Institutional intermediaries have grown massively in American capital markets since the mid-20th century. Where investors used to buy and sell stocks and bonds directly, they now overwhelmingly invest through mutual funds, hedge funds, private equity funds and similar vehicles. Fully three quarters of the common stock of America’s public companies now belongs to institutional intermediaries. Broker-dealers and credit rating agencies have changed as well as they have weathered the financial crisis and searched out new opportunities in markets dominated by intermediaries. This essay, written for the Conference on the New Special Study of Securities Markets at Columbia Law School, identifies the key regulatory challenges posed by the changing role of America’s institutional intermediaries. We survey existing legal and economic research in the area and suggest new areas for regulatory reform and scholarly inquiry. We cover registered investment companies, such as mutual funds, private investment funds, such as hedge funds and private equity funds, and credit-rating agencies and broker-dealers.

1. REGISTERED INVESTMENT COMPANIES

We begin with registered investment companies. Registered investment companies, or “RICs,” consist mostly of the open-end mutual funds that dominate household investing and are publicly registered and regulated under the Investment Company Act of 1940. America’s system for regulating these vehicles has been, by almost any measure, an enormous success. Since Congress adopted the Investment Company Act (the “ICA”) and its sister statute, the Investment Advisers Act (the “IAA”) in 1940, the mutual fund industry has grown massively, progressing from its origins as a niche industry for wealthy northeasterners to a vast behemoth. Investment funds publicly registered under the ICA now reach 43% of American households and comprise

---

1 Greenfield Professor of Securities Law, Harvard Law School, fferrell@law.harvard.edu; Professor of Law, Yale Law School, john.morley@yale.edu.
some $18 trillion in assets (Investment Company Institute 2016, p. 112). Even more remarkable is that the investment fund industry’s growth has brought with it remarkably few problems. Though the 1920s and 1930s witnessed extensive fraud and abuse and saw the dramatic bankruptcy of dozens of large publicly traded investment funds, the years since the ICA and IAA began regulating investment funds in 1940 have been impressively quiet. Even during the once-in-a-generation catastrophe of the financial crisis of 2008 and 2009, remarkably few publicly registered investment funds collapsed from debt or illiquidity. Fraud, though an ever-present challenge, has never been pervasive or widespread. The ICA and IAA have become models for regulatory statutes around the world.

Despite their enormous success, however, the ICA and IAA are now significantly out of date. Drafted when the investment fund industry was only about 15 years old, the ICA and IAA reflect only a rough understanding of what an investment fund was and what the investment management industry would become. The main targets of the ICA’s detailed provisions were closed-end funds, but in the years since 1940, the industry has become overwhelmingly dominated by open-end funds. And the statute’s drafters had no inkling of the future rise of private funds, such as hedge funds and private equity funds. The ICA was written before modern law and economics and even before modern portfolio theory, and it reflects little awareness of some of the foundational principles of organizational economics.

The Securities and Exchange Commission has dealt with the weaknesses of the ICA and IAA admirably, but there have always been limits to the SEC’s authority. The ICA permits the SEC enormous power to grant exemptions from the statute\(^2\) and the SEC has often used this exemption power to effectively rewrite many portions of the statute. But rewriting is not the same as reimagining, and there yet remain many innovations that Congress and the SEC could pursue. Our goal here is to consider those possibilities and how we might learn more about them.

a) **Definition of an Investment Company**

The first task is to define what exactly an investment company is. The challenge here is not to distinguish between a public investment company and a private investment company, but to distinguish between an investment company and an operating company. Instead of probing the

\(^2\) Investment Company Act § 6(c).
difference between the Vanguard S&P 500 index mutual fund and a hedge fund, in other words, we wish here to probe the difference between the Vanguard S&P 500 index mutual fund and Microsoft. The line between an investment company and an operating company matters, because it determines who has to comply with the ICA (either by registering under the ICA or by staying private and issuing securities only to a limited class of investors). For a public company that does not think of itself as an investment fund, having to comply with the ICA can be disastrous.

Though the dividing line between an investment company and an operating company may seem obvious, it has been surprisingly difficult to draw in practice.\(^3\) Intuition tells us that an investment fund is a business that *invests*. But every business invests in some sort of asset, whether factories, brands or human capital, and so the ICA has to say more. The ICA thus defines an investment company not just as a business that invests, but a business that invests in a particular kind of asset, namely *securities*. Section 3 of the statute says that an investment fund, among other things, is an issuer that “is... engaged primarily...in the business of investing, reinvesting, or trading in securities,” or which “owns...investment securities having a value exceeding 40 per centum of the value of such issuer’s total assets.”\(^4\)

This focus on securities ownership seems intuitive at first, but it often produces strange results that bump awkwardly against a deeper sense of what an investment fund is. Consider, for instance, PepsiCo. Everyone would agree that PepsiCo is an operating business, not an investment fund, but it is hard to say exactly why. Though it would seem that most of Pepsi’s assets consist of factories and brand names, in fact most of its assets are securities. The reason is that PepsiCo does not actually own the factories and brands directly—instead it owns securities in operating subsidiaries that own the factories and brands directly. If securities ownership is the defining essence of an investment fund, therefore, it is hard to see why PepsiCo is not an investment fund.

The ICA’s solution to this conundrum is to say that PepsiCo is not an investment fund, because its securities in the operating subsidiaries represent control stakes, rather than minority

---

\(^3\) The term “investment company” has never been popular in common usage. Prior to Congress’ adoption of the ICA in 1940, investment vehicles were commonly known as “investment trusts” (Morley, 2016). The term “investment company” was thus a neologism invented by the ICA’s drafters to satisfy the industry’s desire for a term that was neutral as to entity form. The term never stuck in popular usage, however, and nowadays people commonly use the word “fund,” rather than company. We will often follow this practice in this paper except when referring specifically to the statutory definition of an investment company in section 3 of the ICA.

\(^4\) Investment Company Act §§ 3(a)(1)(A), (C).
stakes, and this makes PepsiCo’s securities different from an investment company’s securities.\textsuperscript{5} This makes sense, until we compare PepsiCo to a private equity fund. Just like PepsiCo, a private equity fund also holds securities that represent control stakes. And yet a private equity fund is clearly an “investment company” within the meaning and intent of the ICA (though it avoids having to comply with the ICA by remaining private). So if a private equity fund is an investment company, why not PepsiCo? Though we can go on to find legal details in section 3 of the ICA that permit us to distinguish between PepsiCo and a private equity fund, it is not clear whether any of these details truly matter, or whether they are just bandages added on to the ICA to cover an obvious hole.\textsuperscript{6}

Consider also Yahoo!, the internet search and technology company. As the value of Yahoo!’s internet search business has declined, the company’s main asset has become its large and valuable stakes in other companies, such as Alibaba, the Chinese online retailer. Though Yahoo! thinks of itself as an operating business, under section 3 of the ICA, Yahoo! technically qualifies as an investment company, because its securities in other companies now comprise more than 40% of its assets (Gorta, 2017, February 28).\textsuperscript{7} Yahoo! has not actually had to comply with the ICA, because it received a special exemption from the SEC early in its life. But there is no real principle behind this exemption, just the SEC’s discretionary decision to let Yahoo! go because of a vague intuition that Yahoo! is an operating company for reasons that no one can clearly articulate.

Think also about Microsoft. Early in its history, Microsoft raised cash to fund research and development operations and then invested the cash in securities as it waited for the research and development program to use the cash up. Microsoft clearly intended to operate a real business, but because none of Microsoft’s intellectual property had any significant accounting value, the investment securities in which Microsoft had parked its cash quickly became the only significant item on the company’s balance sheet. From the perspective of assets, Microsoft was

\textsuperscript{5} Investment Company Act § 3(b).

\textsuperscript{6} The deciding detail is the fact that a private equity fund, unlike PepsiCo is “engaged in the business of investing, reinvesting, or trading in securities,” whereas PepsiCo is engaged in the business of making snack foods and soft drinks. Investment Company Act § 3(a)(1)(A). But this just begs the question, which is what exactly it means to “invest, reinvest, or trade in securities” and how this differs from carrying on an operating business through the ownership of securities.

\textsuperscript{7} ICA § 3(a)(1)(C).
Indistinguishable from a closed-end bond fund. Like Yahoo!, Microsoft also received a special exemption from the SEC, again with no discernible principle behind it.

And what about a pension fund? Just like an investment company, a pension fund’s main business is to invest in securities. So why is a pension fund not an investment company under the ICA? The answer, obviously, is that a pension fund is subject to a different regulatory regime, the Employment Retirement Income Security Act, or ERISA. But if we believe that the primary and essential characteristic of an investment fund is its tendency to invest in securities, then it is not obvious why this different regulatory regime should exist. A pension fund and an investment company are surely different—but not in terms of the one key characteristic that the ICA singles out as the defining essence of an investment company. So why should we regulate them differently? It must be because of one of the many differences that the ICA does not single out.

The ICA and SEC have employed a number of patches to address each of these problems, so that neither PepsiCo, Yahoo!, Microsoft nor the nation’s many pension funds have had to comply with the ICA. Still, the need for patches and fixes in such obvious cases suggests that there is something deeply wrong with the ICA definition. The core features of the ICA definition do not match our intuition about what an investment fund truly is. To make up for the ICA definition’s basic incoherence, the patches and fixes have grown so numerous and complex that the definition of an investment company in section 3 of the ICA now spans nearly 3,000 words and 94 separately numbered subsections, paragraphs and subparagraphs. And even then, the SEC has only been able to spare some companies, such as Yahoo! and Microsoft, by granting them special one-off exemptions, without any attempt to ground the exemptions in principle, regulation or statute. The definition of an investment fund has become a Rube Goldberg contraption, covered in duct tape and Elmer’s glue.

The result is a system of often incoherent differences in treatment. To deal with future iterations of the Microsoft problem, for example, the SEC adopted Investment Company Act Rule 3a-8, which exempts certain companies from the Investment Company Act if they invest large amounts of cash in securities and then quickly spend the cash on research and development activities. This rule works fine as far as it goes, but its logic is unclear and its effects are unfair. If indeed a company exempted by the rule holds large amounts of securities, which is the sine

---

qua non of an investment company under ICA section 3, then why should the company be exempted? And why is the exemption limited to companies that spend money on research and development, rather than on other activities? What if a company raises cash to explore for diamonds or develop video content for a web site? Neither of these activities qualifies as “research and development” and so neither would get the exemption in Rule 3a-8. But there is no obvious reason why.

This confusion forces companies to live with frustrating levels of uncertainty. As recently as February 2017, Yahoo! fought off a shareholder suit in federal district court alleging that the company was an investment company under the ICA. Though the company had received a special one-off exemption from the SEC early in the company’s life, the plaintiffs argued that the company lost the exemption as more and more of its balance sheet became devoted to securities. Since there is now no discernible legal reason why Yahoo! should not be an investment company other than the SEC’s unprincipled discretionary grant of an exemption, Yahoo!’s status under the statute has become precarious.

Future research should search for a more elegant and principled solution. The persistent difficulties with the section 3 definition suggest that securities ownership is not, in fact, the sole essential feature of an investment fund, and so perhaps we ought to look more deeply. One possibility, identified by Morley (2014), is to focus on organizational structure. Morley argues that the most salient feature of an investment fund is not its tendency to invest in securities, but its tendency to separate its investment holdings from its management structure. Almost everything we commonly think of as an investment fund in the ICA regime has a unique tendency to combine two distinct enterprises: a fund holding investment assets, and a management company or adviser holding workers, computers, office space and other managerial assets. These two distinct enterprises are not merely different entities in the same parent-subsidiary family—they are distinct businesses with distinct owners. This pattern of bifurcated organization creates an array of unique contractual and regulatory challenges, such as the possibility that a manager can face conflicts of interest by simultaneously working for multiple clients at the same time. It is perhaps these organizational challenges, more than any unique characteristic of securities ownership, that demand the special regulatory attention of the ICA.

When coupled with a focus on securities ownership, this bifurcated pattern of organization may do better job of identifying investment companies than securities ownership alone. The
thing that makes PepsiCo, Microsoft, Yahoo! and a pension fund different from an investment company is that none of these businesses divides its investments from its management like an investment company. Though some of these businesses employ multiple entities, all of their entities exist in the same parent-subsidiary structure under common ownership. None of these businesses could be said to have a truly external adviser like a hedge fund or mutual fund.

In future study, the Commission might examine how this pattern of bifurcated organization maps onto the existing investment company definition and how useful it would be in close cases, such as Yahoo!. Researchers might also search for other, more elegant ways of defining an investment company in addition to this method focused on organization.

b) Management Fees

Once we define what an investment company is, we next have to figure out how to regulate it. And no issue in investment company regulation has been more hotly debated than the fees charged by managers. Though management fees have declined profoundly since the ICA’s adoption in 1940, many commentators nevertheless argue that fees must decline even further before they fairly reflect the value of the services that investment companies provide.

Financial economists have assembled a mountain of theory and evidence on the fees charged by open-end mutual funds and there is still a robust debate about whether fees are too high. In this paper, however, we will try to avoid this debate. Rather than staking out a position on whether the mutual fund market is competitive or not, we will focus instead on the legal details of how to craft effective fee regulation. Though people may disagree about the big question of whether the market is competitive, we believe that almost everyone can agree that if the government is going to implement fee regulation, it ought to do so in a way that is maximally effective and minimally intrusive. We will thus ignore the grand economic question of whether fees are excessive and focus instead on the technical legal question of how exactly one might build an effective fee regulatory system in practice, assuming such a system is thought to be necessary.

The first task is to learn more about the least competitive segments of the mutual fund market. It seems fair to say, based on the last twenty years of economic research, that although most mutual fund managers charge reasonably competitive fees, at least some do not. We need
not march into the center of the battlefield over fees by expressing an opinion on just how numerous the uncompetitive managers are. It is enough to say merely that the inability of investors to understand and rationally react to fees in experimental settings suggests that even if large management complexes such as Vanguard, Fidelity, State Street and Blackrock charge highly competitive fees (as indeed they almost certainly do), there nevertheless may be some significant number of managers who do not charge highly competitive fees. (Choi, Laibson & Madrian, 2010; Mercer, Palmiter & Taha, 2010; Beshears, Choi, Laibson & Madrian, 2009; Kozup & Pagano, 2008). Carhart’s (1997) famous paper provides support for this view. Though Carhart found that very few managers could overperform year after year, a distressing number of managers tended to underperform year after year.

More knowledge about the worst funds would be useful, because although researchers such as Coates and Hubbard (2007) have taught us a tremendous amount about the competitiveness of the mutual fund market overall, we still know too little about what lies in the market’s darkest corners. And these dark corners are where regulation is most urgently needed.

There is still much we could learn about the worst funds. Research has already taught us, for example, that funds in the highest decile of fees tend to be concentrated among management companies in the lowest decile of size (Curtis & Morley, 2014a). But we do not yet know what investors in these high-fee funds are like. Did they initially understand and accept the high fees in exchange for elaborate add-on services? Or did they initially invest with low fees and then lethargically stay put as the fees crept slowly higher? Do the highest-fee funds have institutional share classes with large institutional investors? Do extremely high-fee funds tend to be older than other funds or younger? Is performance more likely to persist at the high end of the fee distribution than in the middle? And do high-fee funds tend to provide additional services beyond mere portfolio management that independently warrant the high fees?

Beyond understanding the worst-performing and highest-fee funds, a second task will be to learn more about the precise details of how exactly investors respond to fee disclosures. Even if there are sound theoretical reasons not to worry about the weak correlation between past and future returns, as Berk and Green (2004) have argued, everyone can surely agree that there is dysfunction lurking in investors’ tendency to misunderstand disclosures, because experimental evidence strongly indicates that many investors do not understand fees or their importance (Choi,
One way to improve fee regulation, therefore, would be to experimentally test different forms of disclosure. There is still much to be learned here, because although experimental researchers have already taught us which sorts of disclosures do not work, they have not much explored which sorts of disclosures do work. We know that existing forms of disclosure do a bad job of pushing investors to low-fee funds, but we do not know whether other forms of disclosure would do a better job. The SEC or private researchers should therefore test different forms of disclosure by presenting experimental subjects with a variety of different forms. Insights from cognitive psychology could be used to theorize how different forms of disclosures might affect investors, and then the effect of the disclosures could be tested by seeing how investors respond to them. A good template for this kind of research would be the work of Bertrand and Morse (2011), who identified potential borrowers from payday lenders and assessed how different disclosures forms affected a borrower’s likelihood of taking out a payday loan in the future. Another template is the work of Beshears, Choi, Laibson & Madrian (2009), who randomly gave experimental subjects either a standard mutual fund prospectus or the new summary mutual fund prospectus mandated by the SEC in 2009 and then compared how the two documents influenced investors’ choices (with sadly disappointing results). Researchers could take a step beyond the Beshears, Choi, Laibson & Madrian work by imagining entirely new forms of disclosure and seeing how they work.

A third task will be to revise and repair the excessive fee liability provisions of section 36(b) of the ICA. Congress added section 36(b) to the ICA in 1970 by saying that a fund’s adviser has a fiduciary duty to the fund regarding fees the adviser receives from the fund. Congress left the content of this fiduciary duty unexplained, but the courts subsequently supplied the content, most recently in Jones v. Harris Associates, a 2010 case in which the Supreme Court formally adopted a standard that had long prevailed among the circuit courts. Jones said that a fee violates an adviser’s fiduciary duty if it is “so disproportionately large that it bears no reasonable relationship to the services rendered and could not have been the product of arm’s length

9 Jones v. Harris Assocs. L.P., 559 U.S. 335 (2010); Gartenberg v. Merrill Lynch Asset Mgmt., Inc., 694 F.2d 923, 928 (2d Cir. 1982).
bargaining.” Prior judicial opinions in the circuit courts developed a list of six factors to aid judges in their assessment of this general standard.⁹

Section 36(b) may be the single most written-about topic in all of investment management regulation—Curtis and Morley (2014b) collect the literature in a two-page long footnote.¹¹ Most of this literature tends to focus on the grand economic question of whether the mutual fund market is competitive and whether excessive fee liability is therefore desirable.¹² But in this paper, we will steer clear of this grand debate, focusing on the more practical—and in many ways more important—question of how exactly excessive fee liability should be crafted, assuming it is thought to be necessary.

The clearest empirical picture of section 36(b)’s present functioning comes from a 2014 study by Curtis and Morley (2014a). The picture is not encouraging. Curtis and Morley statistically examined every section 36(b) case filed between 2000 and 2009 and found that although there is some positive correlation between fees and the likelihood that a fund would be targeted for a suit, the correlation was fairly weak. Ultimately the strongest predictor of whether a fund would be targeted was not its fees, but the size of its adviser. Lawyers tended to go after the biggest advisers, rather than the most expensive funds. Indeed, almost no funds managed by the smallest one third of advisers were ever targeted, even though these were the funds that tended to charge the highest fees.

There were other troubling findings as well. One was that funds targeted for excessive fee lawsuits did not tend to reduce their fees after being targeted. Another was that the rate of settlements bore no discernible relationship to the level of fees. Overall, the results were consistent with a pattern of scattershot filing of complaints observed by the study’s authors. Plaintiffs’ law firms tended to pursue excessive fee suits by filing standard-form complaints against dozens of fund managers at a time, suggesting that targeting decisions were guided more by access to plaintiffs with standing to bring suit than by careful assessments of actual fee levels.

¹⁰ These include,
(a) the nature and quality of services provided to fund shareholders; (b) the profitability of the fund to the adviser-manager; (c) fall-out benefits; (d) economies of scale; (e) comparative fee structures; and (f) the independence and conscientiousness of the trustees.

¹¹ Johnson (2008) also provides a nice summary.

¹² For a sample, see the recent articles by Birdthistle (2010); Coates (2010); Coates & Hubbard (2007); Cox and Payne (2005); Fisch (2010); Freeman, Brown and Pomerantz (2008); Freeman and Brown (2001); Henderson (2010); Johnsen (2010); Langevoort, (2005); and Ribstein (2010).
Besides its poor functioning in practice, there are others reasons to worry about how section 36(b) operates in its mechanical details. In a legal analysis of section 36(b) published separately from their empirical analysis, Curtis and Morley (2014b) identified a number of mechanical flaws in the statute and judicial doctrine. These flaws seemed to involve unwitting mistakes of drafting and construction, rather than deliberate policy choices. One problem was that the judicial standard for assessing liability provides little space for an obviously low-fee fund to avoid a protracted lawsuit. The Supreme Court’s standard in *Jones* emphasizes that comparisons to other funds are not dispositive, meaning that a defendant advisor cannot resolve a case on a motion for summary judgment by showing that its fees are lower than those of its competitors—even if this is indisputably true. The problem with this standard is that it makes even low-fee funds vulnerable to the risk that a suit will proceed beyond a motion for summary judgment. Fidelity, for example, tends to charge very low fees, but was recently mired in section 36(b) litigation for years. The vulnerability of low-fee funds to suits is distressing not just because it is unfair, but also because it undermines the purposes of section 36(b) by diminishing the incentive to reduce fees. If an adviser cannot reduce its odds of getting sued by reducing its fees, then why then should it reduce its fees?

A further problem is the remedies. Section 36(b) restricts recoveries to the excessive portion of fees paid by a fund during the period beginning one year prior to the commencement of a suit and continuing to the suit’s termination. This recovery provides too little incentive for lawyers to pursue the most meritorious lawsuits. As noted, the highest-fee funds tend to be affiliated with the smallest advisers, and yet the smallest advisers were almost never targeted during the period of Curtis and Morley’s empirical study, probably because the potential recoveries on the highest-fee funds—and thus the potential rewards to plaintiffs’ attorneys — were too small. The only way to encourage lawyers to go after the highest-fee funds would thus be to increase the recoveries or the attorneys’ fees. The statute’s prohibition on punitive damages is also a problem, because it undercuts deterrence. The worst that can happen to an adviser that charges excessive fees is that the adviser has to give the excessive fees back. So why should the adviser not at least try to charge excessive fees?

Yet another problem is that when a recovery is finally paid, section 36(b) pays it to the wrong investors. By statute, an adviser who loses a section 36(b) suit must pay the recovery to the *fund,*

---

13 Investment Company Act § 36(b)(3).
rather than to the *investors*. But by the time the adviser pays the recovery to the fund, many of the investors who paid the fees will already have redeemed and left the fund, leaving behind new investors who did not pay the fees. Since fees accrue and are taken out of a fund on a daily basis, only an investor who holds shares at the moment a fee is charged actually ends up paying the fee. An investor who invests later cannot be said to have paid the fee. Hence, when a manager pays a 36(b) recovery directly to a fund, the payment benefits only investors who hold shares at the moment the recovery is paid—not the investors who held shares at the time the fees were paid. Put bluntly, the recoveries go to the wrong people. This problem was almost certainly not foreseen or understood by the statute’s drafters, because the problem results from the quirky way in which mutual funds are bought and sold. In an operating company, the mismatch in share ownership between the time of an alleged wrong and the time of a resulting recovery poses no problems, because the expected value of a recovery is baked into the share price. An outgoing shareholder can share in a recovery even if she sells before the recovery is received, because she will be able to sell at a price that reflects the recovery’s expected value. This neat solution disappears in a mutual fund, however, because a mutual fund’s share price does not reflect expectations about events that will happen in the future. The fund’s share price is mechanically tied to net asset value, and net asset value does not include the expected value of future litigation recoveries. Section 36(b) should therefore be amended to give recoveries not to a fund, but to the investors who held shares at the time an excessive fee was paid. Other class action recoveries in mutual funds are already paid this way.

A further mechanical flaw is the statute’s awkward treatment of sales loads. Section 36(b) exempts sales loads from liability, apparently on the logic that sales loads are regulated by FINRA. This distorts the analysis of a fund’s annual management fees, however, since an annual management fee is impossible to understand without first considering the sales load that

14 Section 36(b) requires this by providing that a suit may be brought by an investor “on behalf of” a fund, rather than on behalf of investors. Investment Company Act § 36(b).
15 In the worst-case scenario, a savvy investor could steal much of a fund’s recovery by investing one day before the recovery is received and redeeming one day after. Since the stock price the day before the recovery is received will not go up in expectation of the recovery, the investor can net a large chunk of the recovery, by just redeeming at the higher value after the recovery has been received.
16 Section 36(b)(4) (“This subsection shall not apply...to sales loads for the acquisition of any security issued by a registered investment company.”)
17 At the same time Congress added Section 36(b) to the ICA, it also added section 22(b)(1), empowering FINRA to regulate sales loads. Investment Company Amendments Act of 1970, Pub. L. No. 91-547, §§ 12, 22, 84 Stat. 1413, 1422-23.
preceded it. An annual fee that was preceded by a massive sales load is more likely to be unreasonable than an annual fee that was preceded by a small sales load. The exclusion of sales loads is also inconsistent with the inclusion of Rule 12b-1 fees. Rule 12b-1 fees are widely understood to be substitutes for sales loads, since, like sales loads, they can be used to pay commissions to brokers and other distribution expenses. Unlike sales loads, however, Rule 12b-1 fees are not exempt from section 36(b). The only way to explain this inconsistency is that Rule 12b-1 did not exist in 1970, when Congress added section 36(b) to the ICA, and Congress has simply failed to update the statute in the years since.

Beyond tweaking disclosure and fine-tuning section 36(b), a final area of research for fee regulation might be to consider some grander reform. It may well be that no amount of mandatory disclosure or fiduciary fee liability can adequately stamp out high fees. And so perhaps researchers and the SEC ought to imagine some more ambitious scheme of price regulation. Perhaps, for instance, the law could impose “smart” fee caps that adjust as a function of the fees of similar funds (e.g., any fee more than two standard deviations for funds in the same style category could be inherently suspect). Or perhaps extremely aggressive disclosures, akin to the surgeon general’s warnings on cigarette packs, could be appended to the prospectuses of unusually high-fee funds. In any case, if one seriously believes fees to be a problem, there may be more imaginative ways of limiting them than the law has imagined so far.

c) Capital Structure

Beyond tweaking disclosure Another area of interest is the regulation of investment company capital structure. Roughly speaking, the ICA regulates capital structure along three dimensions: (1) the sources; (2) form; and (3) amount (Morley, 2013). The ICA regulates sources of capital by prohibiting open-end mutual funds from taking on debt obligations to anyone other than banks.\(^\text{18}\) The ICA also regulates the form of debt capital by specifying that the only kind of security an open-end fund can issue is a single layer of common stock.\(^\text{19}\) The ICA further regulates the amount of debt capital by saying that in an open-end mutual fund, the ratio of total

\(^{\text{18}}\) ICA § 18(f)(1).
\(^{\text{19}}\) ICA § 18(a).
assets to bank-loan principal must always equal or exceed 3/1.\textsuperscript{20} The statute has separate rules for borrowing by closed-end funds.\textsuperscript{21}

One of the more important consequences of these capital structure regulations is to prevent a mutual fund from ever issuing a debt security. Though few people tend even to notice, it is a striking fact of American capital markets that one cannot buy a bond in a mutual fund.

The first question to ask is whether these restrictions even make sense. Though the rules might seem at first to be basically sensible, it turns out to be very difficult to come up with a satisfying explanation why. We could imagine many plausible rationales for regulating mutual fund capital structure. But the real-life details of mutual fund regulation are not actually consistent with any of them. The ICA is internally inconsistent. Morley (2013) explores the various rationales for capital structure regulation and their limitations.

One plausible rationale for the capital structure restrictions is that they guard against systemic risk. But this argument does not tell us why the restrictions apply only to publicly registered funds, such as mutual funds, and not to private funds, such as hedge funds. If borrowing generates systemic risk, then it does so just as surely in a hedge fund as in a mutual fund. The world already saw the evidence of this in the debt-driven collapse of the hedge fund Long Term Capital Management (Lowenstein, 2000). So why exempt hedge funds from borrowing restrictions? The answer the law gives is that section 3 of the ICA, which says that an investment company can choose to stay private so long as it has only a small number of wealthy and sophisticated investors. But the number and wealth of investors are only important if our concern is the welfare of the investors, not if our concern is the welfare of the financial system. A fund with a small number of wealthy investors can spill just as much risk into the financial system as a fund with a large number of middle class investors. Systemic risk is often compared to environmental pollution in its tendency to spill out market externalities. And what we have in ICA capital regulation, essentially, is a rule that says only rich people can dump sludge into rivers.

Another possible rationale for the capital structure regulations is to protect mutual fund investors from excessive risk in their portfolios. The ICA diminishes the riskiness of mutual fund investments for mom and pop investors by reducing the risks of leverage. But this, too, is

\textsuperscript{20} ICA §18(f)(1).
\textsuperscript{21} E.g., ICA §§ 18(a)(1)(A); (a)(2)(A).
inconsistent with the overall design of the ICA, because the ICA permits every other imaginable form of portfolio risk. The ICA places essentially no limits on the riskiness of portfolio assets other than the limits on borrowing. A fund can put all of its assets in lottery tickets and penny stocks if it so wishes. It can bet everything on out-of-the-money call options on the common stock of a single bankrupt issuer. In the grander scheme of investment management regulation, therefore, the distinction between leverage and the myriad other sources of investment risk is arbitrary.

Yet another potential rationale for the capital structure rules is that open-end mutual funds are mechanically incapable of issuing debt. Perhaps, one might argue, the redeemability of open-end fund shares prevents an open-end fund from issuing senior securities for reasons having to do with accounting difficulties or other technical problems. But this is plainly not true. Several open-end mutual funds issued debt securities before the ICA was passed in 1940 and industry leaders at the time saw no problem with their doing so (Morley, 2013). And even today, mutual funds still maintain debt obligations other than securities, including loans from banks. Hedge funds, which are also open-ended, incur debt obligations in elaborate variety.

Given the weakness of the various rationales we might imagine to support the mutual fund capital structure rules, it may be worth exploring whether the rules should be changed, perhaps even by permitting open-end mutual funds to issue debt securities. We know that people already buy debt securities from closed-end funds and operating companies, and we also know that people buy common stock in money market funds in large part because they believe it to be similar to debt securities. Why then should we not permit mutual funds to experiment with a single class of a simple kind of debt security? Though mutual funds do not often borrow money from banks (Warburton, 2017), they might see in debt securities an opportunity for profits through underwriting and sales fees.

Permitting mutual funds to issue debt might offer the possibility of a less fragile alternative to money market funds. Part of what makes a money market fund so fragile is that, like every other type of open-end mutual fund, a money market fund can issue only a single kind of security under the ICA. And since money market funds try to structure their common stock to functionally resemble debt, the capital structure restrictions in the ICA have the perverse effect of forcing money market funds to become effectively 100% leveraged—the only security they can issue is a common stock that functionally resembles debt. This strategy of total leverage has
no parallel in any other financial institution (or operating business, for that matter), and it naturally makes money market funds extremely fragile. A business with no real equity is a house made of very thin glass. If the ICA were reformed to permit a money market fund to adopt a more rational capital structure that combined a layer of debt with a layer equity, money market funds might become more stable, since the equity could cushion the debt.

Another problem of practical importance is the regulation of leveraged derivative instruments, for which the SEC has recently proposed new rules (whose status in the early days of the Trump administration is now uncertain). It is unclear what exactly the derivatives rules hope to accomplish. If the answer is that derivatives pose special risks to investors, then this is hard to square with the ICA’s permissiveness toward other risky investments. Remember that under the ICA, a mutual fund can invest in basically anything, no matter the risk. So why obsess over the risk of derivatives? And if the reason for regulating derivatives is that they pose risks to the financial system, then this, too, is hard to square with the existing regulatory scheme, since other open-end funds—namely hedge funds—can still use derivatives indiscriminately. Mutual funds have managed to use their freedom to invest in risky securities remarkably responsibly over the past 90 years. They have had very few bankruptcies or liquidity crises, even during the once-in-a-generation havoc of the financial crisis of 2008. It is hard, then, to see why mutual funds should not be trusted to invest in derivatives with the same sober responsibility with which they have long invested in other risky assets—and with which hedge funds are freely permitted to invest.

Of course, more research on the riskiness of derivatives in mutual funds would be helpful. There is some useful research already (Koski & Pontiff, 1999), but the challenge, as in so much of economic research, is to develop a study that avoids the pervasive problem of endogeneity.

d) Voting

Of course, another area of concern is shareholder voting. Prior to Congress’ passage of the ICA in 1940, shareholder voting was common in closed-end funds, but not in open-end funds (Morley & Curtis, 2010). Open-end funds generally offered their

---

shareholders no right to vote on any matter, including the election of directors. The ICA then imposed a system of mandatory shareholder voting in director elections and other matters for both types of funds.\textsuperscript{23} Our view is that although the shareholder voting requirements continue to make sense for closed-end funds, they should be eliminated for open-end funds.

The case for eliminating shareholder voting in open-end funds has been made by Morley and Curtis (2010), who argue that the problem with the right to vote in an open-end fund is simply that no investor will ever use it. Drawing on the analytical framework of exit, voice and loyalty proposed by Albert Hirschman (1970), Morley and Curtis argue that since mutual fund investors have an unusually strong form of exit, they will almost never tend to use their right to voice. Indeed, they are \textit{even less interested} in voice than the highly passive shareholders of ordinary public companies are. A similar point has also been made by Fama and Jensen (1983).

Open-end shareholders’ disinterest in voice stems from the fact that an investor in an open-end fund has a much stronger right of exit than an investor in an ordinary operating company. Unlike an operating company investor, a mutual fund investor does not sell her shares—she redeems them from the issuing fund for cash. When a shareholder in an open-end fund redeems, the fund pays the shareholder a cash amount equal to the net asset value, or NAV, which is the value of the portion of the fund’s assets that corresponds to each share. The upshot is that an open-end fund shareholder can basically pull her money out of the fund. She can demand that the fund give her back what she originally contributed, with the result that the fund will no longer own it. This stands an open-end fund shareholder in contrast an operating company shareholder. An ordinary company shareholder can sell her shares, but she cannot remove the assets that underlie the shares from the company’s possession and control. When a shareholder in General Electric sells to another shareholder, the shares may change hands, but the factories, cash and brand names all remain locked inside the company. General Electric pays out nothing.

Morley and Curtis (2010) walk through a technical analysis that shows why the redemption in open-end funds discourages shareholders from voting. We need not work through this technical analysis here, however, because we can see the basic intuition by simply thinking about how a mutual fund shareholder might choose between exit and voice when the shareholder believes her fund to be in decline. Imagine that an open-end fund shareholder decides that her fund’s management is bad and ought to be fired. The shareholder has two options. She can either

\textsuperscript{23} Investment Company Act § 16.
(1) run a proxy contest and rally other shareholders to vote the managers out, or (2) fire the managers on her own by just redeeming her shares unilaterally. The two options are substitutes for one another, because the ultimate result is similar—either way, the current managers will no longer have control of the shareholder’s money. Because redeeming removes the shareholder’s money from the fund—the fund has to liquidate the portion of its assets that corresponds to the shares and pay out the cash—redemption will terminate the managers’ control over the shareholder’s assets just as surely as firing the managers.

Given that voting and redeeming tend to produce very similar outcomes, which option will a shareholder tend to choose? The answer is that the shareholder will almost always choose redemption, because it is much, much cheaper. As the last eighty years of corporate law scholarship have taught us, firing a manager is costly, because voting is costly. Winning a proxy contest requires a company’s shareholders to act collectively through voting, and rallying shareholders to collective action takes huge amounts of time and money. Proxy forms have to be prepared and mailed and minds have to be persuaded. Few shareholders are willing to spend all of this money, not only because the amount of money necessary is so large, but also because any shareholder who does incur all of these costs will end up reaping only the small portion of the gains that corresponds to her shares. She might foot the entire bill for a proxy contest but own only 10 percent of the fund’s shares.

A mutual fund shareholder will thus rarely choose to pay the costs of activism, because redemption will offer a much cheaper alternative. As we have seen, redemption achieves the same basic result of removing a shareholder’s money from a manager’s control, but unlike voting, redemption does not require collective action. Redemption is a choice each investor can make unilaterally, without any need to coordinate with others. There is no collective action problem in a mutual fund, because redemption requires no collective action. As a result, proxy contests in mutual funds are virtually unheard-of. To our knowledge, no director election has ever been contested by shareholders in the 90-year history of the open-end mutual fund industry.24

The availability of redemption thus stands a mutual fund investor in stark contrast to an ordinary company. Ordinary company investors rarely use their right to vote, because of the

---

24 The handful of proxy contests that in open-end funds in the last ninety years have all originated in internal struggles inside of management firms or between boards and management firms. Shareholders have never initiated these contests and have never been materially involved in waging them.
costs we have already seen. But unlike in mutual funds, in ordinary companies, shareholder activism is not unheard-of; it is merely unusual. Sometimes when the circumstances are right in an ordinary company, the vote becomes a serious tool for change. Every now and again, an activist hedge fund or other investor will spend the costs necessary to run a proxy contest and force a change. And the reason is that there is no alternative. Even though collective action is costly, there is no option for unilateral action. An activist hedge fund that wants to get rid of the management cannot withdraw its money from the management’s control by redeeming. It must get control of the board. And since the costs of bad management sometimes exceed even the high costs of collective action, shareholders in ordinary companies occasionally pay the costs of collective action and use the shareholder vote.

One might argue that redemption is not obviously better than voting, because redemption is costly. Redemption requires time, knowledge, and sometimes the premature realization of tax liabilities. All of this is surely true, but this does make the vote appealing in a mutual fund, because the judgment here is comparative, rather than absolute. Redemption may be costly, but voting is even costlier. If getting on the web and opening an account with Vanguard sounds costly, imagine paying Kirkland & Ellis to run a proxy campaign. A shareholder may fail to redeem because the costs of redemption are too high. But that shareholder will also fail to vote for the very same reasons.

Mutual fund shareholders thus exhibit an extreme of passivity that exceeds even the much-discussed passivity of ordinary public company investors. In an ordinary public company, small and unsophisticated investors tend to find it rational not to vote. But in a mutual fund, even the large and sophisticated investors will fail to vote, because they will always find it easier instead to redeem. This is why proxy contests in mutual funds are not just rare, as they are in ordinary companies, but completely unheard-of.

The experience of other types of funds illustrates how redemption undercuts incentives to vote. Compare an open-end mutual fund to a closed-end fund. A closed-end fund is subject to the

25 Taxes offer pose a more interesting problem, but as Morley and Curtis show, taxes affect so few people that when an investor who cannot redeem for tax reasons tries to lead a proxy contest, most of the other investors who might have supported her will already have redeemed and left before the vote is held. At any given time, a mutual fund’s shareholder base tends to include only people who are (1) satisfied, (2) apathetic, or (3) locked in for tax reasons. Everyone who was unsatisfied, attentive and free to move will already have redeemed. Thus, unless the investors who are locked in by taxes comprise a majority all on their own, they will never be able to cobble together a critical mass for change from among the other investors, because all of the other investors will either be satisfied or apathetic.
same governance and voting requirements under the ICA as an open-end fund, and it tends to attract the same kinds of small investors (Hanley, Lee & Seguin, 1996). But unlike open-end funds, closed-end funds are beehives of shareholder voting activity (Bradley, Brav, Goldstein & Jiang, 2010). There is so much shareholder activism in closed-end funds that published newsletters chronicle the activism in daily detail. The incredible vivacity of shareholder governance in closed-end funds is directly attributable to the closed-end funds’ refusal to permit redemptions. Because closed-end funds do not permit redemptions, their shares trade on stock exchanges at prices that reflect expectations about how managers will use their assets in the future. These prices naturally diverge from NAV and it is precisely this divergence that motivates shareholder activism.\textsuperscript{26}

Further support comes from the experience of hedge funds. Like mutual funds, hedge funds also permit redemptions. And so, like mutual funds, hedge funds also experience no significant shareholder voting. A standard hedge fund operating agreement prohibits shareholder voting altogether. This is not because hedge fund investors tend to be sophisticated and few in number, because if anything, sophistication and small numbers ought to render the vote more valuable in hedge funds, not less. Sophistication and small numbers ought to diminish the collective action problem that plagues shareholder voting in public companies, seemingly making shareholder voting even more useful in hedge funds than in ordinary companies. And yet, the only conditions under which hedge funds commonly permit shareholder voting are when redemptions have been suspended, presumably because when redemptions are available, voting is just not very useful.

One might argue that we should not eliminate voting, because some small investors might lack the sophistication or awareness to redeem their shares efficiently. But we have already seen the answer to this objection above: though small shareholders may not redeem, they also will not vote for the very same reasons. Voting presupposes even more sophistication than redemption. And so if one believes that small shareholders need protection, voting is a terrible way to provide it. The suggestion that we eliminate voting is thus not an argument that we should stop regulating mutual funds. Indeed, it is precisely the opposite. The implication of this understanding of voting is that regulation is quite urgent, precisely because we cannot step back and let investors protect themselves by voting.

\textsuperscript{26} The idea is to buy at a price below NAV and then later force the fund to liquidate at a value close to NAV.
The reason voting made its way into the ICA initially is because at the time of the ICA’s adoption in 1940, the investment fund industry was dominated by closed-end funds, rather than open-end funds. In the late 1930s, when the ICA was being drafted, there were only a handful of open-end funds in existence, and the experience of the industry had been dominated by scandals among closed-end funds in the early 1930s (Grow, 1977). And, of course, in these closed-end funds, voting made sense, because closed-end funds did not offer redemption rights. It should therefore come as no surprise that when the SEC proposed the voting requirements in 1940, it had its eyes mainly on the closed-end funds. And while the closed-end funds supported the requirements, the open-end funds opposed them (Morley, 2012). Indeed, the only reason the open-end funds agreed to drop their opposition was that the SEC promised to include a grandfather provision in the statute that exempted all open-end funds then in existence out of the voting rules (Morley & Curtis, 2010). The grandfather exemption still appears in the ICA today.27

The voting requirements in open-end funds generate significant costs. One cost is the tremendous expense of proxy solicitation. Because shareholders in open-end funds are so extraordinarily apathetic, getting enough shareholders to return their proxies to form a quorum is very challenging. This is why the open-end mutual fund industry lobbied so hard for an exemption from the new NYSE rule prohibiting broker voting in uncontested director elections.28 Shareholder voting also often does more harm to shareholders than good. The ICA says, for example, that a board cannot fire a fund’s adviser unless the shareholders approve. But this greatly weakens a board’s leverage over the adviser by dramatically increasing the costs of firing the adviser. Imagine if the board of a company like Microsoft had to get a shareholder vote every time it wanted to fire the CEO.29 A further cost of shareholder voting is that it provides an unhelpful distraction. The belief that shareholder voting meaningfully protects shareholders in mutual funds may lull Congress and the SEC into a false sense of security that shields the industry against more meaningful regulation.

27 Investment Company Act § 16(c).
29 The effect is even worse in a mutual fund than in Microsoft, since any shareholder who is unhappy with the management of a mutual fund will tend to redeem before the vote takes place, depriving the fund of a critical mass of investors who would vote for change.
Much of shareholder voting has already been eliminated in mutual funds by clever lawyering and industry lobbying. Directors, for instance, tend to be able to appoint most of their own successors and rarely have to stand for annual elections. The little shareholder voting that remains is thus unhelpful, confusing and unnecessary.

**e) Directors**

Much of shareholder voting Mutual fund shareholders’ extraordinary disinterest in shareholder voting raises deep questions about another aspect of mutual fund governance: boards of directors. Given that shareholders never meaningfully participate in mutual fund director elections, is it appropriate to think of mutual fund directors as shareholder representatives? If it is not appropriate to think of directors as shareholder representatives, then where do directors derivate their legitimacy? Why should they have the authority to make important decisions? Given their extreme distance from any meaningful chance of shareholder voting, mutual fund directors are perhaps best regarded as being similar to the directors of autonomous nonprofits. Many commercial nonprofits, such as hospitals, operate without any system of member voting, and their directors appoint successors in perpetuity without election by constituents. This is essentially how mutual fund boards operate. One sensible reform might thus be to openly acknowledge the reality of mutual fund board autonomy by eliminating the fiction of shareholder elections, just as nonprofit hospitals do.

Another reasonable reform might be to eliminate boards of directors altogether. Boards have come under serious attack by a number of legal scholars (Fisch, 2010; Krug, 2013; Langevoort, 2005; Morley & Curtis, 2010; Ribstein, 2010). And there is ample precedent for open-end funds operating without boards. Open-end funds generally did not have boards of directors before the ICA required them, and open-end funds that are not regulated by the ICA, such as hedge funds and certain categories of exchange-traded funds, often operate without boards even now.

To be sure, the case for eliminating boards is not nearly as strong as the case for eliminating shareholder voting, because boards of directors arguably serve some useful functions. Still, even these functions deserve serious scrutiny, because they may well be better served in other ways. For example, the ICA and its administrative rules give directors a major compliance function by requiring them to make certain technical decisions and to oversee a fund’s adherence to certain
aspects of regulation. This compliance role might be filled more effectively, however, by
dedicated compliance professionals. The SEC already implicitly acknowledged as much after the
market-timing and late-trading scandals of the early 2000’s when it began requiring all funds and
advisers to hire Chief Compliance Officers. These CCOs may well be more competent, more
focused, and more diligent than directors in handling the complex details of compliance, because
directors usually serve only part-time and often lack the expertise and inclination to grapple with
the details of accounting and portfolio valuation that make up the bulk of compliance
monitoring. Perhaps the SEC could further empower CCO’s by letting them function
independently, without the ostensible oversight of a board of directors that has no real claim to
being a shareholder representative. This is how hedge funds operate now, and it works
reasonably well. Further research into the role and effectiveness of CCOs in both hedge funds
and mutual funds would be tremendously useful.

Another function boards arguably serve is the negotiation of fees. Section 15(c) of the ICA,
which was added by Congress in 1970, requires a board to re-approve a management contract
every year. Ostensibly, this process of re-approval offers a board a chance to renegotiate and
reduce management fees. In practice, however, the 15(c) process probably offers little benefit.
Mutual fund boards almost never use the 15(c) process to fire their managers, in part because the
shareholder voting requirement would make it almost impossible to do. As a consequence, a
board’s bargaining leverage with its managers in the 15(c) process is extremely limited.
Additionally, there is a distressing ambiguity about whose interests the 15(c) process actually
serves. Perhaps the main output of the 15(c) process is a written record that an adviser can later
use in litigation under section 36(b) (Knickle, 2011). The Gartenberg/Jones standard that
governs excessive fee liability under section 36(b) says that one factor in determining a fee’s
excessiveness is whether the fee was established by a strong bargaining process between a fund’s
board and adviser. A board that extensively documents its bargaining with its adviser thus
provides the adviser with evidence that the adviser can later use in litigation against
shareholders. Ironically, therefore, the main effect of a board’s diligence in the 15(c) process is
to cut off the rights of the very shareholders the board is supposed to protect.

30 Compliance Programs of Investment Companies and Investment Advisers, Release No. IC-26299, (Feb.
5, 2004).
Mutual funds could easily live without the 15(c) process. Section 15(c) was not part of the ICA until 1970, and the 15(c) process has no analogue today in hedge funds, even though hedge fund investors’ size and sophistication ought to make board governance even more useful for them than for mutual fund investors. The 15(c) process also has no parallel in markets for other products and services that are also characterized by strong exit. Mutual fund advisers are much like other professionals, such as lawyers and accountants, and in the markets for the services of these other professionals, consumers tend to be protected by direct regulation of ethics, prices, and quality, rather than by boards of directors. No one would ever argue that the customers of H&R Block should be represented in price negotiations by a board of directors whom they elected, even though the customers of H&R Block are basically the same demographic we worry about in mutual funds.

f) Performance Compensation

Much of shareholder voting The time is perhaps also ripe to reconsider the restrictions on performance compensation. In the nearly eighty years since the ICA and IAA were adopted, the popularity of performance compensation has exploded. Performance compensation is now ubiquitous in all manner of contracts, including executive employment contracts and private investment funds. And yet performance compensation remains very difficult for registered investment companies.

The ICA and IAA impose two sets of restrictions. One includes provisions added to the IAA by Congress in 1970, which say an adviser can only charge a performance-based fee if (1) the performance is measured relative to a benchmark and (b) the fee is symmetric in the sense that the adviser loses from underperforming the benchmark as much as it gains from overperforming the benchmark.31 These restrictions have received significant attention from researchers (Drago, Lazzari & Navone, 2010; Elton, Gruber & Blake, 2003; Das & Sundaram, 2002).

The second restriction is less well-known. Section 22(g) of the ICA, which Congress passed as part of the original statute in 1940, prohibits a mutual fund from issuing shares as compensation for services. The practical effect is to prevent a fund from paying the adviser with

---

31 Investment Advisers Act § 205(b)(2).
shares of the fund. For reasons that are unclear, this restriction applies only to open-end funds and not to closed-end funds.

Both of these restrictions ought to be re-examined. The main reason is the frequency with which private funds adopt performance-fee schemes that would violate these restrictions. Private equity and hedge funds are not subject to the restrictions in the IAA, and so they almost always pay fees on a performance basis and often pay them in the form of equity interests equivalent to shares. If the sophisticated investors in private tend to prefer performance-based compensation, then perhaps the unsophisticated investors in mutual funds should, too. The example of private funds is especially forceful, because with several decades now separating us from Congress’ adoption of the performance compensation restrictions, it is hard to see what the original logic of the restrictions was.

Another reason to doubt the value of these restrictions is that they can be avoided fairly easily. The downside risk that a symmetric performance fee so assiduously imposes can mostly be hedged away with derivatives. And the restriction on paying shares as compensation can be gotten around by just paying a manager in cash and then requiring the manager to spend the cash on shares.32

Of course, there may be reasons to think that asset-based fees are more efficient in mutual funds than performance fees. This is largely what the research noted above tries to explore. But showing why a performance fee might be unwise or inefficient is not the same as showing why it ought to be illegal. And there is little high-quality empirical evidence on the efficiency of performance fees, because of the profound and unavoidable problem of endogeneity and self-selection. In any case, further attempts at empirical evidence and theoretical modeling would be most helpful.

g) Inter-Client Conflicts

Much of shareholder voting As noted above, one of the features that distinguishes an investment fund from an operating company is the investment fund’s tendency to employ an external manager. Instead of hiring a CEO directly, an investment fund receives its management

32 This practice appears to be fairly common for fund directors and individual portfolio managers (Cremers, Driessen, Maenhout, Weinbaum, 2009).
from a different company with a different set of owners. One consequence of this tendency toward external management is that a manager can have other clients in addition to a particular fund. In a large advisory complex, a manager might have hundreds or even thousands of different clients, ranging from hedge funds, private equity funds, mutual funds and closed-end funds to myriad un-pooled accounts for individual investors. The trouble with this managerial promiscuity is that all of these many clients can come into conflict with one another. Since a manager has a separate fiduciary duty to each client, favoring one client over another in the allocation of any resource raises thorny problems of fiduciary duty. Already, much has been written about these conflicts.\(^3^3\)

At first, each of these conflicts can seem deeply alarming. How could a mutual fund be getting a fair deal when, for example, its adviser also operates a hedge fund that pays a performance fee that gives the adviser greater incentive to favor the hedge with investment opportunities and other resources? On further examination, though, these conflicts seem less worrying. Though it is tempting to pluck out one conflict at a time and express alarm about its risks, the truth is that conflicts are pervasive. Virtually everything an investment manager does raises a conflict among its clients. The most obvious conflicts involve investment opportunities, but clients can collide in myriad other ways as well. Every time a manager assigns an employee to serve one client, decides the order in which to execute trades, or chooses whether to make a purchase for one client that would trigger 13D filing obligations for other clients, the manager is facing a conflict of interest. Even the allocation of computer equipment and office space involves a conflict among clients, since a computer allocated to one client is a computer not allocated to another client. Conflicts are everywhere, and there is no escaping them.

Rather than picking out one conflict at a time and wringing our hands about it, therefore, we ought to think more holistically about why investors so willingly permit all of these conflicts in the first place. It cannot be the case that all inter-client conflicts are bad, because even

---

33 See Bhattacharya, Lee and Pool (2013) (conflicts arising when a manager uses a “fund of funds” to invest in and subsidize the family’s other funds); Birdthistle and Henderson (2009) (conflicts arising when a manager’s different funds or proprietary trading strategies invest at different levels of the same portfolio company’s capital structure); Davis and Kim (2007) (conflicts arising when mutual fund managers seek pension management business from operating companies in which the fund managers’ mutual funds invest); Gaspar, Massa & Matos (2006) (conflicts arising from strategic shifting of performance between various funds); Nohel, Wang & Zheng (2010) (conflicts arising when mutual fund management companies also operate hedge funds); Cici, Gibson & Moussawi (2006) (same).
sophisticated institutional clients routinely permit them. And so there must be some deeper logic that renders all of these conflicts acceptable.

To date, however, we lack a fully fleshed-out theory to tell us what that logic might be. Morley (2014) offers a preliminary account by identifying some reasons why inter-client conflicts might be more efficient in investment advisory complexes than ordinary companies. Morley argues that elements of fund organization, including exit rights and the limits on fund investors’ control over managers, render inter-client conflicts less problematic in investment funds than ordinary companies. Still, Morley’s account is short on details, and researchers might make major progress by trying to model them.

The absence of a clear theory of inter-client conflicts has often rendered regulation incoherent. The ICA and IAA and their administrative rules tend to grab onto particular kinds of conflicts without a discernible rationale for why some of them are worse than others, producing confusion and unfairness. Section 17 of the ICA, for example, prohibits a manager from investing jointly in a transaction alongside a registered fund, but permits an unlimited number of the manager’s other clients to do so. And in 2004, the SEC began requiring a fund to disclose the number of other clients the fund’s portfolio managers work for, as well as whether the other clients paid the manager performance fees. But the SEC did not require a fund to disclose whether the other clients had the same investment objectives as the fund in question, even though this is critical to knowing whether there could be a conflict.

One especially valuable place to look for guidance might be the inter-client conflict restrictions that prevail by contract in private funds. Restrictions on inter-client conflicts vary significantly between hedge funds on the one hand and private equity funds on the other (Morley 2014), but among each type of fund the restrictions tend to be highly standardized. Hedge funds tend to be extremely permissive towards inter-client conflicts. Their operating agreements almost always waive the corporate opportunity doctrine, technically permitting an adviser and its other clients to invest in anything without regard to whether something might be an investment opportunities of the fund. Private equity funds, by contrast, are much more restrictive. They typically prohibit a manager from actively investing the assets of more than one private equity fund in a particular investment objective at any given time. Private equity funds also adopt

elaborate restrictions on the ability of managers and their other clients to invest alongside the funds. The width and clarity of the divergences between private equity and hedge funds suggest that there may be a deep structural logic to the regulation of inter-client conflicts, and that it depends on many aspects of a fund’s structure and investing objectives that may or may not be present in mutual funds.

Conflicts among clients are likely to intensify in the future. The investment management industry is consolidating and growing more concentrated (Elhauge, 2016), which may mean that the biggest investment management complexes will have more clients and thus more conflicts among them. Additionally, the client bases of investment managers are fragmenting as more and more investors opt out of pooling arrangements and choose instead to have their money be managed individually (Clayton, 2017). More than ever, regulators will need a clear understanding of when to intervene and when to let the market find an equilibrium on its own.

**h) The Long Shadow of Closed-End Funds**

So far, we have focused almost entirely on open-end funds, since they hold the vast bulk of the registered fund industry’s assets. But closed-end funds are worth thinking about, too, in large part because they cast a long shadow over open-end funds. When the SEC drafted the ICA in the late 1930s, it focused overwhelmingly on closed-end funds. Until the very end of the 1930s, closed-end funds were vastly larger in both number and assets than open-end funds (Grow, 1977; Morley, 2012). Open-end funds were a niche business centered in a handful of small management firms in Boston, while closed-end funds were a big business centered in large investment banks in New York. Size was not the only reason the SEC obsessed over closed-end funds. Closed-end funds had also piled up a horrific record of abuse during the bull market of the late 1920s and the bust of the early 1930s, and it was the SEC’s years-long investigation into these misdeeds in the late 1930s that became the basis of the ICA.

Because the open-end fund industry was so small in the late 1930s, and because the experience with the closed-end fund industry was so much more salient, the SEC and other industry observers overwhelmingly saw the statute’s purposes as having to do with closed-end funds. (Fink, 2008; Morley, 2012). Today, the core of the regulatory scheme is identical for both types of funds, with a slant toward the needs of closed-end funds.
This is unfortunate, because the industry is now composed overwhelmingly of open-end funds. Indeed, open-end funds hold about 70 times more assets than closed-end funds (Investment Company Institute, 2016). The differences between closed- and open-end funds thus warrant serious scrutiny. The ICA was written decades before the work of Henry Hansmann, Oliver Hart, Albert Hirschman, Oliver Williamson and others demonstrated the importance of exit rights in contractual relationships, and so perhaps the drafters of the ICA might be forgiven for failing to understand the profound significance of redemption rights in open-end funds. But the time has now come for a sweeping reassessment. It is unclear how exactly regulation ought to differentiate between open- and closed-end funds, but researchers should set themselves the task of figuring it out.

One obvious area of interest is shareholder voting and board governance. As noted above, voting makes much less sense in an open-end fund than a closed-end fund. Another area of interest is the role of boards in defending closed-end fund managers against shareholder activism. In closed-end funds, boards tend to exercise very little supervisory oversight, since the funds are dominated by their external advisers. But boards nevertheless play a major role in building and defending the anti-takeover fortifications that make it difficult for shareholder activists to hold fund advisers accountable. Is this a useful function for boards to serve? That is a hard question that deserves a serious answer.

Taxation also requires serious scrutiny. Though in this paper we have generally avoided taxation, the differences in taxation between open- and closed-end funds are too glaring to ignore. Tax law treats open- and closed-end funds identically, but it has radically different consequences for each. In order to avoid entity-level income taxation, tax law requires a RIC to distribute all of its income every year. This creates some very weird problems, which manifest in open- and closed-end funds in different (though equally weird) ways. In an open-end fund, taxation creates a problem of tax overhang, in which the failure of open-end fund share prices to adjust to expectations means that the expected value of tax liabilities is not reflected in share prices (Barclay, Pearson & Weisbach, 1998). Closed-end funds face a different problem. In a closed-end fund, the income distribution requirement gradually bleeds the fund of its assets by requiring the fund to pay out its income, even though a fund that trades at a discount at NAV can’t easily recover the distributed income by issuing new shares (Morley, 2012). The distribution requirement thus works as a kind of ratchet, forcing the fund to pay out all of its
income in good years even as it suffers losses in bad years. The differences between open-end and closed-end funds were never seriously considered when Congress firsr adopted the tax system in the late-1930s and early 1940s (Morley, 2012).

Another area of possible difference is disclosure. Disclosure works better in closed-end funds than in open-end funds, because closed-end fund shares tend to be traded on securities exchanges. Closed-end funds’ tendency to trade on exchanges means that their shares can be bought and sold at prices that reflect efficient market estimates of value even if only a small number of investors actually read and understand the fund’s disclosures.35 One can rely on the market price of an exchange-traded security as an accurate estimation of the security’s value, even if one did not personally read the fund’s annual report. This is the logic, for example, of the fraud-on-the-market theory of reliance in securities class actions. Open-end fund shares, by contrast, are not traded on exchanges, and their prices thus do not automatically adjust to efficient levels to reflect supply and demand conditions. This is a consequence of redemption rights. A fund’s NAV stays fixed regardless of how many people buy or sell shares on a given day and a fund’s fees only adjust when a board and adviser make a decision to change them, not when demand increases or declines. Unlike the shares of closed-end funds and other exchange-traded securities, there is no price quotation scheme by which mutual fund fees and NAV automatically update to reflect the opinions of sophisticated investors. This means that the efforts of a few investors to read and understand an open-end fund’s disclosures do not generate the same positive externalities in an open-end fund as they do in a closed-end fund. One cannot rely on an open-end fund’s share price as an accurate estimate of the fund’s expected value, because the price does not necessarily reflect sophisticated investors’ estimations of expected value. Every investor in an open-end fund has to form an opinion for himself. In this regard, an open-end fund is like a consumer product, which also requires individual consumers to form their own estimates of a product’s value.36 And just as with a consumer product, with a mutual fund, disclosure often demands too much sophistication from purchasers, rendering it deeply problematic as a tool for regulation.

35 Of course, some economists doubt whether closed-end fund share prices reflect the valuations of sophisticated investors (Lee, Shleifer & Thaler, 1991). But at least in theory, there is a possibility that a few sophisticated investors could move prices in an inefficient direction in a closed-end fund.

36 Of course, Schwartz (1988) argues that not every consumer has to comprehend a disclosure in order for the disclosure to produce efficient outcomes in a product market. But the percentage of buyers who have to comprehend the disclosure is no doubt much higher in a product market than a securities market.
In addition to exploring how regulation could better accommodate open-end funds, researchers should also explore whether regulation could better address the problems of closed-end funds. Though the puzzle of closed-end fund share pricing has endlessly fascinated financial economists, we know of no academic paper specifically devoted to the challenges of closed-end fund regulation. Research on the unique regulatory problems of closed-end funds is urgently necessary, because the nature of the closed-end fund business has changed since 1940. In 1940, everyone still expected that closed-end funds would continue to dominate the industry they had started and would one day return to trading at premiums to NAV—as they did before 1929. Now that closed-end funds overwhelmingly trade at discounts to NAV, however, there are deep questions about how closed-end funds should be regulated and about whether they should even be permitted to exist at all. Because closed-end funds commonly trade at discounts to NAV, an IPO of a closed-end fund is generally a terrible investment. An IPO investor buys at NAV, with a virtual guarantee that the stock will quickly trade below NAV. Evidence indicates that a distressing share of IPO investors tend to be individuals who are directed to purchase the closed-end fund IPOs by their brokers (Hanley, Lee & Seguin, 1996). Closed-end funds also have a tendency to coerce investors through the use of dilutive rights offerings, in which they sell shares to existing investors at prices below NAV. These rights offerings operate much like poison pills by forcing investors to buy or else be diluted. Closed-end funds may have many other regulatory challenges, and they present a deep vein for legal researchers to mine.

2. PRIVATE FUNDS

So far we have focused entirely on publicly registered investment funds. But the big story in investment management in the last several decades, of course, has been the rise of private funds, such as hedge funds, private equity funds, and venture capital funds. The important fact about these funds for regulation is that the drafters of the ICA never foresaw them (Grow, 1977; Morley, 2012). The prospect that a large industry of private funds might someday challenge the dominance of registered funds never entered the debates about the ICA. The dramatic rise of private funds thus raises fundamental questions about whether the law ought to change.

a.) The Public/Private Distinction
The natural starting point is the set of exemptions that permits private funds to stay private. The key dividing line between public and private appears in section 3(c) of the ICA, which, among other things, exempts a fund from registration under the ICA if the fund has fewer than 100 investors or if its investors are all large enough to satisfy the definition of a “qualified purchaser” in section 2(a)(51).37

The main question is whether the exemptions for private funds should be expanded. The reasons to consider expansion are several. First, there are tremendous tax advantages to remaining private. The taxation of registered investment companies in the United States is almost always less favorable than the taxation of private funds, for both investors and managers (Brunson, 2015; Brunson, 2013; Coates, 2009). Private funds are taxed as partnerships, even as registered investment companies are taxed under their own unique system. And given the choice between partnership and RIC taxation, almost no rational investor would ever choose RIC taxation. We therefore ought to consider expanding the range of funds that can qualify as private fund as a way of broadening access to the favorable tax treatment of private funds.

Another reason to doubt the line between public and private in the ICA is that it differs from the line between public and private in the other securities laws, and there is no obvious reason why. The ICA and the other securities laws employ similar concepts for distinguishing public and private funds, but they establish the threshold dollar and number amounts differently. Under the Exchange Act, for example, an individual can qualify as an “accredited investor” with just one million dollars in assets. But under the ICA, an individual can only become a “qualified purchaser” with at least five million dollars in assets. Similarly, the Exchange Act permits a company to remain private with as many 500 “non-accredited” investors; the ICA draws the line at just 100 “non-qualified purchasers.”38 We know of no effort, historical or modern, to justify the differences.

It is also unclear why the number of investors in a fund should matter if the fund offers redemption rights like an open-end fund. In an operating company, where investors are locked in and cannot redeem, the number of investors matters because it affects the collective action

---

37 ICA §§ 2(a)(51) (defining “Qualified Purchaser”); 3(c)(1), (7) (exempting funds with fewer than 100 investors and funds with only qualified purchasers).
38 Both an ICA and an operating company will become public under the exchange act if it exceeds 2,000 investors of any kind. Exchange Act § 12(g)(1).
problem. As the number of investors increases, so does the investors’ tendency to become rationally apathetic, heightening the need for regulation to protect the investors’ collective interests. In an open-end fund with redemption rights, however, the collective action problem does not matter, because there is no such thing as collective action. An open-end fund investor can withdraw her money unilaterally, and so, as observed above, open-end fund shareholders have very little reason ever to vote collectively. So then why should it matter whether an open-end fund has a hundred investors or a thousand?

Permitting more funds to stay private might also offer small investors the prospect of better investment returns. Many aspects of RIC regulation, such as the restrictions on leverage, the requirements for frequent and costly redemptions, and the prohibition on incentive fees arguably prevent RICs from achieving the same investment returns as private funds. Even if we do not have any faith in private fund managers’ ability to outperform the market by skill, we might nevertheless think they could outperform RICs simply by virtue of being free of the burdens of regulation. The drain some regulations place on investment returns creates a real tension between protecting public fund investors and impoverishing them.

b.) Inter-Client Conflicts of Interest

Like publicly registered funds, private funds also pose serious problems of inter-client conflicts of interest. When a manager simultaneously operates many different private funds and client accounts, the interests of all these funds and accounts inevitably clash. As observed above, the issue is growing more important because of the increasing fragmentation of pooled vehicles into unpooled individually managed accounts. The fragmentation is intensified in private funds by the growing popularity of side letters, which require a manager to give special treatment to one investor over others in a pooled fund. The most sustained academic treatment of this challenge comes from Clayton (2017), who describes the fragmentation and explores the conditions that might be required for it to be efficient. More research on the empirical extent and theoretical risks of inter-client conflicts would be valuable.

3. INVESTMENT FUNDS AS INVESTORS
We have focused so far mostly on investment funds as issuers of securities, but investment funds also play an important role as investors in securities. A great deal has been written about how well investment funds function as investors in other companies, but the most urgent academic research at the moment involves the recent discovery of astonishing statistics about the concentration of holdings in the hands of the largest investment management companies. Work by a number of authors has shown that the investment management industry has grown both much larger and much more concentrated in recent years, causing the biggest investment managers to control astoundingly large stakes in American public companies. (Anton, Ederer, Giné & Schmalz, 2016; Azar, 2016, Azar, Reina & Schmalz, 2016; Azar, Schmalz & Tecu, 2016; Fichtner, Heemskerk & Garcia-Bernardo, 2016).

Some of this research suggests that the largest managers’ stakes are so vast and so widespread that they might be tilting toward monopolization. In one paper, José Azar, Martin Schmalz, and Isabel Tecu (2016) show that as a handful of large investment managers came to control increasingly large stakes in every major airline, the intensity of price competition among the airlines decreased. Labeling this phenomenon “horizontal ownership,” a number of legal scholars have argued that although the declines in competition likely do not reflect direct collusion by investment managers, overlapping ownership by investment managers should nevertheless be regulated as an antitrust problem (Elhauge, 2016; Posner, Morton & Weyl, 2017).

The investment management industry, understandably, tends to believe there is nothing to worry about. And even many academics argue that there are yet reasons to remain skeptical of the academic evidence. Still, even if one feels inclined to accept the investment management industry’s skepticism, it is important to take the issue of horizontal ownership seriously. Though no one can tell the future, the potential impact of concentrated ownership is vast. It is now possible to foresee a day when the clients of two or three large investment managers might hold thirty percent of the shares of a majority of America’s large public companies. Though the intentions of investment managers may be entirely benign, the power that these growing stakes will bring is too significant to ignore.

39 José Azar et al., Anti-Competitive Effects of Common Ownership 1 (Ross School of Business, Paper No. 1235, 2016)
Researchers should thus continue to describe the concentration of ownership and seek out new ways to draw causal inferences about its consequences. Researchers should also continue exploring the antitrust aspects of the issue, as many authors have already done. And beyond the antitrust law, researchers should explore the implications of corporate and securities law. Morley (2017), for example, argues that the power of large investment managers is inherently limited, because conflicts among clients make it difficult for an adviser to convert holdings into direct control. Activism by one client inevitably generates costs and headaches for other clients, and this creates persistent and often unmanageable fiduciary conflicts. And some of these costs come from securities regulation, such as the way in which one client’s activism can increase the Form 13D filing obligations of other clients. The tendency of the costs of activism to spill across different clients may be one reason why activist hedge funds tend always to be managed by small advisors with only a handful of clients. Large advisers, such as Goldman Sachs and Fidelity, have never managed activist hedge funds, for example, and probably never will. These complications and many others warrant exploration as we try to understand what a future of growing ownership concentration will bring.

4. CREDIT RATING AGENCIES

A topic of extensive public debate, academic study, and market commentary has been the role played by credit rating agencies (CRAs) in the fixed income markets. Concerns over the quality of credit ratings first received widespread attention in 2001 when Enron bonds still had investment grade ratings from the three major CRAs (Moody’s, S&P and Fitch) just five days before Enron’s bankruptcy filing (White 2013). However, the far more powerful impetus for this interest in CRAs has been the credit ratings provided to structured finance securities (mortgage-backed securities (MBS), asset-backed securities (ABS), and collateralized debt obligations (CDO)) in the lead up to the financial crisis of 2007-2008 by the dominant CRAs (Moody’s, S&P’s, and to a lesser extent, Fitch) and the subsequent very poor performance of many of these securities during the crisis. For instance, some 95% of all credit rating downgrades over the 2007-2008 period were of MBS, ABS or CDOs. ABS CDOs constituted over 40% of the total write-downs of financial institutions during this period.
Importantly, these downgrades of structured finance securities and their poor performance included a substantial number of triple-A rated obligations. Below is a figure from Benmelech & Dlugosz (2009a) (figure 4a) reflecting the number of downgrades of structured finance securities and the percent that were triple-A:

![Total number of downgrades and number of AAA-structured finance securities downgrades](image)

The ratings on corporate debt by contrast have by comparison fared far better historically, despite the experience of the Enron bonds. As a general matter corporate debt ratings over the 1985-2009 period have become more conservative and in fact are arguably too conservative (in light of market pricing). Baghai, Servaes & Tamayo (2014). During the 2001-2002 recession, a period that was especially trying for corporate bonds with a wave of corporate bankruptcies occurring during this time, there were a significant number of corporate bonds downgraded at the time. However, virtually none were triple-A rated corporate debt. Moreover, for those corporate bonds that were downgraded, the number of downgrade notches (how far the bond was downgraded) during the 2001-2002 recession were far smaller than that experienced by structured finance securities during the financial crisis of 2007-2008. Benmelech & Dlugosz (2009a) conclude during this time that “corporate bond ratings were well calibrated to the underlying economic risk of the issuer.” As for the financial crisis, significantly fewer triple-A
corporate bonds were downgraded during the financial crisis relative to structured finance instruments.

Given all this, it is not surprising that public debate, academic study, and market commentary concerning the CRAs have tended to focus on ratings of structured finance securities, rather than corporate debt securities. Also not surprisingly, much of the empirical work (although certainly not all) has tended to focus on the ratings of structured finance securities and their performance during the financial crisis. The policy debate over CRAs has therefore generally focused on whether investors, and the market more generally, relied to their detriment on flawed credit ratings of structured finance instruments. In unpacking this general debate several issues present themselves.

First, to what extent did the market in fact rely on CRAs’ ratings, perhaps most importantly in pricing debt securities such as MBS, ABS, and CDOs? The answer to this question is of interest regardless of the ultimate answer. If the market did rely to some significant extent on ratings and, moreover, the ratings were somehow flawed this raises an obvious policy concern in terms of the market being misled and capital being misallocated. But even if the market did not actually rely on these ratings (or, equivalently, ignored these ratings when they are inconsistent with other evidence) this merely raises a different set of policy concerns: how well are the CRAs serving their role as information intermediaries if the market views the ratings as superfluous? And if CRAs are not playing a valuable information intermediary role for whatever reason then why should one use these ratings for a wide variety of regulatory purposes as was in fact the case historically (and to some significant extent still today)?

Second, assuming the market does rely to some significant on ratings, to what extent were the credit ratings flawed (measured from an ex ante perspective, rather than using the benefit of hindsight)? The focal point in the literature and public debate here has been on ratings accuracy. One additional issue of particular importance can be folded into this discussion and that is of systematic risk. To the extent that CRAs made the same (mistaken) judgments for a variety of structured finance securities does this create a source of systematic risk? In other words, does uniformity in ratings judgment turn a mistaken ratings judgment in a particular case into a systematic issue for the marketplace?

Third, if the market has relied on flawed ratings, what is the proper regulatory/policy response? Obviously first understanding what problems exist and why is the necessary predicate
for any potential regulatory changes. Before discussing (at a general level) these three issues, we will first briefly discuss at a basic level the function of CRAs in the marketplace and in the regulatory scheme.

a) Credit Rating Agencies: Function and Regulation

CRAs produce ratings that are designed to reflect credit risk of debt securities (default risk and/or expected loss). While credit risk is obviously an important source of risk for debt securities, it is not the only one (such as liquidity and duration risk). In providing information concerning credit risk, CRAs can serve a valuable information intermediary role for debt markets. Debt markets are both enormous and heterogeneous (Rhee 2015) which speaks to the potential economies of scale and scope enjoyed by such an information intermediary. Given the repeated game nature of debt ratings, the reputational capital of CRAs is one potentially important market mechanism that can help ensure ratings quality (see, e.g., Holthausen & Leftwich (1986)). Importantly, CRAs, such as Moody’s, existed long before the regulatory use of credit ratings indicative of a market-based role for these entities. During this time (up until the late 1960s/early 1970s), investors would purchase these ratings from CRAs. Presumably clients of CRAs wished to purchase ratings given their informational content, i.e. CRAs were valued information intermediaries.

But any consideration of the role of CRAs in the modern era must also reflect the fact that credit ratings are used extensively by regulators for a wide variety of purposes, i.e. the regulatory treatment of a particular debt issuance might well be a function of its rating. Partnoy (2006). In short, favorable ratings can help ensure favorable regulatory treatment. This effect is potentially quite separate and apart from these ratings providing valuable information to the debt markets concerning credit risk, even though favorable regulatory treatment might itself have real pricing effects. The regulatory issue of credit ratings had its origins in the 1930s when regulators first used them in regulating commercial banks (White 2013) but truly blossomed in the 1970s, roughly contemporaneous with the move from investors paying for ratings to those ratings being paid for by the debt issuers. Perhaps most importantly, in 1975, the regulatory category of “nationally recognized statistical rating organization” (NRSRO) was created with the credit
ratings of CRAs receiving the NRSRO designation being used for a variety of regulatory purposes.

The regulatory uses of credit ratings include the calculation of minimum capital requirements for a variety of financial institutions (insurance, banking and broker-dealers), pension fund asset allocations, and Basel II bank capital requirements. Also, money market mutual funds have been required by the Investment Company of 1940 to hold highly rated assets. Hunt (2009) identifies 44 SEC rules and forms that incorporate credit ratings. The regulatory use of ratings is not confined to federal law. State regulators also often use ratings for investment limits on regulated entities, such as insurance companies, as do international bodies as well.

Concern over CRAs and their regulatory use has triggered two legislative responses. In 2006 Congress passed the Credit Rating Agency Reform Act mandating transparency and ensuring procedural regularity in the SEC’s designation of NRSROs. Ultimately seven new CRAs received NRSRO designation, in addition to Moody’s, S&P and Fitch. The second legislative response is contained in the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 (Dodd-Frank) which mandated the cessation of regulatory use of credit ratings in federal law.

Not only have ratings been used for regulatory purposes but also contractual purposes in the marketplace. This is yet another role that can be played by an information intermediary: distilling information into clear signals of credit risk that market actors, if they so choose, can use in contracting. Consider three examples. First, the need for institutions to post additional collateral can result from changes in the ratings of the collateral. These ratings-based triggers for posting of additional collateral is an important issue in considering the role of CRAs during the financial crisis of 2007-2008. Second, asset managers often rely on ratings in managing their portfolios with ratings of portfolio securities often being explicitly incorporated into the investment management agreements that govern the construction of these portfolios. For instance, a survey of pension plan sponsors reported that 75% have minimum rating requirements and 50% limits on portfolio distribution by credit rating. Cantor, Gwilym & Thomas (2007). Third, ratings can play an important role in determining whether a debt covenant has been violated.

b) Do the Markets Rely?
Adelino (2009) finds that the yields on MBS issued in the lead-up to the financial crisis did in fact contain important information predicting future downgrades and defaults above and beyond that reflected in the credit rating. Interestingly, yields were more effective at predicting downgrades than credit ratings for MBS rated below AA, i.e. the market appears to rely on information beyond that reflected in the credit rating. This complements the finding of Ashcraft, Goldsmith-Pinkham and Vickrey (2010) who find that credit ratings did not reflect all available (negative) information concerning the risk of default.

Importantly, however, Adelino (2009) findings do not hold true for triple-A MBS tranches. Here, Adelino (2009) finds that the market did to a significant extent rely just on the credit rating. The market emphasis on triple-A ratings as a determinant of pricing is consistent with a number of other papers in the literature that argue that triple-A MBS investors were generally passive and/or less informed. These investors often had a preference to purchase securities believed to be information insensitive making triple-A rated securities a natural fit. See Pagano & Volpin (2010). On a related note, Coval, Jurek & Stafford (2009) document that the many senior (triple-A rated) CDO tranches were in fact more risky than that implied by either their credit rating or their pricing. The argument that triple-A investors to some significant degree were generally passive and/or less informed is also consistent with Ashcraft, Goldsmith-Pinkham and Vickrey (2010) who find that MBS ratings inaccuracy was most severe during the height of the market i.e. the time at which investors (or least a segment of investors) were arguably most likely to be passive and/or less informed. In other words, there is some evidence that the market relied on the triple-A ratings of structured finance instruments potentially at the expense of additional (negative) information relevant to pricing. The role that the extensive regulatory usage of ratings potentially plays in this connection is an interesting one: to what extent (if any) did the regulatory importance placed on the triple-A rated designation result in increased investor passivity? If regulations suggest that a triple-A designation is sufficient to establish that the security is information insensitive, to what extent does this substitute for investor due diligence?

In thinking about potential investor clientele effects of highly rated securities, it is worth noting that significant purchasers of non-agency triple-A MBS during the 2007-2008 period were overseas investors (25.2%), Fannie Mae, Freddie Mac and the Federal Home Loan Bank (18.8%), money market managers (13.8%), and insurance companies (7.6%) (Lehman 2008). Triple-A securities was also commonly used in the repo markets, arguably because the parties to
these transactions wanted to use an information insensitive asset for collateral (an asset for which
the parties to the repo transaction generally did not need to worry about the fundamental value so
as to minimize transaction costs). It is also worth noting that outside the context of the financial
crisis, it is more challenging to empirically study the informational content of credit ratings with
respect to triple-A securities given the very low level of downgrades and default for these
securities historically (combined with the relatively recent development of the structured finance
market). This raises the general issue of how generalizable are the results from the multiple
studies of one particular, albeit important, historical episode: the rating and performance of
structured finance instruments before and during the financial crisis.

Turning to corporate debt, there have been a long series of papers, starting with Katz (1974),
generally finding that corporate debt reacts negatively to negative credit rating events such as
downgrades, i.e. ratings are viewed as having informational content. See, e.g. Wansley,
Glascock & Clauretie (1992). Whether the market views the credit rating as having information
because it reveals new negative information concerning the firm’s prospects (“information
channel”) or, alternatively because of the regulatory and contractual effects of a change in credit
rating (“non-information channels”) is not always addressed in these papers. One interesting
paper that does is Kisgen & Strahan (2010) which finds a meaningful pricing effect on corporate
debt due to non-information channels. On a similar note, Sangiorgi & Spatt (2017) discuss the
fact that Moody’s recaliberation of its municipal bond ratings had real effects on municipal bond
pricing even though the recaliberation was not based on a new evaluation of information
concerning default/expected losses. In other words, municipal debt ratings appear to have pricing
effects outside the informational channel.

From a policy perspective it can be important to separate out the two potential pricing effects
of ratings: the information channel would be consistent with CRAs providing new value-relevant
information to the market filling a valuable information intermediary role to the debt markets,
whereas the non-informational channels would reflect a choice to have the credit rating trigger
certain real effects for the firm (or the municipality in the case of municipal bonds). If the
pricing effects are a function of the latter, then one could ask the question whether this choice,
particularly if it is a regulatory choice, is a wise one.

c) Competition and Revenues
One important set of policy issues concerns how the industrial organization of the CRA market impacts ratings quality along two related dimensions: competition (meaningful choice among CRAs) and the nature of that competition (competing for business from debt issuers).

Turning first to competition, while there have been a total of ten CRAs that are NRSROs, approximately 96% of all ratings are provided by just three CRAs: Moody’s Investor Services, Standard & Poor’s, and Fitch Rating (with Fitch Rating being significantly smaller than the other two). Providing ratings, along with ancillary services such as consulting, was quite profitable for these entities, with Moody’s profits tripling between 2002 and 2006. As of 2006, 44% of its profits came from structured finance.

Whether this apparent lack of choice among CRAs (given the largely duopoly structure of the CRA market) should raise policy concerns is a complicated policy question with no obvious answer: more competition could (i) result in greater concern by a CRA for its reputational capital as a source of competitive advantage (positive for ratings accuracy); (ii) reduce the value of a CRA’s reputation capital because the business is itself less profitable (negative for ratings accuracy); and (iii) create more opportunities for “rate shopping” by debt issuers (negative for ratings accuracy). Some of the empirical evidence on shopping will be reviewed later (in the section entitled Ratings Shopping). See also generally Sangiorgi & Spatt (2017) on this topic.

Even assuming the relative lack of CRA choice reduces ratings accuracy, this hardly ends the analysis. At least two further difficult questions present themselves: First, what accounts for the concentrated nature of the industry and, second, assuming a concentrated CRA industry how can (if at all) regulation improve the situation? The answer to the first question is nonobvious. Is the relative lack of competition due to: (i) the benefit of being designated a NRSRO and the resulting regulatory treatment as some have argued; (ii) the reputational capital of the dominant CRAs which cannot be easily replicated; and/or (iii) economies of scale and scope for information intermediaries? As for (i), the NRSRO designation by itself cannot explain the dominant positions of Moody’s and S&P among the ten NRSROs. As for (ii), both Moody’s and S&P have been heavily criticized for the accuracy of their credit ratings for structured finance securities up to and during the financial crisis, but nevertheless continue to retain their dominant position. See Sangiorgi & Spatt (2017). As for (iii) merely invoking the notion of economies of scale and scope, while certainly plausible in the context of a centralized information intermediary
serving a large and heterogeneous market, hardly provides the level of elucidation needed to understand exactly the nature of the barriers of entry limiting competition. Sangiorgi & Spatt (2017) point out that the Dodd-Frank Act might have increased barriers to entry and reduced competition by imposing costly regulations on CRAs that might be difficult for a new entrant to bear and amortize. As for the second question, regulation as a substitute for the disciplining effect of competition (or perhaps regulation encouraging competition), one needs to examine in more detail the potential areas arguably necessitating consideration of regulatory intervention.

The nature of CRA competition is an important issue as well: CRAs compete with each other for the ratings business of debt issuers. Generally, the issue here is traditionally thought of as CRA conflicts of interest. Simply put debt issuers generally want higher ratings for their securities assuming that they can credibly communicate those ratings to the market. In terms of conflicts of interest, all else being equal, CRAs would have an incentive to cater to issuers and particularly to large issuers that control significant ratings business.

Here it is important to emphasize a fundamental difference in the nature of issuers in the structured finance market relative to the corporate and municipal debt markets: there is far more concentration among issuers of structured finance instruments than there is in these other markets. For instance the top five issuers of private MBS in 2006 constituted approximately 40% of the market. He, Quian, & Strahan (2011); see also Frenkel (2010); Faltin-Traeger, Johnson & Myer (2010). This literature documents that debt issuers are more likely to go back to the same CRA if they received a positive rating. They report that large issuers of triple-A securities had significantly less subordination for those securities (i.e. those securities were more risky holding all else constant) than triple-A securities issued by smaller issuers. Moreover, they find that a CRA is less likely to downgrade a security than a different CRA if that CRA has rated more securities by that issuer. He, Qian & Strahan (2011) document that structured securities receive more favorable ratings if the issuer is large, i.e. a more important potential source of revenues to the rating agency. And, moreover, MBS sold by large issuers performed significantly worse during the financial crisis than did similarly rated securities of smaller issuers.

The tradeoff a CRA faces between avoiding reputational harm due to inaccurate/noisy ratings and generating ratings business from issuers can also vary not only based on the ratings business controlled by the issuer but also over the course of the business cycle. Bar-Isaac & Shapiro (2010) argue that CRAs face higher potential short-term profits from providing inaccurate ratings
during good economic times as revenues from providing ratings (and ancillary services) are greater at this point. And, as was mentioned previously, the dominant CRAs became far more profitable between 2002-2006, with a very large percentage of those profits derived from the structured finance business. Moreover, detecting inaccurate/noisy ratings are more difficult during good economic times given the low default rates. Finally, the cost of human capital to perform the analysis necessarily to provide accurate ratings is likely to be more costly during good economic times. In short, good economic times might be the right moment for CRAs to draw down on their reputational capital. A number of the papers we cite in the following sections look at precisely this issue: rating accuracy during the years of good economic performance and performance during the financial crisis.

**d) Complexity and Ratings Quality**

The more complex a securities is to model, the more likely there is to be “noise” in the rating. This obvious observation is simply a reflection of the fact that complexity can increase the range of reasonable judgments that underlie a rating decision and, hence, increases the likelihood that different CRAs come to different conclusions. Even without a conflict of interest, differences in ratings due to genuine differences can create an incentive for rate shopping (which we will discuss in the next section). Moreover, Mathis, McAndrews & Rochet (2009) modeling predicts that a CRA may decide to assign inflated ratings when the CRA derives a large percentage of its income from rating securities with high complexity.

Complexity and extensive regulatory use of ratings can interact in a way that could also result in rating inaccuracy. In the model of Opp, Opp & Harris (2013) CRAs might rationally decide not to incur the expense of actually figuring out complex securities but rather just sell a high rating to an issuer for its regulatory benefit. In this model, investors and the market are not fooled as to the informational value of the rating, but the rating nevertheless has value due to regulatory usage. Consistent with this model, Stanton & Wallace (2017) document ratings inflation in the commercial MBS market for triple-A securities in response to changes in the regulatory treatment of triple-A commercial MBS. When triple-A commercial MBS received significantly more favorable regulatory treatment, the value of these securities increased (yields decreased) substantially.
The issue of complexity suggests yet another difference between the structured finance and the corporate debt markets. Mark Adelson, Director of Structured Finance Research at Nomura Securities, testified before Congress: “The complexity of a typical securitization is far above that of traditional bonds. It is above the level at which the creation of the methodology can rely solely on mathematical manipulations. Despite the outward simplicity of credit-ratings, the inherent complexity of credit risk in many securitizations means that reasonable professionals starting with the same facts can reasonably reach different conclusions.” (quoted in Skerta & Veldkamp 2009). In other words, complexity can be more of an issue for the structured finance market.

One particular form of complexity is the lack of sufficient historical data to calibrate the parameters used for generating a rating. The ABS CDO market which performed very poorly during the financial crisis of 2007-2008 was a new market with the first ABS CDO only being issued in 1999. Calomiris (2008) discusses the limited data available on the performance of subprime mortgages in a severe downtown. Ashcraft, Goldsmith-Pinkham & Vickery (2010) document that securitizations with more interest-only loans, for which historical performance data was limited, tended to perform worse during the financial crisis. Benmelech & Dlugosz (2009a) document that while a far smaller market, structured finance securities performed quite poorly during the 2001-2002 recession.

e) Ratings Shopping

One widely expressed concern that could lead to ratings inaccuracy is the ability of issuers to shop for favorable ratings from different CRAs. The extent of this phenomenon could theoretically depend, among other factors, on: (i) the number of CRAs that an issuer could go to if the issuer did not like the initial rating (i.e. competition increasing the number of places to shop thereby reducing rating accuracy); (ii) likelihood of rating inaccuracy/noise at a particular CRA; and (iii) the degree to which the market uncritically relies on ratings (or in Bolton, Freixas & Shapiro’s (2012) parlance the existence of “trusting investors”). As for (ii), the incentive to shop might increase if there is likely to be variation in ratings due to rating inaccuracy/noise due to factors such as complexity. As for (iii), rating shopping for an inaccurate rating presumably has more value to the shopper the extent to which the resulting rating will be accepted by the relevant investors. Importantly, Bolton, Freixas & Shapiro (2012) model CRA rating incentives
and predict outcomes in line with these three factors. Their modeling assumes a duopoly CRA market structure and, moreover, that there is a segment of investors who are passive and/or uninformed investors and rely on credit ratings. In this model, these investors do not have Socratic knowledge: they don’t know they don’t know and just accept the rating at face value. It is worth noting that this concern with ratings shopping could still exist even if all CRAs were solely focused on providing the most accurate rating possible, i.e. there were no conflicts of interest. Different CRAs might genuinely have different views.

These three factors collectively suggest that rating shopping might be a particular issue in the structured finance market relative to other debt markets. For instance, factor (i) suggests that the issue of shopping in debt markets is less likely to be an issue as most corporate debt is routinely rated by both the dominant CRAs: Moody’s and S&P. The same pattern of dual ratings does not hold as strongly for structured finance securities (although dual rating is common).

Factor (ii) indicates that to the extent there is rating inaccuracy or rating noise already, rating shopping is likely to be more of a concern. Complexity as a source of CRA disagreement will arguably be less severe for the debt than the structured finance market. As Skreta & Veldkamp (2009) model, security complexity can be an endogenous response. Issuers knowing they can rate shop more effectively with complex securities have an incentive to increase complexity for that reason. Municipal bonds are an interesting example in terms of complexity. Municipal bonds tend to be quite heterogeneous giving rise to some potential complexity. If there is rating noise or inaccuracy for other reasons, perhaps due to conflicts of interest, this could also give rise to an incentive for ratings shopping.

Whether factor (iii) cuts in favor of being more concerned about ratings shopping in the structured finance market depends on whether one views the investor clientele effects (discussed earlier) in the structured finance market stronger than in other markets. This is not to suggest rating shopping is not possible in the corporate debt market. Indeed, Becker & Milbourn (2011) find that CRA competition (Fitch entering the market to compete with Moody’s and S&P for ratings business) in the corporate debt market appears to reduce ratings quality.

There have been several empirical papers on the topic of rating shopping. Benmelech & Dlugosz (2009a) found that ABS CDOs that had a single rating were more likely to be downgraded and have more severe downgrades than dual rated ABS CDOs. That being said, most ABS CDOs have more than one rating (80% of the total) and it is this market that
experienced the worst ratings downgrades and performance during the financial crisis. A potential counter-consideration to this observation is that even for dual rated ABS CDOs, ratings quality is impaired given that CRAs know that issuers could rate shop thereby inducing them to provide more favorable ratings than would otherwise be the case. If this were true, then a comparison of single versus dual-rated securities would not necessarily capture the full impact of rating shopping. Adelson, Sun, Nikoulis and Manzi (2002) found that ABS rated by S&P alone were more likely to be downgraded than dual rated ABS during the 2001-2002 recession. Griffin and Tang (2012) on the other hand failed to find evidence that CDOs with more than one rating performed better.

**f) Non-model Adjustments to Ratings**

The SEC report (2008) investigating credit ratings of MBS and CDOs found, among other things, that the rating agencies “made out of model adjustments and did not document the rationale for the adjustment.” On a related note, the SEC (2008) also found that “None of the rating agencies had specific written procedures for rating RMBS and CDOs.” Griffen & Tang (2012) found that for CDOs the CRAs would regularly make a positive adjustment to its main model resulting in a larger AAA tranche size with larger positive adjustments correlated with higher subsequent downgrades. These findings suggest that the discretion created by the ability to make non-model based adjustments was problematic for these securities during this time period. One interpretation might be that there should have been more reliance on models, rather than less. Of course, the tradeoff between discretionary non-model adjustments versus more reliance on models is likely to be a complicated one, including the fact that applying a model to a particular situation might itself require judgment and hence the use of discretion.

**g) Systematic Risk**

One particular source of concern is the interaction of rating inaccuracy/noise and systematic risk. If highly rated (such as triple-A securities) of structured finance instruments, such as CDOs, are held by passive and uninformed investors who might also be risk-averse then having these highly rated securities being substantially downgraded and falling in price all at the same time
could create a systematic problem for the market. Benmelech & Dlugosz (2009b) document that for CLOs, a popular type of CDO, the structures exhibited a high degree of similarity. They further note that the models CRAs would use to rate CDO securities was easily available to debt issuers potentially enabling them all to arrive at similar structures that would ensure the desired rating outcome. Benmelech & Dlugosz (2009b) speculate that “the uniformity of CLO structures is driven by a boiler-plate model.”

All this implies that if the CRA modeling is inaccurate (or for that matter the modeling is ex ante accurate but ex post ratings have to be downgraded due to unexpected states of the world) this inaccuracy will not be a one-off but rather will be reflected across CDO structures generally. So when one fails there will be many more likely to follow. This has consequences for a wide range of institutions, such as financial institutions posting these securities as collateral. It is in this context that the extensive regulatory and contractual use of ratings rears its head again. If institutions have to dispose of their downgraded triple-A securities for regulatory or contractual reasons, if debt covenants are violated as a result of ratings downgrades, and so on, then the failure of structured finance structures due to common mistakes in ratings modeling will have additional negative knock-on, and potentially systematic, effects.

h) Regulatory Menu

Pulling together several major strands in the literature on CRA incentives and behavior reviewed above, it appears that concerns over ratings quality should be heightened during good economic times for complex structured instruments, especially when the securities are highly rated, given potential investor clientele effects, and when the instruments are issued by larger debt issuers. Of course one could largely surmise this by looking at the performance of highly rated structured products during the financial crisis. This is not to suggest the theoretical and empirical literature is not valuable, it is in fact crucial, but rather to note that the empirical literature is drawing broad lessons largely (although certainly not exclusively) in terms of what can go wrong from an in-depth investigation of one particular historical episode. This raises the question of whether future problems in this space are likely to be sufficiently similar. For instance, to the extent that the problem with rating accuracy for structured finance instruments was a function of limited historical data on how pools of subprime mortgages that were being
structured would perform in a severe downturn (and the willingness of some investors to take those ratings at face value), this would obviously not hold true in the future.

Two further aspects of the literature are worth highlighting: the first observation is that the evidence for systematic problems with CRA ratings is far weaker with respect to the corporate and municipal debt markets. See generally Baghai, Servaes & Tamayo (2014). The second observation is the importance of ratings for triple-A securities in the context of structured finance. These securities are of particular interest given: (1) given the apparent market reliance on these ratings (presumably reflecting the nature of the investors in this marketplace); (2) the substantial academic literature questioning the accuracy of these ratings; (3) the potential implications for systematic risk arising from the similar structuring of these securities; (4) the concentrated nature of the issuers in this market; and (5) the large percentage of MBS and CDO structures (often in excess of 60% of the overall structure) that have historically been rated triple-A.

In terms of potential regulatory changes/issues, any number of possibilities have been proposed, including (but certainly not limited to):

(i) Upfront disclosure of any ratings received so as to combat ratings shopping;
(ii) Full removal of all regulatory uses of ratings (not just under federal law), perhaps replaced with more market-based measures;
(iii) Increased liability for CRA;
(iv) Regulatory oversight of CRA analytical methods used to generate ratings;
(v) Greater disclosure of CRA internal processes;
(vi) Movement away from the existing issuer-pays (and issuer-chooses) model, perhaps to an investor-pay model.

With respect to all these proposals as formulated above, one question is whether they are overbroad in covering the debt and municipal markets (and other debt markets serviced by the CRAs) and not just structured finance. In other words, is the evidence for credit ratings problems in these markets sufficiently severe to justify regulatory intervention? A few brief comments (offered in seriatim) on the above regulatory proposals might be helpful demonstrating the complexity of the choice.
(i) One important set of issues here is implementation. How would “soft” conversations exploring the possibility of a rating between an issuer and a CRA be addressed? Moreover, the literature indicates that one potential issue with ratings shopping is selection of a CRA that is known to have a view that is favorable for a particular debt instrument. If this is known in advance of any actual rating being provided then rating shopping could still occur without even “soft” conversations. Moreover, to the extent that CRAs disclose their ratings processes (see proposal (v)), this could improve the ability of issuers to select a CRA without having any interaction. Finally, in terms of the importance of such a change, one would obviously have to form a judgment as to how important overall rating shopping is as an empirical matter.

(ii) The removal of the regulatory use of ratings (as is now required under federal law) itself raises a set of important questions. First, the removal of the required use of ratings does not imply that ratings would not be used in conjunction with other evidence by regulators in forming judgments, such as in the area of prudential regulation. See White (2010). To this extent, there would still remain a potentially important regulatory impact of ratings, as a de facto if not a de jure matter. Also, assuming decreased regulatory reliance on ratings, what implications does this have for the appropriate level of regulation of CRAs (such as proposals (iv) and (v))? Does this imply that they should be viewed as private market actors providing a business service or is regulation still needed given their continuing impact in the marketplace?

(iii) Putting aside any first amendment questions, increased liability (such as Section 11 liability for ratings provided in a registration statement) also raises a host of issues, including all the standard ones concerning the ability of the legal system to assess on an ex ante basis the reasonableness/appropriateness of a particular rating decision. Depending on how and when liability was imposed, increased liability could be argued to have the effect of increasing accuracy (more information) or, alternatively, reducing the willingness of CRAs to provide ratings at the margin (less information) and perhaps a bias towards lower ratings than would otherwise be the case (less information) in order to reduce disappointment ex post that could lead to litigation. So far CRAs have successfully avoided incurring Section 11 liability despite the Dodd-Frank Act.

(iv) & (v): One important question that would need to be considered is whether regulatory imposition of a similar set of “best practices” and/or disclosure of techniques could induce greater conformity in the rating processes. This could ensure that any mistake is compounded
throughout the system. Moreover, greater conformity might reduce the incentive for a firm to enter the CRA market as a new competitor if such an entrant would merely be replicating what is already currently on offer. Whether such an impact on competition is a positive or a negative is itself another difficult question as the earlier discussion on competition emphasized. One particular comment in terms of (v) is the widely expressed view that CRAs were “too” model-based at the expensive of seeing the bigger picture. The work of Griffen & Tang (2012), however, suggests that qualitative (and perhaps ad hoc) adjustments to models were an issue. Of course no one could object to having “better” models, but this is hardly a helpful observation as a basis for policy.

(vi) In terms of the investor-pays model, it is entirely possible that an investor-pays model could also suffer from conflicts of interest. If a large investor has significant stakes in a particular debt instrument a ratings downgrade might not be in the investor’s interest (or a rating downgrade for a group of debt instruments that the investor owns). Or an investor might desire, in spirit of Opp, Opp & Harris (2013), an inflated rating for regulatory or contractual purposes (such as avoiding limits placed on the investor in terms of their holdings). On a separate note, what would be the impact of selective disclosure of ratings (disclosure to just paying investors) on price discovery in the debt markets? Obviously investors would not pay for ratings if they are publicly disclosed anyway. See generally White (2010); Coffee (2011). One proposal is for the SEC or some other governmental body to choose the CRA to do a particular rating with the issuer still paying. Obvious questions with this proposal are the standard ones concerning the incentives and information capacity of governmental actors and how fees in such a system would be set. One could also imagine a regulatory capture scenario with such a structure.

5. Broker-Dealers

Broker-dealers are clearly important intermediaries in the financial markets. In particular, they are important intermediaries in terms of how households and retail investors participate in the financial markets, including retirement savings, direct ownership of stock and as an important source of financial advice. There are approximately 4,000 registered-broker-dealers with some 100 million investor accounts. SIFMA (2016).
The central role played by broker-dealers as a financial intermediary for households and retail investors directly implicates important issues of investor protection which in turn is a core mission of both the Securities and Exchange Commission and the Financial Industry Regulatory Authority (“FINRA”), the self-regulatory organization which oversees broker-dealers subject to SEC oversight. The centrality of investor protection issues as a focal point of broker-dealer regulation is further heightened by the dual roles that broker-dealers often play as a financial intermediary: broker-dealers act as both the agent for retail investors (hence the “broker” portion of broker-dealer) and as a potential trading counterparty for retail investors’ orders (hence the “dealer” portion of broker-dealer). To be sure, there are other issues besides investor protection raised by broker-dealers’ role in the financial markets, such as the issue of systematic risk, see, e.g., Duffie (2010), but these will not be covered in this section.

Not surprisingly the original Special Study spent a considerable amount of time, resources, and ink on the topic of broker-dealers. Chapter III of the original Special Study is entitled: “Broker-dealers, Investment Advisors and their Customers – Activities and Responsibilities” and extensively covers broker-dealer sales practices. Broker-dealer sales practices are still very much relevant regulatory topics such as: the training and oversight of broker-dealers that interact with the investing public; the contours of the “suitability” requirement (the general requirement that broker-dealers only recommend securities that are suited for investors being solicited); and (mis)incentives resulting from broker-dealer commission-based compensation. Naturally, some broker-dealer issues addressed in the original Special Study are no longer relevant, most notably those arising from the fixed broker-dealer commission schedule of the day (abolished in 1975).

It is fair to say that while an important and longstanding regulatory topic, broker-dealer sales practices as a general matter have not received significant academic attention (unlike, for example, other topics we have discussed such as mutual funds and CRAs). This raises an important point concerning any future Special Study: if broker-dealer sales practices are going to be covered, as was the case before, it is likely that a significant additional amount of work would have to be undertaken to do so given the relative dearth of academic work which traditionally would serve as a starting point. That being said, there are a few areas where academic work has been done, particularly in recent years, shedding light on the interface between broker-dealers and the investing public, perhaps most notably on the important topic of the market for financial advice for retail investors.
a) Suitability versus Fiduciary Obligations

The formulation and scope of the legal responsibilities owed by a broker-dealer to the investing public is just as topical today as it was at the time of the original Special Study. The original Special Study found that “some segments of the [broker-dealer] industry appear to be earnestly promoting high standards of selling while others seem only to be earnestly promoting sales” and recommended (among other things) that “greater emphasis should be given by the Commission and the self-regulatory bodies to the concept of suitability of particular securities for particular customers.” There have recently been a number of proposals to move from a suitability standard to a fiduciary standard for broker-dealers. For instance, the SEC in its 2011 Study on Investment Advisors and Broker- Dealers (“SEC Broker-Dealer Study”) recommended that:

The standard of conduct for all brokers, dealers, and investment advisers, when providing personalized investment advice about securities to retail customers (and such other customers as the Commission may by rule provide), shall be to act in the best interest of the consumer without regard to the financial or other interest of the broker, dealer, or investment adviser providing the advice.

The SEC Broker-Dealer Report also recommended that the Commission promulgate rules and provide interpretive guidance on what this standard actually entails. The Department of Labor proposed regulations (now suspended) that impose fiduciary obligations on broker-dealers offering investment advice for assets held in IRAs.

A natural starting point before discussing proposals like this is an understanding of the current governing standards in this area. The Supreme Court’s decision in SEC v. Capital Gains, 375 U.S. 180 (1963), decided around the same time that the original Special Study came out, held that “investment advisors” pursuant to the Investment Advisors Act have fiduciary obligations that run to their clients. This fiduciary obligation creates both a duty of care and a duty of loyalty. Under the duty of care, investment advisors must provide investors only recommendations that serve their clients’ best interests. Under the duty of loyalty, investment advisors must disclose to clients any potential conflicts of interest, such potential conflicts resulting from the firm’s compensation arrangements.

40 Chapter 3, Special Study
In contrast, broker-dealers that are not considered “investment advisors”, while not having a fiduciary obligation under the *Capital Gains* decision, nevertheless have numerous legal obligations. The SEC in summarizing the regulatory purpose behind these obligations has consistently emphasized the goal of investor protection:

The broker-dealer registration and associated regulatory requirements of the Act, as well as those of the self-regulatory organizations, provide important safeguards to investors. Investors are assured that registered broker-dealers and their associated persons have the requisite professional training and that they must conduct their business according to regulatory standards. Registered broker-dealers are subject to a comprehensive regulatory scheme designed to ensure that customers are treated fairly, that they receive adequate disclosure and that the broker-dealer is financially capable of transacting business.41

These broker-dealer requirements include: (1) providing only recommendations that are “suitable” for the customer; (2) complying with FINRA’s “know your security” requirements; (3) ensuring that customers’ orders receive “best execution”; (4) complying with broker-dealer rules governing “markups” on securities; see Ferrell (2008); (5) acting consistent with an implied representation of fair dealing (sometimes referred to as the “shingle theory”); (6) comporting with FINRA’s rules requiring broker-dealers to “observe high standards of commercial honor and just and equitable principles of trade”; (7) complying with the nontrivial registration and qualification requirements of Section 15 of the Exchange Act of 1934;42 and (8) complying with various affirmative obligations to disclose information in conjunction with a broker-dealer recommendation.

The distinct regulatory regimes for investment advisors and broker-dealers raises two immediate questions: (1) when are broker-dealers deemed “investment advisors” and hence have a fiduciary duty under *Capital Gains*; and (2) given the host of rules that broker-dealers currently face, what conduct exactly is permitted under existing broker-dealer regulation that would be barred under a fiduciary obligation? In other words, how much of a practical difference would a

41 See SEC, Persons Deemed Not to Be Brokers, SEC Release No. 34-20943 (May 9, 1984)
42 Under Section 15 registered broker-dealers are subject to a host of compliance requirements and obligations such as “meeting certain standards of operational capability and standards of training, experience, competence, and other qualifications established by the SEC; becoming a member of a self-regulatory organization; being subject to investigations, inspections, and disciplinary actions by the SEC; complying with minimum net capital requirements, customer protection rules, specific recordkeeping, financial compliance, and financial reporting requirements. Registered Broker-Dealers are also subject to the general antifraud and anti-manipulation provisions of the federal securities laws and implementing rules, as well as specific antifraud requirements.” Colby, Scwhartz & Zweihorn, What is Broker-Dealer? (2015).
move towards imposing on broker-dealers a fiduciary obligation represent? A surprising amount of the discussion swirling around various proposals to extend fiduciary obligations to broker-dealers are silent on exactly how existing broker-dealer legal obligations would be altered.

As to the first question, there are various scenarios where broker-dealers are currently deemed to have a fiduciary obligation. To generalize and simplify existing law, these include situations where the broker-dealer: (1) specifically charges separately for investment advice (as opposed to a general wrap/fee on the account) and by virtue of this fact are deemed an “investment advisor”; (2) are consider fiduciaries under state law (such as California); or (3) has investment control over a discretionary account (sometimes these take the form of a wrap account that combine brokerage and discretionary management with fees being based on assets under management). Approximately 18% of all broker-dealers are in fact registered as “investment advisors.”

As to the second question, Langevoort (2009) points out that disclosure of broker compensation arrangements that give rise to conflict of interests concerns, and other information on broker-dealers’ potential conflicts of interest, is the most likely candidate for situations where the existence of a fiduciary obligation might make a practical difference. Consistent with this, the Investment Advisors Association letter to the SEC on fiduciary obligations identifies the following as broker-dealer obligations that do not presently exist but would under a fiduciary obligation: “Brokers recommending and selling investment products to customers would have to disclose all fees, compensation, and other incentives they earn from the advice . . . Brokers would have to disclose not only information about investment products they recommend, but also information about themselves, including conflicts of interest.”43 To take an often discussed example, a broker-dealer could arguably recommend to a customer a high-load mutual fund for which they receive undisclosed payments from the fund despite there being a more attractive alternative investment, perhaps a no-load mutual fund. In other words, the broker-dealer, perhaps out of self-interest, might recommend a high-load mutual fund that could be considered a “suitable investment” but may not be the “best investment”. The high-load mutual fund type example is a scenario that has loomed large in the Department of Labor’s proposed (and now suspended) fiduciary rule.

---

43 Investment Advisor Association, Letter to SEC (August 30, 2010).
That being said, the broker-dealer suitability requirement does ensure some meaningful degree of consistency between the interests of a customer and the security being recommended. FINRA in its 2012 suitability guidance explained that the suitability rule “prohibits a broker from placing his or her interests ahead of the customer’s interests” such as a “broker whose motivation for recommending one product over another was to receive larger commissions.” This guidance is in line with enforcement cases over the years that have repeatedly found broker-dealer recommendations to be unsuitable because of a failure to properly factor in costs. One question given this articulation of the suitability requirement (and others like it) concerns the likelihood of a divergence between the suitability requirement and what is in a customer’s “best interests” under a fiduciary standard in any given set of circumstances. How important are these “gaps” as a practical matter?

In terms of the overlap between broker-dealer regulation and fiduciary obligations, one could approach the question by starting with a focus on what broker-dealer’s fiduciary obligations might look like. The Supreme Court long ago observed that:

To say that a man is a fiduciary only begins analysis. To whom is he a fiduciary? What obligations does he owe as a fiduciary? In what respect has he failed to discharge these obligations?

What exactly does it mean to say that broker-dealers have a fiduciary obligation? To take an important example, how would the answer to this question be affected by the ability of a broker-dealer to act as principal in a customer’s trade? The ability of a broker-dealer to act as principal gives rise to a potential conflict of interest/disclosure issues in terms of execution quality. Importantly, the Dodd-Frank Act does not appear (at least explicitly) to authorize the SEC to impose section 206(3) of the Investment Advisors Act on broker-dealers which is the provision of the Act that requires investment advisors acting as a principal to provide disclosure and customer consent for each and every transaction. See generally Wrona (2012). Presumably then, extension of fiduciary obligations to broker-dealers would not include Section 206(3) investment advisor restrictions on principal trading.

---

44 FINRA, Regulatory Notice 12-25, at 3–4 (May 2012),
45 See, e.g., Department of Enforcement v. Belden (NAC August 13, 2002) (“We find that [the broker-dealer] made an unsuitable recommendation to his customer. [The customer’s] purchase of Class B shares, instead of Class A shares, resulted in significantly higher commission costs . . . ”)
46 SEC v. Chenery, 318 U.S. 80 (1943)
In determining the possible content of broker-dealer fiduciary obligations, the question is complicated by the tremendous variation in the size of broker-dealers (from one local office to multinational operations); the variation in the range of services provided by broker-dealers (from full-service to execution services only); the sophistication of the customers (from retail to highly sophisticated institutional actors); and the range of compensation arrangements employed (from solely commission-based to assets under management fee-based). Tellingly, the SEC Broker-Dealer Study recommends alongside an extension of the fiduciary obligation to broker-dealers that the “Commission should engage in rulemaking and/or issue interpretive guidance addressing the components of the uniform fiduciary duty.” But the meaning and impact of extending fiduciary duties turns on the content of this future rulemaking/interpretative guidance making it difficult to evaluate the merits and demerits of that Study’s recommendations.

The upshot of this discussion is that an important issue for a new Special Study in the area of broker-dealer regulation would be to document and identify areas where existing broker-dealer regulation and fiduciary law would likely diverge as a practical matter and then assess whether these areas of divergence are problematic and best addressed via imposition of a fiduciary obligation.

b) Compensation and Financial Advice

A recurring and longstanding concern with the current state of affairs is the incentive effects of commission-based brokerage compensation (the dominant form of brokerage compensation). Commission-based compensation will be defined for these purposes to include side payments from financial product providers to broker-dealers for marketing and selling their product. Concerns over brokerage compensation arrangements appear prominently in the original Special Study. The Special Study stated the concern this way:

The general rule of commission compensation for sales efforts creates two problems: The salesman is economically motivated to persuade customers to enter into as many transactions as possible, thereby creating the danger of excessive trading or churning; he also benefits most from sales of those securities for which the rate of commission i.e. highest, and is thus motivated to recommend purchases of securities without sufficient regard for their merit or suitability for a particular customer.47

47 Special Study, Chapter 3, p.254.
Concerns over brokerage commission-based compensation has continued unabated in the fifty plus years following the Special Study. While there is still the traditional concern with the “churning” of investors’ accounts induced by commission-based compensation, a broader concern is that the financial advice provided by broker-dealers might be biased as a result of compensation arrangements, such as side payments from financial product providers. The CFA Institute in a survey of its membership reported that 64% believed that the “fee structures of investment products drive their sales to customers rather than their suitability requirements.” (CFA 2009).

There have been from time to time over the years various proposals for moving towards a compensation regime based on fees as a percentage of assets under management, as is the case with investment advisors – approximately 95% of investment advisors are compensated based on assets under management (IAA (2014)) – rather than commissions and side payments for particular transactions. For example, in 2009 the UK Financial Services Authority considered a proposal requiring that “advisor firms to be paid by advisor charges: the rules do not allow advisor firms to receive commissions offered by service providers.” Concerns over the incentive effects of commission-based brokerage compensation became a particular focal point for the SEC during the 1990s. In particular, the Tully Committee Report of 1995 took the view that fee-based, rather commission-based brokerage compensation better aligned broker-dealer incentives. Tully (1995). These concerns lead the Commission to exclude from the category of “investment advisor” many broker-dealers that used fee-based, rather than commission-based, brokerage compensation. This exclusion was motivated to a significant extent by a desire to encourage broker-dealers to adopt fee-based compensation arrangements by excluding these broker-dealers from the category of “investment advisor” and therefore fiduciary obligations.48

Interestingly, some commentators take the opposite approach arguing that an across-the-board imposition of fiduciary duties on broker-dealers would help address the type of undesirable incentive effects identified by the original Special Study. This approach raises several questions: first, to what extent should the focus be squarely on brokerage compensation arrangements (and/or disclosure of these compensation arrangements), rather than tackling these particular issues indirectly through the imposition of a fiduciary duty? In this connection it is

48 The SEC’s rule was later vacated in 2007 by the United States Court of Appeals for the D.C. Circuit in Financial Planning v. SEC, 482 F.3d 4181.
worth noting that the Dodd-Frank Act poses a potential constraint on such a backdoor approach: the Act states that a broker’s commission “shall not, in and of itself, be considered a violation of [any fiduciary duty] applied to a broker-dealer.” And, indeed, it has long been possible for investment advisors to charge commissions consistent with the Investment Advisors Act, although this is infrequent.

Putting aside the general policy debate over extending fiduciary obligations to broker-dealers, there has been academic work, particularly in recent years, relevant to considering broker compensation arrangements and broker-dealer provision of financial advice. The following identifies some strands of that literature:

First, there has been empirical work investigating how retail investors interact with broker-dealers and investment advisors. It appears as if investors do in fact commonly receive financial advice. Hung, Clancy, Dominitz, Talley, Berrebi & Suvankulov (2008) report that 73% of investors in their sample consulted a financial adviser before purchasing shares or mutual funds. Chater, Huck & Inderst (2010) report in a survey that most investors in their sample were essentially ignorant of financial advisors’ conflicts of interest. Finally, there is some empirical evidence that investors most in need of financial advice often do not actually follow the investment advice actually given. See Bhattacharya, Hackethal, KAESLER, Loos, and Meyer (2012). In short, while investors often receive financial advice there are questions as to the extent to which they evaluate that financial advice in light of potential biases or necessarily even follow the advice proffered.

Second, there is evidence that financial advice can in fact be biased due to conflicts of interest. For instance, Hackethal, Inderst, and Meyer (2011) find that using a financial advisor was associated with increased turnover/churning of investor accounts. Interestingly, this study thus implies that investors do rely on their financial advisor, even if not necessarily to their benefit. Other papers on this topic include Anagol, Core & Sarkar (2013); and Mullainathan, Noeth & Schoar (2012). Noteworthy for present purposes is the fact that these studies typically do not separately analyze broker-dealers and investment advisors. See e.g., Mullainathan, Noeth & Schoar (2012) (“The specific advisers we are looking at in this study are retail advisers whom the average citizen can access via their bank, independent brokerages, or investment advisory firms.”).
Third, there has been interesting work done analyzing the incentive effects of various compensation arrangements on financial advisors. Inderset & Ottaviani (2009) model broker-dealers’ incentives to sell unsuitable financial products and find that this incentive can significantly increase when employees at a broker-dealer are paid both for finding new customers and selling financial products to customers, rather than having these tasks undertaken by different individuals at the firm. Among the most important work in this area has been that of Gabaix & Laibson (2006) who present a model in which “myopic” customers do not realize that a financial advisor can have an incentive to sell high-priced financial products even though the financial advisor benefits from the high prices charged (such as through the provision of a side payment from the product seller to the broker-dealer). As a result, firms will have an incentive to lower the prices that myopic customers actually observe while “shrouding” the high fees these customers actually pay. In this model, competition does not necessarily result in myopic customers being better served. Essentially competitors who might educate the myopic customers will not benefit from these customers defecting to their firm as these customers will now know how to avoid unnecessarily high fees. In a similar vein, Inderset & Ottaviania (2012) find that when customers are naive about broker-dealer conflicts of interest, there is an incentive to increase the unobserved prices and fees charged these customers for a financial product.

Interestingly, in this literature, if customers are not naive, which is to say that they understand that broker-dealers have an incentive to sell high-priced products, broker-dealers receiving side payments from financial product providers can in fact be efficiency-enhancing. Commissions based on selling a financial product can provide an incentive to broker-dealers to learn more about the financial product that they might market and sell. The possibility of earning a commission provides a financial motivation to work hard in contrast to earning a guaranteed fee based on assets under management which could induce shirking.

Putting aside incentive effects, commission-based compensation can simply be more cost-effective for some investors than a fee-based account, for example for investors that do not trade very often. Indeed, there have been enforcement cases against broker-dealers for inappropriately recommending fee-based accounts given the added expense such accounts can entail for some customers. It is also possible that commission-based accounts are not only the best choice but the only choice for some investors of more modest means. If a customer’s assets under management are modest then a fee-based account (assuming one is applying a typical percentage
of assets) might not provide sufficient compensation for services rendered. Consistent with this, empirical evidence indicates that lower-wealth individuals tend to have commission-based accounts, including those with investment advisors. See Dean & Finke (2012). This fact dovetails with the concern that a legal regime that makes commission-based arrangements less available could reduce the availability of needed financial advice for investors with more limited means.

In both the Gabaix & Laibson (2006) and Inderst & Ottavalia (2012) models, a critical factor in terms of the impact of a broker-dealers’ potential conflicts of interest on customers is how sophisticated those customers are with respect to understanding the incentives of the broker-dealer. The importance of this factor could argue for a regulatory distinction between sophisticated and unsophisticated investors (and raise the question of how well this distinction is captured by the commonly drawn regulatory distinction between institutional and retail customers). Consistent with the implications of these models, and the importance of focusing on customer sophistication, are the findings of Bergstesser and Beshears (2010) who document that less financially sophisticated borrowers tended to purchase adjustable-rate mortgages at a higher rate, mortgages which then went on to have higher rates of foreclosure.

Fourth, there has been interesting research on complexity and embedded fees in financial products relevant to broker-dealer marketing and selling practices. Consider two papers from this literature. Celerier & Vallee (2015) analyze 55,000 structured products marketed to retail investors over the 2002-2010 period. They report that the more complex the structured product is the more profitable that structured product is to the financial institution selling it and the worse it tends to perform ex post. Henderson and Pearson (2011) find significant overpricing of a sample of popular structured products. This raises the general investor protection concern over fees being embedded in these structures. See Bethel & Ferrell (2007).

c) Proposals and Regulatory Menu besides Fiduciary Duties

Enhanced Disclosures: One could image any number of disclosure enhancements to the existing regulatory regime. One interesting proposal was presented in a 2010 FINRA concept release. The proposal here was to require broker-dealers to make publicly available disclosures concerning potential conflicts of interest at the very outset of any advisor-customer relationship,
much as investment advisors currently due on Form ADV. Imposition of such disclosures would arguably represent removing an important area of divergence between current broker-dealer and investment advisors’ obligations that motivate arguments for imposition of fiduciary duties.

In the Inderst & Ottaviania (2012) model, mandatory disclosure of potential broker conflicts of interest (such as receipt of side payments) can improve investor outcomes if the disclosure turns naive investors into “wary” or sophisticated investors (investors that take into account the incentive effect these conflicts create). But whether enhanced disclosure will have this salutary result (or to what extent) is non-obvious. There is room for skepticism. See, e.g., Beshears, Choi, Laibson and Madrian (2009). On a behavioral note, Loewenstein, Cain & Sah (2011) shows that disclosure of conflicts of interest in some circumstances could perversely actually increase the bias in financial advice with customers being more willing to follow that advice. One additional consideration that bears on assessing the ability of some retail customers to become more sophisticated in light of new information is the failure of retail investors who actively trade to learn from their mistakes, such as the need for diversification. On the other extreme, it is also possible to imagine a customer reaction that overemphasizes the importance of the disclosure, such as broker-dealer receipt of side payments, at the expense of other relevant dimensions of the decision. See Hackethal & Inderst (2013).

Even with disclosures, broker-dealers could simply vertically integrate by providing both financial advice and the financial product thereby rendering unnecessary the transfer of a side-payment. Indeed, as discussed, one of the defining features of broker-dealers is the ability to engage in transactions as a principal. This observation calls into question the effectiveness of a more stringent prohibition such as a bar or limits on side-payments even if one were to put aside efficiency-enhancing reasons for such payments. Consideration of more substantive non-disclosure based regulation of broker compensation arrangements leads naturally into consideration of best execution and markup requirements as a potential regulatory tool to address potential conflicts of interest.

Uniformity and Investor Confusion: A common critique of existing broker-dealer regulation is the need for equal treatment of broker-dealers and investment advisors insofar as they are engaged in the same activity, i.e. providing financial advice. The SEC Broker-Dealer Study, for instance, states that its recommendations, including the proposed extension of fiduciary obligations, are “intended to make consistent the standards of conduct applying when retail
customers receive personalized investment advice about securities from broker-dealers or investment advisors.” But, once again, a threshold question is the actual extent of divergence between these two bodies of law as a practical matter.

Given that one is not creating a regulatory regime on a blank slate, but rather dealing with long-standing existing structures, one would need to consider whether pursuing uniformity for its own sake is worth the costs in terms of disruption and transition. And it is at least an open question as to how much uniformity can be achieved even with across-the-board fiduciary obligations given the differences between broker-dealers and investment advisors that are likely to remain, such as on the issue of principal trading and enforcement mechanisms (more of which later). It is interesting to note in this connection that SIFMA supported the adoption of an uniform standard of conduct but then stated that what conduct served the “best interest of the customer” should be based on existing case law and guidance developed under Section 206 only for investment advisors, not broker-dealers.

Related to the common argument for the need for uniformity is the argument that the existing distinction between broker-dealers and investment advisors generates investor confusion over the responsibilities of broker-dealers. Indeed, this appears to be the main empirical foundation for the SEC Broker-Dealer Study’s recommendation to extend fiduciary duties. But it is hardly surprising that there is investor confusion over complex legal rules, a confusion that would surely exist regardless of whether fiduciary obligations are extended or not. The more relevant question is how regulation can be substantively improved so as to ensure better investor outcomes.

**Enforcement:** A critical component of any regulatory regime focused on financial intermediaries is examination and enforcement. Here, the differences between broker-dealers and investment advisors might be as important as any differences between existing broker-dealer obligations and the investment advisor fiduciary obligation. Interestingly, these differences can been used to argue that broker-dealer regulation is actually more demanding than that of investment advisors.

Broker-dealer examinations largely occur under the umbrella of its self-regulatory organization (SRO) FINRA. Broker-dealers are also occasionally examined by the Commission as well. FINRA examinations of broker-dealers can and often do lead to disciplinary action ranging from deficiency letters to loss of FINRA membership. Investment advisors, on the other
hand, are examined by the SEC’s Office of Compliance Inspections and Examinations. This results in significant differences in the frequency of examinations of broker-dealers relative to investment advisors. Over half of all broker-dealers are examined every year by FINRA whereas approximately 9% of investment advisors are examined by the SEC. Indeed, the ratio of assets under management by investment advisors per SEC examiner has increased from $42 billion in 2004 to $83 billion per examiner in 2010.

In terms of enforcement, there are also disparities. The SEC regularly brings enforcement actions against both broker-dealers and investment advisors. But on top of this, FINRA also regularly brings enforcement actions against broker-dealers. Wrona (2012) estimates for 2009 the total number of disciplinary actions against broker-dealers at 1,102 (SEC plus FINRA) and for investment advisors a total of 76 (SEC actions). In 2016 FINRA brought a total of 1,434 disciplinary actions against registered broker-dealers and individuals. While comparing the mere number of disciplinary actions ignores important information such as the value of those actions or the resources incurred in bringing them, there does appear to be a significant gap between broker-dealers and investment advisors along the dimension of enforcement.

In considering the level of examinations and enforcement for investment advisors it is worth bearing in mind that the size of the investment advisor universe (which includes broker-dealers registered as investment advisors) is enormous: there are approximately 11,000 firms registered as investment advisors with over $61 trillion in client assets and 27.8 million clients. IAA (2014). The nature of the clients served varies widely from pensions funds to high-net worth individuals to more typical retail investors.

A potential issue for a new Special Study therefore is whether these disparities between broker-dealers and investment advisors along the dimensions of examination and enforcement are cause for concern. There have been repeated calls, starting with the Special Study itself, for investment advisors to form their own SRO with similar enforcement and examination functions that are now played by FINRA for broker-dealers. The US Treasury Department in its Treasury Blueprint for a Modernized Financial Regulatory Structure likewise recommended a SRO for investment advisors. Treasury (2008). The Dodd-Frank Act required the SEC to consider whether such an investment advisor SRO would improve the examination and enforcement process with the resulting SEC study arguing for augmenting resources at the SEC’s Office of Compliance Inspections and Examinations.
References

Adelino, M. (2009). Do Investors Rely only on Ratings, MIT dissertation


CFA Institute (2009). European Union Member Poll on Retail Investment Products


Investment Company Institute (2016)


Kisgen, D.J., Strahan, P.E. (2010). Do regulations based on credit ratings affect a firm’s cost of capital?, *Review of Financial Studies* 23(12)


 Lehman Brothers Securitized Products Research (2008), Residential Credit Losses


Morley (2017). *Too Big to Be Active*


SIFMA (2016) *Fact Book*


