The Industrial Organization of Execution, Clearing and Settlement in Financial Markets

Craig Pirrong
Bauer College of Business
University of Houston

October 9, 2007

Abstract. The execution, clearing, and settlement of financial transactions are all subject to substantial scale and scope economies which make each of these complementary functions a natural monopoly. Integration of trade, execution, and settlement in an exchange improves efficiency by economizing on transactions costs. When scope economies in clearing are more extensive than those in execution, integration is more costly, and efficient organization involves a trade-off of scope economies and transactions costs. A properly organized clearing cooperative can eliminate double marginalization problems and exploit scope economies, but can result in opportunism and underinvestment. Moreover, a clearing cooperative may exercise market power. Vertical integration and tying can foreclose entry, but foreclosure can be efficient because market power rents attract excessive entry. Integration of trading and post-trade services is the modal form of organization in financial markets, which is consistent with the hypothesis that transactional efficiencies explain organizational arrangements in these markets.

1 Introduction

There is an old saw in the military that amateurs talk tactics, but professionals talk logistics. There is a clear parallel in financial markets; although the logistics of the financial markets–clearing and settlement–attract little popular attention, they are matters of primary importance to market professionals. Of late, however, the organization of clearing and settlement functions has assumed a greater public profile. In particular, the integration of the clearing, settlement, and execution of transactions within a single firm has drawn considerable criticism and anti-trust scrutiny, especially in Europe.[1] This was also a central issue in the Chicago Mercantile Exchange-Chicago Board of Trade merger.

The primary objection to the "vertical silo" model of financial trading is that due to the substantial scale and scope economies in clearing, an exchange integrated into all three functions can deny access to its clearing and settlement facilities to foreclose entry by another firm offering execution services in the products traded on the integrated exchange. It has long been understood, however, that vertical integration can also be an efficient way to organize transactions, because it can eliminate double marginalization, mitigate holdups, and provide superior incentives to invest in specific assets. Understanding the motives for vertical integration in financial markets and its welfare effects therefore requires a detailed microanalytic analysis of the economics of trading, clearing and settlement.

This article examines these issues theoretically and empirically. It derives

[1]UK Competition Commission (2005). The Competition Commission found that merger of the LSE and either Deutsch Börse or Euronext would give the combined entity "the incentive and ability to foreclose entry and expansion in the UK at the trading level." European industry groups have also expressed opposition to vertical integration on competitive grounds. French Association of Investment Firms (2006).

several conclusions:

- Trade execution, clearing, and settlement are all subject to pervasive economies of scale, and have strong natural monopoly tendencies. Moreover, specific assets and coordination problems imply that potential competition is unlikely to discipline the market powr of these natural monopolies, and that competition for them will be dissipative.

- Integration of clearing, settlement, and trade execution mitigates inefficiencies resulting from multiple-marginalization and *ex post* opportunism, but is unlikely to lead to a first-best outcome.

- Vertical integration is more efficient than supplying trade execution, clearing, and settlement by separate, for-profit firms, and hence the latter arrangement is unlikely to be observed in practice.

- Cooperative ownership by market users of clearing or settlement functions can mitigate double marginalization if these cooperatives admit all qualified participants and rebate surpluses to members in proportion to their output. However, a cooperative may have an incentive to restrict entry or charge supra-competitive prices, which results in double marginalization. Moreover, arms length relationships between trade execution venues and clearing and settlement cooperatives results in transactions costs that integrated exchanges do not incur.

- Clearing and settlement exhibit strong scope economies. There are also sources of scope economies in trade execution, but it is by no means clear that these execution scope economies are as strong as those that characterize clearing and settlement. To the extent that there are scope diseconomies in trade execution, integration can result in

suboptimally small clearing and settlement operations, suboptimally large execution operations, or both.

- Clearing and settlement cooperatives are more likely to be observed when scope economies in these services are more extensive than scope economies in trade execution. Exchange ownership stakes in the cooperative economizes on the transactions costs of this form of organization.

- Although in theory an incumbent integrated exchange can exploit its control of clearing to foreclose of entry into execution, if execution is potentially highly competitive it has no incentive to do so. If execution is not highly competitive (due to network effects, as is plausible) then foreclosure can be efficient because it prevents dissipative entry, or superfluous, because the incumbent's execution business is not vulnerable to entry even absent tie between clearing and execution.

- Historically, most exchanges that execute transactions in securities or derivatives have integrated into clearing and settlement. Moreover, when clearing and settlement entities are not wholly owned by a single exchange, the exchanges they serve almost always have ownership and voting stakes in them.

- Vertical integration remains the dominant form of organization in securities and derivatives markets. There is only one prominent example of an exchange that has no ownership in or control of its clearing and settlement entity.

- Clearing and settlement cooperatives serve some markets. These cooperatives typically provide services to multiple exchanges (and some-

times over-the-counter markets) where transactions are executed. Exchanges typically have an ownership stake in these firms.

In sum, as transactions cost economics inspired-theory predicts, exchanges that execute financial transactions typically exercise considerable ownership and control over clearing and settlement. Only where economies of scope in clearing and settlement are substantially more extensive than economies of scope in execution have exchanges' control over these functions diminished.

The organization of the trading of securities (and to a lesser degree derivatives) has been a subject of enduring controversy. The network nature of liquidity means that financial trading faces many of the same challenges and conundrums as other industries, such as telecommunications and electricity, where network effects are also present (and more widely recognized) (Pirrong 2002a, 2005). Clearing and settlement contribute additional sources of scale and scope economies that further challenge competition in financial markets. As in any network industry, there are no easy organizational answers. Suffice it to say that there are strong economic reasons to integrate trade execution, clearing, and settlement, and these functions have been commonly integrated in practice. Although this integration is unlikely to produce a first best outcome, because the integrated entity is likely to exercise market power, alternative forms of organization do not obviously increase competition appreciably, and plausibly create other distortions.

The remainder of this paper is organized as follows. Section 2 describes the clearing and settlement of securities and derivatives transactions. Section 3 analyzes the economics of trade execution, clearing and settlement, and demonstrates that there are extensive scale and scope economies in each of these functions. This section also examines how these scale and

scope economies affect the efficient organization of execution, clearing, and settlement. It also analyzes vertical foreclosure. Section 4 presents historical evidence relating to the organization of firms engaged in trade execution, clearing, and settlement. Section 5 briefly summarizes the article.

2 Trade Execution, Clearing, and Settlement

The completion of a financial transaction typically involves a variety of complementary activities.

The first function is the execution of a transaction; that is, the consummation of an agreement between a buyer and a seller. In over-the-counter markets, buyer and seller typically complete deals over the phone. In exchange markets, orders to buy and sell are directed to a central marketplace– the exchange. In a traditional floor-based, open outcry exchange, orders to buy or sell are represented by agents (floor brokers) on the exchange floor, or by exchange members physically present on the exchange dealing on their own account. Buyers and sellers (or their agents) on the exchange floor agree to the terms of a transaction through a negotiation or auction process. In newer, computerized exchanges, orders are routed electronically to a central computer which matches buy and sell orders based on priority algorithms.

Once the buyer and seller agree to terms, a transaction must be cleared. The clearer first verifies that all terms submitted by the buyer and seller match. In most centralized markets, the clearing entity is then substituted as a principal to the transaction, becoming the buyer to the seller, and the seller to the buyer. That is, the clearer becomes the central counterparty ("CCP") that bears the risk of default by those with whom it transacts.[2]

[2]This process is somewhat intricate. See Edwards (1983) or Pirrong (2006) for detailed

That is, CCPs bear performance risk.

In their role as CCP, clearers–typically referred to as "clearinghouses"– engage in a variety of activities, including: calculation and collection of collateral (margin); determination of settlement obligations (that is, the determination of what each party owes or is owed in money and delivery obligations); determination of default; collection from defaulting parties; and remuneration of participants in the event of a default. The CCP usually nets the obligations of those for whom it clears. That is, it determines the net amount each part owes or is owed; since a party may owe money on some transactions, and be owed money on others, netting typically reduces the flows of cash (and securities) between transacting parties. As will be seen, this netting function is economically very important.[3]

Clearers service the financial intermediaries who broker customer orders, and who sometimes trade on their own account. That is, clearinghouses serve as a central counterparty only to so-called "clearing brokers," and collect margins, collect and disburse variation payments, and charge fees from/to these brokers. They typically do not deal directly with the ultimate buyers or sellers for whom the brokerage firms serve as agents.

Settlement is the process whereby parties discharge their contractual obligations to pay cash or deliver securities.[4] At one time, settlement agents

descriptions of this "novation" process. Not all organized exchanges have CCPs. CCPs have been widespread in derivatives markets since the late-19th and early-20th centuries, but have been introduced in equity markets only more recently.

[3]See The Optimal Structure for Clearing and Settlement in the EU: Citigroup Reponse to the Communication on Clearing and Settlement from the European Commission for a detailed description of clearing. Some non-CCP clearinghouses also engage in netting.

[4]Some derivatives transactions involve the delivery of securities or financial instruments. For instance, a corn futures contract not previously offset is settled by the delivery

facilitated the physical delivery of stock certificates, bonds, or other delivery instruments. Presently, delivery is performed by debiting or crediting the securities and cash accounts of the counterparties to transactions. This typically involves the maintenance of a central register that records ultimate ownership of securities.

A securities or derivatives transaction involves all three functions. Thus, these functions are complementary, and the demand for each service is a derived demand. This has important implications for the organization of financial markets, as is discussed in the next section.

3 Scale and Scope Economies in Trading, Clearing and Settlement

3.1 Scale Economies in Trading: The Liquidity Network

The execution of transactions in financial instruments (including securities and derivatives) is subject to substantial economies of scale due to the nature of liquidity. It is typically cheaper to execute transactions in markets where large numbers of other transactors congregate.

There are a variety of formal models that demonstrate that trading of financial instruments is subject to network economies that cause average trading costs to decline with the number of traders.[5] These trading costs include the bid-ask spread and the price impact of trades. Moreover, the

of a shipping certificate; at one time it was settled by delivery of a warehouse receipt. As another example, Treasury note or bond futures contracts not previously offset are settled by delivery of a United States Treasury security.

[5]Pirrong, (1995), Pagano (1989), Admati and Pfleiderer (1988), Economides and Siow (1988).

extant empirical evidence is consistent with these predictions.[6]

Informed trading is the crucial source of these network economies. Informed trading imposes adverse selection costs on those who do not possess private information. The uninformed mitigate their exposure to adverse selection by congregating on a single trading venue.

These models imply that the trading of financial instruments is "tippy." That is, trading activity in a particular instrument should gravitate to a single platform or exchange. With multiple exchanges, the exchange with the larger number of participants exhibits lower trading costs. This attracts traders from the smaller exchanges, which exacerbates the cost disparities, attracting yet more defections to the larger venue. Absent strong clientele effects, in equilibrium this process results in the survival of a single exchange.[7]

Empirical evidence is consistent with this tipping hypothesis (Pirrong, 1999, 2002). In practice, it is known that sometimes trading in financial instruments (notably equities) fragments, with a given security being traded on several venues. However, I have shown theoretically that this fragmentation is typically a form of "cream skimming" whereby orders submitted by those who are verifiably uninformed are executed off-exchange, while all orders that are not verifiably uninformed are submitted to a dominant exchange (Pirrong, 2002). For instance, off-exchange block trading mechanisms attempt to screen out the informed traders and limit participation to those whom are unlikely to have private information about valuations.

[6]Pirrong, (2006b) summarizes this evidence.

[7]Cross-border trading restrictions are one potential source of clientele effects. In the days before efficient telegraphic or telephonic communication, geographic proximity gave rise to clienteles. Arnold, Hersch, Mulherin, and Netter (1999).

Extensive empirical evidence shows that trades executed away from the primary exchange typically have less information content than those executed on the primary exchange.[8] Thus, theory and empirical evidence suggest that trading activity that is not verifiably uninformed tips to a single venue. Put differently, price discovery is a natural monopoly.

Potential competition is unliely to discipline strongly the market power of the natural monopoly in the trading of a particular product. Exchanges must incur sunk costs in specific assets to enter. A traditional open outcry (floor) exchange must construct a specialized trading facility that has no use other than that for which it is designed. Moreover, floor traders invest in specific human capital that is of little use in other professions. Modern electronic exchanges create specialized trading systems involving investments in hardware and (especially) software that has little (if any) value in other uses. In addition, the customers of electronic exchanges invest in linkages customized to a particular exchange to connect it. Thus, both open outcry and computerized trading exchanges incur sunk costs, and customers incur costs to switch exchanges. Finally, for both open outcry and electronic exchanges, to compete on liquidity an entrant must attract the near simultaneous defection of a large number of traders on an incumbent exchange. Coordinating this movement is costly, and these coordination costs are sunk once incurred (Pirrong 1995). Sunk costs in physical trading infrastructure and human capital, switching costs, and coordination costs, all impair the ability of an entrant to compete with an incumbent monopoly trade execution venue.

[8]Bessembinder and Kaufman (1997), Easley, Kiefer, and O'Hara (1996), Smith, Turnbull, and White (2001). Recently, intermediaries hae created "dark pools" of liquidity that match orers at prices determined in other markets. These dark pools also serve as cream skimming venues.

3.2 Scale Economies in Clearing and Settlement: The Single Product Case

There are a variety of sources of scale economies in clearing and settlement. With respect to clearing, there are operational scale economies, and importantly, economies of scale in bearing default risk.

To understand the sources of scale economies in bearing default risk, first consider a simple model of a one-product market. A major cost of operating the clearing system is the capital the CCP requires to absorb defaults and thereby insure performance on non-defaulting contracts.[9] A clearinghouse is able to ensure that all contracting parties that are "in the money" receive what they are owed as long as its capital is larger than the losses from defaults by holders of contracts that are "out of the money." This capital is costly. A simple formal model demonstrates that the amount of capital required to ensure the clearinghouse's performance in response to a given set of defaults is subadditive. That is, the amount of capital required to ensure that one clearinghouse performs is smaller than the amount of capital that is required to ensure performance by two or more clearinghouses serving the same set of customers and (collectively) experiencing the same set of defaults.

To see this, consider first the case with several clearinghouses. At time 0 clearinghouse i is the central counterparty to transactions with customers $j = 1, \ldots, N$. These deals all mature (expire) at time 1. At time 1, customer j pays to or receives from clearinghouse i an amount given by the random variable $\tilde{v}_{ij}$. If $\tilde{v}_{ij} > 0$, the customer owes the clearinghouse money (because the price has moved against this trader); if $\tilde{v}_{ij} < 0$ the clearinghouse owes

[9] Barzel (1997) argues that the role of equity capital is to bond the performance of contracts. It definitely serves this purpose in clearing.

the customer. At time 1, customer j has equity (gross of his obligations to clearinghouse i) equal to the random variable $\tilde{V}_j \geq 0$. The clearinghouse receives the entire contractual payment only if $\tilde{V}_j \geq \tilde{v}_{ij}$. If $\tilde{V}_j < \tilde{v}_{ij}$, the customer cannot pay what he owes, so the clearinghouse receives $\tilde{V}_j$ instead of the full contractual amount. Thus, the payoff to the clearinghouse in his dealings with customer j is $\min[\tilde{v}_{ij}, \tilde{V}_j]$. Note that the payoff to the contract with customer j is an option on the minimum of two risky assets.

At time 1 clearinghouse i has capital given by the random variable $\tilde{E}_i$, and the clearinghouse owes payments to counterparties on deals such that $\tilde{v}_{ik} < 0$. (Note that $\min[\tilde{v}_{ik}, \tilde{V}_k] = \tilde{v}_{ik} < 0$ for such deals.)

All counterparties with contracts that are "in the money" receive full payment from the clearinghouse if and only if:

$$\tilde{Z}_i = \sum_{j=1}^{N} \min[\tilde{v}_{ij}, \tilde{V}_j] + \tilde{E}_i \geq 0 \tag{1}$$

If this inequality does not hold, the clearinghouse defaults. In this case, all customers with $v_{ij} < 0$ receive less than the full contractual payment.

Consider another clearinghouse h that transacts with customers $k = 1, \ldots, K$ at time 0. This clearinghouse does not default if:

$$\tilde{Z}_h = \sum_{k=1}^{K} \min[\tilde{v}_{hk}, \tilde{V}_k] + \tilde{E}_h \geq 0 \tag{2}$$

Now consider the losses from default when i and h merge. No customer of the merged clearinghouse who is owed money receives less than full payment if:

$$\sum_{j=1}^{N} \min[\tilde{v}_{ij}, \tilde{V}_j] + \tilde{E}_i + \sum_{k=1}^{K} \min[\tilde{v}_{hk}, \tilde{V}_k] + \tilde{E}_h \geq 0 \tag{3}$$

It is readily evident that there are some states of the world (i.e., some possible realizations of $\{\tilde{v}_{ij}, \tilde{v}_{hk}, \tilde{V}_j, \tilde{V}_k, \tilde{E}_i, \tilde{E}_h\}$) such that (3) holds, but (1),

(2), or both do not, i.e., (1) and (2) are sufficient, but not necessary, for (3) to hold. In essence, default costs are subadditive. Losses from a clearinghouse default have the characteristics of the payoff to an option on a portfolio. The default costs of N separate clearinghouses equal the value of N of options on N portfolios. The default costs if the N clearinghouses merge are the same as the payoff on single option on the N portfolios. The subadditivity in default costs obtains because an option on a portfolio is less costly than a portfolio of options (Merton, 1973). Therefore, due to the non-linearity of the clearinghouse payoff function, customers lose from defaults in fewer states of the world if the clearinghouses merge.

Equivalently, merger reduces the amount of equity capital required to generate the same expected customer loss from clearinghouse default that would be incurred with multiple clearinghouses serving the same set of customers. Thus, just as there are network effects in trade execution, there are network effects in clearing as well; the more customers that join a particular clearing "network," the lower that network's costs.

There are other sources of scale economies in clearing. In particular, a clearer must create the necessary software and information technology infrastructure. The software investment in particular exhibits scale economies. Moreover, maintenance and operation of the clearing IT does not vary strongly with the number of transactions cleared.

The creation of multiple clearers for the same instrument imposes additional costs. Multiple clearers must maintain costly communication links to match transactions when the counterparties submit trade details to different clearinghouses. They must also maaintain financial linkages to permit the flow of cash between counterparties who use different clearers. The number of communication and banking linkages, and hence the costs thereof, rise

geometrically with the number of clearing firms. Multiple clearers also require costly collateral to bond performance with one another. Inter-clearer margin would be unnecessary with a single clearinghouse.

Settlement costs are also largely fixed. As with clearing, settlement requires the creation of a software and hardware infrastructure, both of which (especially software) involve a large fixed component.[10] Econometric analysis also documents large scale economies in settlement (Schmeidel, Malkamaki, Tarkka, 2002). Multiple settlement entities also necessarily create and maintain costly communications and financial linkages, and require costly collateral to bond settlement performance risk.

3.3 The Organizational Implications of Sale Economies in Trading and Post-Trade Operations

As noted in section 3.1, execution of transactions in a particular financial instrument is subject to strong network/liquidity-driven scale economies, and that as a result there are strong natural monopoly tendencies in trading. Similarly, the analysis of scale economies in clearing and settlement strongly suggests that these functions are also natural monopolies. Operational effects, liquidity economies from netting, and perhaps most importantly, default cost effects all exhibit strong scale economies. Costs are higher with multiple clearinghouses than with a single clearinghouse.

These cost factors influence the efficiency of alternative forms of organizing financial transactions. One such alternative is for separate firms to perform the trade execution, clearing, and settlement functions, and for these firms to be organized as for-profit ventures.

[10]One major European settlement entity, CREST, estimated the cost of establishing a settlement depository at GBP 100 million.

If (a) clearers are separate from exchanges, and (b) multiple clearers attempt to compete, in equilibrium only one clearinghouse survives. The bigger clearinghouse can always undercut the smaller one's prices, and attract all of the latter's customers. This induces tipping of clearing to a single provider.

Nor is potential competition likely to constrain severely the natural monopoly clearer. Much of the clearer's costs are sunk. For instance, the information systems (which represent a large fraction of the clearer's costs) are specialized to the clearing function, and have little value in alternative uses. Moreover, a particular clearer creates a unique interface through which it interacts with customers. Customers incur costs to connect to this interface, and incur switching costs to connect to a new clearer. These customer costs are sunk once incurred. As with trade execution, moreover, a challenger to an incumbent monopoly clearer must coordinate the near-simultaneous defection of the incumbent's customers to overcome the latter's scale advantage; these costs are also sunk once incurred. These sunk costs (those incurred by the clearing firm and its customers) give the incumbent clearing monopoly competitive advantages over those attempting to supplant them.

Competition in a natural monopoly service like clearing is plausibly modeled as a war of attrition. This competition is dissipative. Scale economies are not fully exploited and fixed costs are duplicated, and monopoly pricing is the ultimate outcome. (Tirole, 1988.)

The same dynamic prevails at the settlement level. Again, sunk costs and coordination problems sharply constrain potential competition, likely permitting the settlement monopoly to exercise market power.

Thus, if clearing, settlement, and trade execution are supplied by firms that specialize in a single function, the strong scale economies in each tend to

result in the survival of a single firm in each function, each of which has some market power. Moreover, competition for these monopolies can wastefully dissipate market power rents. Due to the aforementioned complementarity of clearing, settlement, and execution, separate ownership, control, and pricing of these functions therefore creates a tri-lateral monopoly problem. This, in turn, creates the potential for inefficiencies.

First, due to complementarity, multiple-marginalization problems arise. Independent price setting by the three firms results in a price that exceeds the monopoly price that an integrated monopoly firm would charge.

Second, even if the exchange, clearer, and settlement agent enter into a contract (or set of contracts) that prices each firm's services in a way that avoids multiple-marginalization and ensures that the ultimate customer of financial transaction services pays the monopoly price (which maximizes the rent to be divided between the three entities), wasteful rent seeking and opportunism can arise. Recall that each entity employs specific capital, and that this capital is likely to be quite durable. These considerations lock the (putatively separate) suppliers of execution, clearing, and settlement services into long term, trilateral relationships. Due to the enduring nature of the relationships, the parties are likely to rely on long term contracts to govern their interactions. However, the specific assets of the clearer, exchange, and settlement firm give rise to quasi rents, and each firm has the incentive to engage in *ex post* opportunism to expropriate them. That is, even if the parties sign long term contracts, they have an incentive to violate the contract or evade performance in order to expropriate these quasi rents. Unpredictability in the economic environment makes complete contracts impossible, and parties can exploit this incompleteness in an attempt to profit at the expense of their contracting partners. This rent seeking utilizes real

resources.

Integration of the complementary trading functions abolishes the deadweight losses arising from multi-marginalization and opportunism. Although integration does not result in a first best outcome (because the integrated entity is a monopoly, and presumably charges supermarginal cost prices) it offers some advantages over a dis-integrated structure because it avoids the costs associated with inefficient pricing and rent seeking. It can also adapt to unpredictable changes in conditions, such as technology or regulatory shocks, that challenge contractual governance of the relationships between distinct execution, clearing, and settlement firms.

This is not to say that vertical integration is free. Due to their inability to precommit to a high powered incentive system, integrated divisions are typically operated subject to low powered compensation schemes that attenuate incentives to reduce costs and innovate. Moreover, information asymmetries between managers give rise to costly information rents and the use of low powered incentives.

Thus, standard transactions cost considerations imply that integration of trade execution, clearing, and settlement offers increases the rents to be split, makes consumers better off, and reduces deadweight losses due to the elimination of double marginalization and opportunism. Thus, integration is plausibly a second-best response to the natural monopoly characteristics of trading, clearing, and settlement.

Although vertical integration is a well-recognized way to mitigate transactional hazards, there are other ways to organize firms and to govern relationships between them in order to control transactions costs in the presence of small-numbers and specific asset problems such as those inherent in trade execution, clearing, and settlement.

For instance, a user cooperative can eliminate multiple marginalization problems; consumer cooperatives are a well-known response to market power (Hannsman, 1996).[11] Recall that brokerage firms utilize clearing services. These brokerages can form a cooperative firm that supplies clearing.[12]

Formally, consider a set of brokerage firms $i = 1, \ldots, N$ that form a clearing cooperative. The fixed cost of the clearing service is F and the marginal cost is c_c, a constant.[13] The clearing cooperative charges member broker a fixed membership fee of f_i, and a fee of P_c per trade cleared. The cooperative pays its members a patronage dividend of $P_c - c_c$ on each trade cleared. Moreover, it chooses the fixed fees so that total fixed payments just equal the fixed cost of operating the clearinghouse. Moreover, I assume that the cooperative chooses the f_i to maximize the number of brokers that participate, subject to the constraint that the cooperative covers its fixed costs. Thus, the cooperative operates on the open access principle, and just breaks even.

Clearing and execution services are consumed in the fixed proportion of one-to-one (an extreme case of complementarity.) The demand for trading services is $D(Q)$, where Q is the quantity of trades executed and cleared.

Brokers process and manage customer trades for execution and clearing. Broker i incurs a cost $C_i(q)$ to process q trades, with $C_i' > 0$, $C_i'' > 0$. There is a monopoly execution venue. For simplicity, the marginal cost of trade execution is zero.

[11] Hausman, Leonard, and Tirole, (2003) presents a model showing how a non-profit cooperative can induce an efficient outcome in a network industry.

[12] A similar argument can be applied to clearing, or to execution.

[13] This is for convenience only. The model can be extended readily to permit c_c to be a (decreasing) function of the size of the clearing network.

The monopoly execution venue chooses its price P_E. Since execution and clearing are consumed in fixed proportions, the derived demand for brokerage services is $D_B(Q) := D(Q) - P_E$.

Brokers are perfect competitors who take the price of brokerage services P_B as given. Each broker i chooses the quantity of transactions to clear q_i to maximize:

$$\begin{aligned} \Pi_i &= P_B q_i - C_i(q_i) - q_i P_C + q_i(P_C - c_c) - f_i \quad (4) \\ &= P_B q_i - C_i(q_i) - q_i c_c - f_i \quad (5) \end{aligned}$$

The fourth term on the right hand side of (4) is the patronage dividend received by broker i; note that due to this dividend, each broker acts as if the price of clearing services is marginal cost.

Given the exchange's choice of P_E, in the competitive equilibrium

$$P_B = C_i'(q_i) + c_c = D_B(\sum_{i=1}^{N} q_i)$$

where $C_i'(q_i) = C'(q_j)$, for all i, j. Consequently, when determining its price (or equivalently, its output), the execution venue perceives its derived demand to be the difference between market demand for trade execution/clearing and the marginal cost of clearing (including the marginal costs that brokers incur to process trades). This implies that the exchange chooses the same output as if it were to integrate into clearing. Thus, the cooperative alternative eliminates double marginalization, just as integration does.

This does not mean that this alternative is as efficient as, or more efficient than, integration. Several potential problems arise, including:

- The clearing cooperative cannot internalize all benefits from investments to improve productivity or improve service quality because some

of these benefits accrue to the monopoly exchange. For instance, if the cooperative invests in technology to reduce c_c, and this investment is non-contractable, the exchange's derived demand rises. In response, the exchange raises the price of execution, thereby capturing some of the cost reduction. This reduces at the margin the cooperative's incentives to invest, and leads to underinvestment.

- The foregoing analysis assumes that (a) the cooperative implements an open access policy, and (b) P_B is set competitively, that is, the cooperative does not enforce a broker cartel. Both assumptions are subject to challenge. For instance, the New York Stock Exchange and other financial exchanges were non-profit mutuals that enforced broker cartels that set minimum commissions and restricted entry by limiting the number of memberships.[14] Elsewhere I show that a cooperative natural monopoly firm can exercise market power, and allow its members to earn economic rents, by restricting membership. In these models, a particular service is subject to increasing returns.[15] A coalition of suppliers forms a cooperative that performs this service for its members. The cooperative charges these members a fee just sufficient to cover the fixed cost that gives rise to the scale economy. The members then compete for customers (those who need to clear transactions, in this instance). Due to the presence of scale economies, there

[14]Philipson and Posner (2001), and Hansmann (1996) argue that even non-profits may exercise market power even though they cannot distribute profits to their owners. For instance, they can charge supracompetitive prices for goods over which they have market power, and use the resulting profits to subsidize the production of other goods for their owner-members.

[15]In Pirrong (1999) the increasing returns arise from a fixed cost. In Pirrong (2002), network effects create a scale economy.

is a critical size of membership that is (a) smaller than optimal, but (b) just large enough so that any other competing cooperative is too small to cover its fixed costs. Due to its smaller than optimal membership, the cooperative's output is inefficiently small; this effectively results in double marginalization. Thus, to avoid this possibility, it is necessary to constrain the cooperative's ability to limit membership. This is not a straightforward task, as in the case of clearing and settlement (which involve mutualization of some risks) it is economically sound to impose financial requirements on members to mitigate moral hazard and adverse selection problems; it is no mean feat to determine whether a given financial requirement is justified as a prudent way to maintain the solvency of the clearing and settlement firm, or is instead set inefficiently high in order to restrict membership. Moreover, due to the complementarity of trade execution and clearing, when deciding on the profit maximizing membership, the cooperative ignores the impact of the resultant output restriction on the derived demand for the execution venue's services; this causes a double marginalization inefficiency.[16]

[16]The form of the cooperative's payout policy is also important. The formal model in the text assumes that a brokerage receives a rebate only on its purchases of clearing services. The appendix presents another model in which each broker receives a fixed fraction of the difference between the clearing cooperative's revenues and its costs. This would be the case where the clearing firm is a for-profit firm owned by the brokers where the equity share of each broker is fixed. In this case, each broker's payout depends on the output of the others. The model also assumes Cournot competition between the brokers. In this case, setting P_C above marginal cost affects the output of the brokerage firms, which is not true when the patronage dividend depends only each broker's own output. The appendix demonstrates that it is possible to choose the shares accruing to each broker and $P_C > c_c$ to produce the monopoly outcome (conditional on the execution venue's choice of price P_E). Thus, the for profit cooperative can facilitate collusion between the brokers. This is a result similar to the well-known finding that a revenue pool (perhaps run by a joint sales office) can implement a monopoly outcome. This results in double marginalization.

- Separation of trade execution and post-trade services can impede coordination. For instance, a change in a trading or clearing system (such as the addition of a new product for trading, or the offering of a new clearing or trading functionality such as straight-through processing) often requires changes to both the clearing and trading systems. The incentives to adopt efficient changes may not be well aligned when trade execution and post-trade services are carried out by different firms. Similarly, sometimes there is a need to coordinate responses to shocks (such as a market crash) or regulatory changes. Implementation of such changes requires negotiation across firm boundaries, which can provide an opportunity for hold up to extract the quasi rents that arise from specific investments. This impairs incentives to introduce efficiency-enhancing innovations or to respond efficiently to shocks.

- Effectively operating as a non-profit, the clearing firm's management is subject to low-powered incentives.

- If the clearing entity cannot finance fixed costs through the use of fixed assessments (due to information asymmetries, for instance), and therefore must charge a per unit fee $P_C > c_c$, there is double marginalization as the clearer's markup over marginal cost drives the exchange's derived demand for execution services below that which prevails under integration.

A vertically integrated exchange is not vulnerable to expropriation of the returns to investment, or to holdups that impede coordination. The integrated exchange has no incentive to limit brokerage participation in the clearinghouse for strategic purposes, as this reduces the derived demand for its services. On *a priori* grounds it is not possible to determine whether in-

centive power is weaker in an integrated exchange than with an (effectively) non-profit clearer. However, on balance, unless the costs of low powered incentives for an integrated firm are substantially higher than for the post-trade processor, integration dominates supply of post-trade services by a cooperative.

These problems with the cooperative solution can mitigated by extending control and ownership rights in the cooperative to the exchange. That is, shared governance–partial integration–is one means of attenuating the transactions costs associated with the separation of trade execution and post-execution service providers.

In sum, although a vertically integrated exchange that offers trade execution, clearing, and settlement services does not result in a first best outcome, alternative arrangements in which clearing and settlement are separated from execution incur deadweight costs as well. These alternatives might have some merit, as compared to vertical integration, to the extent that regulation or cooperative ownership of one segment of the industry (such as clearing and settlement) facilitates competition in another (such as trade execution), and even then only to the extent that the associated efficiency gains outweigh any efficiency losses that arise in a disintegrated industry.[17] However, in the case of financial transactions, each of the three segments of the industry has strong natural monopoly elements. The creation of a clearing cooperative, for instance, does not eliminate the centripetal force of liquidity that gives exchanges that execute exchanges considerable market power. Thus, a clearing/settlement cooperative does not eliminate the

[17]This is arguably the case in electricity, where transmission is arguably a natural monopoly but generation is plausibly competitive. Nonetheless, as documented in Michaels (2003), vertical disintegration in electricity has not led to obvious improvements in welfare, and may indeed have impaired efficiency.

liquidity-based market power of a trade execution venue, but incurs costs from low powered incentives, weak incentives to reduce costs, or entry restrictions, or some combination thereof; again, this arrangement is preferable to integration only if these costs are lower than the transactions costs (arising from low power incentives, for instance) incurred by the integrated firm.[18]

3.4 Scope Economies

There are strong scope economies in clearing and settlement. There are also sources of scope economies in trade execution. These scope economies influence the efficient organization of financial trading.

The analysis of section 3.2 can be interpreted to demonstrate that a combination of multiple clearinghouses, each clearing a distinct set of products, reduces the capital necessary to generate the same level of customer loss to default. There are other sources of scope economies in clearing. For instance, clearinghouses universally use collateralization–"margin"–to bond contract performance. I now demonstrate that the total amount of margin required to achieve a given risk of default for a particular customer is smaller if multiple products are cleared together, with gains and losses being net across all of that customer's positions. That is, netting reduces the deadweight costs of collateral, and netting opportunities are greater, when multiple products are cleared together.

First note that collateral is costly. Firms must post margin in cash or other liquid instruments, and typically they must hold more of these lower-yielding instruments than they would in the absence of a collateral requirement; this lower yield is an opportunity cost of collateral. Call c the

[18]Competitive implications of integration are discussed in more detail in section 3.6.

cost per unit of margin. Consider a customer who trades two products. The customer's cash flow on product $i = 1, 2$ is the random variable x_i, where as in section 3.2 $x_i > 0$ ($x_i < 0$) means that the customer makes (receives) a payment to (from) the clearinghouse. The customer must post margin m_i on product i. It may be the case that $x_i - m_i > 0$. That is, the price on the product may move so much that the customer's margin is exhausted because the payment he owes exceeds his collateral. In this case, the clearinghouse must secure this margin shortfall from the customer. The customer may be bankrupt, however, in which case the clearinghouse must draw on its own capital to make whole those who have made money on contract i, or must spend resources to obtain the additional funds from the customer. I assume that covering a margin shortfall is costly. The cost of covering shortfall S is $f(S)$, where I assume only that $f' > 0$.[19]

Optimal choice of margin minimizes the sum of margin opportunity costs plus the expected cost of shortfall. If products 1 and 2 are cleared separately, total margin shortfall is:

$$S_S = \max[x_1 - m_1, 0] + \max[x_2 - m_2, 0]$$

and total expected cost of margin is:

$$C_S = c(m_1 + m_2) + Ef(\max[x_1 - m_1, 0] + \max[x_2 - m_2, 0])$$

With a single clearer for the two products, total margin shortfall is:

$$S_I = \max[x_1 + x_2 - m_1 - m_2, 0]$$

This expression reflects the ability to net gains and losses across the two products. The single clearer's total expected cost of margin is:

$$C_I = c(m_1 + m_2) + Ef(\max[x_1 + x_2 - m_1 - m_2, 0])$$

[19]To account for aversion to shortfall risk, one may also assume $f'' > 0$.

The appendix demonstrates that S_S exhibits first order stochastic dominance over S_I. This in turn implies that

$$Ef(\max[x_1 - m_1, 0] + \max[x_2 - m_2, 0]) > Ef(\max[x_1 + x_2 - m_1 - m_2, 0])$$

Thus, for a given choice of margin (and hence the opportunity costs of margin), expected margin shortfall costs are smaller with a single clearer. Alternatively, the single clearer can achieve the same expected shortfall costs with smaller m_1+m_2, and hence smaller opportunity cost, than can multiple clearers. Thus, the ability to net across multiple positions reduces the costly collateral required to achieve a given level of margin shortfall costs. This creates a scope economy.

There are other sources of scope economies in clearing and settlement. These include:

- Increasing the precision of information about risk. A clearinghouse can only observe the positions in contracts it clears. Thus, if a customer trades some contracts cleared by A and others cleared by B, A will be ignorant of the customer's positions at B, and *vice versa.* For instance, if A clears product 1 and B clears product 2, A's payoff is $\min[x_1, \tilde{k}x_2+V]$, where V is the customer's equity, and $\tilde{k}$ is the number of units of product 2 the customer trades. Note that A does not know $\tilde{k}$ if it clears only product 1. In contrast, if A clears both products, it knows k, and receives a payoff $\min[x_1 + kx_2, V]$. The reduction in information that results from clearing a subset of a customer's positions imposes additional risks on the clearinghouse. A similar analysis holds for clearinghouse B. Risk averse clearinghouses (and costly capital makes a clearinghouse risk averse) will demand compensation for this additional risk. Consolidating clearing of multiple products into a

single clearinghouse improves information, reduces this risk–and hence reduces the associated compensations.

- Economizing on the need for cash. Participants in the clearing and settlement system typically have some trades on which they owe money, and other trades on which they are owed money. Netting cash flows permits paying the net amount, reducing the amount of cash required to meet obligations. This reduces the need for costly short term credit and liquidity. Moreover, exceptional liquidity demands can create systemic risks in the banking and payment system, so reducing the need for liquidity concomittantly reduces settlement risks. Netting economies are greater, the greater the set of products netted.

  This economy can have systematic implications. For instance, in the 1987 stock market crash, some firms that had large losses on index futures positions had large mark-to-market gains on index options positions. These products were cleared seperately, however, and hence the gains on one position were not netted against the losses on the other. The inability to net across positions increased these firms' need for cash to meet margin calls precisely at a time when the liquidity of the banking system was strained (Tamarkin, 1993).

- System costs. The software and hardware necessary to effect settlement of one security is readily utilized to settle trades in other securities. Similarly, clearing systems (notably software) can be utilized to clear a large number of instruments. Although hardware capacity must increase with the number of instruments cleared and settled, thereby increasing costs, system design and software costs are largely invariant to the number of instruments cleared and settled.

It should be noted that these clearing and settlement scope economies extend to products that are traded on over-the-counter markets as well as exchange traded products. For instance, government bonds, corporate bonds, and many derivatives are traded in OTC markets rather than on exchanges. Moreover, these products are often intermediated by the same firms that provide brokerage and account clearing services on exchange traded products. Clearing and settlement systems that service both exchange and OTC markets can exploit the scope economies just identified.

There are also sources of scope economies in trade execution. Some of these economies are more pronounced in computerized exchanges than traditional floor-based markets, while others are more important in open outcry markets.

In floor based markets, especially for derivatives, there are some economies to executing transactions in different instruments on the same trading floor. For instance, spreading is an important trading activity in many markets. As an example, in energy markets, the purchase of crude oil futures and the simultaneous sale of gasoline or heating oil futures–a "crack spread"–is a common trade. As another example, futures options traders often hedge options positions with an offsetting futures position–a "delta hedge." These transactions can be executed more rapidly and accurately when both "legs" of the transaction are traded on the same floor. This creates an economy of scope. As an illustration that this source of scope economy is more important in open outcry environments, spread trading (especially during contract rolls) and options hedges now accounts for most floor volume in financial futures contracts.

In electronic markets, a single computer trading system can handle transactions in many instruments. Indeed, once the trading software is created,

the cost of adding additional trading instruments is relatively small. Moreover, trading multiple instruments on a single platform reduces the costs that customers who trade them incur to connect to the market. Typically, users must create a customized interface for each exchange that they deal with. Trading multiple instruments on a single platform, rather than several, reduces the number of interfaces that customers must create and maintain. Moreover, analogous to floor-based exchanges, it is typically easier to execute spread trades on a single system, than across systems.

3.5 Scope Economies and the Organization of Trading

If trade execution is subject to scope economies, or at least is not subject to scope diseconomies, integration of a multi-product exchange with a multi-product clearer and multi-product settling agent avoids multi-marginalization and *ex post* opportunism problems. Thus, absent diseconomies of scope in trading, the theory predicts the formation (through merger or organic growth) of multi-product exchanges with integrated clearing and settlement, rather than the existence of distinct execution, clearing, and settlement entities. This arrangement mitigates transactions costs.

Matters become more complex if there are diseconomies of scope in trading, or if there is a constraint on merging exchanges that execute trades in different products.[20] In this case, there is a trade-off. An integrated multi-product exchange can realize scope economies in clearing and settlement, but incurs costs due to scope diseconomies in execution. Whether integra-

[20]Pirrong (1999) shows that the mutual structures of traditional open outcry exchanges impede consolidation even in the presence of economies of scope. Merger increases the competition that members face, and as a result they may decline to merge even if merger yields scope economies. Moreover, differences in law and regulation, or legal impediments, often increase the costs of merging exchanges located in different countries, and can preclude such mergers altogether.

tion remains the (second best) efficient solution in this instance depends on the relative costs of alternative arrangements.

One alternative is for the several exchanges to own jointly an entity that clears and settles their transactions. This permits the realization of scope economies in clearing and settlement, without incurring scope diseconomies in trade execution, and avoiding double marginalization and opportunism problems.

If it is efficient for intermediaries to trade across a variety of exchanges, or across exchange and OTC markets, another alternative is the formation of an intermediary-owned cooperative that provides clearing and settlement services for products traded on multiple execution venues. As noted above, this solution incurs some costs that integration does not. These costs can be mitigated by giving trade execution venues an ownership and control stake in the clearing and settlement entity. The cooperative solution (perhaps with exchange participation in governance) is more likely to be observed when scope economies in post-trade services are more extensive than the scope economies in execution. This, in turn, is more likely when financial firms efficiently supply intermediation services across a variety of centralized and OTC markets.

3.6 Vertical Foreclosure

The foregoing analysis focuses on the efficiency enhancing aspects of integration of execution, clearing, and settlement. Concerns have been raised, however, that integration is inefficient because it can impede competition. Specifically, the potential for an integrated exchange to foreclose entry into execution by denying access to clearing lies at the heart of criticisms of ex-

change mergers.[21] In theory, by denying access to clearing ("tying" clearing and execution) an incumbent integrated exchange can raise the costs that another exchange incurs to compete in providing execution services.

The incentive to engage in strategic tying, and the efficiency effects thereof, depend on the nature of competition in execution. Putatively competitive (marginal cost) pricing at one link of the value chain (such as, execution) just permits a monopoly at one of the other links (such as, clearing) to capture the entire monopoly rent. Indeed, the natural monopolist of any one service would prefer competition in the markets for the other services. If the others are subject to scale diseconomies, for instance, vertical integration and foreclosure/exclusion reduces the profitability of the monopoly service. Consequently, foreclosure is self-defeating if one of the complementary functions is highly competitive (Posner, 1976; Bork, 1978). Thus, if execution is potentially highly competitive, and clearing is a natural monopoly (as some exchange integration critics argue) tying cannot have anti-competitive effects, and is instead motivated by reductions in transactions costs.

There are models in which inefficient monopolizing foreclosure can occur.[22] These models are not particularly applicable to the financial trading

[21]Larry Harris, Breaking the Futures Monopoly, Forbes Magazine, November 6, 2006. Meyer S. Frucher, Bearish on Chicago, Wall Street Journal, November 20, 2006. Both argue that the Chicago Mercantile Exchange's control of clearing impede the entry of competitive trade execution platforms, and both advocate the disintegration of clearing and execution. Similar arguments have been made in Europe, particularly in criticism of the integrated Deutsche Börse. See the sources cited note 1 *supra*.

[22]Whinston (1990), Riordan (1998), Salop and Scheffman (1983). The Riordan model applies to a market in which there is a dominant firm and competitive fringe. Cream skimming competition has some features of a competitive fringe, although there are differences. Moreover, in Riordan the upstream industry is competitive and not characterized by extensive economies of scale which is not the case in clearing. Whinston (who analyzes tying specifically, but who argues that a vertically integrated firm that refuses to deal with a competitor is equivalent to tying) shows that anticompetitive vertical restrictions is not

context because they typically do not assume network economies (or other pervasive scale and scope economies) at multiple levels of the marketing chain.

Carlton and Waldman's (2002) model of foreclosure in a network industry comes closest to capturing some–but not all–of the salient features of financial markets. In particular, it includes network economies in one of the complementary goods in the model economy (though not in both.) Due to this feature, competition in this good is imperfect, which can distort the incentives to enter; due to the natural monopoly arising from network effects and lock-in, the entrant earns a rent that may exceed the social value of entry.

The Appendix presents a model that alters Carlton-Waldman to make it more reflective of conditions in financial exchange markets. In particular, whereas Carlton-Waldman model competition in the market for a durable good subject to network economies, trading services are not durable. Therefore, the model invokes different sources of the customer lock in that is crucial to the Carlton-Waldman results. Moreover, the model focuses on execution-only entry because most critics of integration argue that clearing is a natural monopoly, and that the most efficient form of competition is from execution-only venues; in contrast, Carlton-Waldman assume that the good that the complementary good (clearing, in this interpretation) is produced subject to constant returns to scale, and is therefore potentially perfectly compet-

profitable in the standard fixed proportions case even when there are increasing returns to scale. These conditions are particularly applicable in the execution-clearing-settlement situation. Hart and Tirole (1990) shows that integration that reduces output can occur when a firm cannot commit to limit output because price information is non-public. In most of these models, a firm must be able to commit to a tie. Whinston, for instance, argues that a firm can commit to a tie through product design. Such physical ties are not feasible in the financial market context.

itive. Furthermore, Carlton-Waldman impose an exogenous restriction on integrated entry that is not plausible in an exchange context, so the model also explores the economics of entry by integrated firms.[23]

The model retains salient features of Carlton-Waldman, however. Specifically, the execution market is subject to network economies, and tips to the most efficient supplier of execution services. This firm earns an economic rent that arises from the natural monopoly aspect of trade execution. If a firm enters the execution market only, and the incumbent sells clearing services (that is, it does not tie clearing and execution), the two firms split a market power rent. If the incumbent ties, it can foreclose execution-only entry, but may face entry by an integrated firm. Moreover, the entrant incurs fixed costs; these can be viewed as representing the costs of creating execution or clearing systems, or of coordinating the defection of the incumbent's customers.

The main results of the analysis are:

- An integrated incumbent sometimes has an incentive to tie clearing and execution to deter execution-only entry.
- Although for some parameterizations tying deters efficient entry, sometimes foreclosing entry is efficient because entry is dissipative. Because of imperfect competition in execution, the entrant earns a rent. As a result, the private returns to entry may exceed the social value of entry.
- The entrant may decide to enter both clearing and execution markets, and tie provision of these services, even when this is inefficient. The ability to capture the rent that the monopoly clearer would retain in

[23]As an example, two of the (failed) entrants into US futures markets (Eurex and Euronext.LIFFE)

the event of execution-only entry can make entry into both clearing and execution profitable even when it is inefficient.

In sum, the effects of tying are ambiguous even when one ignores the transaction costs efficiencies of integration. The rent captured by the entrant into execution means that entry can be overcompensatory. Consequently, even if foreclosure occurs, one cannot determine *a priori* that this is inefficient because the network effect-induced imperfection of competition in execution distorts entry incentives.[24]

Thus, although foreclosure through the tie-in of clearing and execution of an incumbent exchange is theoretically possible, it is not possible to determine *a priori* that this foreclosure is inefficient; indeed, foreclosure can enhance efficiency, and the conditions under which this result obtains plausibly hold in financial trading markets. If trade execution is potentially highly competitive, foreclosure is not a profitable strategy; the clearing monopolist can extract all the rent by pricing clearing services appropriately, and actually has an incentive to encourage entry by a more efficient execution venue. Conversely, if competition in execution is imperfect due to network effects (as argued above), foreclosure may be efficient because entry is dissipative. Put differently, the natural monopoly aspects in trade execution that arise from network effects undermine traditional arguments opposing foreclosure

[24]If the execution market is perfectly competitive because it is not subject to network economies, in Carlton-Waldman the clearing monopolist has no incentive to tie clearing and execution services if the entrant can enter only the execution business. This result is similar to the classic Director-Posner-Bork theories discussed above. The incumbent may tie if the entrant can supply clearing in the second period. Critics of integration of clearing and execution argue that but for the tying of these services, execution is potentially highly competitive, but that clearing is a natural monopoly service that is prohibitively expensive to enter. That is, they discount the importance of network effects in execution and assert that execution-only exchanges are the most likely entrants. If these conjectures are correct, however, *per* Carlton-Waldman integration and tying are benign, and the incumbent has no incentive to use them to foreclose execution-only entry.

through integration, or alternatively suggest that integration is not adopted as part of a foreclosure strategy, but instead has an efficiency rationale.

Moreover, it should be noted that the conditions in the model in which foreclosure are profitable are implausible in the financial trading context. In particular, they involve a peculiar asymmetry which (but for the tie) favors an execution-only entrant over an established incumbent's execution services. In this model, as in Carlton-Waldman, tying is profitable because a successful execution-only entrant induces the market to tip to its services, and subsequent to tipping has market power due to customer lock-in. That is, customers are intially *not* locked into the incumbent's execution services, but become locked into the entrant's; the incumbent ties in order to lock customers into its execution services instead of the entrant's. However, inasmuch as incumbent exchanges have been in business for decades, and in some cases, for more than a century, it is a stretch to argue that if lock-in is a possibility that their customers are not locked in already. And if they are locked in, the incumbent exchange has a competitive advantage even absent a clearing tie; in this case, an execution-only incumbent faces little threat of entry, and a clearing tie is often superfluous to protect the execution monopoly. If customers are not locked into incumbent exchanges with long histories, why would they become locked into an entrant, permitting that firm to earn a market power rent?

In brief, network effects in both execution and post-trade services preclude definitive determination of the competitive and efficiency implications of integration and tying. The network effects in execution that tend to make this function a natural monopoly distort entry incentives. As a result, foreclosure is not *per se* inefficient. Moreover, the conditions in which inefficient foreclosure can occur are not readily evident in the trading of financial in-

struments. If execution is potentially highly competitive because customers are not locked into a particular execution venue, foreclosure is typically not a profit maximizing strategy for a clearing monopolist. Conversely, if lock in is an important consideration that protects execution venues from entry, clearing ties are superfluous to deter entry into execution.

4 Empirical Evidence

Vertical integration has long been, and remains, the rule in financial markets. Table 1 details the organization of execution and clearing for a set of equity and derivatives exchanges that account for virtually all trading in these instruments around the world: an appendix available from the author presents a comprehensive description of the ownership of clearing and settlement arrangements around the world. This table shows that most equity and derivatives exchanges own and operate their clearing and/or settlement entities, as well as provide execution services. There are several instances in which separate entities supply execution and post-execution services, but in virtually all of these cases the execution venue either has an ownership stake and governance role in the post-execution service provider, or there is considerable overlap in the membership of the two entities. Of late, user cooperatives are assuming a more important role in post-execution services, especially where economies of scope are important. The ubiquity of integration, and the shift away from integration when scope economies are important, is consistent with the hypothesis that efficiency considerations drive the organization of financial exchanges.

Thus, vertical integration of trade execution, settlement, and clearing in a single firm–an exchange–is the modal form of organization in centralized securities and derivatives markets. In most cases, the clearing and settlement

operation is a division or wholly owned subsidiary of the exchange where transactions are executed. In most of the remaining instances, the execution venue has an ownership stake or governance role, or both, in the clearing and settlement entities. The exception that proves the rule is the London Stock Exchange, and even this entity operated its own settlement division until mismanagement of technology induced regulators and users to set up a separate venture.

Pervasive economies of scope in clearing and settlement are leading to diminished exchange roles in some clearing and settlement entities, such as DTCC and NSCC in the United States, and LCH.Clearnet in Europe. The scope economies in clearing and settlement extend across multiple exchanges, and also across centralized exchange and decentralized OTC markets. In particular, the consolidation in banking and intermediation, whereby large intermediaries (such as, Goldman Sachs or Citigroup or HSBC) participate in myriad exchange and OTC markets, has increased these scope economies in clearing and settlement. This provides a strong incentive to consolidate clearing and settlement across exchanges and OTC markets. This has raised the opportunity cost of vertical integration and exchange control over clearing and settlement, relative to the alternative form of organization, clearing and settlement cooperatives owned and operated by users of clearing and settlement services. The decline in exchange ownership and control over clearing and settlement entities that span exchange and OTC markets is consistent with this change in relative costs.

In American futures markets, the move to electronic trading has enhanced economies of scope in trade execution. The two largest futures exchanges integrated their clearing functions (by contract) in 2003, and agreed to merge in 2006. The merger was completed in 2007. The merged entity

plans to clear through the CME-owned clearinghouse. In this instance, economies of scope in clearing and trade execution allow economizing on transactions costs in clearing through vertical integration, without sacrificing scope economies in execution.

Thus, consistent with the theory outlined in section 3.3, which states that exchange ownership of execution, clearing, and settlement operations economizes on transactions costs, such integration is the primary means of organizing these functions except where scope economies in clearing and settlement encompass markets where scope economies in execution are absent (such as across exchange and OTC markets.)

5 Summary and Conclusions

The completion of a securities or derivatives transaction involves several complementary activites, each characterized by considerable scale economies arising from fixed costs and network effects. Indeed, each of the three basic activities–execution, clearing, and settlement–exhibit strong natural monopoly tendencies.

This poses challenges to the organization of financial trading. Supply of these functions by separate firms can give rise to multi-marginalization problems and opportunistic holdups. Integration of these functions into a single firm–an exchange–can economize on these costs.

Indeed, integration is the default mode of organization of securities and derivatives trading. The vast majority of exchanges operate their own clearing and settlement operations, or have a significant ownership and control stake in them. Recently, growing disparities in scope economies between clearing and trade execution have put strains on this traditional mode of organization, contributing to calls to de-integrate these activities. Where

this de-integration has occurred, clearing and settlement are usually supplied by user-owned and governed cooperatives, most of which operate (*de facto* or *de jure*) as non-profits that rebate fees to members proportional to their volume.

Although integration has also given rise to assertions of foreclosure, the pervasive scale economies in each activity make it unlikely that disintegration alone is sufficient to improve substantially the competitiveness of centralized trading of financial instruments. Moreover, disintegration threatens to result in inefficiencies due to multiple markups and opportunism. Thus, transaction cost considerations, rather than market power concerns, should be the primary focus of any analysis of financial exchange organization. Any attempt to improve competition in trading of securities and derivatives must address simultaneously the impediments to competition in execution, clearing, and settlement; disintegration alone is an inadequate remedy for market power in financial trading.

Table 1 Panel A The Organization of Clearing Equity Exchanges		
Exchange	Integration	Cooperative
NYSE		O (pre-2007)
AMEX		O (pre-2007)
NASDAQ		O (pre-2007)
Euronext	X (pre-2000)	O (pre-2007)
Deutsche Börse	X	
LSE		N
Scandanavian Markets	X	
SWX	X (post-2007)	O (pre-2007)
TSE	X	
OSE	X	
HKSE	X	
SET	X	
ASE	X	

In this table, an "X" in the Integration column indicates that the executing exchange owns its clearinghouse. An "O" in the Cooperative column means that the executing exchange obtains clearing services from a cooperative firm in which it has an ownership stake. An "N" in this column indicates that the executing exchange obtains clearing services from a cooperative in which it does not have an ownership stake. NYSE, AMEX, NASDAQ and Euronext announced plans to divest their ownership in their clearers in 2006. SWX announced its merger with its clearer, SIS, in 2007.

Table 1 Panel B The Organization of Clearing Derivatives Exchanges		
Exchange	Integration	Cooperative
CBOT		O (pre-2003)
CME	X	
ICE	X	N (pre-2007)
MGE	X	
NYMEX	X	
NYBOT	X	
US Options		O
Canadian Derivatives		O
AEX-Optibeurs	X	
BELFOX	X	
MATIF	X	
MONEP	X	
Eurex	X	
MIF-MTO IDEM	X	
OM	X	
SOFFEX	X	
LIFFE		O
LME		O
IPE		O
TSE	X	
TIFFE	X	
SCE	X	
SGX	X	
HKFE	X	
TFE	X	
SFE	X	

A Cartellization Through a For-Profit Clearing Cooperative

Consider a clearing cooperative that charges a price P_C per unit to its members for clearing services, and then distributes net surplus to the broker-

owners in fixed shares. Broker i receives share $\alpha_i < 1$ of surplus, with $\sum_{i=1}^{N} \alpha_i = 1$. All other notation is identical to that in the formal model in section 3.3.

In contrast to that model, here the brokerage firms are Cournot competitors. Broker i chooses q_i taking q_j, $j \neq i$ as given, to maximize:

$$q_i D_B(\sum_{j=1}^{N} q_j) - C_i(q_i) - P_C q_i + \alpha_i[\sum_{j=1}^{N} q_j(P_C - c_c) - F]$$

The term $\sum_{j=1}^{N} q_j(P_C - c_c) - F$ is the total profit of the cooperative.

The first order conditions are:

$$D_B(\sum_{j=1}^{N} q_j) + q_i D'_B(\sum_{j=1}^{N} q_j) = C'_i(q_i) + (1 - \alpha_i)P_c + \alpha_i c_c$$

Note that output choice depends on P_C and α_i, because they influence each firm's perceived marginal cost (on the right-hand-side of the expression), whereas in the non-profit cooperative P_C does not affect output choices. Thus, the for-profit cooperative can choose the clearing fee and the ownership shares to affect these choices.

Call $\{q_i^*\}_{i=1}^{N}$ the set of outputs that maximizes the brokers' aggregate profit–the "perfect cartel" output vector. The clearing cooperative can choose P_C and the α_i to solve:

$$D_B(\sum_{j=1}^{N} q_j^*) + q_i^* D'_B(\sum_{j=1}^{N} q_j^*) - C'_i(q_i^*) = (1 - \alpha_i)P_c + \alpha_i c_c$$

There are $N + 1$ equations (the N first order conditions and the constraint that the shares add to 1) and $N + 1$ unknowns (P_C and the shares). For a well-behaved demand function it is possible to solve these equations for a clearing price and profit division shares so that the Cournot competitors choose the perfect cartel outputs. If the execution exchange has market power, this results in extreme double marginalization.

Thus, although the surplus of the cooperative is rebated to broker-members in both the for-profit and non-profit cooperatives, the method of the rebate matters. Paying surplus proportional to fixed shares rather than according to each individual firm's output changes the nature of the competitive interaction between the brokers.

It should be noted that if brokers are Cournot competitors but the cooperative pays broker i a patronage dividend equal to $q_i(P_C - c_c)$, double marginalization occurs, but the broker cooperative's output exceeds the perfect cartel output for $N > 1$. The Cournot competitors price their services above marginal cost, and hence include a markup on c_c, that the integrated exchange does not. There is still a double markup because P_B exceeds $C_i'(q_i)$ in this case. Therefore, if the brokers are Cournot competitors, vertical integration can improve efficiency by mitigating (but not eliminating) double marginalization.

Thus, supplier ownership of clearing can either mitigate or exacerbate double marginalization problems, depending on the cooperative's payout policy and the competitiveness of the brokerage sector. A for-profit cooperative clearer can induce severe inefficiencies.

B Stochastic Dominance Proof

Consider first the case where $x_1 - m_1 > 0$. In this case:

$$S_I = x_1 - m_1 + \max[x_2 - m_2, m_1 - x_1]$$

and

$$S_S = x_1 - m_1 + \max[x_2 - m_2, 0]$$

Since $m_1 - x_1 < 0$ in this case, $\max[x_2 - m_2, 0] \geq \max[x_2 - m_2, m_1 - x_1]$, with a strict inequality for some values of x_2. Thus, for any shortfall level

$\bar{S} \geq 0$, there are values of x_2 such that $\bar{S} \geq S_I$ but $S_S > \bar{S}$, but no values of x_2 for which the reverse is true.

Now consider the case $x_1 - m_1 < 0$. In this case

$$S_I = \max[x_1 - m_1 + x_2 - m_2, 0] \leq \max[x_2 - m_2, 0] = S_S$$

with strict inequality for some values of x_2. Again, this implies that there are values of x_2 such that for any shortfall level $\bar{S}$, $S_I \leq \bar{S}$ but $S_S > \bar{S}$, but no values for which the converse is true.

Thus, $\Pr[S_I \leq \bar{S}] > \Pr[S_S \leq \bar{S}]\ \forall \bar{S}$. *QED.*

C Tying and the Efficiency of Entry

Here I modify Carlton-Waldman to derive a model of the incentives of an integrated incumbent exchange to engage in vertical foreclosure. I focus initially on an execution-only entrant, inasmuch as critics of vertical integration commonly assert that clearing is a natural monopoly, and that competition in execution would be enhanced by opening access to the incumbent's clearing facility; in contrast, Carlton-Waldman examine a market in which the monopolized complementary good is not a natural monopoly, but would be perfectly competitive but for some exogenous constraint. Moreover, the exogenous constraint on simultaneous entry into clearing and execution like that assumed in Carlton-Waldman is inapplicable in financial trading, so after discussing execution-only entry I turn my attention to the integrated entry case.

Some modifications to Carlton-Waldman are required to fit salient features of financial markets. Carlton-Waldman analyze the case of a durable good subject to network economies. In Carlton-Waldman's two period model, the good that consumers purchase in the first period is usable in

the second, and contributes to the total size of the network in the second period. This durability "locks in" first period purchasers into their initial network choice in the second period, and thereby affects second period competitive interactions.

This is not relevant in a financial trading context. Trading services are not a durable good. Therefore, in the present model, which also has two periods, the same consumers purchase in each period. I invoke a different source of lock in than do Carlton-Waldman.

There are two firms, an integrated incumbent that supplies both trading and clearing, and an execution-only entrant. The surplus a consumer receives by purchasing execution services from the entrant is $v(N_E) + \Delta$, where N_E is the total number of customers who use the entrant. Moreover, $v'(N_E) > 0$, due to network economies that derive from the nature of liquidity. Similarly, surplus from purchasing from the incumbent is $v(N_I)$, where N_I is the total number of customers who use the incumbent, with $v'(N_I) > 0$. The parameter $\Delta \geq 0$ is a measure of the superiority of the entrant's execution services. If $\Delta > 0$, the entrant's services are superior, and given the size of the network, consumers are willing to pay more for the entrant's execution services than the incumbent's. There is a total of N customers.

The marginal cost of execution services is zero. The marginal cost of clearing services is also zero.

The entrant incurs a cost of E_C to enter the market. This cost incorporates the expense of building a trading system, and any costs that the entrant incurs to coordinate the defection of customers from the incumbent. If the entrant incurs this cost, consumers can choose freely where to trade in the first period. If the entrant does not incur this cost, the customers are

locked into the incumbent exchange.

If the entrant succeeds in inducing the market to "tip" to it during the first period, the incumbent can attempt to regain control of order flow in the second period by incurring a cost of I_C. If the incumbent incurs this cost, customers can choose to trade on the incumbent's execution system even if they defected to the entrant in the first period. If customers defected in the first period, and the incumbent does not pay this cost, customers are locked into the entrant in the second period. Thus, E_C can be viewed as the cost of overcoming initial customer lock-in, and I_C can be interpreted as the incumbent's cost of overcoming lock-in if the market initially tips to the entrant.

The timing of the model is as follows. First, the incumbent exchange decides whether or not to tie clearing and execution. If the incumbent ties, the entrant cannot offer execution services because clearing and execution must be consumed in fixed proportions. Second, the entrant decides whether to enter. Third, if the incumbent has not tied, the entrant and incumbent choose their prices for execution services, and the incumbent chooses a price for clearing services. Fourth, customers choose which exchange to trade on in the first period. Fifth, if customers have chosen to patronize the incumbent in the first period, the incumbent chooses whether to incur the I_C in the second period. The incumbent and the entrant can also bargain at this time over surplus to be divided between them. Sixth, if the execution market has tipped to the entrant, the incumbent does not pay I_C, and the entrant and the incumbent have not reached a bargain over surplus, the entrant and the incumbent split the surplus between them evenly (as would occur under Nash bargaining.)

Consider first the division of surplus if the market has tipped to the

entrant. (I will show momentarily that the market tips to one supplier or the other, so tipping outcomes are the only equilibrium-relevant ones.) Here total surplus is $N\Delta + Nv(N)$, and the entrant and the incumbent (who continues to supply clearing services) split this amount; there is a bilateral monopoly here, and the monopolists of the two segments split the total available rent.

If consumers choose to patronize the incumbent's execution services in the first period, the incumbent captures the total surplus of $Nv(N)$ in the second period.

Now consider the bargaining between the entrant and the incumbent assuming the execution market has tipped to the entrant in the first period. Following Carlton-Waldman, I assume that the entrant and incumbent evenly split the surplus attributable to (a) the entrant's superior product and (b) the entrant's now lower costs (lower because the incumbent must incur a cost to re-enter execution). By incurring cost I_C and coordinating the defection of the customers from the entrant, surplus is $Nv(N) - I_C$. When customers obtain execution services from the entrant, total surplus is $N\Delta + Nv(N)$. The difference in surplus is $I_C + N\Delta$. As a result of the bargaining, the entrant receives $.5N\Delta + .5I_C$, and the incumbent receives the balance of the surplus $.5N\Delta + Nv(N) - .5I_C$.

The basic idea here is that the incumbent's ability to compete for execution order flow in the second period constrains the entrant's pricing power even if the execution market has tipped to the entrant in the first period. The costlier it is for the incumbent to re-enter the execution market (due to the stickiness of order flow, parameterized by I_C), the greater the entrant's bargaining power and the greater its second period rent if the market tips its way in the first period.

Given the bilateral monopoly in the second period that results if the market tips, consumers receive no surplus in the second period. Without loss of generality, if the market does not tip, consumers who choose the entrant in the first period receive surplus of $s_E(N_E)$ if a total of N_E consumers choose the entrant. Similarly, those who choose the incumbent in the first period receive a surplus of $s_I(N_I)$ if N_I choose the incumbent in the first period. Consumers are atomistic (as in Carlton-Waldman), so when making his decision each consumer takes N_E and N_I as given.

Now consider consumers' choices on where to trade in the first period. When the entrant chooses an execution price of P_E, the incumbent chooses an execution price of P_I, and the incumbent's clearing price is P_C, an (atomistic) consumer chooses to patronize the entrant if:

$$\Delta + v(N_E) - P_C - P_E + s_E(N_E) > v(N_I) - P_C - P_I + s_I(N_I).$$

If the reverse is true, the atomistic consumer chooses to patronize the incumbent. Note that if this expression holds for one customer, it holds for all, so the market tips to the entrant in this case; in the reverse situation, the market tips to the incumbent.

As in Carlton-Waldman, and as is customary in the network industry literature, I assume that consumer choices are coordinated to maximize their surplus. Therefore, if

$$\Delta + v(N) - P_C - P_E > v(N) - P_C - P_I$$

the market tips to the entrant. Otherwise, it tips to the incumbent.

Now consider the pricing of services when the incumbent does not tie. In order to preclude "virtual" ties, I constrain $P_I \geq 0$.[25] Without loss of

[25] As in Carlton-Waldman, due to network economies the clearing monopolist can some-

generality, I set $P_I = 0$. Therefore, to get any execution business, under the coordination assumption, $P_E \leq \Delta$ as otherwise the entrant would get no business. Moreover, the pricing equilibrium is not unique. If the entrant chooses a price of P_E, the incumbent's best response is to choose $P_C = \Delta + v(N) - P_E$. For the purpose of determining entry decisions, I assume that the incumbent and the entrant believe that any outcome $P_E \in [0, \Delta]$ is equally likely. That is, the entrant expects to sell at a price of $.5\Delta$ in the first period.

Therefore, in equilibrium, in the absence of tying, if entry occurs, the market tips to the entrant in the first period, the entrant monopolizes execution, and the incumbent monopolizes clearing.

The entrant's profit is $\Pi_E = .5N\Delta + .5N\Delta + .5I_C$. The entrant captures half of the surplus associated with its superior product in each period. It also captures surplus equal to half of the incumbent's re-entry cost. The incumbent's profit is $\Pi_I = N\Delta + 2Nv(N) - .5I_C$. By tying, the incumbent earns $2Nv(N)$.

Note that the incumbent faces conflicting incentives with regard to tying. On the one hand, the ability to capture some of the surplus associated with the entrant's superior product provides an incentive to permit entry. This is the basic "Chicago School" argument that efficient supply of the complementary good raises the monopolist's derived demand. On the other hand, the fact that the entrant obtains some market power if the market tips to it, and hence can extract some market power rents in the second period that the incumbent would otherwise capture, provides an incentive for the

---

times profitably deter entry into execution by selling execution services below cost, and recouping the lost revenue by raising the price of clearing services. CW refer to this as a virtual tie.

incumbent to deter entry through tying. This is the basic Carlton-Waldman motive for tying.

If $2N\Delta \geq E_C$, entry is efficient. Three cases are relevant here:

- $I_C > 2N\Delta > E_C$. Here, tying makes the incumbent better off, so it ties. This deters entry that would occur but for the tie, so an inefficient outcome results.

- $2N\Delta > I_C > E_C$. Tying makes the incumbent worse off. It will not tie, and the efficient outcome results.

- $2N\Delta > E_C > I_C$. Tying again makes the incumbent worse off. However, entry is sometimes unprofitable in this case. Therefore, the incumbent does not tie, but entry may not occur nonetheless.[26]

If $2N\Delta < E_C$ entry is inefficient. Again, three cases are relevant:

- $I_C > E_C > 2N\Delta$. The incumbent ties, but this is efficient. Market power makes entry inefficiently profitable. The entrant can capture a rent due to the lack of competition in execution in the second period due to tipping and the cost of overcoming the stickiness of order flow. Therefore, in this case tying prevents dissipatory entry.

- $E_C > I_C > 2N\Delta$. Entry is unprofitable even absent a tie.

- $E_C > 2N\Delta > I_C$. Entry is unprofitable even absent a tie.

Therefore, this simple model suggests that tying may not occur *for strategic purposes,* and if it does, its efficiency effects are ambiguous. Tying may

[26]There is an incentive for the entrant with the superior technology and the clearing monopolist to merge in this instance.

be profitable and preclude efficient entry, but tying may be profitable and deter inefficient, dissipative entry. Due to the imperfection in competition in execution in the second period attributable to customer lock-in, the entrant secures a market power rent. This reduces the profitability of clearing, and if this reduction is large enough provides the incumbent an incentive to tie. However, this market power rent may also overcompensate the entrant. That is, entry may occur for rent seeking reasons even when it is inefficient.

Note that tying for strategic reasons is often superfluous. If entry costs are sufficiently high, that is, if E_C is sufficiently high relative to $N\Delta$ and I_C, the entrant cannot recoup its costs even if the incumbent does not tie.

Further note that inefficient tying and dissipative entry both require $I_C > E_C$. That is, the incumbent's cost to overcome the order flow advantage of a successful entrant in the second period is greater than the entrant's cost of overcoming the incumbent's order flow advantage in the first period. This condition seems implausible. It would seem much more difficult for an entrant to overcome the order flow advantage of an established incumbent exchange, than for an established incumbent to re-capture business from a successful entrant. At the very least, it would seem that the situation is symmetrical.

Thus, a theoretical model that incorporates salient features of the trading of financial instruments, notably the importance of network effects and customer lock-in, provides little support for a blanket condemnation of tying. Network economies and lock-in make competition imperfect. In such a second-best world, entry is not always efficient, and hence tying that prevents entry is sometimes efficient. Moreover, the most plausible outcome is that tying is strategically superfluous; if an entrant incurs a greater cost to wrest order flow from an incumbent than the incumbent would incur to

wrest it back, tying cannot be inefficient, and the incumbent may be immune from entry into execution even if it does not tie.

It should be noted that these basic results are similar to those in Carlton-Waldman. Due to the market power rent that an entrant can earn, entry can be inefficient in their model as well.

Finally, this model ignores transaction cost and double marginalization inefficiencies that result from non-integration of clearing and execution. To the extent that an entry bears only a fraction of these costs, entry incentives are further distorted, and the efficiency benefits of tying are enhanced.

The foregoing analysis assumes that integrated entry is not feasible. In fact, some exchanges (e.g., Eurex US) have offered bundled execution and clearing services during their attempts to enter a market. It is straightforward to show that (a) tying cannot affect an integrated entrant, and (b) due to the presence of market power rents, such entry may be dissipative. Such entry is best modeled as a "War of Attrition" (Tirole, 1988).

D Empirical Evidence on Exchange Organization

D.1 US Equity Markets

The first stock clearing and settlement system was initiated by the Philadelphia Stock Exchange in 1870 (Shultz, 1946). In 1892 the New York Stock Exchange implemented a system for clearing and settlement of stock trades. Initially, the exchange clearinghouse only cleared and settled stock balances; money balances were not netted or cleared (Schabacker, 1930). Netting and clearing of money balances commenced in 1920. Prior to 1918 the clearinghouse was a department of the NYSE. In 1918, the exchange organized a new clearing corporation, that was a wholly-owned subsidiary of the NYSE Shultz (1946).

Other US equity exchanges, including the American Stock Exchange and the various regional exchanges, operated their own clearing branches. In the 1960s, the Wall Street paperwork crisis led the exchanges to explore initiatives to facilitate clearing and settlement. The first was the creation in 1973 of the Depository Trust Company that immobilized stock certificates and introduced book entry delivery and settlement. The second was the establishment in 1976 of the National Securities Clearing Corporation, through the merger of the stock clearing affiliates of the NYSE, the American Stock Exchange, and the National Association of Securities Dealers. The NSCC was, and is, a CCP. Since its introduction, the regional exchanges have eliminated their stock clearing operations and joined the NSCC.[27] The NSCC was originally owned by the exchanges.[28] In subsequent years, "participant firms," mainly clearing brokers, have acquired ownership stakes. By 2006, the NYSE, NASD, and AMEX owned 36.4 percent of the DTCC. In 2006, exchanges' common shares were redistributed to participant members. The exchanges continue to hold preferred shares in DTCC, and have the right to name a director to the board.[29]

This change reflects the dramatic evolution in the DTCC in the years since its founding. Originally established to clear and settle stock transactions, in the past 30 years it has added services to clear government securities, mortgage backed securities, and some derivatives trades, none of which are traded on organized exchanges. This expansion allowed the DTCC to

[27] www.dtcc.com/AboutUs/history.htm.

[28] *Id.*

[29] Steve Letzler, DTCC Completes Change in Share Ownership, www.dtcc.com/Publications/dtcc/apro06/share.htm.

exploit scope economies across these various instruments, as participants in these markets largely overlap. This has reduced the benefits of exchange ownership and control, and exchange ownership has indeed declined accordingly.

The DTCC provides clearing services at cost. Specifically, it prices its services to generate only the revenues necessary to "liquidate current production costs, provide for a continuance of product enhancement and development, provide for a discount when volume levels are equal to or greater than projections and provide for retained earnings as directed by the Board" (Depository Trust and Clearing Corporation, 2006). The NSCC rebates fees to clearing firms when revenues are in excess of projected costs.It is therefore quite similar to the cooperative modeled in section 3.3, and the exchanges have a role in its organization and governance.

D.2 US Futures Markets

- *Chicago Board of Trade.* Between 1883 and 1925, The Chicago Board of Trade operated a department to net margin payments. This entity did not serve as a central counterparty, and hence had no liability in the event of a default. Instead, it merely calculated members' net margin obligations, and facilitated the transfer of margin payments between members. In 1925, the exchange formed a separate corporation, the Board of Trade Clearing Corporation. This firm did serve as a CCP.

  BOTCC was (and remains) a Delaware for-profit corporation.[30] Originally, exchange rules required each of the CBT's members to purchase shares in the clearinghouse, and allowed only CBT members to utilize

[30]http://www.clearingcorp.com/bylaws/certinc.html. It is now called the Clearing Corporation, following the severing of its relation with the CBT.

BOTCC's clearing services (Kroszner, 2006). That is, each CBT member cleared his own trades through BOTCC and only through BOTCC. Over time, however, it became evident that there were advantages for brokerage firms to clear for multiple members as well as for public customers, so the requirement that each CBT member own a stake in BOTCC was eliminated. Those CBT members who did not remain as BOTCC owners used BOTCC members to guarantee their contracts. However, CBT members had the right to join BOTCC (subject to the clearer's minimum capital requirements.) BOTCC established separate capital requirements for sole proprieterships, partnerships, and corporations, thereby facilitating membership for all types of CBT member firms. Until 2003, CBT rules required exchange members to clear through BOTCC exclusively.

As a for-profit corporation, BOTCC was able to pay dividends. However, it has never done so.[31] Instead, any surpluses were retained as capital to be used to cover the costs of any defaults. Based on trading volumes and open interest, BOTCC's board of directors determined how much capital was required to meet the company's anticipated default obligations. To the extent that clearing revenues net of expenses exceeded the desired increase in capital, BOTCC cut the clearing fees it charged its member-users. Since the cost of the capital required to meet defaults is a major expense of a CCP, BOTCC effectively, though not formally, acted as a cooperative that charged prices which

[31]The following description of BOTCC and its operations was obtained from an interview with Kevin R. McClear, Vice President, General Counsel and Secretary of the Clearing Corporation.

just covered costs similar to that modelled formally *supra.*[32] Moreover, BOTCC adjusted the ownership shares of member-brokers with volume; the shares of firms losing (gaining) volume relative to total volume were reduced (increased).

Although the CBT and BOTCC were separate companies, whose ownership did not overlap completely after the BOTCC's early days, several factors served to align the incentives of the trade execution venue (the CBT) and the clearer (BOTCC), thereby preventing opportunism and double marginalization. Specifically, (a) CBT members–including individual "local" traders who dominated CBT governance–had the ability to become BOTCC members, (b) until 2000, voting on BOTCC was one member-one vote, rather than one share-one vote, and (c) all members of BOTCC were required to own CBT seats. Thus, although BOTCC rules specified a relation between a firm's share ownership and the volume of trades it cleared, with large firms having larger share ownership than small firms, each member had the same voting share regardless of ownership share.

These provisions limited the ability of large brokerage firms to take actions (e.g., restricting entry, or engaging in opportunism) that benefitted BOTCC's large corporate broker-owners and adversely affected other CBT members. The individual proprieterships and partnerships that dominated control of the CBT could become (and became)

[32]BOTCC was somewhat similar to mutual insurance companies, as analyzed by Hansmann (1996). Hansmann notes the possibility that insurance mutuals have a tendency to accumulate excessive capital due to the inability of members to control managers. That problem was presumably less severe at BOTCC since it had a relatively small number of members (147 for an extended period), each of which actively interacted with BOTCC on an ongoing basis, and who had a strong incentive to avoid excessive capitalization.

BOTCC members, and exercise control disproportionate to their ownership due to the one member-one vote rule. The adjustment of ownership shares to reflect volume shares also tied a firm's share of BOTCC surplus to its own volume, rather than giving each member a fixed share. This mechanism made BOTCC more like the efficient non-profit cooperative described in the main text than the potentially cartelizing clearing entity modeled in the appendix.

At the initiative of the large corporate brokerages, BOTCC voting rules were changed from one member-one vote to one share-one vote in 2000. Relations between BOTCC and the CBT became strained soon thereafter. In 2003, BOTCC entered into an agreement to clear trades for the Swiss-German exchange Eurex, which was entering the US market in direct competition with the CBT. In response, the CBT rescinded its rule requiring its members to clear through BOTCC, and required them instead to clear through the Chicago Mercantile Exchange clearing division under terms of a contract between the two exchanges. In 2006 the CME and CBT agreed to merge.

The CBT-BOTCC relationship represents an interesting governance structure. The companies were formally separate, and for many years did not have a direct contractual relationship. However, the ability of CBT members to become BOTCC owners, and the disproportionate voting power of the individual CBT members who dominated the governance of the exchange, damped the conflicts between these entities. However, the consolidation of the brokerage industry exacerbated the disparity between the ownership shares and voting shares of BOTCC members. The large brokerage firms forced a change in the voting rules that reduced the power of non-broker CBT members. The relation-

ship between the two firms ended acrimoniously soon thereafter, and the CBT entered into a formal contract with the CME, and eventually agreed to merge with that exchange.

- *Chicago Mercantile Exchange.* The Chicago Mercantile Exchange established a clearing division in 1919 (Tamarkin, 1993). The exchange continues to maintain and operate a Clearing House.The exchange is required to utilize surplus funds to satisfy Clearing House obligations in the event that the Clearing House's own resources are inadequate.

- *Minneapolis Grain Exchange.* The clearinghouse is a department of the Minneapolis Grain Exchange (Rules and Regulations of the Minneapolis Grain Exchange).

- *Kansas City Board of Trade.* The Kansas City Board of Trade owns all of the outstanding shares of the Kansas City Board of Trade Clearing Corporation (Kansas City Board of Trade Clearing Corporation Charter and By-Laws) Membership in the clearinghouse is open to any KCBT member in good standing.

- *New York Mercantile Exchange.* NYMEX operates a clearing department (New York Mercantile Exchange Bylaws) The Clearing House operates under direction of the exchange Clearing House Committee, and the NYMEX president is manager of the clearinghouse (New York Mercantile Exchange Rules) Membership in the Clearing House is open to any member of the exchange.

- *New York Board of Trade.* The Commodity Clearing Corporation is the CCP for the NYBOT; it is a wholly owned subsidiary of the exchange. Membership in the Clearing Corporation is open to any

member of the NYBOT.

D.3 US Options Markets

The first US options exchange, the Chicago Board Options Exchange, formed its own clearinghouse when the exchange was launched in 1973. In 1975, two additional exchanges began trading options. At this time, these exchanges collaborated to form the Options Clearing Corporation. The exchanges owned OCC in equal shares. In subsequent years, additional exchanges have commenced options trading, and each has become an owner of the OCC. Thus, the OCC is a completely exchange-owned entity. Moreover, OCC prices its services at a level sufficient to meet expenses. If the OCC Board deems that revenues exceed its funding needs, it rebates fees to clearing members.

D.4 Canadian Derivatives Markets

The Canadian exchanges that trade derivatives, the Montreal Exchange, the Toronto Stock Exchange, the Vancouver Stock Exchange, and the Toronto Futures Exchange, jointly own the Canadian Derivatives Clearing Corporation, which clears futures and options trades for these exchanges (Bank for International Settlements, 1996).

D.5 European Equity Markets

- *Euronext.* Euronext was formed by the merger of the Dutch, French, and Belgian stock exchanges. Prior to the merger, each of these exchanges operated a clearing subsidiary (European Competition Commission DG ("ECCDG" hereafter), 2004). After the merger in 2000, the exchanges established Clearnet SA as their CCP. Euronext owned Clearnet. In 2003, Clearnet and LCH (the London Clearinghouse)

agreed to merge. LCH cleared for London futures exchanges, and was owned by the London Metal Exchange, the London International Financial Futures and Options Exchange, the International Petroleum Exchange, and clearing brokers, with the exchanges having a minority stake (LCH.Clearnet Annual Report, 2005) After the Clearnet-LCH merger, the merged firm LCH.Clearnet was owned by users and exchanges, with Euronext holding a 41.5 percent stake (with voting limited to 24.9 percent), LME 2.7 percent, IPE .9 percent, market users 45.1 percent, and Euroclear (a clearing bank) holding the remaining 9.8 percent (ECCDG).

LCH operated along the lines of the cooperative model analyzed formally in section 3.3 (ECCDG). However, LCH.Clearnet operates as a for-profit entity (ECCDG). The analysis of section 3.3 suggests that this could pose multi-marginalization problems, although the ownership stake and voting influence of the exchanges would tend to dampen the incentive to engage in this conduct.

- *Deutsche Börse.* Deutsche Börse owns 50 percent of Eurex Clearing, AG, which clears all DB stock trades. In 2002, DB purchased Clearstream, which settles all DB equity transactions (ECCDG).

- *London Stock Exchange.* The LSE operated a Settlement Department beginning some time prior to World War II, and re-established this department in 1947 after a war-induced hiatus (Michie, 1999). The exchange continued to operate its own settlement department until 1993, when its failure to implement the automated Taurus settlement project caused the Bank of England to intervene and foster the establishment of a separate settlement entity, CRESTCo(Michie, 1999; Eu-

roclear, Euroclear Submission in the Deutsche Borse/Euronext/LSE Matter, 2005). CRESTCo was owned and governed by market users, and returned excess earnings to users (Euroclear, 2005).

LSE adopted a CCP for its SETS trading system in 2001. SETS trades were cleared through LCH, and settled at CRESTCo. LSE trades are now cleared by LCH.Clearnet.

LSE is unique in that it has no ownership stake or governance role in the firms that clear or settle its trades. LSE is also in the process of encouraging competition in clearing. Beginning in 2007, exchange users will be able to clear through LCH.Clearnet or SIS x-clear (London Stock Exchange, 2006). Although the pervasive scale and scope economies make the survival of two competing clearinghouses problematic, LSE's desire to encourage competition in clearing is understandable as a way of reducing its vulnerability to holdup and multiple marginalization.

- *Scandanavian Markets.* HEX Integrated Markets Oy operates exchanges in Sweden, Finland, Latvia, and Estonia. All of these exchanges clear and settle through APK, a HEX subsidiary (ECCDG).

- *SWX.* Heretofore the Swiss Stock Exchange SWX has not had a central counterparty service. Instead, the exchange guaranteed member trades (Swiss Stock Exchange, 2006a). The exchange is in the process of launching a CCP service. Like LSE, the SWX will offer users the choice of either LCH.Clearnet or SIS. In this new arrangement, all SWX participant firms must become a member of either LCH.Clearnet or SISClear (Swiss Stock Exchange, 2006b).

  SIS operates as a non-profit that rebates fees to members (SIS Group

Annual Report, 2005). It is owned by Swiss banks, with UBS AG and Credit Suisse Group holding a majority stake.[33] SWX is owned by 55 banks, each of which have an equal vote in the exchange.[34] The requirement that SWX members join SIS (or LCH.Clearnet), the considerable overlap of ownership between SWX and SIS, and the non-profit organization of SIS make the impending Swiss arrangement quite similar to that of the Chicago Board of Trade and BOTCC at the time of the latter's creation. Common ownership of SWX and SIS mitigates conflicts between the entities, and the non-profit form of SIS addresses double markup problems. SIS also provides clearing services in instruments other than Swiss stocks, and hence exploits scope economies.

In May, 2007, SWX and SIS announced their intention to merge. Thus, when the merger is consummated (in late-2007, if the merger proceeds according to plan), the combined entity will be vertically integrated into clearing and execution, although SWX users will still have the option to utilize LCH.Clearnet clearing if they choose. The stated reasons for the merger are illuminating, and consistent with the transactions cost theory presented above. The entities state that the integrated firm can respond more efficiently to rapid technological change, changing customer needs, and regulatory shocks.[35] Recall from section 3.3 that theory suggests that integration facilitates coordination of responses

[33]http://www.group.sisclear.com/sis/index/sisgroup/sisgroup-shareregister.htm.

[34]Bourse Turns its Back on Merger Mania, swissinfo, November 24, 2006.

[35]Merger of SWX, SIS and Telekurs, Media Conference of 15 May 2007. http://www.swxgroup.com/dowload/media_releases/2007/media20070515b_presentation_en.pdf.

to technological and regulatory shocks.

D.6 European Derivative Markets

- *AEX-Optibeurs.* The European Options Clearing Corporation, a wholly owned subsidiary of Amsterdam Exchanges, NV, the operator of AEX-Optibeurs, provides clearing services for this exchange (BIS).

- *Belgian Futures and Options Exchange.* The Clearing is performed by a department of the exchange (BIS).

- *Marché à Terme Internationale de France.* While MATIF was independent, in 1990 it acquired the clearinghouse Banque Centrale de Compensation SA. This subsidiary cleared all MATIF trades (LCH.Clearnet Annual Report, 2006.

- *Marché des Options Negociables de Paris.* This exchange was owned by Societé des Bourses Franąis. The SBF's clearing department cleared MONEP deals (BIS).

- *Detusche Terminbörse and Eurex.* Eurex Clearing, a subsidiary of Eurex, clears contracts on this market (ECCDG).

- *MIF-MTO and IDEM.* Contracts on these Italian exchanges are cleared by Cassa di Compensazione e garunzia, a separate corporation. Cassa is owned by 21 clearing members, all of whom are exchange members. Moreover, all exchange members that trade derivatives must be Cassa members (BIS)¿

- *OM.* The exchange and clearing facility are both owned by OM Gruppen AB (BIS).

- *SOFFEX.* The Clearinghouse is a department of the exchange (BIS).

- *LIFFE, LME, IPE.* LCH (and latterly, LCH.Clearnet) clears for these exchanges.[36] Prior to the LCH-Clearnet merger, the three exchanges owned collectively 25 percent of LCH (BIS). The exchanges have a smaller ownership stake in the merged clearing entity.

- *ICE.* The Intercontinental Exchange ("ICE") purchased the IPE in 2005, and initially continued obtaining clearing through LCH. In 2007, however, ICE announced its intention to create its own clearinghouse in London.

D.7 Asian and Australian Equity Markets

- *Tokyo Stock Exchange.* The TSE clears derivatives transactions through the Japan Securities Clearing Corporation. Before 1997, it did not serve as a CCP (BIS). JSCC is now a full-fledged CCP. The TSE owns 86.3 percent of JSCC, and five other Japanese equity and derivatives exchanges own the balance.[37]

- *Osaka Securities Exchange.* The OSE clears through JSCC, of which it owns 9.5 percent.[38]

- *Singapore Exchange.* The clearinghouse is a department of the exchange.[39]

---

[36]IPE is now a subsidiary of the IntercontinentalExchange, and goes by the name of ICEFutures.

[37]http://www.jscc.co.jp/english/about/index.html.

[38]http://www.ose.or.jp/e/stocks/index.html.

[39]http://info.sgx.com/SGXWeb_CORPCOM.nsf/NEWDOCNAME/Background_On_SGX.

- *Hong Kong Exchanges.* Hong Kong Exchanges and Clearing, LTD, operates both trade execution and clearing functions.[40]

- *Stock Exchange of Thailand.* The Thailand Securities Depository Company, Ltd., provides clearing and settlement services to SET. It is a wholly owned subsidiary of the exchange.[41]

- *Australian Securities Exchange.* Formed by the merger of seven regional stock markets in 1997, the ASX originally operated a default guarantee fund (rather than a CCP) through the Securities and Exchanges Guarantee Corporation, which managed the National Guarantee Fund. The SECG was a wholly owned subsidiary of the of the ASX.[42] In 2002, ASX restructured its clearing and settlement operations and created a CCP called the Australian Clearing House.[43] It also formed an entity called the Australian Settlement and Transfer Corporation Pty Ltd to settle its transactions. The ACH and ASTC are subsidiaries of ASX.[44]

D.8 Asian and Australian Derivative Markets

- *Tokyo Stock Exchange.* The TSE clears derivatives transactions through the Japan Securities Clearing Corporation. As noted above, the TSE

[40] http://www.hkex.com.hk/infra/dcass/dcass%20e.pdf.

[41] http://www.set.or.th/en/operation/clearing/clearing_p1.html.

[42] http://www.segc.com.au/pdf/ngf_information_booklet.pdf.

[43] http://www.asx.com.au/about/pdf/ach_221102.pdf.

[44] http://www.asic.gov.au/asic/pdflib.nsf/LookupByFileName/ASX_assessment_report_feb_2006.pdf/$file/ASX_assessment_report_feb_2006.pdf.

owns 86.3 percent of JSCC.

- *Tokyo International Financial Futures Exchange.* A department of TIFFE clears its trades (BIS).

- *Singapore Commodity Exchange.* The clearinghouse is a department of the exchange.[45]

- *Singapore Exchange.* The clearinghouse is a department of the exchange.[46]

- *Hong Kong Exchanges.* Hong Kong Exchanges and Clearing, LTD, operates both trade execution and clearing functions.[47]

- *Thailand Futures Exchange.* The TFEX is a wholly-owned subsidiary of Stock Exchange of Thailand, and clears through another wholly-owned subsidiary.

- *Sydney Futures Exchange.* Sydney Futures Exchange Clearing Corporation Pty Ltd, a subsidiary of the Sydney Futures Exchange, provides clearing services for the SFE.[48]

---

[45] http://www.sicom.com.sg/index_sub.asp?content=aboutus

[46] http://info.sgx.com/SGXWeb_CORPCOM.nsf/NEWDOCNAME/Background_On_SGX.

[47] http://www.hkex.com.hk/infra/dcass/dcass%20e.pdf.

[48] http://www.sfe.com.au/index.html?content/clearing/intro.htm.