The Macrostructure of Electronic Financial Markets

Craig Pirrong
College of Business Administration
Oklahoma State University
Stillwater, OK 74078
pirrong@okstate.edu

January 29, 2002

Abstract. Investor-owned, for-profit electronic exchanges differ in many ways from traditional intermediary-owned, non-profit exchanges, but both are subject to the centripetal force of liquidity. Due to the nature of liquidity, competition between investor-owned electronic exchanges that offer different functionalities is of the winner take all variety if traders uniformly prefer one system over the other, with the winner being the firm that offers the preferred functionality obtaining a natural monopoly. The prices and profits of the winning firm are determined by the value traders place on the superior functionality. Switching costs also influence the profitability of the winning exchange. Whereas traditional intermediary-owned exchanges enhance member profits by restricting the number of intermediary-members, thereby restricting the supply of liquidity, the investor-owned electronic exchange has an incentive to maximize liquidity supply in order to maximize derived demand for its services.

1 Introduction

Trading of equities, futures, and options traditionally has taken place on floor-based "open outcry" exchanges. In the last decade and a half, however, there has been a pronounced movement to electronic trading. Outside the United States floor trading has been almost completely eclipsed. Within the United States, electronic trading has made serious inroads, especially in OTC equre some derivatives. Moreover, new Business-to-Business ("B2B") exchanges have been developed to trade non-traditional commodities and derivatives (e.g., computer chips and chemicals).

Traditional open outcry exchanges are non-profit "clubs" of intermediaries who supply brokerage and market making services. Exchange members derive their profits through the provision of these services rather than through a claim on any earnings generated by the exchange's fees. Exchange fees are set so that the exchange covers costs. The non-profit constraint precludes the exchange from distributing any profits to its members, which attenuates the incentive to charge prices in excess of cost, and exchange owners typically prefer lower exchange fees to encourage volume and thereby enhance their trading and brokerage profits.

Some electronic exchanges (such as Eurex and the Intercontinental Exchange) are for-profit exchanges owned largely or completely by traditional intermediaries. These exchanges are similar to traditional exchanges in key respects, most notably in that important intermediaries dominate governance, and profit from their dealings on the exchange as well as through their claim on the exchange's cash flows.

However, electronic trading has the potential to transform liquidity supply by allowing non-traditional intermediaries to supply market making and liquidity services. Whereas the traditional exchange depends on a specialized cadre of liquidity suppliers (such as specialists and "locals"), anyone with a computer and sufficient cash can supply liquidity in an electronic market. In the extreme, as conjectured by Fisher Black as early as 1971, no specialized "dealers" or market makers are required; patient investors who trade *via* limit order can supply liquidity instead.[1] The recent experience in OTC equities provides a vivid example of the potential for non-traditional market makers to compete effectively against traditional dealers in an electronic environment. The promulgation of new Order Handling Rules by the SEC dramatically lowered the costs of "day traders" and others to supply liquidity by entering limit orders through electronic communications networks (ECNs). Competition from this new source of liquidity suppliers dramatically reduced bid-ask spreads and sharply eroded the profits of traditional dealers in OTC stocks (Barclay *et al,* 1999; Weston, 2000). Similarly, there are no dealers or specialists on the Paris Bourse's CAC system; traders compete to supply liquidity *via* limit order (Biais *et al,* 1995).

This experience suggests that electronic trading has the potential to undermine the dominant role that traditional liquidity suppliers and other intermediaries have heretofore exercised on most exchanges. In particular, it is now feasible to create investor owned, for-profit businesses that offer (a) remote electronic market access to end users and market makers, (b) auto-

[1] Glosten (1994) also presents a model in which all liquidity is supplied by limit order investors.

mated execution of transactions, and (c) integrated clearing of transactions. Such electronic exchanges themselves offer no brokerage or market making services (nor do their owners), but instead earn their profits through the fees they charge to market users. For instance, in the futures markets the Chicago Mercantile Exchange (CME) and the London International Financial Futures and Options Exchange (LIFFE) have both implemented restructuring plans that eliminate the traditional link between trading privileges and exchange ownership. In equities, NASDAQ has become a for-profit, investor owned exchange. As another example, many B2B exchanges are investor owned public companies with no necessary linkage between ownership and trading privileges. Thus, open outcry (and intermediary owned electronic) exchanges and investor-owned electronic exchanges operate under fundamentally different business models. This suggests that the nature of competition between exchanges may differ substantially in open outcry and electronic environments.

At the same time, both electronic and open outcry exchanges are subject to common fundamental economic forces. In particular, the nature of liquidity in financial markets exerts a decisive impact on both types of exchange. Liquidity provides an incentive for the concentration of trading in a single market. That is, liquidity effects make financial trading a network industry. Understanding the ramifications of liquidity effects is therefore central to the understanding of the organization of financial markets.

Pirrong (1999, 2001a, 2001b) uses a standard market microstructure model to show how intermediaries that form a non-profit exchange can exploit the network aspects of financial trading to exercise market power and earn economic rents. Specialized intermediaries that operate on exchange floors can create a trading network that is uneconomically small by restricting exchange membership, but which is large enough to preclude competition from any other network. Restricting the number of intermediaries who can trade on the exchange generates rents for those members. This theory predicts concentration of trading on a single exchange.[2] Moroever, the model can be readily interpreted as characterizing the operation of an intermediary-owned electronic exchange.

This article uses a similar microstructure model to examine how network effects influence the "macrostructure" of electronic financial markets. It presents a model of competition between for-profit, investor-owned electronic exchanges. The exchanges develop differentiated trading systems and compete in prices to attract trading activity. As in an open outcry exchange, liquidity-driven network effects induce the market to "tip" to one exchange; that is, the centripetal force of liquidity attracts all trading activity to a single exchange. In the electronic environment, conventional equilibrium selection criteria imply that trader preferences regarding system functionality determine the exchange that wins this contest, the price that winning exchange can charge for its services, and the profits that its owners earn. The market "tips" to the system that traders prefer, and the price that this monopoly exchange can charge is limited by the substitutibility of the two system; the greater trader preference for one system over another, the greater the prices and profits of the winning system.

[2] Pirrong (2001a, 2001b) shows that multiple trading venues can survive if one venue can restrict its dealings to uninformed traders. In such a fragmented market, all those who cannot reliably demonstrate that they are uninformed trade on a single market. This prediction is broadly consistent with observed financial structure.

A comparison between the model of this paper and models of intermediary-owned markets (Pirrong 1999, 2001a, 2001b) points out a major difference in the nature of inter-exchange competition in intermediary-owned and investor-owned electronic markets. In the intermediary-owned market, the liquidity supply capacity of the intermediaries that join to form an exchange is the decisive factor in determining which exchange dominates; the exchange whose intermediaries have the largest risk bearing capacity captures all the order flow. In an investor-owned electronic trading environment, in contrast, technology and functionality are the decisive factors; the exchange that offers the preferred technology wins. Moreover, whereas an intermediary-owned exchange has an incentive to restrict intermediary participation to increase their trading profits, a for-profit electronic exchange dependent on fees for revenues typically does not have such an incentive. Restricting market maker participation reduces the profits that the for-profit exchange can capture. Instead, the for-profit investor-owned exchange has an incentive to encourage market maker participation in order to increase derived demand for the exchange services, thereby allowing it to earn higher revenues from fees charged to end users.

Thus, the model studied herein implies that the electronic financial marketplace is a "winner take all" environment, and that competition is primarily technological. The electronic financial market is therefore very much a "new economy" network industry such as computer software, even though the fundamental force that makes it so is as old as financial markets themselves–liquidity.

The analysis also implies that the costs of coordinating the movement of traders from one exchange to another–switching costs–will help determine the winner who takes all and the profits the winner earns. Moreover, it implies that exchanges may attempt to utilize contractual means (such as loyalty contracts or exclusive dealing contracts) and other strategic devices to influence switching costs.

The remainder of this article is organized as follows. Section 2 presents a formal model of the microstructure of the trading process and competition between exchanges. Section 3 solves the model. Section 4 examines the model's implications for the evolution of industry structure. Section 5 discusses the circumstances under which multiple electronic exchanges can co-exist. Section 6 presents a brief summary.

2 The Model

The study of the macrostructure of a financial market should be grounded on a specific model of the microstructure of the trading process. This article employs a variant of the canonical Kyle-type microstructure model as its analytical foundation. The details of the model follow.

2.1 The Asset

There is a single asset or financial contract available for trading. The true value of the traded instrument (which is not public knowledge) is v. The unconditional distribution of v is normal with mean of 0 and variance σ^2.

2.2 The Traders

Three types of agents desire to trade the instrument.

- There are K risk neutral informed traders who know v. Informed traders trade via market order, and can buy or sell multiple units of the asset.

- There is a large (but finite) number of uninformed traders–"noise traders"– who trade for portfolio balancing or hedging purposes. Each noise trader desires to submit a market order to buy or sell a single unit of the asset. The probability noise trader j wants to buy is .5 and the probability noise trader j wants to sell is $p = .5$. Noise trader demands are independent.

  Noise trader j will submit a market order if and only if the cost of executing the order is less than b_j. The cost of execution equals the sum of (a) the expected deviation between the market price and the value of the asset multiplied by an indicator variable equal to 1 for buys and -1 for sells, and (b) any transaction fee levied by the exchange. Noise traders are indexed in order of decreasing willingness to pay, i.e., $b_i > b_j$ for $i < j$. The reservation prices b_j define an inverse demand function $D(q) = b_q$.

  The foregoing implies that net noise trader demand (noise trader buys minus noise trader sells) is the difference between binomial random variates. For n noise traders, the probability that there are K_S sales is $B(n, K_S, p)$ and the probability that there are K_B buys is $1 - B(n, K_S, p)$ If n is sufficiently large, these binomial probabilities converge to the normal distribution.[3] Therefore, to simplify the analysis, I assume that

[3] As a rule of thumb, when $p = .5$, the normal distribution is a good approximation of the binomial for $n > 5$.

the the number of noise traders is sufficiently large so that net noise trader demand when $\hat{N}$ submit orders is a normal random variable with mean 0 and variance $\hat{N}$.[4] Thus, the variance of the sum of several noise trader's demands is equal to the number of noise traders. Noise trader demand and the value of the asset are orthogonal.

- There is a set of potential liquidity suppliers (also referred to as market makers) $\mathbf{L} = \{1, 2, \ldots, L\}$. Each liquidity supplier $j \leq L$ is risk averse, with a constant absolute risk aversion coefficient α_j. Equivalently, the risk tolerance of intermediary j is $t_j = 1/\alpha_j$. Moreover, wlog $t_j > t_k$ for $j < k$. That is, intermediaries are ordered by decreasing risk tolerance. The total supply of risk bearing capacity (i.e., aggregate risk tolerance) is $T^A = \sum_{i=1}^{L} t_i$.

  The analysis of Pirrong (2001a, 2001b) implies that if market makers can form coalitions with binding contracts, a coalition that offers $.5T_A + \epsilon$ in risk bearing capacity will obtain a monopoly because (a) no other coalition can offer lower trading costs, and (b) due to the network aspects of a financial market, all trading activity concentrates on the exchange offering the lowest trading costs.[5] Moreover, the members of this coalition have no incentive to expand coalition membership because this would erode their trading profits. This article rules out such dominant coalitions. The intuition is that such coalitions may

[4] If K_S sell orders are submitted, net order flow is $z = \hat{N} - 2K_S$. The expected value of z is 0, and the variance is $4E[K_S - E(K_S)]^2 = 4\hat{N}(p)(1-p) = \hat{N}$.

[5] The Pirrong (2001a, 2001b) model assumes that liquidity costs are the only costs that traders incur. If there are other costs, such as order routing costs or clearing costs, an exchange with less than $.5T_A$ of market maker risk bearing capacity can acquire a monopoly if its non-liquidity costs are smaller than those of its competitors.

be plausible in a traditional open outcry system in which a relatively small number of specialized intermediaries can efficiently supply liquidity, but that as the OTC equity experience demonstrates, electronic trading permits the participation of such a large number of spatially dispersed market makers that no coalition of large intermediaries can amass the requisite share of market making capacity to forestall entry by a competing exchange. In essence, I assume the inevitability of Black's (1971) vision of a market that relies upon numerous individual and institutional investors, rather than specialized market makers, to supply liquidity.

2.3 The Exchanges

Two firms, Firm 1 and Firm 2, have the opportunity to invest in creating electronic exchanges by paying a cost c.[6] Each of the firms can choose one of two system designs–design A or design B. These designs offer different functionality. When they must choose the firms do not know which design traders will prefer. If noise traders prefer A to B (respectively, B to A) *ceteris paribus* they are willing to pay δ more to trade on system A rather than system B (respectively, system B rather than system A). That is, if a trader's reservation price on system A is b, his reservation price on system B is $b - \delta$. At the time of the investment choice, the firms believe that there is a probability of .5 that noise traders will prefer design A and an equal probability that noise traders will prefer design B.

This assumption is intended to reflect the fact that trading systems are

[6] It is straightforward to extend the analysis to three or more firms.

heterogeneous and complex. Electronic trading systems compete on functionality and there is considerable scope for innovation regarding system features, and considerable uncertainty regarding the value traders place on alternative bundles of features.

If Firm $k = 1, 2$ invests c and creates a system, it must choose a per trade fee f_k that it charges for execution of a market order and a "terminal rental fee" of R_k that it charges market makers. The per trade fee is levied on market order submitters. The exchanges cannot distinguish between uninformed and informed traders, and serve all market orders on a non-discriminatory basis. Market makers pay no per trade fee. Instead, they pay a fixed access charge that is independent of their scale of trading.

The assumption that market makers pay no per trade fee serves two purposes. First, it is realistic. Many electronic exchanges recognize that taxing market makers can seriously impair liquidity. As a consequence, exchanges often charge discounted per trade fees for those who supply liquidity by posting firm quotes. Globex (a large electronic exchange) charges liquidity suppliers no per trade fee in order to attact liquidity suppliers. As another example, the NYSE's Automated Bond System charges per trade fees only on agency trades. Such exchanges typically charge system "members" some form of fixed access charge. Second, this assumption makes the analysis tractible. Market microstructure models of the Kyle variety are tractible because linear equilibria exist, and are often the only equilibria. If market makers pay a per trade fee a linear equilibrium cannot exist and the model is intractible.

2.4 Timing

The timing of actions in the model is as follows:

- Firms 1 and 2 choose whether to invest c to create an electronic exchange. Each exchange chooses either design A or design B.
- The Firms learn whether traders prefer design A or B.
- The Firms choose their per trade fees and access charges.
- Traders choose which exchange to patronize. Informed traders can submit orders on multiple exchanges. Noise traders must choose the exchange (if any) to which they direct their market order. Market makers must choose a which exchange (if any) to join.
- Noise traders and informed traders submit market orders. Each market clears in a batch auction.

3 Market Equilibrium

As usual, the analysis of this model proceeds backwards from the last stage as follows. First, trading cost on each exchange is determined as a function of the exchange selections of the noise traders and market makers. Second, I show that this trading cost function implies that trading costs on exchange i are decreasing in the number of noise traders that select that exchange. This implies that in any stable equilibrium, all noise traders choose to trade on the same exchange. Third, I show that this "tipping" result implies that all market makers choose the same exchange as the noise traders. Fourth, the standard equilibrium selection criterion implies that noise traders coordinate on the exchange that minimizes their trading costs (given exchange choices of f_i and R_i). This, in turn, determines the relation between exchange prices and exchange revenues. Fifth, traditional Bertrand-style competition between exchanges determines equilibrium prices. If exchanges choose the same system, neither earns any revenue in equilibrium, whereas if one exchange creates the favored system and the other does not, the favored system captures 100 percent of the market and can charge a price in excess of marginal cost. Sixth, given this outcome a standard non-cooperative game analysis implies that either (a) one exchange will invest in creation of system A and the other will invest in the creation of system B, or (b) only one firm will invest.

3.1 Trading Costs

The most complex case obtains when each exchange selects a different system. To analyze this case, assume that Firm (exchange) 1 has chosen technology A and Firm (exchange) 2 has chosen system B, and that *ceteris paribus* traders prefer system A.

Assume initially that market makers have chosen the exchange to patronize such that the total risk tolerance (the sum of the risk tolerances) of the members of exchange 1 is T_1, and the total risk tolerance of exchange 2 is T_2. Assume initially that N_1 of the noise traders have chosen to trade on exchange 1, and $N_2 = N - N_1$. Define z_i as the net uninformed order flow on exchange i.

Analysis of equilibrium proceeds in the standard way. Upon learning v

the informed traders conjecture that the price on exchange i, $i = 1, 2$ is a linear function of order flow:

$$P_i = \lambda_i(\sum_{k=1}^{K} w_{ik} + z_i) \tag{1}$$

where w_{ik} is the order that the informed trader k submits to exchange i, z_i is net noise trader demand on exchange i, and λ_i is a constant. λ_i measures the sensitivity of the security's price to variations in order flow. Its reciprocal is referred to as market "depth;" greater depth (smaller λ_i) desirable because it implies lower transactions costs for noise traders.

Given this conjecture of a linear price function, the informed trader l chooses w_{il}, $i = 1, 2$ to maximize:

$$V_i = w_{il}E[v - \lambda_i(w_{il} + z_i + \sum_{k \neq l} w_{ik})] - f_i|w_{il}| \tag{2}$$

where the expectation is taken over z_i. Since v and z_i are orthogonal, the symmetric solution of the informed traders' maximization problems implies:

$$w_{il} = \beta_i(v - f_i) = \frac{v - f_i}{(K+1)\lambda_i} \qquad v > f_i$$

$$w_{il} = \beta_i(v + f_i) = \frac{v + f_i}{(K+1)\lambda_i} \qquad v < -f_i$$

$$w_{il} = 0 \qquad -f_i \leq v \leq f_i$$

That is, $\beta_i = 1/[(K+1)\lambda_i]$. β_i measures the intensity of informed trading. Define the expected value of v^2 conditional on $|v| > f_i$ as:

$$\sigma_i^2 = 2\int_{-\infty}^{f_i} v^2 \frac{e^{-.5(v/\sigma)^2} dv}{\sqrt{2\Pi\sigma^2}} < \sigma^2$$

The variance of informed trader order flow is $K\beta_i^2\sigma_i^2$.

Noise trader orders depend on λ_i, f_i and N_i. If $\hat{N}_i \leq N_i$ uninformed traders submit orders on exchange i, total expected execution costs (net of per trade fees) incurred by noise traders on exchange i is $X_i(\hat{N}_i, T_i) = E(P_i - v)z_i$. Since (a) $P_i = \lambda_i(z_i + I_i)$ (where I_i is informed order flow on exchange i), and (b) noise trader orders are orthogonal to v and hence to I_i, $X_i(N_i, T_i) = \lambda_i \hat{N}_i$ since $E(z_i^2) = \hat{N}_i$. Thus, the per noise trader expected execution cost (including the per trade fee) is:

$$x_i(\hat{N}_i, T_i) = \frac{X_i(\hat{N}_i, T_i)}{\hat{N}_i} + f_i = \lambda_i + f_i$$

All noise traders j with $b_j > x_i(\hat{N}_i, T_i)$ who have selected exchange i will submit a market order. Thus, $\hat{N}_i$ is decreasing in λ_i and f_i (and is bounded above by N_i).

Total order flow on exchange i is

$$I_i \equiv K\beta_i(v - f_i)1_{[v>f_i]} + k\beta_i(v + f_i)1_{[v<-f_i]} + z_i$$

where $1_{[v>f_i]}$ is an indicator variable equal to one if $v > f_i$ and zero otherwise, and $1_{[v<-f_i]}$ is an indicator variable equal to one if $v < -f_i$ and zero otherwise. Conditional on order flow, liquidity supplier j chooses his trade y_j to maximize his certainty-equivalent profit. Formally:

$$E\Pi_j = \max_{y_j}\{y_j E[v - P|I_i + z_i] - \frac{.5\hat{\sigma}_i^2 y_j^2}{t_j}\} \tag{3}$$

where $\hat{\sigma}_i^2$ is the variance of v conditional on total order flow on exchange i, and P is given by (1). The first term inside the brackets is the market maker's expected profit from a trade of y_j units. The second term adjusts for the risk of holding y_j units; $\hat{\sigma}^2 y_j^2$ is the variance of j's wealth, and $-.5/t_j$ is the cost per unit of variance.

The market makers estimate $E[v|I_i + z_i]$ using the regression of v on $I_i + z_i$.[7] Thus,

$$E[v|I_i + z_i] = \frac{K\beta_i\sigma_i^2}{K^2\beta_i^2\sigma_i^2 + \hat{N}_i}(I_i + z_i) \tag{4}$$

Moreover, by (1), $E[P|I_i + z_i] = \lambda_i(I_i + z_i)$, and

$$\hat{\sigma}_i^2 = \frac{\hat{N}_i\sigma_i^2}{K^2\beta_i^2\sigma_i^2 + \hat{N}_i} \tag{5}$$

Therefore,

$$y_j = \frac{t_j[\frac{K\beta_i\sigma_i^2}{K^2\beta_i^2\sigma_i^2+\hat{N}_i} - \lambda_i](I_i + z_i)}{\hat{\sigma}_i^2} \tag{6}$$

Call $\mathbf{L_i}$ the set of intermediaries on exchange i. Market clearing implies:

$$z_i + \sum_{j\in\mathbf{L_i}} y_j + I_i = 0. \tag{7}$$

Thus,

$$\frac{T_i[\frac{K\beta_i\sigma_i^2}{K^2\beta_i^2\sigma_i^2+\hat{N}_i} - \lambda_i](I_i + z_i)}{\hat{\sigma}_i^2} + I_i + z_i = 0 \tag{8}$$

where $T_i = \sum_{j\in\mathbf{L_i}} t_j$. This, in turn, implies:

$$\lambda_i = \frac{\hat{\sigma}_i^2}{T_i} + \frac{K\beta_i\hat{\sigma}_i^2}{\hat{N}_i} \tag{9}$$

[7] In the traditional Kyle-type model, v and z_i are assumed to be normally distributed, which implies that this regression is the best estimate of the conditional expected value. In the present model, $K[\beta_i(v - f_i)1_{v>f_i} + \beta_i(v + f_i)1_{v<-f_i}]$ is not normally distributed. Moreover, z_i is only asymptotically normal. A hyper-rational Bayesian market maker would use the true distributions of these variables to estimate the conditional mean and variance. Appealing to the Central Limit Theorem and the properties of the binomial distribution, I assume that there are sufficient numbers of noise traders and informed traders such that the order flow is approximately normal. More specifically, the market makers cannot economically discern the differences between the true distribution and the normal distribution, and hence act as if the order flow distribution is normal; that is, the cost market makers incur to implement the full Bayesian analysis exceeds the benefits derived from more exact forecasts of risk and return.

This can be rewritten as:

$$\lambda_i = \frac{\hat{\sigma}_i^2}{T_i} + \frac{K\hat{\sigma}_i^2}{(K+1)\lambda_i \hat{N}_i} \tag{10}$$

Pirrong (2001a, 2001b) proves that λ_i is decreasing in $\hat{N}_i$. Recall that $\hat{N}_i$ is decreasing in λ_i. Thus, determination of the equilibrium λ_i requires solution of a fixed point problem as illustrated in Figure 1. The horizontal axis graphs $\hat{N}_i$. The vertical axis is in dollars. The downward sloping curve labelled "$b(N)$" depicts the reservation price of the marginal noise trader. The downward sloping curve labelled "x" depicts per trader execution cost as a function of $\hat{N}_i$. It is possible to show that λ_i is bounded away from zero, and is infinite when $\hat{N}_i = 0$. Thus, if $b(.)$ is decreasing in $\hat{N}_i$ and there exists some $\hat{N}_i^*$ such that $b(\hat{N}_i) = 0$ for $\hat{N}_i > \hat{N}_i^*$, at the rightmost intersection of the two functions the $b(.)$ curve will cross the $x_i(.)$ curve from above. This defines $\hat{N}_i$ and hence defines S_i, λ_i, and x_i.

3.2 Equilibrium Trader Choice

Given N_1, N_2, T_1, and T_2, one of three conditions holds:

1. $x_1(N_1, T_1) > x_2(N_2, T_2) + \delta$

2. $x_1(N_1, T_1) < x_2(N_2, T_2) + \delta$

3. $x_1(N_1, T_1) = x_2(N_2, T_2) + \delta$

If (1) holds, the fact that x_i is decreasing in N_i implies that some noise traders will shift from exchange 1 to exchange 2. This will increase the disparity between execution costs on the two exchanges, and hence lead to further defections to exchange 2. Similarly, if (2) holds, all noise traders will defect from exchange 2 and move to exchange 1. Possibility (3) is not a stable equilibrium. If any noise trader (or market maker for that matter) switches exchanges, the "tipping" process will begin leading to all noise traders concentrating on a single exchange.

Thus, in any stable equilibrium all noise traders will select the same exchange. Since (as is shown in Pirrong 2001a, 2001b) market maker profits on exchange i are strictly increasing in N_i, in equilibrium no market maker will choose the exchange with no noise traders. Moreover, if all noise traders congregate on exchange i, the equilibrium T_i will be determined as in Figure 2 since (as again shown in the appendix) given N_i market maker profit on exchange i is decreasing in T_i. This figure represents T_i on the horizontal axis and the profit of the marginal market maker on the vertical axis. The equilibrium T_i equates the profit of the marginal market maker to R_i. Thus, T_i is a function of R_i. Moreover, since (a) market maker profit is a function of S_i, and (b) S_i is a function of f_i, T_i is also a function of f_i.

There are two possible stable equilibria in this model. In the first equilibrium all noise traders and market makers trade on exchange 1; in the second, all trade on exchange 2.

This "tipping" phenomenon is typical of network models. Under the assumptions of this model, liquidity considerations effectively make securities trading a network industry. Traders congregate where others trade.

Multiple equilibria are common in network models. To solve for equilibrium decisions at earlier stages of a game that embeds a network game at a later stage it is necessary to specify an equilibrium selection criterion.

The standard equilibrium selection mechanism in this context is to assume coordination among the decision making of the agents whose decisions have a positive feedback effect; the coordination of their decisions determines which equilibrium obtains.[8] Here the noise traders' actions exhibit positive feedback, so I assume that they coordinate their choice to maximize their welfare. This leads them to choose the exchange with the smallest possible execution cost.[9]

3.3 Exchange Pricing

Due to the "tipping" equilibrium, given prices one exchange knows that it will get all of the order flow and the other exchange will get nothing. Thus, one exchange will earn zero revenue. In Bertrand-like fashion, given the prices of the other exchange this exchange has an incentive to reduce its prices in an attempt to capture the entire market.[10]

If the exchanges have chosen different systems the firm that has invested in the favored technology can always undercut its rival. Again without loss of generality assume that exchange 1 has chosen the favored A technology; regardless of the prices that exchange 2 charges, exchange 1 can always find

[8] See Shy (2000), Katz and Shapiro (1986), and Fudenberg and Tirole (1999) for discussions of equlibrium selection in multi-stage network games. Fudenberg and Tirole state that the coordination assumption is the standard selection criterion. The model of Farrell and Saloner (1985) produces a coordination outcome in a dynamic network game with symmetric information.

[9] For equities, listing choice can serve as a coordination mechanism. A firm can choose to list its stock on the exchange that offers the lowest potential execution cost because this maximizes the value of its shares.

[10] One technical aside is relevant here. Note that because an increase in f_i reduces σ_i^2, holding $\hat{N}_i$ fixed λ_i is decreasing in f_i. However $d\lambda_i/df_i > -1$, so $dx_i/df_i > 0$. That is, even though λ_i is decreasing in f_i, execution costs are increasing in f_i. Thus, an exchange must lower its fees to reduce execution costs.

a per trade fee and terminal rental charge such that noise traders choose exchange 1. Exchange 2 has an incentive to cut prices to $f_2 = 0$ and $R_2 = 0$. Given this choice by exchange 2, exchange 1 can charge a per trade fee of $f_1 = \delta - \epsilon$ and $R_1 = \Pi_L$, where Π_L is the profit earned by the most risk averse market maker L assuming that the exchange chooses a price $f_1 = \delta$ and all noise traders and market makers trade on exchange 1. That is, exchange 1 can extract all rent earned by the marginal market maker and still capture the entire market. The exchange can also choose $R_1 > \Pi_L$. This reduces the number of market makers that choose exchange 1, which raises the execution costs of the noise traders. To ensure that it still undercuts the execution costs available on exchange 2, to charge $R_1 > \Pi_L$ exchange 1 must charge $f_1 < \delta$. Thus, there is a locus of points $[f_1, R_1]$ passing through $[\delta, \Pi_L]$ that gives the prices that the exchange offering the preferred system can charge in equilibrium.[11]

Figure 3 depicts such a locus. It is not possible to determine in general the point along it that maximizes the profit of exchange 1; for instance, the exchange may earn a larger profit by charging $R_1 > \Pi_L$ if inframarginal market makers earn large rents due to a large disparity between the risk aversion of the marginal and inframarginal liquidity suppliers. The additional terminal revenues earned by charging $R > \Pi_L$ may exceed the per trade fees the exchange must forego when it raises trading costs by pricing the least

[11] For equities, an exchange may also be able to charge listing fees. Indeed, the exchange may choose $f = 0$ and $R = 0$ and charge a listing fee that capitalizes the flow of δN^*, where N^* is the number of noise traders who submit orders when $f = 0$ and $R = 0$. This generates a larger revenue than that produced by charging a per trade fee of δ and a listing fee of zero because volume is greater with the smaller fee. The use of a listing fee allows the exchange to reduce deadweight loss caused by supermarginal cost pricing and to capture the additional surplus.

efficient liquidity supplier out of the market. Hereafter, Π^* will denote the profit earned by the firm with the favored technology from the maximizing choice along this locus.

If the exchanges choose the same system design, the Bertrand-style undercutting implies that neither firm can earn positive revenues in equilibrium.

If only one firm invests in a system, it is obviously a monopoly. In this case, the firm can earn a $\overline{\Pi} > \Pi^*$ if it has created the system that traders prefer, and a profit $\underline{\Pi}$ (which may be smaller than Π^*) if it has invested in the disfavored system.

3.4 Exchange Technology Choice

It is clearly not an equilibrium for both exchanges to select the same system. If they do, each incurs a cost of c but earns no revenue. Thus, there are two possible pure strategy equilibria:

1. One firm invests in system A and the other invests in system B.

2. Only one firm invests.

The first possibility obtains if $.5\Pi^* - c > 0$. In this case, each firm's expected revenue is $.5\Pi^*$ because there is a fifty-percent chance that its system will prove superior, and its cost of creating this system is c. The second possibility obtains if $.5\Pi^* - c < 0$. In this case, neither firm expects to break even if both invest (even if they choose different systems). Thus, only one firm will invest in the pure strategy equilibrium and earn an expected profit of $.5\overline{\Pi} + .5\underline{\Pi} - c$.[12]

[12]Of course, if this expression is negative neither firm will invest.

3.5 Welfare

As is common in network models with fixed costs, welfare analysis is complicated. If $\delta > 0$, the winning exchange charges a price that exceeds marginal cost, which leads to a costly output distortion; noise traders trade less than they should. However, the firm must earn revenues that exceed variable costs to cover the fixed costs of creating the system; supermarginal cost prices or a positive rental fee or both are needed. If δ is sufficiently small, only one firm will invest, in which case large (monopolistic) output distortions will obtain. If δ is sufficiently large, both firms with invest, and the larger the value of δ the larger the output distortion. In general, the equilibrium in the model is unlikely to be first best or even second best. That is, equilibrium investment decisions and pricing typically will not maximize the sum of noise trader surplus, market maker profit, informed trader profit, and exchange operator profits net of exchange investment subject to the Ramsey pricing constraint that exchange prices and fees just cover the costs of the optimal investment.

4 The Evolution of Industry Structure

The formal model implies that the network effects attributable to liquidity will lead to the survival of a single exchange. The winner who takes all offers the technologically superior system, and its profitability depends on how much traders value this functional superiority over its rival.

More generally, the analysis implies that the competition between investor-owned for-profit electronic financial markets is primarily technological in nature. That is, an incumbent exchange can be unseated only if a rival develops a system with functionality that users value more than that offered by the incumbent. Moreover, exchange market power is determined by its qualitative superiority over its rivals; the greater this superiority, the greater the price that the firm can charge.

Presumably the winner in the initial rivalry between electronic systems will enhance system performance through continued innovative efforts and learning-by-doing facilitated by feedback from system users. Learning-by-doing and information flows attributable to close relationships with system users are likely to provide advantages to the incumbent firm over potential future rivals.

A new rival can supplant the incumbent only by developing a system that users prefer by an amount that exceeds the cost of switching from the existing system. In fact, this advantage must exceed switching costs sufficiently to allow the new exchange to cover the costs of developing the system. Thus, the prices that the incumbent exchange can charge, and hence the likelihood of entry, depend on these switching costs.

In other network industries, such as computer software, switching costs can be large. There are several likely sources of switching costs in electronic financial markets. First, backoffice, recordkeeping and database functions may be very costly to transfer from one system to another. Second, the incumbent exchange may have private information about its customers (including their financial condition and their servicing needs) that (a) reduces its costs of serving them, and (b) is costly for the rival to develop. Third, users may require retraining to use the rival's system. Many electronic exchanges have attempted to reduce these switching costs by creating application programming interfaces (APIs) that allow system users to employ their own interfaces and to connect the trading system with their own trade processing and recordkeeping systems. Fourth, although the model assumes that it is costless to coordinate the movement of traders to the new system, in actuality this may be costly.

Although switching costs may make protect an incumbent and give it considerable pricing power, there are historical instances in financial markets in which a rival has supplanted an incumbent and "tipped" the market. The most spectacular example of this is the Bund (i.e., German government bond) futures market. Bund futures were initially traded via open outcry on the London International Financial Futures and Options Exchange (LIFFE). In 1992 group of German banks created the rival Deutsche Terminborse (DTB–later Eurex) based on an electronic trading platform. In the first several years of DTB/Eurex's existence, it was able to secure only about one-quarter of trading in Bund futures, with its order flow originating primarily from the German banks that owned it. In 1997-1998, however, Eurex undertook a concerted effort to induce the coordinated defection of firms from throughout Europe and North America trading Bund futures on LIFFE. Through a campaign of fee cuts, fee holidays, and intensive marketing, Eurex succeed in tipping the market rapidly. By the end of 1998, virtually 100 percent of all Bund futures trading had migrated to Eurex. This, in turn, induced LIFFE to abandon floor trading and to attempt to compete with Eurex by developing an innovative electronic trading platform of its own (LIFFE Connect).

Given the importance of switching costs and coordination, one can expect that electronic exchanges may attempt to use contractual terms and loyalty contracts to raise switching and coordination costs and thereby preserve and perhaps increase their rent streams. For instance, by signing long term exclusive contracts with 50 percent (plus ϵ) of the uninformed traders, the incumbent exchange can preclude competitive entry. Under certain conditions, this strategy can succeed in equilibrium because those signing the contracts fail to internalize the cost attributable to a decline in competition that results from these contracts.[13]

The network economics of financial trading can also help explain recent developments in US markets. Most important, incumbent open outcry exchanges have either developed their own electronic platforms (such as NYMEX Access) or entered into strategic alliances with electronic exchange operators (such as the CBOT-Eurex alliance) even though their members are strongly in favor of retaining open outcry markets. These actions are strategically sensible ways to extend the life of open outcry markets. By moving first and making sunk investments in electronic systems the exchanges may preempt entry by rivals and thereby preserve open outcry for some time.

5 The Survival of Multiple Electronic Exchanges

The formal model makes the very strong prediction that only one exchange trading a particular asset or contract will survive. This is the consequence of the model's particular assumptions. Under other conditions multiple exchanges may survive.

First, the model assumes that exchanges treat informed and uninformed

[13] For examples of the potential for anticompetitive exclusion in the presence of network effects, see Rasmusen *et al* (1991) and Segal and Whinston (2000). See Whinston (2001) for a summary of this literature.

orders the same. In fact, it may be possible to identify some (but not all) uninformed traders. Screening can take place in a variety of ways. Some traders may get a reputation for being uninformed; block markets use reputation to identify some uninformed traders. Alternatively, some trading mechanisms may help screen for the uninformed; uninformed traders may be more patient than informed ones, and thus more willing to trade on systems that do not offer immediacy.

Pirrong (2001a, 2001b) shows that an exchange that limits its trading to the verifiably uninformed can survive side-by-side with an exchange that does not so limit its dealings and instead serves all market order submitters on a non-discriminatory fashion. Thus, in the electronic markets we may observe a primary exchange that deals in a non-discriminatory fashion along with one or more alternative trading venues that use various mechanisms to identify the uninformed.

Second, the model assumes that system users are homogeneous. That is, they all prefer system A to system B (or *vice versa*). Given the myriad dimensions on which trading systems can differ, it quite possible that some traders prefer A and some prefer B. Under these circumstances, two systems can survive, with one system catering to each type of customer (Shy, 2001). Some network models imply that these exchanges would be likely to operate an intermarket linkage because this weakens price competition between them. These network models imply that sellers of systems designed to cater to different clienteles find it profitable to make their systems compatible in order to mitigate price competition. Financial exchanges can make their systems "compatible" by linking them and pooling order flow and liquidity. For instance, the exchanges could route all orders to a central facility for execution. By doing so, each exchange has no incentive to cut prices to increase the size of its order flow "network." This attenuation of price competition can raise joint profits.

Such linkage would represent an interesting contrast to traditional intermediary-owned exchanges, which have been notably reluctant to create intermarket linkages. For instance, the NYSE and regional exchanges created the Intermarket Trading System only under pressure from Congress and the SEC, and successfully resisted the creation of a Central Limit Order Book (CLOB) that would have fully integrated exchange order flows.[14] Similarly, SEC rules forced NASDAQ to increase the integration of ECNs into the NASDAQ market. As another example, US options exchanges abstained from cross listing option issues until compelled to do so by legal and regulatory pressure from the SEC and DOJ. Even after acquiescing to cross listing they are reluctant to create an options CLOB or its effective equivalent.

The reluctance of intermediary-owned exchanges to make their markets "compatible" through intermarket linkages plausibly reflects the fact that the owners of these exchanges profit primarily through trading against order flow, rather than through sharing in the revenue generated through exchange fees. In the case of a traditional non-profit exchange, the non-distribution constraint prevents the exchange from distributing any of the proceeds from exchange fees to its members. Thus, for intermediary-owned exchanges, sharing order flow does not provide the strategic benefit of softening price compe-

[14] ITS creates a very weak linkage between markets. See Blume, Siegel, and Rottenberg (1993) for a discussion of ITS and the resistance of the NYSE to creating a more complete linkage based on a central limit order book.

tition, but instead serves only to increase the competition for order flow that their market maker owners face. Consequently, if trader heterogeneity makes it possible for multiple exchanges to survive, for-profit, investor-owned electronic exchanges are likely to be more likely to pursue intermarket linkages than their non-profit predecessors.

6 Summary and Conclusions

A major technological revolution is ongoing in financial markets; traditional floor-based trading in equities and derivatives on non-profit exchanges is giving way to electronic trading on for-profit exchanges. Moreover, although some electronic exchanges are owned by the market makers who trade on them, others are investor-owned or in the process of transitioning to investor ownership. This article examines the macrostructure of a financial market when for-profit investor-owned exchanges compete with one another.

Fundamental microstructural considerations imply that if traders that use exchange services have uniform preferences regarding the functionality of competing trading systems, equilibrium market structure is of the "winner take all variety," with the system embedding the preferred functionality capturing 100 percent of order flow in a particular asset or financial claim. Liquidity exerts a centripetal force that attracts all trading to a single market, and if decisions of market participants can be coordinated, all trading activity will gravitate to the market with the superior technology. The prices and profits of the victorious exchange depend on how much traders value its superior functionality; the more they are willing to pay to trade on one system than another, the higher the prices and the greater the profit of the preferred exchange.[15]

This analysis suggests that the nature of competition and market structure in electronic financial markets will resemble competition and market structure in other technologically dynamic network industries, such as computer software. Technological improvement will be the main locus of competition. Moreover, switching and coordination costs will determine the degree of market power that incumbent exchanges possess. In addition, one can expect incumbent exchanges to attempt to increase these costs through various contractual means.

A comparison of the analysis of this article with that contained in Pirrong's (2001a, 2001b) analysis of competition in traditional intermediary-owned exchange markets suggests both similarities and differences. The fundamental similarity is that both are effectively winner take all network markets due to the nature of liquidity.

The primary difference is that the winner in the intermediary-owned market is the exchange that can assemble a dominant coalition of specialized market makers, whereas the winner in the for-profit electronic exchange environment is the one that can develop the superior trading technology. The traditional exchange exploits liquidity-induced network effects by restricting the number of member market makers; the reliance of such exchanges on a relatively small population of specialized market makers makes the formation of a dominant coalition of market makers feasible. In contrast, as noted by

[15] Glosten (1994) presents a model of an electronic limit order book which also predicts that the winning exchange is a natural monopoly. Glosten does not analyze the pricing of exchange services or the possibility of differences in functionality across competing exchanges. Moreover, liquidity suppliers in Glosten are risk neutral.

Black (1971) and Glosten (1994), an electronic environment has the potential to allow virtually anyone with a computer and cash to supply liquidity by submitting limit orders; this seriously undermines the ability of a "club" of market makers to dominate the market. The model predicts that the for-profit, investor-owned electronic exchange (a) will exploit network effects by charging super-marginal cost prices and (b) rather than limiting access to liquidity suppliers, desires to encourage their access so as to raise the derived demand for its services. This also implies that differences in liquidity between electronic and open outcry markets are not exclusively due to differences across systems in transparency, information flows, or the cost (including the "free option" cost) of submitting limit orders. Liquidity differences can also arise across trading platforms with different ownership structures because exchanges owned by liquidity suppliers may have an incentive to constrain liquidity suppliers whereas those not so owned do not. If there is a systematic relation between ownership structure and trading technology, this effect can create a systematic difference in liquidity costs on open outcry and electronic markets.

This analysis has policy implications. Traditional exchanges have largely escaped anti-trust scrutiny. The analysis of this article suggests that electronic exchanges may not. For-profit electronic exchanges are likely to possess market power, exhibit structural monopoly characteristics, and are will be tempted to employ contractual methods to maintain this market power that have attracted antitrust scrutiny in other markets.

References

[1] Admati, A., and P. Pfleiderer. "A Theory of Intraday Patterns: Volume and Price Variability." *Review of Financial Studies* 1 (1988): 3-40.

[2] Admati, A., and P. Pfleiderer. "Sunshine Trading and Financial Market Equilibrium." *Review of Financial Studies* 4 (1991): 443-481.

[3] Barclay, M., W. Christie, J. Harris, E. Kandel, and P. Schultz. "The Effects of Market Reform on the Trading Costs and Depths of NASDAQ Stocks." *Journal of Finance* 54 (1999): 1-34.

[4] Biais, B., P. Hillion, C. Spatt. "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse." *Journal of Finance* 50 (1995): 1655-1689.

[5] Black, F. "Toward a Fully Automated Exchange." *Financial Analyst's Journal* (1971).

[6] Blume, M., J. Siegel, and D. Rottenberg. *Revolution on Wall Street.* New York: W. W. Norton, 1993.

[7] Brown, D., and Z. Zhang. "Market Orders and Market Efficiency." *Journal of Finance* 52 (1997): 277-307.

[8] Chowdhry, B., and V. Nanda. "Multi-Market Trading and Market Liquidity." *Review of Financial Studies* 3 (1991): 483-511.

[9] Farrell, J., and G. Saloner. "Standardization, Compatibility, and Innovation." *Rand Journal of Economics* 16 (1985): 70-83.

[10] Fudenberg, D., and J. Tirole. "Pricing Under the Threat of Entry by a Sole Supplier of a Network Good." Unpublished Manuscript, Harvard University (1999).

[11] Glosten, L. "Is the Electronic Limit Order Book Inevitable?" *Journal of Finance* 49 (1994): 1127-1161.

[12] Katz, M. and C. Shapiro. "Technology Adoption in the Presence of Network Externalities." *Journal of Political Economy* 75 (1986): 822-841.

[13] Kyle, A. "Continuous Auctions and Insider Trading." *Econometrica* 53 (1985): 317-355.

[14] Pagano, M. "Trading Volume and Asset Liquidity." *Quarterly Journal of Economics* 104 (1989): 255-274.

[15] Pirrong, C. "The Industrial Organization of Financial Markets" Theory and Evidence." *Journal of Financial Markets* 2 (1999): 329-358.

[16] Pirrong, C. "Third Markets and the Second Best." Working paper, Oklahoma State University (2001a).

[17] Pirrong, C. "Securities Market Macrostructure." Working paper, Oklahoma State University (2001b).

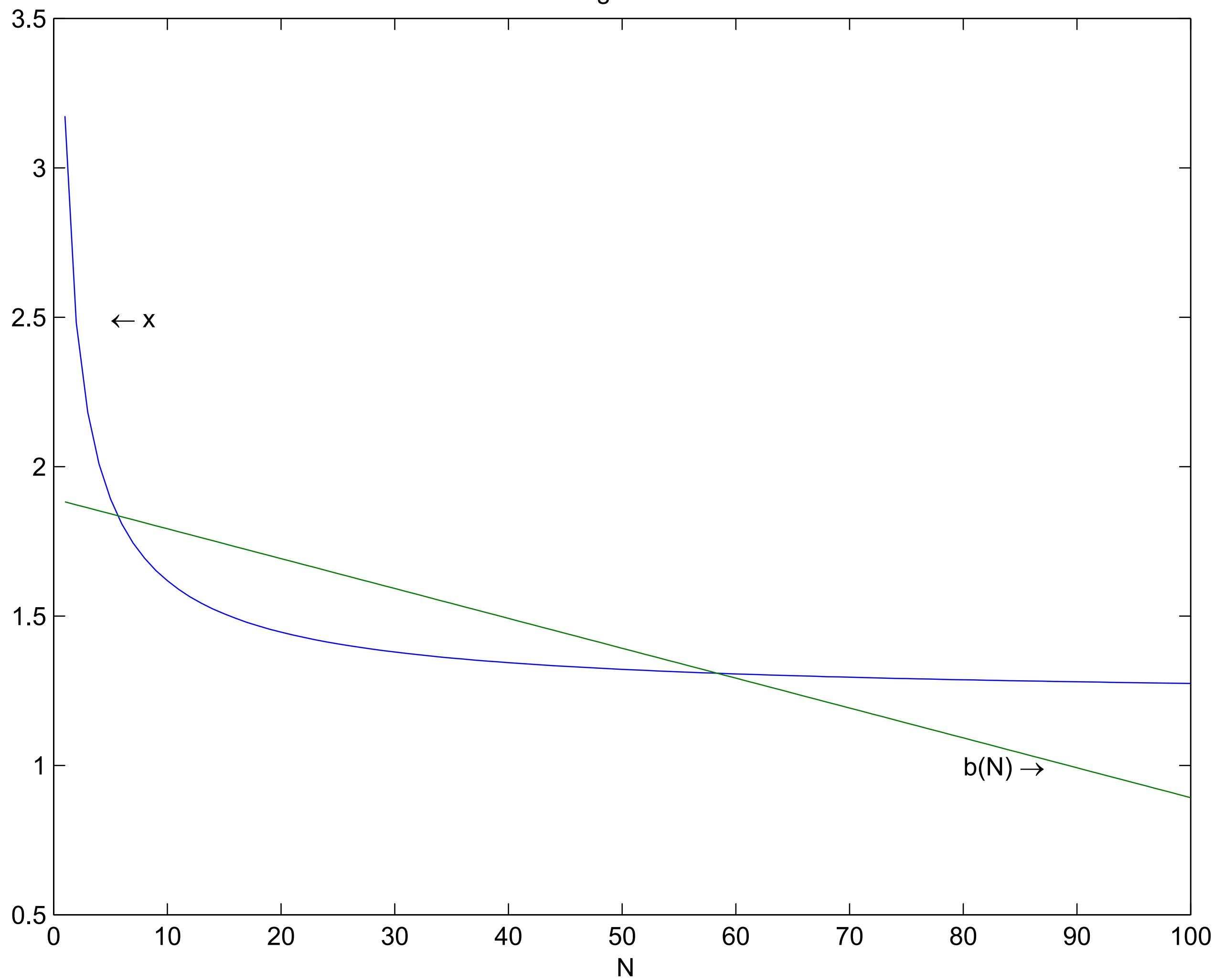
[18] Rasmusen, E., J. Ramseyer, and J. Wiley. "Naked Exclusion." *American Economic Review* 81 (1991): 1137-1145.

[19] Segal, I., and M. Whinston. "Naked Exclusion: A Comment." *American Economic Review* 90 (2000): 296-309.

[20] Weston, J. "Competition on the Nasdaq and the Impact of Recent Market Reforms." *Journal of Finance* 55 (200): 2565-2598.

[21] Whinston, M. "Exclusivity and Tying in *U.S. v. Microsoft:* What We Know, and Don't Know." *Journal of Economic Perspectives* 15 (2001): 63-80.

Figure 1



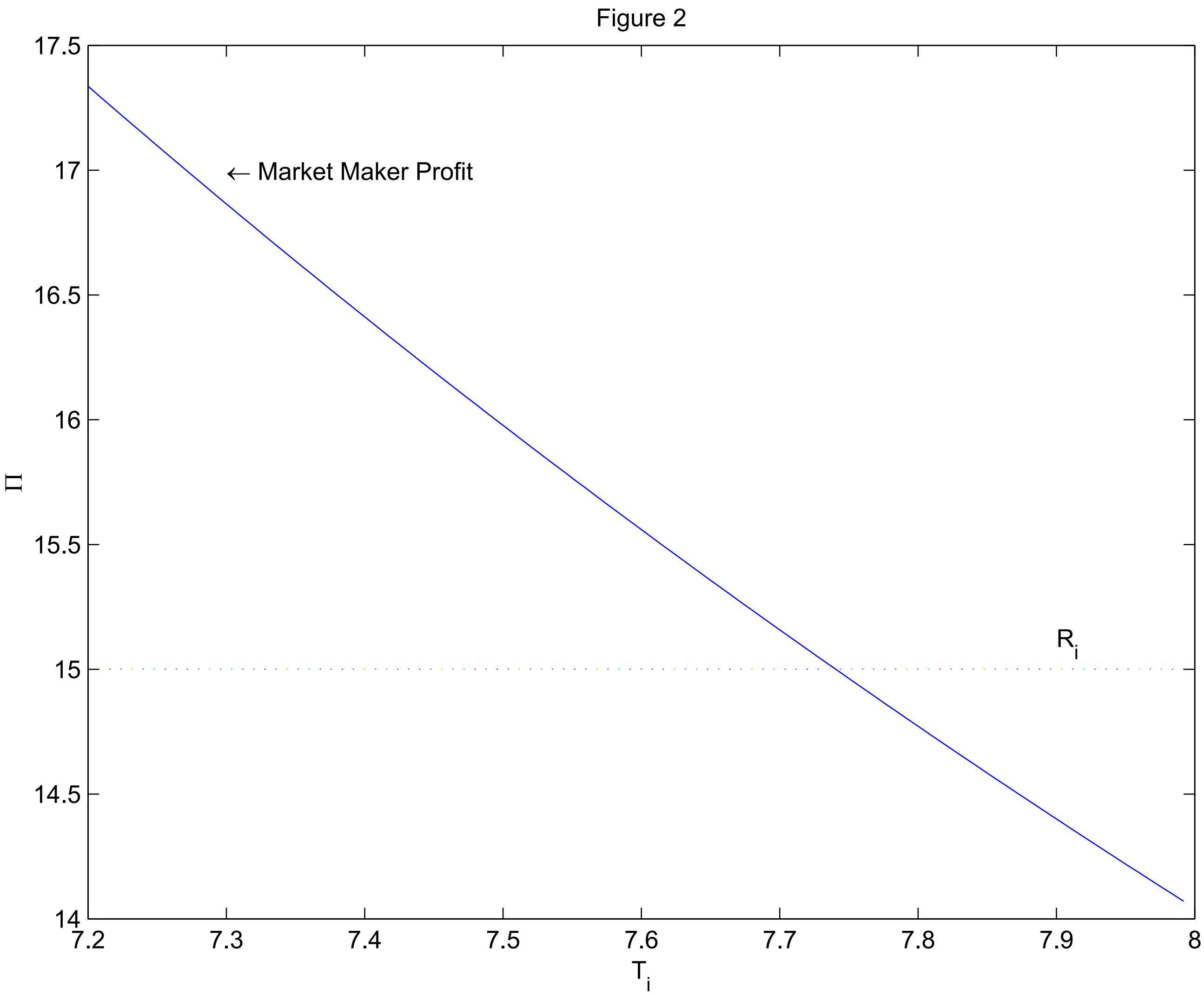


Figure 2
17.5
17
16.5
16
15.5
15
14.5
14
7.2
7.3
7.4
7.5
7.6
7.7
7.8
7.9
8
Π
T_i
← Market Maker Profit
R_i

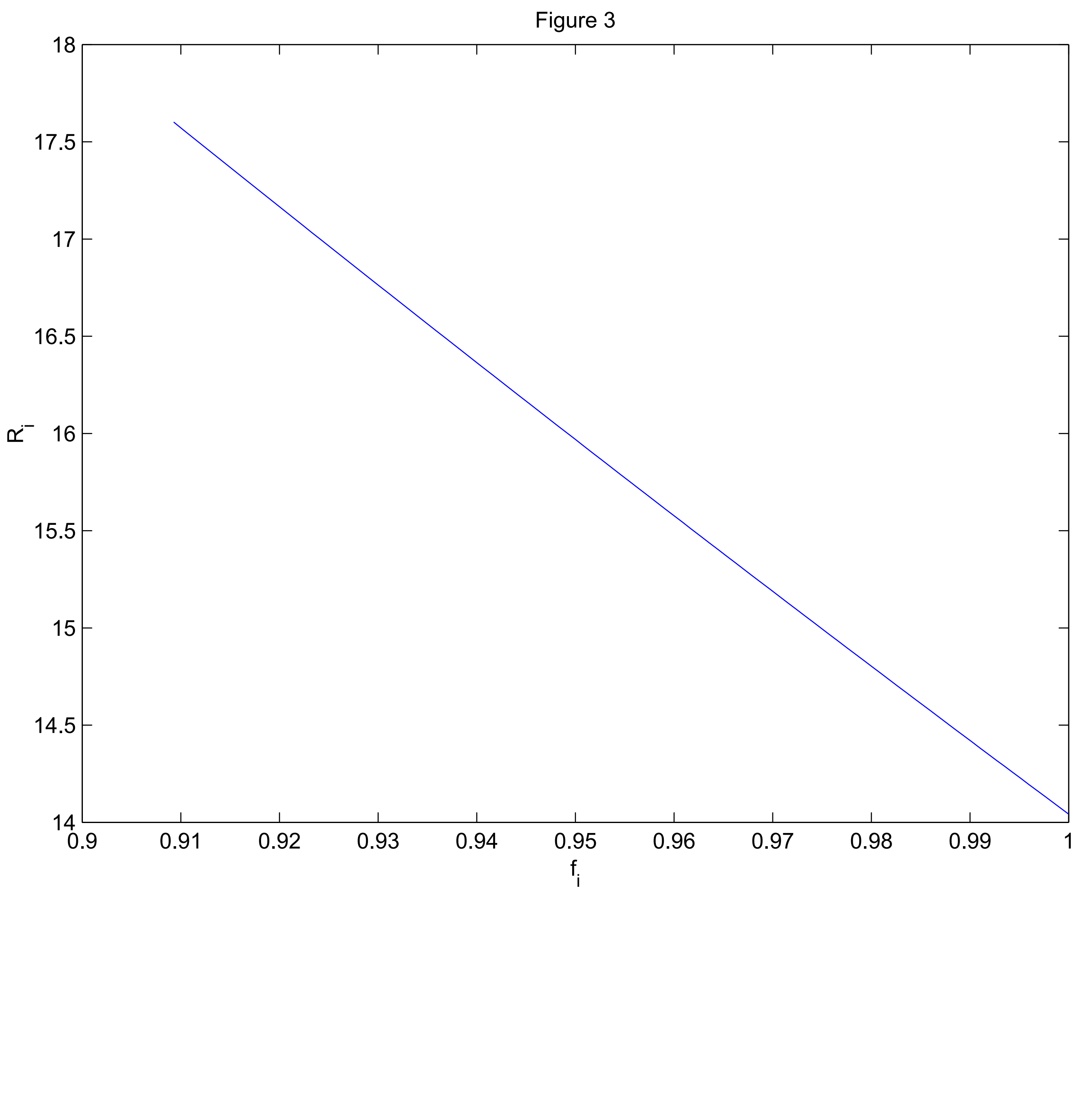


Figure 3
18
17.5
17
16.5
16
15.5
15
14.5
14
0.9
0.91
0.92
0.93
0.94
0.95
0.96
0.97
0.98
0.99
1
R_i
f_i