9.1 Introduction

The stock market bubble of the 1920s was accompanied by questionable conduct by security issuers, underwriters, brokers, and investment companies. Stock in sham companies was issued and pushed on novice investors by aggressive stockbrokers, and the prospects of established firms were knowingly exaggerated.¹ Shareholders in investment companies had their assets diluted by self-dealing managers.² The subsequent crash motivated the creation of the institutions and laws that form the core of modern US financial regulation.

During the late 1990s and early 2000s, history repeated itself on a smaller scale. A set of abuses by accountants, equity analysts, brokers, and investment companies motivated a major new law (the Sarbanes-Oxley Act of 2002), new rulemaking by the Securities and Exchange Commission (SEC), and newly vigorous enforcement of existing laws and rules by the SEC and other regulators. It also led to a surge in interest in further refining finan-

¹ For example, in the conference report accompanying the 1933 Securities Act, the House of Representatives (1933, 2) claims that “fully half or $25,000,000 worth of securities floated during this period [the decade following World War I] have been proved to be worthless.” While this is a claim about the ex post value of these securities after the 1929–1933 market decline, it clearly reflects a belief that many of these securities were of questionable value ex ante.

² See Baumol et al. (1990) and Securities and Exchange Commission (1992) for more details.
cial regulation, especially among generalists. As a crude proxy of generalist interest, figure 9.1 plots mentions of the phrase “Securities and Exchange Commission” in the *New York Times*. Mentions spiked with the collapse of Enron in late 2001 to levels not seen since the 1930s, and they remained at high levels for about four years.

The first draft of this chapter was written and presented at an NBER conference in 2005, as this peak of interest was receding. Interest in financial regulatory reform was revived after the financial crisis of 2008. While the specifics of the policy discussions of 2002 to 2005 and 2008 to 2012 have differed, the underlying market failures that provide a rationale for regulation are surprisingly consistent.

To illustrate this point, I have left the main body of the chapter largely as it was written in 2005, but will conclude by relating this discussion to issues that emerged after 2008. The chapter provides an overview of the regulation of the US retail securities and investments industry from the perspective of an industrial organization economist. It discusses the economic rationale for regulation, the institutions and laws that emerged after the 1929 crash, and then turns to a discussion of more current issues. Given the vastness of the field, this discussion is focused on three issues with parallels in other industries: the regulation of pricing, antitrust, and firm boundaries and their implications for conflicts of interest.\(^3\) With the benefit of postfinancial

\(^3\) Readers interested in a more comprehensive survey of the field are referred to Seligman (2003) or the legal textbook *Coffee and Seligman* (2002). *Baumol et al.* (1990) and *Securities*
Retail Securities Regulation in the Aftermath of the Bubble

In crisis hindsight, I have added a discussion of a fourth issue: the potentially perverse effects of competition when important aspects of product quality, such as the safety of investments, are unobserved by most consumers. This issue emerged in the literature before the financial crisis, but has become more widely discussed since.

The remainder of the chapter is outlined as follows. Section 9.2 discusses the scope and economic size of the retail securities and investments industry, while section 9.3 discusses the underlying reasons why it might require regulation. Section 9.4 provides a brief overview of the main institutions and laws, while sections 9.5 through 9.7 discuss the three current issues just outlined. A conclusion follows.

9.2 Size and Scope of the Retail Securities Industry

Financial services, broadly construed, are a larger piece of the economy than many economically literate Americans realize. Most regular newspaper readers are aware that US health care expenditures are about 15 percent of GDP and that this ratio is about 1.5 times higher than in other advanced countries (OECD 2005). This figure is the centerpiece of an active debate about the extent to which it reflects high quantity and quality, high prices and economic rents, or waste.

Fewer are aware that the corresponding figures for financial services are about as high.4 Because financial services are an intermediate good as well as a final good, a direct comparison of expenditure data is not meaningful. Table 9.1 reports that gross value added of the financial intermediation sector (which includes banking, insurance, and securities) is 8.1 percent of GDP for the United States and an average of 5.1 percent in the rest of the G7. For comparison, gross value added figures for “health and social work” are provided. Gross value added excludes purchases of materials, services, and capital equipment, and so these figures are not directly comparable to the more commonly quoted expenditure data, but they do suggest that finance and health care are roughly comparable in size.

An alternative measure of the sector’s size is provided by revenue data from the 2002 Economic Census. The total revenue of the financial intermediation sector in 2002 is $2.7 trillion, or about 25 percent of GDP (table 9.2). This figure includes both revenue from interest on loans and double counts revenue from intermediate goods and services, so it overstates the size of the sector. The Economic Census data show that the sector accounts for

---

4. After the financial crisis of 2008, the size of the financial services sector has received more discussion (e.g., Philippon and Reshef 2009; Gennaioli, Shleifer, and Vishny 2012; Greenwood and Scharfstein 2013).
12.8 percent of revenue, 10 percent of payroll, and 6 percent of employment reported in all industries. The last two ratios do not suffer from double counting and the first includes it in both numerator and denominator, so these are better indicators of its share in the economy. The conclusion that the sector is about as large as health care still seems at least roughly valid.

Of the $2.7 trillion in revenue reported in the census, about $400 billion falls into the scope of this chapter: securities and investment products purchased by (but not necessarily exclusively by) retail investors. The activities of this industry can be roughly thought of as a value chain with four steps (figure 9.2). First, securities are underwritten and distributed to their initial owners, for example, through an initial public offering (IPO). Second, securities are traded on secondary markets by both proprietary traders and brokers acting as agents for either individual investors or portfolio managers. Third, many securities are purchased and held by investment products such as mutual funds or variable annuities. Fourth, the investment products are sold to retail investors by financial advisors, brokers, banks, insurance agents, or mutual fund companies. In some cases, stages of the process are bypassed. For example, some investors bypass stages 3 and 4 by purchasing securities directly through discount brokers, or bypass stage 3 but not 4 by purchasing securities on the advice of a full-service broker. Likewise, some investors bypass stage 4 by buying investment products such as mutual funds directly from the fund’s manager. Investors and investment funds also often bypass stage 2 by investing directly and holding new issues, especially for bonds with illiquid secondary markets. Even given these exceptions, the four-stage value chain is a useful organizing framework, particularly given that, as discussed later, laws, rulemaking, and regulatory bodies are organized around this delineation of activities.

Table 9.3 provides a product level breakdown of revenue accounted for by

<table>
<thead>
<tr>
<th></th>
<th>Gross value added in financial intermediation (%)</th>
<th>Gross value added in health and social work (%)</th>
<th>Total national expenditure on health care (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>8.1</td>
<td>6.9</td>
<td>15.0</td>
</tr>
<tr>
<td>Equal-weighted average of rest of G7</td>
<td>5.1</td>
<td>5.9</td>
<td>9.2</td>
</tr>
<tr>
<td>Canada</td>
<td>5.9</td>
<td>5.4</td>
<td>9.9</td>
</tr>
<tr>
<td>France</td>
<td>4.3</td>
<td>7.6</td>
<td>10.1</td>
</tr>
<tr>
<td>Germany</td>
<td>3.8</td>
<td>6.1</td>
<td>11.1</td>
</tr>
<tr>
<td>Italy</td>
<td>5.4</td>
<td>4.5</td>
<td>8.4</td>
</tr>
<tr>
<td>Japan</td>
<td>7.0</td>
<td>n/a</td>
<td>7.9</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>4.4</td>
<td>5.9</td>
<td>7.7</td>
</tr>
</tbody>
</table>

Sources: OECD National Accounts for value added; OECD Health Data for total health care expenditures. National accounts data is from 2001 for Canada and the United Kingdom; health expenditure data is from 2002 for Japan and the United Kingdom.
### Table 9.2: Size of US financial intermediation industries, 2002

<table>
<thead>
<tr>
<th>NAICS code</th>
<th>Revenue ($ millions)</th>
<th>Payroll ($ millions)</th>
<th>Employees (thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finance and insurance</td>
<td>2,732,546</td>
<td>377,236</td>
<td>6,534</td>
</tr>
<tr>
<td>521</td>
<td>Monetary authorities—central bank</td>
<td>28,909</td>
<td>1,234</td>
</tr>
<tr>
<td>522</td>
<td>Credit intermediation and related activities</td>
<td>1,061,126</td>
<td>148,211</td>
</tr>
<tr>
<td>5221</td>
<td>Depository credit intermediation</td>
<td>598,871</td>
<td>97,143</td>
</tr>
<tr>
<td>52211</td>
<td>Commercial banking</td>
<td>481,231</td>
<td>79,924</td>
</tr>
<tr>
<td>52212</td>
<td>Savings institutions</td>
<td>78,840</td>
<td>10,311</td>
</tr>
<tr>
<td>52213</td>
<td>Credit unions</td>
<td>37,397</td>
<td>6,503</td>
</tr>
<tr>
<td>52219</td>
<td>Other depository credit intermediation</td>
<td>1,404</td>
<td>404</td>
</tr>
<tr>
<td>5222</td>
<td>Nondepository credit intermediation (credit card issuers, leasing, etc.)</td>
<td>403,913</td>
<td>36,617</td>
</tr>
<tr>
<td>5223</td>
<td>Activities related to credit intermediation (loan brokerage, transaction processing, etc.)</td>
<td>58,342</td>
<td>14,451</td>
</tr>
<tr>
<td>523</td>
<td>Securities intermediation and related activities</td>
<td>325,184</td>
<td>105,549</td>
</tr>
<tr>
<td>52311</td>
<td>Investment banking and securities dealing</td>
<td>104,011</td>
<td>31,486</td>
</tr>
<tr>
<td>52312</td>
<td>Securities brokerage</td>
<td>104,812</td>
<td>36,428</td>
</tr>
<tr>
<td>52313</td>
<td>Commodity contracts dealing</td>
<td>3,905</td>
<td>835</td>
</tr>
<tr>
<td>52314</td>
<td>Commodity contracts brokerage</td>
<td>2,881</td>
<td>1,045</td>
</tr>
<tr>
<td>5232</td>
<td>Securities and commodity exchanges</td>
<td>3,213</td>
<td>721</td>
</tr>
<tr>
<td>5239</td>
<td>Other financial investment activities</td>
<td>106,363</td>
<td>35,034</td>
</tr>
<tr>
<td>52391</td>
<td>Miscellaneous intermediation</td>
<td>10,359</td>
<td>3,054</td>
</tr>
<tr>
<td>52392</td>
<td>Portfolio management</td>
<td>65,483</td>
<td>22,244</td>
</tr>
<tr>
<td>52393</td>
<td>Investment advice</td>
<td>15,098</td>
<td>5,473</td>
</tr>
<tr>
<td>52399</td>
<td>All other financial investment activities</td>
<td>15,423</td>
<td>4,263</td>
</tr>
<tr>
<td>524</td>
<td>Insurance carriers and related activities</td>
<td>1,294,941</td>
<td>120,683</td>
</tr>
<tr>
<td>52593</td>
<td>Real Estate Investment Trusts—REITs</td>
<td>22,386</td>
<td>1,559</td>
</tr>
</tbody>
</table>

**Source:** 2002 Economic Census.

**Note:** The economic census includes only REITs from NAICS (North American Industry Classification System) code 525, excluding, for example, pension funds.
The origination of securities accounted for $18.8 billion in revenue, which is primarily divided among investment banking ($11.7 billion), commercial banking ($4.4 billion), and brokerage ($2.4 billion) establishments. Proprietary trading yielded $40.9 billion, with commercial and investment banking establishments each accounting for about $17.5 billion. The profitability of proprietary trading, particularly by entities engaged in other client business, is a recent source of concern for regulators, for reasons discussed more following. Brokerage and related products such as investment research accounted for $201 billion in revenue, although less than half of this was earned by brokerage establishments, with commercial and investment banking establishments

5. Insurance companies also offer products with investment characteristics. A variable annuity is an investment product in its accumulation stage, while fixed annuities and whole life insurance also have investment aspects to them. Total annuity revenue for insurance establishments in 2002 was approximately $200 billion, but since this could not be separated into variable and fixed annuities, I exclude it from the revenue figures in this section.

6. As in other industries, an establishment refers to all the activities of a particular firm at a particular location. So, for example, if the small asset management arm of a large commercial bank was housed in its own location, it would be classified as a fund management establishment, while if it were housed in that bank’s headquarters building, it would usually appear as asset management activities of a commercial banking establishment. In some cases, however, the Census Bureau separates the activities of a common owner into multiple establishments even when they are collocated. The decision to do so appears related to the ease with which the activities can be cleanly separated. “When two or more activities were carried on at a single location under a single ownership, all activities generally were grouped together as a single establishment. The entire establishment was classified on the basis of its major activity and all data for it were included in that classification. However, when distinct and separate economic activities (for which different industry classification codes were appropriate) were conducted at a single location under a single ownership, separate establishment reports for each of the different activities were obtained in the census” (US Census Bureau 2005, A-1).
<table>
<thead>
<tr>
<th>Product line code</th>
<th>Product</th>
<th>Commercial banking (5221)</th>
<th>Investment banking (52311)</th>
<th>Securities brokerage (52312)</th>
<th>Investment management and advice (5239)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banking products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>550</td>
<td>Loan income</td>
<td>419,611</td>
<td>1,081</td>
<td>113</td>
<td>1,771</td>
<td>422,576</td>
</tr>
<tr>
<td>552</td>
<td>Nonloan credit products</td>
<td>324,557</td>
<td>1,081</td>
<td>113</td>
<td>1,771</td>
<td>327,522</td>
</tr>
<tr>
<td>570</td>
<td>Deposit accounts</td>
<td>38,456</td>
<td></td>
<td></td>
<td></td>
<td>38,456</td>
</tr>
<tr>
<td>571</td>
<td>Cash management</td>
<td>21,783</td>
<td></td>
<td></td>
<td></td>
<td>21,783</td>
</tr>
<tr>
<td>572, 573</td>
<td>Document payment products (i.e., cashier’s checks, money orders) and retail forex</td>
<td>2,244</td>
<td></td>
<td></td>
<td></td>
<td>2,244</td>
</tr>
<tr>
<td>Securities products</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Securities origination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5531</td>
<td>Public equity</td>
<td>88,021</td>
<td>89,880</td>
<td>100,472</td>
<td>100,428</td>
<td>378,800</td>
</tr>
<tr>
<td>5532</td>
<td>Public debt</td>
<td>4,353</td>
<td>11,680</td>
<td>2,362</td>
<td>357</td>
<td>18,754</td>
</tr>
<tr>
<td>5533</td>
<td>Private placement equity</td>
<td>3,400</td>
<td>5,256</td>
<td>1,556</td>
<td>177</td>
<td>10,389</td>
</tr>
<tr>
<td>5534</td>
<td>Private placement debt</td>
<td>849</td>
<td>5,085</td>
<td>584</td>
<td>109</td>
<td>6,627</td>
</tr>
<tr>
<td>5535</td>
<td>Private placement debt</td>
<td>13</td>
<td>630</td>
<td>84</td>
<td>69</td>
<td>796</td>
</tr>
<tr>
<td>5534</td>
<td>Private placement debt</td>
<td>92</td>
<td>709</td>
<td>138</td>
<td>2</td>
<td>942</td>
</tr>
<tr>
<td>Brokering and dealing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>554</td>
<td>Debt instruments</td>
<td>42,415</td>
<td>48,618</td>
<td>78,480</td>
<td>7,071</td>
<td>176,584</td>
</tr>
<tr>
<td>556</td>
<td>Equity</td>
<td>29,429</td>
<td>8,932</td>
<td>12,199</td>
<td>1,427</td>
<td>51,988</td>
</tr>
<tr>
<td>557</td>
<td>Derivatives</td>
<td>1,723</td>
<td>27,948</td>
<td>51,474</td>
<td>3,502</td>
<td>84,647</td>
</tr>
<tr>
<td>558, 560, 561</td>
<td>Other products (currency, commodity pools, correspondent products)</td>
<td>6,324</td>
<td>3,758</td>
<td>2,615</td>
<td>543</td>
<td>13,241</td>
</tr>
<tr>
<td>562 to 564</td>
<td>Financing related to securities (securities lending, repurchase agreements)</td>
<td>1,298</td>
<td>200</td>
<td>8,004</td>
<td>585</td>
<td>10,087</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(continued)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9.3  (continued)

<table>
<thead>
<tr>
<th>Product line code</th>
<th>Product</th>
<th>Commercial banking (5221)</th>
<th>Investment banking (52311)</th>
<th>Securities brokerage (52312)</th>
<th>Investment management and advice (5239)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proprietary trading</td>
<td></td>
<td>17,715</td>
<td>17,299</td>
<td>3,438</td>
<td>2,463</td>
<td>40,915</td>
</tr>
<tr>
<td>565</td>
<td>Debt instruments</td>
<td>12,675</td>
<td>11,950</td>
<td>552</td>
<td>556</td>
<td>25,733</td>
</tr>
<tr>
<td>566</td>
<td>Equities</td>
<td>905</td>
<td>2,104</td>
<td>658</td>
<td>850</td>
<td>4,518</td>
</tr>
<tr>
<td>567</td>
<td>Derivatives</td>
<td>2,739</td>
<td>2,636</td>
<td>2,112</td>
<td>313</td>
<td>7,800</td>
</tr>
<tr>
<td>568,569</td>
<td>Other</td>
<td>1,397</td>
<td>610</td>
<td>115</td>
<td>743</td>
<td>2,864</td>
</tr>
<tr>
<td>Trust, asset management, and financial planning</td>
<td></td>
<td>21,727</td>
<td>12,215</td>
<td>15,699</td>
<td>88,659</td>
<td>138,300</td>
</tr>
<tr>
<td>574</td>
<td>Trust fiduciary fees</td>
<td>14,098</td>
<td>135</td>
<td>116</td>
<td>10,559</td>
<td>24,909</td>
</tr>
<tr>
<td>577</td>
<td>Financial planning and investment management</td>
<td>7,629</td>
<td>12,080</td>
<td>15,583</td>
<td>78,100</td>
<td>113,391</td>
</tr>
<tr>
<td>5771</td>
<td>Businesses and governments</td>
<td>4,691</td>
<td>7,652</td>
<td>3,446</td>
<td>26,164</td>
<td>41,954</td>
</tr>
<tr>
<td>5772</td>
<td>Individuals</td>
<td>2,937</td>
<td>4,428</td>
<td>12,137</td>
<td>51,935</td>
<td>71,437</td>
</tr>
<tr>
<td>Other products</td>
<td></td>
<td>93,049</td>
<td>13,175</td>
<td>4,719</td>
<td>6,042</td>
<td>116,986</td>
</tr>
<tr>
<td>575,576</td>
<td>Financial market clearing products and ACH</td>
<td>1,810</td>
<td>125</td>
<td>493</td>
<td>1,878</td>
<td>4,305</td>
</tr>
<tr>
<td>578</td>
<td>Other products</td>
<td>91,239</td>
<td>13,050</td>
<td>4,227</td>
<td>4,164</td>
<td>112,681</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>598,871</td>
<td>104,011</td>
<td>104,812</td>
<td>106,363</td>
<td>914,057</td>
</tr>
</tbody>
</table>

Source: 2002 Economic Census.

Note: Bold type denotes the scope of this chapter.
dividing the other half roughly evenly. Asset management and financial planning accounted for $138 billion. Both managing a collective investment vehicle such as a mutual fund and providing investment advice to an individual investor are regarded “investment advice,” and revenues from the two sources are not distinguished in the Economic Census. Of the $138 billion, $71 billion is earned by establishments engaged in portfolio management or investment banking establishments, while the remaining $67 billion is earned by investment advising, commercial banking, brokerage, and investment banking establishments. The former is presumably largely portfolio management, while the latter is presumably largely financial advising to retail customers.

The $400 billion in total revenue represents about 3.6 percent of GDP and 2 percent of the stock of financial market assets held by households. As with health care, it is impossible to infer over- or underspending from the $400 billion number alone, but these figures are useful in roughly sizing the economic importance of the sector. Financial services play a special role in capital formation in an economy, and inefficiency in financial services can have disproportionate effects on welfare. For example, suppose that this 3.6 percent of GDP either includes 1 percent of GDP in pure waste or, alternatively, reflects an underinvestment in intermediation services that leads to a misallocation of capital that leads to a net waste of 1 percent of GDP.

Recall from the Solow (1956) growth model that the steady state capital-output ratio is equal to \( s/(d + n + g) \), where \( s \) is savings as a percent of GDP, \( d \) is the depreciation rate, \( n \) is population growth, and \( g \) is total factor productivity growth. Taking reasonable values for the last three parameters of \( d = 5\% \), \( n = 1\% \), and \( g = 2\% \) implies that waste or misallocation that lowers the savings rate by 1 percent of GDP reduces the steady state capital-output ratio by 12.5 percent. Assuming Cobb-Douglas production with a capital share of 0.3, this lowers steady state output per capita by about 6 percent.

This 6 percent reduction, which occurs over time as a lower net savings rate leads to slower accumulation of capital, is in addition to the direct waste

7. The census year 2002 was a trough year for the securities industry. Data from the Service Annual Survey for 2000 to 2004 reveal that revenue for securities firms with employment (about 90 percent of the total by revenue) fell from $385 billion in 2000 to $293 billion in 2002 and then recovered to $349 billion in 2004. Product-level data suggests that the declines and recoveries in underwriting, proprietary trading, brokerage, and asset management/advice were roughly proportional. The stock of financial market assets held by households is calculated as $19.6 trillion for 2002 by taking total financial assets ($29.7 trillion) less bank deposits ($4.0 trillion), equity in noncorporate business ($5.2 trillion), and insurance reserves ($0.9 trillion). Source: Federal Reserve, Flow of Funds Accounts for the United States, table L100.

8. For comparison, French (2008, 1537) estimates that “the fees and expenses paid for mutual funds, the investment management costs paid by institutions, the fees paid to hedge funds and funds of funds, and the transaction costs paid by traders” account for about 0.75 percent of the value of equity outstanding. The difference between this and the 0.09 percent that French estimates would be that the cost of passive investing corresponds to just over 1 percent of GDP.

9. If the Cobb-Douglas production function is \( Y = A * K^a * L^{(1-a)} \), then \((Y/L) = (Y/K)^{[a/(1-a)]} * A^{[1/(1-a)]}\).
of the 1 percent of GDP. In short, the most common argument made in favor of lower taxation of capital, that it leads to more capital accumulation and thus higher returns to labor, can also be made in favor of an efficient financial sector. The desirability of an efficient financial sector is usually uncontroversial, however. The key questions are where are the inefficiencies and whether better designed regulation can address them?

9.3 The Economic Rationale for Securities Regulation

A necessary, but not always sufficient, condition for regulation to be useful is for the unregulated competitive equilibrium to not be socially optimal. The standard conditions for a competitive equilibrium to be optimal are full and symmetric information, rational agents, the absence of externalities, and competitive behavior. Each of these conditions is arguably unmet in at least part of the securities industry, giving rise to potential rationales for regulation.

9.3.1 Imperfect Information

Gathering value-relevant information about securities is costly, leading even rational investors to choose to be not fully informed in equilibrium (Grossman and Stiglitz 1980). Delegation to expert agents is often the response, and as discussed earlier, a typical investor invests via multiple layers of agents: a financial advisor to select investment managers, investment managers to select securities and exercise any voting rights, brokers to trade those securities, and managers of the securities’ issuers to produce returns.

As alignment of incentives in these agency relationships is often imperfect, delegation and imperfect information can give rise to moral hazard problems. Financial advisors may prioritize recommending the investment products of their employer (Christofferson, Evans, and Musto 2005). They may also make recommendations based on sales commissions or other considerations provided by the fund family. Managers of investment products

10. This simple Solow model exercise actually understates the importance of an efficient and effective financial sector in several ways. Savings is exogenous in the Solow model; if savers react to financial-sector inefficiencies by reducing their saving, the effects could be greater. In addition, the Solow model assumes that capital accumulation and technological progress are independent. It thus ignores the fact that new firms and new vintages of capital equipment are a primary means through which new technologies are developed and deployed, respectively.

11. For an example of the former, see Council of Economic Advisers (2003, chap. 5).

12. This prioritization is sometimes less straightforward to detect than one might assume, given the practice of using investment product brands that differ from the brand under which the adviser conducts business but are nonetheless owned by the same firm (e.g., an advisor in a First Union bank branch selling an Evergreen mutual fund).

13. Generally, doing so is legal so long as the payments are disclosed. Recent regulatory cases illustrate the boundaries of what is permitted. Some advisors sold higher commission “B share” versions of a mutual fund to investors without disclosing that lower commission “A share” versions were available (e.g., SEC administrative case 3-11179 against IFG Network Securities). Others sold funds from an approved list that mutual funds made undisclosed payments to
may engage in activities that lower the returns of their portfolios but benefit them privately, such as tolerating arbitrage trading in exchange for investments in other high-fee funds (Zitzewitz 2003). They may also reallocate returns into portfolios with higher incentive-based compensation (Cici, Gibson, and Moussawi 2006) or more return-sensitive investors (Gaspar, Massa, and Matos 2006) through a variety of techniques. Brokers may place trades in a manner that reduces clients’ execution quality but provides them with a private benefit.\(^{14}\) And, of course, issuing firms’ managers may manage in a way that places their own interests ahead of their shareholders.

Regulation can seek to limit agency problems in several ways. Merit regulation proscribes certain practices, investment products, or fee levels. Examples include laws prohibiting front running (trading in advance of one’s clients to profit from the impact their trades have on prices) or Ponzi schemes. Antifraud regulation can help make voluntary disclosure about agents’ performance and practices credible by prosecuting agents who lie. Mandatory disclosure regulation can both require disclosure of performance and practices and impose standards to make such disclosures more comparable.

Imperfect information can also give rise to adverse selection problems. In the absence of regulation, agents who engage in behavior that benefits themselves at the expense of their clients may be more profitable and if so will have more incentive to market their products aggressively. This can create a lemons problem (Akerlof 1971) in which bad agents and products drive out good ones. Regulation can potentially help, again by either mandating disclosure or prohibiting practices.

9.3.2 Investor Behavior

The question of whether efforts to protect consumers should focus on disclosure requirements or on merit regulation that restricts products and behavior is a central debate in financial regulation. As discussed later, the SEC generally favors the former, while the state regulators who enforce antifraud statutes tend to take the latter approach. As Zingales (2004) emphasizes, an advantage of disclosure regulation is that its costs are usually smaller than those of merit regulation, which risks limiting innovation.

A problem, though, is that for the unsophisticated investors most in need of protection, the benefits of additional disclosure may be small too. Whether regulation requiring disclosure is effective depends on whether

\(^{14}\) Examples include internally matching client orders with proprietary or favored-client trades at disadvantageous prices, bundling orders with informed order flow from hedge fund clients or proprietary traders, routing orders to lower volume exchanges in exchange for payments for order flow, and illegally front running clients’ transactions.
investors can make use of the information. When buying securities or investments, many consumers are unaware of the most basic information that is disclosed. In addition, they exhibit behavioral biases, particularly naïveté about the incentives of experts.

For many financial products, the majority of customers do not understand the rather central concept of a “price.” For example, in a 2002 survey by Vanguard and *Money* magazine, only 25 percent of investors correctly identified the expense ratio as the annual fee they pay for a mutual fund (on a multiple choice question with no guessing penalty). Likewise, an OCC/SEC survey reported on by Alexander, Jones, and Nigro (2001, 164) found that only 19 percent of mutual fund investors reported knowing the (approximate) expense ratio of their largest fund investment. Hortacsu and Syverston (2004) find that a large proportion of investors choose S&P index funds as if they had very high search costs. An alternative interpretation of their results would be that investors observe price imperfectly when choosing their funds (Busse, Elton, and Gruber 2004), or misunderstand the strong negative relationship between fees and after-fee performance (Carhart 1997 and others). Barber, Odean, and Zheng (2004) find that investors react more to fees that are salient, such as front-end sales commissions that are deducted from their investment at time of purchase, than to fees that are less salient, such as expenses or deferred commissions that are deducted over time. Choi, Laibson, and Madrian (2010) find that undergraduate and MBA students at top schools fail to choose the lowest expense ratio index fund even when furnished with information on fees, in part because of a belief that past returns are informative about future returns.

One might think that investors do not need to understand expense ratios, since expense ratios are deducted from net returns, and investors should care primarily about net returns. The problem with this logic is that past net returns are almost uncorrelated with future net returns, and a low expense ratio is by far the best single predictor of high future returns. Mutual funds are unlike many other products in that future “quality” (at least as measured by before-expense returns) is close to uncorrelated with past quality, but quality is negatively related to price. Studies of investor demand for funds (Chevalier and Ellison 1997 and others) suggest that many investors appear to invest as if they expect quality (returns) to be positively serially

---

15. In contrast, a recent Investment Company Institute (2006) survey found that 74 percent of investors claimed to have reviewed the expense ratio before making their most recent mutual fund investment. Possible reconciliations of these results include: (a) investors may have become more sensitive to fees since 2002; (b) the ICI sample was more sophisticated than the Vanguard-Money or OCC-SEC samples; or (c) the ICI survey asked whether investors had review the fund’s fees, but did not test this knowledge.

16. Mutual funds are unlike many other products in that quality (at least as measured by after-expense returns) is close to uncorrelated with past quality, but quality is strongly negatively related to price. Studies of investor demand for funds (Chevalier and Ellison 1997 and others) suggest that many investors appear to invest as if they expect quality (returns) to be positively serially correlated and to get more when they pay more, as one might expect if these investors were applying their experience from other products to mutual funds.
correlated and to get more when they pay more, as one might expect if these investors were applying their experience from other products to mutual funds (Mullainathan, Schwartzstein, and Shleifer 2008).

A similar percentage in the Vanguard-Money survey misunderstood loads (sales commissions paid to the broker who sells a fund). Along with the salience issue discussed by Barber, Odean, and Zheng, this might help explain the recent popularity of “B” shares, in which the broker’s commission is deducted gradually from shareholder’s assets as opposed to being deducted from their investment upfront. As mentioned earlier, it is alleged that brokers misrepresent “B” shares as being no-load funds or steer investors into “B” shares where there are lower commission alternatives.

Among investment products, however, mutual fund fees are perhaps the most transparent. Variable annuities carry a variety of fees that are in many cases collectively large enough to pay sales commissions of 5 to 10 percent of the amount invested. In July 2004, the New York Times reported on the sales of a set of extremely disadvantageous contractual mutual fund and life insurance products on military bases (Henriques 2004). In both types of products, the fees that finance sales commissions are not deducted from an investor’s investment upfront in a transparent manner, but instead are spread across various administration fees, expenses charged to the underlying investments, and fees for death benefits that are well above the cost of a comparable amount of term life insurance.

In brokerage accounts, many investors understand commissions, and are less likely to understand other trading costs such as the bid-ask spread and how it is affected by order handling rules. When investors buy bonds from a brokerage at no commission, many do not realize that the brokerage is charging a markup that usually exceeds the commission on comparably sized stock transactions (see, e.g., Harris and Piwowar 2006). Likewise, when investors buy shares in public offerings, some are unaware that the company is paying an underwriting commission on the proceeds, creating a wedge between the amount they pay and the funds that management is able to invest on their behalf. Some have argued (e.g., Lee, Shleifer, and Thaler 1991) that the fact that investors buy closed-end fund IPOs at a premium to net asset value despite the fact that these funds typically trade at a discount several months later provides an example of investors misunderstanding these issues.

Apart from difficulty understanding prices, the field of behavioral finance has documented a variety of psychological biases that affect consumers when making financial decisions. Investors, especially males, trade too

17. See, for example, Securities and Exchange Commission (2004, 2).
18. A full review of the field is well beyond the scope of this chapter—Shefrin (2002), Barberis and Thaler (2003), and Shiller (2003) provide excellent summaries. The findings of behavioral finance about consumer behavior in this industry has motivated some to consider the implications of boundedly rational consumer behavior in other industries, see, for example Gabaix and Laibson (2006).
frequently (Shefrin and Statman 1994; Odean 1998). Investors also react to news inefficiently. At short-to-medium time horizons (e.g., one year) investors suffer from the disposition effect, holding on to losing investments too long and selling winners too quickly (Shefrin and Statman 1985). This is the reverse of what would be optimal given the tax treatment of capital gains and the long-standing findings of momentum in stock prices at the one-year time horizon (Jegadeesh and Titman 1993). Investors also display the disposition effect in their mutual fund investments, holding on to underperforming mutual funds despite the fact that these funds tend to repeat their underperformance (Carhart 1997; Kacperczyk, Sialm, and Zheng 2008). A psychological reason for avoiding selling a losing investment is that it creates cognitive dissonance—booking a loss is an acknowledgement that the initial investment was a mistake.19 Firing a financial advisor that one once trusted requires a similar acknowledgement and creates a stickiness that some advisors may exploit.

Many investors also appear to be excessively influenced by and naïve about the incentives of financial advisors, equity analysts, and the financial media. Across a variety of metrics, financial advisors choose funds for their clients that are no better than the funds no-load investors choose for themselves (Bergstresser, Chalmers, and Tufano 2009), and advisors are particularly unlikely to advise a client to sell a persistently underperforming fund offered by their employer (Christofferson, Evans, and Musto 2005). Alexander, Jones, and Nigro (2001) report that many investors have misconceptions about the sign of the correlation between expenses and future returns, the degree of persistence in mutual fund returns, and whether money market funds are FDIC insured, and that in some cases they acquire these misconceptions from their financial advisors.20 One of the strongest predictors of mutual fund inflows is high 12b1 fees; 12b1 fees are collected from investors and mostly used to finance payments to the brokerage or advisor that recommended the fund (Reid and Rea 2003).21 Mutual fund recommenda-

19. Investors overreact to positive news at longer time horizons (e.g., three to five years), buying stocks that have performed well in the last three to five years and pushing up their prices to the point where they underperform in the future (De Bondt and Thaler 1985 and 1989). This can also be rationalized as being due to cognitive dissonance if investors window dress their own portfolios, removing long-term losing stocks and buying stocks they wish they had bought earlier.

20. For example, 35 percent of investors in money market mutual funds who used a broker believe that these funds are insured, and 23 percent of those report being told this by their broker (180). The number of investors who believe in a positive relationship between expenses and returns outnumber those who believe in a negative relationship (19.9 percent to 15.7 percent); the margin widens to 21.0 to 14.0 for investors who invest only through intermediaries (banks, brokers, insurance companies, or retirement plans) (165). Twenty-four percent of investors expect a fund with a good performance in the previous year to have above average performance in the next year (166).

21. A 12b1 fee refers to a fee that a fund can charge its shareholders to pay for the marketing and distribution of fund, authorized under rule 12b1 promulgated by the SEC under the Investment Company Act of 1940. The fund’s trustees must conclude that doing so is in the
tions in personal finance magazines are associated with significant future inflows, despite the fact that positively mentioned funds perform no better than average in the future and that mentions are correlated with a fund family’s past advertising (Reuter and Zitzewitz 2006). Investors in stocks react to media reports, even when they contain no new information. One of the most famous examples is the fourfold increase in the stock price of EntreMed that followed a front-page *New York Times* story, despite the fact that the potential breakthrough in cancer research highlighted in the article had been published in *Nature* and written up in other newspapers (including the *Times*) over the prior five months (Huberman and Regev 2001). CEO interviews on CNBC from 1999 to 2001 were accompanied by a 1.65 percent stock price appreciation that mean reverted over the next day (Kim and Meschke 2011; see also Busse and Green 2002). The discounts of foreign closed-end funds (the difference between the price of a fund and the value of its underlying assets) react to whether and how extensively foreign news is reported in the US press (Klibanoff, Lamont, and Wizman 1998). Media-savvy issuers appear to exploit these biases, by directing media attention to the most favorable earnings metric (Dyck and Zingales 2005) and by announcing bad news on Friday afternoons (Bagnoli, Clement, and Watts 2004; DellaVigna and Pollet 2009). Investors’ reliance on the media has also been exploited to include trading in advance of media coverage and the use of the media to manipulate asset prices.22

There are limits to the extent to which regulation can protect investors from their own biases or a lack of sophistication. As with regulation designed to address information problems, regulatory responses to investor behavior have generally taken two different approaches. First, merit and antifraud regulations protect the least sophisticated investors by restricting the availability of certain types of securities or financial services that are viewed as particularly abusive (e.g., Ponzi schemes) and limiting others to sophisticated investors (e.g., hedge funds). Both the SEC and self-regulatory bodies such as the National Association of Securities Dealers (NASD) regulate the behavior of investment professionals such as stockbrokers and investment advisors, particularly the exploitation of investors’ naiveté and biases. Second, regulations force the disclosure of certain characteristics of issuers and investments to ensure that sophisticated investors have access to a certain interests of shareholders, for example, by generating enough asset growth to allow a management fee reduction that more than offsets the fee. While there is strong evidence that 12b1 fees are correlated with inflows, some have questioned whether this growth leads to reduction in management fees sufficient to provide a net benefit to shareholders (e.g., Walsh 2004). As a result, more recent justifications of 12b1 fees have argued that they benefit shareholders because they are used to pay brokers for services provided to shareholders.

22. Examples include the insiders who provided tips on the content of the *Wall Street Journal*’s Heard on the Street and *Business Week*’s Inside Wall Street columns and financial columnists who have allegedly recommended stocks they hold positions in.
minimum level of information and in some cases mandate certain standardized formats to increase the salience of the information to investors.\(^{23}\)

9.3.3 Externalities

Two examples of externalities that potentially provide a rationale for regulation are free riding in monitoring and so-called preference externalities. Monitoring corporate or investment managers generates benefits that are shared by other investors. It therefore suffers from a potential free rider problem. This free rider problem is partially addressed through three mechanisms. First, the pricing of securities or investments in the secondary market can create an incentive for a shareholder to acquire a large stake and then monitor management, internalizing the benefits of their monitoring in proportion to their stake. As an example, investors have recently purchased stakes in underperforming closed-end mutual funds at a discount and then forced management to redeem all or some of their shares at net asset value. Some of the benefits of this form of monitoring spill over to the other shareholders of the fund, suggesting that it will be underprovided by the market.

Second, shareholders exert monitoring through boards of directors. Corporate security issuers have boards of directors that monitor management. Investments such as mutual funds are formally organized as companies, and they are required to have a board of directors whose responsibilities include hiring, monitoring, and negotiating fees with the investment manager. The desired level of independence of both corporate and investment company board members is a matter of active debate. Tufano and Sevick (1997), Del Guercio, Dann, and Partch (2003), and Zitzewitz (2003) provide evidence that investment company board independence is correlated with shareholder-friendly fee and valuation policies. The SEC recently issued a rule requiring that 75 percent of investment company board members and the board chair be independent of employment or other business relationships with the investment manager, although this rule has been challenged in court. On the other hand, some question whether boards are necessary for investments (Tkac 2004), invoking the fact that investment companies in other countries do not have boards (Damato, Reilly, and Richardson 2004). Khorana, Servaes, and Tufano (2005 and 2009) compare mutual fund industries across countries, finding that the industry is larger and fees are lower in countries with stronger investor protections, including boards of directors.

Third, the media and other third-party experts can potentially play a monitoring role. Media publications motivated by subscription revenue or analysts interested in building followship in the markets have an incen-

---

\(^{23}\) For example, the SEC requires that mutual fund prospectuses contain at their beginning a “Risk-Return Summary” that includes information on fees and past performance. It also regulates the reporting of past performance to limit the extent to which fund companies can distort their track records by manipulating the time period reported.
tive to provide high-quality information to their clients. At the same time, these experts may have other, conflicting motivations. As mentioned earlier, Reuter and Zitzewitz (2006) find a correlation between the mutual fund recommendations of personal finance magazines and past advertising. They also find that the publications overweight past returns and underweight fees when determining which fund to recommend, which might be regarded as a form of proindustry bias. Lin and McNichols (1998) and Michaely and Womack (1999) find a correlation between analysts' security recommendations and their employer's underwriting business, and this relationship has been extensively probed by regulators in recent years.

Investor's preferences can also impose externalities on other investors. George and Waldfogel (2003) argue that when newspaper readers are homogeneous they create positive externalities for one another by enlarging the market and generating scale economies. When they are heterogeneous, however, they can generate negative externalities. George and Waldfogel (2006) provide an example, arguing that the entrance of the New York Times to a newspaper market “spreads ignorance and apathy” by attracting educated readers away from the local paper, making it optimal for the local paper to reduce national coverage and appeal to less-educated readers. Both externalities are present in securities and investment markets. In investments, Vanguard arguably plays the role of the New York Times, attracting expense ratio–sensitive investors and lowering the average fee sensitivity of other firms' clients.

Regulation can and does address these externalities in several ways. On some issues regulators play the role of monitor themselves, by enforcing rules against certain behavior. By mandating boards and regulating the independence and election of their members, regulators can make the collective action problem cheaper for investors to solve. By mandating disclosure, regulators can facilitate the monitoring roles of both boards and outsiders such as analysts and the media. Regulators can also address externalities arising from the bifurcation of markets into products targeting sophisticated and unsophisticated investors, either by aiding investors who seek to become sophisticated or by limiting the exploitation of the unsophisticated.

24. In contrast, Miller (2006) finds that the media's coverage of accounting fraud in an industry is not related to the industry's propensity to advertise.

25. Arguably another potential example of proindustry bias in the financial media is that fact that academic studies documenting the extent of stale price arbitrage in mutual funds (e.g., Goetzmann, Ivkovic, and Rouwenhorst 2001; Greene and Hodges 2002; Zitzewitz 2003) were known to reporters at major publications, and yet they were discussed extremely rarely until the announcement of New York attorney general (NYAG) Eliot Spitzer's investigation in September 2003. Two notable exceptions were Stone (2002) and Carnahan (2003), although it should be noted that even these articles appeared only in the online editions of Business Week and Forbes, respectively. Other articles discussed the issue, but framed it in a way that buried the lead (e.g., “Monitoring Trades for the Good of the Fund,” New York Times, April 9, 2000). After the announcement of the NYAG's investigation, the financial media did report on the issues thoroughly.
9.3.4 Competitive Behavior

Most financial industries have free entry and large numbers of competitors, and so there is a temptation to assume that they are close to perfectly competitive. At the same time, some of the institutional features that industrial economists normally associate with soft competition are present in these industries. Especially following the relaxation and ultimately the repeal of the separations between commercial banking, insurance, and securities in the Glass-Steagall Act, many financial services firms compete against each other in multiple markets, which can facilitate soft competition (Bernheim and Whinston 1990). In addition, agency relationships (e.g., steering financial advisory clients toward or away from a competitor’s offerings) may provide an inexpensive means of rewarding or punishing a firm for behavior in another market. In many settings, prices or fees are readily observable to one’s competitors, making secret discounting more difficult to implement. For example, underwriting fees are disclosed in offering documents, investment fees are disclosed in prospectuses, and spreads charged market makers are readily observable by other market makers. One should not necessarily expect free entry to lead to tough competition; as Hsieh and Moretti (2003) illustrate in their study of residential real estate brokerage, free entry can be consistent with established firms earning economic rents, although some of the rents may be wastefully dissipated through nonprice competition and business stealing effects.

A market failure is a necessary condition for regulation to be optimal, but it is not always sufficient. Market imperfections must be weighed against the imperfections of the legislative and regulatory institutions responsible for rulemaking and enforcement. This motivates turning to a discussion of the main laws and institutions of US financial regulation.

9.4 The Main Laws and Institutions

The core of modern federal financial regulation is formed by four laws passed during the Great Depression: the Securities Act of 1933 (the “1933 Act”), the Securities Exchange Act of 1934 (the “1934 Act”), the Investment Company Act of 1940 (the “1940 Act”), and the Investment Advisors Act of 1940.26 These four acts each regulate a stage in the value chain previously discussed: respectively, they regulate the issuance of securities,27 the brokerage and secondary trading of securities and the ongoing disclosure requirements of their issuers, investment companies (open and closed-end mutual funds),

26. This brief overview of securities regulation draws heavily on Coffee and Seligman (2002), who I refer readers to for more detail.

27. Along with the 1933 Act, the Trust Indenture Act of 1939 also governs the issuance of bonds.
and investment advisors (including both advisors who manage client assets directly as well as those who manage the assets of investment companies).

The 1933 Act requires the registration of securities with the SEC (subject to certain exemptions, e.g., for private placements that are not made available to the public) and requires the delivery of a prospectus to investors. Given that investors have a favorable cause of action if the issuer makes materially misleading statements or omissions in its offering documents, the disclosure in offering documents is generally much more extensive than ongoing disclosure by issuers. This generates two substantial costs to an initial offering of securities: (1) the fees and other costs associated with generating and delivering these documents, and (2) the competitive costs of the extensive disclosure of business information that is usually involved.

The 1934 Act establishes annual and quarterly disclosure requirements for companies, requires SEC preclearance of proxy statements for shareholder votes, and establishes a self-regulatory system for stock exchanges and brokers. The stock exchanges and the NASD, which self-regulates stockbrokers, are both overseen by the SEC. The 1934 Act (also referred to as the “Exchange Act”) gives the SEC broad rulemaking authority to prescribe practices of broker-dealers as “manipulative, deceptive, or otherwise fraudulent.” The 1934 Act has been amended by Congress multiple times—examples include the 1964 Securities Acts Amendments (which extended disclosure requirements to large over-the-counter [i.e., public, but not stock exchange–listed] firms); the 1970 amendment creating the Securities Investor Protection Corporation (which provides FDIC-like insurance for brokerage accounts); the Securities Act Amendments of 1975 (which deregulated brokerage commissions); the Foreign Corrupt Practices Act of 1977 (prohibiting bribery by public companies); the Insider Trading Sanctions Act of 1984 and Insider Trading and Securities Fraud Enforcement Act of 1988; the Private Securities Litigation Reform Act of 1995 (which sought to limit certain types of shareholder class action lawsuits); and the Sarbanes-Oxley Act of 2002 (SOX).

The Sarbanes-Oxley Act has been both controversial and an active current research topic and thus merits additional discussion. Most provisions of SOX appear to be a direct response to specific accounting abuses at firms such as Enron and Worldcom. SOX creates a self-regulatory body to regulate the accounting profession, restricts the provision of consulting and other services by an audit firm to an audit client, and requires the rotation of the lead audit partner every five years. For issuers, SOX requires audit committees to be composed entirely of independent directors and requires CEOs and CFOs to certify the firm’s accounting numbers and face disgorgement of compensation and stock trading profits and criminal sanctions for misleading earnings or knowingly false statements. SOX requires the SEC to develop rules requiring companies to report on the adequacy of internal controls, rules requiring attorneys appearing before the SEC to report
security laws violations, and rules governing the independence of security analysts. It also tightens rules on stock trading by directors and executives, extends the statue of limitations for securities fraud, and enhances protections for corporate whistleblowers.

SOX has been heavily criticized by the business community and some scholars for making external and internal auditing more expensive and onerous. Eldridge and Kealey (2005) report that average audit fees for a sample of 648 Fortune 1000 companies increased from $3.5 million to $5.8 million from 2003 to 2004, and they attribute most of this increase to SOX. Leuz, Triantis, Wang (2008) and Carney (2006) argue that costs associated with SOX may have encouraged some firms to delist. Against this cost is the benefit firms with clean accounting received from a restored investor confidence. Li, Pincus, and Rego (2008) and Rezaee and Jain (2006) found positive stock price responses to the act. Engel, Hayes, and Wang (2007) find more positive event returns for larger firms, as one might expect given that the costs of SOX increase more slowly with firm size than the benefits. This early evidence suggests that, for better or worse, SOX has significantly “raised the bar” for being a public company.

The 1940 Act regulates open- and closed-end mutual funds. Mutual funds are far more important than when the 1940 Act was passed: in 2003 equity mutual funds accounted for 19.7 percent of household equity holdings and money market funds accounted for 21.2 percent of household holdings of cash equivalents (demand deposits, time deposits, etc.). The 1940 Act contains provisions designed to protect shareholders from dilution by fund managers. It requires that investment companies have a board of trustees, that they annually review the management contract for the fund, and that a majority of these trustees be independent of the investment advisor. It establishes the fiduciary duties of the trustees and the investment advisor. It also establishes rules governing transactions in shares of open-end mutual funds designed to ensure that investors transact at prices that reflect fair market values.

Although the 1940 Act does include some regulation of behavior, like the 1933 Act and 1934 Act, it relies primarily on disclosure. As Jackson (1997,
Retail Securities Regulation in the Aftermath of the Bubble

535) puts it: “the 1940 Act relies on disclosure-based regulation more than any other comparable regulatory structure in the United States.” This is notable in that whereas the 1933 and 1934 Act regulate securities markets where arbitrage ensures that sophisticated investors will have significant influence on asset prices, the 1940 Act regulates investments that are designed primarily for unsophisticated investors. Mutual funds cannot be sold short, and so market efficiency requires that full information and rationality be possessed by all investors, not merely a relatively small number with access to sufficient arbitrage capital.

Finally, the Investment Advisors Act requires registration of investment advisors managing a substantial amount of client assets in either investment companies or separate accounts. It also prohibits fraud and certain deceptive practices and limits the circumstances under which the advisor can receive incentive compensation. Until recently, SEC rules exempted advisors with a limited number of “accredited” (i.e., wealthy enough to be assumed to be sophisticated) clients from registration. The Dodd-Frank Act eliminated this exemption in 2012 and required sufficiently large advisors to hedge funds and other private funds to register with the commission.

The SEC has the primary responsibility for enforcing and promulgating new rules under these acts.31 It is organized around these acts, with the Division of Corporate Finance having primary responsibility for the 1933 Act, the Division of Market Regulation for the 1934 Act, and the Division of Investment Management for the 1940 Act and Investment Advisors Act. These divisions support the commission in its two major channels for policy making: the promulgation of new rules under the acts and responding to parties requesting that the commission take “no action” against a novel practice. Enforcement is handled by its own division, and these four divisions are supported by functional offices (the Office of General Counsel, Office of Chief Accountant, and Office of Economic Analysis).

The SEC grew considerably in the immediate aftermath of Sarbanes-Oxley, in terms of both staff and budget (figure 9.3). It also engaged in a significant amount of new rulemaking. A number of the more important new rules have involved increased disclosure by investment companies and advisors.32 In some cases, enhanced disclosure requirements were adopted as a compromise in lieu of either direct dictation of practices (e.g., on fair

31. The SEC was also charged with enforcing the Public Utility Holding Company Act of 1935, which restricted interstate and nonregulated holdings of regulated utility companies, before its repeal by the Energy Policy Act of 2005. Even before then, Coffee and Seligman (2002, 70) had noted that this is “no longer an important statute because the SEC has largely deregulated the field” through rulemaking and enforcement policy.

32. For example, investment advisors are now required to disclose how they voted shareholder proxies (SEC Rule IA-2106). Investment companies are required to disclose their after-tax returns (33-8010) and to provide information about portfolio managers, including the factors used to determine their compensation (33-8458), about how the trustees determined the appropriateness of management fees (33-8433), about the availability of front-load commission discounts (33-8427), and about their policies regarding market timing, fair value pricing, and selective disclosure of portfolio holdings (33-8408).
value pricing) or more meaningful disclosure (e.g., of portfolio manager salaries, as opposed to the factors used to determine them). In addition, certain practices that were viewed as harmful to shareholders have been prohibited\textsuperscript{33} and fiduciary duties have been clarified.\textsuperscript{34} The SEC has promulgated rules as needed to implement SOX. It has also used rulemaking to implement decimalization, to relax short-selling rules (Regulation SHO), and to limit selective disclosure by companies, particularly to equity analysts (Regulation FD).

As mentioned before, the 1934 Act provides for the SEC to delegate primary regulatory authority to self-regulatory organizations (SROs): the stock exchanges self-regulate themselves, Finra (formerly the NASD) regulates its broker-dealer members, and the new Accounting Oversight Board created by SOX regulates the accounting profession. In each case, the SEC

\textsuperscript{33} For example, investment companies are now prohibited from directing brokerage commissions to firms as a reward for selling fund shares (SEC Rule IC-26591), as this was viewed as fund advisors using shareholders’ assets to reward brokers for an activity that primarily benefits the advisor.

\textsuperscript{34} For example, SEC Rule IA-2106 requires that investment companies vote shareholder proxies in their own shareholders’ interest. Although fund trustees already had a fiduciary responsibility to ensure that advisory and other fees charged to a fund were appropriate, and SEC Rule 33-8433 formally only requires additional disclosure of the basis of that decision, some have argued that in practice it is likely to reinforce trustees’ fiduciary responsibilities in this area.
Retail Securities Regulation in the Aftermath of the Bubble

holds ultimate regulatory authority. A similar structure exists for derivatives, where the Commodity Futures Trading Commission (CFTC) acts as the ultimate regulator, but delegates self-regulatory authority to exchanges such as the Chicago Mercantile Exchange (CME). Generally, cooperation in this system is amicable, but there are exceptions, with the forced replacement of the NASD leadership following the Nasdaq market maker collusion scandal being a prime example.

Before modern federal securities regulation began in the 1930s, most states had their own regulations. These are often called “blue sky” laws, and they typically focus on the prevention of fraud by brokers, investment advisors, and securities issuers. They require registration by brokers and advisors and of newly offered securities, and the resulting registration fees provide a source of revenue that no states choose to forego. Apart from their revenue collection role, state securities laws declined in importance in the 1980s and 1990s, when a number of states dropped merit regulation of securities offerings, and the National Securities Market Improvement Act of 1996 preempted state registration requirements for exchanged listed securities. This trend has reversed in the last five years, particularly as former New York attorney general (NYAG) Eliot Spitzer has used the broad authority given him in New York State’s Martin Act of 1921 to pursue allegedly fraudulent activity by equity analysts, mutual funds, and insurance companies.

The NYAG’s activity in the last four years has created competition between state and federal regulators.\(^3\) Whereas some states (e.g., Michigan) have explicitly rejected the suggestion that they investigate securities issues in parallel with the SEC, many others (e.g., California, Colorado, Kansas, Massachusetts, New Hampshire, New Jersey, North Carolina, West Virginia, and Wisconsin) have investigated in parallel or in advance of the SEC. Some have criticized the activities of New York and the other states, and indeed the Securities Fraud Deterrence and Investor Restitution Act of 2003 included language that would have preempted the Martin Act, had it passed (Macey 2005). At the same time, Eliot Spitzer and other state regulators have explicitly cited regulatory capture at the SEC in motivating action by the states.\(^3\)

This revives a long-standing critique of the SEC and the SROs as reflecting the interests of industry, particularly in more aggressive enforcement action against misconduct by rogue individuals (broker fraud, insider trading) than against more systemic forms of misconduct (analyst conflicts, mutual fund compliance issues, earnings management). Those concerned about regulatory capture worry about two sources: top-down and bottom-up. A potential source of top-down is the natural political influ-

---

35. Romano (2001) discusses the potential benefits of competition across regulatory venues that issuers and investors could select (e.g., stock exchanges of different countries). The current competition between the state and federal governments is subtly different in that it involves competition between institutions to regulate the same venue.

ence of so large an industry. The partisan divide of the SEC over several recent regulators’ proposals has also revived interest in the partisan political economy of the SEC (e.g., Zitzewitz 2002). A source of bottom-up capture is the staffing approach of SROs and the SEC. Turnover rates for attorneys, accountants, and compliance examiners at the SEC are more than twice those for comparable-level employees elsewhere in the federal government, including in bank regulation (Securities and Exchange Commission 2002). As Woodward (2001, 100) argues, the “best, and best by a wide margin, post-SEC employment opportunities [are] working for the regulatees.” A “revolving door” staffing model where employees work short tenures in the government and then transition to higher-salaried positions in industry can be successful in attracting talented individuals at a reasonable cost, but has been long regarded as a source of regulatory capture (Stigler 1971).

Following the abovementioned review of the scope of the securities industry, the economic rationale for regulation, and the main laws and institutions, I now turn to three recent issues in the regulation of financial services and markets that have parallels in other industries.

9.5 The Regulation of Pricing

In November 2003 in testimony before multiple congressional committees, Eliot Spitzer called attention to the “$70 billion in management and advisory fees” paid by mutual fund investors in 2002 that “are in addition to significant costs—such as trading costs—that are passed on to investors.” Spitzer cited the difference between advisory fees charged by the same firm to retail and institutional accounts reported by Freeman and Brown (2001), and he cited weak fund governance as the root cause of both the mutual fund share trading scandal and what he regards as an excessive level of fees: “We know that directors and managers breached their duties to investors in every conceivable manner. As regulators and lawmakers, our duty to investors is to investigate every manifestation of that breach and to return to investors any and all fees that were improper or inappropriate. This includes the fees that the managers received during the very time that they were violating their fiduciary duties to investors.” Spitzer proposed disclosure of the precise dollar amount of fees paid to each investor, a strengthening of fund trustee’s

37. Opensecrets.org lists the securities industry as the fourth-largest political donor; it probably also accounts for some portion of the donations of the top industry, the legal profession. The securities industry’s influence is not monolithic, and on many issues different parts of the industry have different interests (see, e.g., the analysis of interest group competition in Kroszner and Stratmann 1998).

38. A recent Government Accountability Office report (GAO-05-385, “Mutual Fund Trading Abuses: SEC Consistently Applied Procedures in Setting Penalties, but Could Strengthen Certain Internal Controls”) found that the SEC did not have a system in place for ensuring that departing staff’s next employer did not present a conflict of interest.


40. Ibid., 3.
fiduciary duties with respect to fees, most favored nations clauses preventing retail mutual funds from charging more than institutional accounts, and competitive bidding for advisory contracts. Lacking the jurisdiction to act on any of these proposals, Spitzer negotiated fee reductions with several mutual fund companies as part of subsequent settlements of share trading allegations.

The mutual fund industry and the SEC were not especially receptive to this line of argument. The SEC did not participate in the fee reduction portion of the mutual fund settlements, even when all other aspects of the settlement negotiations were coordinated. Regarding retail-institutional fee differences, the industry argued (convincingly) that servicing retail clients was more expensive per dollar invested than servicing institutional clients and (perhaps less convincingly) that this accounted for the entire difference in fees charged. Requiring the disclosure of fees paid by individuals was included at one point in a House of Representatives bill, but removed in committee. The SEC did require disclosure of trustees’ rationale for the advisory fees charged, and some expect this to increase pressure from boards for fee reductions.

Any evidence of pressure created by this disclosure for lower fund expenses was slow to emerge from the data. The asset-weighted average expense ratio declined less than two basis points from 2002 to 2006 (table 9.4). While the combined market shares of Vanguard, Fidelity, and American Funds (three large fund families with lower than average expense ratios) increased from 27.3 to 35.2 percent from 2002 to 2006, this was offset by an increase in the average expense ratio charged by other funds in the industry. As discussed earlier, even if some investors became more sensitive to fees during this period and switched assets to lower expense ratio firms, if these were on average the most fee-sensitive clients at their original firms, their departure would have reduced the average fee sensitivity of clients at the other firms, increasing the optimal price. It is also possible that the increased regulatory activity during the 2002 to 2006 time period increased marginal (as opposed to fixed or sunk) costs and that this offset the effects of any greater fee sensitivity. In contrast, average fees did begin to decline after 2006, after the regulatory pressure on price had subsided.

Downward regulatory pressure on price, whether via the direct regulation of prices as in natural monopoly industries such as electricity or cable television or via indirect measures such as those proposed by the NYAG, is generally considered to have several potential side effects. First, if product quality is noncontractable and thus cannot also be regulated, price regulation can lead to lower-than-efficient levels of quality. For example, price regulation may encourage funds to substitute cheaper anonymous managers for more expensive star managers. Massa, Reuter, and Zitzewitz (2010) report that a trend toward anonymous management is already in progress (driven, they argue, by a desire to avoid competition for star managers from the booming hedge fund industry) and that anonymous
teams manage less actively (as proxied by portfolio turnover) and produce slightly lower returns.

Alternatively, fund managers have other means of charging shareholders for their services, outside of the expense ratio. For example, rather than seeking out lower “execution only” commissions, they can place stock trades at brokers who provide benefits to the advisor. Examples of these benefits can be allocations of IPO (Reuter 2006), which are not always allocated to the funds whose trading produced them (Gaspar, Massa, and Matos 2006). Sales support for the advisor’s funds, or “soft dollar” credits that are officially supposed to be used to finance purchases of research, but in practice have been used for office space, periodical subscriptions, computer equipment, and travel expenses. Benefits can also be given by the broker to the advisors’ employees; the recently alleged excessive gift giving by Jefferies Securities to Fidelity employees provides an example.41 Fund advisors can also divert shareholder assets by allowing stale price arbitrage trading in

Table 9.4 Mutual fund industry fees, 2002, 2006, and 2010 ($ millions)

<table>
<thead>
<tr>
<th>Fund family</th>
<th>Expense ratio revenue</th>
<th>Total net assets</th>
<th>Asset-weighted average expense ratio (basis points)</th>
<th>Market share (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fidelity</td>
<td>4,593</td>
<td>658,704</td>
<td>69.7</td>
<td>11.6</td>
</tr>
<tr>
<td>Vanguard</td>
<td>1,413</td>
<td>568,286</td>
<td>24.9</td>
<td>10.0</td>
</tr>
<tr>
<td>American funds</td>
<td>2,570</td>
<td>332,297</td>
<td>77.3</td>
<td>5.8</td>
</tr>
<tr>
<td>Rest of industry</td>
<td>33,598</td>
<td>4,137,096</td>
<td>81.2</td>
<td>72.6</td>
</tr>
<tr>
<td>Total</td>
<td>42,174</td>
<td>5,696,383</td>
<td>74.0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>2006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fidelity</td>
<td>7,657</td>
<td>1,159,840</td>
<td>66.0</td>
<td>12.5</td>
</tr>
<tr>
<td>Vanguard</td>
<td>2,281</td>
<td>1,103,192</td>
<td>20.7</td>
<td>11.9</td>
</tr>
<tr>
<td>American funds</td>
<td>7,346</td>
<td>990,507</td>
<td>74.2</td>
<td>10.7</td>
</tr>
<tr>
<td>Rest of industry</td>
<td>50,371</td>
<td>6,044,359</td>
<td>83.3</td>
<td>65.0</td>
</tr>
<tr>
<td>Total</td>
<td>67,655</td>
<td>9,297,898</td>
<td>72.8</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fidelity</td>
<td>7,588</td>
<td>1,322,583</td>
<td>57.4</td>
<td>12.3</td>
</tr>
<tr>
<td>Vanguard</td>
<td>2,545</td>
<td>1,445,017</td>
<td>17.6</td>
<td>13.4</td>
</tr>
<tr>
<td>American funds</td>
<td>7,581</td>
<td>985,509</td>
<td>76.9</td>
<td>9.1</td>
</tr>
<tr>
<td>Rest of industry</td>
<td>50,662</td>
<td>7,036,011</td>
<td>72.0</td>
<td>65.2</td>
</tr>
<tr>
<td>Total</td>
<td>68,376</td>
<td>10,789,120</td>
<td>63.4</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: Data are from the CRSP Survivor-Bias-Free US Mutual Funds Database. Total net asset figures are end of year. Expense ratio revenue is the expense ratio reported in the CRSP multiplied by total net assets. Variable annuity units and exchange-traded funds are excluded, but both index and actively managed open-ended funds are included.

their funds, by engaging in cross-trades between portfolios at systematically advantageous prices, and by front running personal or favored-portfolio assets ahead of mutual fund trades. Most of these devices are either illegal or at least discouraged by regulators, but nevertheless, at least in principle one might worry that downward regulatory pressure on prices leads advisors to increase their use.

A second consequence of downward price regulation can be shortages. For mutual funds, which have high fixed costs at the firm level but low marginal costs, a “shortage” is most likely to take the form of a reduction in efforts to sell funds to shareholders with small account sizes. For investors who would not find their way to a less-aggressive marketer of (lower cost) funds, but would instead invest in cash equivalents such as bank deposits, this could lead to a welfare loss resulting from a lower than optimal exposure to equity markets.

This is, however, a commonly made argument that is easy to overstate. Expectations of the future equity premium that are derived from current valuations are lower than the historical US return premium commonly cited by industry. For example, Fama and French (2002) estimate a forward-looking equity premium of 2.5 to 4.3 percent as compared with a 7.4 percent premium calculated from historical returns. Suppose that the individual we are concerned with is a canonical mean-variance investor faced with dividing her portfolio between riskless cash/bonds at the risk-free rate and equities with normally distributed returns and a 3.4 percent (pre-expense) expected premium (taking Fama and French’s midpoint). Suppose also that, if she can invest at comparable cost in either asset, she will want to hold an approximately market portfolio of 50 percent equities and 50 percent cash/bonds. It is straightforward to show that an advisor who places this investor in a 50/50 portfolio, but charges her 0.85 percent of her assets annually for the service, leaves her as well off in certainty-equivalent utility terms as if she had invested costlessly in cash on her own.42 By comparison, the asset-weighted average expense ratio for “C class” shares for 2004 from the CRSP Mutual Funds data set is 1.75 percent.43 Given these fees and expectations about the equity premium, it is hard to argue that the typical advisor is offering better certainty-equivalent utility than a bank CD that pays the risk-free rate.

42. The certainty equivalent utility of a mean-variance investor who invests \( s \) of their assets in a risky asset with normally distribution returns and \( 1 - s \) in the risk-free asset is given by
\[
\text{CEU} = w[(1 + f) + sp - s^2rv] - e,
\]
where \( w \) is initial wealth, \( f \) is the risk-free rate, \( p \) is the expected equity premium, \( r \) is the variance of risk asset returns, \( v \) is a risk aversion parameter, and \( e \) is the expense ratio paid. For \( s = 0.5 \) to be optimal, \( p \) must equal \( rv \). Assuming a \( p \) of 3.4 percent, \( s = 0.5 \) and \( e = 0.85 \) percent yields the same utility as \( s = e = 0 \).

43. C class shares are advisor sold, but compensate the advisor using a 12b1 fee that is included in the expense ratio, rather than using a front-end or back-end load. I focus on C shares since calculating the total annual fees paid to both fund manager and advisor does not require an assumption about holding period. Given their asset-weighted average loads and expense ratios of 5.0 percent and 0.93 percent, one would reach a similar conclusion about the overall fee levels of A shares if one assumed a holding period of six years.
A third, related consequence of downward price regulation can be exit. The increase in regulatory scrutiny in the last few years has increased fixed (as well as sunk) costs for mutual fund families and has probably also reduced the use of some of the non-expense-ratio sources of revenue described earlier. Thus one might expect some pressure for consolidation in the long run, but perhaps surprisingly there is not much evidence of this yet. The number of unique management companies offering funds captured by the CRSP data set has declined from 683 in 2000 to a low of 623 in 2004, but half percent of this decline was from 2000 to 2002 and thus was presumably more related to the stock market decline than to increased regulatory pressure.

The welfare costs of fund advisor exit depend crucially on what one assumes about consumer behavior. If we assume that consumers would like to maximize the risk-adjusted returns on their investments but do so imperfectly due to information and cognitive limitations, then we can analyze welfare by examining the implications of exit for shareholder returns. The firms most likely to be induced to exit by downward regulatory pressure on price are small, high-expense ratio firms, and studies of the determinants of fund returns find that these firms produce the lowest returns, even before deducting expenses (e.g., Carhart 1997). This suggests that in the mutual-fund context, regulatory-induced exit can be good for consumers. On the other hand, if consumers are fully rational and have perfect information about ex ante expected returns, then any fund they buy or continue to hold must be welfare maximizing for them. The exit of a fund firm deprives its clients of their first choice and thus, by assumption, must reduce the welfare of these consumers.

Of course, even if one views returns as an adequate proxy for shareholder welfare, one might still have concerns about policies that induce exit and raise the minimum-required scale for entry in an industry. An increase in industry concentration might reduce competitive intensity in the industry, although concentration in this industry is low enough that one might not expect the exit of a small number of high-cost firms to significantly affect behavior.

On the other hand, increased entry barriers might also limit the future entry of innovative firms. The importance of this effect depends on the extent to which one views the industry as mature. Mutual funds appear to be relatively mature. A comparison of the ranking of top mutual fund families in terms of assets in the CRSP Mutual Funds database in 1992 and 2004 suggests that there has been little turnover (table 9.5). Six of the top seven in 1992 were also in the top seven in 2004 (Evergreen has replaced Merrill

44. For example, one reason why a customer might rationally buy high-expense funds with low ex ante expected returns is if the quality of services that are bundled with the fund are high. Collins (2005) argues that differences in service quality explain the price dispersion in index fund expense ratios reported on by Busse, Elton, and Gruber (2004) and Hortascu and Syverson (2005).
Lynch), although the order of families three through seven has changed slightly. Among the top twenty firms in 2004, Schwab and Barclays are the only firms that have moved up the rankings significantly other than through mergers. Hedge funds, in contrast, have experienced extremely rapid growth during this time period.

9.6 Antitrust

The best known financial services antitrust case is undoubtedly the case against the Nasdaq market makers in the mid-1990s. The case was initiated after Christie and Schultz (1994) reported that odd-eighths quotations (i.e., a market offering to trade a stock at 47 1/8 instead of 47 or 47 1/4) were extremely rare for a subset of Nasdaq stocks. After an investigation, the Department of Justice alleged that the avoidance of odd-eighths quotes was collusive behavior designed to increase average market maker spreads.

Several features of market making may have facilitated collusion. First, market makers observe each other’s price quotations; cheating against any collusive arrangement would thus be readily detected. Second, avoiding odd-eighths was a focal arrangement that allowed for a distribution of quantity while minimizing the need for conferring. Avoiding odd-eighths quotations was particularly focal given that the minimum tick size on Nasdaq had only recently been reduced from one-quarter. Third, under preferencing agreements with sources of order flow (e.g., brokerages), many market makers had the right to handle any order flow at the current best bid and ask prices offered by any other market maker (the National Best Bid and Offer, or NBBO). This functioned as a “meet-or-release” clause; so long as the preferred market maker was willing to match, a market maker who undercut the current NBBO could not attract any of the preferenced order flow. This significantly reduced the returns to “cheating” on any collusive arrangement. Fourth, market makers competed in multiple markets, so cheating in one market could be punished in another. As Christie and Schultz (1995) discuss,
an early response to an odd-eighths quotation was often a phone call to the trader’s boss, where such punishments were reportedly explicitly threatened.

In addition, the average retail investor’s understanding of the bid-ask spread component of transaction costs was limited, and many of the institutional investment managers, who presumably did understand bids and ask, had business units that were beneficiaries of any collusion. Furthermore, the rents from collusion were shared through a system known as payment for order flow. In exchange for signing the abovementioned preferencing agreements, sources of order flow (such as brokerages) received per share payments. Table 9.6 shows minimum tick sizes and average gross trading revenue and order flow payments per share for 1995 to 2003 for Knight Securities, the largest publicly traded pure-play market maker. In 1995 to 1996, Knight paid about one-third of its trading revenue for order flow.

As a result of the antitrust enforcement action, odd-eighths avoidance was abandoned, reducing the effective minimum tick size for stocks where there had been collusion. The collusion case also focused attention on the effects of tick size on investors’ transaction costs and further reductions in minimum tick size followed, to 6.25 cents in June 1997 and to 1 cent in early 2001. As predicted by models such as Kandel and Marx (1998) that emphasized minimum tick size as source of market maker rents and payment for order flow, tick size reductions have reduced both market profitability and order flow payments (table 9.6).

Another market in which price transparency and multimarket contact potentially facilitate collusion is in underwriting and syndicated lending. Placing a new issue into the market requires access to a broad network of potential investors, especially since issuers prefer to place it with investors more likely to hold long term. As a result, several investment banks are usually required to manage and market an offering. Underwriting fees are

<table>
<thead>
<tr>
<th>Year</th>
<th>Minimum tick size</th>
<th>Market maker trading revenue</th>
<th>Payment for order flow</th>
<th>Order flow payment share of revenue (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>12.5</td>
<td>1.47</td>
<td>0.55</td>
<td>37</td>
</tr>
<tr>
<td>1996</td>
<td>12.5</td>
<td>1.71</td>
<td>0.65</td>
<td>38</td>
</tr>
<tr>
<td>1997</td>
<td>Reduced from 12.5 to 6.25 in June</td>
<td>1.45</td>
<td>0.37</td>
<td>26</td>
</tr>
<tr>
<td>1998</td>
<td>6.25</td>
<td>1.03</td>
<td>0.21</td>
<td>21</td>
</tr>
<tr>
<td>1999</td>
<td>6.25</td>
<td>1.04</td>
<td>0.17</td>
<td>16</td>
</tr>
<tr>
<td>2000</td>
<td>6.25</td>
<td>1.03</td>
<td>0.16</td>
<td>15</td>
</tr>
<tr>
<td>2001</td>
<td>1</td>
<td>0.32</td>
<td>0.06</td>
<td>19</td>
</tr>
<tr>
<td>2002</td>
<td>1</td>
<td>0.15</td>
<td>0.03</td>
<td>22</td>
</tr>
<tr>
<td>2003</td>
<td>1</td>
<td>0.09</td>
<td>0.01</td>
<td>15</td>
</tr>
</tbody>
</table>

Source: Knight Securities S-1 and 10K statements.
typically a whole-number percentage of the funds raised (e.g., 7 percent for an initial equity offering, 3 percent for high-yield debt). Underwriting business is reportedly extremely profitable for the bank, and competition for it is typically hard fought, but nevertheless discounts from the standard underwriting fees are rare. Any underwriter who secured business through discounting underwriting fees would be unable to do so in secret, since underwriting fees are disclosed in offering documents. The amount of extra business an underwriter could gain through discounting would be limited by the issuer’s desire for wide distribution. And competing banks could punish the discounter, by encouraging clients to exclude the discounter from other syndicates and by encouraging brokerage clients and asset managers (including any asset managers within the same firm) to avoid purchasing an issue whose underwriting business was obtained by discounting.45

The difficulties of discounting underwriting fees lead banks to compete along other dimensions. For example, issuers will demand that banks bundle low-margin products such as revolving credit lines to obtain the higher-margin underwriting business. Alternatively, commercial banks will demand inclusion in investment banking business as a condition of their lending. The latter practice is known as “tying,” and the NASD has argued that it violates the Bank Holding Company Act Amendments of 1970, which prohibit banks from extending credit on the condition that borrowers engage in other business with the bank. Commercial banks have in turn argued that this form of tying is actually procompetitive in that it creates a nonprice means of competing for underwriting business.

Other forms of nonprice competition for underwriting business have allegedly included biases in analyst opinion and even presumably illegal bribes of management. Investment banks have also been accused of biasing their analyst coverage in order to win underwriting business, which would help explain the correlation between analysts’ opinions and their firm’s investment banking business found by Lin and McNichols (1998) and Michaely and Womack (1999). In the “spinning” cases, banks such as Credit Suisse First Boston were accused of allocating shares in underpriced IPOs to executives of firms in order to win their underwriting business.

Another example of collusion on one dimension of price being at least partly undone by competition on other dimensions is the pre-1975 era of fixed commissions. In the Buttonwood Tree Agreement of 1792 that formed the New York Stock Exchange (NYSE), the NYSE members agreed on

45. Some have argued that institutional investors avoided buying Google when it was offered because of their use of a Dutch auction process and a small number of underwriters and their negotiation of a 3 percent underwriting fee. Although Google used a modified Dutch auction that allowed it to price its shares below the market clearing price, creating an incentive for investors to participate in the offering, investment banks may have viewed a successful Dutch auction as a threat, since if it becomes the common mode of offering it would reduce the importance of underwriters’ distribution networks.
minimum commissions: “We the Subscribers, Brokers for the Purchase and Sale of Public Stock, do hereby solemnly promise and pledge ourselves to each other, that we will not buy or sell from this day for any person whatsoever, any kind of Public Stock at a less rate than one-quarter percent Commission.” The NYSE and, after its 1908 founding, the American Stock Exchange maintained fixed commission structures. The 1934 Act gave the SEC oversight of brokerage commissions, but under the guise of self-regulation, the commission allowed the exchanges to exercise their authority over commissions.

Agreements on commissions only applied to trades on the stock exchanges, but the exchanges prohibited their members from off-exchange trading. Nevertheless, “third market” firms developed that specialized in handling off-exchange block trades for institutional investors at discounted commissions. This resulted in undesirable market fragmentation, leading the SEC to first press the exchanges to offer quantity discounts and then, in 1971, to require that commissions on large orders be set competitively (the ceiling was set at $500,000 in April 1971 and lowered to $300,000 in April 1972). The deregulation of large-trade commissions helped motivate a class of small investors to bring a class-action antitrust suit alleging that fixed commissions were price fixing in violation of the Sherman Act. In Gordon v. NYSE (1975), the US Supreme Court ruled that since the 1934 Act had explicitly given the SEC authority to regulate commissions, this superceded the antitrust laws. The decision was quickly made moot, however, by the fact that commissions were deregulated in 1975 by Congress (via the aforementioned Securities Acts Amendments of 1975) and the SEC.

During the era of fixed commissions, brokers engaged in nonprice competition by offering free research. In addition, institutional clients would negotiate “give ups,” where, in lieu of a discount, a portion of their commission would be paid to another broker who in turn provided the investor with free services (such as research or computer services). A group of third-party research firms developed who earned most of their revenue from these give ups. At the time of commission deregulation, these third-party firms feared that investment managers’ fiduciary duties would prevent them from paying commissions large enough to finance “give ups” and that managers would be unwilling to pay for research directly. In response to lobbying by asset managers and third-party research firms, Congress added a safe harbor, allowing asset managers to pay above-market commissions if they determine that the commission was reasonable given the combined brokerage and research services provided. “Give ups” were renamed “soft dollars,” but their economic purpose changed. They were no longer a form of nonprice competition that undermined fixed commissions, but instead became a device for asset mana-

gers to use client assets to purchase research (and other services) through a less transparent means than including its cost in the expense ratio.47

A consequence of the Gordon decision is that the extent to which the Securities Acts preempt the antitrust laws with respect to the securities industry is uncertain and depends crucially on the specific issue at hand. This question is important in part because regulatory capture theory would predict that enforcement of antitrust-related issues by a multi-industry regulator (like the DOJ or FTC) would be more aggressive than by a single-industry regulator (such as the SEC). In Gordon, the court found that Congress had explicitly discussed the stock exchanges’ fixed commission agreements when writing the 1934 Act, and that their decision to give the SEC primary regulatory authority over commissions carried an implied antitrust immunity (Coffee and Seligman 2002, 646). In contrast, in the Nasdaq Market Makers case brought by the Department of Justice, which alleged practices that were not discussed by Congress when delegating authority to the SEC, the courts did not find that the antitrust laws were preempted.

9.7 Conflicts of Interest and Boundaries of Firms

We have turned conflicts of interest into synergies.
—Jack Grubman, former telecom analyst at Citigroup, in an e-mail, as quoted by Eliot Spitzer.

The proceeding discussion highlights some of the advantages for a firm participating in multiple financial services businesses. Many financial products are complements, and integrated providers should have incentives to provide them on more attractive terms for the usual reason (the elimination of double marginalization). There are also no doubt considerable synergies on the production side. Integration may make otherwise collusive markets more competitive, as if there is tacit collusion on one dimension of price, providing related products can increase one’s ability to engage in nonprice competition. For example, a brokerage salesforce and research department give investment banks an advantage in competing for underwriting business, while the deposit base needed to finance lower-margin bank loans does the same for commercial banks.

At the same time, there are reasons for integration that are less benign from a regulatory perspective. Acting as an agent in industry A may create the opportunity to bias one’s actions in order to generate business benefits in industry B, potentially at the expense of the industry A client. For example, asset managers can use their power to vote shareholder proxies as leverage

47. While most discussions of soft dollars find this problematic (e.g., Siggelkow 2004), Horan and Johnsen (2008) argue that the ability of managers to pass on the costs of research in a less-than-transparent manner is beneficial, in that it offsets what would otherwise be an incentive to underinvest in research.
in obtaining underwriting or other business (Davis and Kim 2007). In-house brokers or financial advisors can help sell an asset manager’s funds instead of lower-fee or better-run alternatives. An in-house broker can allow an asset manager to internalize the benefit of commissions for trades done on behalf of their client, perhaps creating an incentive to overtrade a portfolio. In-house proprietary traders may be able to benefit from a brokerage or investment management business, by illegally front-running client portfolio trades, stepping in front of client limit orders, or otherwise exploiting information gained from clients’ trading activities. In-house proprietary trading can also benefit from improved execution quality resulting from the bundling of informed proprietary trading order flow with the presumably less-informed order flow from client’s brokerage accounts or large managed portfolios, at the cost of worse execution for the less-informed orders. Furthermore, when punishing firms that defect against standard industry practices, it is helpful to be able to do so in multiple lines of business.

Most of this second category of synergies also represent conflicts of interest. These conflicts involve the trade-off of one client’s interests for the interests of either another, favored, client or the firm itself. In some cases, this trade-off of interests can be accomplished across firm boundaries through explicit payments. For example, “directed brokerage” was used as a substitute for fund selling by in-house brokers, and soft dollars, especially if used for nonresearch expenses, can be used to allow asset managers to internalize the profits from portfolio trading commissions. But bringing these trade-offs inside firm boundaries is helpful for several reasons. First, it eliminates the need for explicit payments that are potentially subject to regulatory or client scrutiny. Second, common ownership can provide a credible commitment to clients expecting favoritism that a contractual relationship might not. For example, an underwriting client expecting favorable opinions from an analyst is likely to be more assured of getting them if the analyst and the investment banker are employees of the same firm, as opposed to simply having a business relationship. Likewise, clients may invest in hedge funds run side by side with mutual funds because they expect the differences in fee structures to produce favoritism in their favor. Especially if hedge fund investors are more cognizant of the potential for such favoritism than mutual fund investors, firms running funds side by side may realize net marketing advantages.

Ironically, it was precisely these conflicts of interest that motivated the Glass-Steagall Act of 1933, which legally separated banking, securities, and

---

48. Suppose a client submits a limit order to buy at stock at $47.00 or better. A broker can “step in front” of this order by placing a limit order to buy at $47.01. If the broker’s order is filled, the broker has the option to either hold the order and gain any market appreciation or, if demand for the stock weakens, sell to the client at $47.00.

49. For a useful taxonomy of conflicts of interest within and across financial services business lines, see Walter (2004).
insurance. While reversing the 1999 repeal of Glass-Steagall is not being widely contemplated, the trend toward convergence that the repeal reflected has certainly slowed, and perhaps even begun a reversal. In the summer of 2005, Citigroup swapped its asset management business for Legg Mason’s brokerage business. The stated reason for the deal was to eliminate the regulatory risks arising from common ownership of asset management and brokerage. It remains to be seen whether this deal will begin a broader trend.

9.8 A Fourth Issue: Competition with Unobservable Quality

In many industries, competition can exacerbate problems that arise from the imperfect observability of product quality. Imperfect observability provides a rationale for minimum standards in industries as diverse as construction, food, pharmaceuticals, and transportation, particularly for dimensions of quality like safety, where quality affects the probability of rare but very adverse outcomes. Past work has shown that reputational concerns can act as a bond and limit the temptation to lower quality (e.g., Klein and Leffler 1981; Shapiro 1982), but that competition can exacerbate them by decreasing profits, and thus returns to maintaining reputations, as well as by increasing the return to lowering costs (see, e.g., Kranton 2003 on the general issue and Borenstein and Zimmerman 1988 and Rose 1990 on airline safety).

For example, in banking, depositors can observe the interest rate offered by a bank, but not the risk of losses due to bank failure. This problem is addressed by regulations mandating that banks provide deposit insurance, as well as by regulations limiting the riskiness of bank’s investments. While banking is discussed in chapter 8, an analogous problem affects the investments that are the subject of this chapter. Investors can readily observe the past returns of an investment, as well as its realized past risk, as captured in variance of past returns. But investors often cannot observe unrealized risks. If an investment earns a higher return by accepting exposure to low-probability events, then investors may observe the returns without observing the risk, and competition for returns may exacerbate pressures to take such risks.

Two examples that arose in the financial crisis are credit and liquidity risk. During the middle of the decade, AAA-rated structured debt securities offered higher yields than Treasuries or AAA-rated corporate bonds. As we all know now, and as many knew at the time, flawed assumptions about levels and correlations of default risks made by rating agencies allowed tranches of pools of risky loans, such as residential subprime mortgages, to be rated AAA even when they carried substantial default risk (e.g., Coval, Jurek, and Stafford 2009). Because AAA-rated structured yields were often above borrowing costs faced by investment vehicles, it was possible to increase current yields further using leverage.
One example of a mutual fund that did so was the Oppenheimer Core Bond Fund, which was included as the conservative option in many state “529” college savings plans. The fund had very low variance returns in its twenty-year history prior to 2008, and held largely highly rated bonds. It clearly was judged as low risk by the states that included it as an investment option. Yet it lost almost 50 percent of its value during the financial crisis. The fund charged higher than average fees for its category, thus to maintain an attractive yield, it needed exposure to higher yielding credits, which it added using total return swaps on mortgage-backed securities. Competition on an observable dimension (yield net of fees) may have exacerbated the temptation to reduce quality on an unobservable dimension. The SEC later disciplined Oppenheimer for not adequately disclosing this added risk, although the $35 million settlement was less than 2 percent of investor losses, which totaled approximately $2.5 billion.50

Even if holdings disclosure requirements are complied with, they may not adequately inform investors about risk. One issue with holdings disclosures is that they are periodic, giving managers the opportunity to “window dress” their portfolios around disclosure dates. For example, Morey and O’Neal (2006) find that bond mutual funds’ exposure to credit risk, as measured using correlations of fund returns with bond indices, decreases around portfolio disclosure dates, suggesting that funds shift out of risky bonds in order to reduce their apparent exposure to credit risk.

For hedge funds, which are not required to disclose holdings other than long positions in equities, the primary means of inferring risk is from the variance of monthly returns. Particularly in asset classes that trade infrequently or with wide bid-ask spreads, opportunities exist to smooth returns (Goetzmann et al. 2002; Getmansky, Lo, and Makarov 2004; Bollen and Pool 2009).51 Smoothing of returns may not only mislead investors about past realized risk, but can also create incentives for redemption after a market decline, as investors who anticipate smoothing will expect managers to overvalue assets temporarily after a market decline. This has obvious averse implications for systemic risk. The refinement of policies affecting risk disclosures and portfolio valuation is likely in the years to come.

50. See “Oppenheimer Funds to Pay $35 Million to Settle SEC Charges for Misleading Statements during Financial Crisis” (available at http://www.sec.gov/news/press/2012/2012-110.htm). Investor loses of $2.5 billion are from June 2008 to March 2009, and are measured by multiplying beginning of month assets by monthly returns. Both the settlement and the investor loss figures also cover a second fund (Oppenheimer Champion Income Fund), which was accused of similar practices. Oppenheimer also paid $100 million to settle a class action lawsuit over the same issues (see https://www.oppenheimercoresettlement.com/ and https://www.oppenheimerchampionsettlement.com/).

51. While papers on hedge funds are limited to analysis of return time series, Cici, Gibson, and Merrick (2011) show more directly that bond mutual funds smooth returns by switching between valuing bonds using bid and bid-ask midpoints.
9.9 Conclusion

Financial regulation has been basically reactive in the last decade. Sarbanes-Oxley, Dodd-Frank, and many of the significant SEC rules have been adopted in response to revelations of specific abuses, such as accounting fraud, mutual fund late trading, selective disclosure, insider trading, and market maker collusion. Even the most noteworthy deregulation, the gradual relaxation and finally repeal of the Glass-Steagall Act, was partly a response to a series of mergers between the industries the act was designed to keep separate. Given the increasing emphasis on compliance in most financial services firms in the last few years, the rate of revelation of new scandals is likely to slow. This should create the opportunity to think more proactively about what financial regulation should be attempting to accomplish.

Could thinking proactively in 2005 have yielded a less severe financial crisis? Some have characterized the financial crisis as an unexpectedly toxic combination of known problems. For example, in their dissent to the Financial Crisis Inquiry Report, Hennessey, Holtz-Eakin, and Thomas (2011), while generally more skeptical about regulation than the authors of the majority report, highlight the role played by a lack of transparency about holdings and risk exposure. They also note that while “credit rating agencies erroneously rated mortgage-backed securities and derivatives as safe investments, . . . buyers failed to look behind the credit ratings and do their own due diligence” (418). They noted that these buyers were “in theory, sophisticated investors” (426), but they left unsaid the fact that these buyers were almost entirely managers of third-party assets, and thus that agency problems might be behind the lack of due diligence.

Compulsion of due diligence by regulators is impractical, of course. The larger question is whether a lack of due diligence arises from competition over past returns that create temptations for exposures to risk. Competition over past returns is natural among active managers, while for passive managers competition is equally naturally more over fees and other costs. As discussed in sections 9.2 and 9.3, many investors pay a financial planner to sell them a mutual fund or annuity, pay the fund manager management and administration fees, and pay commissions and transactions costs for active management that is, on average, both aggressive and unsuccessful in generating positive risk-adjusted returns. Perhaps the largest and most controversial outstanding question about financial regulation is whether this represents an efficient market outcome or a market failure, and, if the latter, whether regulation should do more to correct that failure.

If one decides that it should, the next question would be how: how to change laws to correct existing market failures without creating new ones, and how to reform institutions so that they reinforce rather than undermine this goal. The first question is nontrivial. Disclosure about fees and conflicts
of interest appears ineffective in influencing the behavior of many investors. On fees, the strengthening of mutual fund boards’ fiduciary responsibilities to aggressively negotiate on investors’ behalf also does not appear to have led to a significant reduction, at least in the short term. This raises the question of whether more direct regulation of price levels is desirable, either through outright price limits or through the strengthening of suitability requirements for broker recommendations. This hinges in large part on whether it could be implemented without the side effects that accompany it in other contexts.

But the more difficult question is arguably the institutional one. Both the approach of self-regulatory delegation and the staffing model for the SEC lead these institutions to reflect the interests of the industries they regulate. These interests may be well aligned with the public interest in disciplining the behavior of rogue individuals, but are likely to be much less so in correcting systemic market failures that are also sources of economic rents. Both the contrast with Eliot Spitzer over the last few years and the aggressive prosecution of the Nasdaq price fixing case by the DOJ in the 1990s suggest that multi-industry regulators might be less prone to capture. The SEC is currently organized around the industries it regulates, and while this specialization is no doubt useful for building industry expertise, a more generalist-oriented staffing model, in which staff develop expertise that creates future employment opportunities in multiple parts of the securities industry, may reduce at least some of the forces contributing to capture.

A second large and controversial question is whether regulation should continue to encourage, or instead discourage or attempt to reverse, convergence. Many financial services are complements in both their production and consumption, and convergence should allow for many genuine synergies: in production, product innovation, the reduction of search costs via one-stop shopping, and the potential elimination of double marginalization. At the same time, the presence of agency relationships in most services means that convergence may frustrate the policing of conflicts of interest by bringing them inside firm boundaries. Is it optimal to locate in the same firm the underwriters of securities and the third-party managers charged with deciding whether to invest in them? Do the problems associated with convergence outweigh the benefits? Research enumerating and economically sizing them would be especially helpful in answering this question.

As the length and recentness the bibliography that follows demonstrates, the postbubble and postfinancial crisis years have seen the popular interest in refining financial regulation matched by academic interest. Like policy, academic research is often reactive, exemplified by the many papers that usually follow a major policy change such as SOX, Regulation Fair Disclosure, and Dodd-Frank. By helping policymakers understand the economics of the securities business, including the nature of competition and the incentives faced by firms and agents, however, academic research can help policymakers prospectively identify changes that would lead to better outcomes.
Competition, incentives, and the effects of regulation are central issues in industrial organization (IO). While research in and debate about securities regulation is often dominated by specialists, given the centrality of what are essentially IO issues, the generalist readers of this volume are likely to also have a contribution to make.

References


