CONTENTS

1

Introduction

From New York to London, from Chicago to Tokyo, and from Frankfurt to Sydney, capital markets the world over have undergone revolutionary changes during the past two decades. The frenzied activity of traders buying and selling stocks and other financial instruments on the trading floors of the New York Stock Exchange (NYSE), the London Stock Exchange (LSE), and the Chicago Board of Trade—traditional icons of global capitalism—has been replaced by algorithmic trading and supercomputers tucked away in gigantic nondescript "datacenters" in out-of-the-way places such as Mahwah, New Jersey; Aurora, Illinois; and Basildon, outside London. Trading has become extraordinarily complex and opaque, with trading speeds no longer measured in minutes or seconds but in time units beyond human perception: milliseconds (thousandths of a second), microseconds (millionths), and even nanoseconds (billionths). By way of comparison, a millisecond is to a second as one second is to 11.6 days; and a nanosecond is to a second as one second is to 31.7 years. The blinking of the human eye takes about 400 milliseconds, and a nerve impulse reaches the brain in about 80 milliseconds—near eternities compared with the speed of modern communications and trading.

Technological advances have scaled up imperceptible and previously irrelevant time differences into operationally manageable and enormously profitable business opportunities for those with sufficient high-tech trading tools. These tools include the fastest private communication and trading lines, the most powerful computers, and sophisticated algorithms (algos) that are capable of speedily analyzing incoming news and trading data and determining optimal trading strategies in microseconds. High-tech trading also relies on possession of gigantic collections of historical and real-time market data. One Chicago-based market operator is said to possess a collection that contains "the rough equivalent of approximately 100 times the amount of data included in the Library of Congress."[1] The storage, management, organization, and analysis of such big data require enormously costly and complex systems that only a small number of large operators can afford.

But there is another central factor that has contributed to the extraordinary complexity of capital markets: *market fragmentation*. At the dawn of the twenty-first century, the NYSE was the world's preeminent equity market, listing companies from all over the globe. Today, the NYSE is no longer dominant; its overall share of the domestic market dropped from 80 percent to about 24 percent over the past decade.[2] Trading in U.S. equity markets is now split between 12 public (also called "lit") exchanges and many more off-exchange trading venues, including about 40 so-called dark pools (see below) and over 200 internalizing broker-dealers.[3] This fragmentation is a feature not only of equity markets but also of other markets, including options markets and foreign exchange markets. And the trend is global—fragmented capital markets are a growing reality in Europe as well as parts of Asia.

In this hyperfast fragmented global marketplace, algos battle algos for trading dominance (i.e., preferential execution position), and the most sophisticated trading supercomputers deal not only in securities but increasingly across asset classes, including futures, fixed income, currencies, and commodities, and across hundreds of markets and dozens of countries. A retired regulator with a distinguished 15-year record at the helm of two major

financial regulatory organizations recently confessed to me that he no longer understands how these complex capital markets really work. The average investor is even more in the dark about these markets. When an investor sends an order to buy or sell a stock by the click of the mouse, the order may take a lightning journey through a maze of dark pools and exchanges before being filled. How does the investor know that on the journey to execution the order was treated fairly and was filled at the best available price? Adding to market complexity is the extraordinary explosion in order traffic—from millions of orders daily 10 years ago to many billions today.

A comprehensive examination of the functioning of these capital markets of today is opportune and should matter to all of us—for the health of these markets affects our savings and pensions and ultimately has profound implications for the general welfare as well as for equality and justice in society.

Some argue that the recent transformations have introduced, on the whole, greater efficiency through enhanced market competition, resulting in narrower spreads and reduced commissions, for the benefit of investors. Others, however, are sceptical. In a 2014 U.S. survey, a striking 70 percent of financial industry participants said that today's capital markets are not fair to investors; only 18 percent felt they were fair.[4] Many other recent surveys show a persistent majority of buy-side market participants (i.e., asset managers and managers of hedge funds, pension funds, and trusts) expressing negative views on overall market quality.

Telltale signs that all is not well occasionally make the newspaper headlines, including the Flash Crash of May 2010, when the U.S. equity markets dropped 9 percent in value, for no obvious reason, only to fully recover within 30 minutes, or the similarly dramatic and mysterious flash crash of the British pound in October 2016. Less noticed, however, are so-called mini flash crashes—large erratic price swings in individual stocks over milliseconds—which are a daily occurrence in today's fragmented markets. According to one source, about 18,500 mini flash crashes occurred between 2006 and 2010 in U.S. stock exchanges alone.

These unfavorable opinions of the market and worrisome recent events need explanation. Are they linked to the recent changes that have taken place in the markets? Economists traditionally view markets as simple coordination systems that facilitate the efficient exchange of goods and services between buyers and sellers. This view is a helpful starting point for understanding how markets operate, but it is incomplete for understanding why they change and who benefits or loses from changes in market structure. In this book I explain the dramatic recent transformations and events in capital markets and assess how they affect core public policy objectives such as investor protection, as well as market transparency, fairness, and efficiency. I do this by offering a new analytical lens through which to view these events and transformations, based on a reconceptualization of markets.

Markets are more than simple coordination systems or "disembodied" meeting places of demand and supply. They are organizations governed by their own rules and regulation. Moreover, markets are deeply political organizations or governance systems where contending groups of members or stakeholders are frequently embroiled in intense battles to shape market rules and structure according to their own narrow preferences. These contending groups are not necessarily equal in power, and sources of power may quietly change over time, thereby altering bargaining power. In short, power politics must be at the heart of any analysis of markets. Power is central to explaining markets both in the sense that general power politics arguments about who wins or loses apply to market settings, and in the sense that markets themselves are political institutions governed by power relations.

Furthermore, just as with any political system, some markets are well governed and others are poorly governed. A main challenge in capital markets is not primarily the expense and hazards of forging contracts between buyers and sellers, it is opportunistic behavior key market operators—so-called market makers who stand between buyers and sellers (see chapters 3 and 4 and glossary). Their vantage point at the center of the market gives them access to privileged information about order flow, prices, and

market trends that they may be tempted to use for private gain at the expense of their clients. Good market governance seeks to anticipate and preempt such opportunistic behavior. More broadly, good governance is about managing conflicts of interest for the long-term benefit of all in society. It ensures fair, orderly, and efficient markets. Bad governance is about exploiting conflicts of interest for a quick profit, thereby surreptitiously transferring wealth from the weak in society to the powerful. If designed smartly, these exploitative schemes are practically invisible and silent, and can last for many years.

Good governance does not mean that everything is always perfect. Accidents and lapses do happen, and individual bad apples or occasional criminal gangs can cause damage to society. However, good governance is a system where the "governors" and stakeholders have a strong incentive to punish bad apples and criminals, as well as to invest in norms, rules and regulations, and policing and compliance systems in order to deter opportunistic or parasitical behavior by a few and safeguard and protect the interests of the many. For a capital market, reputation can be a powerful incentive for good governance.

Bad governance, by contrast, rewards bad behavior. Deception, lying, obfuscation, and misrepresentation are pervasive in bad governance. The creation of exploitative schemes by particularly powerful actors to benefit themselves is rational in a system of bad governance because the chances of getting caught are tiny and the reputational or material consequences of such behavior are largely insignificant while the profits from such schemes are high.

I argue that markets vary in the quality of their governance. If markets can manifest either good governance or bad governance, the question is what explains the difference? What explains when reputational concerns will trump power and create an incentive for the formation of good governance systems? What explains when and why power can trump reputational concerns?

I show that good capital market governance prevailed, on the whole, during most of the twentieth century (see chapter 3). Over the past decade and a half, by contrast, bad governance has been

on the rise (see chapters 4 and 5). Market makers have fewer obligations, market surveillance is neglected or impossible, enforcement is rendered ineffective, and new technologies are no longer used primarily to improve market governance but to offer lucrative preferential market access to select clients, often in undisclosed or hidden ways. Specifically, although some of the evidence remains partial, I show that information asymmetries and secrecy—often deliberate governance-design strategies—have enabled a small but powerful group of unscrupulous market operators to milk conflicts of interest at the expense of the unsuspecting investing public.

Powerful actors claim to "innovate" to achieve greater market competition for the benefit of all in society. In reality, the modern fragmented markets that they have constructed (see chapter 2) tend to undermine competition. Fragmentation produces many "shallow" pools of liquidity (see glossary)—a proliferation of public exchanges, broker-dealer dark pools, and other private off-exchange trading places—that enable the powerful to more easily extract private rents on the back of hoodwinked investors.

Latent in the minds of many victims of these strategies is a belief that "modern" markets are technologically determined and that technological progress must be good. But new technology is neither bad nor good per se. Its social value is solely determined by the incentives or motives of the users of this technology. I show that in the old system of centralized markets, the dominant exchange had a strong reputational incentive to use technology for the benefit of all investors; in today's fragmented markets, by contrast, costly new technology is often used by powerful market operators in quiet and nearly invisible ways to maximize their profits at the expense of ordinary investors.

This book thus offers a new—and sobering—perspective on why capital markets have fragmented, as well as when and why algorithmic capital markets (i.e., "instantaneous" electronic trading) may fail the public.

A Deeply Puzzling Market Transformation

For over two centuries, securities markets in all major countries tended toward greater concentration. Concentration of trading in one large organized public market or trading "pool" seemed natural and inevitable in the wake of improvements in information and communication technologies. Consider, for example, the case of the United States: in the first half of the nineteenth century, securities trading was largely local, and all large cities, including Boston, Philadelphia, New Orleans, Chicago, and San Francisco, had their own exchanges. The size of these exchanges reflected the size and wealth of the local population; the NYSE was only marginally larger than the Boston or Philadelphia exchange.[5] It derived its slight edge from its location in the most populous U.S. city and in the center of a prosperous region. With its large port, New York was a principal channel of international commerce, and the opening of the Erie Canal in 1825 helped it become a major hub of interregional trade. As a result, the local catchment area of the NYSE comprised the largest number of affluent investors.[6]

With the advent of the ticker-tape machine, enabling the speedy diffusion of NYSE stock prices throughout the country, and continuous improvements in telegraph technology during the second half of the nineteenth century, the catchment area of the NYSE expanded rapidly.[7] Increasingly, major companies in Boston, Baltimore, Philadelphia, and other cities sought listings on the large NYSE, and the bulk of share trading gradually moved to New York.

The reason for the expansion is so-called user-network effects. The greater the number and variety of users of a given exchange market, the more attractive the market is to new or potential users, since new buyers or sellers are more likely to find a counterparty to a transaction in a large market than in a small one. A central market naturally has the highest concentration of orders: it has the greatest trading depth (volume of bids and offers) as well as breadth (range of tradable securities); in other words, it has the highest liquidity. In addition, highly liquid markets both reduce investment risks, by making it easy to quickly enter or exit a trade, and lower the cost

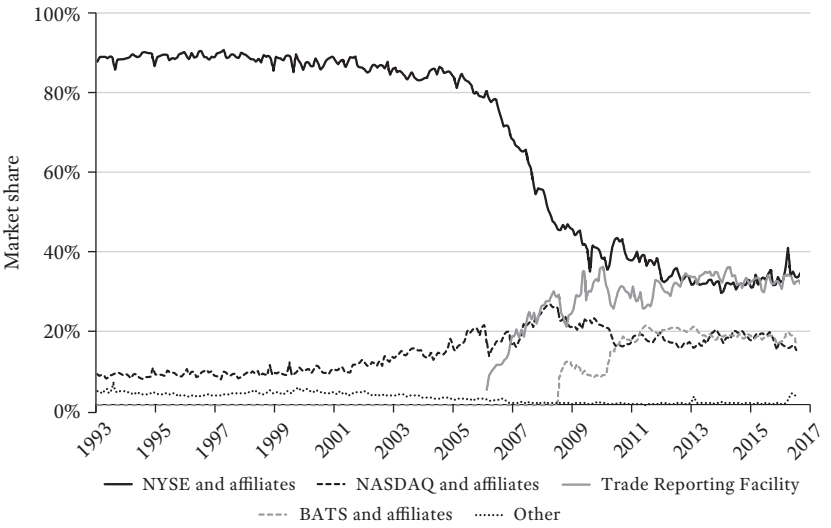

FIGURE 1.1. NYSE-listed equities trading on various markets. *Source:* Mnuchin, Steven, and Phillips, *A Financial System That Creates Economic Opportunities: Capital Markets*, Washington, DC: U.S. Department of the Treasury, 2017, 53.

of trading, since competition among buyers and sellers narrows the spread—that is, the price difference between the best bid and offer.[8] Unsurprisingly, therefore, "liquidity begets liquidity. It [is] a fundamental law of markets, like gravity. The bigger the flow of trades, the stronger the pull."[9] And NYSE's pull proved irresistible. Centralization of the securities market was accomplished by the end of the nineteenth century: about two-thirds of all domestic trading now took place on the NYSE, while the rival exchanges in Boston and Philadelphia saw their shares reduced to 6.5 percent and 3.5 percent, respectively.[10]

The NYSE maintained its dominant position for over 100 years. Then, quite suddenly, the apparent iron law of liquidity begetting liquidity no longer seemed to apply, and market centralization was replaced by fragmentation. NYSE-listed stocks trade today in dozens of separate markets (see figure 1.1).[11]

It is important to emphasize that this development was by no means limited to the U.S. equity market. Elsewhere, too, market centralization was replaced with a common pattern of market

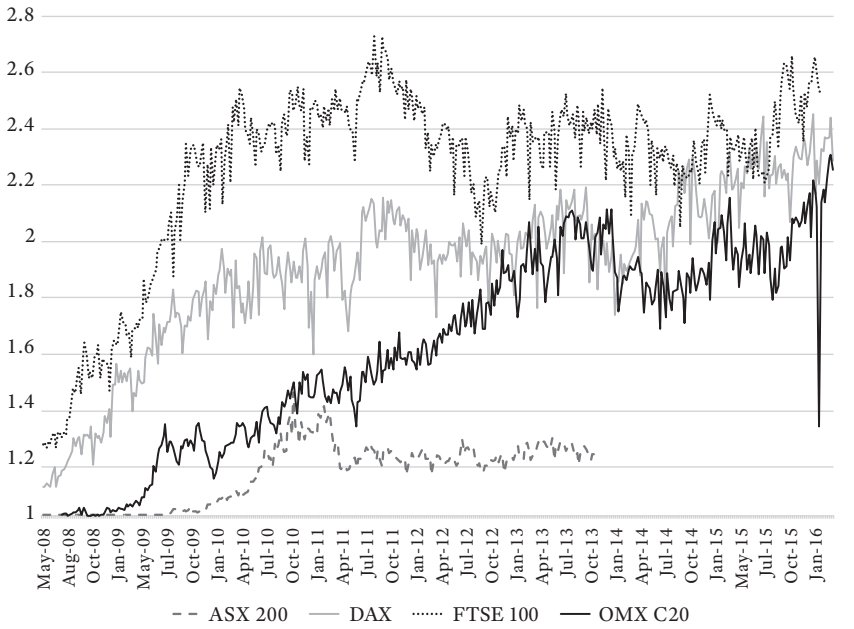

FIGURE 1.2. Market fragmentation outside the United States. *Source:* Private data received from Fidessa Group Plc in 2016.

fragmentation. Figure 1.2 illustrates this graphically, by plotting the Fidessa Fragmentation Index (FFI) for stocks included in four geographically diverse equity indices: the ASX 200 (Australia), DAX (Germany), FTSE 100 (United Kingdom), and OMX C20 (Denmark). Summarized briefly, the FFI represents the number of trading venues an investor is likely to have to visit to achieve best execution for an order—accounting for both the number of venues on which a given security is traded and their relative market share—such that a higher FFI score indicates greater market fragmentation. It is a commonly used aggregate-level measure of market structure. The picture that emerges from these statistics is clear: from a starting point of relative centralization in 2008, all four indices experienced a significant shift toward greater fragmentation over the subsequent decade. While this development was comparatively moderate in Australia, it was all the more pronounced elsewhere, with FFI scores effectively doubling over the period. This mirrored

the pattern observed in the United States with the fall from dominance of the NYSE. Across the globe, then, a new reality of market fragmentation rapidly emerged. What explains this deeply puzzling transformation of market structure?

The conventional narrative of securities market transformation in the United States runs as follows: the NYSE was *the* market—monopolistic and organized as a private club or a not-for-profit "utility." As such, it had little incentive to innovate or fix inefficiencies. In the words of a prominent set of authors: "The NYSE's members…preferred to milk a profitable franchise rather than aggressively innovate.… The supreme irony…[is that] throughout its history this bastion of the free-enterprise system has been operated as the very model of a socialist collective. That is why the NYSE has rarely been in the forefront of innovative services to investors."[12]

And like a socialist collective, according to this narrative, its fate was sealed; it would not withstand the pressure from a wave of highly competitive new-age electronic trading platforms spawned by the computer revolution. To survive, it had to shed its old "private club" ways and incorporate as a for-profit company—a process referred to as "demutualization"—and use the injection of public capital to completely overhaul its trading systems. From this process, facilitated by the regulators, an efficient new market structure with multiple modern exchanges fiercely competing for business emerged in the mid-2000s. A similar narrative has been used to explain the reasons behind equity market transformation in other major economies.

In this narrative, investors are the principal beneficiaries owing to narrower trading spreads and lower commissions. In sum: "The old [monopolistic] membership association structure fail[ed] to provide the flexibility and financing needed to compete in today's competitive environment. Over the long run, for-profit stock exchanges run by entrepreneurs and disciplined by profit-seeking investors should produce better-funded organizations with greater ability to…adapt to a fast-changing marketplace."[13]

This deeply entrenched conventional view is flawed. It is little more than a moral free market tale with a happy denouement.

First, market dominance should not be confused with monopoly. A monopoly restricts supply in order to raise prices and increase profits. The NYSE was never a monopoly; it had competition and could not increase commission charges and spreads at will without losing business to over-the-counter (i.e., off-exchange) traders or new rivals. In the nineteenth century, for example, NYSE fiercely competed for business with several New York–based trading venues, including the New Board, the Open Board of Stock Brokers, the National Stock Exchange, the New York Mining Stock Exchange, the Consolidated Exchange, and the New York Curb Exchange.[14] In the second half of the twentieth century, NYSE similarly came under intense competitive pressure from the over-the-counter market, the regional exchanges, and automatic electronic trading platform providers.

Second, the claim that NYSE failed because it was organized as a member-owned cooperative or "socialist collective" and thus lacked an incentive to innovate is manifestly wrong from a historical perspective. Survival in the competitive securities industry has always depended on NYSE's ability to continually improve its governance system and incorporate the latest technologies in support of its operations, as detailed in chapter 3. Critics are quick to point to one or another episode where the Exchange seemed complacent, failing to quickly spot a lapse in good governance or missing a chance to be ahead of the competition. The relevant issue, however, is not whether the NYSE was faultless or always at the forefront of innovation but whether it had a (market) structural incentive to improve and innovate. The NYSE was in a position similar to that of other old established corporate giants, such as General Electric or General Motors. They all had the occasional adjustment crisis but survived in their respective competitive markets on the back of a powerful incentive to innovate.

NYSE was the first exchange in the world to install stock tickers (in 1867) and telephones on the trading floor (in 1878). It had other firsts. In 1930, it introduced a high-speed ticker service, supplanting it in 1964 by an even more rapid ticker system capable of displaying 900 characters per minute. In 1966, it launched a fully

automated system that transmitted trade and quote data from the floor. Six years later, NYSE teamed up with the American Stock Exchange to form the Securities Industry Automation Corporation to develop computer support systems and provide consulting services in automated systems to the financial industry.

NYSE's response to the looming competitive threat from new electronic trading venues (so-called electronic communications networks, or ECNs) was greater automation. In 1976, it introduced the Designated Order Turnaround (later SuperDOT) electronic transmission system. In 1983, NYSE launched an electronic order book, providing for partial computerization of small buy and sell orders. By the late 1990s, over 90 percent of NYSE trades were handled electronically.[15] In 2000, finally, the NYSE introduced NYSE Direct+, a new system of automatic order execution. Investors were given the choice to enter orders either through SuperDOT for potential price improvement in the auction market managed by the floor specialists, or through the automatic execution system, where orders were executed directly against contra-side bids or offers. By the end of the 1990s, no trading floor brokerage fees applied to orders sent through SuperDOT, and NYSE-related transaction fees and commissions represented only 2.6 percent of the total amount paid by individual and institutional investors in brokerage commissions for NYSE-listed securities.[16]

Total NYSE spending on new technologies in the 1990s was over $2 billion.[17] This investment paid off. At the dawn of the twenty-first century, the NYSE was the world's preeminent equity market: average daily trading volume grew from 157 million shares in 1990 to 781 million in 1999—in a trading system with a technical capacity of processing 4 billion shares a day. Over the same period, listings more than doubled to nearly 3,000, including over 400 from 48 other countries. NYSE-listed companies had a staggering total market capitalization of $19.8 trillion—$12.8 trillion for American stocks and $7 trillion for foreign stocks. By comparison, the next largest equity market outside the United States was the Tokyo Stock Exchange, running a distant second with a market capitalization of $3.7 trillion. This was followed by the London Stock Exchange with

$2.8 trillion, Euronext with $2.4 trillion, and the Deutsche Borse with $1.2 trillion.[18] These trends and figures belie the image of a complacent, lazy, inefficient, or moribund market organization.

In short, it was not monopoly complacency then, or "membership socialism," or any disease of organizational old age that killed the dominant exchange model and changed market structure. The answer must lie elsewhere.

Power Politics and Market Governance

The mysterious death of the world's leading equity market, the old NYSE, is forensically investigated in chapter 2. The key finding is that power politics within the NYSE killed it—a plot by a coterie of powerful insiders who had grown weary of the traditional way of organizing trading, viewed the old model increasingly as contrary to their economic interests, and quietly pushed for a different market structure more aligned with those interests. Technology plays a central role in explaining the rise in power of some NYSE members, but, in the final analysis, it is power politics—not technological change per se—that explains the end of organized market dominance and the advent of a fragmented market reality.

There is a simple logic to the plot. Traditionally, the typical member of NYSE was the small broker partnership. NYSE membership comprised about 600 such partnerships during the first half of the twentieth century. Some member firms handled retail clients, others looked after wholesale customers, still others specialized in market making in specific stocks, thus serving as brokers of (retail and wholesale) brokers, and so on. The picture that emerges from the empirical investigation in chapter 2 is that of a membership teeming with a multitude of relatively small and highly specialized operators. They were essential cogs in a sophisticated market-constituting "machine" and depended for their livelihood on trading operations on the floor. They thus had a strong vested interest in the success and good reputation of their market. Each member had an equal voice on key regulatory and policy matters, and no single member group was dominant or prevailed.

The face of the NYSE membership changed in the second half of the twentieth century as a result of successive waves of mergers and acquisitions triggered by the computer revolution and changes in membership rules that allowed public companies, notably highly capitalized banks, to become members. The long-standing balance of power within the membership disappeared. In its place emerged a hierarchy of economic power, with a few giant broker-dealers at the top, including Merrill Lynch, Goldman Sachs, Morgan Stanley, JP Morgan, UBS, Credit Suisse, Barclays, and Deutsche Bank. Their big size made them latent markets; that is, they had such an enormous client base that they could profitably match client orders in-house, rather than routing the orders to NYSE and paying a fee for trade execution. Only unmatched orders would be sent to the Exchange. They could further profit by setting up proprietary trading desks and trading against client orders.

Big size was a source of power because it reduced dependence on the Exchange, even while the Exchange remained heavily dependent on big members for liquidity. The larger the members, the greater the potential gains and savings from running in-house markets and the stronger these members' incentive to push for a breakup of the old system. A democratic form of market governance, where a numeric majority of small members had an operational voice and voting rules failed to reflect economic importance, was of no interest to them. It held them back in their desire and ambition to freely expand their business. In the early 2000s, they decisively moved against the old NYSE. Death came swiftly and ushered in an era of market fragmentation.

This plot travels well beyond the NYSE. Most formerly dominant market organizations, including the London Stock Exchange and leading commodity exchanges in the world, share many of the key organizational features of NYSE, and they underwent similar transformations. So why should we be concerned about such change in market structure?

Market transformations matter because they shape the incentives of market organizations to invest in either good or bad governance. As noted above, good market governance is about managing

conflicts of interest for the long-term benefit of all in society. Well-governed market organizations perform several socially beneficial functions. They create standardized financial instruments (like stocks and bonds), facilitate trading among strangers, transfer ownership as well as risk, and, perhaps most importantly, provide "price discovery"—that is, they produce price information that accurately reflects the true value of a security or its underlying asset. Accurate price information is a public good. "[It] help[s]…to allocate the economy's scarce capital to the most promising potential real investment projects and…improv[es]…the utilization of the economy's existing productive capacity through optimizing the signals provided to management about investment decisions and the signals given to boards and shareholders about the quality of management decisions."[19] Good governance reduces the likelihood of opportunistic market manipulations that distort the production of accurate price information.

Good market governance is expensive, however, necessitating extensive investments in developing stringent private rules, robust surveillance, and strict enforcement. A dominant exchange, I argue, is likely to have an incentive to make these investments. This is because dominance means public visibility, which, in turn, entails particular reputational vulnerabilities. Fraudulent trading by one exchange member risks damaging the reputation of the entire exchange. And because the potential wealth gained by one member acting opportunistically is generally more than offset by the wealth lost by the many other exchange members as a result of the reputational damage inflicted by the one, a dominant exchange has a particularly strong incentive to invest in robust governance safeguards.

Reputational sensitivity is compounded by the fact that dominance may attract blame even in the absence of fault or culpability. A dominant or focal exchange is all too easy a scapegoat for anything that may go wrong in the wider financial market. The chief economist of the NYSE recognized this problem in the early 1920s, lamenting that "rumors…never start on the floor of the Exchange but outside it; [but] since their effects are principally felt in the

stock market, the opposite is commonly supposed. The tense and imaginative atmosphere of Wall Street is peculiarly liable to magnify trifles into bonanzas or catastrophes—but we must remember that the Stock Exchange and Wall Street are not synonymous."[20]

Evidence of high levels of investment in governance in centralized markets is abundant (and assessed more systematically in chapter 3). The LSE, for example, early on had an unusually stringent admissions policy to protect its reputation. Members had to reapply annually for admission, giving the Exchange "regular opportunity to refuse re-entry to those considered unsatisfactory.... This was in addition to the ability to expel for specific misdeeds or to reject unsuitable new applicants."[21] On the trading floor, deals were made based on honor and the word of Exchange members. "Reputations [were]...more important...than ironbound contracts complete with sinuous codicils."[22] Reneging on a deal with another member resulted in immediate expulsion, not only from the Exchange but from the social life of the city.[23]

A well-governed exchange, however, faces a serious challenge—a so-called free rider problem. A free rider is an actor who is able to benefit from the provision of a public good by another actor without having to incur the cost of creating, providing, or maintaining it. In this case, rival exchanges or off-exchange trading venues may take price information from the well-governed exchange to undermine it. Because they do not incur the considerable expense of creating and maintaining a price discovery mechanism, these rivals can charge lower commissions or offer better prices. In the nineteenth century the dominant NYSE competed for business with many local exchange providers. Some of these rivals were ingenious free riders.[24] For example, in the 1830s the members of the New Board rented a room in a building "next to the...room [occupied by the NYSE], and dug the bricks out of the wall in order that they might see and hear what was going on."[25] Another example is provided in the July 1857 issue of Hunt's *Merchants' Magazine*: "Curbstone brokers have leased a large room directly under that occupied by the [NYSE].... Any transaction [on the NYSE] is known below as

soon as made."[26] And, as expected: "The contracts of many of the curb-stone brokers are infinitely better than many of the [NYSE deals], and are taken more readily by outsiders."[27]

How, then, did dominant exchanges survive in a context of free riding by rivals? The key here is economies of scale that derive from dominance and, in effect, represent a "subsidy" essential to the maintenance of good market governance.

Scale economies stem from two sources. The first source is diminishing marginal costs in market operations. The capital costs of setting up a market organization (i.e., the fixed costs) have historically been substantial. Funds have to be raised to build a trading house and equip it with the most up-to-date trading technologies. However, once the organization is in place with its rules and sophisticated systems, the marginal cost of managing an increase in trading volume (or additional listings) declines over a long stretch, until a maximum operational limit is reached.

This limit is only ever attained in exceptional circumstances. Nevertheless, dominant exchanges, eager to protect their reputations, have to be able to cope with sudden volume surges in moments of great market stress. Writing in the 1950s, Wall Street observer Martin Mayer noted: "One of the reasons the Stock Exchange is so expensive to operate is that nobody ever knows what the volume of trading will be tomorrow; and the Exchange likes to be prepared for the three-million-share days that mean prosperity. Since the average day will probably run under two million shares . . . the place is pretty badly overstaffed."[28]

The significance of scale economies that derive from an initial investment in large fixed market-building assets appeared to wane with the advent of new electronic markets in the 1990s and the subsequent closure of many traditional trading floors. However, the price of advanced trading systems has been escalating of late. This may lead in the coming years to a growing significance of fixed costs and thus the size—and economies of scale—of a market organization (see chapters 4 and 5).

The second source of scale economies derives from user-network effects. As discussed above, the greater the number and

variety of users of a given market organization—speculators as well as retail and institutional investors—the more attractive the market is to new or potential users. Such users (buyers or sellers of a security) are more likely to find counterparties in a large market or "trading network" than in a small one. And as "liquidity begets liquidity," more revenues are generated to finance investments in good market governance. Such governance, in turn, further strengthens the exchange's appeal, solidifying commercial success and securing continuing investments.

In North America, this virtuous dynamic interplay of factors affirmed the position of NYSE as the leading exchange by the early twentieth century. "Increasingly, refinements were added to the operations of the NYSE whereby it became better than any of its rivals.... In terms of the volume of securities to be bought or sold, the speed at which it could be done, [and] the narrowest spread between buy and sell prices...no other exchange in the Americas could challenge the NYSE, and so it attracted business from throughout the nation."[29] And though free riding and price competition remained a persistent challenge, powerful economies of scale allowed NYSE to achieve savings and attract clients to finance expensive governance investment, generating a further pull of business that enabled the Exchange not only to survive but to prosper over the next hundred years, as detailed in chapter 3.

The governance implications of market fragmentation, which is an accelerating trend in today's capital markets, contrast sharply with those of market centralization. In a fragmented system, liquidity no longer flows naturally into a single large pool, and economies of scale are thus much reduced or absent. Competing trading venues now have to balance costly investments in good governance against an overriding new mandate to attract liquidity to survive. Not all market participants are equal in their ability to supply liquidity. Large liquidity providers are enormously powerful in fragmented markets because they can shop around in the bazaar of abundant market options, extracting extensive trading privileges and various other concessions from competing exchanges and off-exchange trading venues. The most powerful liquidity providers include the

traders at large broker-dealer banks—not coincidentally the very same market actors who successfully pushed for fragmentation in the first place (see chapter 2)—and a new generation of high-speed proprietary trading firms. They possess the fastest communications lines, the most advanced computers, and sophisticated algos capable of dispatching huge numbers of orders to trading venues within milliseconds of spotting profit opportunities. When their demands for privileged treatment clash with principles of good governance, the latter become dispensable for trading venues. Institutional investors are second-class citizens in this new market order. Long-standing commitments of traditional exchanges to fairness, equality, and transparency are sacrificed to efforts to curry favor with the powerful. The result is the emergence of bad market governance—a system designed to milk conflicts of interest for the benefit of the powerful at the expense of the investing public.

The evidence for this stark claim is examined in great detail in chapter 5. Although it is empirically challenging to quantify the harm being done to investors, the findings are troubling. They reveal an extraordinary and unprecedented catalog of governance failings by market providers since the onset of fragmentation in the second half of the 2000s: secretive discriminatory operations; undisclosed business practices inconsistent with exchange rules or securities law; ineffective oversight and accountability mechanisms; deliberate strategies to keep regulators and investors in the dark through various forms of deception, including lying, concealing, and spinning; failure to take corrective action even when told to do so by the regulators; and inadequate testing protocols and monitoring procedures to ensure that the operational systems comply with rules and regulations. The cloak-and-dagger and systematic nature of many of the shenanigans quietly facilitated by market providers, and the utter disregard of the architects of such shenanigans for the adverse consequences for investors and for society more generally, are deeply disturbing.

These governance failings are the inevitable by-product of determined and relentless efforts by competing trading venues to win

over the business of high-speed liquidity providers. Trading venues today offer a wide range of special services favoring high-speed traders—at the expense of investors. One such proprietary service is labeled "enriched private data feeds." These feeds contain an astonishing amount of trading information that exchanges collect each time a client places an order, including order execution details, modifications, and cancellations. Although access to enriched data feeds is nondiscriminatory in principle, only relatively few market participants possess the costly hardware and software to standardize raw data feeds and decrypt millions of messages in milliseconds. Millions of investors have no idea that such trading data is being collected and sold to high-speed traders.

A companion service, "colocation," involves placing a high-speed trader's computer server or black box next to the exchange's trade-matching engine in the data center of the exchange, thus reducing the order and information travel time between server and matching engine.

Private data feeds and colocation enable high-speed liquidity providers to anticipate investor order flow, infer short-term price movements of stocks, and trade ahead of other market participants. Individual or institutional investors do not typically possess the resources to pay for these special services or invest in the telecommunications and computer systems needed to support and benefit from them. The high cost of these services crowds out most investors. As a result, investors run the risk of paying more for stocks they buy or receiving less for stocks they sell.

So-called special order types (SOTs) are another service on offer. SOTs are complex buy and sell orders that define how an order is placed in a market, how it is displayed, and how it interacts with other orders. Certain opaque SOTs allow orders of powerful clients to remain hidden and jump the queue to be first in line of execution when the clients wish to enter a trade. This discriminatory treatment of order handling imposes significant costs on investors. Exchanges have produced hundreds of SOTs in the rat race to attract liquidity from high-speed traders.

A third prominent service is the provision of dark pools. These pools were initially designed to enable big institutional investors to place large "block" orders without tipping the market to their intentions, which could trigger adverse price changes. In a lit market, the risk of displaying a large sell order, for example, is that buyers may pull their orders out of the market in the hope of soon buying at lower prices, thereby depressing the price for the seller. Dark pools provided a solution by simply matching large buy and sell orders. In a fragmented market system, however, many dark pools became vehicles of market abuse. By adding a layer of darkness to the invisibility of ultrafast trading, these pools created a practically foolproof environment for opportunistic trading. As will be detailed in chapter 5, the abuse in dark pools has been truly shocking.

This evidence may represent only the tip of the iceberg of market shenanigans. Firms are exploiting conflicts of interest under cover of darkness, so to speak, to the detriment of investors. For regulators and academics alike, governance failings and market manipulation are hard to spot and investigate in today's highly fragmented and complex markets where tens of billions of trades are executed daily at the speed of micro- or even nanoseconds. No watchdog possesses adequate surveillance tools or data analytical capabilities to be able to systematically monitor these markets and deter market abuse.

The book concludes with a few reflections on how to reengineer good market governance in today's capital markets—that is, how to bring back healthy markets that ensure fairness, orderliness, and efficiency. My main propositions may surprise the reader. Failings in market governance are rarely fixed by governmental intervention. It is a considerable challenge for governmental rules to keep apace with changing technology and resulting new market practices. Governments can provide the basic parameters of fair play, most importantly through disclosure regulation or legislation, but market failures are most effectively resolved through market intervention. Here governments can play an important role in enabling

certain market players to contribute to market solutions, by nudging or incentivizing them in the right direction. Specifically, a transformation from a heavily fragmented market toward consolidation or centralization will bring about a simpler and more transparent marketplace. The existence of a dominant market organization exposed to relentless sunlight and persistent competition from ambitious newcomers or free riders will generate better market governance.

INDEX

Page numbers in *italics* refer to figures and tables.