Clearing — a centralized mechanism for confirming and guaranteeing trades in securities and derivatives markets — is often likened to the plumbing of the financial system: it is ignored when it works, and can cause serious problems when it doesn’t. Market insiders have always understood the importance of this arcane aspect of the financial system, however, and recently it has received far more attention on both sides of the Atlantic. In particular, in both the United States and Europe, fierce debates are raging over how clearing should be organized and who should own the clearing function. For the most part, the battle pits exchanges on one hand and large banks and regulators on the other, but even then some exchanges and some regulators take different views than most of their peers.

In a nutshell, the debate centers on whether exchanges should own the clearing entity or whether exchange ownership of the clearing function impedes competition for the execution of financial transactions. Opponents of vertical integration in clearing assert that clearing is effectively a natural monopoly function and that exchange control over this function makes it effectively impossible for new exchanges to compete against incumbents. Proponents of the “vertical silo” model, including the world’s largest derivatives exchanges, Deutsche Börse and the Chicago Mercantile Exchange, respond that exchange ownership of clearing is not an insuperable obstacle to competition, and what is more, is an efficient way to organize financial transactions. Opponents of vertical integration prefer a user- (read “large bank and brokerage”) owned nonprofit clearing cooperative, which they argue will achieve efficiencies in clearing, mitigate market power in clearing, and also encourage competition in the execution of financial transactions. Most exchanges believe that those benefits are chimerical and that such an institution will merely transfer power to large banks, to the detriment of the broader financial system.

This article goes back to basics in an attempt to sort out the competing claims. Based on an analysis of the fundamental economic features of the businesses of executing and clearing trades in financial instruments — the most notable feature being the presence of strong scale and network economies — I conclude that the competitive benefits of “opening up” or “dis-integrating” the clearing function are illusory and that integration can economize on transactions costs. Thus, the economic case for the forced disintegration of the clearing and execution of financial transactions is very weak. Indeed, in my view, such a measure would actually be counterproductive, leading to greater costs than benefits.

The Economics of Executing and Clearing

Completion of a financial transaction, such as the purchase of a stock or the sale of a futures contract, involves several complementary steps. Consider a transaction in a standardized futures contract, such as a Treasury bond contract traded on the Chicago Mercantile Exchange. The buyer and seller must first interact to establish a price at which they are willing to trade. The process of negotiating the price and quantity terms of the transaction is the process of “executing” a trade. Historically, and still today to some extent, exe-
cation took place in “open outcry” auctions on the floor of an exchange. During an open outcry auction, buyers and sellers (or their broker-agents) shouted out the prices at which they wanted to transact in the trading pit or ring of an exchange, reaching agreements via the face-to-face bargaining process. Today, most futures, options, and stock transactions are executed via computers. Buyers submit bids and sellers submit offers to an exchange computer. If the highest price any buyer submits equals or exceeds the lowest price offered by any seller, the exchange computer matches the high bid and low offer, resulting in the execution of a trade.

Once a buyer’s and seller’s orders are matched, the putative deal is submitted for clearing. The “clearinghouse” first checks to make sure that all the details of the trade submitted by the buyer and seller match. Once the match is confirmed, the deal is cleared.

In many markets, including exchange-traded futures markets and many equity markets, the clearinghouse serves as a “central counterparty” in which it becomes the buyer to the seller and the seller to the buyer. That is, the clearinghouse is inserted between the original buyer and seller, assumes performance obligations to each, and looks to each for performance. In very simplified terms, once a deal is cleared, the original buyer and seller are no longer in a contractual relationship with one another; the buyer has a contract to buy from the clearinghouse and the seller has a contract to sell to the clearinghouse. The clearinghouse, in turn, has a contract to sell to the buyer and a contract to buy from the seller. This mechanism standardizes credit and performance risk.

Financial contracts are always vulnerable to default. Consider a firm that has sold November soybean futures in the spring; this firm is typically referred to as a “short.” During the summer, a drought develops, ravaging the soybean crop. The soybean price rises dramatically, imposing losses on the seller. The losses may become so large that the short does not have the financial resources to cover them. In that event, the short defaults and does not perform on his contractual obligation. Because, in dollar terms, the futures market is a zero sum game, the failure of the short to perform means that buyers of November futures (the “longs”) cannot receive all that they are owed. The possibility of default undermines the ability of futures markets to serve their function of facilitating hedging and speculation.

The clearing mechanism insures market users against default. If the seller defaults on his contractual obligation, the buyer looks to the clearinghouse for performance. In the soybean example, the clearinghouse dips into its capital to pay the longs what they are owed, just as an insurance company dips into its capital to compensate a policyholder whose home burns to the ground. That is, the clearinghouse takes on the risk that the parties to a contract will not perform. As a result, individual buyers and sellers are rather indifferent to the creditworthiness and performance risk of those with whom they deal, because they look to the clearinghouse for performance on their contracts. In other words, the clearinghouse effectively guarantees contract performance and standardizes the risk of
The clearinghouse can be viewed as an insurance mechanism whereby credit and performance risks are shared. It sometimes said that clearing “mutualizes” credit and performance risk.

**Economies of Scale and Scope**

It is crucial to understand that both the execution and clearing of financial transactions are subject to strong economies of scale. Moreover, clearing is subject to strong scope economies; scope economies in execution are less clear cut, though it is likely that those economies are stronger in modern, electronic financial markets. Indeed, the scale and scope economies are so strong that each function is arguably a natural monopoly. This has important implications for the efficient organization of financial markets.

First, consider scale economies in execution. It is well known that there is a network effect in the trading of financial instruments; liquidity, it is said, attracts liquidity — people like to trade where a lot of people are already trading. This centripetal tendency in securities and derivatives trading reflects the impact of informed trading. Some market participants have better information than others, and better-informed traders systematically make money at the expense of poorly informed traders. Some less informed traders want to buy and sell nonetheless, to hedge their price risks for instance, but want to minimize their losses at the hands of the better informed. One way they can do this is to “bid” — to trade on the same market. That is, the losses of the uninformed to the informed are greater if they split their trading activity across two markets. Thus, uninformed trading activity tends to “tip” to a single market. Because the informed make money at the expense of the uninformed, they follow the uninformed to this market, and in equilibrium all trading occurs on a single exchange. This tipping process means that larger markets have lower trading costs, as measured by bid-ask spreads and price impact costs, than smaller ones (all else equal).

Of course, the uninformed would prefer not to trade with the informed at all. If some of the uninformed can reliably demonstrate that their trading is not driven by a desire to exploit an information advantage, they may choose to trade at a venue that attempts to exclude informed traders. This process fragments trading and is sometimes pejoratively referred to as “cream skimming.” Recent years have seen the proliferation of such cream-skimming trading venues.

Although cream skimming–driven fragmentation means that multiple trading venues can survive, the “tipping” phenomenon means that price discovery is a natural monopoly. That is, all informed traders and those traders who cannot demonstrate that they are uninformed will congregate on a single exchange. It is very difficult for entrants to compete with an incumbent exchange because, to overcome the incumbent’s trading cost advantage, it is necessary for an entrant to attract simultaneously the bulk of the incumbent’s customers. It is a daunting task to coordinate the simultaneous defection of numerous customers, and successful examples of this are few and far between.

Next consider scale economies in clearing. These scale economies arise from the effects of diversification. In essence, the clearinghouse has a portfolio of risks that it insures. The losses, gains, and capitalizations of a clearinghouse’s customers are not perfectly correlated. As a consequence, the variability of the clearinghouse’s average exposure (the variance of its total exposure divided by the number of risks insured) declines as the number of insured risks increases. That is, the law of large numbers helps reduce the riskiness of the clearinghouse’s portfolio. In essence, increasing the number of risks insured increases the diversification of the clearinghouse’s portfolio, and the well-known diversification effect means that the riskiness of this portfolio (properly scaled) declines with size.

This has implications for the cost of supplying clearing services. The clearer must hold costly capital to perform its insurance function because, as noted earlier, it must dip into its capital to cover the losses of defaulting customers. The higher its capital, the more costly it is to hold. Thus, insuring more risks and diversifying more effectively, the clearer can provide the same level of insurance protection with a smaller amount of capital per customer, the larger the number of positions insured.

Similar factors create scope economies in clearing. Recall that clearings houses collect margins that serve as performance bonds. The clearinghouse can seize a defaulting customer’s margins to cover at least part of the loss. Consider a customer trading two different futures contracts, say oil and Treasury bonds. The payoffs to those contracts are not perfectly correlated, so the variance of the loss across both positions is smaller than the sum of the variances of losses on each of the two positions. That occurs because losses on one position are sometimes offset by gains on the other. If the oil and T-bond contracts are cleared by separate entities, each clearer will assess a margin based on the riskiness of each individual position. In contrast, if the contracts are cleared by a single entity, it can achieve the same exposure to customer default while charging a margin based on the (lower) riskiness of the combined position. Thus, clearing multiple contracts through a single entity reduces the margins required to achieve a given level of customer performance. Because margins are costly (as customers have to hold lower-returning liq-
and derivatives markets. In particular, the strong scale economies and scope economies in trading and execution would create difficulties if clearing and execution were provided by distinct for-profit firms. This is true for two basic reasons.

First, because those activities are highly complementary, the scale and scope economies inevitably result in substantial concentration—and quite plausibly a monopoly—in the provision of each service. Hence, separate provision by for-profit entities results in double monopoly markups. The reason for this is as follows. Traders care about the total cost of executing a transaction. This cost includes the sum of the price of clearing services and the price of execution services. Because those services are complementary, the amount that a customer is willing to pay for execution declines if the price of clearing goes up; that is, the demand for each service is decreasing in the price of the other service. Because of this externality, if separate for-profit entities provide those services, profit maximization results in prices that add up to an amount that exceeds the monopoly price of the combined service. This externality arises because, when setting its price, each entity ignores the effects of its pricing decisions on the demand for the other’s services. Together, the firms charge prices that are higher than the combined price that they would choose to maximize their joint profit. That, in turn, means that separate provision leads to higher prices, smaller output, and lower welfare than provision by an integrated monopoly.

Second, the separate clearing and execution venues would necessarily coexist in a long-term relationship. Even if they entered into a contract designed to mitigate the double markup problem, each party would have an incentive to opportunistically subsequent to the signing of the contract in order to extract rents from the other. Haggling, post-contractual “holdups,” and other forms of rent-seeking impose transactions costs on the parties. The potential for such opportunistic behavior is most acute when the parties need to respond to technological and regulatory shocks that require a coordinated response in order to achieve an efficient outcome.

Vertical integration—the combination of both clearing and execution activities in a single firm—eliminates double markups and is a well-known means to economize on transaction costs. Thus, fundamental economic considerations cause integration to dominate the provision of trading and post-trading services by separate for-profit entities. One should not be surprised to observe such integration in prac-
A clearing cooperative has some disadvantages relative to integration, however. First, even if supply of clearing by a cooperative mitigates double-markup problems, the provision of clearing and execution by separate entities still creates transactions costs that integration avoids. There is the potential for haggling and opportunism between the execution venue and the clearing cooperative. Separate provision is particularly susceptible to impeding the coordinated response to technological and regulatory shocks.

Second, the clearing cooperative can still be used to exercise market power, in which case double-markup problems reappear. Even if it formally is a nonprofit, a cooperative can exercise market power by inefficiently restricting membership. Even if its formal rules specify that it is open to all, in reality, because of the sharing of financial risks in a clearinghouse, it is inevitable that member firms have to meet certain financial criteria. It is possible to restrict membership, and thereby raise the price of clearing services, by choosing inefficiently strict financial standards. Moreover, a cooperative can choose its pricing and surplus redistribution policies in a way that facilitates the charging of supracompetitive prices by its members. For instance, I have shown formally that a clearing utility that distributes surplus to members in fixed shares, rather than in proportion to the number of transactions a member clears, can charge the monopoly price for its services and distribute this monopoly profit to its members. In this way, double-markup problems reappear with a vengeance.

That a nonprofit cooperative can exercise market power is no mere theoretical possibility. For decades, until their recent shift to for-profit status, financial exchanges were typically nonprofit and cooperatively owned. These cooperatives could develop market power by restricting entry. Exchange members earned substantial rents as a result of those restrictions. A clearing cooperative could achieve a similar outcome.

Thus, clearing cooperatives have some serious disadvantages as compared to providing clearing through an integrated exchange. The main potential advantage of a separate clearing cooperative is that it could better realize clearing scope economies by clearing for multiple exchanges. Horizontal mergers of vertically integrated exchanges could achieve those scope economies as well, but that outcome can be less efficient than clearing through a single cooperative if there are diseconomies of scope in execution. In that case, the formation of a single clearer servicing multiple optimally scaled execution venues can fully exploit clearing scope economies without incurring diseconomies in execution. The crucial question then becomes whether those greater scope economies are smaller or larger than the double-markup losses and higher transactions costs that separate provision of clearing and execution services entails.

In sum, vertical integration of clearing and execution makes considerable economic sense. It economizes on double markups and transactions costs. It may not be able to achieve clearing scope economies as efficiently as a clearing cooperative servicing many exchanges, but it is very difficult for an outsider — including a regulator — to determine whether the potential benefit in terms of greater scope economies more than offsets the higher transactions costs and greater potential for double markups inherent in separate provision of clearing and execution. Thus, there is a strong case for vertical integration of clearing and execution. This case is firmly grounded in an understanding of the economics of clearing and execution, basic price theory, and transactions costs economics.

DISSENTING VOICES

Not everybody — including some market participants and regulators in the United States and Europe — sees things quite the same way as the argument I sketched above. In the alternative view, vertical integration is not an efficiency-enhancing form of organization, but is instead a nefarious alternative view, vertical integration is not an efficiency-enhancing form of organization, but is instead a nefarious alternative view, vertical integration is not an efficiency-enhancing form of organization, but is instead a nefarious alternative view, vertical integration is not an efficiency-enhancing form of organization, but is instead a nefarious alternative view, vertical integration is not an efficiency-enhancing form of organization, but is instead a nefarious
highly competitive. As noted above, and in a substantial body of academic research, there are strong centripetal forces driving price discovery and trading activity to a single execution venue. Moreover, the empirical evidence supports this view. Furse’s own exchange is perhaps the best illustration. The London Stock Exchange has obtained clearing services from a third party for years, yet it has never faced serious competition for its execution business and several would-be competitors have failed miserably. Furthermore, its profit margins are among the fattest in business—fatter, in fact, than the integrated Deutsche Börse.

The Justice Department advanced a couple of examples from the equity and options markets to bolster its assertion, but they are unpersuasive to say the least. The New York Stock Exchange has long obtained clearing services from a cooperative and the Options Clearing Corporation. Nonetheless, the fact that major exchanges like the Chicago Mercantile Exchange and Deutsche Börse remain integrated, and that others such as the London International Financial Futures Exchange and the InterContinental Exchange are looking to integrate, is inconsistent with assertions that execution is naturally competitive and that vertical integration is anticompetitive. Those who argue that execution is inherently competitive and that integration is an anticompetitive artifice are essentially arguing that exchanges are stupid and do not understand their own economic self-interest. Of course that is always possible, but I would not take any odds on its truth.

The vertical-integration-as-anticompetitive tactic is a throwback to a long discredited antitrust view. It is perhaps understandable that the FT editors and Clara Furse have fallen for it. It is far less forgivable for the Justice Department’s Antitrust Division to succumb to this fallacy and to express their error in unsolicited comments that directly caused a sell-off in exchange stocks, thereby costing investors billions in market capitalization.

Just separating clearing from execution, using a cooperative clearer, is not sufficient to induce substantial competition in execution.
ENHANCED EFFICIENCY

Because it is implausible, on both theoretical and empirical grounds, that exchanges integrate to foreclose competition in execution, the most compelling rationale for integration is that it enhances economic efficiency by reducing frictions (such as double markups and transactions costs) at the interface between the suppliers of highly complementary services subject to strong scale economies. None of the critics of vertical integration have ever raised the possibility that the combination of execution and clearing enhances transactional efficiency, let alone addressed it seriously or refuted it. (The Justice Department’s failure to address fundamental concepts that have dominated antitrust economics and industrial organization for over 30 years is particularly inexcusable.)

If integration is driven by efficiencies — and to me, the case that it does is compelling and heretofore unchallenged — then regulatory intervention to force disintegration is extremely perverse. Indeed, it is likely to reduce efficiency substantially (by raising transactions costs and double markup–related waste) without appreciably increasing competition in the execution of securities and derivatives transactions. If regulators want to improve financial market efficiency in the worst way — well, disintegration would be it.

Put differently, vertical integration is a rational and efficiency-enhancing response to extensive scale economies in both execution and clearing. A focus on one source of scale economy while ignoring or dismissing the other, as the critics of also have done, recommends policies that are doomed to disappoint their advocates and that are guaranteed to increase the costs of transacting in securities and derivatives markets.

If clearing cooperatives offer advantages that outweigh the other potential sources of inefficiencies, market participants are unlikely to leave that money on the table. They do not need regulators or legislators to compel them to pick it up.

CONCLUSION

In sum, vertical integration of the execution and clearing of financial trades is an efficiency-improving response to pervasive scale economies. That said, it is highly unlikely that integrated exchanges achieve a first-best, optimal outcome. Because of scale economies, it is very difficult to compete against incumbent exchanges and, as a result, they possess and exercise market power. The prices of exchange services are, thus, supracompetitive and exchanges earn monopoly rents. The hefty profit margins that exchanges earn, and their outsized stock prices and multiples, provide evidence that exchanges indeed earn such supracompetitive profits.

The trading of securities and derivatives poses the same policy challenges as other network industries. The government’s less than stellar experiences with attempts to restructure electricity and telecom markets via legislation and regulation intended to address competitive issues should give pause to anyone recommending similar measures in financial trading. The United States has made some strides in breaking the liquidity-driven network effect in stock trading (through the implementation of RegNMS), but this took more than 30 years to achieve and the industry is still in transition, making it too early to judge whether the changes will substantially improve competition in the U.S. stock market in the long run. Europe is just embarking on a similar path, and it is premature to judge from its experience. The mere discussion of competitive issues in U.S. futures markets has only just begun, and the kind of measures necessary to undermine the liquidity advantage of incumbent exchanges (such as mandated market linkages or the creation of a central order book utility) have not been mooted. Thus, it appears that debates over financial trading market structure will be with us for a long time.

It is possible, however, to dispatch one policy nostrum right here and now: the notion that eliminating vertical integration of trade execution and clearing is sufficient to improve competition substantially in the trading of securities and derivatives. Such a policy would entail all pain and no gain. Strong scale economies in trading and clearing both contribute to competitive imperfections in financial trading. Indeed, integration is an economizing response to large scale economies in those complementary activities. Disintegration of clearing and execution through regulatory fiat would not substantially improve competition in trading, but would force market participants to address the challenges inherent in scale economies of both activities in a less efficient way. There are no easy fixes to competitive imperfections in financial trading, and disintegration is not even a partial fix.

Serious attempts to undermine exchange market power must involve measures designed to handle the liquidity network effect — through mandated market linkages, for instance — and constrain the ability of clearers to exploit scale economies and charge supracompetitive prices. A dual track is required because scale economies in both activities foster market power, addressing one without confronting the other is clearly inadequate and will almost certainly make matters worse. But regulators would be advised to tread very carefully, as the devil is truly in the details. Poorly conceived and implemented efforts could result in outcomes worse than the admittedly imperfect status quo.

Readings