition, show that the trading speed slows down when players receive not only shocks to their own endowments but also information about the common component of value. In addition, Du and Zhu (2013) also analyze the implications of the frequency of trading on efficiency.

Relative to Vayanos (1999) and Du and Zhu (2013), who model markets through a double uniform-price auction, our paper presents a methodological contribution as we develop a tractable way to analyze markets with heterogeneous traders. As discussed above, double auctions present a complicated filtering problem in our environment, as players want to know their counterparty. To avoid this problem, we assume that players observe the distribution of supply and demand across players of different

3

sizes, or, equivalently, they can condition their supply and demand on the size of the counterparty. Players can change their behavior depending on whether they face a competitive segment of the market with small investors, or a large counterparty.

Our model, set in a linear-quadratic framework, leads to interesting dynamics. Trading speeds can be characterized conveniently via the eigenvector decomposition of the matrix that describes the rates, at which each player sells his endowment and at which other players absorb these flows. The eigenvectors correspond to misallocations away from efficiency, and the eigenvectors describe the speed at which these misallocations get traded away. We find that misallocations among large players get traded away much more slowly than those among the more competitive segment of the market. The equilibrium price does not depend on total supply uniformly: it is more sensitive to the supply from small traders as they tend to sell their endowments faster. However, as a function of the flow, large traders have a greater price impact. When a large trader sells, the market infers that more is left behind, and so the price drops more.

We can also study welfare using our model. One interesting implication of the model is that liquidity providers, including high-frequency traders, who may make money by front-running large investors, are generally good for welfare. This observation is somewhat at odds with the common view that high-frequency traders are good for small retail investors, but bad for large institutional traders. Indeed, when we introduce into the model new market participants who do not have trading needs on their own, but who participate in the flow to make profits off of price momentum, they do "front-run" large traders but they also change the entire equilibrium dynamics. The general force at work here is that the more market participants there are, the faster the speed of trading, and the lower price impact everyone has. Large players trade faster, in part because they expect to be front-run. The market, anticipating this behavior, reacts less to the trades of large traders: it expects less of the iceberg to be hidden underwater. The entry of liquidity providers does not benefit everyone, however. Clearly, other liquidity providers suffer from greater competition.

This paper is organized as follows. Section 2 presents the baseline linear-quadratic model and discusses trading in environments where players can adjust behavior depending on the size of their counterparty. Section 3 derives equilibrium equations when each trader anticipates their price impact and takes into account aggregate price dynamics. Section 3 also characterizes convergence to efficiency and price dynamics using eigenvectors and eigenvalues. Section 4 introduces a competitive fringe into the model and investigates phenomena such as price momentum and front running. Section 5 analyzes welfare, especially the effects of mergers and high-frequency traders. Section 6 microfounds the linear-quadratic model in a more realistic framework with exponential utility. Section 7 concludes.

2   The Model.

This section lays out a basic linear-quadratic model of trading in a small market with private information about individual preferences. While here we assume quadratic preferences, in Section 4 we present a more realistic model with exponential utility. As we see later, the linear-quadratic model provides an especially tractable special case of the exponential model. It leads to trading dynamics that are expressed cleanly through the market power of individual players, independently of the individual shocks to buy or sell. The exponential model leads to a slightly more complicated system of equilibrium equations, and it also allows not only for private values, i.e. idiosyncratic reasons for trading, but also a common-value component of private information, related to fundamentals and future cash flows.

There are N large traders. The traders get stochastic endowment shocks of an asset. The position X_t^i of trader i evolves according to the equation

$$dX_t^i = -\delta X_t^i \, dt + \sigma_i \, dZ_t^i - q_t^i \, dt, \tag{1}$$

where Brownian endowment shocks Z_t^i have the non-singular correlation matrix

$$\mathcal{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1N} \\ \rho_{21} & 1 & \cdots & \rho_{2N} \\ \vdots & \vdots & & \vdots \\ \rho_{N1} & \rho_{N2} & \cdots & 1 \end{bmatrix} \tag{2}$$

and q_t^i reflect the net trading rates of the asset. The trading rates q_t^i must add up to 0, so that the markets clear. Parameter δ reflects the depreciation rate.

For simplicity, we assume that the players' preferences over asset holdings are quadratic of the form

$$- b^i (X_t^i)^2 / 2. \tag{3}$$

That is, the player experiences a quadratic disutility when his position deviates away from the bliss point, normalized to 0. We denote the vector of the players' risk parameters by $B = [b^1, \, b^2, \ldots b^N]$.

Players have private information about their endowment shocks dZ_t^i. We can interpret this private information, which motivates trade, in various ways. Players could be equity fund managers, who have to manage inflows or outflows of clients money. Trades can be motivated by portfolio rebalancing. Alternatively, asset X could also reflect a particular risk exposure that a market participant can have, such as exposure to interest rate or currency risk. If so, then we can interpret this as a market for options or swaps to hedge these risk exposures. Depreciation δ can be interpreted as the rate at which the trader absorbs the risk that he would otherwise desire to trade.

We can interpret $1/b^i$ as the "risk capacity" of trader i. Players with a lower coefficient b^i are "larger": they can hold larger positions away from their bliss points

at a lower cost. Conversely, players with a higher coefficient b^i are "smaller", and therefore more impatient to trade the shocks to their endowments.

Players are risk-neutral with respect to monetary transfers and they discount their payoffs at rate r. If p_t is the price at which player $i = 1, \ldots N$ sells the flow of q_t^i at time t, then his total payoff is given by

$$E\left[r \int_0^\infty e^{-rt} \left(b^i \frac{(X_t^i)^2}{2} + p_t q_t^i \right) dt \right].$$

First Best. The efficient allocation of assets is proportional to risk capacities. Given any endowments $(X_t^1, \ldots X_t^N)$, it would be efficient for the players to trade immediately to the efficient allocation, under which player i would be holding the quantity

$$\hat{X}_t^i = \frac{1}{b_i \bar{\beta}} \bar{X}_t, \quad \text{where} \quad \bar{X}_t = \sum_{n=1}^N X_t^n \quad \text{and} \quad \bar{\beta} = \sum_{n=1}^N 1/b_n. \tag{4}$$

If endowments were publicly observable, then the players would be able to trade to the efficient allocation immediately. The resulting price would be equal to the marginal disutility of holding an additional unit of asset, which would be the same across all agents. This value is given by

$$\hat{p}_t = \frac{d}{dx_i} \int_0^\infty e^{-rt} \frac{-b_i}{2} (e^{-\delta t} \hat{X}_t^i)^2 \, dt = -\frac{b_i \hat{X}_t^i}{r + 2\delta} = -\frac{1}{(r + 2\delta)\bar{\beta}} \bar{X}_t, \tag{5}$$

where $\bar{\beta}$ is the total risk capacity of the market. Under first best, an extra unit of endowment received by any player has the same price impact of

$$-\frac{1}{(r + 2\delta)\bar{\beta}}. \tag{6}$$

Mechanisms for trading. It is typical to model market trading through a double uniform-price auction, as in Vayanos (1999) and Du and Zhu (2012). However, in our setting, since players are not symmetric, such a mechanism leads to a complicated solution which involves a filtering problem. For reasons that will become clear later, players would want to know not only the price but also the counterparty. They would be making inferences about the distribution of supply and demand, across players of different sizes, from the dynamic properties of prices and through other means.

To avoid these filtering problems, we assume that players observe the flows of all other players, or can condition their demand functions on these flows.[1] To match these requirements, we analyze the following auction format.

---

[1] There is evidence that market participants in practice spend a considerable amount of effort identifying the sources of trades. For example, brokers call each other to find out who traded, and some market-makers pay discount brokerages for the flow from retail investors. Moreover, recently NYSE began allowing orders from retail investors to be marked as such through the Retail Liquidity Program (RLP).

CONDITIONAL DOUBLE-AUCTION. *At each moment of time t, each player i announces a supply-demand function*

$$p = \bar{\pi}^i - \sum_{j \neq i} \pi^{ij} q^j$$

that gives the price at which the player is willing to trade, as a function of the selling rates of all other players (with player i buying the net residual supply). The market maker then determines the price p and the selling rates q^j from the system of equations

$$\sum_{i=1}^{N} q^i = 0, \qquad \forall\, i, \quad p = \bar{\pi}^i - \sum_{j \neq i} \pi^{ij} q^j. \tag{7}$$

Note that the system of equations (7) may have no solutions, or multiple solutions. The market maker may have special treatment for those situations, e.g. the price-flow vector $(p, q) = (0, 0, \ldots 0)$ may be chosen in those situations. Moreover, the players' bids may or may not reveal information about their endowment shocks. We focus on strategy profiles that never lead to degeneracies, and which reveal the players' private endowment shocks. That is, we look for fully separating equilibria of these games.

A profile of strategies is *stationary* if the slopes of the demand functions π^{ij} remain constant, while the intercepts $\bar{\pi}^i$ may depend on the players' endowments. Furthermore, a stationary profile of strategies is *linear* if $\bar{\pi}^i = \hat{\pi}^i X^i$ for an appropriate constant $\hat{\pi}^i$, where X^i is player i's endowment. Obviously, a linear stationary profile such that $\hat{\pi}^i \neq 0$ for all i is revealing (i.e. fully separating).

While the conditional double auction is an intuitive way to model price formation in the market, it is easier to analyze price formation and trade dynamics using a direct revelation mechanism that is (as we show below) strategically equivalent to that auction.

DIRECT REVELATION MECHANISM. *A (stationary, linear) mechanism (P, Q) is a direct revelation mechanism in which at each moment of time t, the market maker asks every trader to announce his endowment X_t^i. The vector of announcements determines the price $p_t = P X_t$ and the vector of trading rates $q_t = Q X_t$ such that the market clears (i.e. the columns of Q must add up to 0). We require the mechanism (P, Q) to be within-period incentive compatible i.e., that truthtelling is a best response for every player for every vector of reports of other traders.*

Note that since the mechanism is linear in reports, for every vector of reports of others, trader i can find a report that implies he does not trade. Hence, in this setup ex-post incentive compatibility implies that (ex-post) individual rationality is satisfied as well.

We now show that any stationary linear equilibrium of the conditional double-auction can be implemented through a direct revelation mechanism and vice versa.

Definition. *Two profiles of strategies in two mechanisms are equivalent if (1) they lead to the same paths of prices and flows, conditional on histories of endowments shocks and (2) after each history, each player with his action can choose among price-flow vectors from the same set under both mechanisms.*

Theorem 1 *Given any stationary linear revealing equilibrium of the conditional double auction, there is an equivalent truth-telling equilibrium of an appropriate direct revelation mechanism, and vice versa.*

Proof. Consider any stationary linear revealing profile of strategies of the conditional double auction, and let us show that truth-telling under an appropriate direct revelation mechanism is equivalent, in the sense that they lead to (1) the same paths of prices and flows, conditional on endowment shocks and (2) choice sets for all players.

Let us derive the system of equations that characterizes the map from conditional demand functions to the price-flow vectors $(p, q_2, \ldots q_N)$. Since $q_1 = -\sum_{i=2}^{N} q_i$, the demand functions of players 2 through N can be written in terms of the flows $(q_2, \ldots q_N)$ as follows

$$p + \sum_{j \neq 1, i} (\pi^{ij} - \pi^{i1}) q^j - \pi^{i1} q^i = \bar{\pi}^i.$$

Thus, the map from conditional demand functions to prices and flows is represented through the matrix equation

$$\underbrace{\begin{bmatrix} 1 & \pi^{12} & \cdots & \pi^{1N} \\ 1 & -\pi^{21} & \cdots & \pi^{2N} - \pi^{21} \\ 1 & \vdots & & \vdots \\ 1 & \pi^{N2} - \pi^{N1} & \cdots & -\pi^{N1} \end{bmatrix}}_{\hat{\Pi}} \begin{bmatrix} p \\ q^2 \\ \vdots \\ q^N \end{bmatrix} = \begin{bmatrix} \bar{\pi}^1 \\ \bar{\pi}^2 \\ \vdots \\ \bar{\pi}^N \end{bmatrix}.$$

Given any non-degenerate strategy profile, this equation must have a unique solution, so the matrix $\hat{\Pi}$ must be invertible. Denote its inverse U and its components by u_{ij}. Any subset of $N-1$ equations has a one-dimensional set of solutions, and so the set of price-flow vectors that any player i has control over is also one-dimensional. Player i can reach any point in that set simply by varying $\bar{\pi}^i$, and so this set is

$$U \begin{bmatrix} \bar{\pi}^1 \\ \bar{\pi}^2 \\ \vdots \\ \bar{\pi}^N \end{bmatrix} + \begin{bmatrix} u^{1i} \\ u^{2i} \\ \vdots \\ u^{Ni} \end{bmatrix} x.$$

For comparison, in a direct revelation mechanism, the map from allocations to prices

8

and flows is determined by the map

$$
\underbrace{\begin{bmatrix}
p^1 & p^2 & \cdots & p^N \\
q^{21} & q^{22} & \cdots & q^{2N} \\
\vdots & \vdots & & \vdots \\
q^{N1} & q^{N2} & \cdots & q^{NN}
\end{bmatrix}}_{Q^P} \hat{X} =
\begin{bmatrix}
p \\
q^2 \\
\vdots \\
q^N
\end{bmatrix},
$$

where $\hat{X}$ represents the vector of reports, q^{ij} represent the entries of the matrix Q and p^i represent the entries of the vector P. Player i controls the ith component of $\hat{X}$. Thus, in order for the second requirement of equivalence to hold, the ith column of U must be collinear to the vector

$$
\begin{bmatrix}
p^i \\
q^{2i} \\
\vdots \\
q^{Ni}
\end{bmatrix}.
$$

Thus, to get from vector P and matrix Q to the equivalent demand functions, we must invert matrix Q^P to obtain a matrix in which *rows* are collinear to the rows of $\hat{\Pi}$ and multiply each row i by a constant α^i so that the first column is a column of ones. Furthermore, to ensure that the vector $(p, q_2, \ldots q_N)$ is in the choice set of all players, and is chosen, given their vector of endowments X, we need that

$$
U \begin{bmatrix}
\bar{\pi}^1 \\
\bar{\pi}^2 \\
\vdots \\
\bar{\pi}^N
\end{bmatrix} = Q^P \begin{bmatrix}
\alpha^1 \bar{\pi}^1 \\
\alpha^2 \bar{\pi}^2 \\
\vdots \\
\alpha^N \bar{\pi}^N
\end{bmatrix} = Q^P X =
\begin{bmatrix}
p \\
q^2 \\
\vdots \\
q^N
\end{bmatrix}.
$$

That is, we need to take

$$
\bar{\pi}^i = \hat{\pi}^i X^i, \quad \text{where} \quad \hat{\pi}^i = 1/\alpha^i.
$$

This leads to a stationary linear revealing strategy profile of the conditional double auction, which is equivalent to the truth-telling strategy profile of the direct revelation mechanism (P, Q).

We must reverse the procedure to get from the conditional double auction to the direct revelation mechanism. That is, we need to start with the matrix $\hat{\Pi}$, invert it, and then multiply each *column* of the resulting matrix by $\hat{\pi}^i$ to obtain the matrix Q^P. This matrix can be split into the vector P and the matrix Q, knowing that the columns of Q must add up to 0. (Note that rows 2 thought N of Q^P are rows 2 through N of Q).

Finally, the argument so far has focused on the equivalence of strategy profiles. Of course, since the set of choices available to each player after each history is the same under both mechanisms, it follows that if we have an equilibrium of the conditional double auction, then truthtelling must be an equilibrium of the corresponding direct revelation mechanism, and vice versa. ∎

From now on we will focus on direct revelation mechanisms, in which truth-telling is an equilibrium, as the representation in terms of P and Q provides a direct map from the players' allocations to prices and flows.

3   Equilibrium Characterization.

This section derives the equations that characterize trading dynamics under stationary linear equilibria in our model.

Under a direct revelation mechanism (P, Q), on the equilibrium path the endowments follow

$$dX_t = -\delta X_t\, dt + \Sigma\, dZ_t - QX_t\, dt,$$

where Σ is the diagonal matrix of the volatilities of endowment shocks and Z is a vector of N Brownian motions with the correlation matrix $\mathcal{R}$ given by (2). The resulting price is given by $p_t = PX_t$.

If player i deviates and reports endowment $y + X_t^i$ instead of X_t^i then the resulting price is $p_t + p^i y$ and the endowment vector follows

$$dX_t = -\delta X_t\, dt + \Sigma\, dZ_t - QX_t\, dt - Q^i y\, dt, \tag{8}$$

where Q^i is the ith column of Q.

Denote the value function of player i by $f^i(X)$, where X is the vector of players' endowments. Then function $f^i(X)$ must satisfy the HJB equation

$$r f^i(X) = \max_y \frac{-b^i}{2}(X^i)^2 + (PX + p^i y)(\underline{Q}^i X + q^{ii} y) + \tag{9}$$

$$\nabla f^i(X)(-\delta X - QX - Q^i y) + \frac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial^2 f^i}{\partial X^j \partial X^k}\sigma^j \sigma^k \rho_{jk},$$

where $\underline{Q}^i$ is the i-th row of Q, q^{ii} is the i-th diagonal entry of Q and ∇f^i denotes the gradient of f^i. In a truth-telling mechanism, $y = 0$ must solve the maximization problem in (9).

We conjecture (and verify) that the players' value functions take the quadratic form $f^i = X^T A^i X + k^i$, where A^i is a symmetric N-by-N matrix and k^i is a constant. Given that, the HJB equation (9) becomes

$$r(X^T A^i X + k^i) = \max_y \frac{-b^i}{2}(X^i)^2 + (PX + p^i y)(\underline{Q}^i X + q^{ii} y) - 2X^T A^i(\delta X + QX + Q^i y) + \sigma^T(A^i \circ \mathcal{R})\sigma,$$

where $\sigma = [\sigma_1, \sigma^2, \ldots \sigma^N]^T$ is the vector of volatilities of individual shocks and $A^i \circ \mathcal{R}$ denotes the Hadamard (i.e. element-wise) product of two matrices.

Taking the first-order condition at $y = 0$, the HJB equation reduces to the following system of equations

$$p^i \underline{Q}^i + q^{ii} P = 2(A^i Q^i)^T, \tag{10}$$

$$k^i = \frac{1}{r}\sigma^T (A^i \circ \mathcal{R})\,\sigma \quad \text{and} \quad A^i((r + 2\delta)I + 2Q) \sim P^T \underline{Q}^i - \frac{b^i}{2}1^{ii}, \tag{11}$$

where we use the notation "$\sim$" to indicate that two matrices have the same diagonals, and the same sums of all symmetric off-diagonal entries, I denotes the N-by-N identity matrix and 1^{ii} denotes the square N-by-N matrix that has 1 in the i-th diagonal position and zeros everywhere else.[2]

The following proposition formally registers the fact that appropriate solutions of equations (10) and (11) indeed lead to equilibria.

Proposition 1 *Consider any solution $(P, Q, k^i, A^i, i = 1, \ldots N)$ of the system (10) and (11) such that $p^i < 0$ and $q^{ii} \geq 0$ for all $i = 1, \ldots N$, and the matrix Q is such that the process X defined by (8) is non-explosive. Then, for all i it is optimal to follow the truth-telling strategy if all other players tell the truth in the direct revelation mechanism given by (P, Q). That is, we have stationary linear equilibrium of the model.*

Proof. See Appendix. ∎

The trading game has degenerate equilibria, in which some or all of the traders are excluded from the market (i.e. the matrix Q consists of zeros in several columns). We are interested primarily in the non-degenerate equilibria, and would like to understand their properties such as the speed of trade, price momentum, and inefficiencies.

Unfortunately, the system of (10) and (11) cannot be solved in closed form in general. However, any equilibrium for a given pair (r, δ) can be adjusted to obtain an equilibrium for any other pair $(\hat{r}, \hat{\delta})$. As the following proposition demonstrates, in general the speed of trading is proportional to $r + 2\delta$ for any set of market powers of individual players.

Proposition 2 *Consider a stationary linear equilibrium (P, Q) of the game with parameters (r, δ) Let $\alpha = (\hat{r} + 2\hat{\delta})/(r + 2\delta)$. Then for parameters $(\hat{r}, \hat{\delta})$, there exists a stationary linear equilibrium in which*

$$\hat{Q} = \alpha Q \quad \text{and} \quad \hat{P} = P/\alpha. \tag{12}$$

---

[2] That is, $A \sim B$ if $a^{ij} + a^{ji} = b^{ij} + b^{ji}$ for all i and j, or, equivalently, if $X^T A X = X^T B X$ for all vectors X.

Proof. Suppose that the value functions of players $i = 1, \ldots N$ are characterized by (A^i, k^i) in the equilibrium (P, Q) under the parameters (r, δ). Then $(P, Q, k^i, A^i, i = 1, \ldots N)$ of the system (10) and (11) and satisfy the conditions of Proposition 1. Define $\hat{A}^i = A^i/\alpha$ and $\hat{k}^i = (r/\hat{r})k^i/\alpha$. Then it is straightforward to verify that $(\hat{P}, \hat{Q}, \hat{k}^i, \hat{A}^i, i = 1, \ldots N)$ solve the system (10) and (11) for parameters $(\hat{r}, \hat{\delta})$ and satisfy the conditions of Proposition 1. Thus, parameters $(\hat{r}, \hat{\delta})$, $(\hat{P}, \hat{Q})$ give a linear stationary equilibrium. ∎

3.1 The Speed of Trade and Price Impact.

The equilibrium is inefficient: players take time to trade towards the efficient allocation even though on the equilibrium path they can infer everybody's desire to trade from their trading behavior. The following proposition illustrates the equilibrium for the case of symmetric traders.

Proposition 3 *If the players have identical risk parameters given by $B = [b, b, \ldots b]$, then in the unique symmetric non-degenerate equilibrium the price is always first best, given by the vector[3]*

$$P = -\left[\frac{b/N}{r + 2\delta}, \frac{b/N}{r + 2\delta}, \cdots \frac{b/N}{r + 2\delta} \right].$$

However, the allocation does not jump to first best immediately, but rather its convergence is given by the matrix

$$Q = \frac{(N - 2)(r + 2\delta)}{2N} \begin{bmatrix} N - 1 & -1 & \cdots & -1 \\ -1 & N - 1 & \cdots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \cdots & N - 1 \end{bmatrix}.$$

That is, the allocation converges towards efficiency at the exponential rate given by $(N - 2)(r + 2\delta)/2$. The players' welfare is characterized by the matrices $(r + 2\rho)A^i$ with entries

$$a^i_{ii} = -\frac{b}{2}\frac{3N - 2}{N^2}, \quad a^i_{ij} = -\frac{b}{2}\frac{N - 2}{N^2} \quad \text{and} \quad a^i_{jk} = \frac{b}{2}\frac{N - 2}{(N - 1)N^2} \quad (13)$$

for $j, k \neq i$. In case shocks Z^i_t are uncorrelated across players, the constant term in the value functions reduces to

$$k^i = \frac{1}{r(r + 2\delta)}\frac{b}{2N^2}\left(-(3N - 2)(\sigma^i)^2 + \frac{N - 2}{N - 1}\sum_{j \neq i}(\sigma^j)^2 \right). \quad (14)$$

---

[3]Of course, there are also degenerate equilibria in which specific rows of Q are set to 0, i.e. specific players are excluded from trade.

Proof. See Appendix. ∎

The result that with symmetric players, trading takes place gradually even though the price converges to first-best immediately, has already been proved in a slightly different context by Vayanos (1999). Of course, the speed of trade is increasing in the number N of players in the market.

Equations (13) and (14) reveal interesting implications about welfare. Even though the utility functions (3) are always negative, some players may sometimes have positive utility in equilibrium. Those players are the *liquidity providers*: they have low endowment shocks σ^i relative to the rest of the market, and so they can make money by catering to the needs of other players.

Next, we want to understand how the equilibrium changes when players have unequal risk capacity. While in general it seems like the equilibrium cannot be characterized in closed form, the following proposition illustrates how the equilibrium changes near the symmetric case.

Proposition 4 *Consider a perturbation of the symmetric case, in which*

$$B = [b^1,\, b^2, \ldots b^N] = [b,\, b, \ldots b] + [\epsilon^1,\, \epsilon^2, \ldots \epsilon^N], \quad \text{with} \quad \sum_{n=1}^{N} \epsilon^n = 0.$$

Then there is an equilibrium in which

$$P = -\frac{1}{N(r+2\delta)}\left([b,\, b,\, \ldots b] + \frac{3N-4}{N(N-1)}\left[\epsilon^1,\, \epsilon^2, \ldots \epsilon^N\right]\right) + O(\epsilon^2) \qquad (15)$$

and

$$Q = \frac{(N-2)(r+2\delta)}{2Nb}\begin{bmatrix} (N-1)b^1 & -b^2 & \ldots & -b^N \\ -b^1 & (N-1)b^2 & \ldots & -b^N \\ \vdots & \vdots & & \vdots \\ -b^1 & -b^2 & \ldots & (N-1)b^N \end{bmatrix} + O(\epsilon^2). \qquad (16)$$

Proof. See Appendix. ∎

Proposition 4 provides a fairly precise approximation of the equilibrium dynamics even when ϵ^n are not small, e.g. on the order of 10% of b. For example, if $r + 2\delta = 1$ and $B = [1.8,\, 1.9,\, 2,\, 2.1,\, 2.2]$, equation (16) predicts that the rates at which players sell their endowments (i.e. the diagonal of Q) are given by

$$[1.08,\, 1.14,\, 1.2,\, 1.26,\, 1.32].$$

The true rates are given by

$$[1.0820,\, 1.1428,\, 1.232,\, 1.2631,\, 1.3224],$$

13

i.e. the approximation error is only about 0.24% in this case. The approximation errors for P and the off-diagonal entries of Q are similar. In particular, the columns of (16) indicate that when player i sells, the flow is absorbed approximately equally by all other traders despite the differences in their risk capacities. Trade among small traders is faster than it is among large traders.

Some of the most interesting aspects of our model are price impact and price momentum. We can estimate those from the approximation given by Proposition 4. First, from (15) we see that the price is more sensitive to the endowments of large players than those of small players. Large players control have market power and they control their trading rates better, while the small players compete with each other and pay much less attention to the impact of their trades on the price.

The traditional definition of market impact measures the sensitivity of the price to the trading flow, rather than endowment. According to this definition, the price impact of player i is given by

$$\frac{p^i}{q^{ii}} \approx \frac{2}{(r+2\delta)^2(N-1)}\left(\frac{b}{N-2} - \epsilon^i\frac{N-2}{N(N-1)}\right),$$

and it is clear that as long as $N > 2$, the market impact of players is increasing with their size (i.e. as ϵ^i decreases).[4] This phenomenon occurs because we assumed that market participants can observe who they are trading against. They know that when large players sell, their sales are a smaller tip of a bigger iceberg. Therefore, their sales signal larger hidden supply, and the price reacts more.

3.2 Eigenvalue Decomposition of Equilibria.

If no further shocks occur, then any misallocation away from efficiency goes away due to the trading rates from the matrix Q (as well as depreciation), according to the equation

$$dX_t = -\delta X_t \, dt - QX_t \, dt.$$

Moreover, if $\delta = 0$, then any misallocation can go away only through trading. In order to understand how quickly different missalocations get traded away to efficiency, it is useful to compute the eigenvector decomposition of the matrix Q. Then eigenvalues give the rates, at which the misallocations from the corresponding eigenvectors get traded away.

The efficient allocation $U^1 = [1/b^1, \ldots 1/b^N]^T$ is always one eigenvector of Q, with the corresponding eigenvalue being 0. In equilibrium, players do not trade if they are at the efficient allocation: if they traded then at least one player would be worse of than if he had not traded at all. If the equilibrium is nondegenerate and the players eventually trade to efficiency, then all other eigenvectors of Q must have components

---

[4]If $N = 2$, then our game does not have any stationary linear equilibria with trade.

that add up to 0 and the corresponding eigenvalues must be positive. Let

$$Q = U\Lambda U^{-1}$$

be the eigenvalue decomposition of Q, where the columns of U are eigenvectors and Λ is a diagonal matrix of eigenvalues, in increasing order. Furthermore, denote by $\Pi = PU$ the vector that prices the eigenvector misallocations. The following theorem transforms the equilibrium conditions into the eigenvector space.

Theorem 2 *In the eigenvector space, equations (11) and (10) can be written as*

$$\hat{A}^i((r + 2\delta)I + 2\Lambda) \sim \Pi^T \underline{U}^i \Lambda - \frac{b^i}{2}(\underline{U}^i)^T \underline{U}^i \quad and \tag{17}$$

$$\Pi(U^{-1})^i \underline{U}^i \Lambda + (\underline{U}^i \Lambda (U^{-1})^i)\Pi = 2((U^{-1})^i)^T \Lambda \hat{A}^i, \tag{18}$$

where the relationship between A^i and $\hat{A}^i$ is given by $\hat{A}^i = U^T A^i U$. The following expressions for the coefficients of $\hat{A}^i$ are equivalent to (17):

$$\hat{a}^i_{jk} = -\frac{b^i u^{ij} u^{ik} + \pi^j u^{ik} \lambda^k + \pi^k u^{ij} \lambda^j}{2(r + 2\delta + \lambda_k + \lambda_j)}, \tag{19}$$

where λ_k is the k-th diagonal element of Λ.

Proof. See Appendix. ∎

Equations (19) provide a convenient direct formula to compute the players' value functions from the pair (P, Q). Otherwise, to obtain the matrices A^i from (11), one has to solve a more complicated system of equations, or obtain A^i via an iterative procedure.

3.3   An Example.

We finish this section by providing a numerically solved example, in order to develop a sense for what equilibria look like in general, away from the symmetric case. Consider a game with five traders, whose risk coefficients are $(b^1, b^2, b^3, b^4, b^5) = (1, 1.5, 2, 2.5, 3)$. Then any allocation X is priced by the vector $P = [-.254, -.329, -.387, -.435, -.476]$. The rates of trading flows are given by the matrix

$$Q = \begin{bmatrix} 0.630 & -0.244 & -0.319 & -0.389 & -0.455 \\ -0.163 & 0.965 & -0.326 & -0.401 & -0.473 \\ -0.160 & -0.244 & 1.289 & -0.405 & -0.481 \\ -0.156 & -0.241 & -0.324 & 1.598 & -0.483 \\ -0.152 & -0.236 & -0.320 & -0.403 & 1.892 \end{bmatrix}.$$

When players have different risk tolerances, the equilibrium pricing vector P does not assign the same weights to the endowments of different players (even though the first-best pricing vector still assigns the same weight to all players). The reason is that players with large risk capacity $1/b^i$ exercise market power by selling their endowments more slowly. They do it in order to get a more favorable price from smaller players: it is less costly for large players than for small players to stay away from their ideal allocations.

The diagonal of Q consists of positive numbers that capture the rates, at which all players sell their endowments. The off-diagonal terms of Q in each column i indicate how the sales of trader i become absorbed by other traders. Interestingly, flows are not absorbed proportionately to risk capacity. Rather, smaller traders absorb a disproportionately large portion of the flows, while large players wait and trade slowly.

As large traders trade more slowly and exercise market power, they also have a greater price impact, defind as the derivative of the price with respect to the *flow* from trader i, i.e. p^i/q^{ii}. In this example the vector of price impacts is

$$\begin{bmatrix} -0.403 & -0.341 & -0.301 & -0.272 & -0.251 \end{bmatrix}.$$

Larger players have a greater price impact in our model because the market can identify the source of trades. Larger players trade more slowly, and hide their true supply. When the market sees sales by a large player, it knows that these trades are just a tip of the iceberg: they expect the selling to continue for a long time and depress the price. In contrast, if the trade came from a small trader who is desperate to sell quickly, the price would fall a lot less for the same volume of trade. These observations explain why in practice market participants want to know the source of trade, and are more willing to trade against the flows of small traders than those of large traders.

The eigenvector decomposition of Q information about the rates at which different misallocations get traded away to efficiency. The eigenvector misallocations, together with the corresponding eigenvalues, are given by

$$U = \begin{bmatrix} 1 & 1 & .218 & .118 & .081 \\ .666 & -.530 & .782 & .214 & .123 \\ .5 & -.221 & -.619 & .668 & .213 \\ .4 & -.143 & -.234 & -.748 & .583 \\ .333 & -.106 & -.147 & -.252 & -1 \end{bmatrix}, \quad \mathrm{diag}\,\Lambda = \begin{bmatrix} 0 & 0.93 & 1.38 & 1.82 & 2.24 \end{bmatrix}.$$

Note the sign pattern in the eigenvectors: they break the market into two sides according to risk capacity, with one side of the market selling and the other, buying. Misallocations among the smallest players get traded away much faster than those among the largest players in the market.

We normalized the eigenvectors of Q so that one unit in total is misallocated in each one, and so that the larger players are sellers. The prices assigned to each

eigenvector are given by

$$\Pi = [\,-1\ \ .119\ \ .098\ \ .086\ \ .078].$$

When large players sell, they have market power to raise prices above first best. The price impact is the greatest in the eigenvector, in which the largest player alone sells to the rest of the market.

4   Trading between Large Players and a Fringe.

Existing literature, such as Vayanos (1999) and Du and Zhu (2012), has shown that in a market with finitely many large traders there is inefficiency as players trade slowly to the efficient allocation. However, in those models all traders are identical; prices do not depend on the allocation of assets and follow martingales.

Many other phenomena can happen in a market with heterogeneous traders, and our model is suitable for explaining these phenomena. Prices may have momentum due to trading between large and small traders. For example, a large seller will try to control the price by choosing an appropriate rate of selling, so that the price will have a downward drift. Price drift leads to other interesting phenomena, such as front-running. Players, who do not have any needs to buy or sell on their own, will attempt to identify sales by large traders in order to sell ahead of the price drop and buy back later. In this section, we illustrate these phenomena using a simplest version of our model: which captures trade between identical large players and a competitive fringe.

The benchmark case of interactions between one large trader and a competitive fringe, which we can solve in closed form, also provides useful bounds that shed a lot of light on dynamics in large markets with a low but positive Herfindahl index. Arguably, the case of a large number of market participants but not of perfect competition, is most relevant empirically. We would like to understand well the properties of these markets, particularly the speed of trade and price momentum. We finish by providing several examples of those markets and explaining how the dynamics in those markets can be understood through the prism of a special case with one large player and a fringe.

4.1   Equilibrium Equations with a Competitive Fringe.

So far, we analyzed a model with a finite number of traders. In this subsection we include a competitive fringe and derive the relevant equilibrium conditoins. We define a competitive fringe as a continuum of traders with a given finite risk capacity b^F. A group of m traders has risk capacity $1/b^F$ if each trader has quadratic disutility function with the same coefficient mb^F. Taking $m \to \infty$, we obtain a competitive fringe. We can include a competitive fringe into our model and, if so, we designate

trader N as the fringe. The common shock among the fringe members is denoted by $\sigma^N \, dZ_t^N$, and the disutility flow of the fringe is

$$-\frac{b^F}{2}(X^N)^2.$$

Individual fringe members may also experience idiosyncratic shocks, but since fringe members trade infinitely fast among each other, those shocks get diversified instantaneously and they do not affect the utility of the fringe.

The HJB equation for the utility $f^N(X)$ of the fringe is given by the same equation as the equation (11) for large traders. However, the first-order condition differs from (10), since individual fringe members can no longer affect the price with their individual actions.

Proposition 5 *If player N represents a competitive fringe, then prices and flows must satisfy the first-order condition*

$$P((r+2\delta)I + Q) + b^F 1^N = 0, \tag{20}$$

where 1^N denotes a row vector with 1 in the N-th position and zeros everywhere else.

Proof. The value function of an individual fringe member, whose allocation x may differ from the allocation of the fringe X^N, is given by $X^T A^N X + (x - X^N)PX + k^N$. Indeed, from symmetry, we know that the optimal strategy of the individual is to sell the excess allocation and align himself with the rest of the fringe. The trade generates income $(x - X^N)PX$.

The value function of the individual must satisfy the HJB equation

$$\max_x \; \frac{-b^F}{2}x^2 - r(X^T A^N X + (x - X^N)PX + k^N) + (PX)(\underline{Q}^N X)$$

$$-2X^T A^N(\delta X + QX) - \delta(x - X^N)(PX) - (x - X^N)P(\delta X + QX) + \sum_{j=1}^N a_{jj}^N(\sigma^j)^2 = 0,$$

where x denotes the individual's choice of asset holdings. The first-order condition is

$$-b^F x - P(rX + 2\delta X + QX) = 0.$$

Since the choice $x = X^N$ must be optimal, it follows that (20) must hold. ∎

4.2 Price Momentum: One Large Player and a Fringe.

We can immediately apply the condition of Proposition 5 to a market with a single large player and a fringe, and obtain a closed-form solution. The following proposition characterizes trading in a model between one large player with risk capacity $1/b_L$ and a competitive fringe with risk capacity $1/b_F$.

Proposition 6 *Consider a market with $N = 2$, in which player 1 is an individual large player and player 2 is a competitive fringe. Then in the unique nondegenerate linear stationary equilibrium, equilibrium prices and the players' trading rates are characterized by vectors*

$$P = -\frac{1}{r + 2\delta}\left[\frac{b_L b_F}{3b_F + b_L}, \frac{b_L b_F + 2b_F^2}{3b_F + b_L}\right], \quad Q = \frac{r + 2\delta}{2}\begin{bmatrix} b_L/b_F & -1 \\ -b_L/b_F & 1 \end{bmatrix}. \quad (21)$$

The welfare of the large trader and the fringe is characterized by matrices

$$A^L = \frac{b_F}{2(r + 2\delta)(3b_F + b_L)}\begin{bmatrix} -3b_L & -b_L \\ -b_L & b_F \end{bmatrix} \quad and$$

$$A^F = \frac{b_F}{2(r + 2\delta)(3b_F + b_L)(2b_F + b_L)}\begin{bmatrix} -(b_L^2 + 5b_L b_F + 5b_F^2) & -b_L b_F \\ -b_L b_F & b_L^2 \end{bmatrix}$$

Proof. See Appendix. ∎

Note that the second coefficient of P is more negative than the first coefficient. This leads to price momentum. Market price depends not only on the total endowment, but also its distribution between the large player and the fringe. A greater allocation to the fringe leads to a lower price.

If assets do not depreciate, i.e. $\delta = 0$, and in the absence of shocks, the initial allocation will converge to efficiency according to the equation

$$d\begin{bmatrix} X_t^L \\ X_t^F \end{bmatrix} = -\frac{r + 2\delta}{2}\frac{b_L + b_F}{b_F}\begin{bmatrix} X_t^L - \hat{X}_t^L \\ X_t^F - \hat{X}_t^F \end{bmatrix},$$

where $(\hat{X}_t^L, \hat{X}_t^F)$ is the first-best efficient allocaiton given by (4). The rate at which any misallocation gets traded away is given by the second eigenvalue of Q,

$$\frac{r + 2\delta}{2}\frac{b_L + b_F}{b_F}. \quad (22)$$

This is also the rate at which the price converges to the first-best price of

$$\hat{p}_t = -(X_t^L + X_t^F)\frac{b_L b_F}{b_L + b_F}.$$

We have

$$dp_t = -\frac{r + 2\delta}{2} \frac{b_L + b_F}{b_F} (p_t - \hat{p}_t).$$

The rate of trading and price convergence (22) decreases as the fringe becomes smaller relative to the large player. The rate of convergence varies from $r/2 + \delta$ when the fringe is small to infinity when the fringe is large (so that the "large" player is like any other member of the fringe). Of course, any misallocation within the fringe gets traded away instantaneously, since the fringe is competitive.

Now, price momentum leads to front-running. In the case of a large player trading against a fringe, imagine that the large player has one unit to sell while the fringe as a whole wants to buy on unit, so that $X_0 = [1, -1]^T$. However, imagine that the demand of one unit is not distributed uniformly and some of the fringe members in fact want neither to buy nor sell. Then, will these fringe members stay at their bliss endowment points, while the large trader sells to other fringe members? Certainly not: all fringe members will trade at time 0 to redistribute their endowments uniformly, and then they will buy from the large trader at proportionate rates. Effectively, the fringe members who start at their bliss point front-run the large trader who wants to sell. These players will sell assets short to other fringe members ahead of the large trader, while the price is high, and then buy back later from the large trader at a lower price.

We illustrate the process of front-running in more detail in the next subsection, using a model with identical large players trading against a fringe.

4.3 Front-Running: Many Large Players and a Fringe.

Consider a market with $N - 1$ large players with identical risk capacities $1/b_L$ and a competitive fringe with risk capacity $1/b_F$. The following proposition characterizes trading dynamics in this market.

Proposition 7 *In a market with $N - 1$ identical large traders and a fringe, trading towards the efficient allocation is characterized by the following eigenvector decomposition*

$$U = \begin{bmatrix} 1/b_L & 1 & \dots & 1 & 1 \\ 1/b_L & -1 & \dots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1/b_L & 0 & \dots & -1 & 1 \\ 1/b_F & 0 & \dots & 0 & 1 - N \end{bmatrix}, \quad \text{diag } \Lambda = [0 \ \lambda \ \dots \ \lambda \ \bar{\lambda}]. \quad (23)$$

The columns of U, eigenvector misallocations, are priced at

$$\Pi = \begin{bmatrix} -\dfrac{1}{r + 2\delta} & 0 & \dots & 0 & \pi^B \end{bmatrix}, \quad \text{where } \pi^B = \frac{(N-1)b_F}{r + 2\delta + \bar{\lambda}}.$$
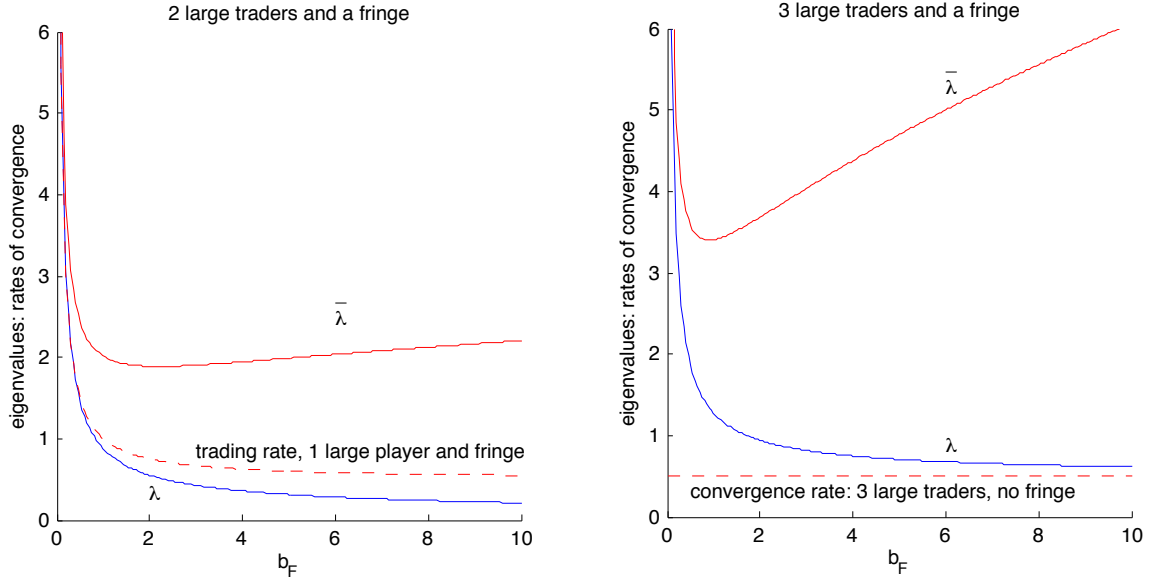
Figure 1: Convergence to efficiency: identical large traders and a fringe.

The rates of convergence to efficiency λ and $\bar{\lambda}$ satisfy equations

$$\left(\frac{\lambda b_F}{r + 2\delta} - \frac{\lambda \pi^B}{N - 1} - \frac{\bar{\lambda}}{N - 1} \frac{b_L - \pi^B \lambda}{r + 2\delta + \lambda + \bar{\lambda}} \right) \frac{b_L}{b_L + (N - 1)b_F} - \frac{\lambda b_L}{r + 2\delta + 2\lambda} \frac{N - 2}{N - 1} = 0,$$

$$\left(\frac{\bar{\lambda} b_F}{r + 2\delta} - \frac{\bar{\lambda} \pi^B}{N - 1} - \frac{\bar{\lambda}}{N - 1} \frac{b_L - 2\pi^B \bar{\lambda}}{r + 2\delta + 2\bar{\lambda}} \right) \frac{b_L}{b_L + (N - 1)b_F} - \frac{\lambda b_L}{r + 2\delta + \lambda + \bar{\lambda}} \frac{N - 2}{N - 1} -$$

$$\left(\lambda \frac{N - 2}{N - 1} + \frac{\bar{\lambda}}{N - 1} \frac{b_L}{b_L + (N - 1)b_F} - \frac{\lambda^2}{r + 2\delta + \lambda + \bar{\lambda}} \frac{N - 2}{N - 1} \right) \pi^B = 0.$$

Proof. See Appendix. ∎

From symmetry, any misallocation within the sector of large traders has no effect on the price. In contrast, any misallocation between large traders and the fringe leads to a price that is different from first best: when large traders are net sellers, the price will be above first best, as illustrated by the positive value of π^B.

Figure 1 illustrates the rates of convergence to efficiency in a market with 2 and 3 identical large traders and a fringe, where we set $r = 1$, $\delta = 0$, and $b_L = 1$. Trading between the large traders and the fringe always takes place at a faster rate of $\bar{\lambda}$ than the speed λ of trading within the group of large traders. The speed of trading is increasing in the number of market participants, but individual eigenvalues may be

nonmonotonic in the risk parameters of individual participants. In the right panel of Figure 3, as the fringe segment becomes small (i.e. as $b^F \to \infty$), the rate of trading among the large traders converges to the rate $(r + 2\delta)/2$ of trading in a market with 3 symmetric players.

Let us illustrate how front-running can happen in this model. Consider a market with two identical large traders and a fringe (i.e. $N = 3$, where the last player is the fringe). Suppose that player 1 wants to sell one unit, while the fringe wants to buy one unit. Player 2 wants neither to buy nor sell, so that $X_0 = [1 \ 0 \ -1]^T$. Then, because player 1 has market power against the fringe, he will sell slowly and charge a price which is above first best (5). This can be seen by the positive last component of Π, which prices the excess holdings of the large players relative to the fringe. Over time, the price will drift down and converge to first best.

Knowing this, of course, player 2 will not stay at his bliss point but rather sell short initially to the fringe, and buy back later. In other words, player 2 will front-run player 1. The mere presence of player 2 in the market completely changes the dynamics between player 1 and the fringe. Player 1 has a lot less control of the price, and so he sells much faster to the fringe. This can be illustrated by the comparison of $\bar{\lambda}$ with the dashed curve in the left panel of Figure 1.

In this example, the initial allocation can be decomposed into eigenvectors in the following way

$$\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \frac{1}{2}U^2 + \frac{1}{2}U^3 = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 1/2 \\ -1 \end{bmatrix}.$$

That is, the allocation $X_0 = [1 \ 0 \ -1]^T$ consists of an imbalance between the large players, and between large players and the fringe. The former gets traded to efficiency much more slowly than the latter, with the corresponding convergence rates $\lambda < \bar{\lambda}$. Thus, at the beginning player 2 will be primarily selling to the fringe, and buying only a little from player 1. The misallocation between the two large players persists a lot longer than the misallocation between the large traders and the fringe.

4.4 Large but Not Perfectly Competitive Markets.

It turns out that the closed-form characterization of the game between a single large player and a fringe (see subsection 4.2) can shed a great deal of light on behavior in large but not perfectly competitve markets.

Arguably, asset markets in practice are large but not perfectly competitive. Trades have price impact, which can be measured empirically. Large market participants spread trades over time to optimize the execution price. Market prices have momentum.

To give a flavor of what our model implies about these markets, we start by looking at a couple of numerical examples. Consider a market with infinitely many market
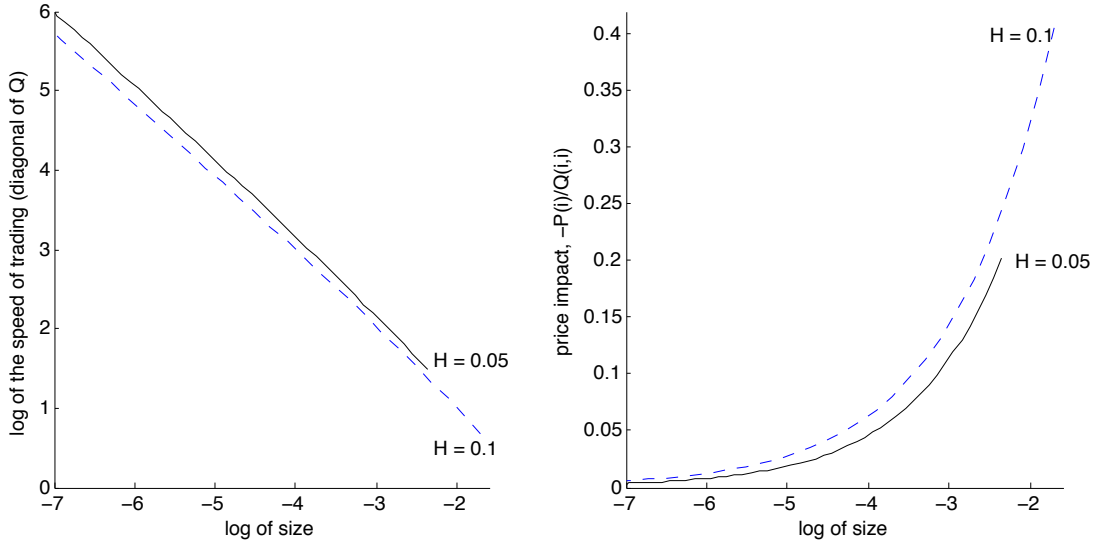
Figure 2: Speed of trading and price impact, as functions of size.

participants that have risk coefficients

$$b, \; b/x, \; b/x^2, \ldots$$

Then the total risk capacity of this market is given by $1/(b(1-x))$. We can normalize the total risk capacity to 1 by setting $b = 1/(1-x)$.

Furthermore, the Herfindahl Index of this market is given by

$$H = \sum_{n=0}^{\infty} \frac{x^{2n}}{b^2} = \frac{1-x}{1+x}.$$

For this market, Figure 2 shows the speed of trading and price impact in this market for $H = 0.1$ and 0.05, and $r + 2\delta = 1$. The left panel plots the logarithm of the corresponding diagonal element of Q against the logarithm of size, measured as the fraction of the whole market. Smaller players trade a lot faster than large traders. However, the speed of trading is not hugely sensitive to the Herfindahl Index: as the index moves half-way to zero, the larger players in the market trade only about 15% faster. It is natural to ask the question: will these players trade infinitely fast as $H \to 0$, or is there a specific upper bound on the speed of trading?

The right panel plots the price impact p^i/q^{ii} against the logarithm of size. Smaller traders have a lot less price impact, because their trades are less "toxic." Moreover, as H falls, players of the same size trade faster and their price impact falls. If a player sells 1% faster, it means that for the same flow, the hidden supply is less by about 1%. In this example, as H falls from 0.1 to 0.05, the price impact of the larger players is about 17% greater.
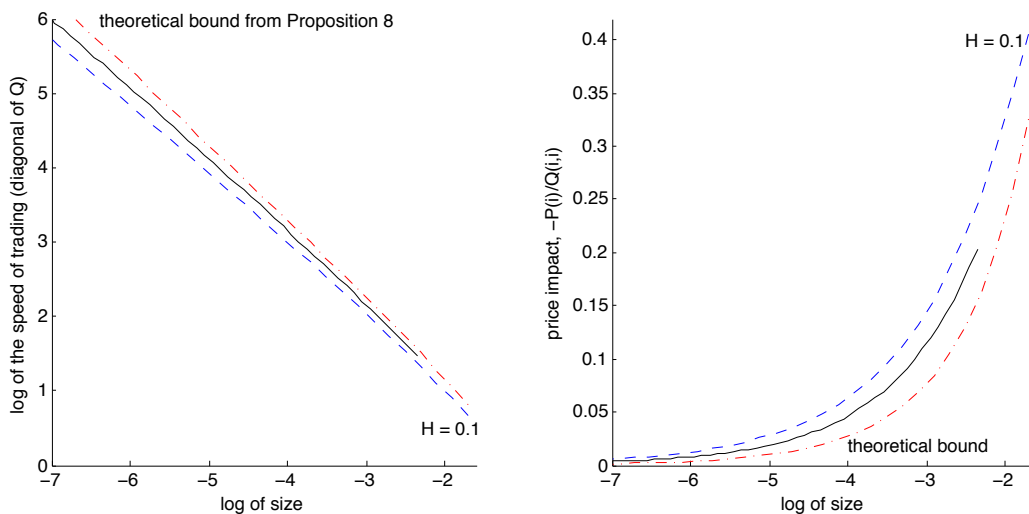
23

Figure 3: Speed of trading and price impact, together with theoretical bounds.

Is there any pattern? It is natural to guess that the market between a single large player and a competitive fringe can give us excellent guidance as to what goes on. In particular, it can provide an upper bound on the speed of trading as well as a lower bound on price impact for a player of any size.

Proposition 8 *In a market with one large player and a competitive fringe, if the size of the large player is x as a fraction of the entire market, then the trading speed of the large player is given by*

$$Q^{11} = \frac{r + 2\delta}{2} \frac{1 - x}{x}$$

and the price impact is given by

$$-P^1/Q^{11} = \frac{2}{(r + 2\delta)^2} \frac{x}{(1 - x)(1 + 2x)}.$$

Proof. If the large player is fraction x of the market, and the total risk capacity of the market is 1, then the corresponding risk coefficients are $b^L = 1/x$ and $b^F = 1/(1 - x)$. Using these in conjunction with Proposition 6, we obtain the desired expressions. ∎

Figure 3 superimposes the theoretical bounds implied by Proposition 8 onto the example in Figure 2. We observe that the bounds provide good estimates of the speed trading and price impact in markets with many participants, which are nevertheless not perfectly competitive. We estimate that in practice, markets for listed equities and options have value of H a lot closer to 0 than those from this example, and so the approximate estimate would be even more precise.

24

5 Welfare.

We can use our model to analyze welfare in the market. In particular, we can study how welfare depends on the amount of competition in the market, the number of players, asymmetries in the market, and potentially market design. While we consider a particular trading mechanism, there are many others - we can study the welfare of different mechanisms, given the players' preferences and information, address the question of optimal mechanism design and think about natural market implementations of the optimal mechanism.

This section focuses on the effects of mergers and entry on welfare. Mergers have costs in our model, as they lead to slower trade due to decreased competition. This leads to inefficiencies, as players have to wait longer to trade shocks. At the same time, when players merge, then there is the obvious benefit of diversification, assuming that the merged players can perfectly share risks within the unit after the merger. Had the players not merged, the risks that they would otherwise diversify within the unit would need to be traded in the open market, with delay.

To do proper welfare experiments, we have to treat the risk coefficients as well as the volatilities of shocks properly. When players j and k merge, the risk capacity of the merged unit has to be the sum of the risk capacities of individual players. That is, the risk coefficient of the unit b satisfies $1/b = 1/b^j + 1/b^k$. Then, taking as given the total endowment of the two players, the sum of individual utilities under first-best sharing of endowments must equal to the utility of the merged unit that holds the total endowment. Likewise, if individually players face shocks with volatilities σ^j and σ^k and correlation ρ_{jk}, then the volatility of shock to the unit is

$$\sqrt{(\sigma^j)^2 + (\sigma^k)^2 + 2\rho_{jk}\sigma^j\sigma^k}.$$

The correlations between the shock to the unit and shocks to the endowments of the remaining players must also be computed appropriatedly.

We begin with a surprising result that in symmetric markets, symmetric mergers have no effect on total welfare. This result holds even when shocks to individual players' endowments are correlated.

Proposition 9 *Consider a market with $N = 2n$ symmetric players, in which each player has risk coefficient $b = 2\beta$ and faces shocks with volatility σ. Shocks to endowments of any two different players j and k have correlation ρ. Then the equilibrium utility of any player at time 0 (before any shocks are realized) is given by*

$$k^i = -\frac{1}{r(r+2\delta)}\frac{b}{2}\frac{2+\rho(N-2)}{N}(\sigma^i)^2. \tag{24}$$

If the players merge in pairs, so that n units with risk coefficients β appear, then assuming perfect diversification of risks within the unit, the equilibrium utility is now

$$-\frac{2}{r(r+2\delta)}\frac{\beta}{2}\frac{2+\rho(2n-2)}{n}(\sigma^i)^2 \tag{25}$$

per unit. The total welfare in the market does not change with merger.

Proof. First, let us confirm (24). According to (11), $rk^i = \sigma^T(A^i \circ \mathcal{R})\sigma$. From Proposition 3, matrix $A^i \circ \mathcal{R}$ takes the form

$$\frac{1}{r+2\rho}\begin{bmatrix} \ddots & -\rho\frac{b}{2}\frac{N-2}{N^2} & \cdots & \rho\frac{b}{2}\frac{N-2}{(N-1)N^2} \\ -\rho\frac{b}{2}\frac{N-2}{N^2} & -\frac{b}{2}\frac{3N-2}{N^2} & -\rho\frac{b}{2}\frac{N-2}{N^2} & -\rho\frac{b}{2}\frac{N-2}{N^2} \\ \vdots & -\rho\frac{b}{2}\frac{N-2}{N^2} & \ddots & \rho\frac{b}{2}\frac{N-2}{(N-1)N^2} \\ \rho\frac{b}{2}\frac{N-2}{(N-1)N^2} & -\rho\frac{b}{2}\frac{N-2}{N^2} & \rho\frac{b}{2}\frac{N-2}{(N-1)N^2} & \frac{b}{2}\frac{N-2}{(N-1)N^2} \end{bmatrix}.$$

Multiplying by σ on both sides, we obtain (24).

Now, when pairs of players merge, then the variance of the shocks that each unit faces is $2(1+\rho)(\sigma^i)^2$. The correlation between the shocks of different units is $2\rho/(1+\rho)$. Thus, the welfare of each unit is now

$$-\frac{1}{r(r+2\delta)}\frac{\beta}{2}\frac{2+\frac{2\rho}{1+\rho}(n-2)}{n}2(1+\rho)(\sigma^i)^2 = -\frac{2}{r(r+2\delta)}\frac{\beta}{2}\frac{2+\rho(2n-2)}{n}(\sigma^i)^2,$$

which confirms (25). ∎

Proposition 9 implies that there are no obvious reasons why mergers in our model may be beneficial or detrimental for all players.[5] However, mergers by some of the players can have mixed effects on the welfare of everyone else as well as within the merged group. These effects depend on market power as well as whether different players experiencing small shocks, and thus providing liquidity, or demanding liquidity. Some of the welfare effects may appear counterintuitive at first. We present several interesting examples in the next subsection. In all of the examples, we assume that shocks are independent and normalize $r + 2\delta = 1$.[6]

5.1 Examples.

We start with a basic question: does market power really help players? While players with market power can control the rate of trading, in order to get a more favorable price from the rest of the market, they are also punished by greater sensitivity of prices to flows as the rest of the market anticipates this behavior. Our first example is a market with a large player and a fringe, which have identical risk coefficients $b^L = b^F = 1$ and face identical shocks $[\sigma^L, \sigma^F] = [1, 1]$. In this case the equilibrium utilities of the large player and the fringe are given by
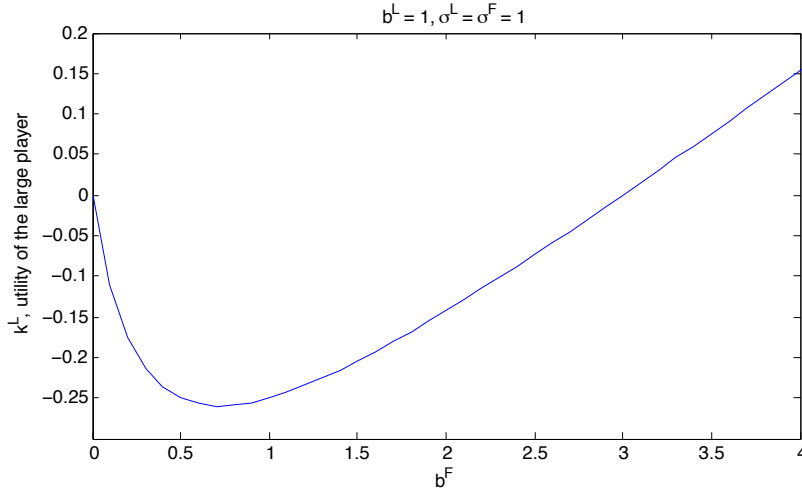
$$[k^L, k^F] = [-0.25, -0.333].$$

Figure 4: Utility of the large player as a function of b^F.

Clearly, market power helps in this example.

Next, we explore how the welfare of the large player depends on the size of the fringe. As we vary b^F in the example above we find that

$$k^L = \frac{(b^F - 3)b^F}{2(3b^F + 1)}.$$

As seen in Figure 4, this funciton is non-monotonic. The large player gets utility 0 when the fringe is large (i.e. $b^F = 0$) as he can offload any idiosyncratic exposure costlessly. This becomes harder as the fringe gets smaller and utility becomes negative. At some point, when the fringe becomes sufficiently small, and desperate to trade as $\sigma^F = 1$, the utility of the large player starts rising and eventually becomes positive. The large player can make profit by trading with the fringe.

Mergers. Let us explore the effects of mergers on the welfare of different players. In the following examples, we start with a market containing a large player with $b^L = 1$ (risk capacity 1) and a fringe with $b^F = 1/2$ (risk capacity 2). Then we merge half of the fringe members to form another large player (hedge fund, H) with risk capacity 1.

If the variances of the shocks before the split are given by $[(\sigma^L)^2, (\sigma^F)^2] = [1, 1]$, then the equilibrium payoffs are

$$[k^L, \ k^F] = [-0.25, \ -0.1875]. \tag{26}$$

---

[5]Proposition 9 assumes that shocks can be fully diversified within the merged unit. If they are not, then mergers would be clearly detrimental in the symmetric model.

[6]Also, when dealing with the fringe, we evaluate its welfare by the formula from the proof of Proposition 6, which assumes perfect risk sharing among fringe members. This is inconsistent with the limit taken in Proposition 9, in which the utility of N players does not converge to first best.

After the formation of the hedge fund, the variances of shocks are $[(\sigma^L)^2, (\sigma^H)^2, (\sigma^F)^2] = [1, 1/2, 1/2]$, and the equilibrium payoffs are

$$[k^L,\ k^H,\ k^F] = [-0.2786, -0.1030, -0.1459].$$

Here, the formation of the hedge fund is bad for everyone, as the utilities of both the fund and the remaining fringe are less than half of the utility of the large fringe prior to merger.

Surprisingly, the fringe as a whole does not need to be worse off as a result of the formation of a hedge fund. Moreover, the hedge fund does not need to be better off than the rest of the fringe, even though both face identical shocks but the hedge fund has market power. Suppose that, in the example above, $[(\sigma^L)^2, (\sigma^F)^2] = [1, 0]$ before the split and $[(\sigma^L)^2, (\sigma^H)^2, (\sigma^F)^2] = [1, 0, 0]$ after. Then the payoffs are

$$[k^L,\ k^F] = [-0.3, 0.05] \tag{27}$$

before the split and

$$[k^L,\ k^H,\ k^F] = [-0.3275, 0.0237, 0.0297].$$

Fringe members compete to provide liquidity to the large player and they get positive payoff only because their risk capacity is bounded. When the hedge fund splits off, it may look surprising that the hedge fund, with its market power, gets a smaller payoff than the fringe. The reason is that both are competing to provide liquidity to the large player, and the heldge fund - with its market power - absorbs the flow from the large player more slowly. The remaining fringe members, of course, free ride.

High-frequency Trading. In our last set of examples, rather than keeping the set of market participants constant, we consider what happens when we allow new players to enter. Specifically, in the examples above with $[b^L,\ b^F] = [1, 1/2]$, we consider the entry of a second large player with risk parameter $b^2 = 1$. The entrant has no individual need to trade as $\sigma^2 = 0$. He only provides liquidity, so we interpret the entrant as a high-frequency trader.

First, if $[(\sigma^L)^2, (\sigma^F)^2] = [1, 1]$ then the entrant changes the vector of utilities from (26) to

$$[k^L,\ k^2,\ k^F] = [-0.2298, 0.0497, -0.1765].$$

The utilities of both the large player and the fringe rise with the entry of the high-frequency trader. While the latter effect confirms our intuition, the former may seem surprising. Conventional wisdom holds that high-frequency traders hurt large institutional investors. What happens here is that while the entrant can front-run the large player, he also changes the entire equilibrium dynamics so that trade is faster. This, of course, benefits the large player.

On the other hand, if $[(\sigma^L)^2, (\sigma^F)^2] = [0, 1]$, i.e. the large player is a liquidity provider, then the entrant obviously hurts the large player. In this case, welfare before entry is given by

$$[k^L,\ k^F] = [0.05, -0.2375],$$

and after entry,

$$[k^L, \ k^2, \ k^F] = [0.0335, \ 0.0335, \ -0.2132].$$

The fringe unambiguously benefits from a competing liquidity provider.

6   A Microfoundation of Quadratic Preferences.

In this section we microfound our model with quadratic preferences by laying out a more natural model with exponential utilities, in which players trade to hedge private shocks that expose them to a common risk factor. We show that the equilibrium equations of the linear-quadratic model match those of the exponential model in the special case when the shocks that expose players to the common risk factor become small. In this sense, the exponential model is more general, but the linear-quadratic model provides a clean special case as the equilibrium dynamics, characterized by the pair (P, Q), depend only on the players' risk capacities and not the sizes of shocks that individual players receive, or the correlation among shocks.

We also extend the model to also allow the shocks to carry information about a common component of value. We confirm the result of Du and Zhu (2013) that is symmetric markets, as the players get more information about common fundamentals, the speed of trade slows down. In general asymmetric markets, equilibrium in this more general setting is characterized by the same set of equations with only one extra term.

6.1   The Exponential Model.

Consider a model in which all players $i = 1, \ldots N$ have exponential utility

$$- \exp(-\alpha^i c_t),$$

where $\alpha^i > 0$ is the coefficient of absolute risk aversion. Players consume continuously and have a common discount rate r, which is also the risk-free rate in the market.

Players have private information about their risk exposure X_t^i to a common Brownian risk factor dW_t. Risk exposure depreciates at rate δ and changes due to shocks $\sigma^i \, dZ_t^i$ for player i, where $Z = [Z_t^1, \ldots Z_t^N]$ is a vector of Brownian motions with the correlation matrix $\mathcal{R}$, but independent of W_t. Risk exposures can also be traded in the market. We consider a linear equilibrium, in which players announce their risk exposures, and given a vector of announcements $\tilde{X}_t$, the trading flows are given by $Q\tilde{X}_t$, and the market price is given by $P\tilde{X}_t$. Then the risk exposures follow

$$dX_t = -\rho X_t \, dt + \sigma dZ_t - Q\tilde{X}_t \, dt$$

and the wealth of agent i follows

$$dw_t^i = (rw_t^i - c_t^i) \, dt + (P\tilde{X}_t)(\underline{Q}^i \tilde{X}_t) \, dt + X_t^i \, dW_t,$$

where c_t^i is the consumption of player i.

Conjecture that the equilibrium value function of player i takes the form

$$-\frac{1}{r}\exp(-r\alpha^i\underbrace{(w_t^i + X_t^T A^i X_t + k^i)}_{v_t^i}). \tag{28}$$

Then

$$dv_t^i = (rw_t^i - c_t^i)\,dt + (P\tilde{X}_t)(\underline{Q}^i\tilde{X}_t)\,dt + X_t^i\,dW_t$$

$$+2X_t^T A^i(-\rho X_t\,dt + \sigma dZ_t - Q\tilde{X}_t\,dt) + \sigma^T(A^i \circ \mathcal{R})\,\sigma\,dt.$$

In order to write down the HJB equation for player i, we must consider $\tilde{X}_t^i$ of the form $X_t + 1^i y$, where 1^i is the i-th coordinate vector and y is the amount by which player i lies.

Then the HJB equation of player i is

$$-\exp(-r\alpha^i v^i) = \max_{c,\,\hat{X}=X+1^i y} -\exp(-\alpha^i c)$$

$$+\alpha^i\exp(-r\alpha^i v^i)\left(rw^i - c^i + (P\tilde{X})(\underline{Q}^i\tilde{X}) - 2X^T A^i(\rho X + Q\tilde{X}) + \sigma^T(A^i \circ \mathcal{R})\,\sigma\right)$$

$$-\frac{r(\alpha^i)^2}{2}\exp(-r\alpha^i v^i)\left(4X^T A^i \Sigma\,\mathcal{R}\,\Sigma A^i X + (X^i)^2\right),$$

where Σ is the diagonal matrix with the elements of σ on the diagonal. The term $4X^T A^i \Sigma\,\mathcal{R}\,\Sigma A^i X$ is the incremental variance of v_t^i from the volatility of the entire vector X_t.[7]

The first-order condition with respect to c is

$$\exp(-\alpha^i c) = \exp(-r\alpha^i v^i) \quad \Leftrightarrow \quad -c = -r(w^i + X^T A^i X + k^i).$$

Given this, the HJB equation simplifies to

$$0 = \max_{\hat{X}=X+1^i y} -r(X^T A^i X + k^i) + (P\tilde{X})(\underline{Q}^i\tilde{X}) - 2X^T A^i(\rho X + Q\tilde{X}) + \sigma^T(A^i \circ \mathcal{R})\,\sigma$$

$$-\frac{r\alpha^i}{2}\left(4X^T A^i \Sigma\,\mathcal{R}\,\Sigma A^i X + (X^i)^2\right). \tag{29}$$

Separating the first-order condition, we obtain matrix equations that characterize stationary linear equilibria in this model. We summarize them in the following proposition.

---

[7]This expression assumes that A^i is symmetric, otherwise the second instance of A^i would need to be replaced with $(A^i)^T$.

Proposition 10 *Stationary linear equilibria of the exponential model are characterized by the equations*

$$P^i \underline{Q}^i + Q^{ii} P = 2(A^i Q^i)^T, \quad rk^i = \sigma^T (A^i \circ \mathcal{R}) \, \sigma, \tag{30}$$

$$and \quad A^i((r + 2\rho)I + 2Q) \sim P^T \underline{Q}^i - \frac{r\alpha^i}{2} 1^{ii} - 2r\alpha^i A^i \Sigma \, \mathcal{R} \, \Sigma A^i. \tag{31}$$

Proof. Equation (29) must hold for all vectors $X \in \mathbb{R}^N$. To ensure that, the coefficients on the constant term as well as the terms of the form $X^j X^k$ must match, and the first-order condition with respect to y must hold at $y = 0$. From those conditions, we obtain (30) and (31). ■

The system of (30) and (31) is different from equations (10) and (11) in the linear-quadratic model only in the term $2r\alpha^i A^i \Sigma \, \mathcal{R} \, \Sigma A^i$. Parameter b^i in the linear-quadratic model corresponds to $r\alpha^i$ in the exponential model, i.e. it reflects the players' capacities to wait and absorb risk waiting for a better price to hedge at. In the limit as $\sigma \to 0$, the equations in the exponential model become identical to those in the linear-quadratic model. This, the linear-quadratic model is a special case of the exponential model. We summarize this finding in the following proposition.

Proposition 11 *Any solution of the linear-quadratic model solves equations (30) and (31) in the limit as $\sigma \to 0$.*

Proof. The conclusion follows immediately, since the term that distinguishes the two sets of equations converges to 0 as $\sigma \to 0$. ■

Even though the exponential case is more general, the linear-quadratic model provides a much cleaner picture of equilibrium dynamics, as the equilibrium equations depend only on the players' risk capacities and not the distribution of shocks. This makes our benchmark case particularly attractive. Nevertheless, in order to provide a more complete picture, we present a couple of computed examples for the general case at the end of this section.

6.2 Extension to Private Information about Fundamentals.

The explicit exponential model makes it clear how we can include private information about fundamentals, i.e. future cash flows to the traded asset. For simplicity, we assume that players learn about fundamentals from the same shocks Z_t^i that affect their individual preferences.[8] We also assume that the signals of all players are

---

[8]If players had learned about fundamentals and their individual preferences from different signals, other market participants would face a filtering problem when figuring out whether trades are motivated by private or common values. This would lead to a more difficult problem, which is important for future research, but beyond the scope of current paper.

equally informative about fundamentals, so that the total supply of the asset $\bar{1}^T X_t$ is a sufficient statistic for all available information about fundamentals, where $\bar{1}$ is a column vector with all coefficients equal to 1.

To be concrete, suppose that the rate of change of the value of the asset is given by

$$dW_t - \kappa\,\bar{1}^T X_t\,dt,$$

where W_t is a Brownian motion the information of all market participants and $\bar{1}^T X_t$ is the total supply of the asset. In particular, if $\kappa > 0$ then whenever any player gets a shock that increases that player's exposure to W_t, the shock also carries bad information about the payoff from holding the asset. The wealth of player i, given his exposure X_t^i and consumption c_t^i, has to follow

$$dw_t^i = (rw_t^i - c_t^i)\,dt - (P\tilde{X}_t)(\underline{Q}^i \tilde{X}_t)\,dt + X_t^i\,(dW_t - \kappa\,1^T X_t\,dt).$$

Maintaining all other assumptions of subsection 6.1, we conjecture value functions of the form (28). Then

$$d\underbrace{(w_t + X_t^T A^i X_t + k^i))}_{\tilde{w}_t^i} = (rw_t^i - c_t^i)\,dt - (P\tilde{X}_t)(\underline{Q}^i \tilde{X}_t)\,dt + X_t^i\,(dW_t - \kappa\,1^T X_t\,dt)$$

$$+2X_t^T A^i(-\rho X_t\,dt + \sigma dZ_t - Q\tilde{X}_t\,dt) + \sigma^T(A^i \circ \mathcal{R})\,\sigma\,dt$$

and, through an analogous sequence of steps, the HJB equation (29) is reduced to

$$A^i((r + 2\rho)I + 2Q) + 2r\alpha^i A^i \Sigma\,\mathcal{R}\,\Sigma A^i \sim -P^T\underline{Q}^i - \frac{1}{2}r\alpha^i 1^{ii} - \kappa\,\underline{1}^i, \qquad (32)$$

where $\underline{1}^i$ is a matrix with ones in the i-th row, and zeros everywhere else. Equations (30) remain the same.

A common-value component can also be included in our linear-quadratic model if we generalize the payoff flow that each player receives (3) to

$$-\frac{b^i}{2}(X_t^i)^2/2 - \kappa\bar{1}^T X_t. \qquad (33)$$

In this case the equilibrium equations are given by (30) as well as (32) with the last term of the left-hand side removed (or with Σ set to 0). For the linear-quadratic model with a common-value component, we are able to characterize the equilibrium in a symmetric market in closed form, extending Proposition 3.

Proposition 12 *In the linear-quadratic model, if all players have identical risk parameters given by $B = [b, b, \ldots b]$, then a symmetric non-degenerate equilibrium exists*

whenever the common-value component $\kappa \in (-b/N, (N-2)b/N)$. In this case, the price of the asset is always first-best and given by

$$P = - \left[\frac{b/N + \kappa}{r + 2\delta}, \frac{b/N + \kappa}{r + 2\delta}, \dots \frac{b/N + \kappa}{r + 2\delta} \right]$$

and trading dynamics are characterized by the matrix

$$Q = \frac{q}{N} \begin{bmatrix} N-1 & -1 & \dots & -1 \\ -1 & N-1 & \dots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \dots & N-1 \end{bmatrix}, \quad with \;\; q = \frac{r + 2\delta}{2} \frac{\frac{N-2}{N}b - \kappa}{\frac{b}{N} + \kappa}.$$

Proof. See Appendix. ∎

This proposition confirms the result of Du and Zhu (2013) that as the common-value component of individual signals increases, trade in equilibrium slows down.

6.3   Examples.

We revive the example from subsection 3.3 to explore how the extra term in (31) affects prices and the rates of trading. Recall that the risk coefficients are $[b^1, b^2, b^3, b^4, b^5] = [1, 1.5, 2, 2.5, 3]$ in that example. We set $r = 0.05$ and $\delta = 0.475$ so that $r + 2\delta = 1$, and the coefficients of absolute risk aversion to $[\alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5] = [20, 30, 40, 50, 60]$ to match that example. Assume that $\mathcal{R} = I$, i.e. shocks to individual players are uncorrelated.

Then, if $\sigma = [0.1, 0.1, 0.1, 0.1, 0.1]^T$, we have

$$P = [-.257, \; -.334, \; -.394, \; -.443, \; -.485].$$

The price sensitivities to the allocations of all players increase slightly. Trading dynamics are now characterized by

$$Q = \begin{bmatrix} 0.619 & -0.243 & -0.318 & -0.388 & -0.454 \\ -0.161 & 0.953 & -0.324 & -0.400 & -0.472 \\ -0.157 & -0.242 & 1.278 & -0.403 & -0.479 \\ -0.153 & -0.237 & -0.320 & 1.589 & -0.481 \\ -0.147 & -0.231 & -0.315 & -0.398 & 1.886 \end{bmatrix}.$$

The speed of trading slows down somewhat, but qualitatively and quantitatively the solution looks similar to our baseline model.

Now, consider $\sigma = [0.1, 0.3, 0.3, 0.1, 0.1]^T$. We raise the fundamental needs to trade of players 2 and 3, while keeping shocks to everyone else the same. Now, players 1, 4

and 5 can provide liquidity to players 2 and 3, and help them share risks. Then the trading dynamics are characterized by the price vector

$$P = [-.265, \ -.358, \ -.426, \ -.463, \ -.507]$$

and the trading matrix

$$Q = \begin{bmatrix} 0.572 & -0.246 & -0.326 & -0.374 & -0.437 \\ -0.146 & 0.951 & -0.322 & -0.374 & -0.445 \\ -0.131 & -0.222 & 1.312 & -0.364 & -0.439 \\ -0.150 & -0.244 & -0.334 & 1.500 & -0.467 \\ -0.145 & -0.239 & -0.330 & -0.387 & 1.787 \end{bmatrix}.$$

The price impact of shocks rises and trade slows down, especially for players who are hit by relatively smaller shocks.

These examples seem to imply that the more general model with exponential utility does not add much intuition about market dynamics on top of what the baseline linear-quadratic model already tells us. Of course, there may be interesting effects that we are overlooking.

7 Conclusion

To be completed.

Appendix

Proof of Proposition 1. We have to prove that the truth-telling strategy maximizes the utility of any player i. For an arbitrary strategy $\{y_t, t \geq 0\}$, which specifies the misrepresentation y_t of player i's allocation for any history $\{X_s, s \in [0,t]\}$ of allocations, consider the process

$$G_t = \int_0^t e^{-rs} \left((PX_s + p^i y_s)(\underline{Q}^i X_s + q^{ii} y_s) - \frac{b^i}{2}(X_s^i)^2 \right) ds + e^{-rt} f^i(X_t).$$

Then the conditions $p^i < 0$ and $q^{ii} > 0$ ensure that $y_t = 0$ maximizes the drift of G_t, and (9) ensures that the maximal drift of G_t equals 0. That is, the process G_t is always a supermartingale, and a martingale under the truth-telling strategy.

Now, since the process X defined by (8) us nonexplosive, it follows that

$$E[e^{-rt} f^i(X_t)] \to 0$$

as $t \to 0$ when player i, as well as everybody else, follow the truthtelling strategies. Therefore, player i's expected payoff under the truthtelling strategy is

$$E\left[\int_0^\infty e^{-rs} \left((PX_s)(\underline{Q}^i X_s) - \frac{b^i}{2}(X_s^i)^2 \right) ds \right] = E[G_\infty] = G_0 = f^i(X_0).$$

Consider any alternative strategy $\{y_t, t \geq 0\}$ that satisfies the no-Ponzi condition $E[e^{-rt} X_t^2] \to 0$ as $t \to 0$. Then for any quadratic value function $f^i(X)$, $E[e^{-rt} f^i(X_t)] \to 0$ as $t \to 0$. It follows then that player i's payoff under this strategy is

$$E\left[\int_0^\infty e^{-rs} \left((PX_s + p^i y_s)(\underline{Q}^i X_s + q^{ii} y_s) - \frac{b^i}{2}(X_s^i)^2 \right) ds \right] = E[G_\infty] \leq G_0 = f^i(X_0).$$

Thus, truth-telling is optimal. This completes the proof of Proposition 1. ∎

Proof of Proposition 3. In a symmetric model with $b^i = b$, a symmetric mechanism (P, Q) has the trade-flow trading matrix:

$$Q = \begin{bmatrix} q & -q\frac{1}{N-1} & -q\frac{1}{N-1} \\ -q\frac{1}{N-1} & q & -q\frac{1}{N-1} \\ -q\frac{1}{N-1} & -q\frac{1}{N-1} & q \end{bmatrix}$$

and price vector $P = [p, ..., p]$. In other words, a symmetric mechanism is characterized by two parameters, q and p. Moreover, the value function of any trader depends only on own holdings and total holdings of others, $X^{-i} = \sum_{j \neq i} X^j$:

$$rf^i\left(X^i, X^{-i}\right) = \max_Y -\frac{b}{2}\left(X^i\right)^2 + p\left(Y + X^{-i}\right)\left(qY - \frac{q}{N-1}X^{-i}\right) + E\frac{df^i\left(X^i, X^{-i}|Y\right)}{dt} \tag{34}$$

Recall that we guessed $f^i = X^T A^i X + k^i$, which in the symmetric model simplifies to:

$$f^i \left(X^i, X^{-i} \right) = k + a_{11} \left(X^i \right)^2 + 2a_{12} X^i X^{-i} + a_{22} \left(X^{-i} \right)^2$$

The change of continuation payoff due to trade is then:

$$E \frac{df^i \left(X^i, X^{-i} | Y \right)}{dt} = 2a_{11} X^i E \left[\dot{X}^i | Y \right] + 2a_{12} \left(X^{-i} E \left[\dot{X}^i | Y \right] + X^i E \left[\dot{X}^{-i} | Y \right] \right) + 2a_{22} X^{-i} E \left[\dot{X}^{-i} | Y \right] + C$$

Since holdings change according to:

$$E \left[\dot{X}^i | Y \right] = -\delta X^i - q \left(Y - \frac{X^{-i}}{N-1} \right)$$

$$E \left[\dot{X}^{-i} | Y \right] = -\delta X^{-i} + q \left(Y - \frac{X^{-i}}{N-1} \right)$$

we get:

$$E \frac{df^i \left(X^i, X^{-i} | Y \right)}{dt} = -2 \left(a_{11} X^i + a_{12} X^{-i} \right) \left(\delta X^i + q \left(Y - \frac{X^{-i}}{N-1} \right) \right)$$
$$-2 \left(a_{22} X^{-i} + a_{12} X^i \right) \left(\delta X^{-i} - q \left(Y - \frac{X^{-i}}{N-1} \right) \right) + C$$

Plugging it back to the optimization problem of reporting X^i, we obtain the following FOC:

$$pq \left(2Y + X^{-i} - \frac{X^{-i}}{(N-1)} \right) - 2 \left(a_{11} X^i + a_{12} X^{-i} \right) q + 2 \left(a_{22} X^{-i} + a_{12} X^i \right) q = 0$$

Evaluated at truth-telling it becomes (after collecting terms with X^i and X^{-i}):

$$\left(pq \frac{N-2}{N-1} - 2qa_{12} + 2qa_{22} \right) X^{-i} + \left(2pq - 2qa_{11} + 2qa_{12} \right) X^i = 0$$

Since we require that the mechanism be ex-post incentive compatible, the FOC has to hold for all X^i, X^{-i}, that is:

$$pq \frac{N-2}{N-1} - 2qa_{12} + 2qa_{22} = 0$$
$$pq - qa_{11} + qa_{12} = 0$$

Finally, matching the coefficients of the value function we get:

$$r\left(a_{11}\left(X^i\right)^2 + 2a_{12}X^iX^{-i} + a_{22}\left(X^{-i}\right)^2\right)$$

$$= -\frac{b}{2}\left(X^i\right)^2 + p\left(X^i + X^{-i}\right)\left(qX^i - \frac{q}{N-1}X^{-i}\right)$$

$$-2\left(a_{11}X^i + a_{12}X^{-i}\right)\left(\delta X^i + q\left(X^i - \frac{X^{-i}}{N-1}\right)\right)$$

$$-2\left(a_{22}X^{-i} + a_{12}X^i\right)\left(\delta X^{-i} - q\left(X^i - \frac{X^{-i}}{N-1}\right)\right)$$

Matching up coefficients, brings the whole system with unknowns $(p, q, a_{11}, a_{12}, a_{22})$ to

$$pq\frac{N-2}{N-1} - 2qa_{12} + 2qa_{22} = 0$$

$$pq - qa_{11} + qa_{12} = 0$$

$$ra_{11} = -\frac{1}{2}b + 2qa_{12} - 2a_{11}(q+\delta) + pq$$

$$r2a_{12} = 2qa_{22} - 2a_{12}\left(2\delta + q\frac{N}{N-1}\right) + 2q\frac{a_{11}}{N-1} + pq\frac{N-2}{N-1}$$

$$ra_{22} = q\frac{2a_{12}}{N-1} - 2a_{22}\left(\delta + \frac{q}{N-1}\right) - p\frac{q}{N-1}$$

This system has two solutions: a degenerate one (i.e., $q = 0$ and no trade) and a regular one:

$$q = (r+2\delta)\frac{(N-1)(N-2)}{2N}$$

$$p = -\frac{1}{N}\frac{b}{r+2\delta}$$

$$A^i = \frac{-b}{2(r+2\delta)N^2}\begin{bmatrix} 3N-2 & N-2 \\ N-2 & -\frac{N-2}{N-1} \end{bmatrix}$$

Given this solution, price at time t is:

$$p_t = PX_t = -\frac{1}{N}\frac{b}{r+2\delta}\bar{X}_t = -\frac{1}{(r+2\delta)\bar{\beta}}\bar{X}_t$$

which is indeed the efficient price (5).

The efficient allocation is $\frac{\bar{X}_t}{N}$ and individual holdings evolve according to:

$$E \frac{d\left(X_t^i - \frac{\bar{X}_t}{N}\right)}{dt} = -\delta \left(X_t^i - \frac{\bar{X}_t}{N}\right) - q\left(X_t^i - \frac{X_t^{-i}}{N-1}\right)$$

$$= -\left(\delta + q\frac{N}{N-1}\right)\left(X_t^i - \frac{\bar{X}_t}{N}\right)$$

$$= -\left(\delta + (r+2\delta)\frac{(N-2)}{2}\right)\left(X_t^i - \frac{\bar{X}_t}{N}\right)$$

Hence trade contributes exponential rate of convergence $(r+2\delta)\frac{(N-2)}{2}$, as claimed. ∎

Proof of Proposition 4. Let us show that the equilibrium equations hold with linear approximations

$$P = \hat{P} + P^\epsilon, \quad Q = \hat{Q} + Q^\epsilon \quad \text{and} \quad A^i = \hat{A}^i + A^{i,\epsilon},$$

where $(\hat{P}, \hat{Q}, \hat{A}^i; i = 1 \ldots N)$ represent the solution of the symmetric model with risk capacities b (from Proposition 3), and

$$P^\epsilon = -\frac{1}{r+2\delta} \frac{3N-4}{N^2(N-1)} \left[\epsilon^1, \epsilon^2, \ldots \epsilon^N\right] + O(\epsilon^2),$$

$$Q^\epsilon = \frac{(N-2)(r+2\delta)}{2Nb} \begin{bmatrix} (N-1)\epsilon^1 & -\epsilon^2 & \ldots & -\epsilon^N \\ -\epsilon^1 & (N-1)\epsilon^2 & \ldots & -\epsilon^N \\ \vdots & \vdots & & \vdots \\ -\epsilon^1 & -\epsilon^2 & \ldots & (N-1)\epsilon^N \end{bmatrix} + O(\epsilon^2)$$

and $(r+2\delta)A^{i,\epsilon}$ has entries

$$a^{i,\epsilon,ii} = -\frac{4(N-1)}{N^3}\epsilon^i, \quad a^{i,\epsilon,ij} = a^{i,\epsilon,ji} = -\frac{N-2}{N^3}(\epsilon^j + \epsilon^i)$$

$$\text{and} \quad a^{i,\epsilon,jk} = \frac{N-2}{N^2(N-1)}\left(\frac{\epsilon^j + \epsilon^k}{N} - \frac{N-2}{2(N-1)}\epsilon^i\right),$$

for $j, k \neq i$. We have to show that the equilibrium equations hold up to terms of $O(\epsilon)$ (inclusively).

Note that (10), given (11), is equivalent to $((r+2\delta)I + Q)A^{ii} = -(b^i/2)1^i$, where A^{ii} denotes the i-th column of A^i. Linearizing this equation, together with (11), near the symmetric solution, we obtain

$$(r+2\delta)A^{i,\epsilon,i} + (Q^\epsilon)^T \hat{A}^{ii} + \hat{Q}^T A^{i,\epsilon,i} = -\epsilon^i/2\, 1^i, \tag{35}$$

and

$$(r+2\delta)A^{i,\epsilon} + (Q^\epsilon)^T \hat{A}^i + \hat{Q}^T A^{i,\epsilon} + \hat{A}^i Q^\epsilon + A^{i,\epsilon}\hat{Q} = \tag{36}$$

38

$$-\epsilon^i 1^{ii}/2 + (\hat{P}^T \underline{Q}^{\epsilon,i} + (\underline{Q}^{\epsilon,i})^T \hat{P})/2 + ((P^\epsilon)^T \hat{\underline{Q}}^i + (\hat{\underline{Q}}^i)^T P^\epsilon)/2.$$

We have

$$(Q^\epsilon)^T \hat{A}^{ii} = -\frac{N-2}{4N^3} \begin{bmatrix} (N-1)\epsilon^1 & -\epsilon^1 & \cdots & -\epsilon^1 \\ -\epsilon^2 & (N-1)\epsilon^2 & \cdots & -\epsilon^2 \\ \vdots & \vdots & & \vdots \\ -\epsilon^N & -\epsilon^N & \cdots & (N-1)\epsilon^N \end{bmatrix} \begin{bmatrix} N-2 \\ \vdots \\ 3N-2 \\ \vdots \end{bmatrix} = \frac{N-2}{2N^2}(\epsilon - N\epsilon^i 1^i)$$

$$((r+2\delta)I + \hat{Q}^T)A^{\epsilon,ii} = \frac{1}{N^2}\left(\frac{N}{2}I - \frac{N-2}{2N}\begin{bmatrix} 1 & 1 & \cdots \\ 1 & 1 & \cdots \\ \vdots & \vdots & \end{bmatrix}\right)\left(-\frac{N-2}{N}(\epsilon + \epsilon^i 1) - \frac{2N}{N}\epsilon^i 1^i\right) =$$

$$\frac{1}{N^2}\left(-\frac{N}{2}\frac{N-2}{N}(\epsilon + \epsilon^i 1) - N\epsilon^i 1^i + \frac{(N-2)^2}{2N}\epsilon^i 1 + \frac{N-2}{N}\epsilon^i 1\right) = \frac{1}{N^2}\left(-\frac{N-2}{2}\epsilon - N\epsilon^i 1^i\right)$$

Combining these expressions, we obtain (35).

Furthermore, to evaluate column i of the left-hand side of (36), we compute

$$\hat{A}^i Q^{\epsilon,i} = \frac{N-2}{4N^3}\begin{bmatrix} \frac{N-2}{N-1} & \cdots & -(N-2) & \cdots \\ \vdots & & \vdots & \\ -(N-2) & \cdots & -(3N-2) & \cdots \\ \vdots & & \vdots & \end{bmatrix}(-\epsilon^i 1 + N\epsilon^i 1^i) = -\frac{N-2}{4N^2}((N-2)1 + N1^i)\epsilon^i$$

$$A^{i,\epsilon}\hat{Q}^i = \frac{N-2}{2N^3}\begin{bmatrix} -\frac{N-2}{N-1}\left(\frac{\sum_{k\neq i}\epsilon_k + \epsilon_1(N-1)}{N} - \frac{N-2}{2}\epsilon^i\right) - \frac{N-2}{N}(\epsilon^1 + \epsilon^i)(N-1) \\ \vdots \\ \frac{N-2}{N}\left(\sum_{k\neq i}\epsilon_k + (N-1)\epsilon_i\right) - \frac{4(N-1)}{N}\epsilon^i(N-1) \\ \vdots \end{bmatrix} =$$

$$\frac{N-2}{2N^3}\begin{bmatrix} -\frac{N-2}{N-1}\left(\frac{\epsilon_1(N-1)}{N} - \frac{N-2}{2}\epsilon^i - \frac{\epsilon^i}{N}\right) - \frac{N-2}{N}(\epsilon^1 + \epsilon^i)(N-1) \\ \vdots \\ \frac{N-2}{N}(N-2)\epsilon^i - \frac{4(N-1)}{N}\epsilon^i(N-1) \\ \vdots \end{bmatrix} = \frac{N-2}{2N^3}\begin{bmatrix} -(N-2)\epsilon^1 - \frac{(N-2)^2}{2(N-1)}\epsilon^i \\ \vdots \\ -(3N-4)\epsilon^i \\ \vdots \end{bmatrix}$$

It follows that column i of the left-hand side of (36) is

$$-\frac{1}{2}1^i\epsilon^i - \frac{(N-2)^2}{4N^2}1\epsilon^i - \frac{N-2}{4N}1^i\epsilon^i + \frac{N-2}{2N^3}\begin{bmatrix} -2\frac{N-2}{N}\epsilon^1 - \frac{(N-2)^2}{2(N-1)}\epsilon^i \\ \vdots \\ -(3N-4)\epsilon^i \\ \vdots \end{bmatrix} \qquad (37)$$

Column i of the right-hand side of (36) is

$$-\frac{1}{2}1^i\epsilon^i - \frac{N-2}{4N^2}(N-1)1\epsilon^i + \frac{N-2}{4N^2}\epsilon - \frac{N-2}{4N}1^i\epsilon^i - \frac{N-2}{4N}\frac{3N-4}{N^2}\epsilon - \frac{3N-4}{N^2(N-1)}\epsilon^i\frac{N-2}{4N}(N1^i-1).$$

Subtracting (37), we obtain

$$-\frac{N-2}{4N^2}1\epsilon^i - \frac{(N-2)^2}{2N^3}\epsilon - \frac{3N-4}{N^2(N-1)}\epsilon^i\frac{N-2}{4N}(N1^i-1) + \frac{N-2}{2N^3}\begin{bmatrix} (N-2)\epsilon^1 + \frac{(N-2)^2}{2(N-1)}\epsilon^i \\ \vdots \\ (3N-4)\epsilon^i \\ \vdots \end{bmatrix}$$

$$= -\frac{(N-2)^3}{4N^3(N-1)}1\epsilon^i - \frac{(3N-4)(N-2)}{4N^2(N-1)}1^i\epsilon^i + \frac{N-2}{2N^3}\begin{bmatrix} \frac{(N-2)^2}{2(N-1)}\epsilon^i \\ \vdots \\ 2(N-1)\epsilon^i \\ \vdots \end{bmatrix}$$

$$= -\frac{(3N-4)(N-2)}{4N^2(N-1)}1^i\epsilon^i + \frac{N-2}{2N^3}\left(2(N-1) - \frac{(N-2)^2}{2(N-1)}\right)1^i\epsilon^i = 0.$$

Next, let us compute column $j \neq i$ of the left-hand side of (36). We have

$$(Q^\epsilon)^T\hat{A}^{ij} = \frac{N-2}{4N^3}\begin{bmatrix} (N-1)\epsilon^1 & -\epsilon^1 & \cdots & -\epsilon^1 \\ -\epsilon^2 & (N-1)\epsilon^2 & \cdots & -\epsilon^2 \\ \vdots & \vdots & & \vdots \\ -\epsilon^N & -\epsilon^N & \cdots & (N-1)\epsilon^N \end{bmatrix}\begin{bmatrix} \frac{N-2}{N-1} \\ \vdots \\ -(N-2) \\ \vdots \end{bmatrix} = \frac{(N-2)^2}{4N^2(N-1)}(\epsilon - N\epsilon^i1^i)$$

$$((r+2\delta)I+\hat{Q}^T)A^{\epsilon,ij} = \left(\frac{N}{2}I - \frac{N-2}{2N}\begin{bmatrix} 1 & 1 & \cdots \\ 1 & 1 & \cdots \\ \vdots & \vdots & \end{bmatrix}\right)\frac{N-2}{N^2(N-1)}\left(\frac{\epsilon^j1+\epsilon}{N} - \epsilon^j1^i - \frac{N-2}{2(N-1)}\epsilon^i1 - \frac{N}{2(N-1)}\epsilon^i1^i\right)$$

$$= \frac{N-2}{N^2(N-1)}\left(\frac{\epsilon^j1+\epsilon}{2} - \frac{N}{2}\epsilon^j1^i - \frac{(N-2)N}{4(N-1)}\epsilon^i1 - \frac{N^2}{4(N-1)}\epsilon^i1^i + \frac{N-2}{4}\epsilon^i1\right) =$$

40

$$= \frac{N-2}{N^2(N-1)}\left(\frac{\epsilon^j 1 + \epsilon}{2} - \frac{N}{2}\epsilon^j 1^i - \frac{N-2}{4(N-1)}\epsilon^i 1 - \frac{N^2}{4(N-1)}\epsilon^i 1^i\right)$$

$$\hat{A}^i Q^{\epsilon,j} = \frac{N-2}{4N^3}\begin{bmatrix} \frac{N-2}{N-1} & \cdots & -(N-2) & \cdots \\ \vdots & & \vdots & \\ -(N-2) & \cdots & -(3N-2) & \cdots \\ \vdots & & \vdots & \end{bmatrix}(-\epsilon^j 1 + N\epsilon^j 1^j) = \frac{N-2}{4N^2(N-1)}((N-2)1 + N1^i)\epsilon^j$$

$$A^{i,\epsilon}\hat{Q}^j = \frac{N-2}{2N^3}\begin{bmatrix} N\frac{N-2}{N-1}\left(\frac{\epsilon^j+\epsilon^1}{N} - \frac{N-2}{2(N-1)}\epsilon^i\right) - \frac{N-2}{N-1}\left(\frac{\sum_{k\neq i}\epsilon_k+\epsilon_1(N-1)}{N} - \frac{N-2}{2}\epsilon^i\right) + \frac{N-2}{N}(\epsilon^1+\epsilon^i) \\ \vdots \\ -(N-2)(\epsilon^j+\epsilon^i) + \frac{N-2}{N}\left(\sum_{k\neq i}\epsilon_k + (N-1)\epsilon^i\right) + 4\frac{N-1}{N}\epsilon^i \\ \vdots \end{bmatrix}$$

$$= \frac{N-2}{2N^3}\begin{bmatrix} \frac{N-2}{N-1}\epsilon^j + \frac{N-2}{N-1}\epsilon^1 + \frac{(N-2)N}{2(N-1)^2}\epsilon^i \\ \vdots \\ -(N-2)\epsilon^j + 2\epsilon^i \\ \vdots \end{bmatrix} = \frac{N-2}{(N-1)N^2}\left(\frac{N-2}{2N}\epsilon^j 1 + \frac{N-2}{2N}\epsilon + \frac{N-2}{4(N-1)}\epsilon^i 1\right)$$

$$+ \frac{N-2}{N^2(N-1)}\left(\frac{N}{4(N-1)}\epsilon^i - \frac{N-2}{2}\epsilon^j\right)1^i.$$

Adding up,

$$\frac{N-2}{N^2(N-1)}\left(\frac{N}{4}\epsilon - N\frac{N-2}{4}\epsilon^i 1^i + \frac{\epsilon^j 1}{2} - \frac{N-2}{4(N-1)}\epsilon^i 1 - \frac{N^2}{4(N-1)}\epsilon^i 1^i + \frac{N-2}{4}1\epsilon^j - \frac{N}{4}1^i\epsilon^j\right)$$

$$+ \frac{N-2}{(N-1)N^2}\left(\frac{N-2}{2N}\epsilon^j 1 + \frac{N-2}{2N}\epsilon + \frac{N-2}{4(N-1)}\epsilon^i 1 + \frac{N}{4(N-1)}\epsilon^i 1^i - \frac{N-2}{2}\epsilon^j 1^i\right) =$$

$$\frac{N-2}{N^2(N-1)}\left(\frac{N^2+2N-4}{4N}\epsilon - N\frac{N-1}{4}\epsilon^i 1^i - \frac{3N-4}{4}1^i\epsilon^j + \frac{N^2+2N-4}{4N}\epsilon^j 1\right)$$

Next, column $j \neq i$ of the right-hand side of (36) is

$$\frac{N-2}{4N^2}\epsilon^j 1 + \frac{N-2}{4N^2}(\epsilon - N\epsilon^i 1^i) + \frac{N-2}{4N}\frac{3N-4}{N^2(N-1)}\epsilon - \frac{3N-4}{N^2(N-1)}\epsilon^j\frac{N-2}{4N}(N1^i - 1) =$$

41

$$-\frac{N-2}{4N}\epsilon^i 1^i + \frac{N-2}{4N^2}\frac{N^2+2N-4}{N(N-1)}\epsilon - \frac{3N-4}{N(N-1)}\frac{N-2}{4N}\epsilon^j 1^i + \frac{N^2+2N-4}{N(N-1)}\frac{N-2}{4N^2}\epsilon^j 1$$

We see that left and right-hand sides match in column j for all $j \neq i$ as well. We conclude that indeed all relevant equilibrium conditions hold up to terms of order ϵ inclusively. ∎

Proof of Proposition 6. With one large trader and the fringe, the mechanism is described by the four parameters:

$$Q = \begin{bmatrix} q_L & -q_F \\ -q_L & q_F \end{bmatrix}, P = [P_L, P_F]$$

Let L_t denote the holding of the large trader and F_t the holding of the fringe. Price at time t is

$$p_t = (P_F F_t + P_L L_t)$$

and the large trader net selling rate is

$$(q_L L_t - q_F F_t)\, dt$$

We now establish existence of and uniqueness of a non-degenerate mechanism. Consider first the fringe optimality condition which in this case simplifies to:

$$-(r+\delta)\, p_t = b_F F_t - E\left[P_F \dot{F}_t + P_L \dot{L}_t\right]$$

The expected changes in holdings are

$$E\left[\dot{L}_t\right] = -\delta L_t - (q_L L_t - q_F F_t)$$
$$E\left[\dot{F}_t\right] = -\delta F_t + (q_L L_t - q_F F_t)$$

The fringe optimality can be hence written as:

$$(r+2\delta)\,(P_F F_t + P_L L_t) = -b_F F_t + (P_F - P_L)\,(q_L L_t - q_F F_t)$$

Since this equation has to hold for all F and L, we must have:

$$P_F\,(r+2\delta+q_F) - P_L q_F = -b_F$$
$$P_L\,(r+2\delta+q_L) - P_F q_L = 0.$$

Now consider the large trader optimality. He chooses an announcement Y_t to maximize:

$$rf\,(L_t, F_t) = \max_{Y_t}\left(-\frac{b_L}{2}L_t^2 + (P_L Y_t + P_F F_t)\,(q_L Y_t - q_F F_t) + E\frac{df\,(L_t, F_t|Y_t)}{dt}\right) \quad (38)$$

42

As usual, we make a guess that the value function is quadratic:

$$f(L, F) = k_0 + a_{11}L^2 + 2a_{12}LF + a_{22}F^2$$

Then

$$Edf(L_t, F_t|Y_t) = 2a_{11}L_tE\left[\dot{L}_t\right] + 2a_{12}(L_tE\left[\dot{F}_t\right] + F_tE\left[\dot{L}_t\right]) + 2a_{22}F_tE\left[\dot{F}_t\right] + C$$

and

$$E\left[\dot{L}_t\right] = -\delta L_t - (q_L Y_t - q_F F_t)$$
$$E\left[\dot{F}_t\right] = -\delta F_t + (q_L Y_t - q_F F_t)$$

The FOC of the maximization problem (38) is:

$$P_L(2q_L Y_t - q_F F_t) + P_F F_t q_L - 2(a_{11}L_t + a_{12}F_t)q_L + 2(a_{22}F_t + a_{12}L_t)q_L = 0$$

Evaluated at truth-telling it becomes:

$$P_L(2q_L L_t - q_F F_t) + P_F F_t q_L - 2(a_{11}L_t + a_{12}F_t)q_L + 2(a_{22}F_t + a_{12}L_t)q_L = 0$$

For it to hold for all (L_t, F_t) we need:

$$-P_L q_L + q_L a_{11} - q_L a_{12} = 0$$
$$P_L q_F - P_F q_L - 2a_{22}q_L + 2a_{12}q_L = 0$$

Finally, matching the coefficients of the large player value function we get a system of equations: :

$$0 = -P_L q_L + q_L a_{11} - q_L a_{12}$$
$$0 = P_L q_F - P_F q_L - 2a_{22}q_L + 2a_{12}q_L$$
$$ra_{11} = 2a_{12}q_L - 2a_{11}(\delta + q_L) - \frac{1}{2}b_L + P_L q_L$$
$$r2a_{12} = 2a_{11}q_F - 2a_{12}(2\delta + q_L + q_F) + 2a_{22}q_L + P_F q_L - P_L q_F$$
$$ra_{22} = 2a_{12}q_F - 2a_{22}(\delta + q_F) - P_F q_F$$
$$-b_F = P_F(r + 2\delta + q_F) - P_L q_F$$
$$0 = P_L(r + 2\delta + q_L) - P_F q_L$$

This system has two solutions: a degenerate one (with no trade) and a regular one:

$$q_F = \frac{1}{2}r + \delta$$

$$q_L = \frac{1}{2}b_L\frac{r + 2\delta}{b_F}$$

$$P_L = -\frac{b_L b_F}{(3b_F + b_L)(r + 2\delta)}$$

$$P_F = -\frac{(2b_F + b_L)b_F}{(r + 2\delta)(3b_F + b_L)}$$

$$A^L = \frac{b_F}{2(r + 2\delta)(3b_F + b_L)}\begin{bmatrix} -3b_L & -b_L \\ -b_L & b_F \end{bmatrix}$$

Finally, the welfare of the fringe can be written as

$$rf^F(F, L) = -\frac{b_F}{2}F^2 + (P_L L + P_F F)(-q_L L + q_F F) + \frac{df^F(F, L)}{dt}$$

Making a guess that

$$f^F(F, L) = k_0^F + a_{11}^F F^2 + 2a_{12}^F LF + a_{22}^F L^2$$

allows us to match coefficients:

$$ra_{11}^F = -\frac{1}{2}b_F + P_F q_F + d_{12}q_F - 2a_{11}^F(\delta + q_F)$$

$$r2a_{12}^F = 2a_{11}^F q_L - 2a_{12}^F(2\delta + q_F + q_L) + P_L q_F - P_F q_L + 2a_{22}^F q_F$$

$$ra_{22}^F = -P_L q_L + 2a_{12}^F q_L - 2a_{22}^F(\delta + q_L)$$

Using the solutions for P and Q we get a unique solution:

$$a_{11}^F = -\frac{1}{2}b_F\frac{b_L^2 + 5b_L b_F + 5b_F^2}{(3b_F + b_L)(2b_F + b_L)(r + 2\delta)}$$

$$a_{12}^F = -\frac{1}{2}b_L\frac{b_F^2}{(3b_F + b_L)(2b_F + b_L)(r + 2\delta)}$$

$$a_{22}^F = \frac{1}{2}b_L^2\frac{b_F}{(3b_F + b_L)(2b_F + b_L)(r + 2\delta)}$$

or

$$A^F = \frac{1}{2}\frac{b_F}{(3b_F + b_L)(2b_F + b_L)(r + 2\delta)}\begin{bmatrix} -(b_L^2 + 5b_L b_F + 5b_F^2) & -b_L b_F \\ -b_L b_F & b_L^2 \end{bmatrix}$$

∎

Proof of Proposition 7. If $\lambda = \bar{\lambda}$ then subtracting the second equation from the first equation, we get an expression whose sign is the same as

$$-\frac{r + 2\delta + \bar{\lambda}}{r + 2\delta + 2\bar{\lambda}}\left(\frac{b_L}{b_L + (N - 1)b_F} + N - 2\right) < 0.$$

To be completed. ∎

Bibliography.

Babus, A. and P. Kondor (2012) "Trading and Information Diffusion in Over-the-Counter markets," working paper, Imperial College London.

Back, K. (1992) "Insider Trading in Continuous Time," *Review of Financial Studies*, 5, 387-409.

Back, K., C. H. Cao, and G. A. Willard (2000) "Imperfect Competition among Informed Traders," *Journal of Finance*, 55, 2117-2155.

Du, Songzi and Haoxiang Zhu (2013) "Dynamic Ex Post Equilibrium, Welfare and Optimal Trading Frequency in Double Auctions," working paper, MIT

Glosten, L. R. and P. R. Milgrom (1985) "Bid, Ask and Transaction Prices in a Specialist Market with Heterogenously Informed Traders," *Journal of Financial Economics*, 14, 71-100.

Kyle, A. S. (1985) "Continuous Auctions and Insider Trading," *Econometrica*, 15, p. 1315-1335.

Kyle, A. S. (1989) "Informed Speculation with Imperfect Competition," *Review of Economic Studies*, 56, 317-355.

Rostek, M. and M. Weretka (2012) "Price Inference in Small Markets," *Econometrica*, 80(2), 687-711.

Vayanos, D. (1999) "Strategic Trading and Welfare in a Dynamic Market," *Review of Economic Studies*

Vives, X. (2011) "Strategic Supply Function Competition with Private Information," *Econometrica*, 79, 1919-1966.