

# The Challenges of Regulating Finance<sup>1</sup>

Jonathan Glover (Columbia University) and Shyam Sunder (Yale University)

## Abstract

When financial regulators attempt to constrain the ability of regulatees to attain their goals, regulatees organize and use the political system to capture the regulators or financially engineer and outmaneuver new regulations. The deep pockets of the financial service industry raises their ability to execute such capture. Compared with other industries, the regulation of financial services presents more difficult challenges since the essence of finance lies in information and symbols, not material objects. Bank capital, assets, liabilities, and the cost of borrowing, for example, have no fixed meaning outside the words used to specify them, and the meanings of words in natural languages are necessarily ambiguous. Regulatees easily win the arms race with regulators by altering transactions, reinterpreting regulations, adding new contingencies unforeseen by regulators, and greater complexity, as we see in the attempts to regulate banking and financial reporting. Evidence to punish financial crimes is difficult to compile, and prosecutors tend to take the easy way out by negotiating deferred prosecution and shareholder fines instead of jury trials of corporations' executives. Perhaps, we should investigate fostering a better financial culture by rewarding good behavior in banking and finance instead of relying merely on punishing bad behavior. Balancing the roles of rules, standards, principles, and norms of the industry with regulatory discretion may also help to yield better outcomes.

Preliminary Draft, August 11, 2017 (not for circulation)

Please send comments to Jonathan Glover ([jg3463@columbia.edu](mailto:jg3463@columbia.edu)) and Shyam Sunder ([shyam.sunder@yale.edu](mailto:shyam.sunder@yale.edu)).

---

<sup>1</sup> Prepared for presentation at International Conference, "Public Authority and Finance: Deregulation and New Regulation," Friday 1st and Saturday 2nd September 2017, University of Paris 1 Panthéon-Sorbonne, Paris.

*“Iron law of corporate physics: For every regulation there is a proportionate proliferation of innovations to circumvent it.”*

*Harold James (2015)*

*“All of finance is essentially a gentlemen’s agreement, and ... the agreement matters more than what the agreement is about.”*

*Pascal-Emmanuel Gobry*

### **Political Economy of Regulation**

Regulating economic activities has been proposed as a solution to remedy the inefficiency and distributional consequences of market failures. Stigler (1971) observed that regulators, such as the Interstate Commerce Commission (ICC), tend to serve the interests of the regulatees (the railroad companies in the case of ICC) instead of their customers to remedy the consequences of the regulatees’ monopoly power. Stigler explained this phenomena by the ability of a small compact group with sufficiently large per capita interest to pool their resources and capture the political process behind regulation to serve their own interests. This happens even if a larger group of consumers lose a much larger total amount. The reason, he argued, is that the per capita loss to individual consumers is too small for them to incur the cost of organizing themselves to effectively pressure the political system to act in the interest of the general public. Although initially proposed to advance public interest, regulation ends up becoming a form of tax paid by public to favor the regulatees (Posner 1971).

The primary targets of regulation in finance concern capital, information and contracts. A bank that takes deposits and lends them out multiple times, creating private money, cannot meet the demands of depositors if too many of them demand their money at the same time. Banking regulation is supposed to ensure that the banks keep at least a minimum percent of their deposits “idle” to maintain some margin of safety against a bank run. An insurance company accepts premiums in exchange for its promise to pay any claims that may arise from the insurance contract in the future, and insurance regulation is supposed to ensure that the premiums remain invested and available for any claims. Disclosure requirements for effective interest rates on bank and credit card loans are an example of information regulation.

Political economy analyses of financial regulation also point to a tendency toward regulatory capture in financial services. The large and growing size and resources controlled by financial service firms enable them to spend, individually and collectively, massive resources to persuade and pressure legislators as well as regulators at the federal and state levels to act in concordance with their interests. Although almost the entire population of the country often stands on the losing side of such regulatory actions, the per capita losses imposed are too small for the consumers to organize themselves in opposition. At the time of this writing in mid-2017, it is hardly surprising that the post 2008 crisis regulations created by the Dodd-Frank Act in 2009 and the Volker Rule are already well in the process of being unraveled through the political process in the United States (see Partnoy 2017; Jopson 2017).

This political manipulation/regulation is well documented so in this paper, we focus our attention on other challenges to financial regulation that arise from the nature of financial services and innovation. We explore the broad nature of accounting and finance; virtual finance; the arms race between financial regulators and financial engineers; the regulation of financial reporting as an example; future challenges in regulating finance; an exploration of using regulation to foster desirable bank, corporate and industry culture; and balancing the roles of rules, standards, principles, and norms of the industry with regulatory discretion.

## The Nature of Accounting and Finance

That financial reporting is largely a matter of convention is reflected by the phrase “generally accepted accounting principles (GAAP)” to describe the governing regime of financial reporting over the past century. Sunder (2016, Chapter 5) examines the social norms and ritual aspects of financial reporting.

In contrast to accounting, finance literature is dominated by the mathematics of financial economics which obscures the social norms that undergird finance. According to Gobry (2017), “all of finance is essentially a “gentlemen’s agreement” (that’s what a vanilla loan is) and that the agreement matters more than what the agreement is about.” Fiat money is the ultimate bubble—it is valued only because it is valued; it has no other basis for its value. The same is true of any security or financial asset—its value arises entirely from one’s expectation about the future and has no verifiable basis. Gobry attributes to Robert Merton these words, claimed to have been said jokingly: “it wasn’t that everyone used (his) options pricing model because it was accurate, but that it was accurate because everyone used it.” This makes drawing inference from “empirical tests” of widely taught theories quite difficult.

In this sense, a belief in a model can be self-fulfilling, and LIBOR (London Interbank Overnight Rate) worked for more than 30 years as the foundation of world-wide lending networks of trillions of dollars because millions of people believed in it. In fact, it had little factual basis because the banks submitting quotes did not always know what their future borrowing costs would be, and/or because the banks manipulated and lied about the quotes even when they knew their costs. Levine (2017) writes:

“Still, yes, Gobry is right that finance is a set of social conventions. And much of the financial industry is a stylized arena for battling over those conventions. One thing that I like to think about Libor in the heyday of its manipulation is that it was supposed to reflect a real market, driven by supply and demand, for unsecured short-term interbank lending -- and it didn't -- but it *did* reflect a real market, driven by supply and demand, for interest-rate derivatives. Some banks would profit from high Libors and so submitted high Libors, other banks would profit from low Libors and so submitted low Libors, and ultimately you got a Libor that crudely reflected the market -- but the market *for Libor, from derivatives traders*, rather than the market *for short-term funding, from banks*. It didn't represent the thing it was supposed to reflect, but it *did* reflect its actual deep purpose, in a useful and market-driven way.”

This amounts to saying that a measurement is to be assessed not on the basis of what it is purported to represent, but on the basis of its outcomes when it is used as an input for business decisions under the prevailing beliefs.<sup>2</sup> To the extent that the essence of finance lies in beliefs and expectations, finance has always been virtual, even before technology advances in computers and communications. The question is: How can this industry in the modern computer age be successfully regulated?

## Virtual<sup>3</sup> Banking and Finance

Modern computer and communications technologies have shifted significant aspects of our lives from the physical domain to the virtual domain in retail banking - lending and financing, payments and transfers, wealth and asset management, markets and exchanges, insurance, and blockchain transactions. Finance has also been influenced by this phenomenon. This is easy to see in money,

---

<sup>2</sup> Ironically, this a stout defense of historical cost accounting over the mark-to-market accounting coming from finance.

<sup>3</sup> Virtual often refers to objects or actions which might have had physical form at some time but have since become disembodied, usually in electronic form, to take an invisible and intangible form.

transactions, accounting and legal records, trading, as well as employment. A few examples may help to illustrate this point further.

In its early form, money consisted of tangible objects which had readily perceptible value in their own right. Metals (gold, silver, and copper), grains (rice and barley), minerals (salt and stones), intoxicants (cigarettes, alcohol and cannabis), and sea shells and cowries are among such examples. With the exception of gold, shells, and stone, most commodity money deteriorated with time, as it was bulky and inconvenient. Metal coins first appeared some half a millennium B.C. in China, India and Lydia though the bulk of early metal coins appeared to suggest that their value may have come mostly from their metal. Gradually, the importance of the metal content of coins declined relative to their nominal value and this created representative or fiat currency. Paper money was almost entirely representative, but still tangible. Government printing presses produced paper money which was transported in bags to banks for distribution. Credit and debit cards led the way to virtualization of money and the paper form began to decline. Most people today see only a small fraction of their income in the form of coins or paper; it is simply transferred among the bank accounts of employers, employees, and merchants at the appropriate time with proper authorization. In recent years, crypto-currencies have taken this process even further through to untraceability.

Like banking transactions, financial transactions also have moved from the tangible to intangible. Securities have changed from printed paper to computerized book entry. Instead of mailing the paper securities to and from brokers, trading now involves little more than entries in the electronic books of various organizations executed almost instantaneously over the phone or computers connected through communication networks. Details of the transaction, e.g., its exact time of execution, counterparties, and the prices available in various markets at the time of execution can be tightly controlled by those who operate various parts of the system (creating some additional regulatory challenges about control and distribution of this information). Checks which could take days or weeks to clear have been replaced by bank wire transfers in many parts of the world. Credit and debit cards, ATM cards, PayPal and ApplePay are other examples of new transactions technologies which expand well beyond financial transactions.

Accounting and transaction records take an intangible form now, as compared to wood, clay, papyrus, bark, parchment, thread, or paper earlier in history. The effort of physically writing in a medium appears to increase the longevity of the medium. Electronic records, easier to create, are also easier to erase. Since most computers can keep backups as well as a complete record of all steps carried out on them, erasure does not necessarily eliminate access to the erased information. Absent backups and process logs, information can be lost irretrievably. In 2001, it took days for Enron and its auditors at Arthur Andersen & Co. (both accused of committing and aiding fraud at Enron) to shred its records, and the amount of time and people involved in the shredding made the event known (Glater and Brick 2002; Leeds 2002). Erasure of electronic records is easier and faster, and not so difficult to conceal (see <http://www.hawaii.edu/askus/706> for example on secure deletion of electronic records). The blockchain technology of decentralized encrypted recordkeeping on multiple computers, initially employed for crypto-currencies, offers new, computer-intensive prospects for the future, but it is too recent for us to know its potential. The same could be said for legal records.

Securities and commodities trading in exchanges and over-the-counter is an important part of the financial sector. Until a few decades ago, this trading took place in face-to-face transactions in exchange floors or over telephone with slips of paper as primary documents used to support record keeping. Today, most trading has moved to a virtual domain so buyers, sellers, and various intermediaries (market makers and brokers, etc.) can sit at their office desks and trade through computer screens and keyboards, or even let artificial algorithmic agents do the trading on their behalf. There are also the so-called “dark pools”—private exchanges in which anonymous institutional investors trade with one another with little transparency, and mostly outside the domain of regulation. Trading

systems include not only buy and sell orders, executions and market information, but also are closely intertwined with many other kinds of information that includes financial and general news, feeds from other markets, financial data about the securities and commodities being traded, and market statistics and indexes. A good part of stock market regulation concerns trade execution and access to various kinds of market and non-market information. Tracking information flow in an electronic world is difficult without access to the system software. The financial services workforce has become diffuse and decentralized. Much of their information flows on public and private networks and these networks are not accessible to regulators. Who has what information, and what was communicated among these people can be made undetectable to others. The increasing role of artificial agents in finance increases these regulation difficulties.

The finance world has changed and is now largely pervaded by rapid innovation. What are the consequences of this innovation, and how can regulation control it if it proves to be harmful?

### **The Arms Race Between Regulators and Financial Engineers**

In the context of physical sciences, engineering, medicine, and the industries that depend on physical products, innovation has a decidedly positive connotation and is generally presumed to be a social welfare-improving phenomenon. The same cannot be said of a great many financial innovations. New financial innovations emerge in response to market demand. These include not only products designed to make markets more complete, efficient, and informationally symmetric, but also products designed to circumvent regulation.

Regulations place constraints on choices that regulatees can make. Some regulations are meant to improve coordination (e.g., requiring all cars to drive on either the left or the right side of the road), while others are intended to improve the quality of products or services (e.g., disclosure of net interest yields on bank loans to their customers). Both kinds of regulatory constraints seek to achieve some broad social goals. The cost, inconvenience, and reduction in their opportunity sets leads many regulatees to see regulations as burdensome and unwelcome impositions. Not surprisingly, they seek ways of preserving their profitability (and achievement of any other goals) by adjusting their businesses and transactions to circumvent the regulations (Exhibit 1).

The speed of innovation is rapid, while regulation is inherently much slower. Often the regulator applies rules to an existing game while the actual game that will be played when the regulation is in place is drastically different than anticipated. With the increasing ease of developing parallel (shadow) markets not governed by the existing regulations, even the nature of the players can be difficult (if not impossible) to anticipate. Bitcoin presents a recent example. It was designed for anonymity which also facilitates illegal commerce; even the identity of Bitcoin's developer is unknown. Unlike the black markets (e.g, alcohol sold in speakeasies during the 1920s and early 1930s when the National Prohibition Act was in place in the U.S. or currency exchanges in the streets of countries with government controlled exchange rates), virtual currencies can be and are designed for invisibility. Faced with the daunting task of regulating such currencies and other nearly invisible products, regulators have essentially three choices. First, they can go the prohibition era route and try to ban such currencies (with a more vigorous enforcement than Prohibition), perhaps with regulators trading in bitcoins in the hopes of identifying the counterparties to those transactions. Prohibition received only mixed public support. Would an aggressive ban of crypto-currency transactions face a similar lack of public support? Second, regulators could try to foster more visible competition from banks. Third, the regulators could require all such transactions to be conducted through government-run exchanges.

Financial regulations set in motion a dynamic process that tends to weaken, if not nullify, their intended effects. Most regulations employ classification as a device to partition the attribute space of the organizations, instruments or transactions. The Glass-Steagall Act of 1933, for example, classified banks into categories of commercial and investment. Securities are routinely classified into categories of

equity, fixed income, treasury bills, repurchase agreements and derivatives, etc. Borrowing, lending, bonding, investing, are examples of the categorization of financial transactions. Categorization has four important, often unintended consequences: alteration, interpretation, addition of new dimensions and additional complexity.

### 1. Alteration

The attribute spaces subjected to regulatory categorization are mostly continuous. Even when they are not strictly continuous, categories often have boundaries which are not clear and therefore subject to interpretation or manipulation. When the attribute space is continuous, categorization of the space requires insertion of one or more thresholds at selected point(s). Every such threshold creates a discontinuity in the relevant dimension of the space. For example, if the regulation specifies that banks with less than X dollars of assets do not have to comply with a requirement and the ones with more than X have to comply, such regulations insert a discontinuity in the asset dimension of the regulatory space. And it has consequences.

If asset space is continuous and the amount of assets held is controllable by the bank, the regulation creates incentives for the bank to determine which side of the regulatory threshold it lies in its current state, and whether it might be advantageous for it to move over to the other side. This decision involves balancing the cost of compliance with the regulation on each side of the threshold and the cost of changing its assets (in this example) in order to shift to the other side of the threshold. If the latter cost is continuous in the magnitude of change, for any finite compliance costs, there will always be a band, immediately adjacent on the “disadvantageous” side of the threshold, from which banks will prefer to move to the other side. This is the essence of the Dye-Glover-Sunder (2015) theorem, illustrated schematically in Exhibit 2.

In Exhibit 2, the x-axis depicts the dimension on which classification is carried out (amount of bank assets in this case), and the y-axis depicts the cost-benefit to the bank. The red dashed line depicts the gross benefit to the bank of being classified as being on the right side of the asset threshold, relative to being on the left side. If the assets of the bank exceed the threshold level, there would be no motivation for the bank to try to change its assets. However, if its assets are less than the threshold level (as shown in the exhibit, the bank will consider changing its assets in light of the regulation. The red dotted line depicts the gross cost of changing its asset level. If before issuance of the regulation, the bank has its assets at the optimal level, changing its assets from the optimal will involve incurring these gross costs (for simplicity and without loss of generality, assumed to be symmetric on either side of the status quo). The thick black line depicts the net costs to the bank (relative to the status quo) of choosing any given level of assets (which is equal to the difference between the gross benefit of being on the right side of the regulatory threshold less the gross costs of adjusting its assets). It is easily seen that the ideal decision of the bank depicted in the exhibit would be to increase its asset level just above the regulatory threshold; by keeping assets below the threshold the bank would leave money on the table, and by exceeding the threshold it would incur unnecessary adjustment costs.

In the United States, the U.S. Congress limits the amount of money the federal government can borrow. In recent decades, whenever the projected borrowing needs of the government exceeded the legal limit, a protracted battle ensued with those who oppose authorizing additional borrowing, or wish to exchange their vote for other concessions. A recent proposal for legally side-stepping the need for congressional approval through financial engineering has been described as follows (Levine 2015): The borrowing constraint enacted by Congress applies to the face value of the debt. The government can replace each \$100 face value of maturing debt by new debt of the same face value, but at a considerably higher coupon rate. Since the market pays a higher price for the higher coupon rate debt, the

government can borrow as much additional money as it wants without seeking new authorization from Congress.<sup>4</sup> This same idea has long been used by desperate corporate CEOs to massage their earnings by recalling low (relative to current market yield) coupon debt at a price discounted below its face value, and reporting the difference as a gain on the early retirement of debt.

One consequence of this alteration strategy shows up in aggregate data which gets bunched up on one (advantageous) side of the regulatory threshold, leaving a hole on the other side. This phenomenon can be seen in Burgstahler and Dichev's (1997) analysis of the changes in net income (scaled by market value) reported for about 64,000 firm-years by U.S. non-financial corporations during 1977-94 (Exhibit 3). Zero change in earnings is a cognitive, not a regulatory threshold - did the earnings increase or decrease - and it is easy to see that far too many observations are bunched up immediately to the right of the zero leaving too few observations just below the zero. This alteration in reported earnings of corporations is a response to an anticipation of how investors might react to lower earnings. Although the earnings are supposed to have been reported according to the generally accepted accounting principles, managers retain discretion within any regulatory structure to make enough non-trivial alterations to reduce the consequences of regulatory constraints that they consider undesirable.

## 2. Interpretation

A second set of regulation consequences arises from the interpretations of regulations. It is not possible to write perfectly clear rules any more than it is possible to write complete contracts. The need for interpretations arises from at least two sources: contingencies and meaning ambiguity.

As Weinberg (1992) points out, it is not possible to provide a complete description of even a piece of chalk, much less for a far more complex financial transaction and instrument. Laws, regulations and contracts are labeled as "incomplete" when they do not ensure that all relevant contingencies yield a unique conclusion.<sup>5</sup> But the space of all contingencies in any natural or social environment is unbounded. It is not possible to imagine a complete set of rules or contracts any more than it is possible to specify the finest detail in a fractal space (because it is always possible to sub-divide them to reveal additional detail). The problem before the regulator then is not how to write a complete set of regulations, but how to choose the contingencies that are included in, or excluded from, the regulation, and the extent of detail included in the chosen contingencies. In this sense, all regulations are intermediate solutions on the incompleteness spectrum. Yet, regulations in accounting and finance generate demand for further clarification and get gradually expanded from a few pages to hundreds, even thousands. Basel I, II, and III and financial reporting standards in the U.S. as well as internationally are good examples of the process and consequences of pursuing evermore detailed contingencies in regulation.

---

<sup>4</sup> Miniting a trillion dollar platinum coin is another financial engineering solution, actively debated since 1992 to address the debt limit. See Balkin (2011).

<sup>5</sup> For example, "The murdering-heir case is one where the law is incompletely detailed." Ehrlich and Posner (1974, p. 260).

Ambiguity in the meaning of words seems to be an essential property of natural language (as opposed to an artificial language such as Python). If a common noun or verb were made precise enough, it could apply only to a specific instance of an object or action, because no two objects (e.g., shirts) or actions (e.g., drinking) are exactly alike. The *Oxford English Dictionary* includes a million or so examples of word usage from its collection of several million words to give the readers a sense of the range of meanings associated with each word. Written regulations, dependent on words' meanings, must necessarily offer a plethora of meanings around them, creating ample room for interpretation and variations in practice.

### 3. Addition of new dimensions

In addition to alteration and interpretation, the most powerful tool of a financial engineer is the addition of new dimensions and contingencies to contracts, instruments and transactions. Even if the regulator finds some way of eliminating room for alteration and ambiguity, no regulator has the capacity to either prevent and anticipate the additional dimensions engineers might add in response to regulation. The vast proliferation of derivative securities over the recent decades has been driven, in no small degree, by the attempts of financial engineers to bypass the regulatory classifications by adding new dimensions as attributes of the securities. Such new dimensions allow financial engineers to claim that the classification scheme of prevailing regulations does not apply to contracts/instruments in the expanded domain, and therefore they are free to treat them in a manner more favorable to them.

### 4. Complexity

Large firms are complex, and we cannot rule out the possibility that this is intentional. In such organizations, it is often difficult to assess regulatory compliance, and to determine responsibility for a decision on an identifiable individual or group. In U.S. law, conviction requires establishing criminal intent beyond reasonable doubt. Modern corporations have been designed to make this prosecutorial task almost impossible. The Sarbanes-Oxley Act (2002) requires CEOs and CFOs to personally certify the accuracy of their financial report. Major restatements in financial reports have hardly decreased since this Act, and yet few senior executives have been punished for violating the law. Subsequent fines have become just a cost of doing business with little associated guilt or social stigma (Soltes 2014).

As Dye et al. (2015) point out, this re-engineering strategy sets up an arms race between regulators and regulatees. Since the latter have fewer constraints and significant monetary rewards linked to bypassing the regulations, they move much faster than the regulators (who are bound by laws and due process requirements). Not surprisingly, the regulators lose.

### **Regulation of Financial Reporting**

Discontinuities created by regulatory classification thresholds facilitate design of financial products aimed at engineering around the regulations. From what he calls his "Bad Bob" years as a financial engineer, Herz (2013, p. 14) gives the following examples of financial engineering meant to undermine (and facilitated by) accounting regulations:

“arrangements designed to boost reported earnings by triggering gain recognition on appreciated assets carried at historical cost basis while retaining the underlying risk and rewards of those assets;

financing techniques involving the issuance of debt-like securities that were treated as equity under the accounting rules;

financings involving equity securities and hybrid securities that were accorded favorable treatment in computing earnings per share;



transactions designed to obtain off-balance sheet treatment, for example under lease accounting and rules applying to special purpose entities;

techniques designed to minimize the dilutive effects of M&A transactions through the use of pooling of interests accounting and other structures that qualified under the accounting rules as common control mergers, joint ventures, and partial combinations; and

transactions to arbitrage the lack of discounting of future cash flows in accounting for insurance loss reserves and deferred tax assets and liabilities.”

Herz argues that the U.S. Financial Accounting Standards Board (FASB), which he chaired, has made progress towards capturing the underlying economics of transactions in accounts and, thus, ameliorated incentives for manipulation. A common theme in most of the examples given above is the introduction of discontinuities<sup>6</sup> in the mapping between the underlying economic attributes of a transaction and its accounting treatment.

The cost of pursuing the underlying economics of transactions, including continuity of accounting treatment, often takes the form of a sacrifice in the verifiability of measurements. By focusing on a few transaction attributes and threshold values, accounting has traditionally preserved the verifiability of measurements. Reduced verifiability provides more opportunities to manipulate measurement and, perhaps more importantly, greater information asymmetry about verifiability and manipulability between financial statement preparers and users (Glover and Levine, 2017). Moreover, transactions, being endogenous, can be designed to be difficult to verify.

Over its history, the FASB’s approach to dealing with financial engineering has primarily been one of writing more detailed rules, eliminating existing attempts to get around the rules in what can be seen as an attempt to move the financial reporting “contract” towards greater completeness. To sketch a model, in an initial period, the standard-setter chooses a contract or rule that maps the transaction attribute space into accounting reports. If endowed with perfect foresight, the standard-setter could write a complete contract and choose the minimal attribute space that will include all future transactions. However, standard-setters have far from perfect foresight. So, financial engineers with clients who prefer an alternative treatment of the transactions redesign them by adding new attributes to transactions that were not anticipated by the standard-setter. Under the existing rule designed for the unmodified transactions, the new transactions are projected into the original transactions space. By the financial engineer’s intent, the resultant accounting reports fail to capture the economic substance of the new transactions. The standard-setter is free to revise the rules or issue “implementation guidance” to deal with the newly invented attributes. With its slow due process, the standard-setter is left struggling to complete the contract with respect to the newly-defined expanded attribute space, with ever more complexity in transactions and the rules or regulations for those complex transactions. Financial reporting for lease contracts and failed regulatory attempts to get lease obligations reported on corporate balance sheets is a good example (Dye et al. 2015).

In the race between scientists and illness, scientists develop new antibiotics and often nature develops immunity to the new antibiotics, making it wise to use antibiotics sparingly and limit any unnecessary exposure of the viruses to the antibiotics. However, when antibiotics are used, they are best used in a sufficiently strong dose to leave no surviving virus to adapt and become immune to the treatment. Financial engineers too seem to thrive on advance knowledge of new regulations from regulators to better devise strategies to defeat them.

---

<sup>6</sup> That is, small changes in the attributes of a transaction should not generate large changes in accounting treatment of transactions, such as on vs. off the balance sheet treatment, or the use of present vs. exit values.

### **Challenges ahead in regulating finance?**

Besides the domestic political power of large financial service firms, their international power is an additional factor in successfully resisting regulation. Keefe (2017) reported that HSBC, a major international British bank, after being caught aiding Mexican drug lords, rogue states, and terrorists for many years, got away with a slap on the wrist. There is zero chance that the top management of HSBC did not direct these activities, and yet, when the U.S. regulators threatened to prosecute HSBC for its repeated flouting of regulations, the “British Chancellor of Exchequer warned U.S. authorities of serious implications for financial and economic stability.” If this threat was real, HSBC was simply another example of being “too big to fail” and “too big to jail.” On the other hand, one cannot rule out the possibility that this was simply an example of British domestic politics preventing its allies in the U.S. government from prosecuting HSBC and its executives for their crimes.

The problem of enforcing regulations in financial services is common and goes well beyond HSBC. After the U.S. savings and loan industry crisis in the 1980s, hundreds of bank executives, including many CEOs were found guilty and went to jail. After the Atlantic Financial Crisis, the Financial Crisis Inquiry Commission asked the U.S. Department of Justice to prosecute dozens of the largest financial services firms. Yet there have been few indictments, trials, or jailings for these major financial crimes, by far the largest in the history of the U.S. Fines (paid by shareholders) and deferred prosecution arrangements are all that the public sees, in many cases repeatedly for the same banks. U.S. laws permit the whole company to be held responsible for the actions of its employee (e.g., Arthur Andersen & Co.), but that does not seem to happen any more. What has changed? After the guilty verdict (subsequently reversed on appeal) on Arthur Andersen & Co. shut it down, the number of large international accounting firms was reduced to four. In spite of repeated major violations, the government could not afford to put another of the large accounting firms out of business because it would reduce the number to just three, and insulate them even further from competition. In the case of banks, the former U.S. Attorney General Eric Holder (2013) “acknowledged that decades of deregulation and mergers had left the U.S. economy heavily consolidated. It was therefore difficult to prosecute the major banks, because indictments could have a negative impact on the national economy, perhaps even the world economy.” This reads like a certificate of immunity against prosecution from the top law enforcement official of the U.S. After all, his former cabinet colleagues Rubin and Summers ran one of the largest US banks, Citicorp.

In addition to such hints of immunity from the top, the U.S. Department of Justice has not supported any inclination the prosecutors may have to take cases against large banks and their CEOs to jury trial. Prosecutorial incentives to jail the CEO of a large company and to win plaudits seem to have changed (Eisinger 2017; Kwak 2017). Prosecutors know that when they put senior executives in the dock, they must battle considerable financial resources of the giant corporations and the battalions of lawyers hired to defend them at shareholder expense. Fines and deferred prosecution agreements seem like a safer route to success for the prosecutors. This is what they seem to be doing since the Financial Crisis, reducing if not eliminating the enforcement power of financial regulations the banks are supposed to comply with.

How are we to deal with these difficult challenges? Let’s explore two options: emphasizing culture and social norms and combining written rules with significant regulatory discretion.

### **Banking and Finance Culture**

Instead of trying to change a bank’s culture after its misconduct is discovered, perhaps more could be done to reward banks with a good existing culture. What is a “good banking culture”? If much of finance is essentially a gentlemen’s agreement (an implicit promises to serve the interests of customers with integrity), a good banking culture would be, at least in part, one in which bankers honor their gentlemen’s agreements.

In terms of the large literature in economics on relational contracts (e.g., Baker, Gibbons, and Murphy 2002; Levin 2003), the loss to the banker from renegeing on a promise and facing a self-enforcing punishment in the subsequent periods must be greater than any short-term gain from renegeing. The most severe punishment is typically the equilibrium of the single-play game — in this context, the clients stop trusting the bankers.

What is the role of regulation in fostering a good banking culture? Regulators seem to have focused on two aspects: capital requirements and deferred bonus payments. But future rents—the prospect of sufficient profitability so bankers want to preserve their franchise, instead of taking what money they can now, and running—is also important.

Conservatism has been the focus of much research in accounting (e.g., Basu 1997 and the subsequent literature), but the focus is typically on analyzing conservatism within a single-period setting. Yet, conservatism is essentially a multi-period phenomenon. An important (and we think largely ignored in research) aspect of conservatism is its ability to facilitate long-run relationships, including gentlemen’s agreements at the heart of finance.

Both capital requirements and deferred bonus payments are related to but different than what accountants call conservatism, sometimes used interchangeably with prudence. Sunder (1997) notes “[t]he presence of uncertainty and the downward bias of measured current-period income, assets, and owner’s equity in the presence of uncertainty seem to be the essential aspects of conservatism.” Accounting conservatism has been the focus of much research in accounting (e.g., Basu 1997 and the large literature that followed him), but the focus is typically analyzing conservatism within a single-period setting. Yet, conservatism is essentially a multi-period phenomenon. An important (and we think largely ignored in research) aspect of conservatism is its ability to facilitate long-run relationships.

It isn’t clear which comes first—conservatism (sometimes used interchangeably with prudence) or a long-term focus. If bankers find it sufficiently profitable to focus on the long-term, they may find it in their own best interest to be conservative (e.g., in their financial reporting choices, bonus plans, and dividend payout policies). Making conservative measurement and other related choices, when they lock-in those choices, can serve to decrease the attractiveness of short-term over long-term actions for bankers. For example, under conservative accounting, poor current performance may be the result of bias, increasing the appeal of sticking around for the inevitable reversal of that bias and the related payouts.

It is important to see regulation focused on culture as fundamentally different from regulation focused on directly preventing misdeeds. This point can be made even with a single-period model of regulation. (We will come back to the importance of long-term relationships shortly.) Let us turn again to the specific case of regulating financial reporting. One way to state the financial engineering problem facing regulators is: how do we transition from a financial engineering culture to a culture of communication and integrity? Preparers may feel pressed to engage in accounting-motivated financial engineering if everybody else does it, and they do not wish to be at a competitive disadvantage. Viewed as a problem of social norms analyzed by Hume (1739), there are multiple equilibria, and preparers are playing a bad (from the regulator’s and society’s point of view) reporting equilibrium. Under this view, the regulator is expected to upset an undesirable equilibrium rather than create a desirable one.

Some SEC practices target, or potentially target, a bad financial reporting culture. One example is the whistleblower program whereby an employee can report illegal behavior without fear of reprisal. A less obvious example is the Office of Chief Accountant’s (OCA’s) “pre-clearance” process, whereby preparers seek approval for the accounting treatment of a proposed transaction. Pre-clearance has been used to target bad reporting innovations before they become industry norms. Unfortunately, the Final Report of the Advisory Committee on Improvements to Financial Reporting (2007) criticized this useful aspect of the pre-clearance process; instead it advocated that the SEC staff emphasize that its pre-clearance process is registrant-specific—a simplistic view of the SEC’s role.

Another example is the SEC Division of Enforcement’s “wild-catting.” Under Stephen Coutler’s direction in the mid-2000s, the Division of Enforcement began targeting entire industries for investigation, even without evidence of a problem beyond some isolated cases. When the investigations occurs sequentially, with a finding of bad reporting at one firm triggering investigations of other firms, wild-catting has desirable incentive properties in upsetting unwanted equilibria. The fines for firms subject to later investigations are implicitly dependent on the earlier findings at other firms. If the investigations are simultaneous, a desirable penalty structure is a team-based one: each firm’s penalty is highest when other firms are also found to be using the bad reporting practice. These (modified) forms of wild-catting turn the power of investigation up if the reporting norms are bad, and turn the power down when the reporting norms are good. (See Baldenius, Glover, and Xue (2015) for a related analysis in the context of bonus pools.) To illustrate the point, consider the following game.

Player 1\Player 2	A	B
A	a, a	c, d
B	d, c	b, b

(The first entry in each cell is row Player 1’s Payoff)

The regulator can investigate each player to discover that player’s true action with probability  $p > 1/2$ . Consider two possible penalty approaches. First, the regulator can treat the players independently and impose a penalty of  $P^I$  on a player when he is found to have chosen action B instead of A. Second, she can treat the players as a team and penalize both players with a penalty of  $P^T$  if and only if the investigation indicates that both players chose B. Suppose that penalties are socially wasteful, so the regulator’s goal is to motivate (A, A), while imposing the smallest expected penalty. Each player’s net payoff is the amount shown in the table above, minus his expected penalty.

Suppose the regulator’s problem is to make (A, A) a Nash equilibrium of the game. That is, without investigation, (A, A) is not a Nash equilibrium:  $d > a$ . Consider individual penalties. To make (A,A) a Nash equilibrium, the penalty  $P^I$  must satisfy:  $a - (1 - p)P^I > d - pP^I$ . The smallest penalty that satisfies this constraint is  $(d - a)/(2p - 1)$ , and the expected penalty (summed across both players) is  $2(d - a)/[p/(1 - p) - 1]$ . If instead the regulator uses a team penalty  $P^T$ , the penalty must satisfy:  $a - (1 - p)^2 P^T > d - p(1 - p)P^T$ . The smallest such penalty is  $(d - a)/[(1 - p)(2p - 1)]$ , and the expected penalty is  $2(1 - p)^2(d - a)/[(1 - p)(2p - 1)] = 2(d - a)/[p/(1 - p) - 1]$ , which is the same as under individual penalties. Intuitively, in motivating Player  $i$ , the investigation into Player  $j$  ( $j$  not  $i$ ) is not informative in the sense of Holmstrom (1979) and, hence, is not useful. However, because of the payers’ risk-neutrality, the team penalty does not increase the expected cost. The team-based penalty is simply a rescaling of the individual-based penalty, and this rescaling is not costly if the players are risk neutral.

Now, suppose that (A,A) is an equilibrium but agents find another equilibrium, (B,B), more appealing:  $a > b$ . The principal now uses investigations and penalties to ensure (B,B) is not an equilibrium. To ensure (B,B) is not a Nash equilibrium, the penalty  $P^I$  must satisfy:  $c - (1 - p)P^I > b - pP^I$ . The smallest penalty that satisfies this constraint is  $(b - c)/(2p - 1)$ , and the expected penalty (summed across both players) is  $2(b - c)/[p/(1 - p) - 1]$ . If instead the regulator uses a team penalty  $P^T$ , the penalty must satisfy:  $c - p(1 - p)P^T > d - p^2P^T$ . The smallest such penalty is  $(b - c)/(2p^2 - p)$ , and the expected penalty is  $P^T = 2(1 - p)(b - c)/[p^2/(1 - p) - p] = [(1 - p)/p]P^I$ . Since  $p > 1/2$ , the expected team-based penalty is smaller.

Intuitively, when the principal is trying to eliminate a bad equilibrium, team incentives allow the regulator to “turn up the power incentives” at the bad equilibrium (B,B), while “turning down the power incentives” at the good equilibrium (A,A). Put in more technical terms, using a team penalty creates a situation in which the size of the penalty is determined by the actions (B,B), which generates penalties with probability  $p^2$ , and (A,B), which generates penalties with probability  $p(1 - p)$ . The penalty is paid out with probability determined by (A,A), which is  $(1 - p)^2$ . Hence, the usual likelihood ratio approach cannot be used to determine the cost, since three different probabilities are involved.

Holmstrom (1979) is about exogenous correlation in states. What we identify here could be termed *equilibrium selection informativeness*. Player  $j$ 's play of B and, hence, the regulator's imperfect investigation into  $j$ 's actions, is informative about player  $i$ 's play in that it provides some indication that the players have selected the (B,B) equilibrium. Importantly, equilibrium selection informativeness, and the reason to tie the players together using team incentives, does not arise if the regulator's goal is to create rather than upset an equilibrium. Holmstrom's condition has nothing to do with equilibrium selection.

While it may seem a stretch to use team-based fines (based on the behavior of multiple firms rather than a single firm), they are implicitly employed in practice. Sequential investigations of preparers are one example. Another is that the SEC sometimes (more often in the past) lets firms off the hook with a warning (a CFO letter) to change their reporting practice when many firms in the same industry have all adopted a practice the regulator objects to. If the SEC is interested in targeting bad norms, then they should instead be tougher on firms when problems are widespread.

A more ambitious role for team-based regulatory incentives is to view the regulators as setting the stage for regulatees to police one another through mutual monitoring and long-term relationships (e.g., “tit-for-tat” play). When economic agents observe each other's actions, it can be optimal to tie them together with joint rewards and penalties that motivate them to mutually monitor each other and punish each other for bad behavior (Arya, Fellingham, and Glover, 1997; Che and Yoo, 2001). The mutual monitoring approach turns multiple equilibria into something to be fostered rather than eliminated, since multiple equilibria are a source of implicit contracting among the agents. Put in terms of relational contracts, the role of such team-based regulatory incentives is to create a situation in which each firm in an industry would rather forego the gains from short-term bad behavior (i.e., forego free-riding on the good behavior of other firms) when that bad behavior would bring retaliation from other firms in the industry (other firms in the industry adopting bad behavior in future periods). The idea can also be applied to financial engineers and regulatees, creating a joint accountability (e.g., requiring financial engineers to provide the SEC with a customer list when the SEC uncovers practices they wish to target). This has the potential to make preparers weary of financial engineers who specialize in products designed to get around a written rule.

The potential benefits of introducing team incentives are large and could overturn much of the current thinking in financial regulation. For example, long-term relationships between regulatees and financial engineers would be encouraged, as would long-term relationships between regulatees and auditors (e.g., no auditor rotation). Instead of fostering collusion (the usual regulatory concern), repeated interaction fosters mutual monitoring.

Under the team incentives approach, a combination of regulatory and self-policing behavior would motivate preparers and financial engineers to behave. A broader point is that various mechanisms (managerial compensation practice, corporate governance, standards, auditing, the SEC's various roles, the courts, mutual monitoring, etc.) interact with each other. To design one without understanding the others will almost surely be dysfunctional. For example, detailed guidance by the rule makers does not help regulators promote an environment of good norms. Guidance narrows room for regulatee discretion, reducing the frequency of future self-governed interactions. This less ambitious role for standard setters and regulators in setting the stage for preparers to mutually monitor each other also

seems to have the advantage of creating fewer opportunities for regulatory capture by constituents. See Revsine (1991) for an excellent discussion on regulatory capture in accounting as part of a broader discussion of what he terms “the selective misrepresentation hypothesis.”

Returning again to the FASB, the term Generally Accepted Accounting Principles, used now to refer largely to the written pronouncements of the FASB in the U.S., seems to suggest something else. Indeed, the term was developed during a time when best practices—social norms of the community—rather than written rules guided reporting. How were such norms upheld before the SEC was created? The answer seems to be through mutual monitoring. To the extent that written rules and standards interfere with mutual monitoring, this is a cost of relying on written regulations.

### **Rules, Standards, Principles, Norms and Discretion**

These terms are used in different fields and literatures with different connotations, and it is not possible to assign to them any “correct” meanings and definitions. Yet, these terms have added much to our understanding of social phenomena and it would be unwise to discard them for the lack of variability of, and unanimity on, their meaning. The best we can do is to discuss these terms and put their connotations in relation to one another in the context of financial regulation.

Most people will agree that rules lie near the specificity end of the specificity-generality spectrum (Ehrlich and Posner 1974). They are written down, and try to cover specific contingencies, e.g., Ex. Law 22: If anyone is committing a robbery and is caught, then he shall be put to death (Code of Hamurabi, 1754 BCE). Specificity itself is a relative term, and one could push further in that direction by demanding a definition of a robbery, e.g., necessary and sufficient conditions for an act to be categorized as a “robbery”, and the event as “being caught.” A rule which appears to be specific when it is formulated, may be revealed to be lacking in specificity, usually when an accused points to (or constructs) an ambiguity. There is no such thing as a perfectly clear rule; interested parties can always point to ambiguities in any rule. What if the presumed “robber” claims that he was only repossessing some property that he thought had been stolen from him earlier? One’s thoughts are private and unobservable. Crime is based on intent and once ambiguity is pointed out, it is difficult for the rule maker to refuse to clarify them. Those who are accused of violating the rules have every reason to discover and point to such ambiguities in their defense. Consequently, in most domains, rules follow a dynamic growth process and the rulebooks rarely get thinner with the passage of time. Every Congressional attempt to “simplify” the Internal Revenue Code in the U.S. has only added to its volume (Bram 2017) The same is true of banking and accounting regulations.

Norms lie near the general end of the rules-norms spectrum. They are rarely written down, and are communicated through a slow acculturation process. Norm violations are identified by judgment—often self-judgment—and have little formal sanctions beyond disapproval or fear of disapproval and internal sanctions.

Standards and principles lie somewhere between rules and norms. Different literatures characterize them in different ways, Ehrlich and Posner (1974, p. 259). To meet the challenge of finance regulation, a new emphasis on culture and social norms of the industry is necessary. Implementation of such a regime requires empowering and entrusting regulators with a significant level of discretionary power; Otherwise, they are guaranteed to lose to financial engineers.

It’s worth pointing out that in the past, banking companies included “trust” as a part of their name. Trust is an essential part of a gentleman’s agreement by definition, and implies a broader relationship with the bank that extends beyond the transaction at hand. Indeed, the label “transactional” connotes a short term focus and *caveat emptor*. In recent decades, “trust” seems to have been removed from the names of most if not all financial service companies. Perhaps, this is no accident.

## References

- Baker, G., Gibbons, R. and Murphy, K.J., 2002. Relational Contracts and the Theory of the Firm. *The Quarterly Journal of Economics*, 117(1), pp.39-84.
- Balkin, Jack M. (July 28, 2011). "Three Ways Obama Could Bypass Congress. *CNN*. (accessed October 3, 2015  
<http://www.cnn.com/2011/OPINION/07/28/balkin.obama.options/index.html>).
- Barlin, Rick. 2017. "Regulation on the Rise as Bitcoin Gains Popularity, *CPA Journal* (July 2017) ([http://www.cpajournal.com/2017/06/30/regulation-rise-bitcoin-gains-popularity/?utm\\_source=RealMagnet&utm\\_medium=Email&utm\\_content=Shyam.Sunder@Yale.edu&utm\\_campaign=7%2E26%2E17%20Trending%20CPAJ%20Stories](http://www.cpajournal.com/2017/06/30/regulation-rise-bitcoin-gains-popularity/?utm_source=RealMagnet&utm_medium=Email&utm_content=Shyam.Sunder@Yale.edu&utm_campaign=7%2E26%2E17%20Trending%20CPAJ%20Stories) accessed July 27, 2017).
- Basu, S., 1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of accounting and economics*, 24(1), pp.3-37.
- Bram, Thursday. 2017. "Why the Complex U.S. Tax Code Won't be Simplified," *Investopedia* (<http://www.investopedia.com/financial-edge/0512/why-the-complex-u.s.-tax-code-wont-be-simplified.aspx> accessed August 11, 2017).
- Burgstahler, David, and Ilia Dichev. 1997. "Earnings Management to Avoid Earnings Decreases and Losses," *Journal of Accounting and Economics*, Vol. 24 No. 1 (December): 99-126.
- Dye, Ronald A., Jonathan Glover, and Shyam Sunder. "Financial Engineering and the Arms Race Between Accounting Standard Setters and Preparers." *Accounting Horizons*, Vol.29, No. 2, 2015, pp. 265-295.
- Eisinger, Jesse. 2017. *The Chickenshit Club: Why the Justice Department Fails to Prosecute Executives*. New York: Simon & Schuster.
- Glater, Jonathan D. and Michael Brick. 2002. "Enron's Collapse: The Overview; Ex-Official Says EnronEmployees shredded Papers," *The New York Times*, January 22, 2002.
- Glover, J. 2013. "Can financial accounting regulators and standard setters get (and stay) ahead of the financial engineers?" Emanuel Saxe Lecture at Baruch College.
- Glover, Jonathan, and Carolyn Levine. 2017. "Information Asymmetries about Measurement Quality," Columbia University Working Paper.
- Gobry, Pascal-Emmanuel (@pegobry). Thread of 36 tweets beginning with "I'm surprised and disappointed that @matt\_levine missed the moral of the Libor story in today's newsletter..." 7:05 AM -7:32 AM, July 27, 2017..
- Jopson, Barney. 2017. "US Regulator Moves to Loosen Volker Rule," *The Financial Times*, August 2, 2017 (<https://www.ft.com/content/86e0ff36-7702-11e7-90c0-90a9d1bc9691> accessed August 2, 2017).
- Keefe, Patrick Radden. (2017). "Why Corrupt Bankers Avoid Jail?" *The New Yorker* (July 31, 2017).
- Kwak, James. 2017. "America's Top Prosecutors Used to Go After Top Executives. What Changed?" *The New York Times*, July 5, 2017.

- Leeds, Jeff. 2002. "Anderson Auditor Details Shredding of Enron Papers," *The Los Angeles Times*. May 14, 2002.
- Levin, J., 2003. Relational incentive contracts. *The American Economic Review*, 93(3), pp.835-857.
- Levine, Matt. 2015. "Federal Debt," in *Bloomberg View: Money Stuff*, July 28, 2017,
- Levine, Matt. 2017. "LIBOR and Society," in *Bloomberg View: Money Stuff*, July 28, 2017, (<https://www.bloomberg.com/view/articles/2017-07-28/libor-phone-smashing-and-hacking> accessed July 28, 2017).
- Partnoy, Frank. 2017. "The Sequel to the Global Financial Crisis is Here," *The Financial Times*, July 31, 2017 (<https://www.ft.com/content/95808118-662e-11e7-9a66-93fb352ba1fe> accessed August 2, 2017)
- Penno, M. 2008. "Rules and accounting: Vagueness in conceptual frameworks." *Accounting Horizons* 22: 339-352.
- Posner, Richard A. 1971. "Taxation by Regulation," *Bell Journal of Economics and Management Science*, Vol. 2, pp. 22-50.
- Sunder, Shyam. 2016. "Rethinking Financial Reporting: Standards, Norms and Institutions", *Foundations and Trends® in Accounting: Vol. 11: No. 1–2*, pp 1-118.
- Stigler, George J. 1971. "The Theory of Economic Regulation," *The Bell Journal of Economics and Management Science*, Vol. 2 No. 1 (Spring 1971), pp. 3-21.
- Weinberg, S. 1992. *Dreams of a Final Theory*. Vintage Books.



Exhibit 1: Cartoon by P.C. Vey, *The New Yorker*, March 9, 2009.



*"These new regulations will fundamentally change the way we get around them."*

Exhibit 2: Threshold Discontinuity Created by Regulation (Source: Dye et al. 2015)

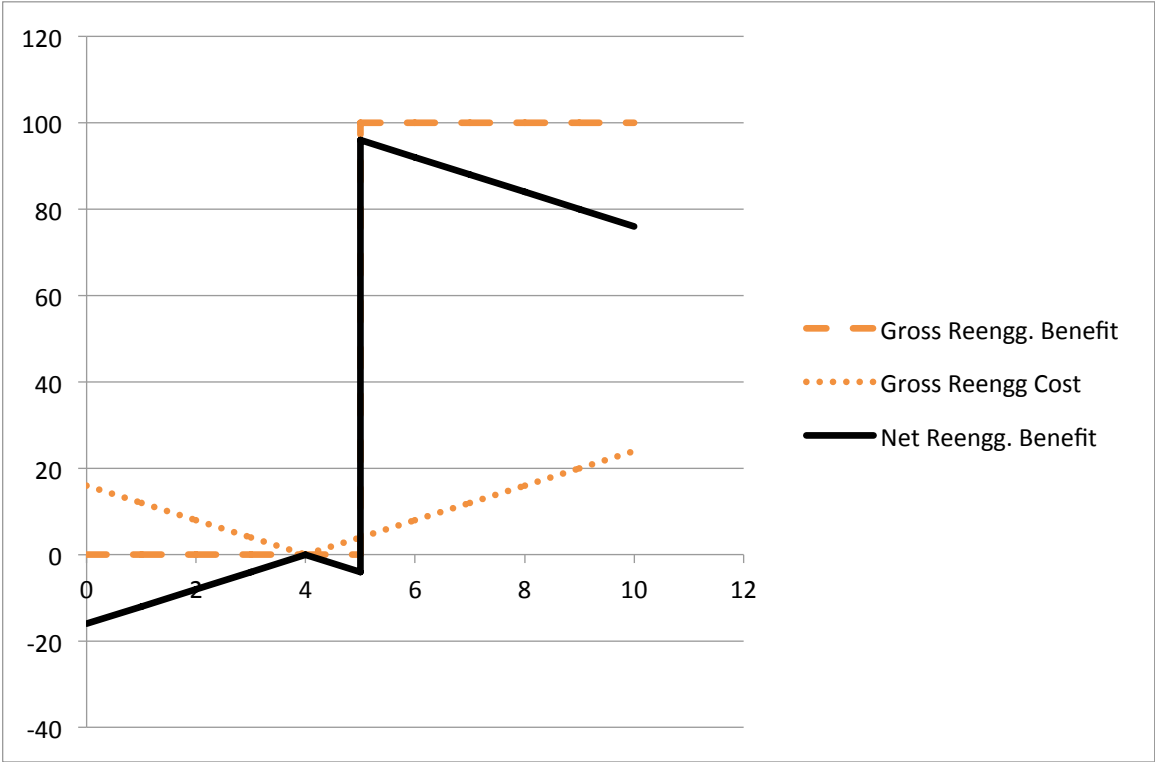
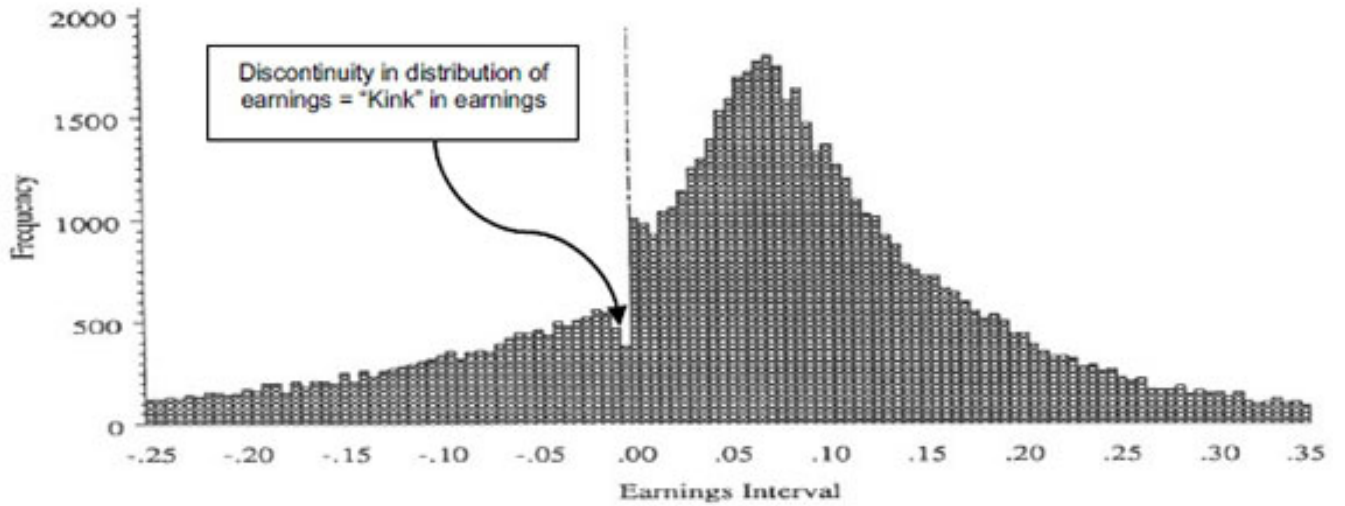


Exhibit 3: Effect of Regulation Thresholds on Population Distribution

**Figure 1**  
The distribution of earnings of US non-financial companies. Earnings: Annual net income scaled by market value of equity at the beginning of the year,  $Earnings_t / MV_{t-1}$



Source: Graph taken from Burgstahler and Dichev (1997b)